

Supporting Information for “Unified Model for Photophysical and Electro-Optical Properties of Green Fluorescent Proteins”

Chi-Yun Lin^{1†}, Matthew G. Romei^{1†}, Luke M. Oltrogge^{1,2}, Irimpan I. Mathews³,
and Steven G. Boxer^{1*}

¹Department of Chemistry, Stanford University, Stanford, California 94305, United States.

²Current address: Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, United States.

³Stanford Synchrotron Radiation Lightsource, 2575 Sand Hill Road, Menlo Park, California 94025, United States.

*Correspondence to: sboxer@stanford.edu

†These authors contributed equally to this work.

S1	Sample Preparation	S3
	<i>Plasmid Construction</i>	S3
	<i>GFP Constructs in This Study</i>	S4
	<i>Dronpa Constructs in This Study</i>	S5
	<i>DNA Sequences</i>	S6
	<i>Amino Acid Sequences</i>	S8
	<i>Synthetic Peptide Design</i>	S9
	<i>Protein Expression and Purification</i>	S10
	<i>Semisynthetic Method for Split GFPs</i>	S10
	<i>Chromophore Analogue Synthesis</i>	S14
	<i>Buffers for Spectroscopic Studies</i>	S18
	<i>Sample Preparation for Stark Spectroscopy</i>	S18
S2	Spectroscopic Methods	S20
	<i>Room-Temperature UV–Vis Absorption and Fluorescence Measurements</i>	S20
	<i>Low-Temperature (77 K) Absorption Measurements and Electronic Stark Spectroscopy</i> ...	S20
	<i>Stark Spectroscopy Data Analysis</i>	S23
	<i>Extinction Coefficient Determination of GFP Mutants</i>	S24
S3	X-ray Crystallography	S25
	<i>Protein Crystallization</i>	S25
	<i>X-ray Data Collection and Structure Determination</i>	S26

S4	The Marcus–Hush Model and Its Implications for Spectroscopic Observables ...	S29
S5	Rebane’s Quadratic Stark Model and Stark Spectroscopy	S37
S6	Stark Spectroscopy and Local Field Factors	S41
S7	Supercharged GFPs: Color and Chromophore pK_a	S44
S8	Further Discussion of Electrostatic Contribution to Excitation Energies and Driving Forces.....	S49
S9	Justification of the Marcus–Hush Model from Resonating Valence Bond Theory	S57
S10	Equivalence of the Su–Schrieffer–Heeger Model and the Marcus–Hush Model ...	S63
S11	Stark Spectra and Fitting of GFP Mutants and Chromophore Analogues.....	S65
	<i>Mutants and Variants of GFP and Dronpa2: Classical Stark Analysis</i>	<i>S65</i>
	<i>GFP Model Chromophores: Violation of Classical Stark Analysis.....</i>	<i>S80</i>
S12	Discussion on Protein Structures	S86
S13	Supplementary Figures.....	S94
S14	Supplementary Tables	S109
S15	References.....	S123

S1 Sample Preparation

Plasmid Construction

The gene sequences of s10:loop:GFP [1], ih:loop:GFP [2], ih:GFP [3], and Dronpa2 [4] are the same as those used in our previous publications. Point mutations were made using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's protocol. The *Aequorea victoria* GFP (avGFP) gene mirrored Prasher's *gfp10* cDNA [5] with the known innocuous mutation Q80R [6]. Supercharged GFP genes in pET-29 were generously provided by the David Liu Lab at Harvard University [7]. All but the supercharged GFP genes were inserted into the pET-15b vector (Novagen) at the NdeI and XhoI restriction sites.

Circular permutants with proteolytic loops were used to facilitate the reconstitution of split GFPs with desired mutations on the synthetic strands or the internal helix (including the chromophore). Circular permutants without loops (ih:GFP) were chosen based on availability. As noted previously [1][8] and implied from the absorption spectra (Figure S4) along with crystal structures (Figure S16), no appreciable differences were detected in spectral features among circular permutants aside from the chromophore pK_a changes. For ih:loop:GFP, the sacrificial proteolytic loop was mutated to a thrombin cleavage site (LVPRGS) because chromophores in R96 mutants matured very slowly, and the trypsin treatment used previously caused complete digestion, presumably due to the lack of structural integrity [9]. C48S and C70A mutations were introduced to eliminate the possibility of disulfide linkage formation during the denaturation step, but the latter caused a significant decrease in expression yield, so cysteine was restored for R96 mutants. V206K in ih:loop:GFP was originally introduced to mitigate aggregation for truncated GFPs [10] and did not result in any appreciable spectral change as expected from the supercharged GFP control experiment (Section S7, Figure S4). Supercharged GFP +36 with negatively charged s10 (+36 s10⁻) was designed to break the symmetry of uniformly supercharged GFPs from Professor David Liu (Figure 3C and Table S8). The replaced nucleotides and amino acids are shown in bold below.

GFP Constructs in This Study

We adopted the nomenclature devised for split GFP circular permutants in our previous works [2]. 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. 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 (see Figure 3B). To facilitate readability, the mutation carried by the synthetic strand is enclosed by parentheses rather than superscripted as in our previous publications. See Figure 1A for the relative positions of critical mutation sites with respect to the chromophore.

GFP Constructs	ih		s4	s7	s10	s11
	65	66	96	148	203	222
avGFP	S	Y	R	H	T	E
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 ₃ Y	E
s10:loop:GFP E222Q	S	Y	R	H	T	Q
s10:loop:GFP T203V E222Q	S	Y	R	H	V	Q
s10:loop:GFP T203Y E222Q	S	Y	R	H	Y	Q
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
supercharged +36	T	Y	R	H	T	E
supercharged +36 s10-	T	Y	R	H	T	E
ih:GFP S65T Y66(3-NO ₂ Y)	T	3-NO ₂ Y	R	H	T	E
ih:GFP S65T Y66(2,3,5-F ₃ Y)	T	2,3,5-F ₃ Y	R	H	T	E
ih:GFP S65T Y66(2,3-F ₂ Y)	T	2,3-F ₂ Y	R	H	T	E

ih:GFP S65T Y66(3,5-F ₂ Y)	T	3,5-F ₂ Y	R	H	T	E
ih:GFP S65T Y66(3-F ₁ Y)	T	3-F ₁ Y	R	H	T	E
ih:GFP S65T Y66(3-Cl ₁ Y)	T	3-Cl ₁ Y	R	H	T	E
ih:GFP S65T Y66(3,5-Cl ₂ Y)	T	3,5-Cl ₂ Y	R	H	T	E
ih:GFP S65T Y66(3-Br ₁ Y)	T	3-Br ₁ Y	R	H	T	E
ih:GFP S65T Y66(3-I ₁ Y)	T	3-I ₁ Y	R	H	T	E
ih:GFP S65T Y66(3-CH ₃ Y)	T	3-CH ₃ Y	R	H	T	E
ih:GFP S65T Y66(3-OMeY)	T	3-OCH ₃ Y	R	H	T	E
ih ^{mat} (65T) · ih:loop:GFP R96M	T	Y	M	H	T	E
ih:GFP S65T H148D	T	Y	R	D	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 ₁ F)) · s10:loop:GFP S65T	T	Y	R	H	4-F ₁ F	E
s10(203F ₅ F) · s10:loop:GFP S65T	T	Y	R	H	F ₅ F	E
s10(203(4-NH ₂ F)) · s10:loop:GFP S65T	T	Y	R	H	4-NH ₂ F	E
ih:GFP S65T T203(3-OMeY)	T	Y	R	H	3-OCH ₃ Y	E
s10:loop:GFP S65T T203Y	T	Y	R	H	Y	E
ih:GFP S65T E222Q	T	Y	R	H	T	Q
ih ^{mat} (65T) · ih:loop:GFP R96M T203Y	T	Y	M	H	Y	E
ih ^{mat} (65T) · ih:loop:GFP R96E E222K	T	Y	E	H	T	K

Dronpa Constructs in This Study

Table S2. Dronpa constructs in this study. Red indicates non-wild-type amino acids. See Figure S21 for the relative positions of critical mutation sites with respect to the chromophore.

Dronpa Constructs	ih		s7	s8	
	62	63	142	157	159
Dronpa	C	Y	S	V	M
Dronpa2 (M159T)	C	Y	S	V	T
Dronpa2 T159E	C	Y	S	V	E
Dronpa2 T159Q	C	Y	S	V	Q
Dronpa2 Y63(3-NO ₂ Y)	C	3-NO ₂ Y	S	V	T
Dronpa2 Y63(2,3,5-F ₃ Y)	C	2,3,5-F ₃ Y	S	V	T
Dronpa2 Y63(2,3-F ₂ Y)	C	2,3-F ₂ Y	S	V	T
Dronpa2 Y63(3,5-F ₂ Y)	C	3,5-F ₂ Y	S	V	T
Dronpa2 Y63(3-F ₁ Y)	C	3-F ₁ Y	S	V	T
Dronpa2 Y63(3-Cl ₁ Y)	C	3-Cl ₁ Y	S	V	T
Dronpa2 Y63(3-Br ₁ Y)	C	3-Br ₁ Y	S	V	T
Dronpa2 Y63(3-I ₁ Y)	C	3-I ₁ Y	S	V	T
Dronpa2 Y63(3-CH ₃ Y)	C	3-CH ₃ Y	S	V	T
Dronpa2 Y63(3-OMeY)	C	3-OCH ₃ Y	S	V	T
Dronpa2 S142A	C	Y	A	V	T

Dronpa2 V157D	C	Y	S	D	T
Dronpa2 V157N	C	Y	S	N	T
Dronpa2 V157L	C	Y	S	L	T
Dronpa2 V157T	C	Y	S	T	T

DNA Sequences

s10:loop:GFP (with critical mutation sites in bold)

ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGTGGCAGCCATATGCTGCCGGATAACCATTATCTGAGC**ACC**CAGACCGTGCTGAGCAAAGATCCGAACGAAGGCACCCGCGGCAGCGGCAGCATTGAAGGCCGCCATAGCGGCAGCGGCAGCAAACGCGATCACATGGTGCTGCAT**GAA**TATGTGAACGCGGCGGGCATTACCCATGGCATGGATGAACTGTATGGCAGCACCGGCGCAGCGCGAGCCAGGGCGAAGAACTGTTTACCGGCGTGGTGCCGATTCTGGTGGAACTGGATGGCGATGTGAACGGCCATAAATTTAGCGTGCGCGGCGAAGGCGAAGGCGATGCGACCATTGGCAAACTGACCCTGAAATTTATTTCCACCACCGGCAAACCTGCCGGTGCCGTGGCCGACCCTGGTGACCA CCTG**AGC**TATGGCGTGCAAGCCTTTAGCCGCTATCCGGATCACATGAAACGCCATGATTTTTTTAAAAGCGCGATGCCGGAAGGCTATGTGCAGGAACGCACCATTAGCTTTAAAGATGATGGCAAA TATAAAACCCGCGCGGTGGTGAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACGAAAGGCACCGATTTTAAAGAAGATGGCAACATTCTGGGCCATAAACTGGAATATAACTTTAACAGCCA TAACGTGTATATTACCGCGGATAAACAGAAAAACGGCATTAAAGCGAACTTTACCGTGCGCCAT AACGTGGAAGATGGCAGCGTGCAGCTGGCAGATCATTATCAGCAGAACACCCCGATTGGCGATG GCGCGGTGCTGTAA

ih:loop:GFP C48S S65T C70A V206K (with critical mutation sites in bold)

ATGGGGCATCATCATCATCATCGCAGCGGCGGCAAACCTACCGGTGCCGTGGCCGACCCTGGTGACCACCTTAACCTATGGCGTGCAAGCGTTTAGCCGCTATGGCACCCGTGGCAGCCTGGTGCCGCGTGGCAGCGGCAGCCCGGATCATATGAAACGCCATGATTTTTTTAAAAGCGCGATGCCGGAA GGCTATGTGCAGGA**CGC**ACCATTAGCTTTAAAGATGATGGCAAATATAAAACCCGCGCGGTGGTGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACGAAAGGCACCGATTTTAAAGAAGATGGCAACATTCTGGGGCATAAACTGGAATATAACTTTAACAGCCATAACGTGTATATTACCGCGGATAAACAGAAAAACGGCATTAAAGCGAACTTTACCGTGCGCCATAACGTGGAAGATGGCAGCGTGCAGCTGGCGGATCATTATCAGCAGAACACCCCGATTGGCGATGGCCCGGTGCTGCTGCCGGA TAACCATTATCTGAGC**ACC**CAGACCAAGCTGAGCAAAGATCCGAACGAAAAACGCGATCACATG GTGCTGCT**GAA**TTTGTGACCGCAGCGGGCATTACACACGGCATGGATGAACTGTATGGCGGCA CCGGCGGCAGCGCGAGCCAGGGCGAAGAACTGTTTACCGGCGTGGTGCCGATTCTGGTGGAACT GGATGGCGATGTGAACGGCCATAAATTTAGCGTGCGCGGCGAAGGCGAAGGCGATGCGACCATT GGCAAACCTGACCCTGAAATTTATTTCCACCACCTAA

ih:GFP C48S S65T (with critical mutation sites in bold)

ATGGGGCATCATCATCATCATAGCAGCGGCGGCAAACCTACCGGTGCCGTGGCCGACCCTGGTGACCACCTTA**ACCTAT**GGCGTGCAAGTGCTTTAGCCGCTATCCGGATCATATGAAACGCCATGATTTTTTTAAAAGCGCGATGCCGGAAGGCTATGTGCAGGAACGCACCATTAGCTTTAAAGATGATGGCAAATATAAAACCCGCGCGGTGGTGAATTTGAAGGCGATACCCTGGTGAACCGCATTGAAC

TGAAAGGCACCGATTTTAAAGAAGATGGCAACATTCTGGGGCATAAACTGGAATATAACTTTAA
CAGCC**CATA**ACGTGTATATTACCGCGGATAAACAGAAAAACGGCATTAAAGCGAACTTTACCGTG
CGCCATAACGTGGAAGATGGCAGCGTGCAGCTGGCGGATCATTATCAGCAGAACACCCCGATTG
GCGATGGCCCCGGTGCTGCTGCCGGATAACCATTATCTGAGC**ACC**CAGACCGTGCTGAGCAAAGA
TCCGAACGAAAAACGCGATCACATGGTGTCTGCTG**GAA**TTTGTGACCGCAGCGGGCATTACACAC
GGCATGGATGAACTGTATGGCGGCACCGGCGGCAGCGCAGCCAGGGCGAAGAACTGTTTACCG
GCGTGGTGCCGATTCTGGTGGAACTGGATGGCGATGTGAACGGCCATAAATTTAGCGTGCGCGG
CGAAGGCGAAGGCGATGCGACCATTGGCAAACCTGACCCTGAAATTTATTTCCACCACCTAA

avGFP

ATGGGTTCATCATCATCATCATCATAGCGGTCTGGTGCCGCGTGGCAGCAGTAAAGGAGAAG
AACTTTTTCACNGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAAT
TTCTGTCAAGTGGAGAGGGTGAAGGTGATGCAACATACGAAAACTTACCCTTAAATTTATTTGC
ACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAT
GCTTTTCAAGATACCCAGATCATATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGG
TTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAGTC
AAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATG
GAAACATTCTTGGACACAAATTGGAATACAACATAACTCACACAATGTATACATCATGGCAGA
CAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTT
CAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCTTTTACCAGACA
ACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGT
CCTTCTTGAGTTTGTAAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAA

supercharged GFP +36 (with critical mutation sites in bold)

ATGGGTTCATCACCACCACCATCACGGTGGCGCTAGCAAAGGTGAACGTCTGTTTCGTGGTAAAG
TACCGATCTTAGTGGAATTAAGGGGCGACGTGAACGGTCATAAATTTAGCGTGCGCGCAAAGG
CAAAGGTGACGCTACCCGTGGTAAATTGACCCTGAAGTTTATTTGCACAACAGGCCAAATTACCC
GTTCCGTGGCCCACCTTAGTGACCACCCTGACCTATGGCGTTCAGTGCTTCAGTCGTTACCCTA
AACATATGAAACGTACGATTTTTTCAAATCAGCCATGCCTAAAGGATATGTTCAAGAGCGTAC
AATCAGCTTCAAGAAGGATGGCAAATATAAAACGCGTGCGGAAGTGAAATTTGAAGGCCGCACA
TTAGTAAATCGTATCAAACCTGAAAGGTGCTGACTTCAAAGAAAAAGGCAACATTTTAGGCCATA
AACTGCGTTATAACTTTAATTCTCATAAGGTGTATATTACGGCCGATAAACGCAAGAATGGTAT
CAAGGCAAAATTCAAATTCGCCATAACGTGAAAGACGGCAGCGTTCAATTAGCGGATCATTAT
CAACAAAACACGCCGATTGGTTCGCGGGCCTGTACTGTTACCT**CGCAACC**ACTACCTGAGCACCC
GTTC**TAAA**CTGAGCAAAGATCCGAA**AGAAAAC**CGCGATCACATGGTTCTGTTAGAATTCGTGAC
CGCTGCAGGCATTAAGCACGGACGCGACGAACGCTACAAGTAA

supercharged GFP -30 (with critical mutation sites in bold)

ATGGGTTCATCACCACCACCATCACGGTGGCGCTAGCAAAGGTGAAGAGCTGTTTGACGGTGTAG
TACCGATCTTAGTGGAATTAGACGGCGACGTGAACGGTCACGAATTTAGCGTGCGCGGCGAGGG
CGAAGGTGACGCTACCGAGGGTGAATTGACCCTGAAGTTTATTTGCACAACAGGCCAATTACCC
GTTCCGTGGCCCACCTTAGTGACCACCCTGACCTATGGCGTTCAGTGCTTCAGTGATTACCCAG
ATCATATGGATCAACACGATTTTTTCAAATCAGCCATGCCTGAAGGATATGTTCAAGAGCGTAC

AATCAGCTTCAAGGACGATGGCACCTATAAAACGCGTGCGGAAGTGAAATTTGAAGGCGACACA
TTAGTAAACCGTATCGAACTGAAAGGTATCGACTTCAAAGAAGACGGCAACATTTTAGGCCATA
AGCTGGAATATAACTTTAATTCTCATGACGTGTATATTACGGCCGATAAACAGGAAAACGGTAT
CAAGGCAGAATTTGAAATTCGCCATAACGTGGAGGACGGCAGCGTTCAATTAGCGGATCATTAT
CAACAAAACACGCCGATTGGTGATGGGCCTGTACTGTTACCT**GACGAT**CACTACCTGAGCAC**GG**
AGTCAGCCCTGAGCAAAGATCCGAAC**CGAAGAC**CGCGATCACATGGTTCTGTTAGAATTCGTGAC
CGCTGCAGGCATTGATCATGGAATGGACGAGCTGTACAAGTAA

Dronpa2 (with critical mutation sites in bold)

ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGTGCCGGGCGGCAGCCATATGG
TGAGCAAGGGCGAGGAGAACAACATGGCCGTGATTAAACCAGACATGAAGATCAAGCTGCGTAT
GGAAGGCGCTGTAAATGGACACCCGTTTCGCATTGAAGGAGTTGGCCTTGGGAAGCCTTTCGAG
GGAAAACAGAGTATGGACCTTAAAGTCAAAGAAGGCGGACCTCTGCCTTTCGCCTATGACATCT
TGACAACTGTGTTCTGT**TAC**GGCAACAGGGTATTCGCCAAATACCCAGAAAATATAGTAGACTA
TTTCAAGCAGTCGTTTTCCTGAGGGCTACTCTTGGAACGAAGCATGAATTACGAAGACGGGGC
ATTTGTAACGCGACAAACGACATAACCCTGGATGGTGACTGTTATATCTATGAAATTCGATTTG
ATGGTGTGAACTTTCCTGCCAATGGTCCAGTTATGCAGAAGAGGACTGTGAAATGGGAGCCAT**C**
CACTGAGAAATTGTATGTGCGTGATGGAGTGCTGAAGGGTGAT**GTTAACAC**CGGCTCTGTGCGCTT
GAAGGAGGTGGCATTACCGATGTGACTTCAAACACTACTTATAAAGCTAAGAAGGTTGTCCAGT
TGCCAGACTATCACTTTGTGGACCACCACATTGAGATTAAGCCACGACAAAGATTACAGTAA
TGTTAATCTGCATGAGCATGCCGAAGCGCATTCTGGGCTGCCGAGGCAGGCCATGGACGAGCTG
TACAAGTAA

Amino Acid Sequences

s10:loop:GFP (with trypsin cleavage sites indicated and critical mutation sites in bold)

MGSSHHHHHSSGLVPR▼GSHMLPDNHYLST**T**QTVLSKDPNEGTR▼GSGSIEGR▼HSGSGSKRDH
MVLH**E**YVNAAGITHGMDELYGGTGGASQGEELFTGVVPILEVELDGDVNGHKFSVRGEGEGDAT
IGKLTTLKFISTTGKLPVPWPTLVTTLSYGVQAFSRYPDHMKRHDFFKSAMPEGYVQERTISFKD
DGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVIITADKQKNGIKANFT
VRHNVEDGSQLADHYQQNTPIGDGPVL

ih:loop:GFP C48S S65T C70A V206K (with trypsin cleavage sites indicated, thrombin cleavage site in red, and critical mutation sites in bold)

MGHHHHHR▼SGGKLPVPWPTLVTTLTYGVAQAFSRYGTRGSLVPR▼**RS**GSPDHMKRHDFFKSAM
PEGYVQ**E**RTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVI
TADKQKNGIKANFTVRHNVEDGSQLADHYQQNTPIGDGPVLLPDNHYLST**T**QTKLSKDPNEKRD
HMLLE**F**VTAAGITHGMDELYGGTGGASQGEELFTGVVPILEVELDGDVNGHKFSVRGEGEGDA
TIGKLTTLKFISTT

ih:GFP C48S S65T (with critical mutation sites in bold)

MGHHHHHHSSGGKLPVPWPTLVTTL**TY**GVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDD
GKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNS**HN**VYITADKQKNGIKANFTV
RHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLS**T**QTVLSKDPNEKRDHMLL**E**FVTAAGITH
GMDELYGGTGGASQGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTTLKFISTT

avGFP

MGSHHHHHSSGLVPRGSSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYKLTTLKFIC
TTGKLPVPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEV
KFEGDTLVNRIELKIDFKEDGNILGHKLEYNFNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSV
QLADHYQQNTPIGDGPVLLPDNHYLS**T**QSALS**K**DPNEKRDHMLL**E**FVTAAGITHGMDELYK

supercharged GFP +36 (with critical mutation sites in bold)

MGHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGDATRGKLTTLKFICTTGKLP
VPWPTLVTTLT**TY**GVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRT
LVNRIKLKGRDFKEKGNILGHKLRYNFN**SH**KVYITADKRKNGIKAKFKIRHNVKDGSVQLADHY
QQNTPIGRGPVLL**PN**HYL**ST****RS**KLSKDP**KE**KRDHMLL**E**FVTAAGIKHGRDERYK

supercharged GFP -30 (with critical mutation sites in bold)

MGHHHHHHGGASKGEELFDGVVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKFICTTGELP
VPWPTLVTTLT**TY**GVQCFS**DY**PDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDT
LVNRIELKIDFKEDGNILGHKLEYNFNSHDVYITADKQENG**IK**A**E**FEIRHNVEDGSVQLADHY
QQNTPIGDGPVLL**PD**DH**YL**ST**ES**ALS**K**DP**NE**DRDHMLL**E**FVTAAGIDHGMDELYK

Dronpa2 (with critical mutation sites in bold)

MGSHHHHHHSSGLVPGGSHMVSKGEENNMAVIKPKDMKIKLRMEGAVNGHPFAIEGVGLGKPF
GKQSMDLKVKEGGPLPFAYDILTTVFC**Y**GNRVFAKYPENIVDYFKQSFPEGYSWERSMNYEDGG
ICNATNDITLDGDCYIYEIRFDGVNF**P**ANGPVMQKRTVKWEP**S**TEKLYVRDGV**L**KGD**V**NTALS
EGGGHYRCDFKTTYKAKKV**V**QLPDYHFVDH**IE**IKSHDKDYSNVNLHEHAEAH**S**GLPRQAMDEL
YK

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-F₁F)): LPDNHYLS (**4-F₁F**)QTVLSKDPNE

s10(203(F₅F)): LPDNHYLS (**F₅F**)QTVLSKDPNE

s10(203(4-NH₂F)): LPDNHYLS (**4-NH₂F**) QTVLSKDPNE

Protein Expression and Purification

The protein expression and purification protocols, including noncanonical amino acid incorporation, were performed as previously described [1][3][4]. While most proteins eluted around 0.1 – 0.2 M NaCl during anion-exchange chromatography, 0.33 M NaCl was required to elute supercharged -30. For supercharged +36 and +36 s10- (Figure 3C), a cation-exchange column (HiTrap 5 mL SP HP; GE Healthcare) was required instead of an anion-exchange column, and the proteins eluted at 0.85 M and 0.80 M NaCl, respectively.

Semisynthetic Method for Split GFPs

All chemicals were purchased from Sigma-Aldrich unless otherwise stated. For s10·~~s10:loop~~:GFP, s10:loop:GFP was first cleaved with 100 units of trypsin (Type III from bovine pancreas, ≥ 10,000 BAEE units per mg; Sigma) per 1 mg of GFP in buffer A (10 mM sodium chloride (CAS 7647-14-5; Fisher), 50 mM Tris hydrochloride (CAS 1185-53-1; Fisher), pH 8.0) for 20 min at room temperature and then purified with anion exchange chromatography [1]. The purified protein was concentrated with spin filtration (10K Amicon Ultra-15 Centrifugal Filters; Millipore) to mM concentrations. Three volumes of 8 M guanidinium chloride (CAS 50-01-1; Fisher) was added to achieve a final concentration of 6 M, and the protein was further concentrated with spin filtration as needed. Immediate loss of fluorescence was observed. The truncated protein ~~s10:loop~~:GFP was carefully separated from s10 with size exclusion chromatography (Superdex 75 10/300 GL; GE Healthcare) equilibrated with denaturation buffer (6 M guanidinium hydrochloride, 300 mM sodium chloride, 30 mM Tris hydrochloride, pH 8.0), and the colored fractions free of s10 (with retention volumes of 8 – 10 mL) were collected and concentrated again via spin filtration to 1 – 2 mM. At this stage, the denatured truncated protein can be flash frozen with liquid nitrogen and stored indefinitely at -80°C.

For reconstitution (Figure 3A), the denatured protein was first evenly spread at the bottom of a beaker, and 100 volumes of at least 50 molar excess of synthetic peptide s10 dissolved in refolding buffer (20 mM sodium carbonate (CAS 497-19-8; Fisher), 20 mM

sodium chloride, 10% glycerol (CAS 56-81-5; Fisher), pH 10.0) was quickly added followed by swirling of the beaker. A fluorescent green color of the reconstituted GFP was readily seen upon refolding (except the T203(4-NH₂F) mutant which was orange). Such a procedure was contrived to minimize the local concentration of truncated protein to prevent aggregation, which occurs at protein concentrations over 20 μM [11], and to ensure sufficient guanidinium dilution. Any attempt to add truncated protein directly to the refolding buffer led to significant precipitation. To maximize reconstitution, the mixture was incubated at room temperature for 3 hours and stored at 4°C. Excess peptide was purified away by anion exchange chromatography using the same protocol described above. Mutants with noncanonical amino acids were made using this semisynthetic approach to avoid the use of specific aminoacyl-tRNA synthetases. For the T203F mutant, this semisynthetic approach was required to avoid chromophore maturation issues. T203 and T203Y mutants were not prepared this way because the split proteins were demonstrated to be spectroscopically indistinguishable from their intact counterparts [11].

For $ih^{mat} \cdot ih:loop:GFP$ (Figure 3B), the internal helix containing the matured chromophore, ih^{mat} , had to be prepared first. Anion-exchange purified $ih:loop:GFP$ C48S S65T C70A V206K was incubated with 100 units of trypsin per 1 mg of GFP in buffer A for an hour at room temperature and then purified via anion exchange again to remove the protease. The denaturation and peptide separation via size exclusion protocol was the same as described above for the s10 counterpart, except the denaturation process could take up to 30 minutes, and the desired colored fractions eluted at larger retention volumes (13 – 15 mL) since they contain the peptide rather than the truncated protein. The fractions with peptide were stored at -20°C and later loaded onto a C₄ gravity column (BAKERBOND speTM Wide-Pore Butyl (C₄) Disposable Extraction Columns; J. T. Baker). The peptide, observed to be orange, was concentrated on top of the column and then washed with 9 column volumes of water to remove extra guanidinium chloride and salt. It was eluted with 80% acetonitrile (CAS 75-05-8; Fisher), 20% water, and 0.2% trifluoroacetic acid (CAS 76-05-1). The collected fractions were colorless due to protonation of the phenolate oxygen and were subsequently lyophilized for two days. The freeze-dried peptide appeared fluffy and light yellow. A 6 mM dimethyl sulfoxide (DMSO, CAS 67-68-5) stock was made with concentration determined by UV-Vis absorption

according to the empirical extinction coefficient $8480 \text{ M}^{-1}\text{cm}^{-1}$ at 280 nm [12] and was stored at -20°C .

The “empty barrel”, ~~ih:loop~~:GFP, carrying the R96 mutation was prepared by cleaving the intact protein ih:loop:GFP R96X with 2 units of thrombin (plasminogen-free, bovine; EMD Millipore) per 1 mg of GFP in buffer A for three hours at room temperature. As mentioned above, trypsin attacked unwanted cleavage sites and could not be used. The cleaved protein was then purified with anion exchange to remove the protease. The denaturation and peptide separation via size exclusion protocol was the same as described above for the s10 counterpart, except low mM DL-dithiothreitol (DTT, CAS 3483-12-3) was added during denaturation prior to column loading, and the desired fractions with retention volumes of 8 – 10 mL were colorless. The empty barrel was concentrated via spin filtration to 1 – 2 mM based on the empirical extinction coefficient $10430 \text{ M}^{-1}\text{cm}^{-1}$ at 280 nm [12] and stored at -80°C . For reconstitution, the denatured protein was thawed and evenly spread at the bottom of a beaker. 100 volumes of at least 2 molar excess ih^{mat} from the DMSO stock dissolved in refolding buffer (20 mM sodium carbonate, 20 mM sodium chloride, 10% glycerol, pH 10.0) was quickly added followed by swirling of the beaker, after which the color became lime from the deprotonated chromophore. The full reconstitution process continued for 3 hours at room temperature, as indicated by the gradual increase in green fluorescence over time. Some aggregated protein floating in the mixture could be spotted and was spun down with a table-top centrifuge (Sorvall ST 16R Centrifuge; Thermo Scientific). The supernatant was stored at 4°C overnight and purified with anion exchange chromatography the next day. The reconstitution yield was estimated to be 50%. Expression of the intact Q94E mutant was attempted but afforded prohibitively low yields.

The identity and purity of all proteins were confirmed with electrospray ionization mass spectrometry (ESI-MS) measured with LC-MS (Waters 2795 HPLC with ZQ single quadrupole MS at the Stanford University Mass Spectrometry (SUMS) facility). The expected and observed masses are summarized in the following table, in which only proteins that have not been previously characterized are listed:

Table S3. Expected and observed mass for each fluorescent protein construct.

Constructs	Expected Mass ^a (Da)	Observed Mass ^b (Da)
GFP mutants		
s10:loop:GFP T203V	30669	30673
ih:GFP T203(3-OMeY)	28119	28128
s10:loop:GFP T203V E222Q	30668	30673
s10:loop:GFP S65T	30685	30690
s10:loop:GFP S65T	25310, 2898	25319, 2899
ih:loop:GFP S65T	24558, 3743	24563, 3744
supercharged +36	28339	28344
supercharged -30	27637	27640
ih:loop:GFP S65T R96M	29278 (immature)	29282
ih:loop:GFP S65T R96M	24533, 4762	24538, 4762
ih ^{mat} (65T) · ih:loop:GFP R96M	24533, 3743	24541, 3744
ih:GFP S65T H148D	28019	28023
s10:loop:GFP S65T T203V	30683	30686
s10:loop:GFP S65T T203V	25310, 2896	25318, 2900
s10:loop:GFP S65T T203H	30721	30724
<u>s10</u> (203F) · s10:loop:GFP S65T	25310, 2217	25320, 2218
<u>s10</u> (203(4-F ₁ F)) · s10:loop:GFP S65T	25310, 2235	25319, 2236
<u>s10</u> (203F ₅ F) · s10:loop:GFP S65T	25310, 2307	25319, 2308
<u>s10</u> (203(4-NH ₂ F)) · s10:loop:GFP S65T	25310, 2232	25319, 2233
ih:GFP S65T T203(3-OMeY)	28133	28140
s10:loop:GFP S65T T203Y	30747	30750
ih:GFP S65T Q94E	28062 (immature)	28065
ih:loop:GFP S65T R96M T203Y	29340	29344
ih:loop:GFP S65T R96M T203Y	24595, 4762	24603, 4764
ih ^{mat} (65T) · ih:loop:GFP R96M T203Y	24595, 3743	24605, 3744
ih:loop:GFP S65T R96E E222K	29275 (immature)	29280

ih:loop:GFP S65T R96E E222K	24530, 4762	24534, 4763
ih ^{mat} (65T) · ih:loop:GFP R96E E222K	24530, 3743	24538, 3744
Dronpa2 mutants		
Dronpa1 (T159M)	28947	28953
Dronpa2 T159E	28945	28950
Dronpa2 T159Q	28944	28950
Dronpa2 S142A	28901	28907
Dronpa2 V157D	28933	28952 ^c
Dronpa2 V157N	28932	28939
Dronpa2 V157L	28931	28938
Dronpa2 V157T	28919	28926

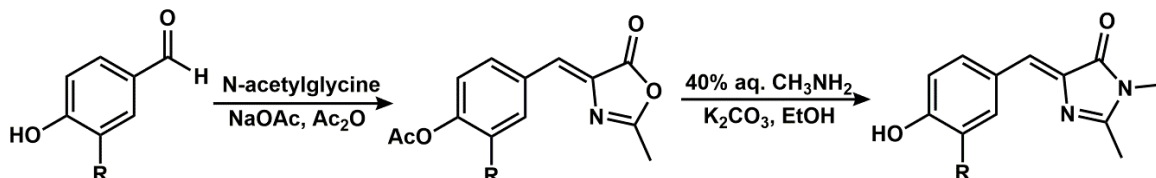
^a Predicted from the primary sequence with N-terminal methionine removed [13].

^b Proteins with ~ 30 kDa have ±10 Da deviations, depending on the protonation states.

^c The protein has mostly an immature chromophore, but spectroscopic studies could still be performed on the small mature fraction.

^d Supercharged +36 s10⁻ aggregated in low-salt conditions, so it was not characterized by ESI.

Chromophore Analogue Synthesis



Scheme S1. Synthesis of chromophore analogues. HBDI: R = H; 3-ClHBDI: R = Cl; 3-BrHBDI: R = Br; 3-CH₃HBDI: R = CH₃; 3-OMeHBDI: R = OCH₃.

The difluoro chromophore analogue ((*Z*)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1*H*-imidazol-5(4*H*)-one, 3,5-F₂HBDI, CAS 1241390-29-3) was purchased from Sigma-Aldrich under the name DFHBI, while the rest required synthesis. The Erlenmeyer–Plöchl azlactone synthesis strategy was adapted from that of the Samie Jaffrey lab at Cornell University [14].

Synthesis of HBDI ((*Z*)-4-(4-hydroxybenzylidene)-1,2-dimethyl-1*H*-imidazol-5(4*H*)-one):

N-acetylglycine (9.59 g, 0.082 mol, CAS 543-24-8), anhydrous sodium acetate (5.00 g, 0.061 mol, CAS 127-09-3), acetic anhydride (50 mL, CAS 108-24-7), and 4-hydroxybenzaldehyde (10.0 g, 0.082 mol, CAS 123-08-0) were stirred at 100°C for 2 hours. The color of the mixture started milky white, turned yellow within minutes, and gradually became clear and dark brown after 20 minutes. The reaction was cooled to room temperature and 240 mL 0°C water was added. Light yellow solid started to precipitate, was collected via vacuum filtration, washed with 0°C 70% ethanol (CAS 64-17-5, Fisher), and dried under vacuum. The yield of the azlactone intermediate was more than 15 g (> 75% yield).

10 g of intermediate was refluxed with 45 mL ethanol, 8.5 mL 40% aqueous methylamine (CAS 74-89-5; Fisher), and 850 mg potassium carbonate (CAS 584-08-7; Fisher) for 4 hours, showing a color of blood red. The mixture was removed from heat, cooled to room temperature, and 425 mL of water was added. Concentrated hydrochloric acid (CAS 7647-01-0; Fisher) was then slowly added to reach pH 3.0, during which the solution went from clear red, to murky yellow, to bright orange. Fumes were observed due to HCl reacting with remaining gaseous methylamine. The mixture was cooled overnight at 4°C to facilitate precipitation. The next day, a yellow precipitate was observed with a red supernatant. The solid was collected and refluxed with a minimal amount of ethanol to reach saturation at the boiling point of ethanol. The saturated solution was subsequently cooled at room temperature, and orange needle-like crystals appeared. The whole mixture was further cooled on ice. The crystals were collected, washed with mother liquor, and dried under vacuum in a desiccator, yielding 3 to 4 g of product (34 – 45% yield). ¹H NMR, GC-MS (*m/z* = 216 as expected), and UV–Vis absorption for both neutral and anionic states (maxima at 373 and 442 nm, respectively) in ethanol confirmed the identity and the high purity.

¹H NMR (400 MHz, CD₃OD):

HBDI: δ 7.990 (d, ³*J* = 8.4 Hz, 2H (meta to OH)), δ 7.010 (s, 1H (bridge)), δ 6.833 (d, ³*J* = 8.8 Hz, 2H (ortho to OH)), δ 3.182 (s, 3H (methyl attached to carbon)), δ 2.383 (s, 3H (methyl attached to nitrogen))

Synthesis of 3-ClHBDI ((*Z*)-4-(3-chloro-4-hydroxybenzylidene)-1,2-dimethyl-1*H*-imidazol-5(4*H*)-one), 3-BrHBDI ((*Z*)-4-(3-bromo-4-hydroxybenzylidene)-1,2-dimethyl-1*H*-imidazol-5(4*H*)-one), 3-CH₃HBDI ((*Z*)-4-(4-hydroxy-3-methylbenzylidene)-1,2-dimethyl-1*H*-imidazol-5(4*H*)-one), and 3-OMeHBDI (((*Z*)-4-(4-hydroxy-3-methoxybenzylidene)-1,2-dimethyl-1*H*-imidazol-5(4*H*)-one):

N-acetylglycine (767 mg, 6.57 mmol), anhydrous sodium acetate (539 mg, 6.57 mmol), acetic anhydride (2.5 mL), and 4-hydroxybenzaldehyde derivatives (6.57 mmol, CAS 2420-16-8 for Cl, 2973-78-6 for Br, 15174-69-3 for CH₃, 121-33-5 under the name of vanillin for OCH₃) were stirred for 2 hours at 90°C and 120°C for electron donating and withdrawing groups, respectively. The amounts of all chemicals were doubled for the electron donating group reactions. Note that higher temperature was necessary to initiate the reactions for electron withdrawing groups. The mixture started colorless, turned yellow within minutes, and gradually became dark brown after 20 minutes. The reaction was cooled to room temperature, cold ethanol (20 mL) pre-chilled on ice was added, and the mixture was subsequently stirred overnight at 4°C. Pale yellow precipitate isolated via vacuum filtration was washed with cold ethanol, hot water, and hexanes (CAS 92112-69-1 and 110-54-3; Fisher) before drying under vacuum. The resulting azlactone intermediates were measured by ESI-MS with methanol in the positive mode, and the masses were consistently higher than the expected value by 55 Da, presumably capturing adducts of the intermediate, methanol, and a sodium ion. Percent yields, expected masses, and observed *m/z* are shown in the table below:

Table S4. Percent yields, expected masses, and observed *m/z* for azlactone intermediates with different substituents.

Azlactone intermediate	Percent yield	Expected mass (Da)	Observed <i>m/z</i> in ESI+ (Da)
3-Cl	77%	279.5	334
3-Br	76%	324	380
3-CH ₃	52%	259	314
3-OCH ₃	56%	275	330

3.56 mmol of intermediate was refluxed with 15 mL ethanol, 1 mL 40% aqueous methylamine, and 700 mg potassium carbonate in an oil bath maintained at 90°C for 4 hours. Color of the mixtures ranged from blood red to cherry red depending on the substituents. Solids started to appear and accumulated during the course of the reaction. The mixture was removed from heat, cooled to room temperature, and further cooled overnight at 4°C. The bright yellow precipitate (orange for OMe) was filtered, washed with cold ethanol, and dissolved in 500 mM sodium acetate pH 3.0 aqueous solution to protonate the chromophore analogue, during which sizzling from CO₂ could be heard. The mixture was transferred to a separatory funnel, and an equal volume of ethyl acetate (CAS 141-78-6; Fisher) was added to extract the chromophore into the organic layer, which was then isolated and dried with anhydrous sodium sulfate (CAS 7757-82-6). The solvent was removed via rotary evaporation and a Schlenk line. The products were confirmed by both ¹H-NMR and ESI-MS in both positive and negative modes. Percent yields, expected masses, and observed m/z are shown in the table below:

Table S5. Percent yields, expected masses, and observed m/z for chromophore analogues (Figures S32 – S35).

Chromophore analogues	Percent yield	Expected mass (Da)	Observed m/z in ESI+ (Da)	Observed m/z in ESI- (Da)
3-Cl	58%	250.5	251, 253	249, 251
3-Br	42%	295	295, 297	293, 295
3-CH ₃	55%	230	231	229
3-OCH ₃	40%	246	247	245

¹H NMR (300 MHz, CD₃OD) (Figures S32 – S35):

3-ClHBDI: δ 8.279 (d, ⁴J = 2.1 Hz, 1H (ortho to Cl)), δ 7.807 (dd, ³J = 8.6 Hz, ⁴J = 2.1 Hz, 1H (para to Cl)), δ 6.944 (s, 1H (bridge)), δ 6.943 (d, ³J = 8.1 Hz, 1H (ortho to OH)), δ 3.185 (s, 3H (methyl attached to carbon)), δ 2.403 (s, 3H (methyl attached to nitrogen))

3-BrHBDI: δ 8.425 (d, $^4J = 1.8$ Hz, 1H (ortho to Br)), δ 7.861 (dd, $^3J = 8.6$ Hz, $^4J = 1.8$ Hz, 1H (para to Br)), δ 6.937 (s, 1H (bridge)), δ 6.927 (d, $^3J = 8.1$ Hz, 1H (ortho to OH)), δ 3.184 (s, 3H (methyl attached to carbon)), δ 2.402 (s, 3H (methyl attached to nitrogen))

3-CH₃HBDI: δ 7.897 (d, 4J not resolved, 1H (ortho to CH₃)), δ 7.793 (dd, $^3J = 8.3$ Hz, $^4J = 2.1$ Hz, 1H (para to CH₃)), δ 6.991 (s, 1H (bridge)), δ 6.792 (d, $^3J = 8.7$ Hz, 1H (ortho to OH)), δ 3.185 (s, 3H (methyl attached to carbon)), δ 2.391 (s, 3H (methyl attached to nitrogen)), δ 2.214 (s, 3H (methyl attached to aromatic ring))

3-OMeHBDI: δ 7.967 (d, $^4J = 2.1$ Hz, 1H (ortho to OCH₃)), δ 7.460 (dd, $^3J = 8.4$ Hz, $^4J = 1.8$ Hz, 1H (para to OCH₃)), δ 7.005 (s, 1H (bridge)), δ 6.835 (d, $^3J = 8.1$ Hz, 1H (ortho to OH)), δ 3.920 (s, 3H (methyl attached to oxygen)), δ 3.189 (s, 3H (methyl attached to carbon)), δ 2.393 (s, 3H (methyl attached to nitrogen))

Buffers for Spectroscopic Studies

The pH 8 buffer is 50 mM Tris hydrochloride and 200 mM NaCl titrated with sodium hydroxide. The pH 10 buffer is 50 mM sodium carbonate and 200 mM NaCl titrated with hydrochloric acid. This moderate amount of salt ensured the solubility of supercharged +36 s10-. The pH 6.5 buffer for avGFP is 50 mM 2-(N-morpholino)ethanesulfonic acid (MES, CAS 1266615-59-1) and 100 mM NaCl titrated with sodium hydroxide.

Sample Preparation for Stark Spectroscopy

Glass forming solvents, such as ethanol or a 1:1 mixture of glycerol and aqueous buffer, are required for low temperature Stark spectroscopy experiments. Protein samples were concentrated with spin filters (Amicon Ultra 0.5 mL, 10 kDa cutoff) and thoroughly exchanged into the appropriate buffers. The concentrated samples were then mixed with an equal volume of glycerol. Model chromophores were freshly dissolved in pH 10 buffer and mixed with an equal volume of glycerol right before measurements to avoid base-catalyzed hydrolysis. The final sample concentrations for Stark spectroscopy were checked with the NanoDrop (ND-1000 Spectrometer; NanoDrop) to ensure a maximum absorbance of 0.2 – 0.9 with 25 μ m path length using the Beer–Lambert law for good signal-to-noise ratio in low temperature absorption. Despite the temperature sensitivity of Tris pK_a ($\Delta pK_a/\Delta T = -0.031$ K⁻¹) [15] and the titratability of GFP S65T mutants [8], no

appreciable absorbance change was observed at 77 K between samples in Tris- or phosphate-based buffers provided that the samples were properly flash frozen.

S2 Spectroscopic Methods

Room-Temperature UV–Vis Absorption and Fluorescence Measurements

PerkinElmer Lambda 25, Lambda 365, and Agilent Cary 6000i UV–Vis spectrometers were used for UV–Vis absorption measurements at room temperature. Both Horiba Fluorolog-3 and PerkinElmer LS 55 fluorescence spectrometers were used for steady-state fluorescence measurements. The emission maxima listed in Table S12 were determined from the peak wavelengths of their corresponding emission profile (Figure S27).

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

The method with slight variations has been reviewed elsewhere [16][17]. The cell for electronic Stark spectroscopy 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 to form an electrode pattern. The windows were separated from each other by a pair of finely cut 27-micron thick Teflon spacers to ensure electrical insulation and held in place with a metal clamp and four adjustable screws at the corners. The interference fringes were optimized by adjusting the screw tightness under a fluorescent lamp to make the windows as parallel as possible, and the path length was determined by the undulation 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 (whose length is adjusted to match the distance between the center of the optical windows and the top of the cryostat) with electrical wires and alligator clips attached to the aforementioned nickel electrodes. The rod was designed with the capability to rotate the sample by any dialed-in angle about the rod axis to adjust the χ angle between probe light polarization and field direction. The whole apparatus was meticulously insulated with electrical tape (Super 88; 3M), and a sample (at most 10 μ L) was loaded carefully into the cell by capillary uptake. The whole rod was then rapidly plunged into an immersion cryostat [18] pre-filled with liquid nitrogen to allow the sample to form a transparent glass upon flash freezing. The cryostat was specifically designed to avoid bubbling of liquid nitrogen and resistive heating of the sample during voltage application. Since any bubbles

in the frozen sample can cause the electricity to arc through the electrodes and lead to premature dielectric breakdown, protein samples with glycerol were centrifuged at 17000 rcf for at least 40 min prior to sample loading. The capacitance of the frozen sample cell was checked through the rod to ensure the voltage application circuit was intact after plunging and field application.

The custom-built spectrometer could be switched between Stark spectroscopy and absorption modes, with the latter dual-beamed (sample and reference channels) to account for the fluctuations in the intensity of the 500 W xenon lamp. Specifically, light from the lamp was dispersed through a monochromator (1681B; SPEX) with an entrance 0.50 mm wide slit and an exit pinhole and split into two channels with a striped mirror scavenged from a UV–Vis Spectrometer (Lambda 2; PerkinElmer). Along the sample channel, the light passed through a Glan–Thompson polarizer, focused onto the sample, and again focused onto a silicon diode detector prepackaged with an operational amplifier (EG&G Judson HUV-1100BQ or Hamamatsu S8745-01) with a custom-made circuit board (OSHPark) connected to a lock-in amplifier (SR830; Stanford Research). Along the reference channel, the light was modulated by a beam chopper (SR540; Stanford Research) and focused onto another silicon diode detector (PDA55; Thorlabs, Inc) with both connected to another lock-in amplifier. The sinusoidal high voltage signal was generated from the sample channel lock-in amplifier 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 – 3.0 kV, which amounts to a peak external field strength F_{ext} of 0.3 – 1.6 MV/cm given the sample thickness.

For Stark spectroscopy, only the sample channel was used. 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 in the dark since it was not locked-in. The overall phase θ of the signal, which is related to the lag between the applied field and the detector signal, was obtained by minimizing $\left| \frac{X}{I} \sin \theta - \frac{Y}{I} \cos \theta \right|$ summed over all wavelengths. 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, assuming

ΔI is at least one order of magnitude smaller than I [16]. Higher-order Stark spectra [16], specifically 4ω , can be obtained by $\Delta A(4\omega) = \frac{8\sqrt{2}}{\ln 10} \frac{\Delta I(4\omega)}{I}$. The application frequency ω (= 203 Hz) was chosen such that it (1) avoids any conventional noise from the power lines, (2) is large enough to minimize $1/f$ noise [17], (3) is small enough to minimize frequency-dependent signal attenuation (especially at 4ω) from the RC circuit connected to the operational amplifier of the detector, and (4) is small enough when compared with the slew rate of the voltage amplifier to prevent distortion of the sinusoidal voltage signal after amplification. To achieve a good signal-to-noise ratio and avoid over-smoothing due to averaging, aside from high sample concentration or high field strength, a wavelength scan rate of 0.3 nm/s and a time constant of 300 ms were chosen. Shorter time constants were tested to ensure there was no over-averaging. The sensitivity of the lock-in amplifier was picked for each scan to avoid detection overload and over digitization. The polarized probe light was set to be horizontal, and depolarization along the beam path was carefully checked before any experiment since strained quartz windows on the cryostat can lead to birefringence and thus depolarize the incoming light. χ angles 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 later analysis. Smaller χ angles could not be sampled because the probe light was severely clipped by the window holder.

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, since a better signal-to-noise ratio was more important in this case). The scanning rate and time constant were set to match those of Stark measurements. The blank sample was prepared by carefully removing the Stark sample from the cell with an air hose and then loading the cell with a 1:1 buffer and glycerol mixture (or ethanol). By comparing the output voltages from both channels with samples and blanks, the absorbance A was determined at $\chi = 90^\circ$ with good reproducibility and 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 ΔA were shown with their corresponding absorbance A scaled to 1 MV/cm with $\chi = 90^\circ$ according to its proportionality to $(F_{ext})^2$ and $(F_{ext})^4$ for the 2ω and 4ω spectra, respectively. 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 $\Delta\mu_{app}$ and the measured angles ζ between difference dipoles and transition dipoles [16]:

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

and

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

where F_{ext} is the strength of the externally applied field through the parallel-plate capacitor. As stated in the main text, the magnitude of a vector quantity is denoted by dropping the vector notation. For example, $\Delta\mu_{app}$ ($= |\Delta\vec{\mu}_{app}|$) represents the magnitude of $\Delta\vec{\mu}_{app}$. The data was processed by the MATLAB code kindly provided by Professor Robert Stanley at Temple University [17]. Simultaneous fitting of ΔA and A at $\chi = 90^\circ$, 70° , and 50° were performed with a minimal number of Gaussians and their analytical derivatives to model the vibronic progression and effectively smooth the absorbance spectra. No real physical meanings were associated with the individual peak positions, 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 ; i.e. one band) 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 S11). Due to nonnegligible contribution from $\Delta\mu_{app}$, no attempt was made to isolate the difference polarizabilities $\Delta\alpha$ from B_χ (Section S5). The uncertainties in $\Delta\mu_{app}$ from both fitting and duplicates amounted to $\pm 5\%$, while those in ζ

were $\pm 5^\circ$, 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 f reflects our lack of control over the magnitude of the local field sensed by the chromophore (Section S6) [16]. Numerical second derivative analysis (Figure S24) was only performed after applying a Savitzky–Golay filter to the absorption spectra for excess noise removal.

Extinction Coefficient Determination of GFP Mutants

The base denaturation method to determine extinction coefficients was employed in this study and described in our previous works [1][4]. Aside from mutants with previously determined values, extinction coefficients were only measured for mutants that exhibit purely anionic absorption at $\text{pH} \leq 10$, summarized in Table S15. The corresponding dipole moments and oscillator strengths were subsequently obtained via the following formulas by integrating from 330 to 550 nm [19]:

$$f_{opt}m = \sqrt{\frac{3000 \ln 10 \, hc}{8\pi^3 N_A} n \int \frac{\varepsilon(\bar{\nu})}{\bar{\nu}} d\bar{\nu}} \quad (\text{S3})$$

and

$$\text{oscillator strength} = \frac{3000 \ln 10 \, m_e c}{\pi e^2 N_A} n \int \varepsilon(\bar{\nu}) d\bar{\nu} \quad (\text{S4})$$

in which h is Planck's constant, c is the speed of light, N_A is Avogadro's number, m_e is the mass of an electron, e is the elementary charge, n is the refractive index ($= 1.33$ for water), and $\varepsilon(\bar{\nu})$ is the extinction coefficient in $\text{M}^{-1} \text{cm}^{-1}$. Unity oscillator strength corresponds to the strongest possible (maximally allowed) absorption. Similar to Stark spectroscopy, a local field factor f_{opt} at optical frequency is necessary in Equation S3 (Section S6) [19].

S3 X-ray Crystallography

Protein Crystallization

All ih circular permutants were exchanged into the pH 8 spectroscopic buffer and concentrated to 10 mg/mL. The s10 split protein complex s10:løøp:GFP was further purified by size exclusion chromatography, during which the protein was exchanged into a pH 8.0 buffer containing 0.1 M NaCl and 50 mM Tris-HCl. The sample was then concentrated to approximately 19 mg/mL. The concentrated protein stocks were centrifuged at 17000 rcf for 10 min prior to use to remove dust and particulates. The hanging drop wells were all manually set using 24-well VDX plates with sealant (Hampton Research) and 22 mm siliconized glass circle cover slides (Hampton Research). Specifically, 1 µL of protein stock was mixed with 1 µL of mother liquor on the glass slide, which was then flipped and sealed above the well containing 500 µL mother liquor.

ih:GFP Y66 and Y66CIY readily crystallized in 34% (wt/vol) PEG 3350 (CAS 25322-68-3; Hampton Research) and 0.15 M ammonium acetate (CAS 631-61-8) within two days at room temperature, while crystallization took about a week in 0.15 M ammonium acetate and 30% PEG 3350. These conditions are very similar to those in which short hydrogen bond mutants (S65T/H148D) crystallized in previous work [3]. Y66CH₃Y and T203OMeY required streak seeding with a cat whisker using crushed ih:GFP S65T Y66 crystals. After forced nucleation, the crystals grew within two to three days to a size of a few hundreds of microns in the longer dimension for conditions spanning 24% to 34% PEG 3350. s10:løøp:GFP crystallized in 0.1 M MES, pH 6.5 with 20% 2-methyl-2,4-pentanediol (MPD, CAS 107-41-5, under the name hexylene glycol) only at 4°C and in the dark due to its photosensitive nature. Needle-like crystals appeared after 3 days of equilibration and reached their maximum size of 200 µm × 20 µm × 20 µm after one month. 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 [20]. Two different cryoprotectants were tested for the ih circular permutants: 0.15 M ammonium acetate with 34% PEG and perfluoropolyether (CAS 69991-67-9; Hampton Research). Crystals from high PEG concentrations (28 – 34%) with perfluoropolyether

gave the best diffraction results. For s10:ih:GFP, 0.1 M MES, pH 6.5 with 45% MPD was used for cryoprotection, though it led to substantial ice formation. The following table summarizes the concentrations of PEG in which the ih protein crystals diffracted to the highest resolution.

Table S6. Optimal PEG 3350 concentration for crystallization of each ih:GFP variant and mutant.

	Y66	Y66CIY	Y66CH ₃ Y	T203OMeY
PEG (% wt/vol)	30	30	28	34

X-ray Data Collection and Structure Determination

X-ray data for all crystals were acquired at SSRL (Menlo Park, CA) [21] at 100 K. Specifically, diffraction intensities of all protein crystals were collected at BL 12-2, except for Y66CIY diffraction, which was collected at BL 14-1. Data processing was performed with XDS [22][23] using the autoxds script [24]. Molecular replacement was performed in PHENIX [25] using the ih:GFP S65T H148D (PDB: 4ZF3) and superfolder GFP (PDB: 2B3P) structures as the search models for ih and s10 circular permutants, respectively. Chromophore restraint files were built using REEL and eLBOW in PHENIX. Numerous rounds of model building and refinement were carried out with Coot [26] and PHENIX. Structures of ih:GFP S65T and S65T Y66CIY have been further refined with PDB-REDO [27]. The overall folds of the proteins, despite being circular permutants, are virtually identical to superfolder GFP (PDB: 2B3P) (Figure S16). The resulting data collection and refinement statistics are summarized in Table S7.

Table S7. X-ray data collection and refinement statistics for GFP variants and mutants.

	ih:GFP S65T	ih:GFP S65T Y66CIY	ih:GFP S65T Y66CH ₃ Y	ih:GFP S65T T203OMeY	s10:leap:GFP
PDB entry	6OFK	6OFL	6OFM	6OFN	6OFO
Data collection statistics					
beamline	BL 12-2	BL 14-1	BL 12-2		
wavelength (Å)	0.8856	0.9795	0.9795		
detector distance (mm)	188	150/130 (two crystals combined)	257	280	302
resolution range (Å)	36.47 – 1.15 (1.191 – 1.15)	35.85 – 1.25 (1.28 – 1.25)	34.92 – 1.48 (1.533 – 1.48)	34.91 – 1.65 (1.709 – 1.65)	26.82 – 2.603 (2.696 – 2.603)
space group	P 1 2 ₁ 1 (No.4)				
unit cell dimensions a, b, c (Å) α, β, γ (°)	51.94, 68.89, 60.95 90.0, 100.7, 90.0	52.22, 68.73, 60.88 90.0, 100.6, 90.0	50.98, 68.81, 60.61 90.0, 101.6, 90.0	48.19, 67.76, 58.85 90.0, 101.5, 90.0	50.75, 51.07, 97.30 90.0, 103.67, 90.0
total observations	1975584 (175350)	2601793 (86948)	925442 (90350)	600171 (56344)	53957 (5216)
unique reflections	146651 (14296)	113139 (7995)	66960 (6627)	43937 (4324)	14664 (1138)
multiplicity	13.5 (12.2)	23.0 (10.9)	13.8 (13.6)	13.7 (13.0)	3.7 (3.6)
completeness (%)	98.2 (96.6)	96.9 (93.6)	97.9 (97.2)	98.2 (97.3)	92.4 (72.5)
mean I/σ	28.4 (1.5)	27.4 (1.5)	19.7 (1.5)	21.3 (1.5)	6.44 (0.9)
Wilson B-factor (Å ²)	15.2	16.1	18.6	21.9	40.1
R _{meas}	0.035 (1.76)	0.058 (1.65)	0.074 (1.44)	0.072 (1.42)	0.192 (1.76)
CC _{1/2}	1 (0.792)	1 (0.721)	1 (0.768)	1 (0.670)	0.991 (0.542)

	ih:GFP S65T	ih:GFP S65T Y66CIY	ih:GFP S65T Y66CH₃Y	ih:GFP S65T T203OMeY	s10:loop:GFP
Refinement statistics					
reflections used	146542	113105	66942	43915	13937
reflections used for R _{free}	7327	5656	3349	2197	693
R _{work}	0.1590	0.1376	0.1597	0.1637	0.2029
R _{free}	0.1844	0.1657	0.1774	0.1906	0.2350
chromophore three-letter code	CRO	4NT	MFV	CRO	GYS
number of non-H atoms	4186	4443	4243	4056	3417
protein	3736	3836	3777	3722	3331
ligand	48	92	46	44	42
solvent	402	515	420	290	44
protein residues	453	458	468	462	447
RMSD bond lengths (Å)	0.014	0.022	0.007	0.006	0.006
RMSD bond angles (°)	1.76	1.73	1.25	1.14	1.11
Ramachandran favored (%)	99.09	98.65	98.24	97.51	96.96
Ramachandran allowed (%)	0.91	1.35	1.76	2.26	3.04
Ramachandran outliers (%)	0.00	0.00	0.00	0.23	0.00
rotamer outliers (%)	1.17	1.35	0.23	0.24	0.00
clashscore	4.89	8.62	3.42	2.55	5.99
Average B-factor (Å ²)	23.52	23.95	29.35	31.41	42.83
protein	22.53	22.49	28.67	31.10	43.00
ligand	17.61	14.88	17.91	20.32	34.42
solvent	33.39	36.47	36.73	37.10	37.35

S4 The Marcus–Hush Model and Its Implications for Spectroscopic

Observables

The goal of this section is to understand the implications of the environment on spectroscopic observables of the GFP chromophore, such as the Stokes shift, Stark tuning rate, and vibronic coupling strength, by treating the chromophore as a mixed-valence system using Marcus–Hush theory. Important results that are applied to data interpretation in the main text will be boxed. First, we will construct the semiclassical Marcus–Hush model for two centers, for which energies are in terms of spectroscopic units (cm^{-1}) to facilitate comparisons between experimental data and theoretical results. The notation follows our previous work on the Stark effect for mixed-valence compounds [28]. The simplified version of Marcus–Hush theory consists of only one (antisymmetric) nuclear degree of freedom q (identified as BLA in this case), which is sufficient to capture the relevant phenomena. The word “semiclassical” implies treating the electron quantum mechanically and the nuclei classically.

For each low-lying valence-bond structure (the P and I form in this work, Figure 6A) in which the electrons prefer different distributions, we can write down the corresponding energy dependence along the nuclear degree of freedom and invoke the harmonic approximation:

$$V_P(q) = \frac{\bar{\nu}}{2}(q + \delta)^2, \quad V_I(q) = \Delta\bar{\nu} + \frac{\bar{\nu}}{2}(q - \delta)^2 \quad (\text{S5})$$

also known as the diabatic “surfaces” (Figure 6B). They possess different minima in q , demonstrating the fact that the equilibrium nuclear configurations are different in order to accommodate different electronic distributions. The difference in minima is denoted 2δ for later convenience. The curvature, $\bar{\nu}$, or the vibrational frequency of the BLA mode, is assumed to be shared between the two diabatic states, which can be justified by the relatively constant absorption maxima differences between the 0–1 and 0–0 bands for chromophores in various environments (Figure 5B, Table S11). The minimum energy offset between the two states is the driving force $\Delta\bar{\nu}$, representing the differential stabilization of the two valence-bond structures in a given environment. If these diabatic states were the energy eigenstates, no electron delocalization would be possible, and we would expect to see only two distinct colors from each structure regardless of the

environment, contrary to the color palette exhibited by the series of GFP mutants. Instead, these two states are coupled electronically through the electronic coupling term V_0 , and thus the potential energy part of the Hamiltonian in the basis of valence-bond structures is as follows:

$$V(q) = \begin{pmatrix} \frac{\bar{v}}{2}(q + \delta)^2 & V_0 \\ V_0 & \Delta\bar{v} + \frac{\bar{v}}{2}(q - \delta)^2 \end{pmatrix} \quad (\text{S6})$$

To solve for the energy eigenstates S_0 and S_1 (or adiabatic states, Figure 6C), which are linear combinations (i.e. mixing) of the diabatic states:

$$|S_0\rangle \equiv |1\rangle = c_P(q)|P\rangle + c_I(q)|I\rangle, \quad |S_1\rangle \equiv |2\rangle = c_I(q)|P\rangle - c_P(q)|I\rangle \quad (\text{S7})$$

we have to diagonalize the potential energy matrix in Equation S6. The resulting adiabatic “surfaces” are

$$V_{1,2}(q) = \frac{\Delta\bar{v}}{2} + \frac{\bar{v}}{2}(q^2 + \delta^2) \mp \frac{1}{2}\sqrt{(\Delta\bar{v} - 2\bar{v}\delta q)^2 + 4V_0^2} \quad (\text{S8})$$

with the lower and higher eigenvalues corresponding to the ground and excited states, respectively. An interesting feature of this model is that the excited state surface is completely determined if the ground state surface is fixed. This assertion can also be seen from the orthogonality of states shown in Equation S7, so we expect properties involving electronic transitions to fully correlate with ground state properties, including the electronic distribution. The fact that knowing the ground state electron density is enough to determine all properties is reminiscent of the first Hohenberg–Kohn theorem in density functional theory [29]. By tuning the strength of the electronic coupling V_0 , we observe qualitative changes in the shape of the ground state surface. It was noted by Treynor and Boxer [28] that these changes occur when $2V_0 = |\Delta\bar{v} + \lambda|$, where $\lambda \equiv 2\bar{v}\delta^2$ is the reorganization energy for the nuclei. Larger V_0 leads to a single minimum, while smaller V_0 causes a double-welled ground state. The former and the latter are usually conveniently classified as Robin–Day Class III and II (Figure 7A), respectively, in the mixed-valence system literature [30].

We justify in the main text that V_0 for the chromophore is sufficient to cause strong electron delocalization, thus resulting in a single minimum on the ground state energy surface. From the derivative of $V_1(q)$, we obtain the ground state minimum $q_{1,\text{min}}$:

$$q_{1,min} = \frac{-\delta(\Delta\bar{\nu} + 2\bar{\nu}\delta q_{1,min})}{\sqrt{(\Delta\bar{\nu} + 2\bar{\nu}\delta q_{1,min})^2 + 4V_0^2}} \equiv -\delta \frac{x}{\sqrt{x^2 + 4V_0^2}} \quad (\text{S9})$$

The Class II/III transition criterion suggests $x (= \Delta\bar{\nu} + 2\bar{\nu}\delta q_{1,min})$ is much smaller than V_0 , allowing for a Taylor expansion in $\frac{x}{2V_0}$:

$$q_{1,min} = \frac{-\delta}{2V_0} \frac{x}{\sqrt{1 + \left(\frac{x}{2V_0}\right)^2}} = -\frac{\delta}{2V_0} x + \frac{\delta}{16V_0^3} x^3 + \dots \quad (\text{S10})$$

Keeping only the first term and solving for $q_{1, \min}$ yields:

$$q_{1,min} \approx -\frac{\Delta\bar{\nu}}{2V_0} \delta \left(1 + \frac{\lambda}{2V_0}\right) = -\frac{\Delta\bar{\nu}}{2V_0} \delta + \mathcal{O}(V_0^{-2}) \quad (\text{S11})$$

Therefore, the Franck–Condon transition energy for absorption is

$$\bar{\nu}_{abs} = V_2(q_{1,min}) - V_1(q_{1,min}) = \sqrt{(\Delta\bar{\nu} - 2\bar{\nu}\delta q_{1,min})^2 + 4V_0^2} \approx \sqrt{\Delta\bar{\nu}^2 \left(1 + \frac{\lambda}{2V_0}\right)^2 + 4V_0^2} \quad (\text{S12})$$

which can be approximated as the 0–0 transition energy if the reorganization energy λ is much smaller than V_0 (as verified by Subsection 2.4 in the main text):

$$\bar{\nu}_{abs} \approx \bar{\nu}_{0-0} = \sqrt{\Delta\bar{\nu}^2 + 4V_0^2} \quad (\text{S13})$$

which has been shown by Olsen and McKenzie [31] based on Platt's formulation.

Analogously, we derive the excited state minimum $q_{2, \min}$:

$$q_{2,min} \approx \frac{\Delta\bar{\nu}}{2V_0} \delta \left(1 - \frac{\lambda}{2V_0}\right) = \frac{\Delta\bar{\nu}}{2V_0} \delta + \mathcal{O}(V_0^{-2}) \quad (\text{S14})$$

and the Franck–Condon transition energy for emission:

$$\bar{\nu}_{em} = V_2(q_{2,min}) - V_1(q_{2,min}) = \sqrt{(\Delta\bar{\nu} - 2\bar{\nu}\delta q_{2,min})^2 + 4V_0^2} \approx \sqrt{\Delta\bar{\nu}^2 \left(1 - \frac{\lambda}{2V_0}\right)^2 + 4V_0^2} \quad (\text{S15})$$

Remarkably, both the absorption and emission energy only differ from the 0–0 transition energy by scaling of the driving forces (Equations S12, S13, and S15), and they all reduce to the same value as anticipated at $\lambda = 0$ or $\Delta\bar{\nu} = 0$. The Stokes shift is thus

$$\begin{aligned} \bar{\nu}_{Stokes} = \bar{\nu}_{abs} - \bar{\nu}_{em} &\approx 2V_0 \left[\sqrt{1 + \left(\frac{\Delta\bar{\nu}}{2V_0}\right)^2 \left(1 + \frac{\lambda}{2V_0}\right)^2} - \sqrt{1 + \left(\frac{\Delta\bar{\nu}}{2V_0}\right)^2 \left(1 - \frac{\lambda}{2V_0}\right)^2} \right] \\ &\approx \frac{\lambda}{2} \left(\frac{\Delta\bar{\nu}}{V_0}\right)^2 + \mathcal{O}(V_0^{-4}) \end{aligned} \quad (\text{S16})$$

The correlation between Stokes shift and absorption energy is obtained by removing the driving force $\Delta\bar{\nu}$ from Equations S13 and S16:

$$\bar{\nu}_{Stokes} = \frac{\lambda}{2V_0^2} [\bar{\nu}_{abs}^2 - (2V_0)^2] \quad (\text{S17})$$

which predicts a linear correlation between $\bar{\nu}_{Stokes}$ and $\bar{\nu}_{abs}^2$ (Figure S29). The parameters V_0 and λ were determined to be 9530 cm^{-1} and 2910 cm^{-1} from Figure 8, respectively. Assuming that the only varying parameter among mutants is the driving force, we deduce the ratio of ranges over which absorption and emission maxima change from Equations S13 and S16:

$$\frac{d\bar{\nu}_{em}}{d\bar{\nu}_{abs}} = \frac{\frac{\partial \bar{\nu}_{em}}{\partial \Delta \bar{\nu}}}{\frac{\partial \bar{\nu}_{abs}}{\partial \Delta \bar{\nu}}} = 1 - \frac{\frac{\partial \bar{\nu}_{Stokes}}{\partial \Delta \bar{\nu}}}{\frac{\partial \bar{\nu}_{abs}}{\partial \Delta \bar{\nu}}} \approx 1 - \frac{\lambda}{V_0^2} \bar{\nu}_{abs} \quad (\text{S18})$$

For GFPs ($\bar{\nu}_{abs} \approx 20000 \text{ cm}^{-1}$) and PYPs ($\bar{\nu}_{abs} \approx 22000 \text{ cm}^{-1}$) the ratios are estimated to be 1/3 and 1/4, respectively, which roughly reproduce the ranges observed from our experimental data (Table S12, Figures 5A and S27) and Philip *et al.* [32], respectively. Interestingly, Marcus–Hush theory also confirms and provides a basis for understanding the claim in Philip *et al.* that the major mechanism responsible for the range difference of absorption and emission maxima is the relative curvature change between S_0 and S_1 across mutants [32], while tuning the vertical energy gap also plays a nonnegligible role.

For electro-optic properties, such as Stark tuning rates and transition dipole moments, we can utilize the Mulliken–Hush approximation, in which the transition dipole moment between the two valence-bond states vanishes. That is to say, the valence-bond states are also dipole moment eigenstates. This is not a necessary assumption, and we can introduce another mixing angle between the true dipole moment eigenstates and the valence-bond states [31], but it renders the formulation rather cumbersome and is shown to be unnecessary for data interpretation (Section S9). The difference dipole moment between the two valence-bond states is denoted $\Delta\mu_{CT}$, with the subscript meaning “charge transfer”. This quantity can be interpreted as the sensitivity of the driving force under the influence of external electric fields applied in the Stark measurements and is also assumed intrinsic to the chromophore:

$$\Delta\bar{\nu}(F_{ext}) = \Delta\bar{\nu} - \Delta\mu_{CT}F_{ext} \quad (\text{S19})$$

in which the zero-field reference point is taken to be no external field. This means external electric fields can bias the driving force. A one-dimensional chromophore is implicitly assumed in this construction since there are only two centers for charge localization, so F_{ext} should be understood as the field projection on the one-dimensional system. Note

that for convenience in the one-dimensional case (this section and Section S5), quantities that are supposed to be vectors or tensors are denoted without the arrows and underscores, e.g. $\Delta\mu_{CT}$ and F_{ext} , and they carry signs. In contrast, when referring to three-dimensional systems, such notations refer to the magnitudes; symbols without the arrows or underscores are unsigned as in Equations S1 and S2. We will occasionally revert back to the full notation for magnitudes (e.g. $|\Delta\vec{\mu}_{CT}|$) when ambiguity arises. By plugging into Equation S13 and expanding in terms of F_{ext} [16], we obtain

$$\bar{\nu}_{abs}(F_{ext}) = \sqrt{(\Delta\bar{\nu} - \Delta\mu_{CT}F_{ext})^2 + 4V_0^2} \equiv \bar{\nu}_{abs} - \Delta\mu F_{ext} - \frac{1}{2}\Delta\alpha F_{ext}^2 + \dots \quad (\text{S20})$$

where the linear coefficient $\Delta\mu$ is the difference dipole moment between the excited and ground states or the so-called ‘‘Stark tuning rate’’:

$$\Delta\mu = \frac{\Delta\bar{\nu}}{\bar{\nu}_{abs}} \Delta\mu_{CT} \quad (\text{S21})$$

and the quadratic coefficient $\Delta\alpha$ is the difference polarizability between the excited and ground states:

$$\Delta\alpha = -\frac{4V_0^2}{\bar{\nu}_{abs}^3} \Delta\mu_{CT}^2 \quad (\text{S22})$$

Equation S21 can be also found in Olsen *et al.* [31]. Note that both electro-optic properties, $\Delta\mu$ and $\Delta\alpha$, depend on $\Delta\bar{\nu}$ through Equation S13, so they are not intrinsic to the chromophore and can vary across mutants. However, as discussed in Section S5, $\Delta\alpha$ is in fact not sensitive to the change in $\Delta\bar{\nu}$ such that treating it as a constant is valid. In other words, this provides a justification for the quadratic Stark approximation for chromophore color tuning in Drobizhev *et al.* [34]. Combining Equations S13 and S21, we derive the correlation between Stark tuning rate and absorption maximum:

$$\bar{\nu}_{abs} = \frac{2V_0}{\sqrt{1 - \left(\frac{\Delta\mu}{\Delta\mu_{CT}}\right)^2}} \quad (\text{S23})$$

We expect zero Stark tuning rate when the chromophore exhibits the reddest possible color and the driving force is zero, which means no dipolar charge redistribution can happen upon excitation when the electron cloud is maximally delocalized (i.e. no bias between the two diabatic states). Further tuning of the driving force to negative values will again blue shift the absorption and increase the Stark tuning rate (Equation S21), which is qualitatively very different from the behavior predicted by a linear Stark effect,

in which the absorption maximum is linearly dependent on the protein electric field experienced by the chromophore [33] (see also Section S5). Note that we include the local field factor f in the corresponding equation in the main text (Equation 2) to emphasize the fact that all experimentally determined dipole-moment quantities are associated with f due to environmental polarization in response to the externally applied field (Section S6). Equation S23 is also invoked by Drobizhev *et al.* (see Supplementary Information of [34], Equation 1) with an explicit consideration of various hydrogen bonding interactions, but the resulting description required distinction between long- and short-range electrostatic interactions, and thus the model introduced too many free parameters to be meaningfully compared with experiments.

To study the degree of charge localization at the phenolate moiety for the ground state (at $q_{1,min}$), we need the coefficients in Equation S7 from diagonalizing the matrix in Equation S6:

$$|c_P(q_{1,min})|^2 = \frac{1}{2} + \frac{\Delta\bar{v}}{2\bar{v}_{abs}} \quad (\text{S24})$$

As anticipated, the larger the magnitude of the driving force $\Delta\bar{v}$, the more localized the electronic distribution (Figure 7B). From the dipole moment operator matrix $\boldsymbol{\mu}$, the transition dipole moment m between the ground and excited states is as follows:

$$m \equiv |\langle S_1 | \boldsymbol{\mu} | S_0 \rangle| = \sqrt{|c_P(q_{1,min})|^2 (1 - |c_P(q_{1,min})|^2)} \Delta\mu_{CT} = \frac{V_0}{\bar{v}_{abs}} \Delta\mu_{CT} \quad (\text{S25})$$

which governs the extinction coefficient of the absorption band. Since the model is intrinsically one-dimensional, the transition dipole moment must be parallel to the difference dipole moment, i.e. $\cos^2 \zeta = 1$, which is not too far from the estimation of $\cos^2 \zeta = 0.75$ to 0.93 via electronic Stark experiments (Figure S10). Such deviation can be accommodated by modelling another diabatic state with slightly less contribution and charge localization away from the oxygen-oxygen axis (Section S9). From Equations S21 and S25, an identity can be obtained:

$$\Delta\mu_{CT}^2 = \Delta\mu^2 + 4m^2 = (\text{Tr } \boldsymbol{\mu})^2 - 4(\text{Det } \boldsymbol{\mu}) \quad (\text{S26})$$

which can also be readily represented in terms of the trace (Tr) and determinant (Det) of the dipole moment operator $\boldsymbol{\mu}$ and thus are invariant under any basis transformation. This implies the validity of Equation S26 even when another mixing angle is introduced between the true dipole moment eigenstates and the valence-bond states (Section S9).

The difference dipole moment between the two valence-bond states $\Delta\mu_{CT}$ is partitioned between the Stark tuning rate $\Delta\mu$ and the transition dipole moment m , the ratio of which is determined by the environment through the driving force $\Delta\bar{\nu}$. The sensitivity of m to the change in $\Delta\bar{\nu}$ may seem similar to that of $\Delta\mu$, but for the same reason as the insensitivity of $\Delta\alpha$ (Equation S22), m is also nearly constant across mutants as shown by their extinction coefficients (Figures 10C and 10D).

To explain the vibronic structure, the difference equilibrium position between excited and ground states is required (Equations S11 and S14) [19]:

$$\Delta R = q_{2,min} - q_{1,min} \approx \frac{\Delta\bar{\nu}}{V_0} \delta \quad (\text{S27})$$

which is expected to vanish when either the driving force or δ is zero. Another related quantity that measures the strength of vibronic coupling is the Huang–Rhys factor S :

$$S \equiv \frac{\Delta R^2}{2} = \left(\frac{\Delta\bar{\nu}}{V_0}\right)^2 \frac{\delta^2}{2} = \left(\frac{\Delta\bar{\nu}}{2V_0}\right)^2 \frac{\lambda}{\bar{\nu}} \quad (\text{S28})$$

where we have assumed that the observed reorganization energy λ can be entirely attributed to BLA as justified in the main text. The integrated intensity ratio of the 0→1 and 0→0 transitions can thus be obtained from their corresponding Franck–Condon factors:

$$\frac{(FC)_{0\rightarrow 1}}{(FC)_{0\rightarrow 0}} = S = \left(\frac{\Delta\bar{\nu}}{2V_0}\right)^2 \frac{\lambda}{\bar{\nu}} \quad (\text{S29})$$

and thus the 0→1 sideband is only prominent when the driving force is large, corresponding to bluer mutants as observed in Figure 5A.

Note that among all electronic-transition-related observables, including absorption maxima (Equation S13), emission maxima (Equation S15), Stokes shifts (Equation S16), Stark tuning rates (Equation S21), transition dipole moments (Equation S25), and vibronic structures (Equation S29), the signs of the corresponding driving forces are intrinsically ambiguous. In other words, localization on either ring gives rise to identical results. The only way to figure out the sign from electronic spectroscopy is to infer the direction of the difference dipole from mutational studies (either by modifying the environment or the chromophore) [4]. The sign cannot be determined solely from electronic Stark spectroscopy due to the limitation from isotropic samples (or any centrosymmetric samples such as crystals with space groups possessing inversion symmetry). Another

method is to probe the vibrational frequencies of the chromophore using Raman spectroscopy [35][36], taking advantage of the opposite bond-order-alternation patterns of the valence-bond structures (Figure 6A), but this is beyond the scope of the current study.

S5 Rebane's Quadratic Stark Model and Stark Spectroscopy

In Drobizhev *et al.* [34], a quadratic Stark effect model was invoked to explain the color tuning behavior of the GFP chromophore. Specifically, $\Delta\alpha$ was assumed to be field independent. In the language of Marcus–Hush theory, this entails an insensitivity of $\Delta\alpha$ to varying driving forces as we hint at in Section S4. Figure S1 shows $\Delta\alpha$ plotted against absorption maximum (Equations S13 and S22), and the variation is indeed small compared with that of Stark tuning rates (Figure 9), hence omission of higher order terms (i.e. difference hyperpolarizability) is valid.

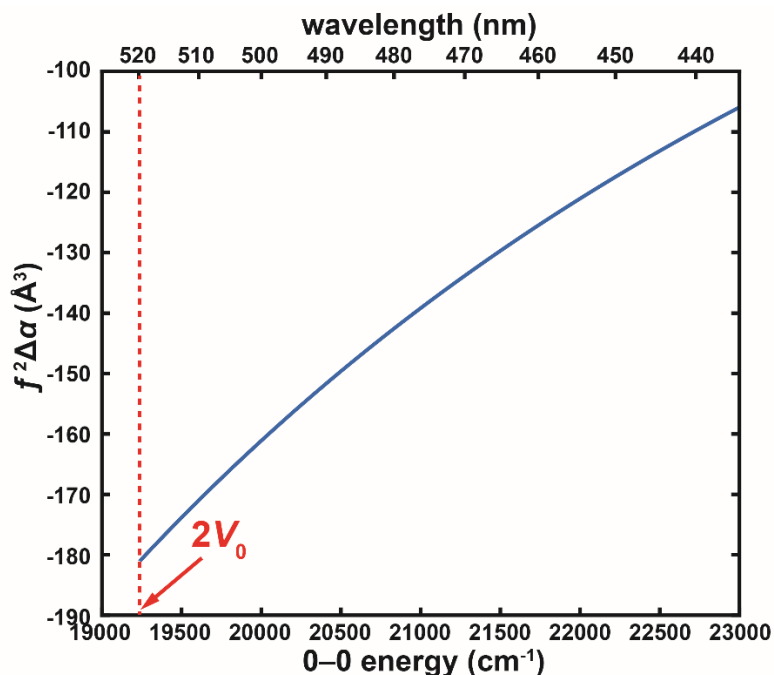


Figure S1. Predicted difference polarizability as a function of 0–0 energy from Equations S13 and S22 using the parameters obtained from Figure 9, in which no *a priori* value of the local field factor f is assumed.

The discrepancy between the difference polarizability shown on the vertical axis in Figure S1 and Rebane's value of -35 \AA^3 determined from two-photon absorption cross-sections originates from the local field effect (Section S6). The existence of $\Delta\alpha$ arises from the coupling of two diabatic states, for which the difference in minimum energies is assumed to obey a strict linear Stark effect (Equation S19). To be specific, perturbing the driving forces with an electric field leads to a change in mixing of the two diabatic states for both ground and excited states, so the resulting Stark tuning rate cannot be constant. The Marcus–Hush model, with the assumption of invariant diabatic states, automatically

predicts $\Delta\alpha$ from V_0 and $\Delta\mu_{CT}$, which are measurable from Stark experiments, and no additional parameter is required (see Equation S32). Furthermore, all the electro-optic parameters are determined accordingly [28]. In contrast, the model in Drobizhev *et al.* treats $\Delta\alpha$ as an additional independent parameter.

If the magnitude of $\Delta\alpha$ is around 35 \AA^3 , which amounts to $0.12 \text{ D}/(\text{MV}/\text{cm})$, it should manifest itself as the first derivative component (B_χ in Equation S1) in Stark spectra [16]. However, such an effect was not mentioned in our previous publication [37] and may have thus led to subsequent confusion by others about the linearity of the chromophore's Stark effect [33]. The Stark spectra of GFP are truly dominated by second derivative components (Section S11, Figure S10), suggesting a large linear Stark effect. Closer examination of the fits reveals nontrivial first derivative components (Figure S10), but by working with isotropic samples in the Stark spectroscopy experiments, it is hard to tease apart the contributions from $\Delta\alpha$ and $\Delta\mu$ as seen in the expression for B_χ [16]:

$$B_\chi = \frac{5}{2} \text{Tr}(\underline{\Delta\alpha}) + (3 \cos^2 \chi - 1) \left(\frac{3}{2} \underline{\Delta\alpha}_m - \frac{1}{2} \text{Tr}(\underline{\Delta\alpha}) \right) + \frac{1}{|\bar{m}|^2} \sum_{ij} [10m_i A_{ij} \Delta\mu_j + (3 \cos^2 \chi - 1) (3m_i A_{jj} \Delta\mu_i + m_i A_{ij} \Delta\mu_j)] \quad (\text{S30})$$

Here we explicitly show the indices because this expression is not limited to one-dimensional cases. In rare occasions (e.g. the vibrational Stark effect of nitriles), components of $\underline{\Delta\alpha}$ can be extracted from B_χ [38], or zero transition polarizability \underline{A} must be assumed [17]. However, such an assumption is especially poor for non-centrosymmetric molecules inferred from perturbation theory (see footnote 1 in [16]), and we will now examine its validity in the context of the Marcus–Hush model. If we consider the field dependence on the transition dipole moment m (Equation S25) through Equation S19, the transition polarizability A and transition hyperpolarizability B can be expressed in terms of the reduced Stark tuning rate $\frac{\Delta\mu}{\Delta\mu_{CT}} \equiv x$:

$$A = \frac{\Delta\mu_{CT}^2}{4V_0} x(1 - x^2), \quad B = \frac{\Delta\mu_{CT}^3}{16V_0^2} (3x^2 - 1)(1 - x^2)^{\frac{3}{2}} \quad (\text{S31})$$

For convenience, we also express $\Delta\alpha$ and m in terms of x :

$$\Delta\alpha = -\frac{\Delta\mu_{CT}^2}{2V_0} (1 - x^2)^{\frac{3}{2}}, \quad m = \frac{\sqrt{1-x^2}}{2} \Delta\mu_{CT} \quad (\text{S32})$$

We need a one-dimensional version for the coefficients of the derivative components, assuming $\zeta = 0$:

$$A_\chi = \frac{2}{5m^2}(A + 2mB)(\cos^2 \chi + \frac{1}{2}), \quad B_\chi = 3(\Delta\alpha + \frac{4A\Delta\mu}{m})(\cos^2 \chi + \frac{1}{2}), \quad C_\chi = 6\Delta\mu^2(\cos^2 \chi + \frac{1}{2}) \quad (\text{S33})$$

The relative contribution of $\Delta\mu$ and $\Delta\alpha$ to the first-derivative component is therefore

$$\frac{4A\Delta\mu}{m\Delta\alpha} = -\frac{4x^2}{1-x^2} \quad (\text{S34})$$

Their contributions are opposite in sign and cancel each other out when x is approximately 0.45, which is the case for the bluer mutants, so a large $\Delta\mu$ can mask the effect from $\Delta\alpha$ in B_χ . Figure S2 plots the contributions from $\Delta\mu$ and $\Delta\alpha$ based on the Marcus–Hush model and shows that extracting $\Delta\alpha$ from B_χ can be unreliable especially for bluer mutants (larger Stark tuning rates).

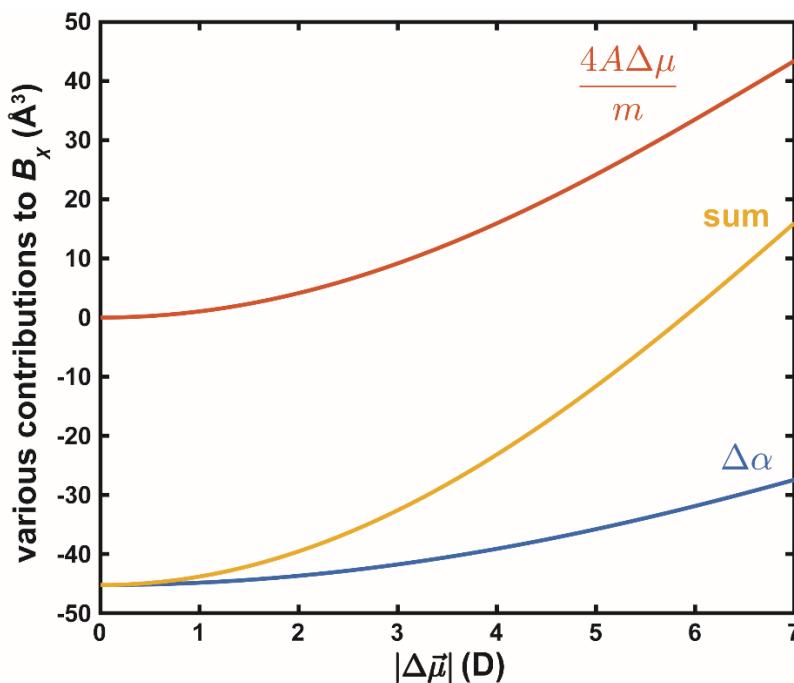


Figure S2. Predicted contributions from $\Delta\mu$ and $\Delta\alpha$ to the first derivative component B_χ of the Stark spectra using Equations S31 and S32. Note that in this plot, f is assumed to be 2 (Section S6), but the ratio of the contributions is constant regardless (Equation S34).

There are three ways to measure $\Delta\alpha$ and avoid the limitations posed by Stark measurements on isotropically immobilized samples: (i) hole-burning Stark spectroscopy [39][40] allowing one to isolate subpopulations with certain orientations [41]; (ii) performing Stark spectroscopy on non-centrosymmetric crystals that fixes the relative

orientation between the chromophores and the external applied field [42]; (iii) mutational studies as in Drobizhev *et al.* to effectively orient the chromophore with respect to the protein field, which is varied via site-directed mutagenesis [34][43]. Using the Marcus–Hush model, upper limits of A_χ , B_χ , and C_χ are estimated to be on the order of $10^{-21} \text{ m}^2/\text{V}^2$, $10^{-15} \text{ m}^2/\text{V}^2 \text{ cm}^{-1}$, and $10^{-11} \text{ m}^2/\text{V}^2 (\text{cm}^{-1})^2$, respectively. For a typical anionic GFP absorption band at 77 K, if the absorbance is around 1, its wavenumber-weighted first and second derivatives are on the order of 10^{-3} and 10^{-6} , respectively. Consequently, the second derivative component is about an order of magnitude larger than the first derivative component, which is three orders of magnitude larger than the zeroth derivative component if not affected by the mutual cancellation of $A\Delta\mu$ and $\Delta\alpha$ terms in B_χ .

S6 Stark Spectroscopy and Local Field Factors

Since all molecules of interest in the Stark experiments are embedded in dielectric media rather than vacuum, the local field experienced by the molecules is not equal to the externally applied electric field that we control during the experiment. Rather, the two fields differ by a scaling factor due to the polarization effect from the dielectric medium. This factor, usually known as the local field factor, is generally a tensor since the medium can be anisotropic, but for convenience it is usually approximated as a scalar assumed to be roughly constant across typical media. The most widely cited theory that accounts for the local field factor was proposed by Lars Onsager [44], who introduced the concept of solvent reaction field to tackle this problem. However, this model is still far from satisfactory (for example, due to uncertainty on the meaning of the cavity radius), and no analytical theory so far can predict local field factors with sufficient accuracy [45], so experimentalists usually include a local field factor f for any observables that necessitate such a conversion. Since major polarization mechanisms, including dipole reorientation, nuclear displacement, and electronic distortion [46], are heavily dependent on the driving frequencies of the incoming electric fields, we expect smaller local field factors at higher driving frequencies due to the inability for mechanisms with lower intrinsic frequencies to respond. For example, during our electronic Stark experiment, the Stark tuning rate $\Delta\mu$ couples to the applied electric field (~ 200 Hz) while the transition dipole moment m couples to the probing visible light (~ 500 THz), so the local field factor f_{opt} associated with the apparent transition dipole moment is expected to be smaller than the f associated with the apparent Stark tuning rate. The former's local field factor should be ubiquitous across all spectroscopic techniques, but it is rarely necessary to isolate the real extinction coefficient [19]. For frequencies beyond which any polarization mechanism can respond, such as X-ray frequencies, the local field factor should reduce to 1.

Stark tuning rates determined via electronic Stark spectroscopy (this work) and two-photon absorption (Drobizhev *et al.* [34]) provide a unique opportunity to test the ratio of local field factors at two distinct frequencies, albeit in different environments. Specifically, the former was performed in a frozen glass of a water/glycerol mixture, while the latter was carried out in liquid buffer at room temperature. For two-photon absorption, one of the photons behaves as the probe light, while the other serves as the field

perturbation just like in electronic Stark spectroscopy. However, the field perturbation during two-photon absorption is at optical frequency (~ 800 nm from a Ti:sapphire laser), and thus the only possible polarization mechanism is electronic, which should correspond to the same local field factor f_{opt} as in the apparent transition dipole moment. During our experiment, an electric field of ~ 200 Hz is applied, and we expect an additional local field factor contribution from the nuclear displacement due to the applied field but not from dipole reorientation because the sample is frozen. Comparison between our and Rebane's data reveals a difference of a factor of 1.6 if no *a priori* estimation for f_{opt} is applied (Figure S3). In other words, if the Lorentz theory for f_{opt} ($= 1.26$ for water [43]) is invoked, f for electronic Stark spectroscopy is 2, which is consistent with the value estimated from vibrational Stark spectroscopy (see addendum in [47]). Another interesting comparison arises from Equation S26, in which fm can be predicted from the electronic Stark measurements and compared against $f_{\text{opt}}m$ obtained from extinction coefficients. In this case, we observe the ratio to be around 1.2 (Figures 10C and 10D), which is slightly less than the aforementioned factor of 1.6. So far we are unsure if this slight inconsistency is universal to all condensed phase systems since we are limited by only investigating GFPs. More systematic studies comparing results from two-photon cross section measurements, electronic Stark spectroscopy, and extinction coefficient determination could lead to an informative estimate of local field factors.

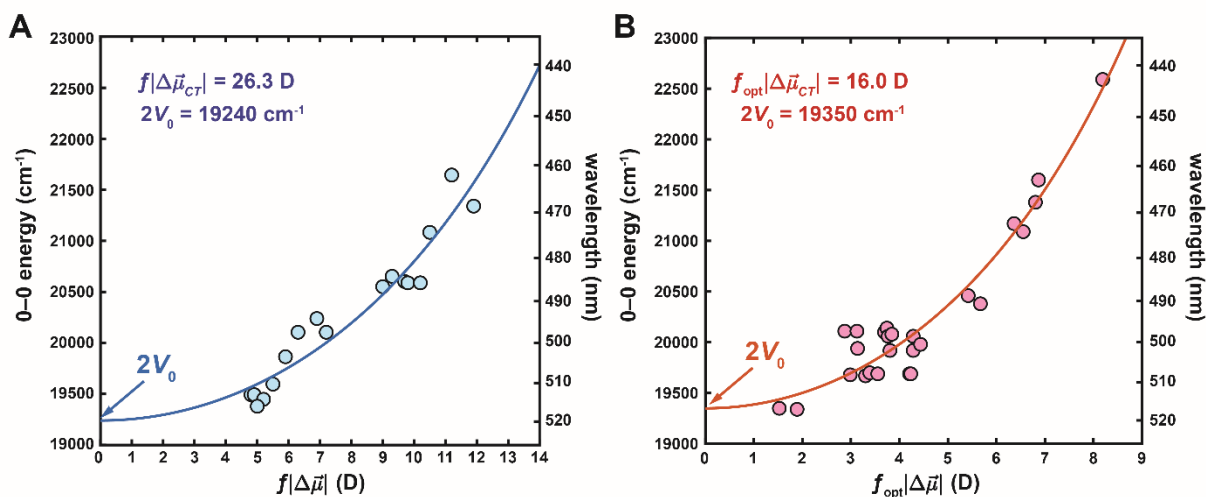


Figure S3. 0–0 energy plotted against the Stark tuning rate for each S65T mutant obtained from (A) electronic Stark spectroscopy and (B) two-photon absorption cross section measurements. The fit curves are based on Equation S23. Data in (B) are

reproduced from ref. [34] with f_{opt} restored. While $\Delta\mu_{CT}$ carry different local field factors from these two approaches, the two estimated reddest possible 0–0 energies $2V_0$ agree well.

S7 Supercharged GFPs: Color and Chromophore pK_a

In this section we assert that while the environment can tune both the absorption spectra and the pK_a of the GFP chromophore, they are not entirely governed by the same factors leading to their poor correlation in most circumstances. To support this statement, we measured the absorption spectra of a series of supercharged [48] and non-supercharged GFPs (Table S8). The mutated residues are all exposed to the bulk solvent, so none of their side chains is in direct contact with the chromophore. Note that supercharged +36 s10⁻ was deliberately designed to avoid the possibility of total cancellation of electric fields from the charged residues (Figure 3C), even though the charges are already not uniformly distributed on the β-barrel for supercharged +36 and -30.

Table S8. Comparison of charged residues in supercharged and non-supercharged GFPs. For the range of pHs used in this work (8.0 and 10.0), the residues listed below are assumed to be positively (blue) and negatively (red) charged. The residues of supercharged +36 s10⁻ within thick borderlines are those inherited from supercharged -30 strand 10, while the rest are based on supercharged +36.

strand	residue number	+36	-30	+36 s10 ⁻	ih:GFP S65T	s10:loop:GFP S65T
N-terminal loop	3	K	K	K		Q
	6	R	E	R		E
	9	R	D	R		T
1	11	K	V	K		V
	19	K	D	K		D
2	26	K	E	K		K
	32	K	E	K		E
	34	K	E	K		E
loop 2-3	39	R	E	R		I
3	41	K	E	K		K
loop 3-ih	52	K	E	K		K
loop ih-4	73	R	D	R		R
	76	K	D	K		D
	79	K	D	K		K
	80	R	Q	R		R
	90	K	E	K		E
loop 4-5	102	K	D	K		D
5	105	K	T	K		K
	111	E	E	E		V
loop 5-6	117	R	D	R		D

6	124	K	E	K	E
	128	R	I	R	T
loop 6-7	133	K	D	K	D
	142	R	E	R	E
7	149	K	D	K	N
loop 7-8	157	R	Q	R	Q
	158	K	E	K	K
8	164	K	E	K	N
	166	K	E	K	T
loop 8-9	172	K	E	K	E
loop 9-10	190	R	D	R	D
	197	R	D	D	D
	198	N	D	D	N
10	204	R	E	E	Q
	206	K	A	A	V
loop 10-11	212	K	N	N	N
	214	K	D	D	K
C-terminal loop	230	K	D	K	T
	233	R	M	R	M
	236	R	L	R	L
	238	K	K	K	G

The UV–Vis spectra and absorption maxima of these species are shown in Figure S4 and Table S9, respectively. The peak maxima are similar (within 5 nm or 200 cm⁻¹), while the chromophores exhibit very different pK_a's. Supercharged +36 and ih:GFP exhibit only the (deprotonated chromophore) B state at pH 8.0, whereas the (protonated chromophore) A state is still present at pH 10.0 for supercharged -30. The measured Stark tuning rates (Table S9) for these species are also relatively close compared to the range of Stark tuning rates observed from the mutants in this study (Figure 9). Furthermore, from an online database of reported fluorescent proteins [49], no correlation between chromophore pK_a and excitation wavelength can be found. For example, mNeonGreen [50] and Citrine [51] both have a pK_a of 5.7 but absorb at 506 and 516 nm, respectively.

For reasons that will be more explicitly stated in Section S8, while supercharging the barrel leads to a change in the electrostatic environment, its influences on the pK_a and color of the chromophore are different. Specifically, since the Franck–Condon excitation process involves a charge transfer from the phenolate to imidazolinone ring,

the color and the Stark tuning rate are more sensitive to the magnitude and sign of the electric potential *difference* between the two rings. The chromophore protonation or deprotonation process, however, involves a net charge gain or loss, so the free energy difference between the A and B states are more sensitive to the *absolute* potential exerted by charged residues. The accumulation of negative charges results in a drastic decrease in the overall electric potential, and the pK_a increases accordingly, but it does not cause a significant potential *difference* between the two rings, and thus the color is minimally affected. The rough correlation between color and pK_a observed in Slocum *et al.* [52] was simply due to sampling at only a single mutational site close to the chromophore (T203), which would be expected to change *both* the absolute potential and the potential difference between the two rings to similar degrees. Only by modifying multiple sites can one decouple these two parameters. This assertion is further confirmed from combining the mutations R96M and T203Y (Figure 11), in which stripping hydrogen bonds at both ends of the chromophore only slightly shifts the absorption maximum while causing an overall destabilization of the chromophore's anionic state.

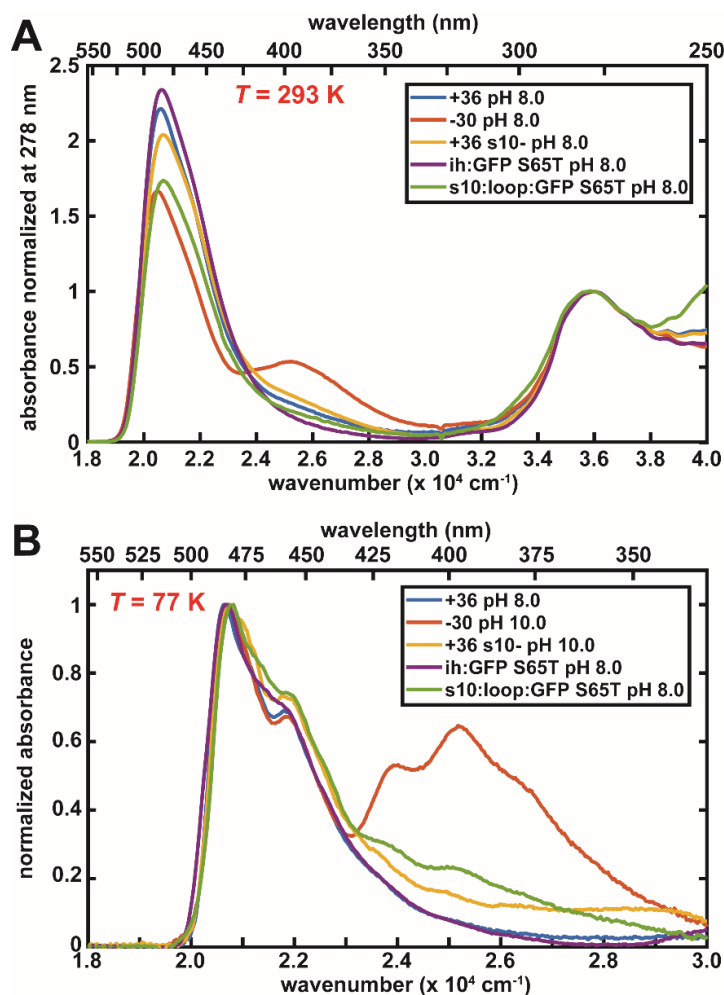


Figure S4. Normalized UV-Vis absorption spectra for supercharged and non-supercharged GFPs at (A) room temperature and (B) 77 K. The former is normalized against the 278 nm band, while the latter is normalized against the peak maximum. The absorption maxima are summarized in Table S9. Note that the scales of the horizontal axes are different for the two panels because the output of the xenon lamp and the sensitivity of the detector are poor below 330 nm for low-temperature absorption measurements (Section S2).

Table S9. Absorption maxima and Stark tuning rates for supercharged and non-supercharged GFPs.

	+36	-30	+36 s10-	ih:GFP S65T	s10:loop:GFP S65T
room-temperature absorption maximum (nm)	485	488	484	485	483

room-temperature absorption maximum (cm ⁻¹)	20600	20500	20700	20600	20700
77 K absorption maximum (nm)	484.5	482.1	482.7	483.3	480.3
77 K absorption maximum (cm ⁻¹)	20640	20740	20720	20690	20820
Stark tuning rate (D)	9.0	9.7	9.8	10.2	9.3

S8 Further Discussion of Electrostatic Contribution to Excitation Energies and Driving Forces

It might be tempting to replace the external field F_{ext} applied by the parallel-plate capacitor in Equation S20 with the field F_{prot} exerted by the protein environment in which the chromophore is embedded to explain the color tuning of the chromophore in various environments [33][34][43]:

$$\bar{\nu}_{abs}(F_{prot}) = \bar{\nu}_{abs} - \Delta\mu F_{prot} - \frac{1}{2}\Delta\alpha F_{prot}^2 + \dots \quad (\text{S35})$$

where the zero-field reference point represents the chromophore in vacuum with the same geometry as in the protein. Analogously, one could also generalize the electrostatic contribution to the driving force from Equation S19, namely

$$\Delta\bar{\nu}(F_{prot}) = \Delta\bar{\nu}_0 - \Delta\mu_{CT}F_{prot} \quad (\text{S36})$$

in which the zero-field value is again evaluated in vacuum. In this section we address potential problems in such generalizations for electronic systems.

Approximating environmental electrostatics as the major contributor for color tuning of electronic transitions is well-founded, since the nuclei are relatively fixed, and it is the electronic redistribution that makes the electronic absorption sensitive to environmental electrostatics [53]. However, in both Equations S35 and S36, only the dipolar contributions are considered, since the spatial distribution of the protein electric field, $\vec{F}_{prot}(\vec{r})$, can be highly inhomogeneous across the electron cloud of the bound chromophore. To see this and its consequences, we reproduce arguments involved in the dipolar approximation, which can be found in standard texts on classical electrodynamics [46][54].

For any electrostatic contribution to the transition energy, which is the energy difference between the S_1 Franck–Condon state and S_0 minimum, the general expression is

$$hc\bar{\nu}_{abs} = \int \rho_e(\vec{r})V_e(\vec{r}) d^3\vec{r} - \int \rho_g(\vec{r})V_g(\vec{r}) d^3\vec{r} \quad (\text{S37})$$

where the subscripts e and g represent electronic excited state and ground state, respectively, ρ is the charge density of the chromophore, and V is the electric potential

exerted by the protein environment or the external field, depending on the “zero-field” reference state (free chromophore in vacuum for the former and GFP without external field for the latter). For simplicity, the electric potential around the chromophore should be similar in the excited state and ground state ($V_e(\vec{r}) = V_g(\vec{r}) \equiv V(\vec{r})$), especially if the nuclear degrees of freedom can fully account for the electric potential based on the Franck–Condon principle. Note that there might be some difference in electronic polarization of the environment in response to the electronic redistribution of the chromophore upon photoexcitation, causing a difference in potential, which we will neglect for now. With this caveat in mind, the transition energy becomes

$$hc\bar{\nu}_{abs} = \int \Delta\rho(\vec{r})V(\vec{r}) d^3\vec{r} \quad (\text{S38})$$

in terms of the difference charge density $\Delta\rho(\vec{r}) (\equiv \rho_e(\vec{r}) - \rho_g(\vec{r}))$, which solely arises from the electron density difference of the chromophore upon excitation according to the Franck–Condon principle. We can now define an origin \vec{r}_0 , around which we can perform a Taylor expansion on $V(\vec{r})$:

$$\begin{aligned} V(\vec{r}) &= V(\vec{r}_0) + (\vec{r} - \vec{r}_0) \cdot \vec{\nabla}V(\vec{r}_0) + \frac{1}{2} \sum_{i,j} (r_i - r_{0i})(r_j - r_{0j}) \frac{\partial^2 V}{\partial r_i \partial r_j}(\vec{r}_0) + \dots \\ &= V(\vec{r}_0) - (\vec{r} - \vec{r}_0) \cdot \vec{F}(\vec{r}_0) - \frac{1}{6} \sum_{i,j} [3(r_i - r_{0i})(r_j - r_{0j}) - (\vec{r} - \vec{r}_0)^2 \delta_{ij}] \frac{\partial F_j}{\partial r_i}(\vec{r}_0) + \dots \end{aligned} \quad (\text{S39})$$

in which \vec{F} is the negative gradient of the potential, δ_{ij} is the Kronecker delta, and the last equality is obtained by recognizing $\vec{\nabla} \cdot \vec{F} = 0$ (Gauss’s law) since the field source is outside the chromophore. Plugging Equation S39 back into Equation S38 and performing the integration yields:

$$\begin{aligned}
hc\bar{\nu}_{abs} &= \Delta qV(\vec{r}_0) - \left[\int \Delta\rho(\vec{r})(\vec{r} - \vec{r}_0)d^3\vec{r} \right] \cdot \vec{F}(\vec{r}_0) \\
&\quad - \frac{1}{6} \sum_{i,j} \left[\int \Delta\rho(\vec{r})[3(r_i - r_{0i})(r_j - r_{0j}) - (\vec{r} - \vec{r}_0)^2\delta_{ij}]d^3\vec{r} \right] \frac{\partial F_j}{\partial r_i}(\vec{r}_0) + \dots \\
&= -\Delta\vec{\mu}(\vec{r}_0) \cdot \vec{F}(\vec{r}_0) - \frac{1}{6} \sum_{i,j} \Delta Q_{ij}(\vec{r}_0) \frac{\partial F_j}{\partial r_i}(\vec{r}_0) + \dots
\end{aligned} \tag{S40}$$

where the difference total charge (monopole) Δq of the chromophore between the ground and excited state is essentially zero because charge is conserved within the π -system, and ΔQ_{ij} is the difference quadrupole moment which interacts with the field gradient, as can be seen, for example, in the quadrupole splitting of Mössbauer spectroscopy and quadrupolar NMR. This is essentially the multipole expansion of electrostatic energy. One thing worth mentioning is that $\Delta\vec{\mu}$ is independent of the choice of \vec{r}_0 when $\Delta q = 0$. For processes that involve net charge gain/loss rather than mere redistribution ($\Delta q \neq 0$), such as redox reactions and protonation/deprotonation, the potential itself should be the governing factor for the reaction energetics, and thus any attempt to correlate chromophore transition energy with its pK_a is unphysical [52]. As we see in the cases of supercharged GFPs (Section S7) and different GFP circular permutants [8], these two quantities are not well correlated. Similarly, the effect of exterior charges on the reduction potentials of hemes have been demonstrated with a man-made system “maquette” [55], and absorption maxima exhibit no obvious correlation with their corresponding reduction potentials for hemes in cytochromes [56].

If the reference state is the chromophore in the GFP environment in the absence of an applied external field, Equation S40 describes the color shift in an electronic Stark experiment with a *homogenous* external electric field. In this case, the field gradient is zero, and thus we can extract from a Stark spectrum the exact $\Delta\vec{\mu}$ (including both the intrinsic and induced components from the protein environment [57]) and the dipole approximation is exact. However, if we were to extend this argument to describe color tuning by the protein environment, where an *inhomogeneous* field is experienced by the chromophore, it is unclear where we should place the *point* difference dipole $\Delta\vec{\mu}$, i.e. where we define the origin \vec{r}_0 . The choice would affect the relative contribution from the

dipolar term and higher order terms, jeopardizing the validity of the dipole approximation. This conundrum also applies when calculating ring current effects on nuclear spins in NMR using the dipole approximation [58]. For the GFP chromophore, some suggest the origin should be at the electrical center of gravity (around the bridge of the chromophore) [34]. It might also be possible to choose an origin to minimize the higher order terms in Equation S40 if Equation S40 is a converging perturbation series. Completely removing the quadrupolar contribution might not always be possible because the difference quadrupole moment or the field gradient tensor both have five degrees of freedom (symmetric and traceless) [59].

The only scenario in which we can safely ignore the higher-order terms and perform the dipole approximation is when the field is sufficiently homogeneous (“quasiuniform” in Landau’s language [54]) at the chosen origin. For diatomic vibrational probes, such an approximation is more suitable than for electronic systems, because the former has a smaller dimension that is comparable to the length scale over which the protein field varies. For example, the substrate carbonyl experiences perturbation in both vibrational transition energy (manifested as its vibrational frequency) and free energy barrier for catalysis across ketosteroid isomerase (KSI) mutants [60]. The correlation between these two energetic quantities can be successfully interpreted as the influence of protein electrostatics in the language of an *effective* electric field, which is the difference in electric potentials experienced by the two charges divided by their mutual distance. This *effective* electric field should not be too far from the *localized* electric field exerted by the protein environment, which also consists of atoms separated by bond lengths comparable to the dimension of the carbonyl. On the other hand, in the case of the GFP chromophore embedded in the protein environment, calculations reveal a large degree of electric potential inhomogeneity across the dimensions of the electronic system, especially at positions where the chromophore and the protein matrix interact through hydrogen bonds [61][62]. Even if an *effective* field is defined through Equations S35 and S36 as a convenient scale to quantify the electrostatic contribution of various protein environments, it is not immediately clear how this field can be inferred from the spatial distribution of the *localized* protein field. *This is why we invoke the notion of driving force in this study to characterize the electrostatic interaction between the chromophore and its*

environment, avoiding the physically unrealistic need to use a single electric field to represent the whole electrostatic environment of the chromophore (see also the next paragraph). If one insists on finding the physical meaning of the effective field and its influence on the driving force, it can be determined using Equation S36 and the definition of driving force:

$$\Delta\Delta\bar{v} \equiv \Delta\bar{v} - \Delta\bar{v}_0 = \int \rho_I(\vec{r})V(\vec{r}) d^3\vec{r} - \int \rho_P(\vec{r})V(\vec{r}) d^3\vec{r} \quad (\text{S41})$$

where $\rho_I(\vec{r})$ and $\rho_P(\vec{r})$ are the charge densities of the diabatic states with negative charge localized on the imidazolinone and phenolate, respectively, and $V(\vec{r})$ is the electric potential exerted by the protein environment. Because the diabatic states have their charge localized at either of the oxygen atoms, the *effective* field is analogueous to the carbonyl case: the difference in electric potentials experienced by the two oxygens divided by their mutual distance:

$$\Delta\Delta\bar{v} \approx -e(V_I - V_P) = -ed_{CT} \left(\frac{V_I - V_P}{d_{CT}} \right) = -\Delta\mu_{CT} F_{eff} \quad (\text{S42})$$

in which $V_I - V_P$ is the difference in potential sensed by the chromophore at the two oxygens, d_{CT} is the oxygen-oxygen distance, and F_{eff} is the corresponding *effective* protein field, which results naturally from the formalism of Marcus–Hush theory and determines the driving force. Even though this field truly governs color tuning, it is coarse-grained and does not reflect the spatial distribution of the protein *localized* field (unless the field were homogeneous). In other words, the *localized* field corresponds to the tangent slope of an electric potential curve, while the *effective* field experienced by a probe can be described as the secant slope. It is only when the physical dimension of the probe is small compared to the length of variation in potential that the secant slope truly approaches the tangent slope. Therefore, a better way to phrase the effective field in this case is to discard the division by distance (which is constant for a given chromophore) and only treat the difference potential as the relevant parameter.

From another perspective, it has been frequently observed via gas phase studies that the peak maximum of the action spectrum, which in theory corresponds to its absorption spectrum, is the same as that measured in the wild-type GFP environment [63]. Similarly, in our R96M/T203Y test (Figure 11), where we deliberately weaken the

interaction on each side of the chromophore, the absorption maximum is again close to that of wild-type GFP. Interpreting these three chromophore environments with the same field parameter seems unphysical, since it does not really reflect the true *localized* field distribution imposed by the environment as its name suggests, and the interaction patterns are drastically different in these cases. This interpretation can also lead to further confusion, suggesting that the wild-type protein environment acts as a vacuum. Thinking instead in terms of driving forces avoids these conundrums because the differential stability of negative charges on the two rings not only emphasizes the tug-of-war nature of chromophore-protein interactions pertaining to their influences on the chromophore's color, but also reflects the ground state electron distribution of the chromophore (Section S4, Equation S24). Similar tug-of-war phenomena and the use of driving forces have been described in the context of several biologically relevant systems involving charge transfer [64][65]. Additionally, driving forces can unify outcomes from modifying either the chromophore (variants) or the protein environment (mutants). Both modifications modulate the electron density of the chromophore through the differential negative charge stability [66], while only the latter is achieved via electrostatic interactions.

The use of driving force to characterize the protein environment can be powerful, since the effects from several point mutations should be reflected in a linearly additive manner based on the following two reasons. First, the contribution from each mutation should be additive due to the superposition principle of electric fields F_{eff} and the linear Stark effect of the driving force (Equation S42), as long as the mutated residue's atomic coordinates in each single amino acid mutant are retained in the double mutant. Second, $\Delta\mu_{CT}$ is constant in different environments, since diabatic states remain constant by assumption. Therefore, we should be able to infer properties of constructs with multiple mutations from the corresponding single mutants (Subsection 2.9 in the main text).

An obvious alternative way to formulate a simple model for the electrostatic interaction between an electronic system and its environment is to break the chromophore into smaller moieties that are more amenable to the application of Equation S35. However, the total charge is not conserved ($\Delta q \neq 0$) within each moiety, and we have to consider the monopolar contribution (Equation S40), which essentially requires reverting back to Equation S38, and minimal insight would be gained.

It is instructive to point out that the failure of the dipole approximation is not only limited to the color tuning problem of pigments in condensed-phase environments, in which the role of the difference dipole moment is largely emphasized. The same exact issue also arises when evaluating excitation energy transfer (EET) rates among pigments [67][68]. Within the framework of the dipole-dipole approximation, namely the Förster theory [69][70], the EET rates are associated with the pigments' transition dipole moments, which can be conveniently determined through absorption or fluorescence measurements. Because the Förster theory is formulated in terms of parameters that are readily accessible via experiments, it is always a first choice for modeling EET rates. However, the dipole-dipole approximation demands intermolecular distances to be larger than the molecular dimensions of pigments, which is severely violated in chromophore aggregates and by the pigment arrangements in many photosynthetic complexes [67][68], thus rendering the parameters (orientations, transition dipole moments, intermolecular distances) ill-defined. Such a breakdown is best illustrated by the examples of EET between a carotenoid and a nearby bacteriochlorophyll in the light-harvesting antenna complex 2 (LH2) [70][71][72][73][74] and between a bacteriochlorophyll and the special pair in the bacterial reaction centers [75][76][77][78]. In both cases, a forbidden transition for either the donor or acceptor is involved (i.e. S_0 to S_1 for the carotenoid and ground state to the upper exciton state for the special pair) due to symmetry, and the Förster theory would predict EET rates orders of magnitude smaller than experimental results. Further investigation using transition densities (cf. the difference density in Equation S38) obtained from high-level calculations reveals that the dipole-dipole approximation tends to overlook molecular details due to averaging (cf. Equation S40) and is thus unsuitable for cases with short intermolecular distances. In other words, by virtue of their small separations, the pigments can sense the details in charge distributions of their neighbors via local interactions instead of the averaged transition dipole moment. The same observation was made by Fritz London when estimating strengths of intermolecular interactions using ground state dipole moments [79]:

“... it is clear that even the dipole terms of this power series must turn out to be quite inappropriate if one has to consider oscillators of some length extended over a large

region of a chain molecule. Another molecule eventually would interact chiefly with one end of such a long virtual oscillator, and this situation would be completely distorted if one were to represent the oscillator by a decomposition into point-form multipoles, all located in the center of the molecule. It would obviously be much more appropriate in this case to represent each oscillator by several distinct poles, “monopoles,” of different sign, suitably located in the molecule, thus directly taking account of the actual extension of the oscillator in question. Such a description may be advantageous and preferable to the multipole representation even in the case of small molecules.”

London’s proposal of dissecting the charge distribution into monopoles was discussed earlier in this section (*vide supra*). However, by resorting to the more sophisticated high-level calculations, one also discards the elegance and intuitiveness of the Förster theory, as highlighted in the review by Scholes and Fleming [68]:

“Both qualitative and quantitative aspects of this issue can be handled by explicitly calculating the Coulombic interaction between the transition densities of donor and acceptor, but now these must be obtained from electronic structure calculations, thus breaking the reliance on only experimentally determined quantities, which is the great strength of the Förster theory.”

S9 Justification of the Marcus–Hush Model from Resonating Valence Bond Theory

Theory

It is possible to derive and justify the use of two limiting electronic states in the Marcus–Hush model from resonating valence bond theory [80][81], in which all possible valence-bond states are considered rather than those with low-lying energies. Here we just consider the anionic allylic states, which have been computationally verified to reasonably capture the photophysical behavior of the GFP chromophore [82]. The π -system of an allylic anion consists of three p orbitals and four electrons, leading to nine electron-paired valence-bond states ($S_z = 0$, Figure S5).

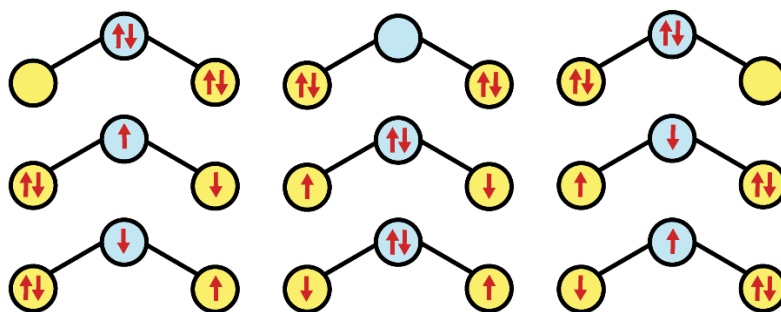


Figure S5. The nine electron-paired valence-bond states in a three-orbital-four-electron model. The yellow and blue circles represent the rings and the bridge of the chromophore (or the terminal and middle carbon atoms of the allylic anion), respectively.

To solve for the resulting singlet eigenstates and their corresponding energies, of which the lowest two will be S_0 and S_1 , we need to determine the energies of the valence-bond states and their couplings. We will follow the simplified scheme of the Hückel–Hubbard model [83], in which the Hückel part (tight-binding model) accounts for the resonance integral β between two neighboring p orbitals (hopping or transfer integral in the condensed matter literature [84]) and the Hubbard part deals with the repulsion J between two electrons if they reside on the same p orbital (Figure S6A). For the GFP chromophore, the repulsion within the ring (terminal p orbitals in the allylic model) should be smaller than that on the bridge (the middle p orbital in the allylic model), so the repulsion term is separated into J_r and J_b , respectively. The energies of the valence-bond states can thus be ranked accordingly (Figure S6B).

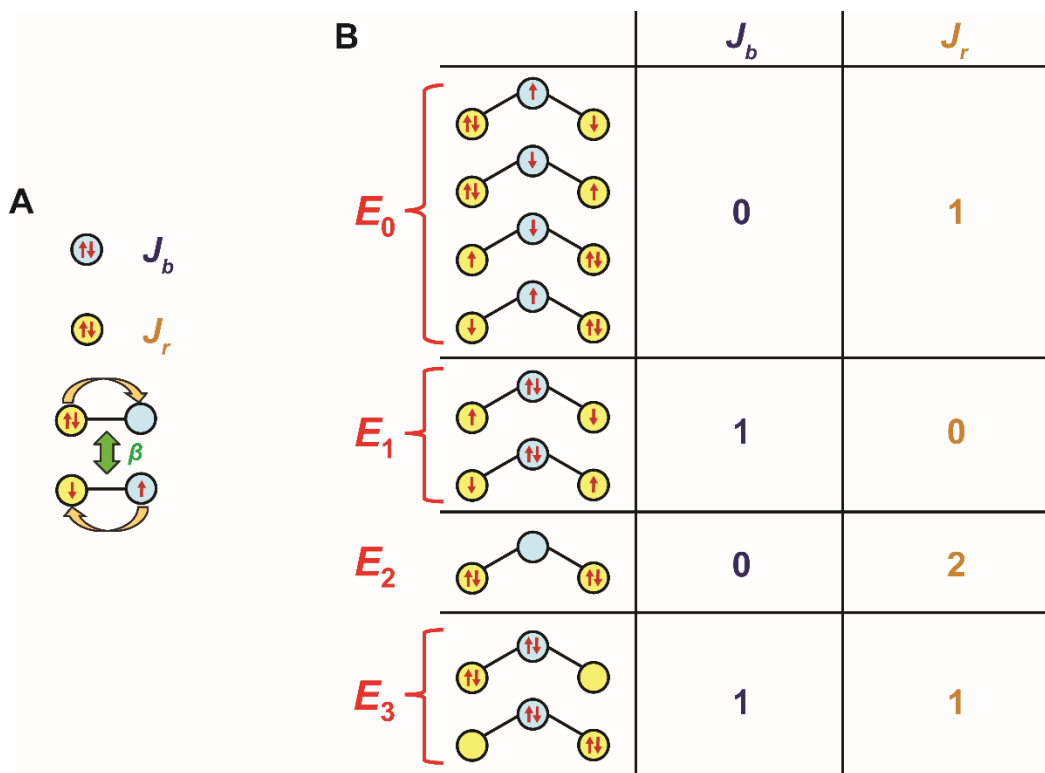


Figure S6. (A) The three parameters for the Hückel–Hubbard model. (B) Energies of the valence-bond states according to the Hückel–Hubbard model in the order of ascending energies. Note that the driving force is not considered here, but it can be incorporated in a straightforward manner.

From the nine valence-bond states, a 9×9 full configuration interaction matrix can be constructed (Figure S7):

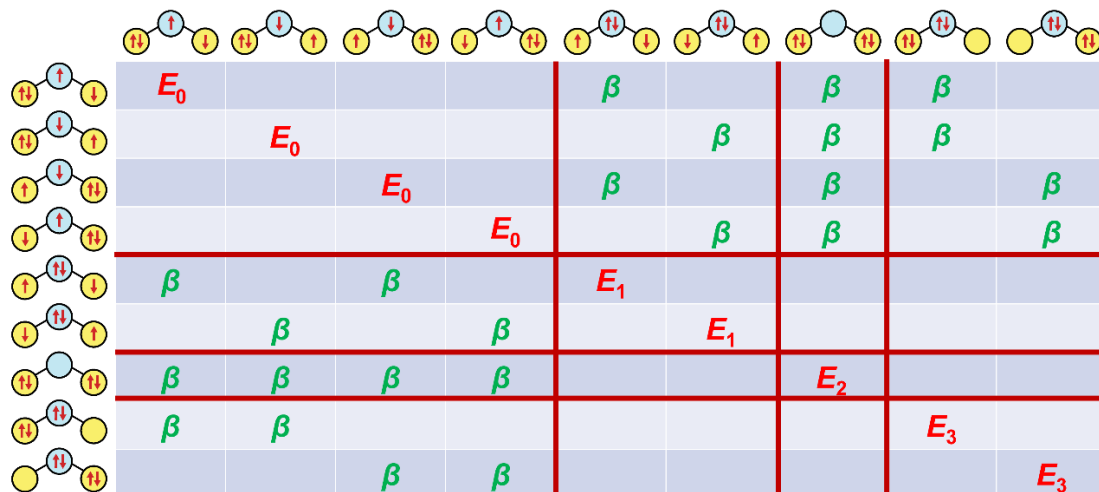


Figure S7. The 9x9 full configuration interaction matrix. The entries are demarcated by red lines to reflect isoenergetic states (Figure S6B). The empty entries are 0 and not shown for clarity.

We expect the linear combinations of the nine states to result in six singlet states and three triplet states (with $S_z = 0$), but the latter can be decoupled to reduce the matrix dimension. Specifically, a pair of singlet and triplet states can be constructed by combining states with the same electron (un)pairing pattern (Figure S8):

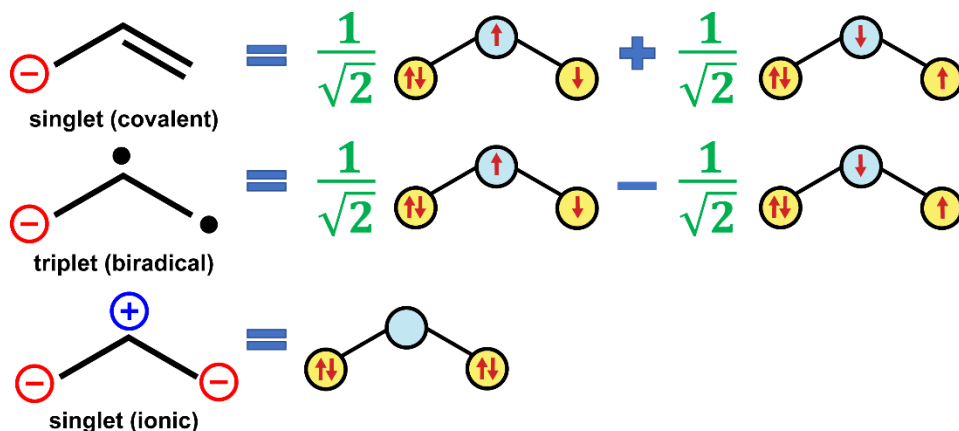


Figure S8. Singlet and triplet states from the combinations of valence-bond states.

After the basis transformation from valence-bond states to singlet and triplet states, the resulting 6x6 matrix spanned by the singlet states is shown in Figure S9:

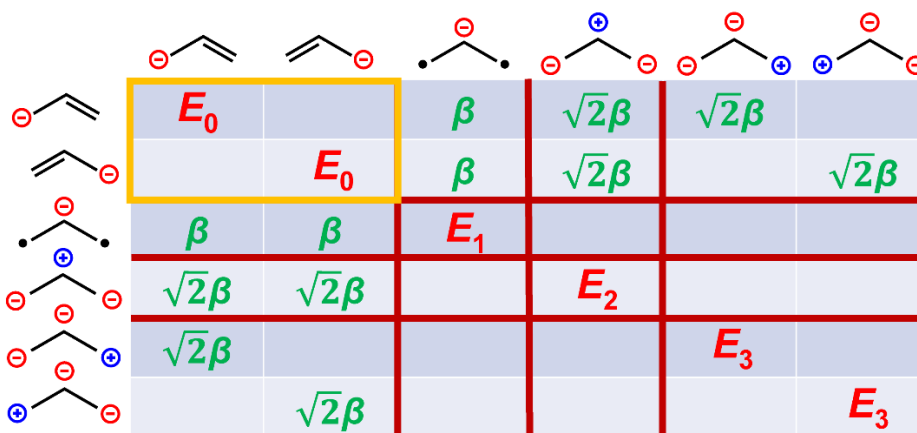


Figure S9. The 6x6 matrix spanned by the singlet states. The entries are demarcated by red lines to reflect isoenergetic states, while the block with the two lowest energy states is highlighted with a yellow border. The empty entries are 0 and not shown for clarity.

Note that we can now diagonalize this matrix to obtain the desired eigenvalues. However, instead of using the brute-force approach, we search for the possibility to reduce this matrix down to a 2x2 matrix while keeping the lowest two eigenvalues. We can then directly express the empirical electronic coupling V_0 in terms of fundamental parameters β and J without having to extract them from the eigenvalues. This approach is the effective-Hamiltonian method [85], and the resulting couplings are effective (renormalized) parameters. The goal is to recapitulate the low-energy physics with a manageable number of degrees of freedom.

The specific method adopted here is the canonical Van Vleck transformation [86]. We first find a unitary matrix that transforms the 6x6 matrix to a block-diagonalized matrix with a 2x2 and a 4x4 block. Since there is no coupling between the resulting two and four states, we can safely remove the latter and describe the desired physics with just two states. Note that this method is also known as quasi-degenerate perturbation theory, so if there are more than two low-lying valence-bond states that are close in energy, it is not possible to block-diagonalize the matrix into a 2x2 block. Therefore, we first need to figure out the relative energies of the singlet valence-bond states based on the number of repulsion terms involved (Figure S6B). For simplicity, we neglect the driving force caused by the environment for now, which can be easily restored. The lowest two states with energies $E_0 (= J_r)$ are degenerate and analogous to those hypothesized using Marcus–

Hush theory (Figure 6A). The third lowest state is the biradical state with energy $E_1 (= J_b)$, while ionic states have higher energies (E_2 and E_3). The energy assignment confirms that it is reasonable to expect a 2x2 effective Hamiltonian.

The use of perturbation theory suggests that the transformed matrix will only be roughly block-diagonal, and the transformation can be performed to any desired accuracy. If a second-order Van Vleck transformation is incorporated, we obtain a 2x2 effective Hamiltonian:

$$H_{eff} = \begin{pmatrix} E' & V_0 \\ V_0 & E' \end{pmatrix} \quad (\text{S43})$$

where

$$E' = -\beta^2 \left(\frac{1}{E_1 - E_0} + \frac{2}{E_2 - E_0} + \frac{2}{E_3 - E_0} \right) = -\beta^2 \left(\frac{1}{J_b - J_r} + \frac{2}{J_r} + \frac{2}{J_b} \right) \quad (\text{S44})$$

and

$$V_0 = \beta^2 \left(\frac{1}{E_1 - E_0} + \frac{2}{E_2 - E_0} \right) = \beta^2 \left(\frac{1}{J_b - J_r} + \frac{2}{J_r} \right) \quad (\text{S45})$$

Interestingly, in the original singlet Hamiltonian (Figure S9), the two lowest energy valence-bond states are not coupled (as highlighted by a yellow border). In order to recapitulate the coupling among other valence-bond states in terms of the two low-lying states, an effective coupling V_0 is generated and has the form of superexchange in charge transfer [87] (see also Creutz, Newton, and Sutin's approach for mixed-valence systems [88]). Its proportionality to β^2 suggests that the two states can "communicate" via two nearest-neighbor electron hopping interactions, while the denominator gauges the ease of such hopping via the energy difference between the intermediate and initial states. Since the absolute energy does not matter in Equation S43, it is identical to the Marcus–Hush model without nuclear degrees of freedom and to resonance color theory [31].

One additional insight we gain from this exercise is that the basis of the effective Hamiltonian (Equation S43) is not exactly the original two low-lying valence-bond states, as we claim throughout the main text. Rather, the states are slightly "contaminated" by the second lowest biradical state from the unitary transformation, so the charge localization for the basis states is slightly closer to the bridge instead of exactly at the oxygen atoms. The degree of contamination is proportional to the coupling V_0 . This is a rationale for why the magnitude of the difference dipole moment between the basis states

$\Delta\mu_{CT}$ tends to be smaller than the value expected from an elementary charge moving across a distance d_{CT} between two hypothesized charge localization centers, especially for Class III or Class II/III borderline mixed-valence compounds such as the GFP chromophore and the Creutz–Taube ion [89]. Within the two-state framework, another possible explanation was explored by Olsen and McKenzie [31] and involves the breakdown of the Mulliken–Hush approximation (Section S4), i.e. the diabatic states are no longer dipole eigenstates. Since an additional mixing angle is required to transform the true dipole eigenstates to the Marcus–Hush diabatic states, we expect a mismatch in determined parameters from observables that require this approximation for interpretation and those that do not. Explicitly, we expect the estimation of V_0 to be different between the Stokes shift-absorption maximum correlation plot (Equation S17, Figure 8) and the absorption maximum-Stark tuning rate correlation plot (Equation S23, Figure 9), because the Mulliken–Hush approximation is unnecessary for the former. However, since we found no discrepancy, this approximation is valid, and we believe that the contamination from other valence-bond states leads to the small observed $\Delta\mu_{CT}$ in the case of the GFP chromophore.

S10 Equivalence of the Su–Schrieffer–Heeger Model and the Marcus–Hush

Model

Another famous model that couples electronic degrees of freedom to nuclear distortion is the Su–Schrieffer–Heeger (SSH) model Hamiltonian, which was originally proposed to explain behaviors of polyacetylene [90]:

$$H = - \sum_{ns} t_{n+1,n} (c_{n+1,s}^\dagger c_{ns} + c_{ns}^\dagger c_{n+1,s}) + \frac{1}{2} \sum_n K (u_{n+1} - u_n)^2 + \frac{1}{2} \sum_n M \dot{u}_n^2 \quad (\text{S46})$$

The model is formulated in the atomic basis n and the electronic spin is denoted s (up or down). The electronic degree of freedom is treated quantum mechanically via the creation and annihilation operators c^\dagger and c , while the coupled nuclear degree of freedom, which is BLA ($u_n = (-1)^n u$), is treated classically in terms of atomic displacement u . The last two terms account for the nuclear potential and kinetic energies, while the first term allows electron hopping between nearest-neighbor atoms, characterized by the parameter t , which is analogueous to β in the Hückel model (Section S9). The vibronic coupling α is encoded in the dependence of t on the nuclear displacement:

$$t_{n+1,n} = t_0 - \alpha(u_{n+1} - u_n) \quad (\text{S47})$$

where t_0 is the hopping integral between atoms without nuclear distortion. The vibronic-coupling-dependent term shows that t becomes larger when the neighboring atoms are closer, and a linear approximation is valid since the magnitude of BLA is generally small [90]. In other words, the SSH model is the Hückel model coupled to nuclear motion. Because this model is semiclassical and incorporates vibronic coupling, we expect it to be equivalent to the Marcus–Hush formalism without driving force. The only difference is the basis involved: SSH is written with the atomic basis, while Marcus–Hush incorporates diabatic states.

Typically for polyacetylene, which is approximated with infinite atoms, the SSH Hamiltonian is solved first by transforming into reciprocal space to obtain the one-electron energy levels [84]. However, for an effective three-atom model (the allylic model), such practice is unnecessary since the matrix is only 3×3 , and no translational

symmetry is involved. We can represent the three-atom form of Equation S46 with the following matrix:

$$H_{SSH}(u) = \begin{pmatrix} 4Ku^2 & -t_0 + 2\alpha u & 0 \\ -t_0 + 2\alpha u & 4Ku^2 & -t_0 - 2\alpha u \\ 0 & -t_0 - 2\alpha u & 4Ku^2 \end{pmatrix} \quad (\text{S48})$$

It is easily seen that the model reduces to the Hückel theory when u becomes 0. The three corresponding “molecular orbital” energies are:

$$E_1 = 4Ku^2 - \sqrt{2t_0^2 + 8\alpha^2u^2}, E_2 = 4Ku^2, E_3 = 4Ku^2 + \sqrt{2t_0^2 + 8\alpha^2u^2} \quad (\text{S49})$$

in the order of increasing energy. When four electrons are filled, the ground and excited state (S_0 and S_1) energies become:

$$E_{S_0}(u) = 2E_1 + 2E_2 = 16Ku^2 - 2\sqrt{2t_0^2 + 8\alpha^2u^2} \quad (\text{S50})$$

and

$$E_{S_1}(u) = 2E_1 + E_2 + E_3 = 16Ku^2 - \sqrt{2t_0^2 + 8\alpha^2u^2} \quad (\text{S51})$$

respectively. Rewriting Equation S8 without driving force yields:

$$V_{1,2}(q) = \frac{\bar{v}}{2}(q^2 + \delta^2) \mp \frac{1}{2}\sqrt{4V_0^2 + (2\bar{v}\delta q)^2} \quad (\text{S52})$$

Given that only relative energies matter, we see a strong resemblance of terms within the square root between Equations S50/S51 and S52: both t_0 and V_0 depict the electronic coupling, both α and $\bar{\delta}$ characterize the vibronic coupling strength, and both u and q describe the nuclear displacement. Based on this analysis, we can arrive at the same answers regardless of the starting model, which also justifies the validity of using only two diabatic states.

S11 Stark Spectra and Fitting of GFP Mutants and Chromophore Analogues

Measuring the Stark tuning rates of the GFP chromophore in different environments allows us to link electron redistribution upon photoexcitation to electrostatic color tuning. This information can be obtained through electronic Stark spectroscopy, in which the absorbance change of an isotropically immobilized sample in the UV–Visible range is recorded under an externally applied electric field and fit using Equation S1, the classical sum-of-derivative Stark analysis [16]. The Stark tuning rate can then be extracted from the coefficient of the second-derivative component (Equation S2). Most data sets, including those from mutants and variants of GFP and Dronpa2, obey the classical Stark analysis. In some cases, this analysis is complicated by band overlap, and this is discussed in detail below; in other cases, notably the model chromophores, the underlying sum-of-derivative formulation [16] may not apply, and this is also discussed. Each panel lists the number assigned to the construct in Table S14 and a shorthand for the species.

Mutants and Variants of GFP and Dronpa2: Classical Stark Analysis

All the Stark spectra below (Figure S10) are consistently dominated by the second derivative lineshape of the absorption spectra (blue trace), even though appreciable first derivative components from the fit can also be noticed (green trace) due to the difference polarizability $\Delta\alpha$. This leads to different Stark tuning rates among the mutants and variants due to differences in the chromophore environment (Section S5). Further evidence supporting the dominance of the $\Delta\mu$ contribution is that the 4ω Stark spectra strongly resemble the second derivative of the conventional 2ω spectra [16] (Figure S11), which has also been observed with polyene dyes [91], bacteriochlorophyll *a* in solution, bacteriopheophytin in bacterial reaction centers, and spheroidene in the LH2 antenna complex [92]. Interestingly, even though the crystal structures of Dronpa2 Y63(3-F₁Y) (PDB: 6NQK, [4]) and ih:GFP S65T Y66(3-Cl₁Y) (PDB: 6OFL, this work) indicate the existence of two rotamers for the phenolate moiety of the chromophore, their corresponding Stark spectra fits (Figure S10, panels 44 and 53) do not demand two different sets of electro-optic parameters, suggesting that the two rotamers are spectroscopically indistinguishable.

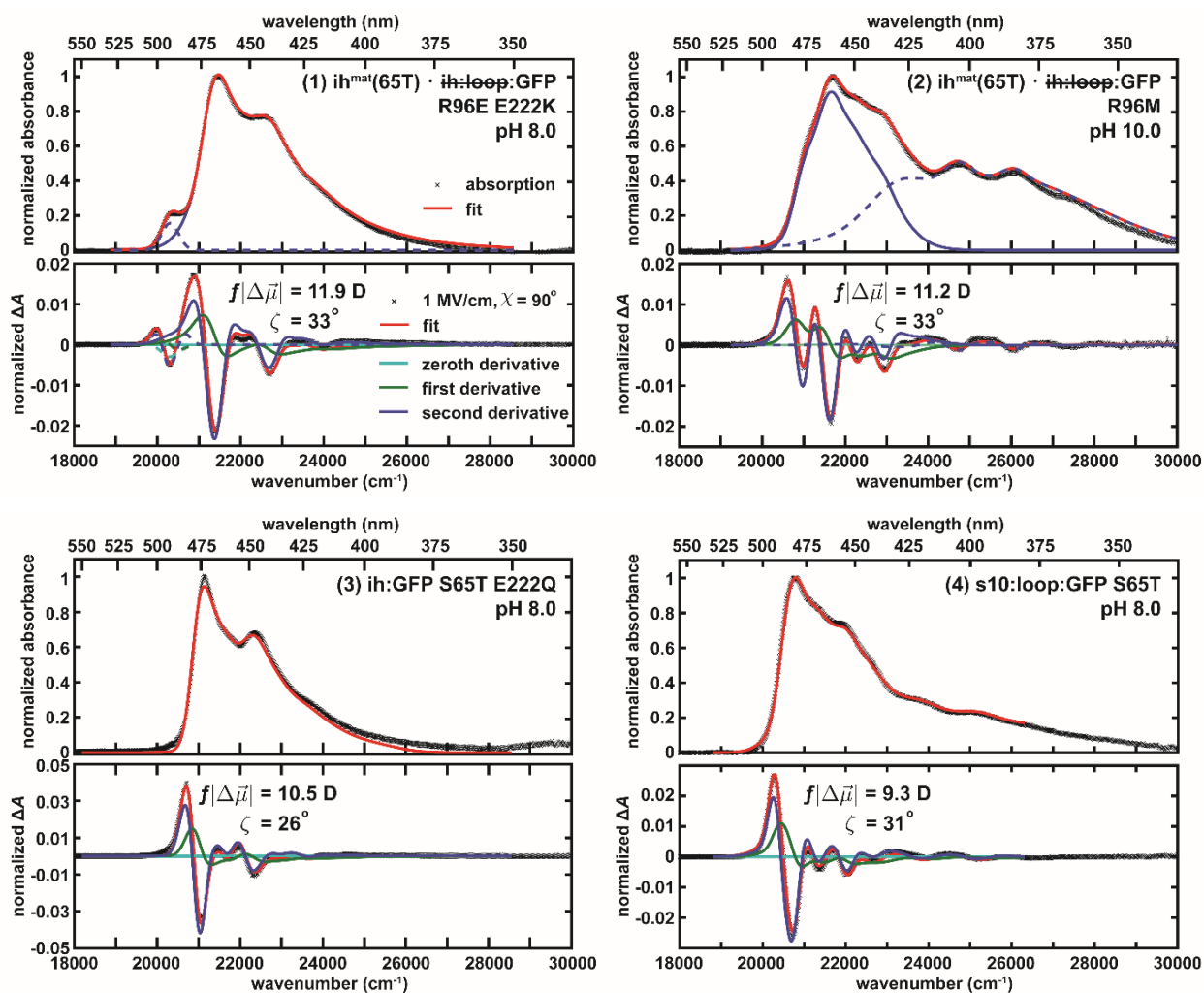
From Figure S10, one can generally see three populations in the order of decreasing absorption energies: the A, B, and I states, whose bands were assigned by comparing a 1.6 K hole-burning study [93] with the canonical scheme for excited-state proton transfer of GFP [94]. In the A and I states, the hydroxyl group of T203 points away from the protonated and deprotonated chromophore, respectively, while in the B state, the hydroxyl group of T203 interacts with the deprotonated chromophore. Further mutant studies show that it is possible to selectively enhance or remove certain states via site-directed mutagenesis [95][96]: E222Q or S65T eliminates the A state at high pH due to decoupling of the proton transfer chain [97]. T203V stabilizes the I state, confirming the difference in T203 orientation between the I and B state environments [98], thus rationalizing the red-shifted I state relative to the B state. These observations are also consistent with the spectra in Figure S10. For example, from the Stark spectrum of s10:loop:GFP E222Q (Figure S10, panel 22), we can conclude that the I state has a smaller Stark tuning rate than the B state, as predicted by the Marcus–Hush model (Equation S23). This is further demonstrated by comparing the Stark spectrum to the second derivative of the absorption spectrum (Figure S12). Mutating T203 into aromatic residues also eliminates the distinction between the B and I states.

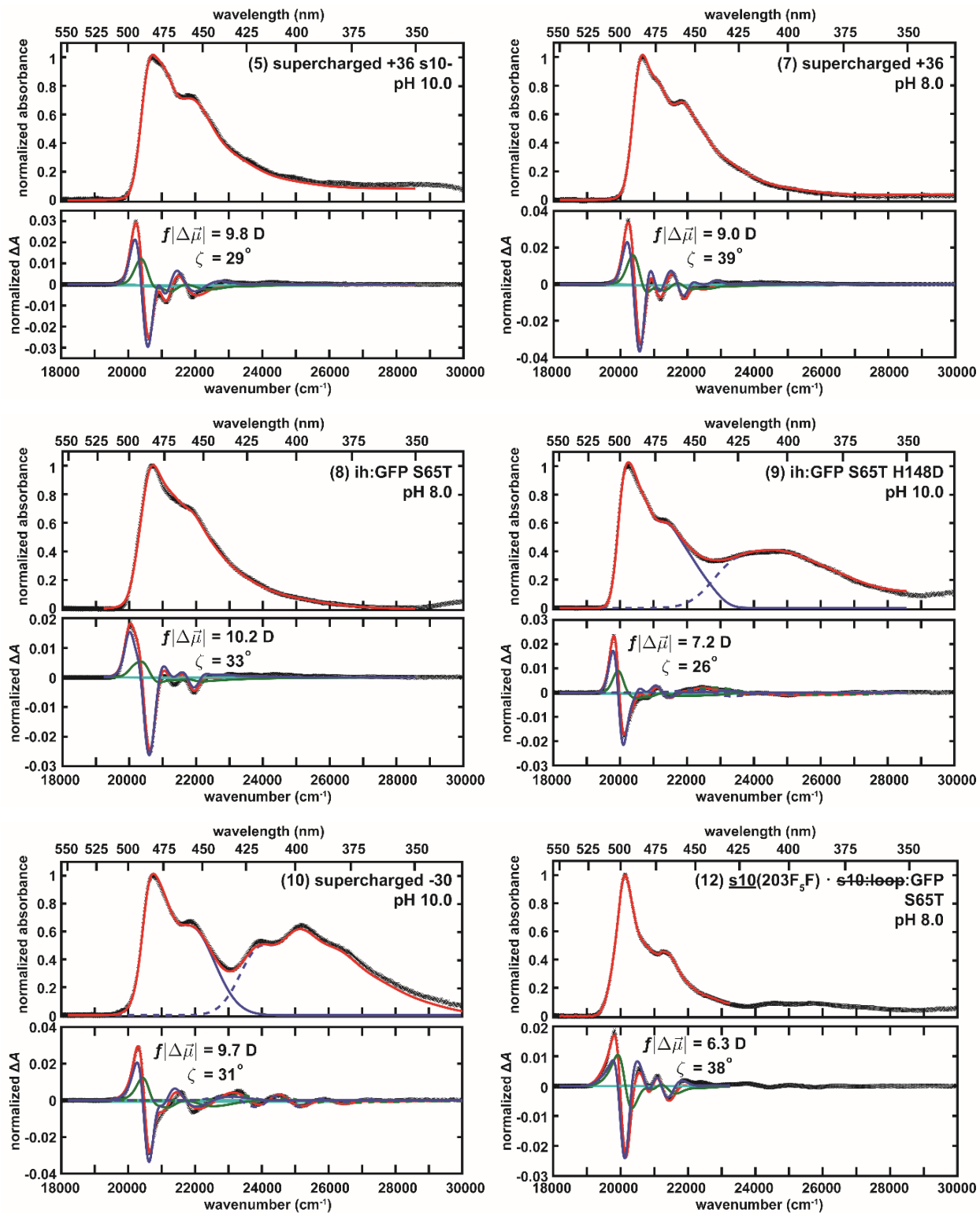
Without prior knowledge of the underlying absorption lineshape for each population, Stark fitting is prone to significant errors from overlapping bands due to cancellations in positive and negative features of derivatives [16]. The wide range of mutations and resultant variations in absorption spectra prove to be very useful as they enable us to isolate the individual populations in a number of cases and more confidently determine their electro-optic parameters. Extensive mutant studies also allow us to reexamine the first set of GFP Stark spectra published two decades ago [37]. To facilitate comparison, we also reproduce the absorption and Stark spectra of avGFP at pH 6.5 (Figure S10, panel 23). The slightly higher A to B state population ratio than that in ref. [37] arises from the difference in concentration between the two samples. Even at the same pH, dimerization enhances the A and I state populations of GFP S65 mutants [96][99]. When normalized, the Stark spectra reported here are larger by a factor of 2 than those in ref. [37], leading to a factor of $\sqrt{2}$ difference in the measured Stark tuning rates, which can likely be explained by the omission of a factor of 2 when processing the raw data (Section

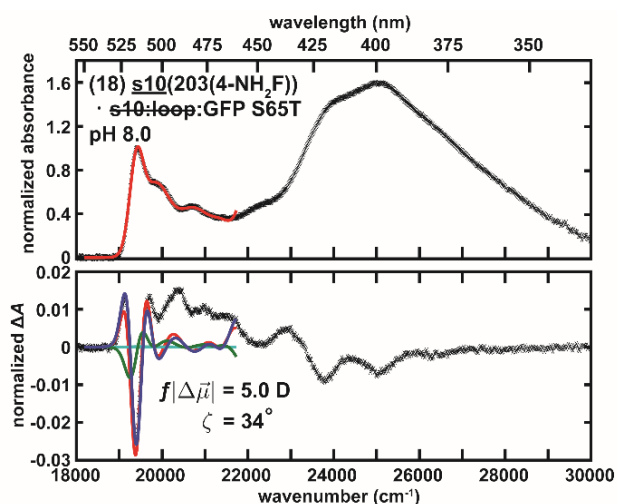
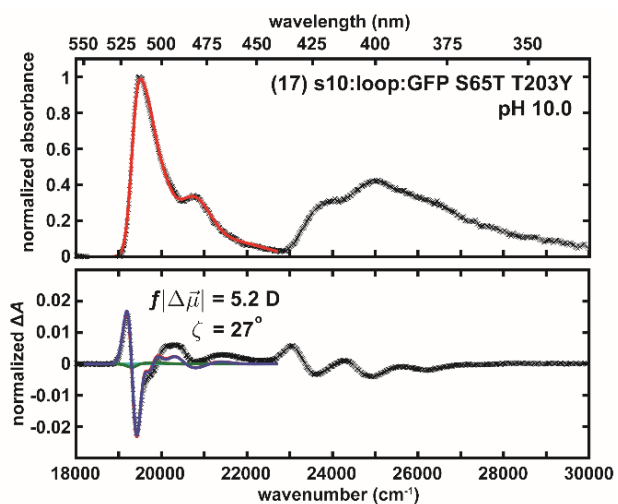
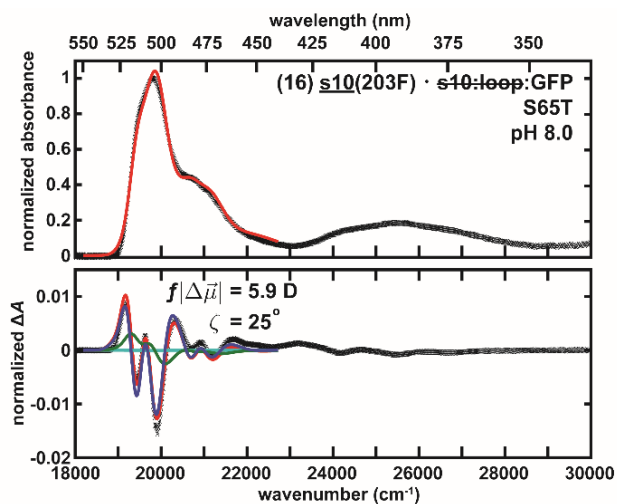
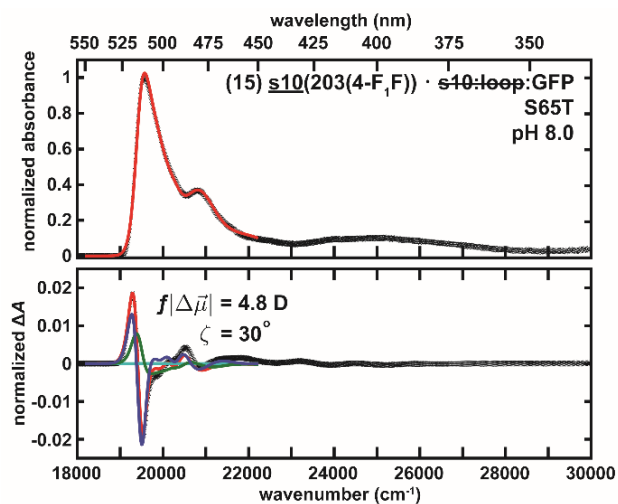
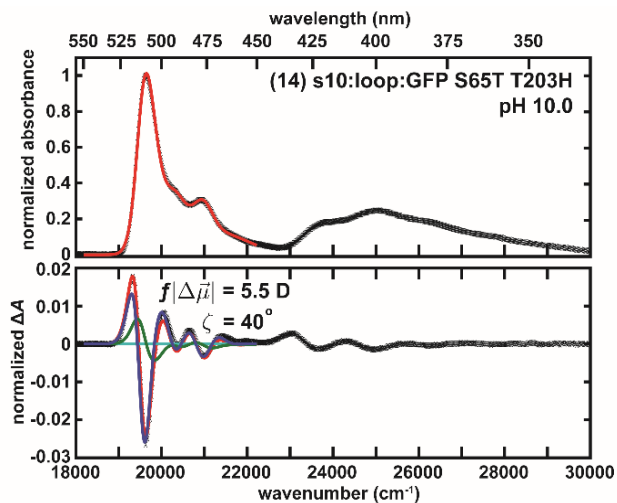
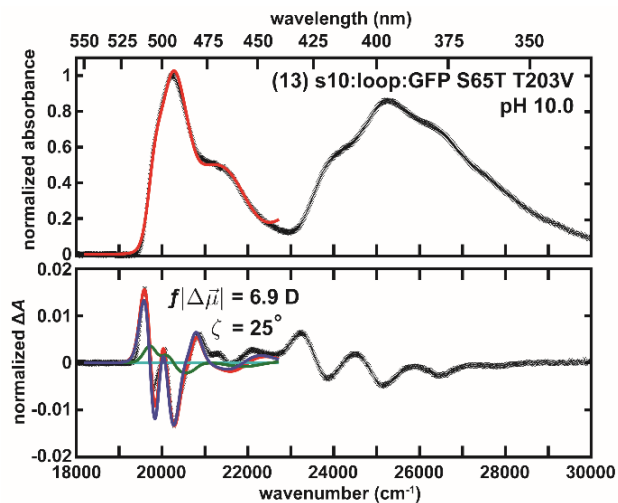
S2). Moreover, based on the conserved vibronic progression across mutants, we can unambiguously assign the vibronic features for the A and B bands (Figure S23), which is not consistent with the vibronic structure assignment from Stark spectroscopy in ref. [37]. Closer evaluation reveals that both A and B bands share almost identical Stark tuning rates such that it is possible to fit the entire Stark spectrum in the visible with just one set of electro-optic parameters for S65 wild-type GFPs (Figure S10, panels 20 and 23). It is not possible to deconvolve the vibronic structure using Stark spectroscopy in this case because any linear combination of bandshapes that accounts for the absorption spectrum can also simultaneously match the Stark spectrum, rendering the determination of individual bandshapes an underdetermined problem. Even with different Stark tuning rates, relying on a simultaneous two-band fit between the absorption and Stark spectra can be misleading, since it still allows for degenerate solutions owing to the partial cancellations of positive and negative Stark features [16][17], discussed further below. Therefore, Stark spectroscopy is not a good primary tool for spectral deconvolution and requires additional input from other techniques, such as fluorescence excitation anisotropy [100] and magnetic circular dichroism [101], or realistic physical models.

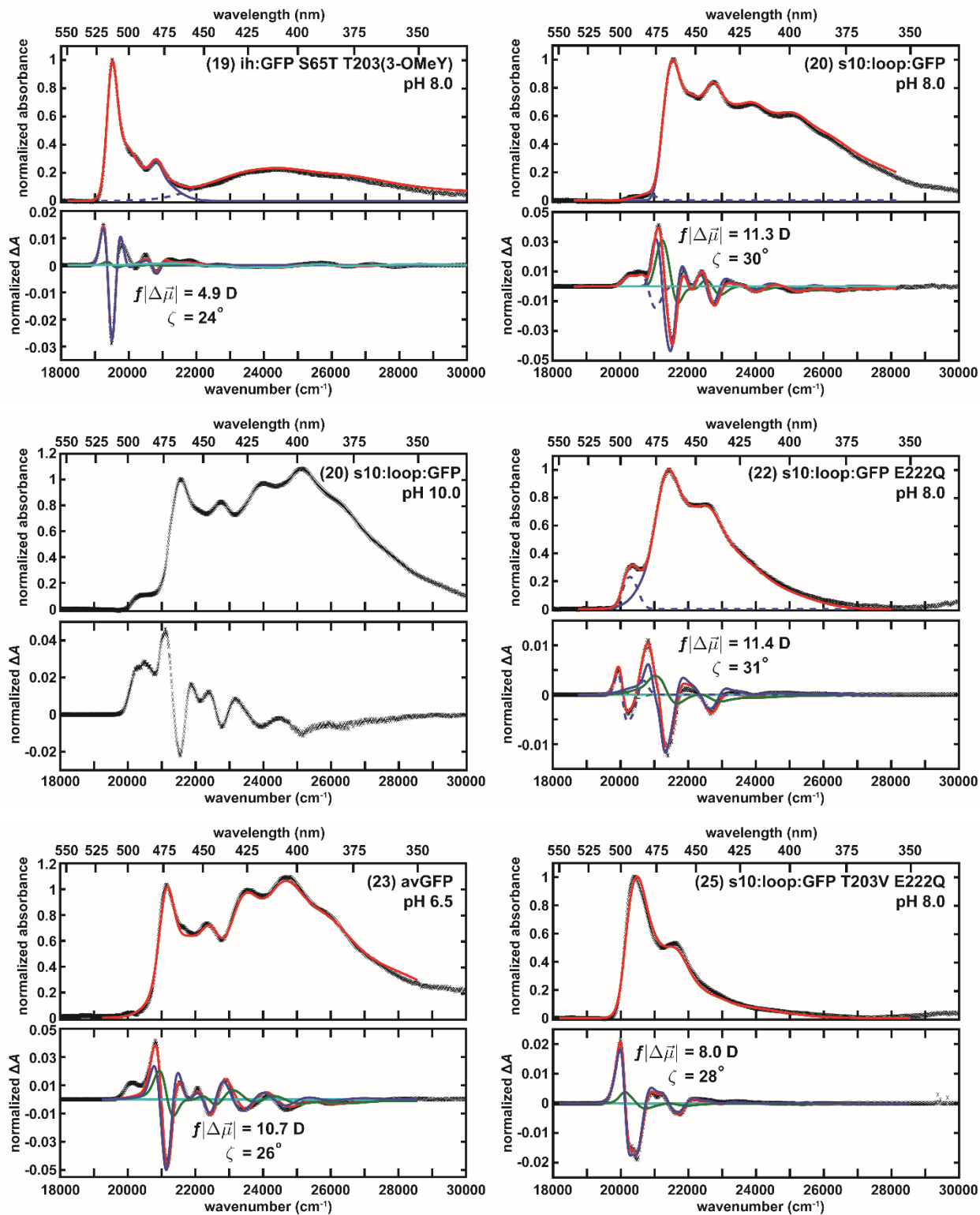
To further isolate the contribution from the B state, one has to shift the population ratio of the A and B states using other means such as pH titration. As an example, we compare the normalized absorption and Stark spectra for s10:loop:GFP S65 at pH 8.0 and 10.0 (Figure S13A). Unexpectedly, the spectra measured at pH 10.0 exhibit a larger A state population, which is due to the aforementioned concentration dependence overpowering the effect from pH [96]. Nevertheless, we still obtain two data sets with different A to B state ratios as shown in Figure S13A. Due to the concentration-dependent dimerization effect on these states (the Stark spectra require mM concentrations due to the short path length), the B state and the A state populations are inversely correlated, while the ratio of A state to I state populations remains constant [96]. This behavior enables us to further separate the I state from the B state. If we assume that the bluer side of the Stark spectra exclusively arises from the A state, since B state absorption should mostly tail off with minimal vibronic features at this region ($24000 - 30000 \text{ cm}^{-1}$), we can scale the pH 10.0 Stark spectrum (Figure S13A, red trace) to the pH 8.0 counterpart (Figure S13A, blue trace) at 24000 cm^{-1} to match the A state contribution.

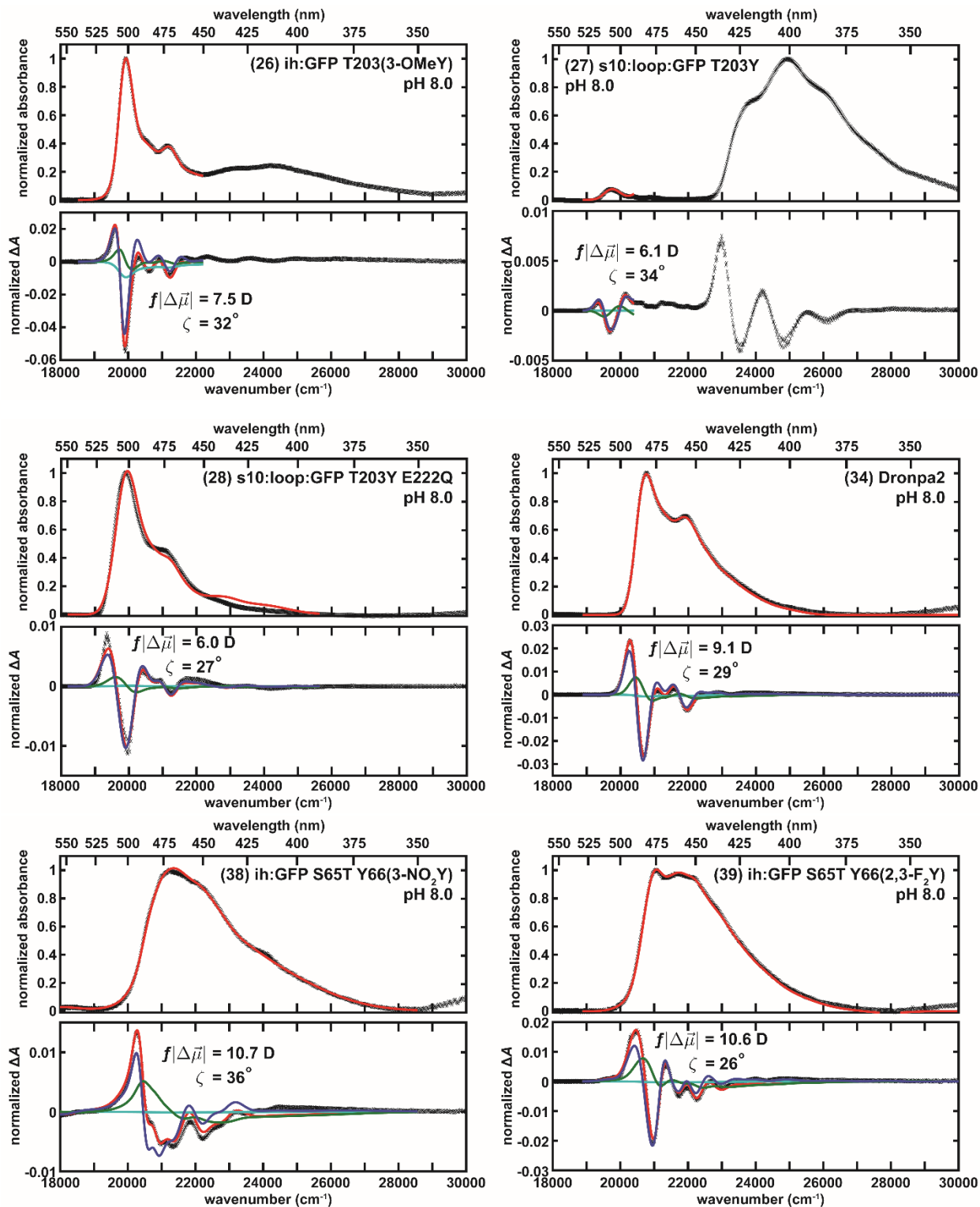
The same scaling factor is also applied to the absorption spectrum. We can then take the difference between the two spectra and remove the contributions from the I and A states, revealing only the B state absorption and Stark spectra (Figure S13A, green traces). Indeed, the B state only inherits two vibronic peaks from the entire s10:loop:GFP vibronic structure, contradictory to the previous assignment from two decades ago. We also compare the normalized s10:loop:GFP absorption and Stark spectra to those of s10:loop:GFP S65T (Figure S13B). The former has a larger magnitude in its Stark spectrum, but this is simply due to its sharper absorption feature rather than possessing a much higher Stark tuning rate (11.3 D vs. 9.3 D for S65 and S65T, respectively).

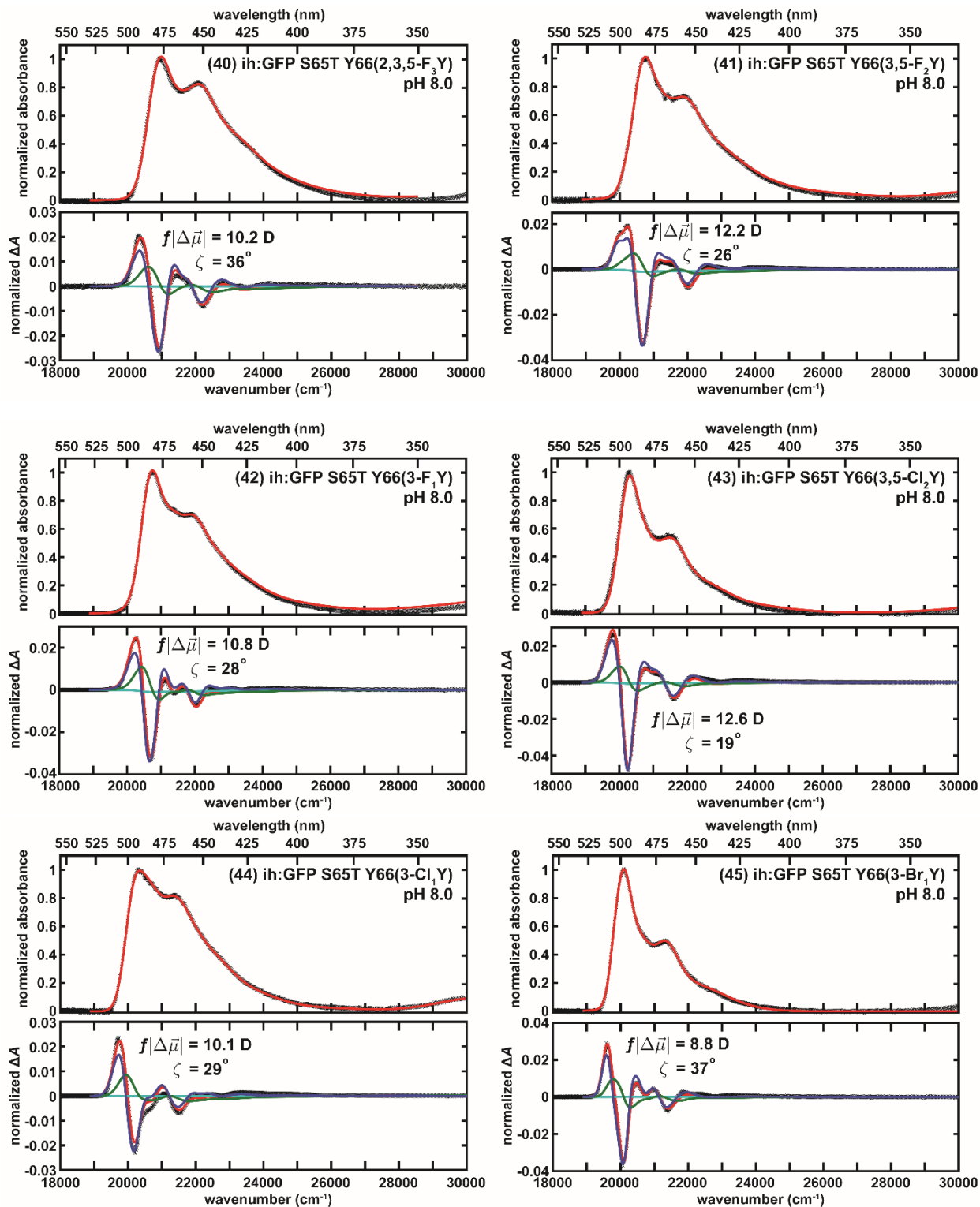


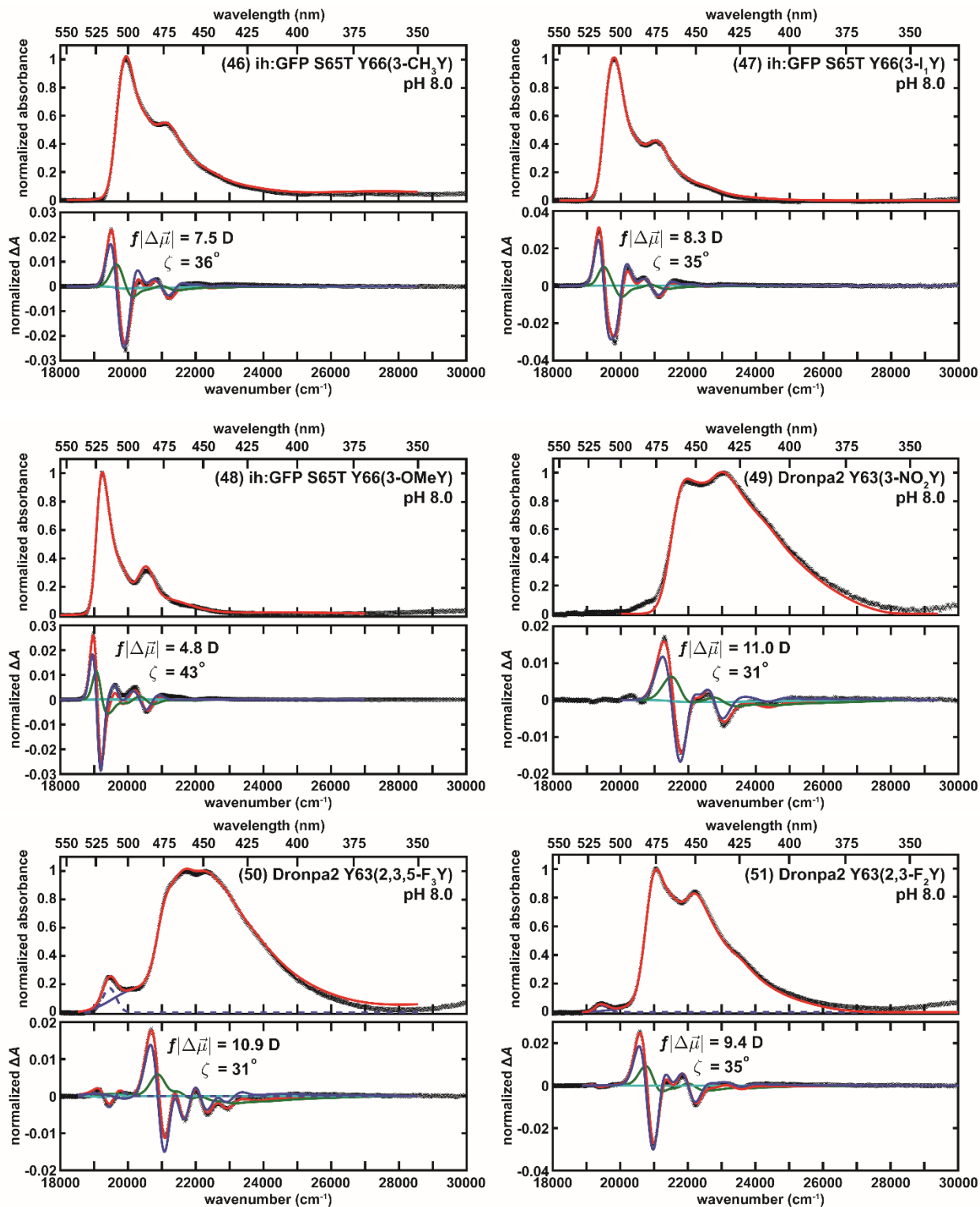


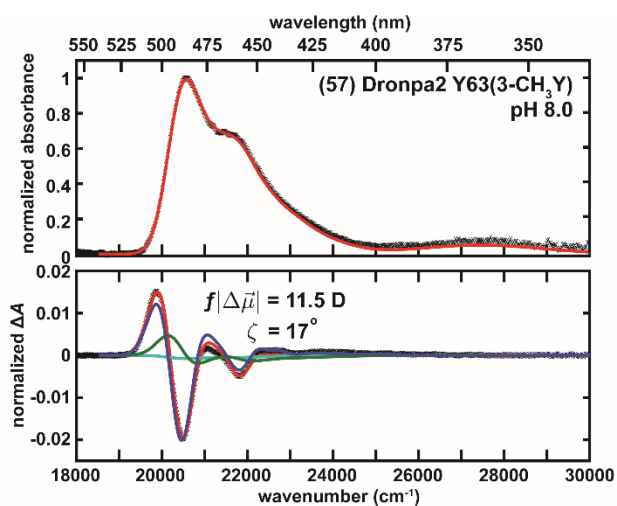
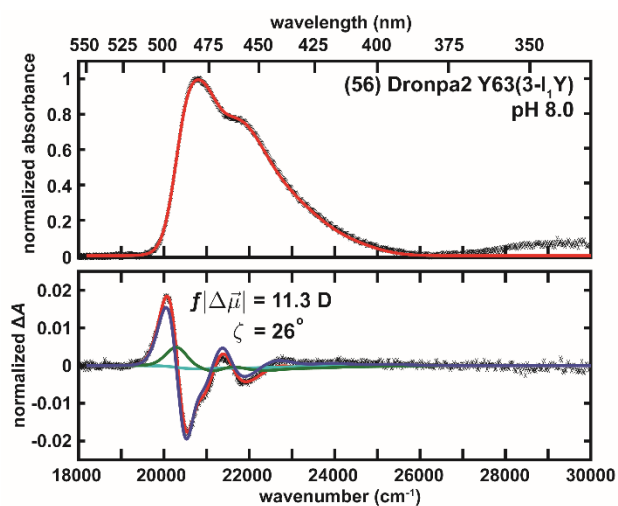
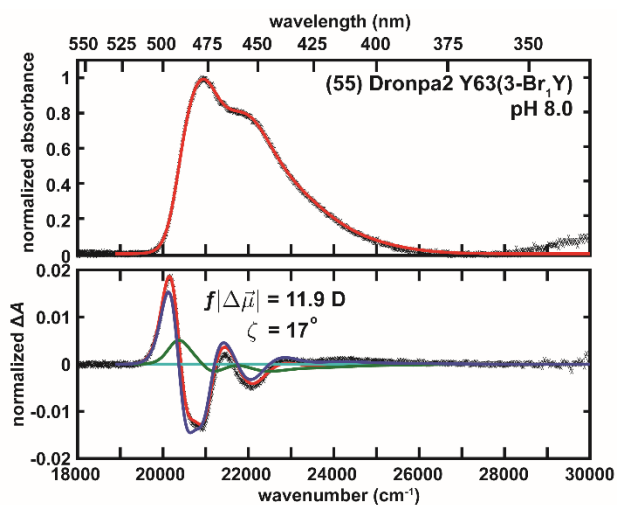
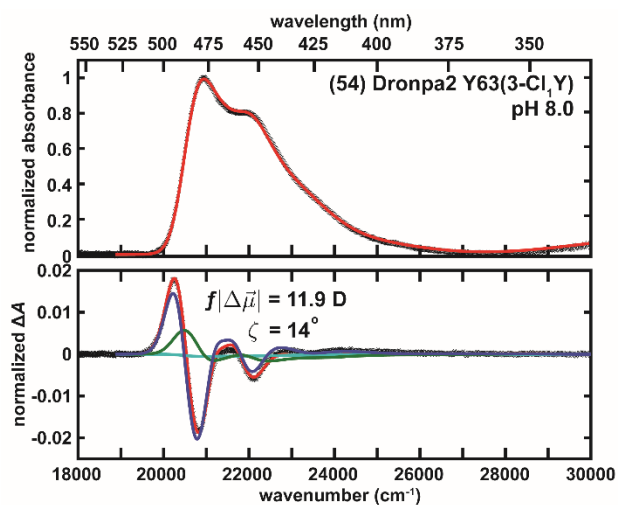
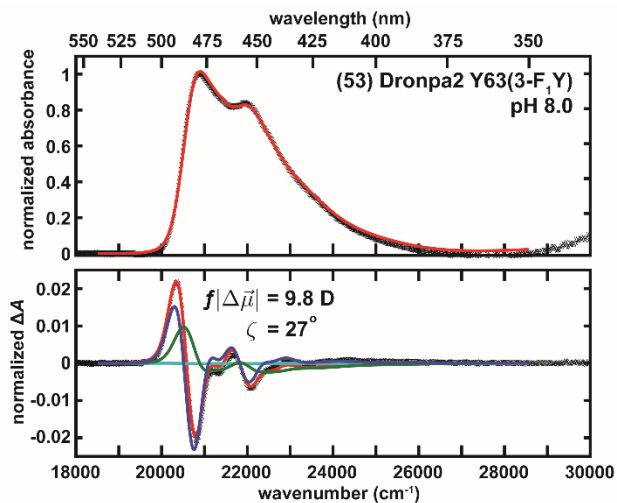
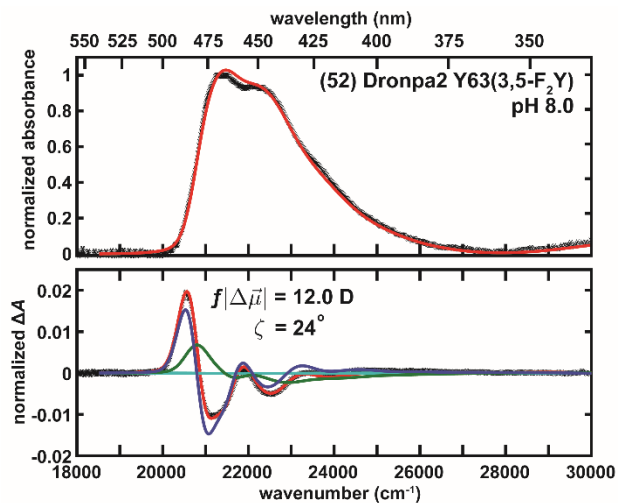












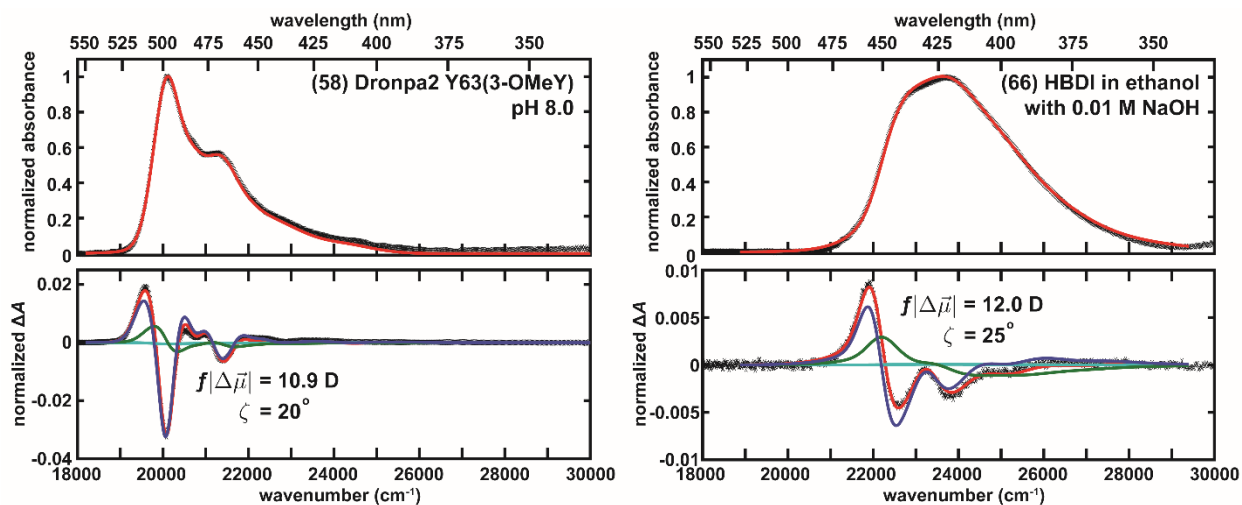
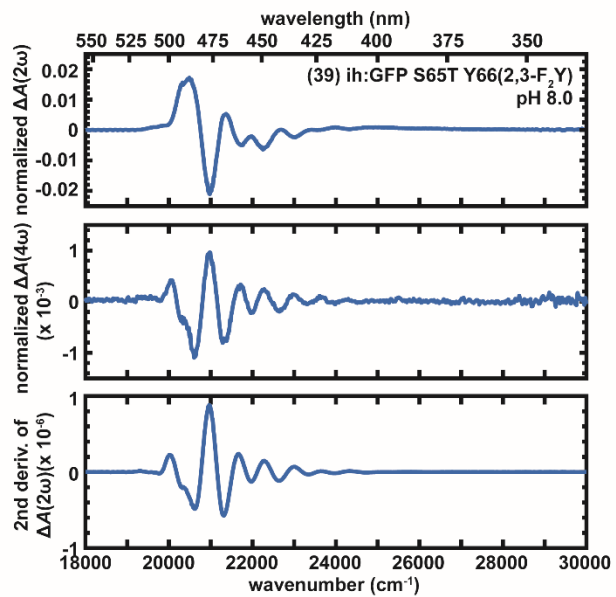
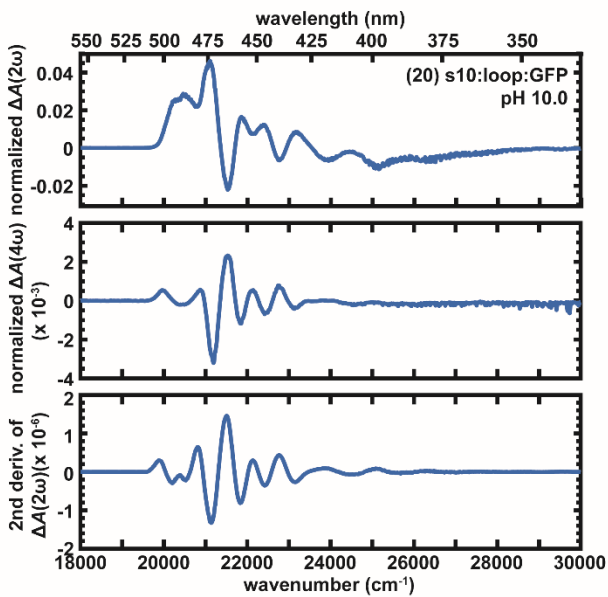
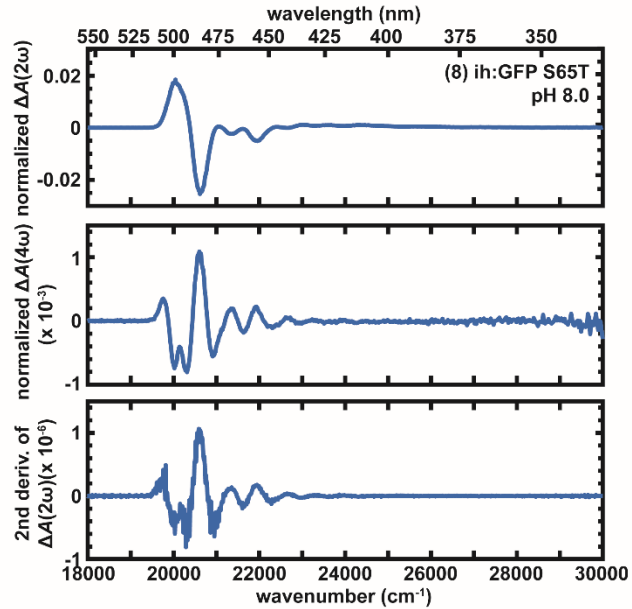
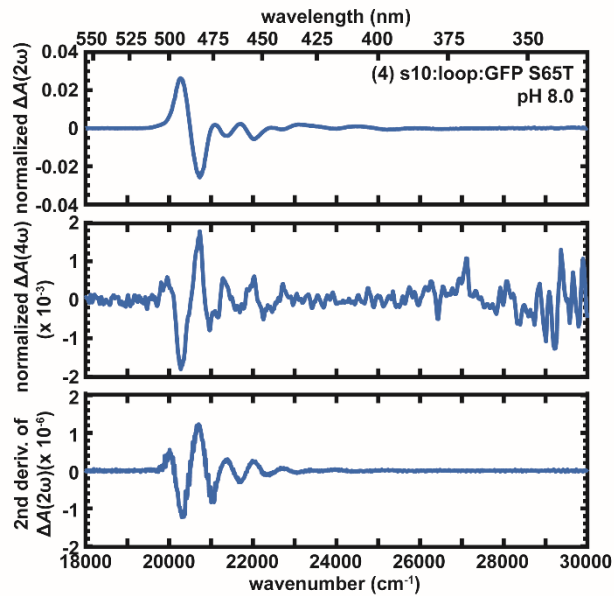


Figure S10. The classical sum-of-derivative analysis for 77 K UV–Vis absorption and Stark spectra. The panels are numbered and listed according to Table S12. The Stark spectra are measured at $\chi = 90^\circ$ and scaled to 1 MV/cm to facilitate comparison. The color scheme of fit lines and data, as shown in the first panel, is consistent throughout the figure. Solid lines represent the band of interest (the anionic state), for which the measured Stark tuning rate ($\pm 5\%$) and the ζ angle ($\pm 5^\circ$) is noted, while dashed lines represent other absorption bands that require simultaneous fitting to extract out electro-optic parameters from the main lower 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 is preferred, even when two bands with distinct origins (such as A and B states) share similar Stark parameters (panels 20 and 23, see also Figure S13). 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 (for example, Figure S12 and discussion that follows). The only protein data set not analyzed using the classical analysis is that of s10:loop:GFP at pH 10.0 (panel 20, pH 10.0) because of the comparable contributions and extensive overlap of A, B, and I bands in the Stark spectrum. Instead, another method is used to obtain the B state Stark tuning rate (Figure S13).



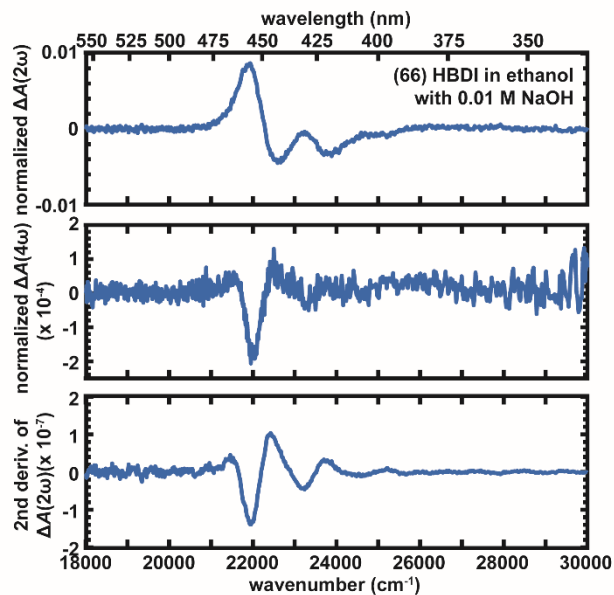
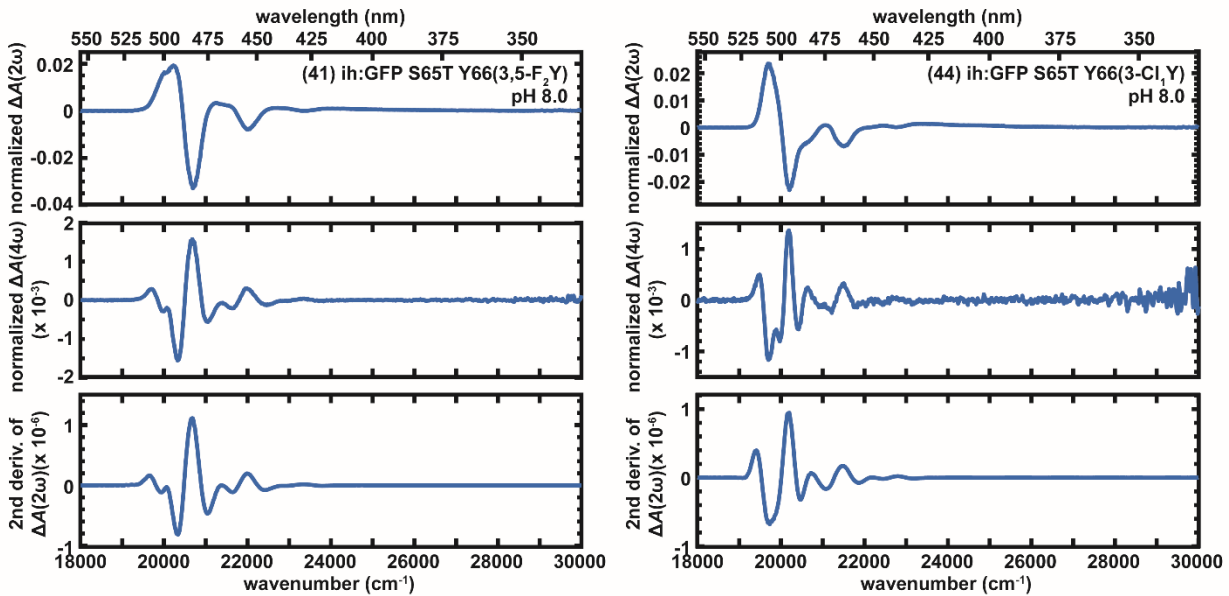


Figure S11. Normalized 2ω Stark spectra, normalized 4ω Stark spectra, and wavenumber-weighted second derivative of the 2ω Stark spectra for GFP mutants, variants, and the model chromophore in ethanol. Note the strong resemblance between the 4ω Stark spectra and the second derivative of the 2ω Stark spectra.

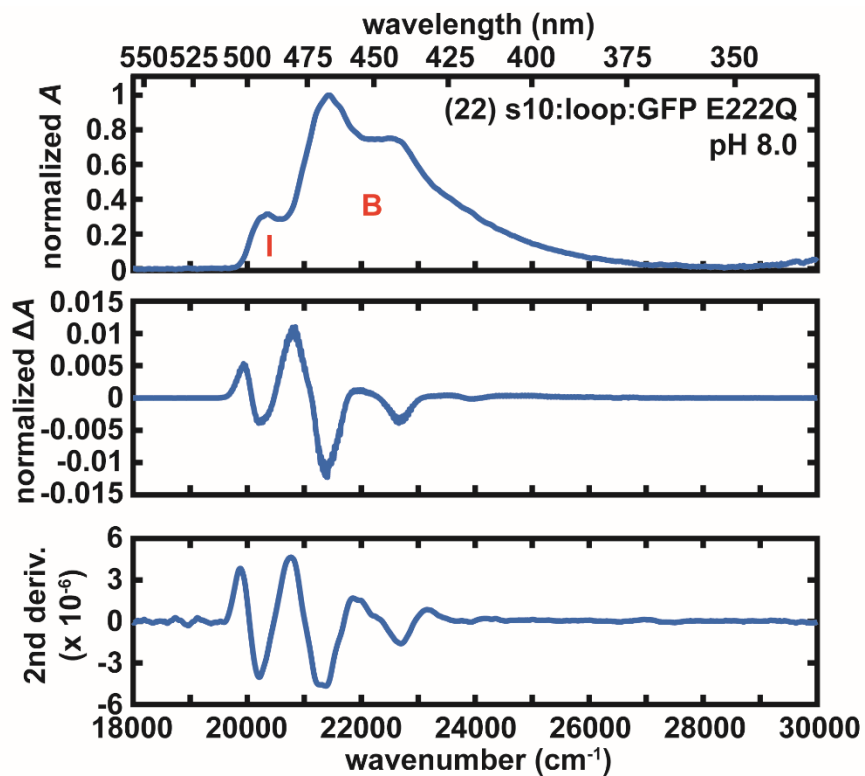


Figure S12. The normalized 77 K absorption spectrum, Stark spectrum (scaled to 1MV/cm applied field), and wavenumber-weighted second derivative of the absorption spectrum for s10:loop:GFP E222Q. By comparing the Stark spectrum and the second derivative, one can conclude that the I state has a smaller Stark tuning rate than the B state (Equations S1 and S2).

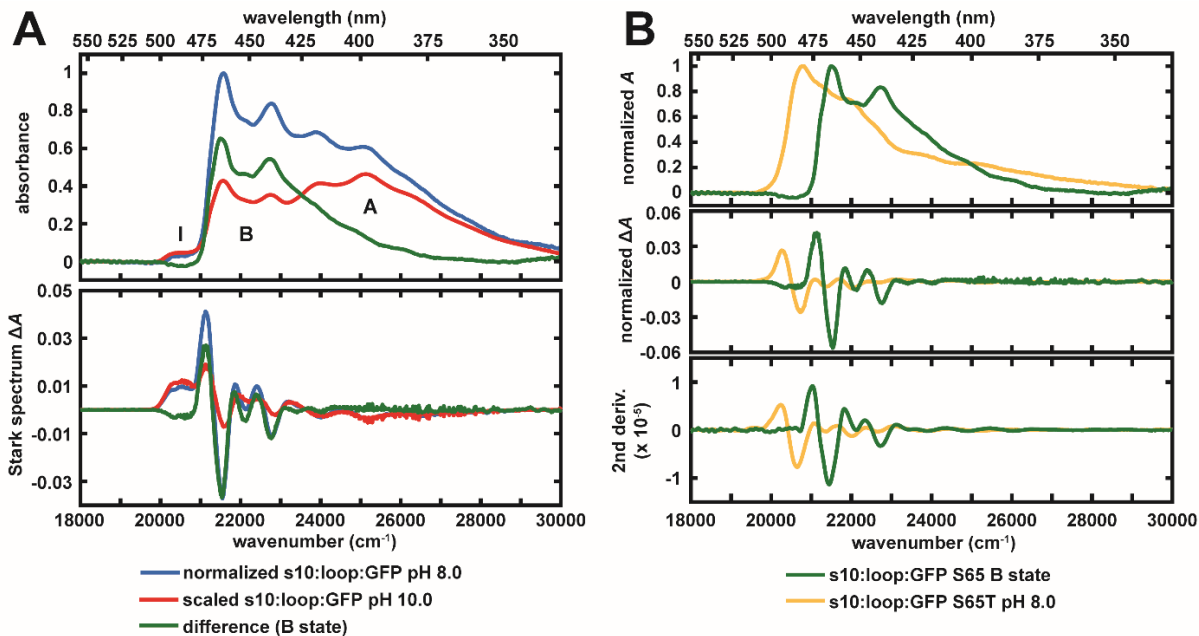
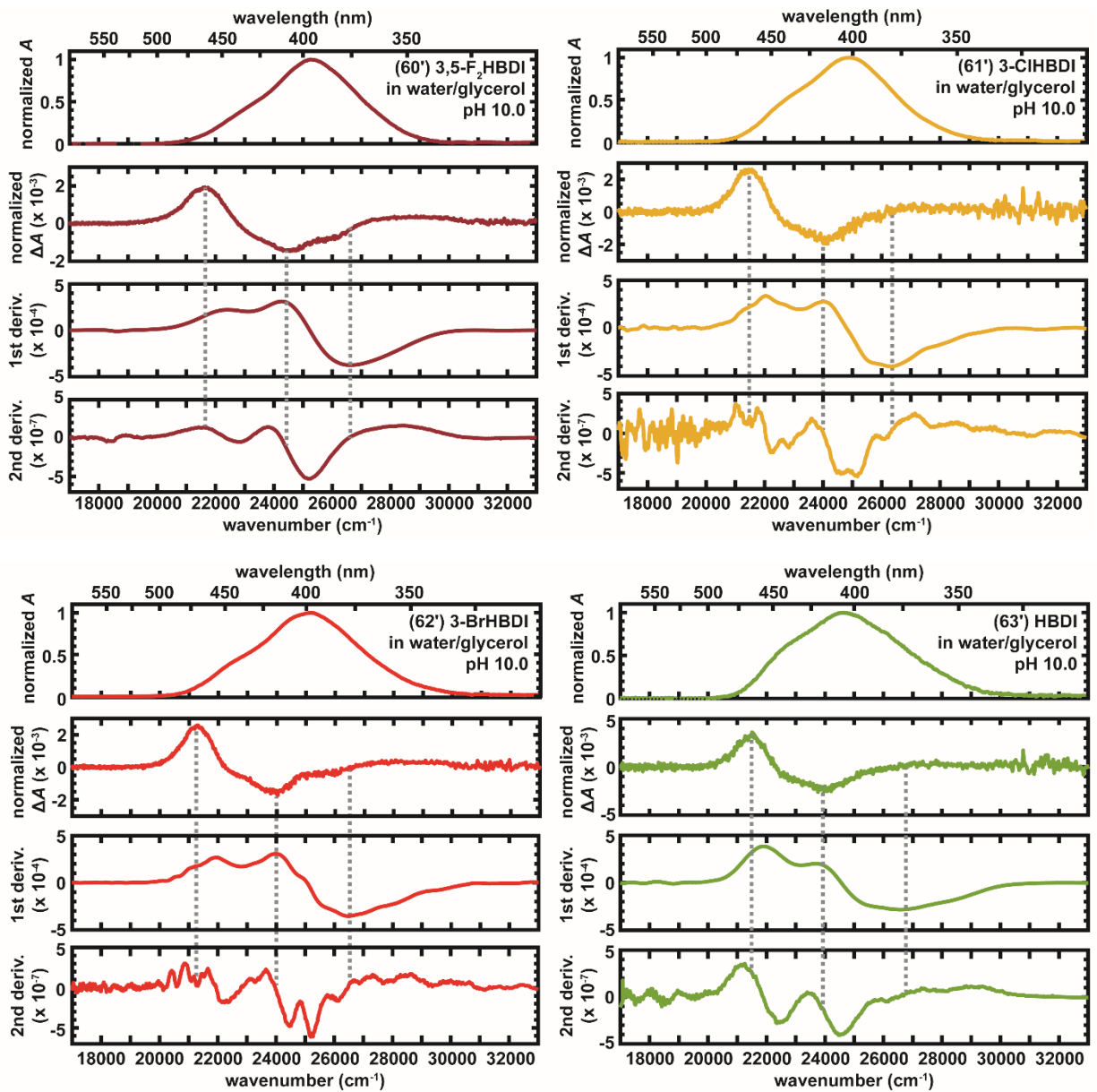


Figure S13. The use of Stark spectroscopy for spectral deconvolution. (A) Comparing the absorption and Stark spectra of s10:loop:GFP at pH 8.0 and 10.0 can effectively isolate the contribution from the B band. (B) Normalized absorption spectra, normalized Stark spectra, and the wavenumber-weighted second derivative of the absorption spectra for the s10:loop:GFP S65 B state obtained from A and s10:loop:GFP S65T at pH 8.0.

GFP Model Chromophores: Violation of Classical Stark Analysis

The strong resemblance between the Stark spectra and the second derivative lineshape of absorption spectra observed for GFP mutants and variants is no longer seen with anionic model chromophores in water/glycerol (Figure S14), especially when comparing the shift in wavenumbers between the most negative features of the Stark spectrum and second derivative. This necessitates the inclusion of other derivative components within the framework of classical Stark analysis while still limiting ourselves to only one set of electro-optic parameters. However, it is not possible to recapitulate these Stark spectra with linear combinations of first- and second-derivative lineshapes (Figure S14), unlike for all GFP mutants and variants (Figure S10). The Stark spectra look as if they can be recapitulated by an extra negative zeroth derivative component. However, the sum-of-derivative analysis fails to varying degrees when simultaneous fits between the absorption and Stark spectra are attempted (one-band fits in Figure S15).

One could resort to two-band fits for analyzing these Stark spectra and find solutions that simultaneously explain both absorption and Stark spectra (two-band fits in Figure S15). This is physically motivated since the band can be strongly inhomogeneously broadened, especially due to the associated large Stark tuning rate. Furthermore, the electro-optic parameters are not necessarily identical across the absorption band [16][17], as has been observed in a hole-burning study of squaraine dyes [41]. However, as described earlier, such an approach allows for too many degenerate solutions. Even if one only picks out solutions with maximal spectral separation, approximately equal population, and minimal features for individual bands, the big differences in Stark tuning rates from the two bands render these solutions suspicious. On the other hand, these fitting difficulties using the classical analysis could suggest the possibility of a non-classical Stark effect such as the intervalence Stark effect of HBDI model chromophores [28][89], arising from the sensitivity in the mixing and populations of the two diabatic states to the fields at the Class II/III borderline. This interpretation is supported by the dominance of the negative zeroth-derivative component in the Stark spectra similarly observed in the Creutz–Taube ion [89] and the fact that the model chromophores' driving forces approach the borderline between the Robin–Day Class II and III (Figures 8 and 9). Deeper analysis requires quantization of the vibrational degrees of freedom [28] in contrast to the current semiclassical treatment (Section S4) and is thus beyond the scope of this study. In contrast, for Class III systems such as all GFP mutants under study, due to the nearly complete electron delocalization from strong electronic coupling, the chromophore can be considered as a single species, and thus its Stark spectra are well-accounted for by the classical sum-of-derivative analysis (Figure S10).



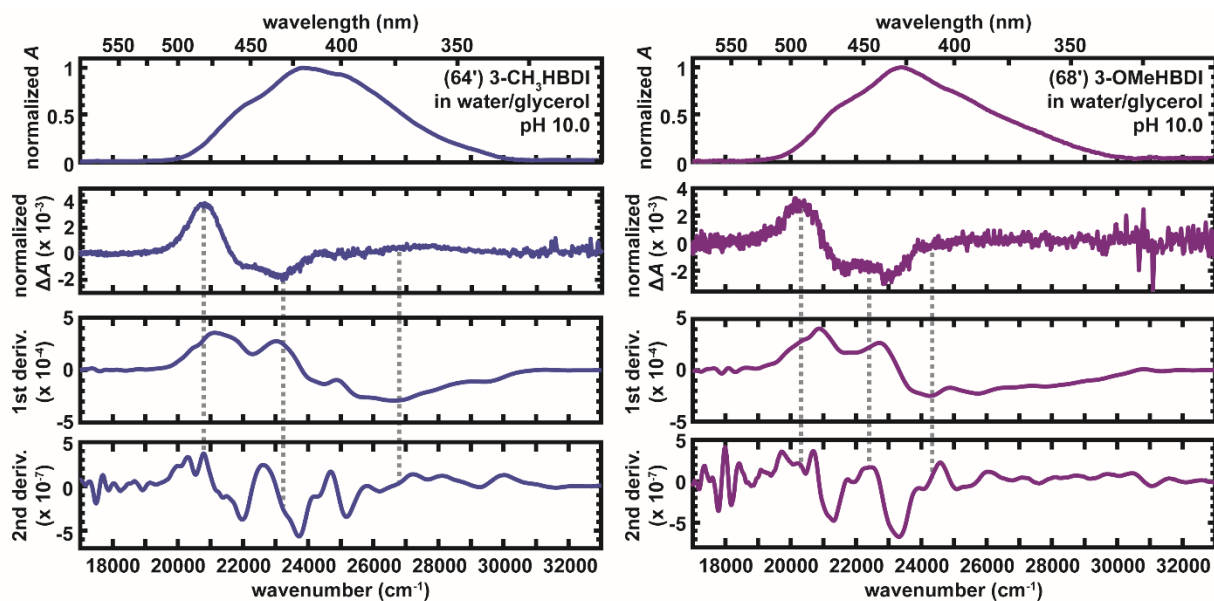
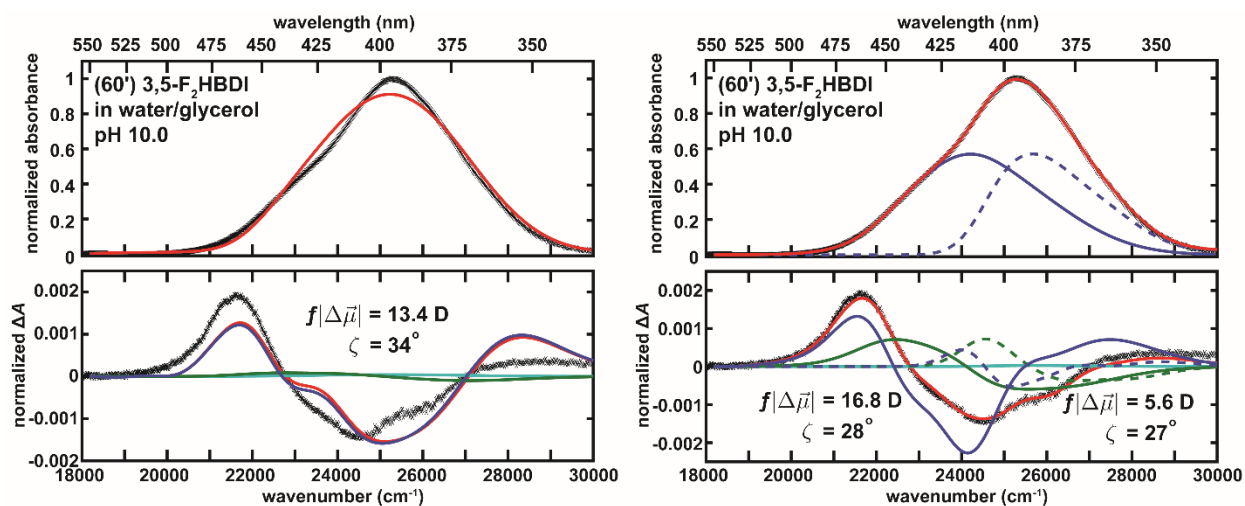
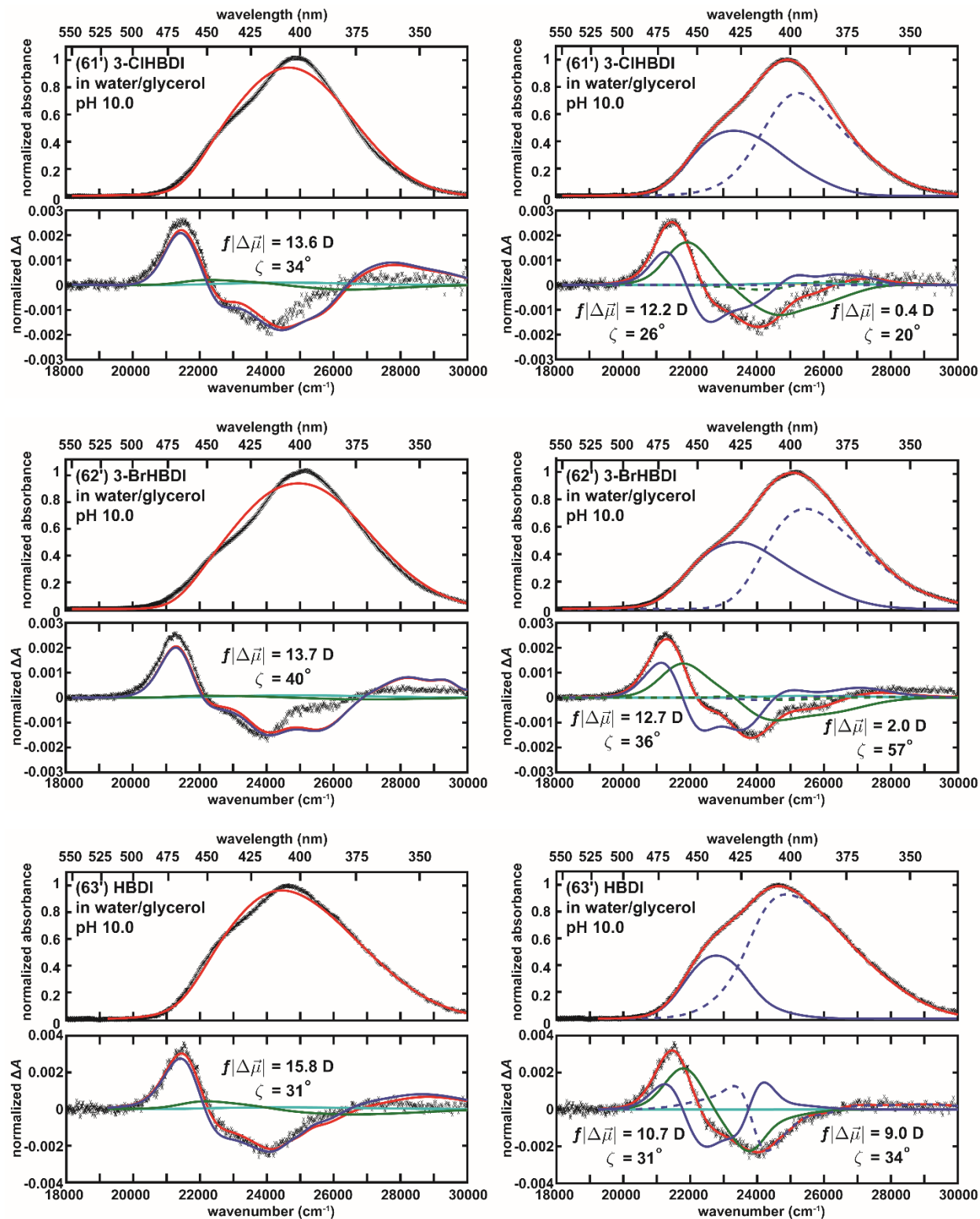


Figure S14. Normalized absorption spectra, Stark spectra, the wavenumber-weighted first derivative, and the second derivative of the absorption spectra for substituted GFP model chromophores in water/glycerol at pH 10.0. Gray dashed lines are shown to illustrate regions that cannot be simultaneously accounted for using classical Stark analysis with one set of electro-optic parameters. The numbering follows Table S14.





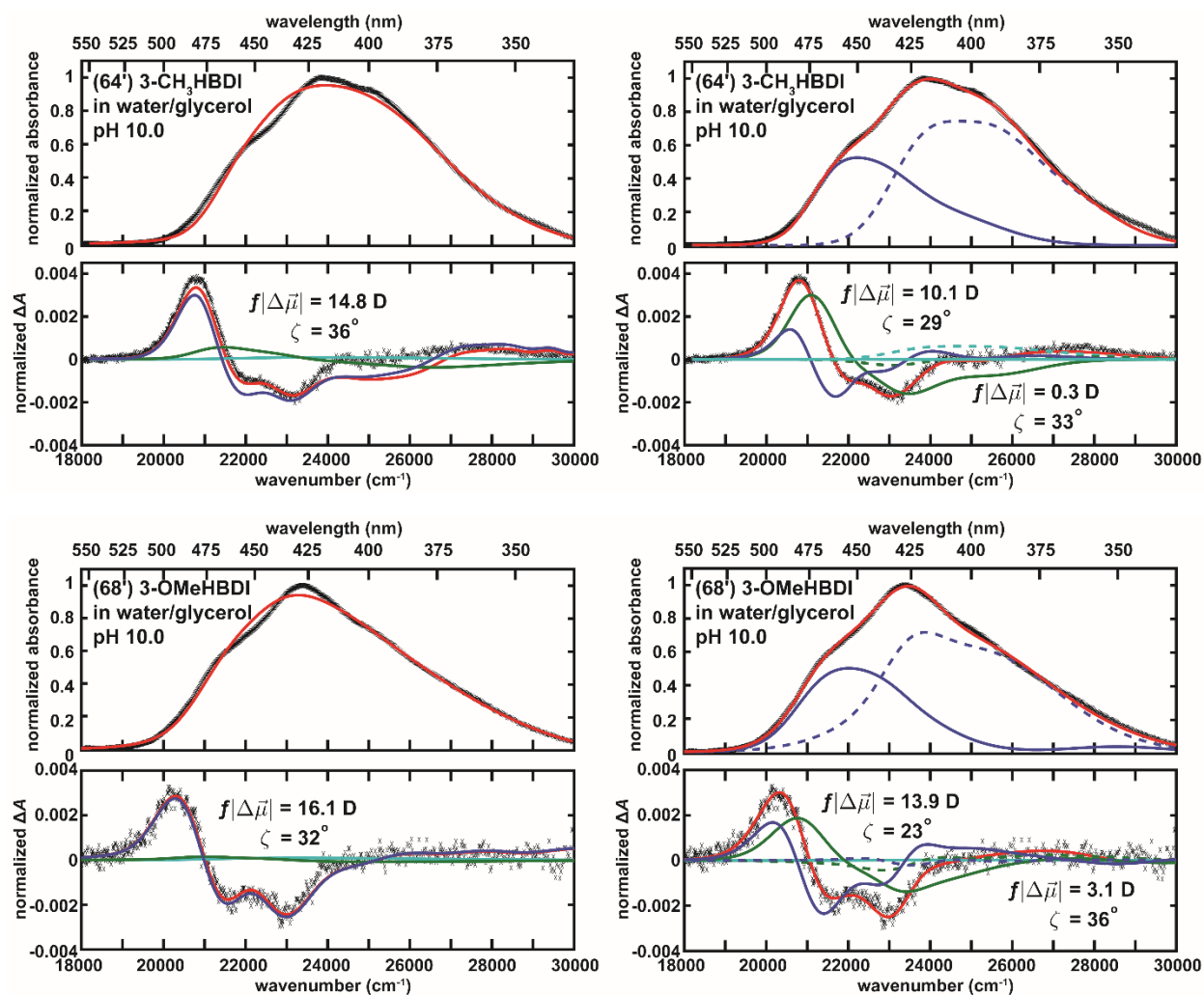


Figure S15. One-band and two-band classical Stark analyses for substituted GFP model chromophores in water/glycerol at pH 10.0. The Stark spectra are measured at $\chi = 90^\circ$ and scaled to 1 MV/cm. The color scheme for fit lines and data points follows the first panel in Figure S10.

S12 Discussion on Protein Structures

In this work, we have extensively utilized circular permutation and split GFP techniques to obtain mutants that are not otherwise accessible, relying on the assumption that these manipulations do not affect the overall structure of GFP and the immediate environment of the chromophore. To justify this assumption, we overlay crystal structures of ih:GFP S65T and s10:∅∅∅:GFP S65, which is a split protein, and compare them to the crystal structure of superfolder GFP (PDB: 2B3P, [102]), which is not circularly permuted (Figure S16). The overall folds for both circular permutants are virtually indistinguishable with that of superfolder GFP (Figure S16A), except for some loop regions. Most notably, the C-terminal end of s10 for s10:∅∅∅:GFP shows a more appreciable deviation from the other two structures, which is not surprising since it is not covalently linked to s11. The largely unresolved artificial loop for linking the C-terminus of s11 and the N-terminus of s1 in circular permutants is also structurally different from the N-terminus and C-terminus of superfolder GFP. Serendipitously, we are almost able to fully identify the whole C-N linker in the structure of ih:GFP S65T T203(3-OMeY), while we are curiously unable to do so in the higher resolution structure of ih:GFP S65T (1.15 Å vs. 1.65 Å) (Figure S17), suggesting that the observed rigidity of this linker might stem from crystal packing rather than an intrinsic property. The immediate environment of the chromophore is also not significantly perturbed upon circular permutation and splitting, especially around the imidazolinone (Figure S16B). E222 in s10:∅∅∅:GFP S65 is slightly displaced compared to the other two structures containing S65T. The mutation of S to T at residue 65 is known to result in rearrangement in the hydrogen-bonding network involving E222 [97]. T203 is the most affected residue and can be displaced up to 1 Å but only leads to subtle absorption maximum shifts of around 3 nm among different circular permutants with the same amino acids in the chromophore's immediate environment (Tables S12 and S14). Combining all the structural and spectroscopic information we know about circular permutants compared to wild-type, we can safely compare the properties of these different constructs and attest our previous claim that circular permutation and splitting minimally affect the chromophore environment [1].

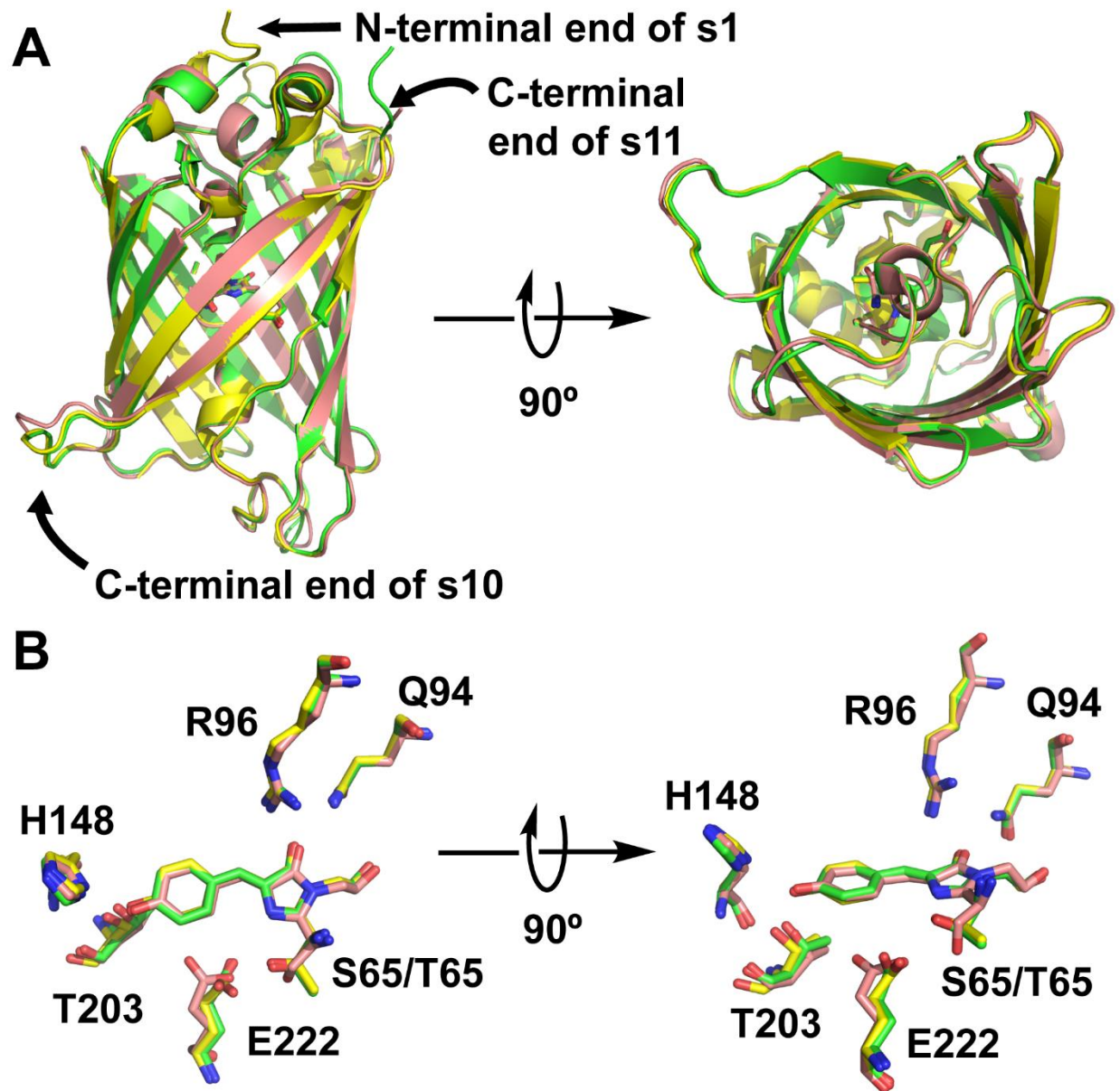


Figure S16. Overlay of circular permutant X-ray structures, including superfolder GFP (PDB: 2B3P, green, [102]), ih:GFP S65T (PDB: 6OFK, yellow, this work), and s10:∅∅∅:GFP S65 (PDB: 6OFO, salmon, this work). (A) The overall folds are virtually identical. (B) The chromophore and its immediate environment exhibit only minor displacements upon circular permutation and splitting.

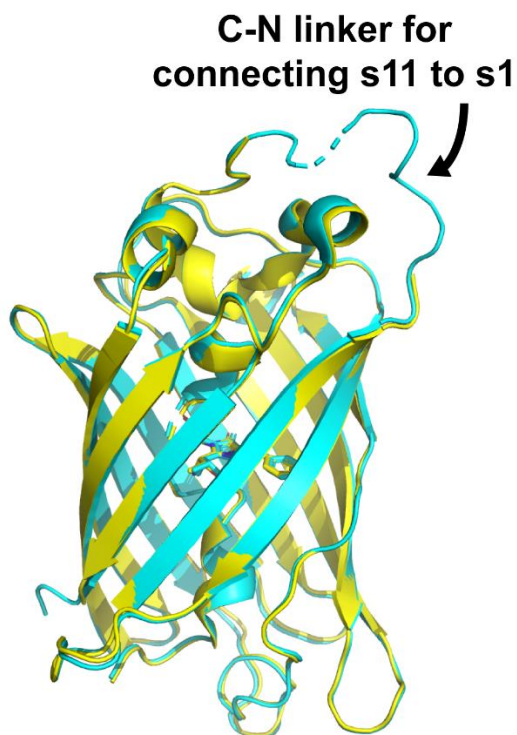


Figure S17. The overlay of ih:GFP S65T T203(3-OMeY) (PDB: 6OFN, cyan, this work) and ih:GFP S65T (PDB: 6OFK, yellow, this work) X-ray structures. The C-N linker is almost fully resolved in the former, while not so in the latter despite a higher resolution (1.15 Å vs. 1.65 Å).

It is also very important to know how introducing substituents to the chromophore perturbs its structure and its local environment, which is the basis of this work and our previous studies [4]. First, introducing a chlorine atom to the chromophore via 3-chlorotyrosine within different protein environments can lead to very different substituent orientations and occupancies (Figure S18). We have shown that the chlorine atom points opposite to the imidazolinone carbonyl when engaging in a short hydrogen bond in ih:GFP S65T Y66(3-ClY) H148D [3], while in Dronpa2 the chlorine lies on the same side as the carbonyl [4]. In ih:GFP S65T Y66(3-ClY), however, the crystal structure shows 80% of the population for the former orientation and the remaining 20% for the latter. There is no sign in any spectroscopy we have conducted, such as Stark spectroscopy, that necessitates the consideration of two distinct populations for this protein, suggesting that the two populations are spectrally indistinguishable (Section S11 and Figure S10, panel

44). It is unclear whether this population ratio is locked during protein folding or if the two orientations are at thermal equilibrium depending on the protein environment and freely exchanging through a ring flip. From an ^{19}F NMR study using a mono-fluorinated chromophore in Dronpa2, we can only set a lower bound of seconds on the population exchange timescale of the chromophore ring flip (an intrinsic limitation of T_1 for ^{19}F [4]), which is not enough to rule out either of the aforementioned scenarios.

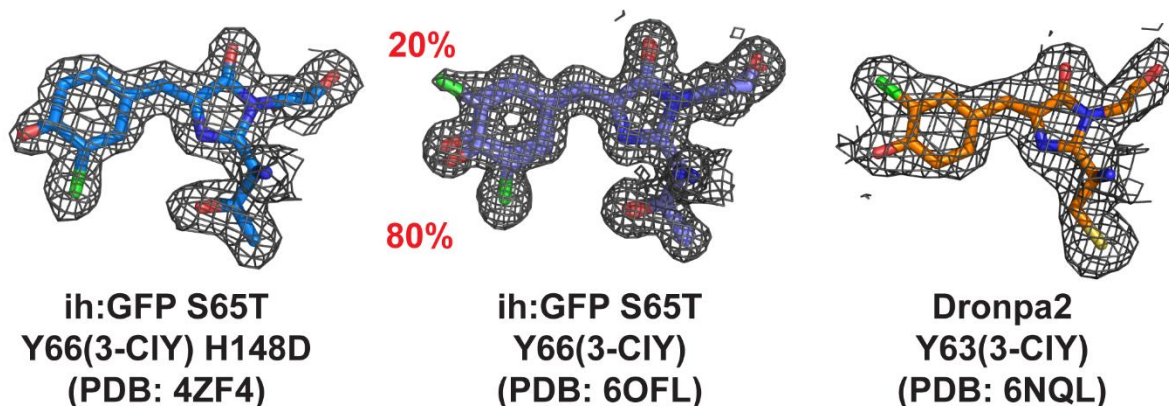


Figure S18. Orientations and populations of the chlorine atoms for mono-chlorinated chromophores in various fluorescent proteins, shown with their corresponding $2mF_o - DF_c$ at 1σ . Note that the variations in mesh densities among different chromophores are due to their different resolutions. Note that various scenarios of chlorine atom orientation have been observed before through 3-ClY incorporated KSI [103][104] and PYP [105], and no obvious environmental factor that leads to these differences can be identified by examining these structures.

In this study, we have also determined high-resolution structures ($< 1.5 \text{ \AA}$) for both chlorinated and methylated chromophores in ih:GFP S65T, allowing us to scrutinize the perturbation on the chromophore's local environment in better detail. In Figure S19, we overlay crystal structures of Y66, Y66(3-ClY), and Y66(3-CH₃Y). We observe little to no residue displacement, the largest of which is T203, amounting to at most 0.3 \AA , which is even smaller than the shift caused by circular permutation (Figure S16). Slight changes in dihedral angles φ and τ along the two bridging bonds (P- and I-bonds, respectively) of the chromophores within various GFP and Dronpa2 mutants/variants can also be distinguished within the resolution of the structures (Table S10). In particular, GFP chromophores tend to be much more planar than Dronpa2 chromophores, and substitutions do not change the dihedral angles by more than 10° , which is insufficient to

significantly affect the electronic coupling V_0 as indicated in the Stokes shift vs. absorption maximum correlation plot (Figure 8). Using the same argument in Section S9, but allowing variations in the dihedral angles, the dihedral angle dependence of V_0 becomes:

$$V_0(\varphi, \tau) = V_0(\varphi = 0, \tau = 0) \cos \varphi \cos \tau \quad (\text{S53})$$

which is also derived in Olsen *et al.* [107]. Given the maximum twisting angles extracted from the crystal structure of Dronpa2 Y63(3-CIY), the penalty factor ($\cos \varphi \cos \tau$) from nonzero dihedral angles is 84%, explaining why values from Dronpa2 variants tend to have a larger spread than the GFP counterparts. Since most GFP and Dronpa2 chromophores possess either lower or similar dihedral angles when compared to those of Dronpa2 Y63(3-CIY), it is reasonable to treat all these chromophores as approximately planar with regards to the electronic distribution.

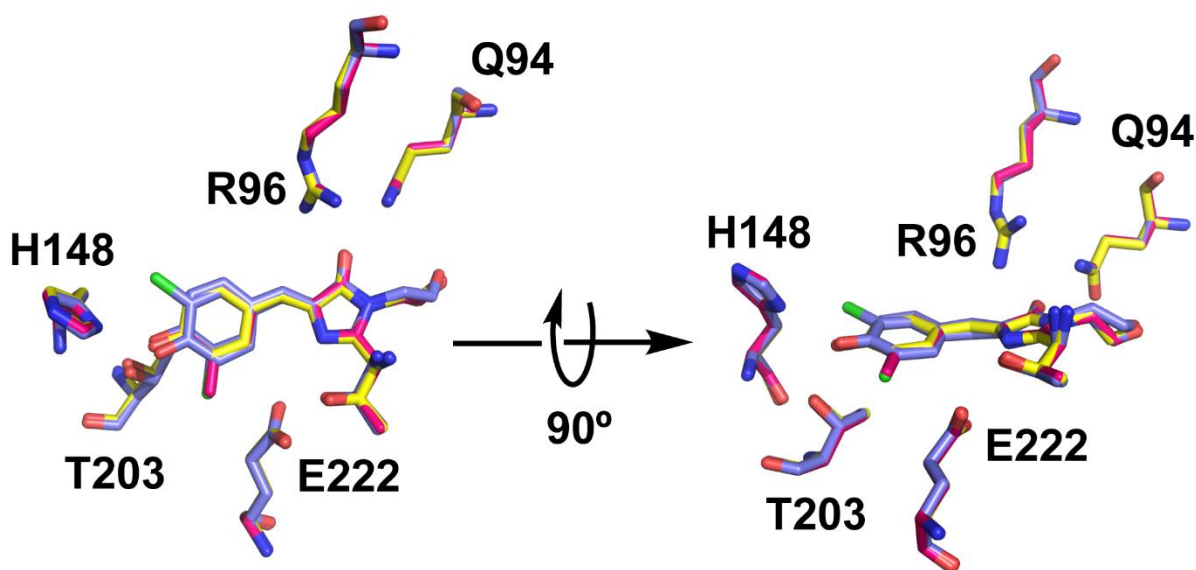


Figure S19. Overlay of GFP variant X-ray structures, including ih:GFP S65T (PDB: 6OFK, yellow, this work), ih:GFP S65T Y66(3-CIY) (PDB: 6OFL, lavender blue, this work), and ih:GFP S65T Y66(3-CH₃Y) (PDB: 6OFM, magenta, this work). Minimal displacements are observed by substituting the chromophore. Note that there are two chlorine orientations in ih:GFP S65T Y66(3-CIY) (Figure S18).

Table S10. Dihedral angles along the chromophore bridging bonds in various fluorescent protein mutants and variants. The dihedral angles are defined in accordance with the convention in Maddalo *et al.* [106].

mutants/variants	dihedral angle φ along the P-bond	dihedral angle τ along the I-bond
superfolder GFP (PDB: 2B3P)	0.4°	1.4°
ih:GFP S65T (PDB: 6OFK)	-8.2°	2.8°
ih:GFP S65T Y66(3-CIY) (PDB: 6OFL)	-10.7° (major orientation) -29.0° (minor orientation) ^a	-0.2° (major orientation) 22.9° (minor orientation) ^a
ih:GFP S65T Y66(3-CH ₃ Y) (PDB: 6OFM)	-13.4°	3.5°
ih:GFP S65T T203(3-OMeY) (PDB: 6OFN)	0.9°	-0.3°
Dronpa2 (PDB: 6NQJ)	-23.7°	18.3°
Dronpa2 Y63(3-CIY) (PDB: 6NQL)	-29.4°	16.1°

^a Since the population for the minor orientation is only 20%, which means little weight has been given to this population during electron density fitting, the corresponding geometries are not as reliable as those determined from the major orientation.

One last interesting feature we learn from protein structures determined in this work is the π - π stacking behavior between T203(3-OMeY) and the GFP chromophore. Replacing T203 with an aromatic residue is a well-known strategy for red shifting the absorption and emission bands of GFP given its ability to π - π stack with the phenolate moiety of the GFP chromophore (Figure 2). In YFP, which contains the mutation T203Y,

such π - π stacking shows an off-centered parallel configuration that is commonly referred to as “parallel-displaced” (Figure S20) [110]. Its preference over the perfectly sandwiched and eclipsed configuration can be elegantly explained by electrostatics through the Hunter–Sander rules for two electron-rich rings [111]. However, when we attach an additional methoxy group to the tyrosine ring, the OMeY ring is no longer parallel to and less displaced from the chromophore’s phenolate moiety such that a nonzero angle offset is evident between either of the C-O bonds of OMeY and the chromophore’s P-bond (Figure S20). Interestingly, the two chains within the asymmetric unit show different configurations of π - π stacking. It has been shown through computation on model systems that the identity of the substituents on both rings can shift the most stable angle offset between the rings and determine whether the configuration is eclipsed or staggered [112], but this study was performed with the sandwich configuration and 1,3,5-substituted rings, which are not directly comparable with our results [113]. Based on these crystal structures, we argue that the use of T203 mutants as model systems for exploring and elucidating π - π stacking is largely overlooked and can be complementary to examining synthetic organic molecules [113].

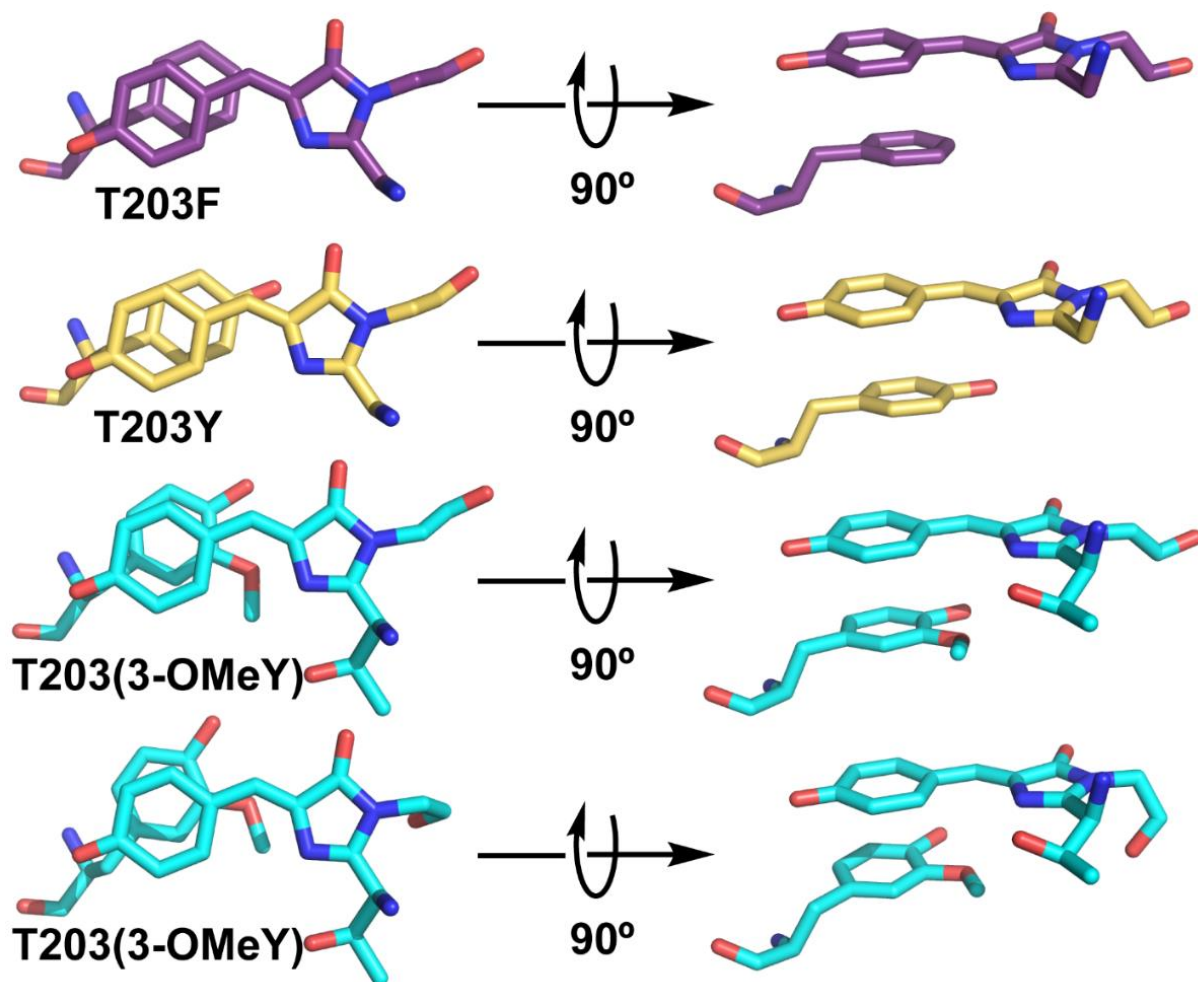


Figure S20. The π - π stacking between the aromatic side chain at residue 203 and the GFP chromophore. The structures from top to bottom are extracted from EYFP S65G T203F (PDB: 3V3D, purple, [108]), avGFP T203Y (PDB: 1YFP, gold, [109]), ih:GFP S65T T203(3-OMeY) (PDB: 6OFN, cyan, this work) chain A and chain B.

S13 Supplementary Figures

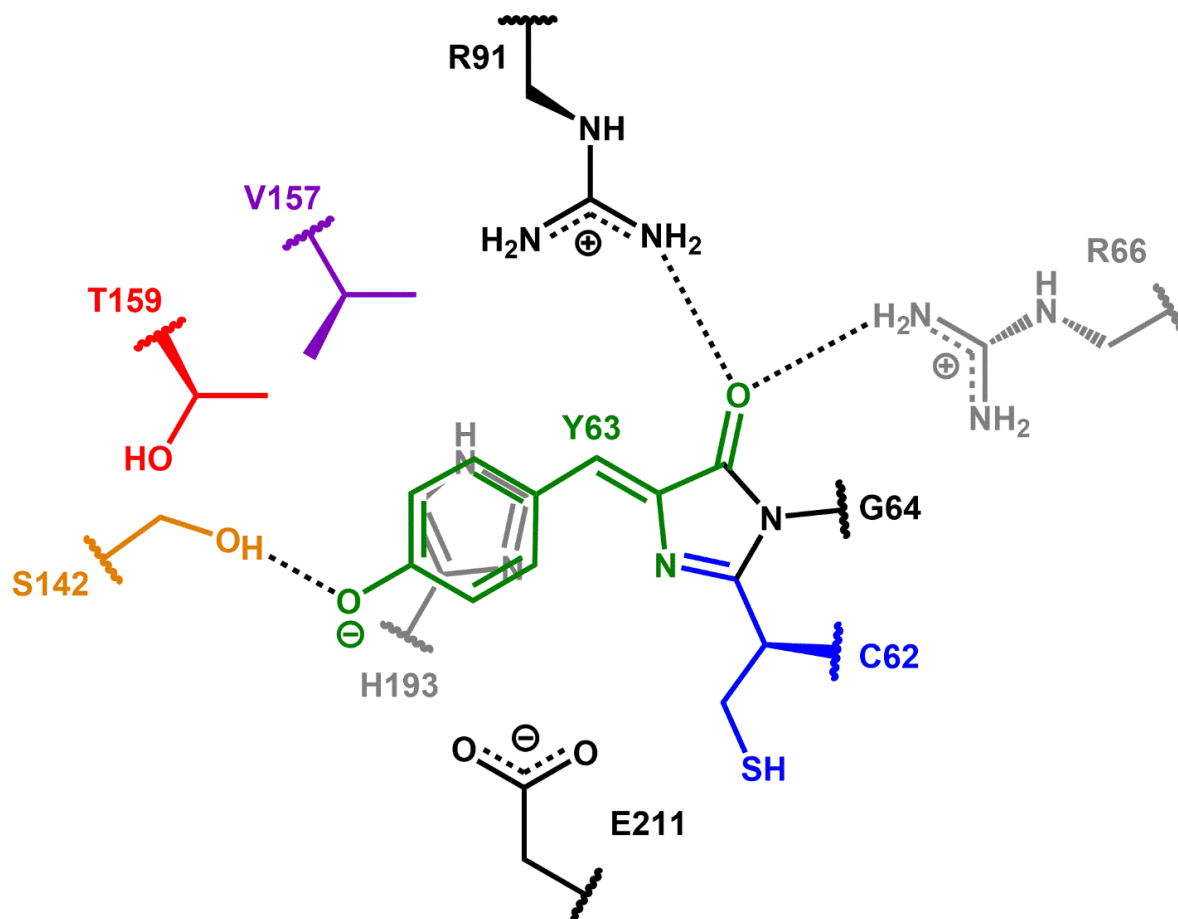


Figure S21. The chromophore binding pocket of Dronpa2 based on the crystal structure determined by Kaucikas *et al.* (PDB: 4UTS, [114]). The colored residues represent those mutated in this work (Figure 8 and Table S12). Gray residues are those behind the chromophore plane.

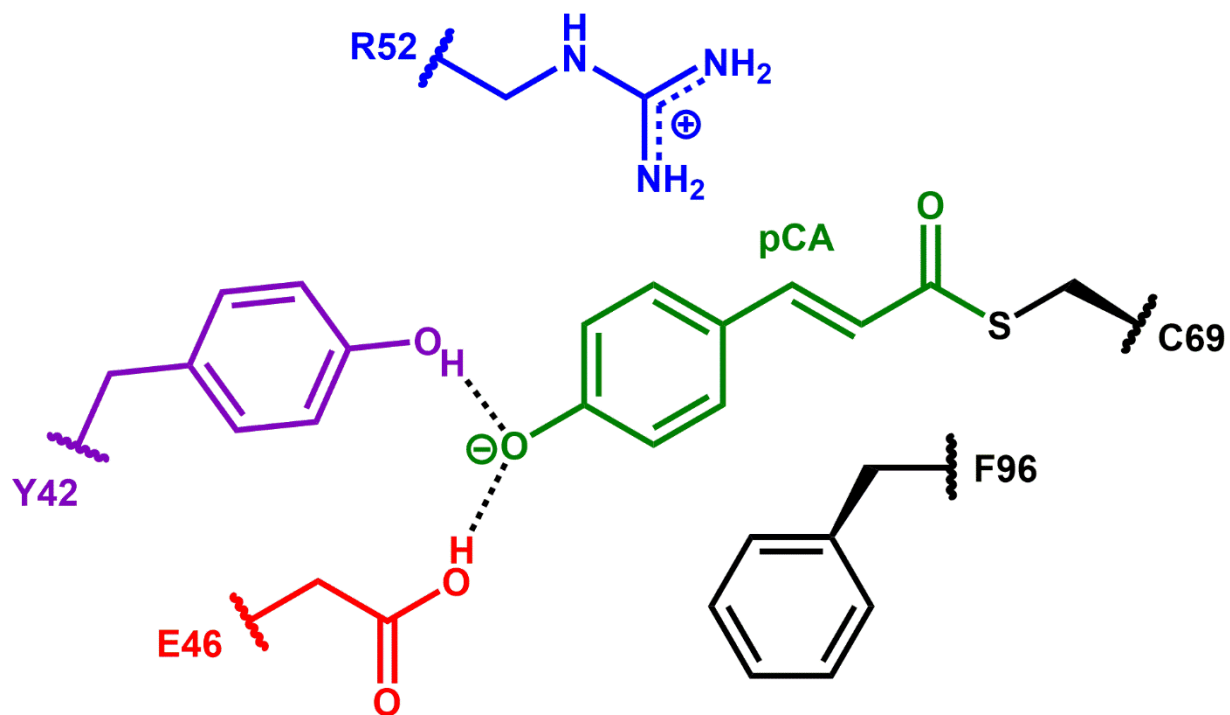


Figure S22. The chromophore binding pocket of *Halorhodospira halophila* PYP based on the high-resolution crystal structure determined by Getzoff *et al.* (PDB: 1NWZ, [115]). The colored residues represent those mutated in the literature cited in this work (Figure 8 and Table S12).

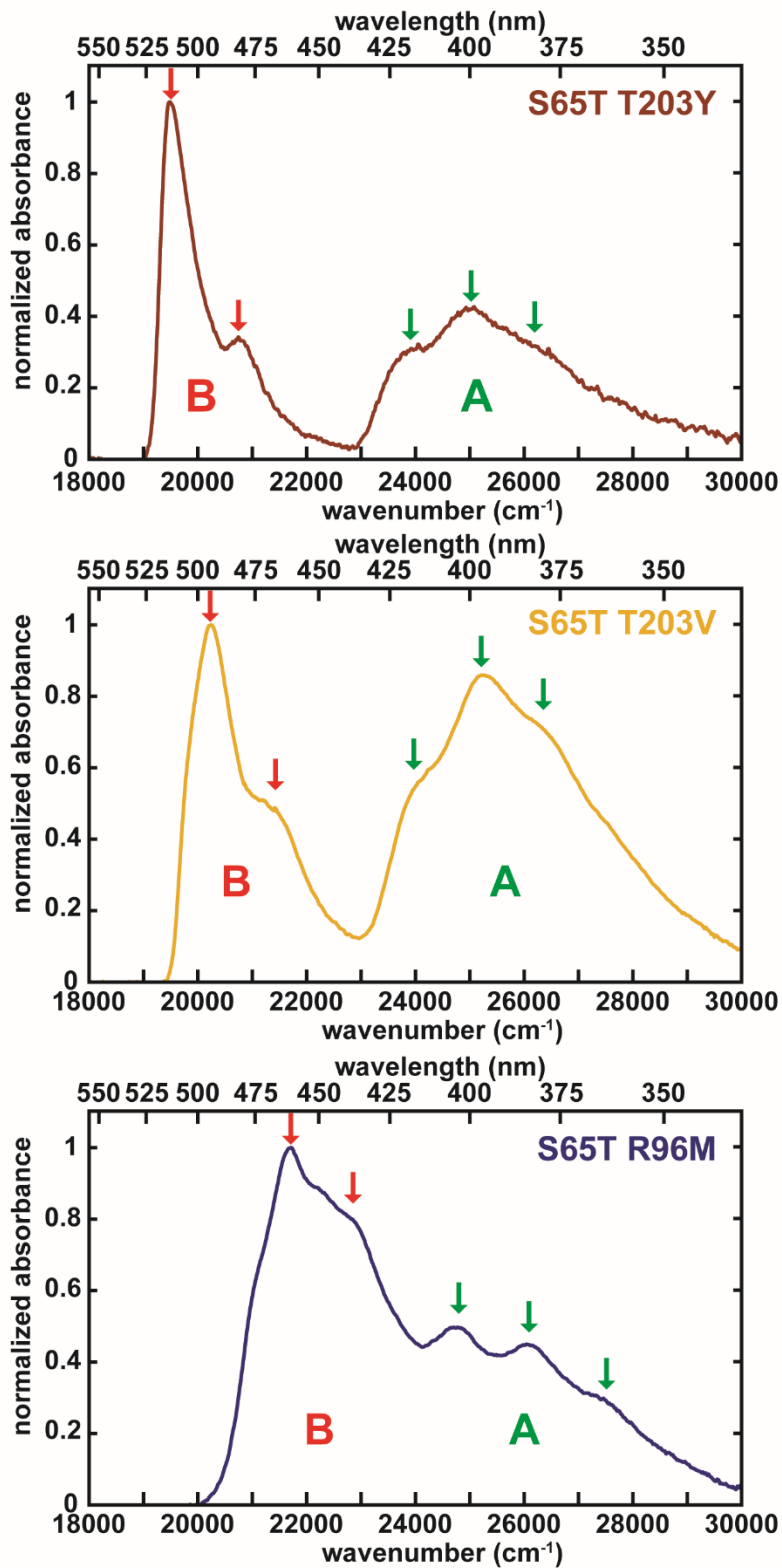


Figure S23. Vibronic structures of A and B bands for representative GFP mutants at 77 K. Three prominent bluer peaks (green arrows) and two redder peaks (red arrows) can be assigned to the A and B bands, respectively.

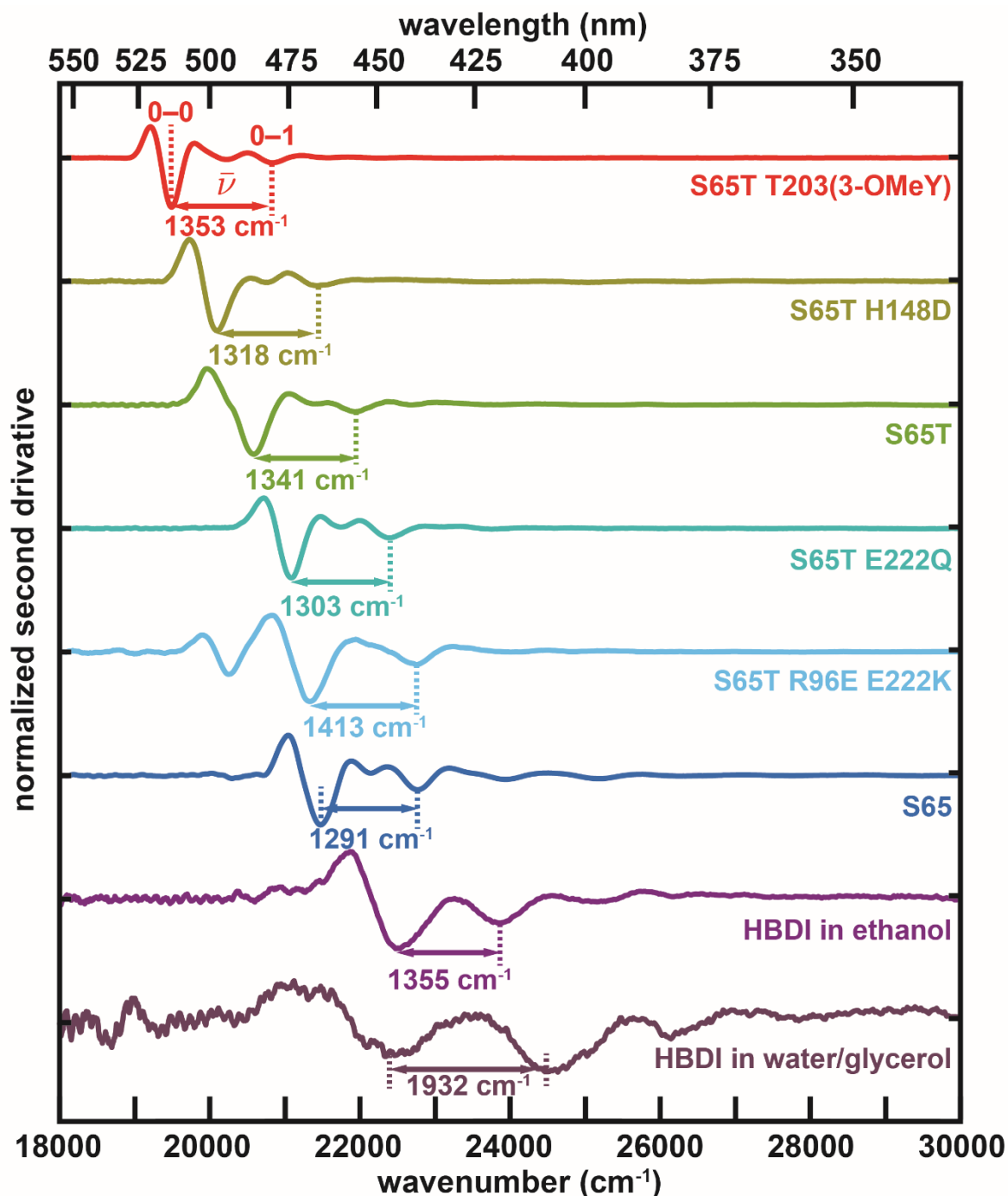


Figure S24. Method for estimating the difference between 0–0 and 0–1 transitions using the second derivatives of 77 K absorption spectra (Figure 5A) of GFP mutants and HBDI. 0–0 and 0–1 energies are assigned from the negative peak positions; this difference corresponds to the dominant BLA vibrational frequency. The color coding follows that of Figure 5A. The reddest negative peak of GFP S65T R96E E222K corresponds to the I state, which is different from the B state by a rotation of T203 (Section S11). The presence of this I band likely results from the denaturation and renaturation of the protein during preparation (Section S1).

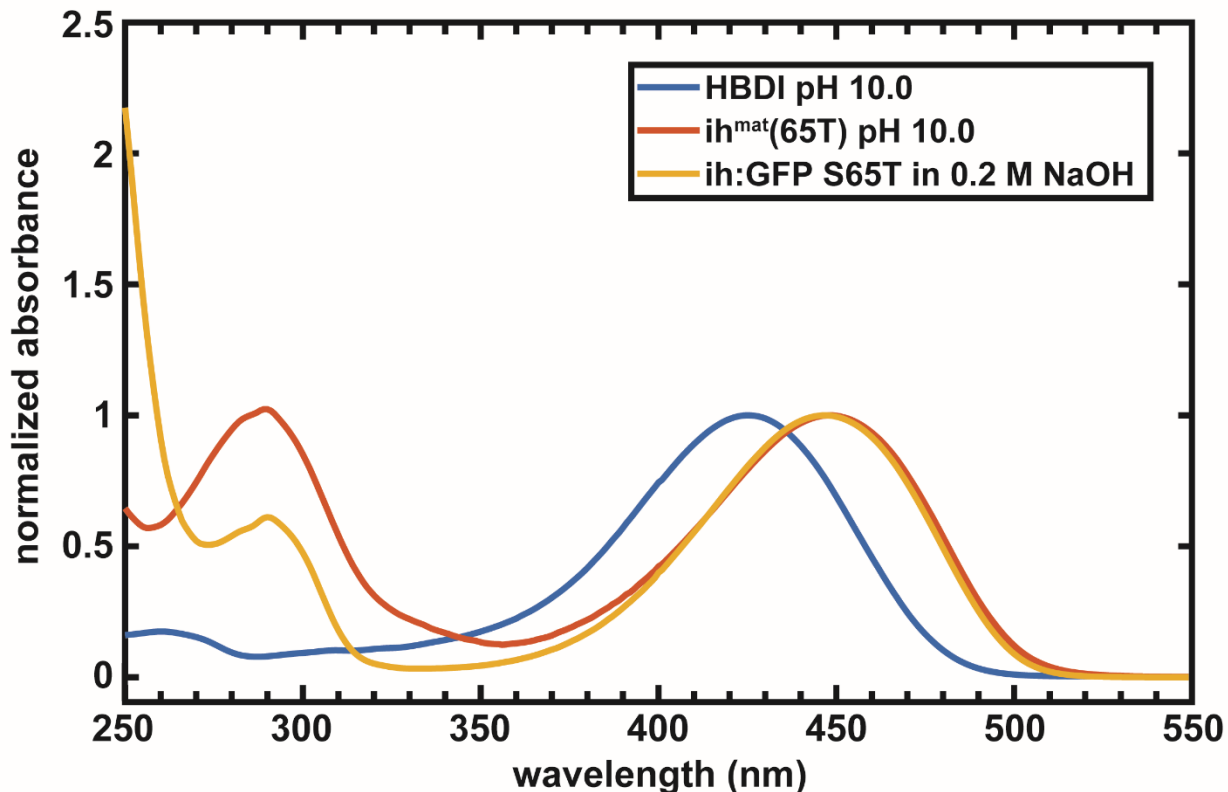


Figure S25. Room-temperature absorption spectra of anionic HBDI, the internal helix containing the mature chromophore (33 residues long if the chromophore counts as one), and denatured ih:GFP S65T in 0.2 M NaOH. The absorption peak of the denatured ih:GFP S65T at 447 nm is usually used as the standard for GFP extinction coefficient determination. Interestingly, the anionic model chromophore HBDI absorbs at 425 nm, which does not match the absorption maximum of the denatured GFP or the isolated internal helix (447 nm). This observation suggests that the immediate environment of the chromophore is not the same when free in water and located in the internal helix, presumably due to the perturbation of the water solvation structure around the chromophore from neighboring residues. A difference in absorption spectra between another model chromophore in water and enzymatically digested GFP has also been noted previously [116]. Similarly, the vibrational frequency of *p*-acetyl-L-phenylalanine's carbonyl when incorporated into a supposedly unstructured 15-residue long polypeptide was measured to be somewhat shifted from that when free, likely reflecting some residual local structure [117].

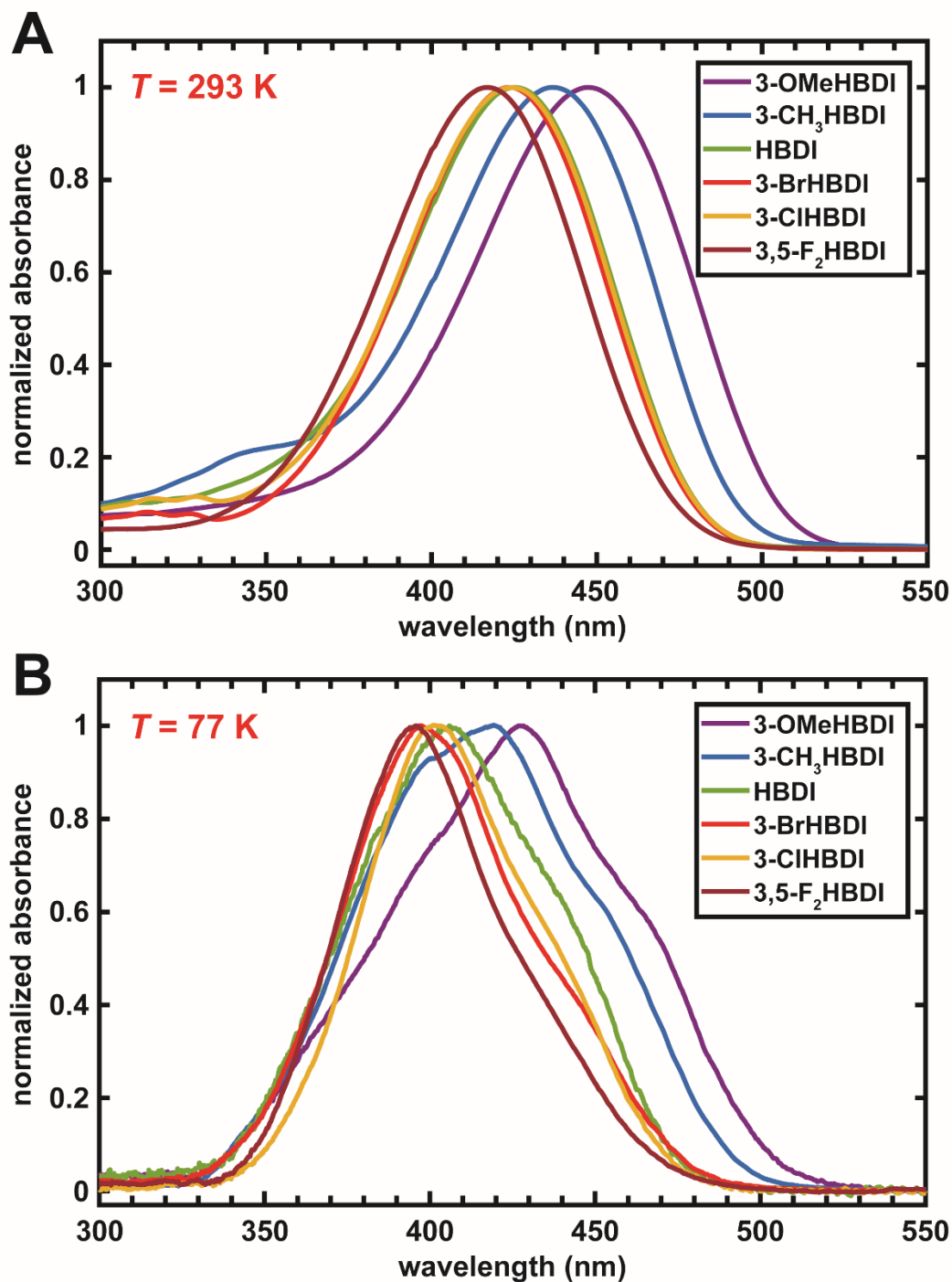


Figure S26. (A) Room-temperature and (B) 77 K absorption spectra of various GFP model chromophore analogues (Figure 4) in carbonate buffer and carbonate buffer/glycerol at pH 10.0, respectively. The absorption peak maximum is red-shifted or blue-shifted when the chromophore is modified by an electron donating or withdrawing group, respectively. The absorption peak maximum is also blue-shifted upon freezing (Figure S28). The color coding follows that of Figure S14.

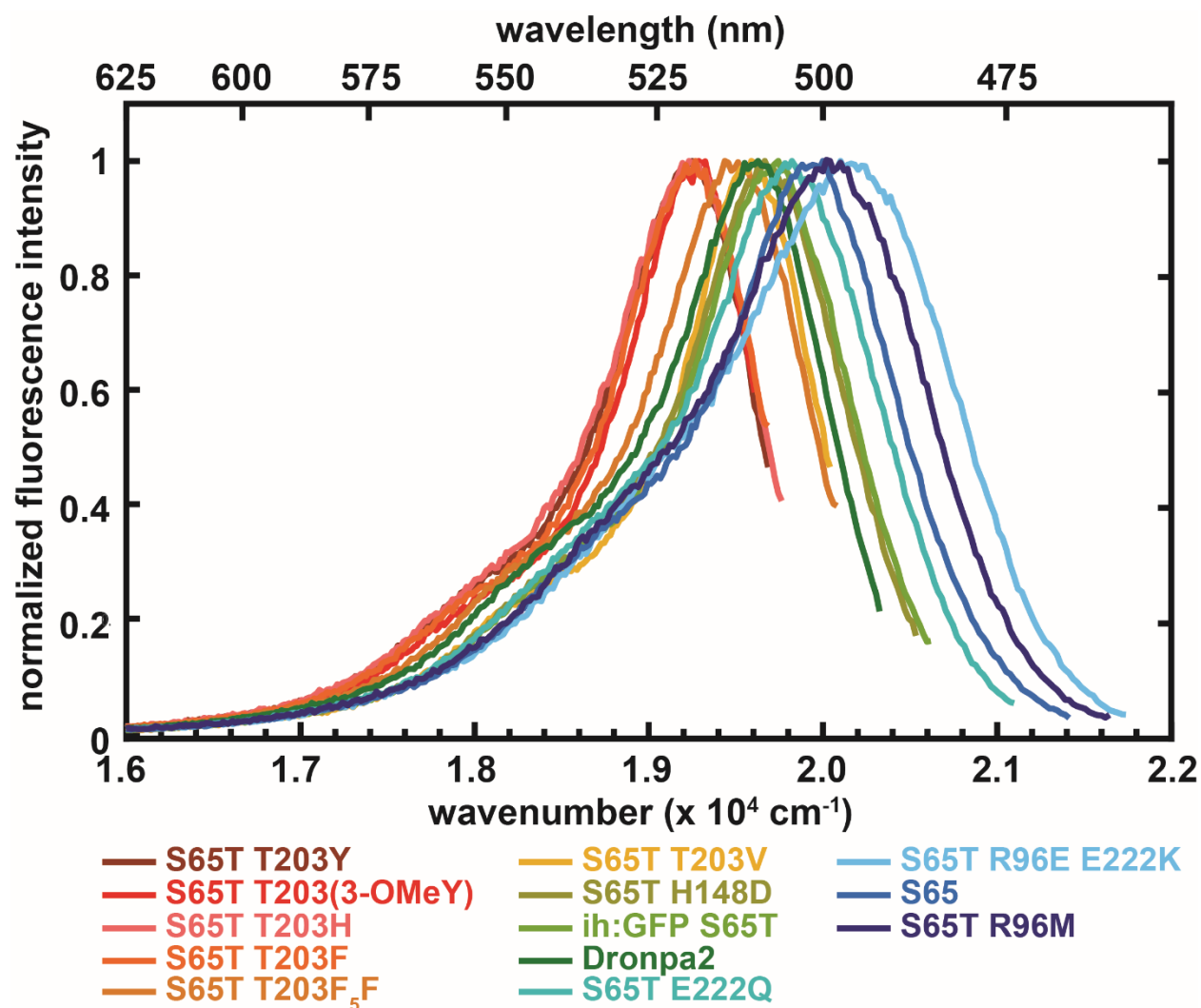


Figure S27. Representative emission spectra at room temperature for GFP mutants listed in Figure 5A with the same color coding. These emission profiles are acquired by excitation at the corresponding absorption maxima, and the emission maxima in Table S12 are subsequently determined from the peak wavelengths. Note that especially for redder mutants, a vibronic sideband (0-1 transition) can be observed due to the narrow peak.

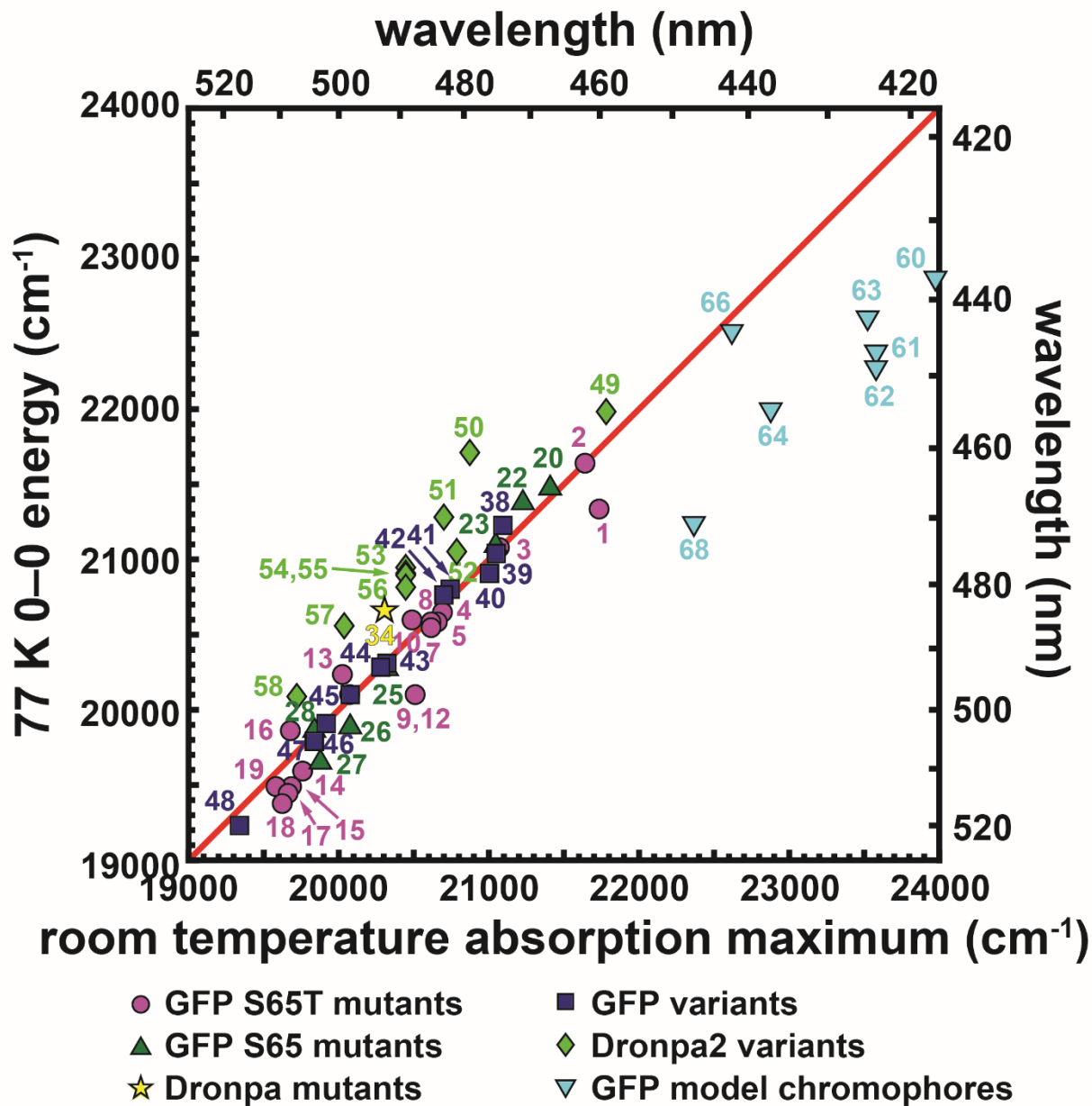


Figure S28. Correlation plot between 77 K 0–0 energy (Table S14) and room-temperature absorption maximum (Table S12) for mutants, variants, and model chromophores of GFP and Dronpa, with the diagonal line (red) as a visual aid. The numerical labels are defined in Table S12. The two quantities are more or less identical for GFP mutants and variants. Dronpa mutants and variants are blue-shifted upon freezing, as noted in our previous publication [4]. For model chromophores in water, because their Huang–Rhys factors are larger than unity (Subsection 2.6), their room temperature absorption maxima correspond to 0–1 transitions.

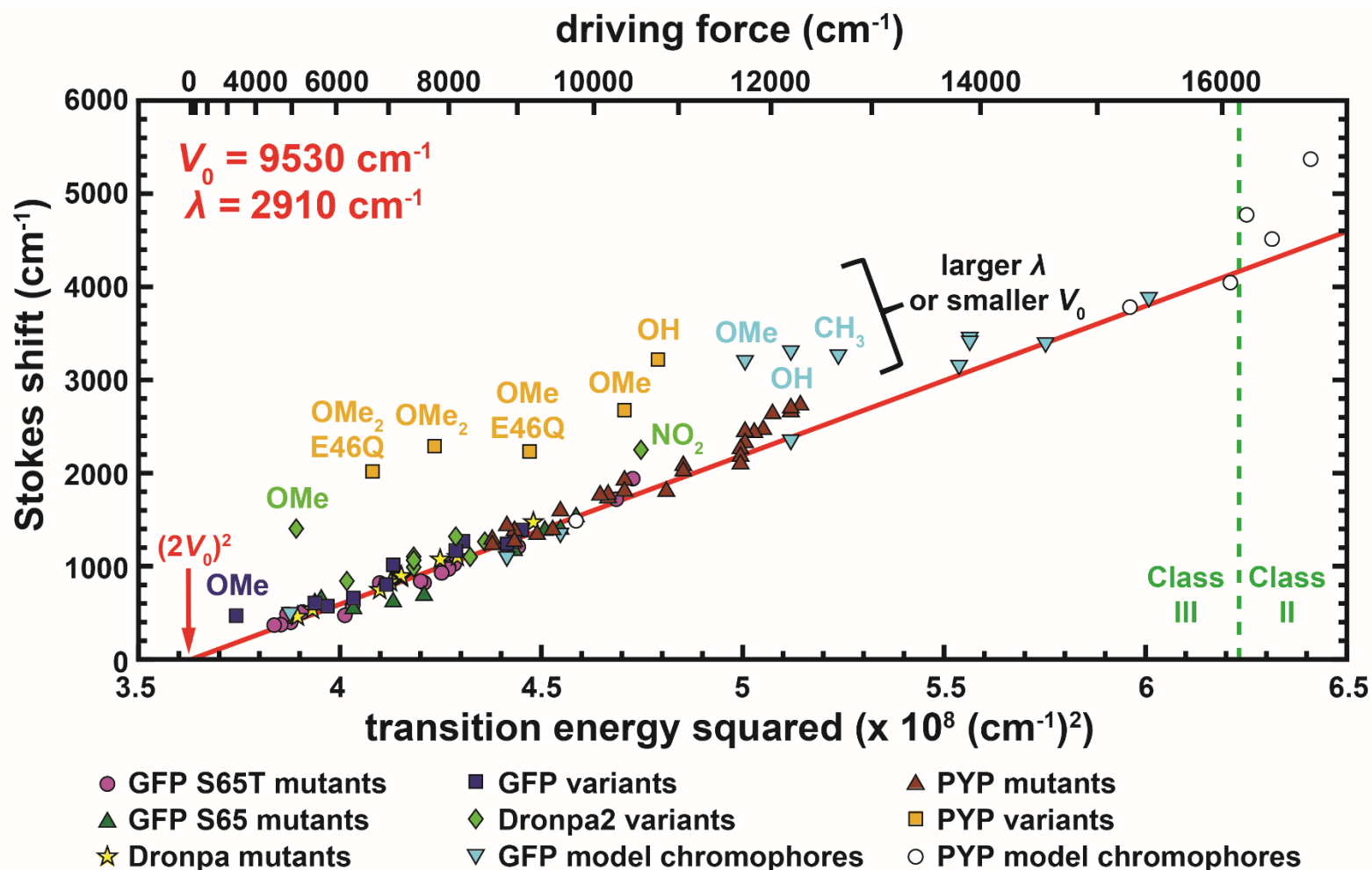


Figure S29. Correlation between Stokes shift and transition energy squared from absorption maxima for all mutants, variants, and model chromophores of GFP, Dronpa, and PYP (Table S12). References of the data not measured in this study are also listed in Table S12. The red line comes from fitting the data for the GFP S65T mutants to Equation S17. The outliers, corresponding to chromophores with mostly strong electron donating or withdrawing groups, are labeled. The distinction between Class II and Class III systems is shown with a green dashed line.

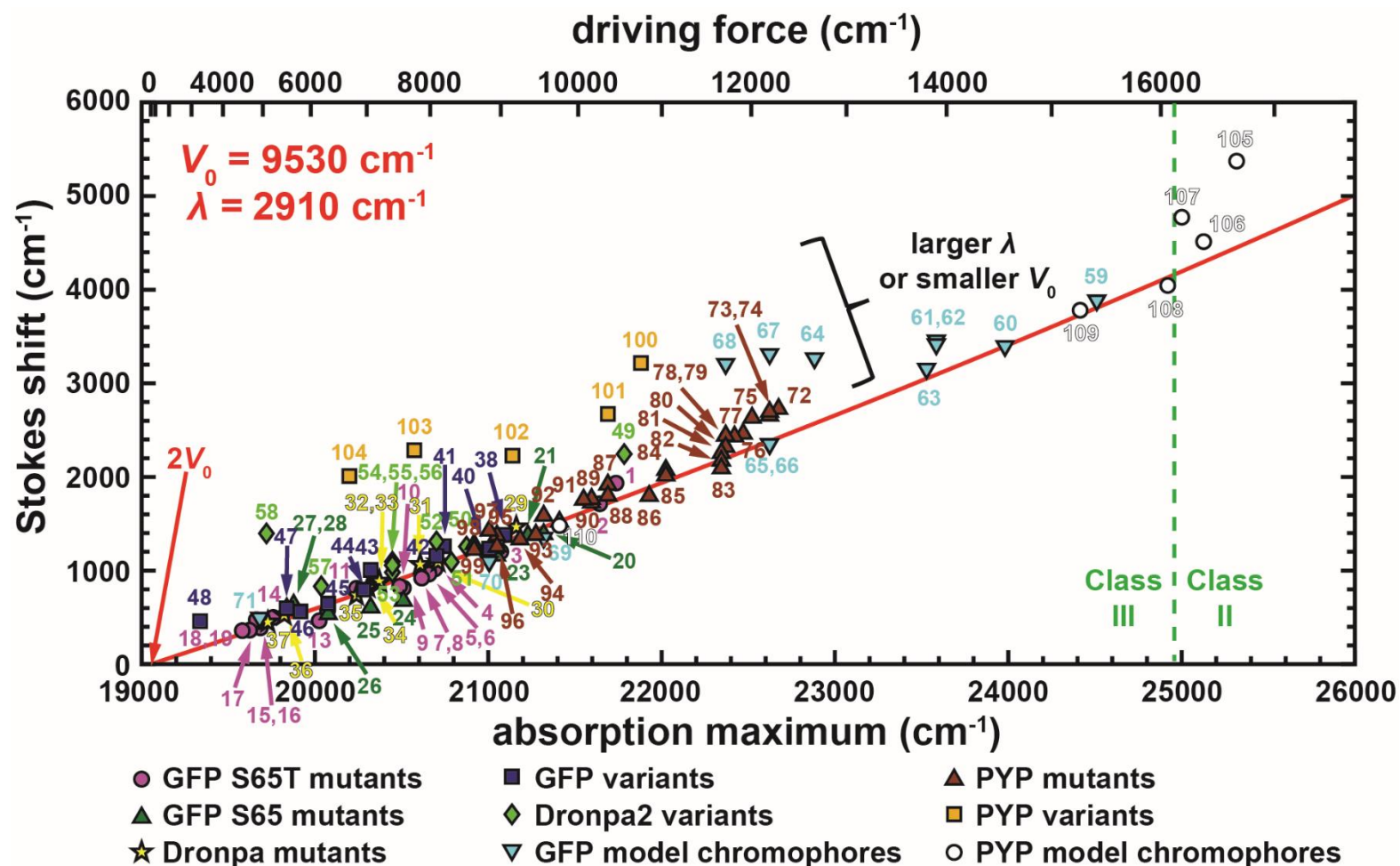


Figure S30. Correlation between Stokes shift and absorption maximum for all mutants, variants, and model chromophores of GFP, Dronpa, and PYP (Table S12). This figure is reproduced from Figure 8 to include numerical labels defined in Table S12. References for the data not measured in this study are also listed in Table S12. The red curve comes from fitting the data for the GFP S65T mutants to Equation S17. The outliers, corresponding to chromophores with mostly strong electron donating or withdrawing groups, are labeled with their corresponding substituents. The distinction between Class II and Class III systems is shown with a green dashed line.

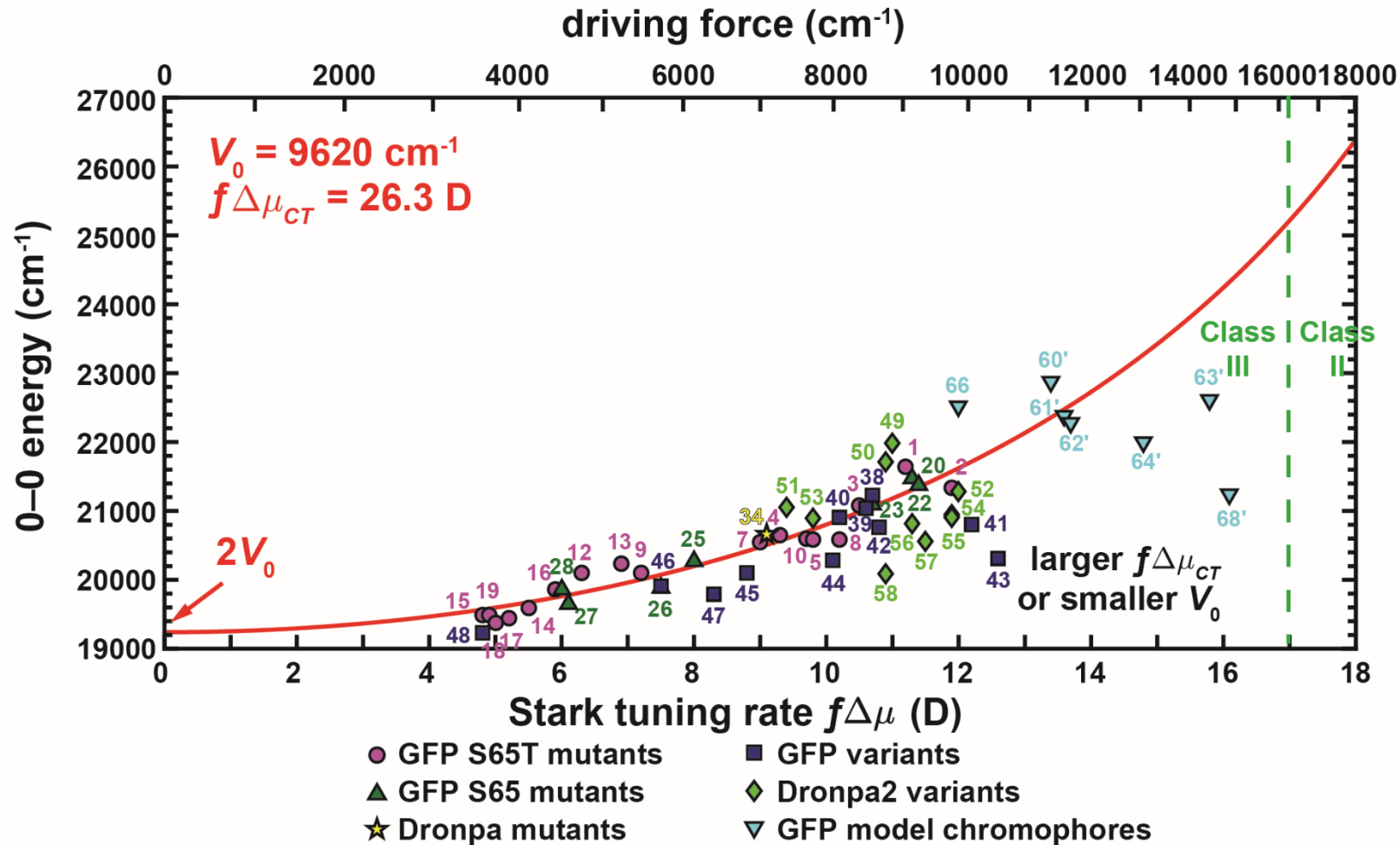


Figure S31. Correlation between Stark tuning rate and 0–0 transition energy for mutants, variants, and model chromophores of GFP and Dronpa (Table S14). This figure is reproduced from Figure 9 to include numerical labels defined in Table S14. The red curve comes from fitting the data for the S65T mutants to Equation S23. The distinction between Class II and Class III systems is shown with a green dashed line. The local field factor f is required as a conversion factor between the observed and the true Stark tuning rates (Section S6).

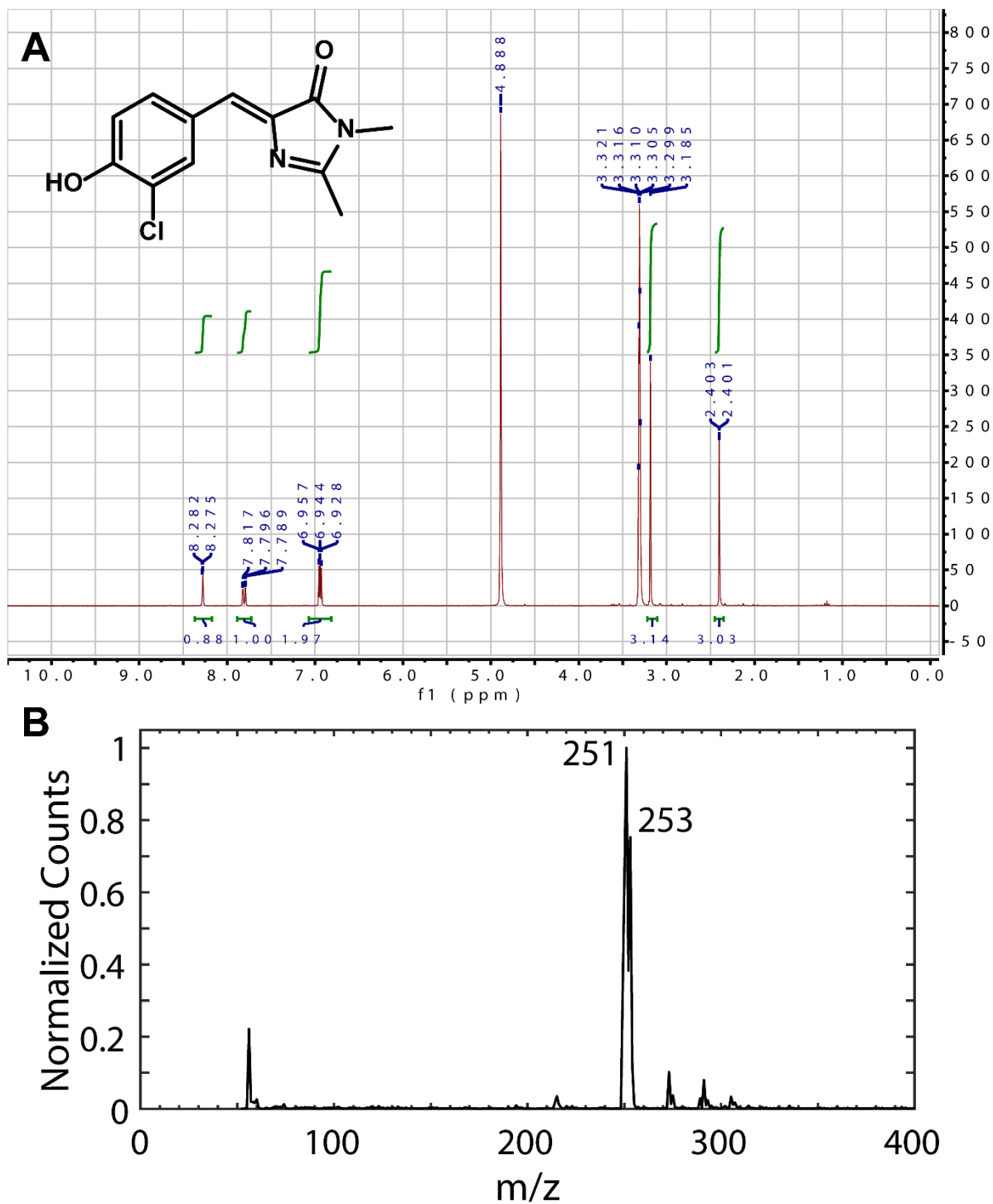


Figure S32. (A) ^1H NMR spectrum (300 MHz, in CD_3OD) and (B) ESI+ mass spectrum of 3-ClHBDI.

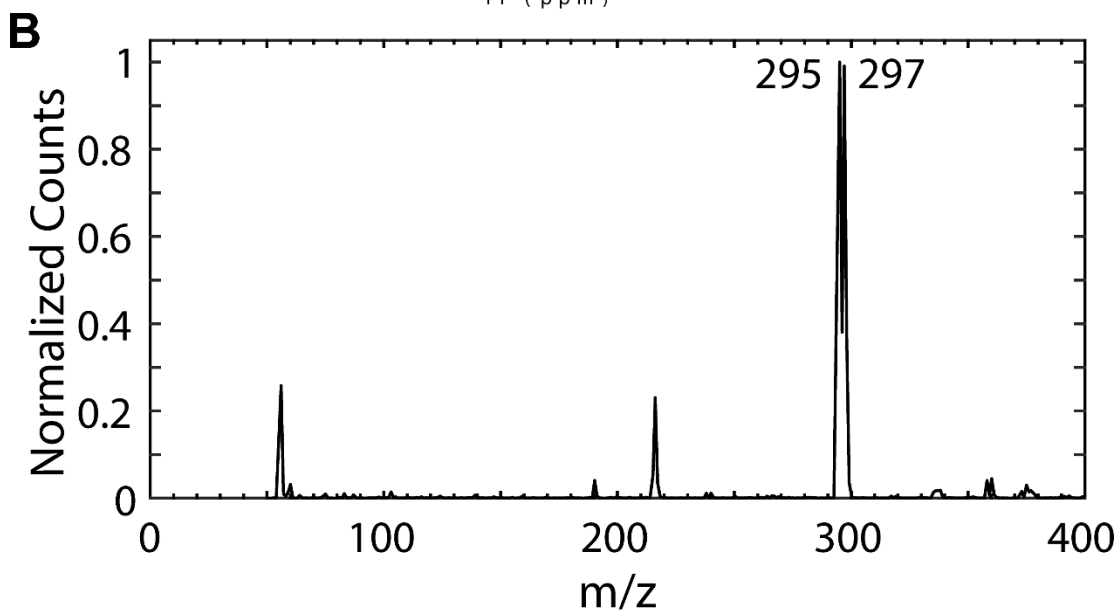
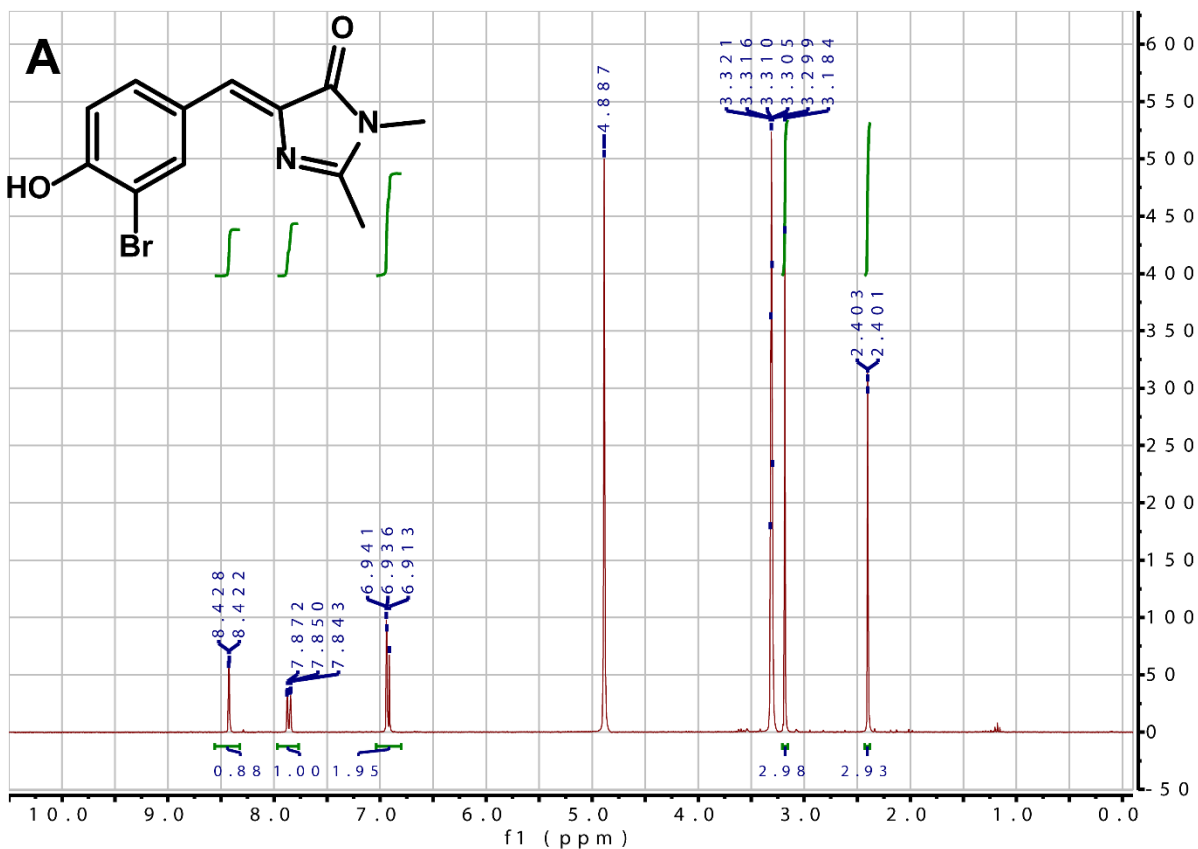


Figure S33. (A) ¹H NMR spectrum (300 MHz, in CD₃OD) and (B) ESI+ mass spectrum of 3-BrHBDI.

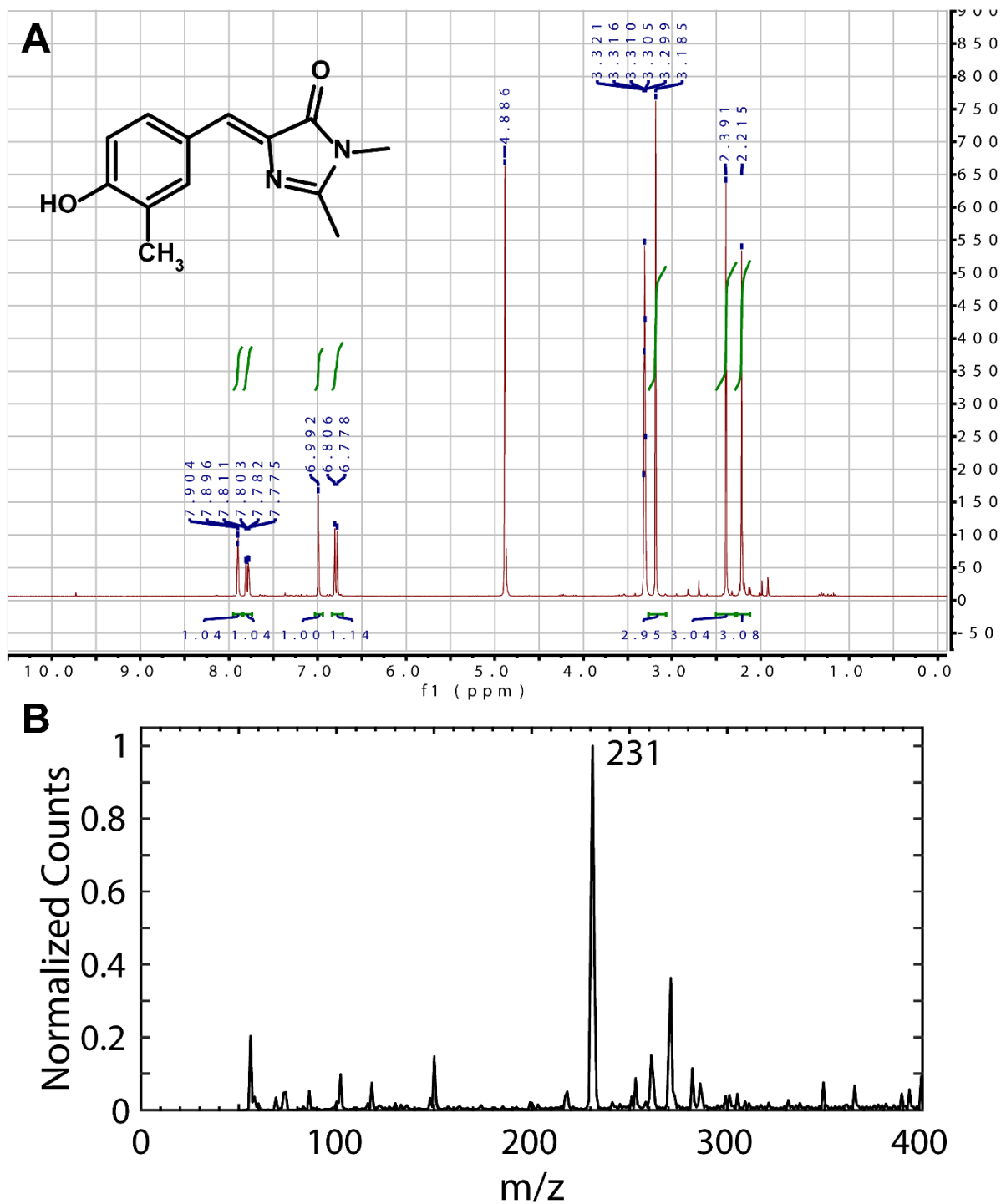


Figure S34. (A) ¹H NMR spectrum (300 MHz, in CD₃OD) and (B) ESI+ mass spectrum of 3-CH₃HBDI.

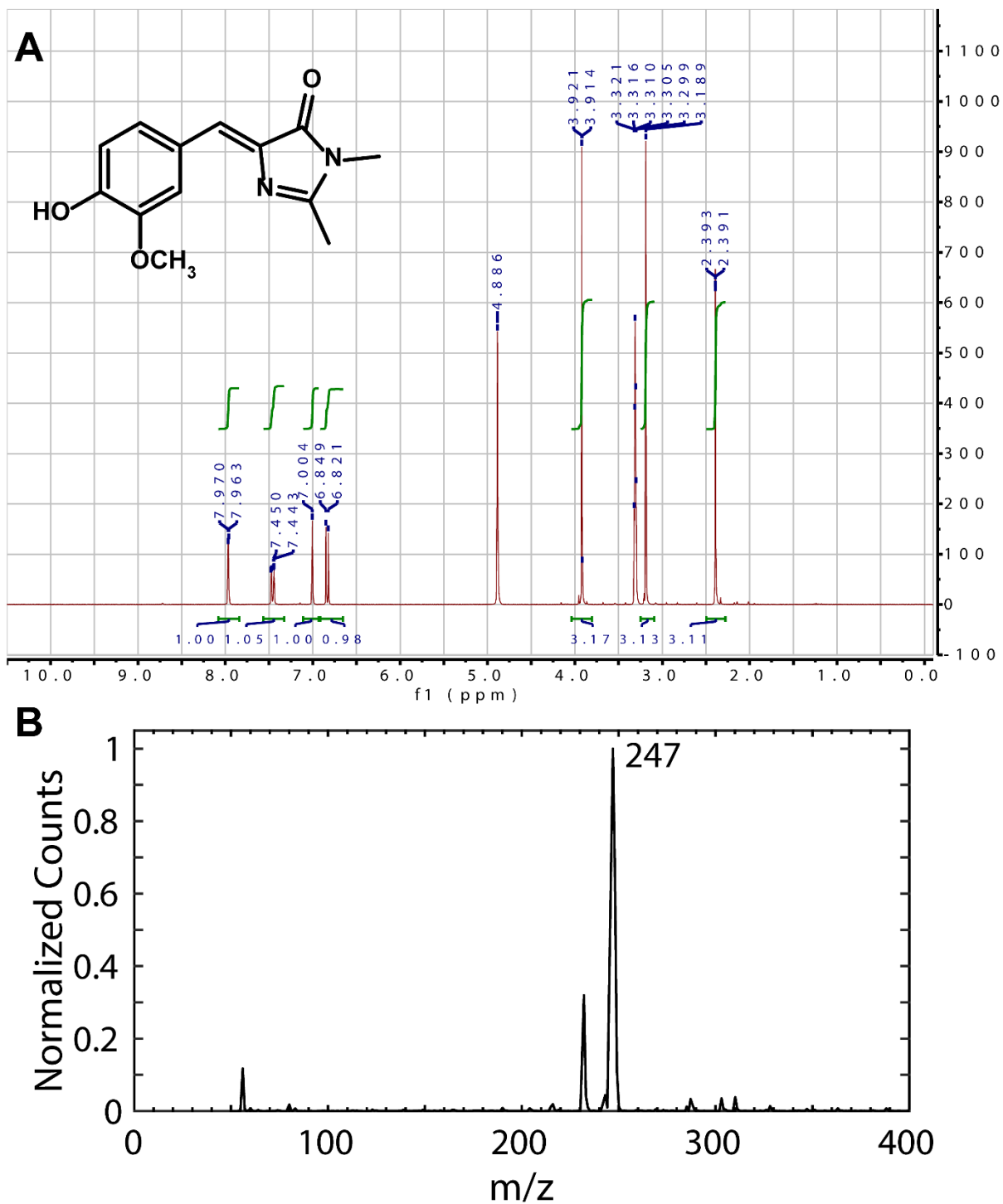


Figure S35. (A) ^1H NMR spectrum (300 MHz, in CD_3OD) and (B) ESI+ mass spectrum of 3-OMeHBDI.

S14 Supplementary Tables

Table S11. Summary of the second-derivative analysis of GFP mutants, Dronpa mutants, and model chromophore analogues at 77 K (Figures S24 and 5B, presented in order of the 0–0 peak maximum). The numbering scheme follows Table S12. Primed numbers are given to water/glycerol mixture instead of water.

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	462.0	21645	22957	1312
1	ih ^{mat} (65T) · ih:loop:GFP R96E E222K	468.6	21340	22753	1413
3	ih:GFP S65T E222Q	474.3	21084	22386	1303
4	s10:loop:GFP S65T	484.2	20653	22002	1350
10	supercharged -30	485.4	20602	21944	1343
5	supercharged +36 s10-	485.7	20589	21973	1384
8	ih:GFP S65T	485.7	20589	21930	1341
7	supercharged +36	486.6	20551	21901	1350
13	s10:loop:GFP S65T T203V	494.1	20239	21617	1378
9	ih:GFP S65T H148D	497.4	20105	21422	1318
12	s10(203F ₅ F) · s10:loop:GFP S65T	497.4	20105	21409	1304
16	s10(203F) · s10:loop:GFP S65T	503.4	19865	21177	1313
14	s10:loop:GFP S65T T203H	510.3	19596	20978	1381
15	s10(203(4-F ₁ F)) · s10:loop:GFP S65T	513.0	19493	20872	1379
19	ih:GFP S65T T203(3-OMeY)	513.0	19493	20846	1353
17	s10:loop:GFP S65T T203Y	514.2	19448	20820	1373

18	<u>s10</u> (203(4-NH ₂ F)) · s10:loop:GFP S65T	516.0	19380	20743	1363
S65 GFP mutants					
20	s10:loop:GFP	465.6	21478	22769	1291
24	s10:loop:GFP T203V	465.6	21478	22722	1244
22	s10:loop:GFP E222Q	467.7	21381	22691	1310
23	avGFP	474.0	21097	22417	1319
25	s10:loop:GFP T203V E222Q	493.2	20276	21673	1397
26	ih:GFP T203(3-OMeY)	502.8	19889	21204	1316
28	s10:loop:GFP T203Y E222Q	503.4	19865	21137	1272
27	s10:loop:GFP T203Y	508.8	19654	20978	1323
Dronpa mutants					
34	Dronpa2	483.9	20665	22002	1337
GFP model chromophore					
63'	HBDI in water/glycerol	442.2	22614	24546	1932
66	HBDI in ethanol	444.0	22523	23878	1355

Table S12. Summary of absorption and emission properties for mutants, variants, and model chromophores of GFP, Dronpa, and PYP at room temperature (Figures 8 and S29).

number	species	absorption maximum (nm)	absorption maximum (cm ⁻¹)	emission maximum (nm)	Stokes shift (cm ⁻¹)	references
S65T GFP mutants						
1	ih ^{mat} (65T) · ih:loop:GFP R96E E222K	460	21739	497.5	1639	this work
2	ih ^{mat} (65T) · ih:loop:GFP R96M	462	21645	502	1725	
3	ih:GFP S65T E222Q	474.5	21075	503.5	1214	
4	s10:loop:GFP S65T	483.2	20695	508.5	1030	
5	supercharged +36 s10-	484	20661	508.5	995	
6	s7:loop:GFP S65T	484	20661	508	976	[1]
7	supercharged +36	485	20619	508	934	this work
8	ih:GFP S65T	485	20619	508	934	
9	ih:GFP S65T H148D	487.5	20513	508	828	
10	supercharged -30	488	20492	509	845	
11	ih ^{mat} (65T) · ih:loop:GFP R96M T203Y	494	20243	515	825	
12	<u>s10</u> (203F ₅ F) · s10:loop:GFP S65T	498.1	20076	513	583	
13	s10:loop:GFP S65T T203V	499.3	20028	511.5	478	
14	s10:loop:GFP S65T T203H	506	19763	519.5	514	
15	<u>s10</u> (203(4-F ₁ F)) · s10:loop:GFP S65T	507.9	19689	518.5	403	
16	<u>s10</u> (203F) · s10:loop:GFP S65T	508.1	19681	520.5	469	

17	s10:loop:GFP S65T T203Y	508.5	19666	521.5	490	
18	<u>s10</u> (203(4- NH ₂ F)) · s10:loop:GFP S65T	509.5	19627	519.5	378	
19	ih:GFP S65T T203(3-OMeY)	510.6	19585	520.5	373	
S65 GFP mutants						
20	s10:loop:GFP	467	21413	503	1533	[1]
21	s10:loop:GFP F145W	469	21322	502	1402	this work
22	s10:loop:GFP E222Q	471	21231	504	1390	[1]
23	avGFP	475	21050	503	1172	this work
24	s10:loop:GFP T203V	487.5	20513	504.5	691	
25	s10:loop:GFP T203V E222Q	492	20325	507.5	621	
26	ih:GFP T203(3-OMeY)	498	20080	512	549	
27	s10:loop:GFP T203Y	503	19881	520	650	[1]
28	s10:loop:GFP T203Y E222Q	504	19841	520	611	
Dronpa mutants						
29	Dronpa2 T159E	472.5	21164	508	1479	this work
30	Dronpa2 V157D	482.8	20713	510	1105	
31	Dronpa2 V157N	485.2	20610	512	1079	
32	Dronpa2 V157T	490.8	20375	513	882	
33	Dronpa2 T159Q	490.8	20375	513.5	901	
34	Dronpa2	492.4	20309	513.5	834	
35	Dronpa2 V157L	494.0	20243	513	750	
36	Dronpa2 S142A	504.4	19826	518.5	539	
37	Dronpa	506.8	19732	519	464	
GFP variants (S65T)						
38	ih:GFP S65T Y66(3-NO ₂ Y)	474	21097	507.5	1393	this work

39	ih:GFP S65T Y66(2,3-F ₂ Y)	475	21053	505.5	1270	
40	ih:GFP S65T Y66(2,3,5-F ₃ Y)	476	21008	506	1246	
41	ih:GFP S65T Y66(3,5-F ₂ Y)	482	20747	513.5	1273	
42	ih:GFP S65T Y66(3-F ₁ Y)	483	20704	512	1173	
43	ih:GFP S65T Y66(3,5-Cl ₂ Y)	492	20325	518	1020	
44	ih:GFP S65T Y66(3-Cl ₁ Y)	493	20284	513.5	810	
45	ih:GFP S65T Y66(3-Br ₁ Y)	498	20080	515	663	
46	ih:GFP S65T Y66(3-CH ₃ Y)	502	19920	517	578	
47	ih:GFP S65T Y66(3-I ₁ Y)	504	19841	520	611	
48	ih:GFP S65T Y66(3-OMeY)	517	19342	530	474	
Dronpa2 variants						
49	Dronpa2 Y63(3-NO ₂ Y)	459	21786	512	2255	this work
50	Dronpa2 Y63(2,3,5-F ₃ Y)	479	20877	510	1269	
51	Dronpa2 Y63(2,3-F ₂ Y)	481	20790	508	1105	
52	Dronpa2 Y63(3,5-F ₂ Y)	483	20704	516	1324	
53	Dronpa2 Y63(3-F ₁ Y)	489	20450	514	995	
54	Dronpa2 Y63(3-Cl ₁ Y)	489	20450	516	1070	
55	Dronpa2 Y63(3-Br ₁ Y)	489	20450	517	1108	
56	Dronpa2 Y63(3-I ₁ Y)	489	20450	516	1070	
57	Dronpa2 Y63(3-CH ₃ Y)	499	20040	521	846	
58	Dronpa2 Y63(3-OMeY)	507	19724	546	1409	
GFP model chromophore and analogues						
59	3-CNHBDI in water	408	24510	485	3891	[118]
60	3,5-F ₂ HBDI in water	417	23981	486	3405	this work

61	3-ClHBDI in water	424	23585	497	3464	
62	3-BrHBDI in water	424	23585	496	3424	
63	HBDI in water	425	23529	491	3163	[118]
64	3-CH ₃ HBDI in water	437	22883	510	3275	this work
65	2,6- (CH ₃) ₂ HBDI in ethanol	442	22624	493.5	2361	[119]
66	HBDI in ethanol	442	22624	495.5	2361	
67	3-OHHBDI in water	442	22624	518	3319	[118]
68	3-OMeHBDI in water	447	22371	522	3214	this work
69	3,5-F ₂ HBDI with aptamer (Spinach)	469	21322	501	1362	[14]
70	3,5- (CH ₃) ₂ HBDI in ethanol	476	21008	502.5	1108	[119]
71	3,5-(t- Bu) ₂ HBDI ^a in ethanol	508	19685	521.5	510	
PYP mutants						
72	E46P	441	22676	501.5	2736	[32]
73	E46D	442	22624	501	2664	
74	E46R	442	22624	502	2704	
75	E46K	444	22523	503	2642	
76	E46Y	445	22472	500	2472	
77	wild type	446	22422	500.5	2442	
78	P68A	447	22371	502	2451	[120]
79	Y98A	447	22371	502	2451	
80	Y98Q	447	22371	499	2331	
81	R52E	447.5	22346	498	2266	[121]
82	R52Q	447.5	22346	496	2185	
83	R52K	447.5	22346	494	2103	
84	E46H	454	22026	501.5	2086	[32]
85	Y42F	454	22026	500	2026	[122]
86	T50V	456	21930	497	1809	[123]
87	E46Q	461	21692	506	1929	[32]
88	E46W	461	21692	503	1811	
89	E46G	463	21598	503.5	1737	

90	E46N	463	21598	504.5	1777	
91	E46S	464	21552	505.5	1769	
92	E46A	469	21322	507	1598	
93	E46F	470	21277	503	1396	
94	E46C	472	21186	504	1345	
95	E46L	475	21053	508.5	1387	
96	E46T	475	21053	505.5	1270	
97	E46M	476	21008	511	1439	
98	E46I	478	20921	509.5	1293	
99	E46V	478	20921	508	1235	
PYP variants						
100	wild type with caffeic acid (3-OH)	457	21882	536	3225	[124]
101	wild type with ferulic acid (3-OMe)	461	21692	526	2681	[125]
102	E46Q with ferulic acid (3-OMe)	473	21142	529	2238	
103	wild type with sinapinic acid (3,5-OMe ₂)	486	20576	547	2295	
104	E46Q with sinapinic acid (3,5-OMe ₂)	495	20202	550	2020	
PYP model chromophore and analogues						
105	pCT ^b in water	395	25316	501.5	5376	[126]
106	pCT in methanol	397.7	25126	484.8	4518	
107	pCT in ethylene glycol	400	25000	494.5	4778	
108	pCT in 1-decanol	401.3	24919	479.2	4051	
109	pCT in ethanol	409.6	24414	484.8	3787	
110	pCT in DMF ^c	467	21413	502	1493	

^a t-Bu = *tert*-butyl

^b pCT = phenyl thioester derivative of *p*-coumaric acid

^c DMF = dimethylformamide

Table S13. Summary of driving forces for mutants, variants, and model chromophores of GFP, Dronpa, and PYP at room temperature, calculated from Table S12 using Equation S13. Outliers in Figures 8 and S29 are excluded due to the lack of estimation for V_0 . The values used for the prediction of S65T R96M/T203Y absorption maximum (Table 2) are shown in red. The localization degree on the phenolate moiety is estimated using Equation S24. The numbering scheme follows Table S12.

number	species	driving force (cm ⁻¹)	localization degree on the phenolate
S65T GFP mutants			
1	ih ^{mat} (65T) · ih:loop:GFP R96E E222K	10470	74.1%
2	ih ^{mat} (65T) · ih:loop:GFP R96M	10270	73.7%
3	ih:GFP S65T E222Q	9010	71.4%
4	s10:loop:GFP S65T	8080	69.5%
5	supercharged +36 s10-	7990	69.3%
6	s7:loop:GFP S65T	7990	69.3%
7	supercharged +36	7880	69.1%
8	ih:GFP S65T	7880	69.1%
9	ih:GFP S65T H148D	7600	68.5%
10	supercharged -30	7540	68.4%
11	ih ^{mat} (65T) · ih:loop:GFP R96M T203Y	6840	66.9%
12	<u>s10</u> (203F ₅ F) · s10:loop:GFP S65T	6320	65.7%
13	s10:loop:GFP S65T T203V	6170	65.4%
14	s10:loop:GFP S65T T203H	5250	63.3%
15	<u>s10</u> (203(4-F ₁ F)) · s10:loop:GFP S65T	4960	62.6%
16	<u>s10</u> (203F) · s10:loop:GFP S65T	4930	62.5%
17	s10:loop:GFP S65T T203Y	4870	62.4%
18	<u>s10</u> (203(4-NH ₂ F)) · s10:loop:GFP S65T	4710	62.0%
19	ih:GFP S65T T203(3-OMeY)	4530	61.6%
S65 GFP mutants			
20	s10:loop:GFP	9770	72.8%
21	s10:loop:GFP F145W	9570	72.4%
22	s10:loop:GFP E222Q	9370	72.1%
23	avGFP	8950	71.3%
24	s10:loop:GFP T203V	7600	68.5%
25	s10:loop:GFP T203V E222Q	7080	67.4%
26	ih:GFP T203(3-OMeY)	6340	65.8%
27	s10:loop:GFP T203Y	5680	64.3%
28	s10:loop:GFP T203Y E222Q	5530	63.9%

Dronpa mutants			
29	Dronpa2 T159E	9210	71.8%
30	Dronpa2 V157D	8120	69.6%
31	Dronpa2 V157N	7860	69.1%
32	Dronpa2 V157T	7220	67.7%
33	Dronpa2 T159Q	7220	67.7%
34	Dronpa2	7030	67.3%
35	Dronpa2 V157L	6840	66.9%
36	Dronpa2 S142A	5480	63.8%
37	Dronpa	5130	63.0%
GFP variants (S65T)			
39	ih:GFP S65T Y66(2,3-F ₂ Y)	8950	71.3%
40	ih:GFP S65T Y66(2,3,5-F ₃ Y)	8850	71.1%
41	ih:GFP S65T Y66(3,5-F ₂ Y)	8210	69.8%
42	ih:GFP S65T Y66(3-F ₁ Y)	8100	69.6%
43	ih:GFP S65T Y66(3,5-Cl ₂ Y)	7080	67.4%
44	ih:GFP S65T Y66(3-Cl ₁ Y)	6960	67.2%
45	ih:GFP S65T Y66(3-Br ₁ Y)	6340	65.8%
46	ih:GFP S65T Y66(3-CH ₃ Y)	5810	64.6%
47	ih:GFP S65T Y66(3-I ₁ Y)	5530	63.9%
Dronpa2 variants			
50	Dronpa2 Y63(2,3,5-F ₃ Y)	8530	70.4%
51	Dronpa2 Y63(2,3-F ₂ Y)	8320	70.0%
52	Dronpa2 Y63(3,5-F ₂ Y)	8100	69.6%
53	Dronpa2 Y63(3-F ₁ Y)	7430	68.2%
54	Dronpa2 Y63(3-Cl ₁ Y)	7430	68.2%
55	Dronpa2 Y63(3-Br ₁ Y)	7430	68.2%
56	Dronpa2 Y63(3-I ₁ Y)	7430	68.2%
57	Dronpa2 Y63(3-CH ₃ Y)	6210	65.5%
GFP model chromophore			
63	HBDI in water	13800	79.3%
66	HBDI in ethanol	12200	77.0%
PYP mutants			
72	E46P	12290	77.1%
73	E46D	12200	77.0%
74	E46R	12200	77.0%
75	E46K	12010	76.7%
76	E46Y	11910	76.5%
77	wild type	11820	76.4%
78	P68A	11720	76.2%
79	Y98A	11720	76.2%
80	Y98Q	11720	76.2%
81	R52E	11670	76.1%
82	R52Q	11670	76.1%
83	R52K	11670	76.1%

84	E46H	11050	75.1%
85	Y42F	11050	75.1%
86	T50V	10860	75.8%
87	E46Q	10370	73.9%
88	E46W	10370	73.9%
89	E46G	10170	73.5%
90	E46N	10170	73.5%
91	E46S	10070	73.4%
92	E46A	9570	72.4%
93	E46F	9470	72.2%
94	E46C	9260	71.8%
95	E46L	8950	71.3%
96	E46T	8950	71.3%
97	E46M	8850	71.1%
98	E46I	8640	70.6%
99	E46V	8640	70.6%

Table S14. Summary of 0–0 positions and Stark tuning rates for mutants, variants, and model chromophores of GFP and Dronpa at 77 K (Figure 9). The 0–0 positions are assigned from the reddest negative peaks in second-derivative analysis (Figure S24, entries are presented in order from blue to red). For model chromophores in water/glycerol, Stark tuning rates are extracted from the classical Stark (sum-of-derivative) analysis (Figure S10), and the parentheses imply non-classical Stark effects from Class II/III borderline behaviors (Figures S14 and S15). The numbering follows Table S12. Primed numbers are given to water/glycerol mixture instead of water.

	species	0–0 position (nm)	0–0 position (cm ⁻¹)	Stark tuning rate (D)
S65T GFP mutants				
2	ih ^{mat} (65T) · ih:loop:GFP R96M	462.0	21645	11.2
1	ih ^{mat} (65T) · ih:loop:GFP R96E E222K	468.6	21340	11.9
3	ih:GFP S65T E222Q	474.3	21084	10.5
4	s10:loop:GFP S65T	484.2	20653	9.3
10	supercharged -30	485.4	20602	9.7
5	supercharged +36 s10-	485.7	20589	9.8
8	ih:GFP S65T	485.7	20589	10.2
7	supercharged +36	486.6	20551	9.0
13	s10:loop:GFP S65T T203V	494.1	20239	6.9
9	ih:GFP S65T H148D	497.4	20105	7.2
12	s10(203F ₅ F) · s40:loop:GFP S65T	497.4	20105	6.3
16	s10(203F) · s40:loop:GFP S65T	503.4	19865	5.9
14	s10:loop:GFP S65T T203H	510.3	19596	5.5
15	s10(203(4-F ₁ F)) · s40:loop:GFP S65T	513.0	19493	4.8
19	ih:GFP S65T T203(3-OMeY)	513.0	19493	4.9
17	s10:loop:GFP S65T T203Y	514.2	19448	5.2

18	<u>s10(203(4-NH₂F)) · s10:loop:GFP S65T</u>	516.0	19380	5.0
S65 GFP mutants				
20	s10:loop:GFP	465.6	21478	11.3
22	s10:loop:GFP E222Q	467.7	21381	11.4
23	avGFP	474.0	21097	10.7
25	s10:loop:GFP T203V E222Q	493.2	20276	8.0
26	ih:GFP T203(3-OMeY)	502.8	19889	7.5
28	s10:loop:GFP T203Y E222Q	503.4	19865	6.0
27	s10:loop:GFP T203Y	508.8	19654	6.1
Dronpa mutants				
34	Dronpa2	483.9	20665	9.1
GFP variants (S65T)				
38	ih:GFP S65T Y66(3-NO ₂ Y)	471.0	21231	10.7
39	ih:GFP S65T Y66(2,3-F ₂ Y)	475.2	21044	10.6
40	ih:GFP S65T Y66(2,3,5-F ₃ Y)	478.2	20912	10.2
41	ih:GFP S65T Y66(3,5-F ₂ Y)	480.6	20807	12.2
42	ih:GFP S65T Y66(3-F ₁ Y)	481.5	20768	10.8
43	ih:GFP S65T Y66(3,5-Cl ₂ Y)	492.3	20313	12.6
44	ih:GFP S65T Y66(3-Cl ₁ Y)	492.9	20288	10.1
45	ih:GFP S65T Y66(3-Br ₁ Y)	497.4	20105	8.8
46	ih:GFP S65T Y66(3-CH ₃ Y)	502.2	19912	7.5
47	ih:GFP S65T Y66(3-I ₁ Y)	505.2	19794	8.3
48	ih:GFP S65T Y66(3-OMeY)	519.9	19234	4.8
Dronpa2 variants				
49	Dronpa2 Y63(3-NO ₂ Y)	454.8	21988	11.0
50	Dronpa2 Y63(2,3,5-F ₃ Y)	460.5	21716	10.9

52	Dronpa2 Y63(3,5-F ₂ Y)	469.8	21286	12.0
51	Dronpa2 Y63(2,3-F ₂ Y)	474.9	21057	9.4
54	Dronpa2 Y63(3-Cl ₁ Y)	477.3	20951	11.9
55	Dronpa2 Y63(3-Br ₁ Y)	478.2	20912	11.9
53	Dronpa2 Y63(3-F ₁ Y)	478.5	20899	9.8
56	Dronpa2 Y63(3-I ₁ Y)	480.3	20820	11.3
57	Dronpa2 Y63(3-CH ₃ Y)	486.3	20563	11.5
58	Dronpa2 Y63(3-OMeY)	497.7	20092	10.9
GFP model chromophore and analogues				
60'	3,5-F ₂ HBDI in water/glycerol	437.1	22878	(13.4)
63'	HBDI in water/glycerol	442.2	22614	(15.8)
66	HBDI in ethanol	444.0	22523	12.0
61'	3-ClHBDI in water/glycerol	446.7	22386	(13.6)
62'	3-BrHBDI in water/glycerol	448.8	22282	(13.7)
64'	3-CH ₃ HBDI in water/glycerol	454.5	22002	(14.8)
68'	3-OMeHBDI in water/glycerol	470.7	21245	(16.1)

Table S15. Summary of extinction coefficients, transition dipole moments, and oscillator strengths for mutants of GFP and Dronpa at room temperature (Figure 10). The numbering scheme follows Table S12.

number	species	absorption maximum (nm)	absorption maximum (cm ⁻¹)	extinction coefficient at peak maximum (M ⁻¹ cm ⁻¹)	transition dipole moment (D)	oscillator strength
S65T GFP mutants						
111	ih:GFP S65T in NaOH solution	447	22371	44100	9.79	1.02
1	ih ^{mat} (65T) · ih:loop:GFP R96E E222K	460	21739	47241	9.61	0.995
4	s10:loop:GFP S65T	483	20704	54657	9.85	1.00
8	ih:GFP S65T	485	20619	57783	9.77	0.986
7	supercharged +36	485	20619	58218	9.82	0.994
9	ih:GFP S65T H148D	488	20492	66216	10.1	1.02
12	s10(203F ₅ F) · s10:loop:GFP S65T	498	20080	70032	9.81	0.976
19	ih:GFP S65T T203(3-OMeY)	511	19569	99462	10.1	0.999
S65 GFP mutants						
22	s10:loop:GFP E222Q	471	21231	51947	9.66	0.981
28	s10:loop:GFP T203Y E222Q	504	19841	77433	9.68	0.960
Dronpa mutants						
34	Dronpa2	492	20325	73709	9.91	0.991

S15 References

- [1] 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. U. S. A.* **2017**, *114*, E2146-E2155.
- [2] Kent, K. P.; Oltrogge, L. M.; Boxer, S. G. Synthetic control of green fluorescent protein. *J. Am. Chem. Soc.* **2009**, *131*, 15988-15989.
- [3] Oltrogge, L. M.; Boxer, S. G. Short hydrogen bonds and proton delocalization in green fluorescent protein (GFP). *ACS Cent. Sci.* **2015**, *1*, 148-156.
- [4] Romei, M. G.; Lin, C.-Y.; Mathews, I. I.; Boxer, S. G. *Manuscript under review.*
- [5] Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* **1992**, *111*, 229-233.
- [6] Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **1994**, *263*, 802-805.
- [7] 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.
- [8] Oltrogge, L. M.; Wang, Q.; Boxer, S. G. Ground-state proton transfer kinetics in green fluorescent protein. *Biochemistry* **2014**, *53*, 5947-5957.
- [9] Wood, T. I.; Barondeau, D. P.; Hitomi, C.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. *Biochemistry* **2005**, *44*, 16211-26220.
- [10] Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.
- [11] Do, K.; Boxer, S. G. Thermodynamics, kinetics, and photochemistry of β -strand association and dissociation in a split-GFP system. *J. Am. Chem. Soc.* **2011**, *133*, 18078-18081.
- [12] Gill, S. C.; von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **1989**, *182*, 319-326.

- [13] 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. U. S. A.* **1989**, *86*, 8247-8251.
- [14] Paige, J. S.; Wu, K. Y.; Jaffrey, S. R. RNA mimics of green fluorescent protein. *Science* **2010**, *333*, 642-646.
- [15] Good, N. E.; Winget, G. D.; Winter, W.; Connolly, T. N.; Izawa, S.; Singh, R. M. M. Hydrogen ion buffers for biological research. *Biochemistry* **1966**, *5*, 467-477.
- [16] Bublitz, G. U.; Boxer, S. G. Stark spectroscopy: applications in chemistry, biology, and materials science. *Annu. Rev. Phys. Chem.* **1997**, *48*, 213-242.
- [17] 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.
- [18] Andrews, S. S.; Boxer, S. G. A liquid nitrogen immersion cryostat for optical measurements. *Rev. Sci. Instr.* **2000**, *71*, 3567-3569.
- [19] Parson, W. W. *Modern Optical Spectroscopy*, 2nd ed.; Springer-Verlag: New York, 2015; pp 123-223.
- [20] Using the SSRL automated mounting (SAM) system:
http://smb.slac.stanford.edu/users_guide/manual/Using_SSRL_Automated_Mounting.html (accessed Jul 5, 2019).
- [21] Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.
- [22] Kabsch, W. XDS. *Acta Cryst.* **2010**, *D66*, 125-132.

- [23] Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Cryst.* **2010**, *D66*, 133-144.
- [24] Using the *autoxds* script:
http://smb.slac.stanford.edu/facilities/software/xds/#autoxds_script (accessed Jul 5, 2019).
- [25] 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.
- [26] Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst.* **2004**, *D60*, 2126-2132.
- [27] Joosten, R. P.; Joosten, K.; Cohen, S. X.; Vriend, G.; Perrakis, A. Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank. *Bioinformatics* **2011**, *27*, 3392-3398.
- [28] Treynor, T. P.; Boxer, S. G. A theory of intervalence band Stark effects. *J. Phys. Chem. A* **2004**, *108*, 1764-1778.
- [29] Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*, 1st ed.; Oxford University Press: New York, NY, 1989.
- [30] Parthey, M.; Kaupp, M. Quantum-chemical insights into mixed-valence systems: within and beyond the Robin–Day scheme. *Chem. Soc. Rev.* **2014**, *43*, 5067-5088.
- [31] Olsen, S.; McKenzie, R. H. Bond alternation, polarizability, and resonance detuning in methine dyes. *J. Chem. Phys.* **2011**, *134*, 114520.
- [32] Philip, A. F.; Nome, R. A.; Papadantonakis, G. A.; Scherer, N. F.; Hoff, W. D. Spectral tuning in photoactive yellow protein by modulation of the shape of the excited state energy surface. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 5821-5826.
- [33] Slocum, J. D.; Webb, L. J. Nitriles probes of electric field agree with independently measured fields in green fluorescent protein even in the presence of hydrogen bonding. *J. Am. Chem. Soc.* **2016**, *138*, 6561-6570.

- [34] 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.
- [35] Marder, S. R.; Perry, J. W.; Tiemann, B. G.; Gorman, C. B.; Gilmour, S.; Biddle, S. L.; Bourhill, G. Direct observation of reduced bond-length alternation in donor/acceptor polyenes. *J. Am. Chem. Soc.* **1993**, *115*, 2524-2526.
- [36] 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.
- [37] Bublitz, G.; King, B. A.; Boxer, S. G. Electronic structure of the chromophore in green fluorescent protein (GFP). *J. Am. Chem. Soc.* **1998**, *120*, 9370-9371.
- [38] Andrews, S. S.; Boxer, S. G. Vibrational Stark effects of nitriles I. Methods and experimental results. *J. Phys. Chem. A* **2000**, *104*, 11853-11863.
- [39] Geissinger, P.; Kohler, B. E.; Woehl, J. C. Electric field and structure in the myoglobin heme pocket. *J. Phys. Chem.* **1995**, *99*, 16527-16529.
- [40] Köhler, M.; Friedrich, J.; Fidy, J. Proteins in electric fields and pressure fields: basic aspects. *Biochim. Biophys. Acta* **1998**, *1386*, 255-288.
- [41] Vauthey, E.; Voss, J.; de Caro, C.; Renn, A.; Wild, U. P. Spectral hole-burning and Stark effect: frequency dependence of the induced dipole moment of a squaraine dye in polymers. *Chem. Phys.* **1994**, *184*, 347-356.
- [42] Hochstrasser, R. M. Electric field effects on oriented molecules and molecular crystals. *Acc. Chem. Res.* **1973**, *6*, 263-269.
- [43] 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.
- [44] Onsager, L. Electrical moments of molecules in liquids. *J. Am. Chem. Soc.* **1936**, *58*, 1486-1493.
- [45] List, N. H.; Jensen, H. J. A.; Kongsted, J. Local electric fields and molecular properties in heterogeneous environments through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 10070-10080.
- [46] Jackson, J. D. *Classical Electrodynamics*, 3rd ed.; Wiley: Hoboken, NJ, 1999.

- [47] Fried, S. D.; Boxer, S. G. Measuring electric fields and noncovalent interactions using the vibrational Stark effect. *Acc. Chem. Res.* **2015**, *48*, 998-1006.
- [48] Lawrence, M. S.; Phillips, K. J.; Liu, D. R. Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* **2007**, *129*, 10110-10112.
- [49] Fluorescent protein properties: <http://www.fpvis.org/FP.html> (accessed Jul 5, 2019).
- [50] Shaner, N. C.; Lambert, G. G.; Chamma, A.; Ni, Y.; Cranfill, P. J.; Baird, M. A.; Sell, B. R.; Allen, J. R.; Day, R. N.; Israelsson, M.; Davidson, M. W.; Wang, J. A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat. Methods* **2013**, *10*, 407-409.
- [51] Griesbeck, O.; Baird, G. S.; Campbell, R. E.; Zacharias, D. A.; Tsien, R. Y. Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. *J. Biol. Chem.* **2001**, *276*, 29188-29194.
- [52] Slocum, J. D.; First, J. T.; Webb, L. J. Orthogonal electric field measurements near the green fluorescent protein fluorophore through Stark effect spectroscopy and pK_a shifts provide a unique benchmark for electrostatics models. *J. Phys. Chem. B* **2017**, *121*, 6799-6812.
- [53] Kaila, V. R. I.; Send, R.; Sundholm, D. Electrostatic spectral tuning mechanism of the green fluorescent protein. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4491-4995.
- [54] Landau, L. D.; Lifshitz, E. M. *Course of Theoretical Physics Volume 2: The Classical Theory of Fields*, 4th ed.; Butterworth-Heinemann: Oxford, UK, 2000.
- [55] Farid, T. A.; Kodali, G.; Solomon, L. A.; Lichtenstein, B. R.; Sheehan, M. M.; Fry, B. A.; Bialas, C.; Ennist, N. M.; Siedlecki, J. A.; Zhao, Z.; Stetz, M. A.; Valentine, K. G.; Anderson, J. L. R.; Wand, A. J.; Discher, B. M.; Moser, C. C.; Dutton, P. L. Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nat. Chem. Biol.* **2013**, *9*, 826-833.
- [56] Liu, J.; Chakraborty, S.; Hosseinzadeh, P.; Yu, Y.; Tian, S.; Petrik, I.; Bhagi, A.; Lu, Y. Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chem. Rev.* **2014**, *114*, 4366-4469.

- [57] Gottfried, D. S.; Steffen, M. A.; Boxer, S. G. Large protein-induced dipoles for a symmetric carotenoid in a photosynthetic antenna complex. *Science* **1991**, *251*, 662-665.
- [58] McConnell, H. M. Theory of nuclear magnetic shielding in molecules: I. Long-range dipolar shielding of protons. *J. Chem. Phys.* **1957**, *27*, 226-229.
- [59] Buckingham, A. D.; Longuet-Higgins, H. C. The quadrupole moments of dipolar molecules. *Mol. Phys.* **1968**, *14*, 63-72.
- [60] 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.
- [61] Thompson, L. M.; Lasoroski, A.; Champion, P. M.; Sage, J. T.; Frisch, M. J.; van Thor, J. J.; Bearpark, M. J. Analytical harmonic vibrational frequencies for the green fluorescent protein computed with ONIOM: chromophore mode character and its response to environment. *J. Chem. Theory Comput.* **2014**, *10*, 751-766.
- [62] Park, J. W.; Rhee, Y. M. Electric field keeps chromophore planar and produces high yield fluorescence in green fluorescent protein. *J. Am. Chem. Soc.* **2016**, *138*, 13619-13629.
- [63] Andersen, L. H.; Lapierre, A.; Nielsen, S. B.; Nielsen, I. B.; Pedersen, S. U.; Pedersen, U. V.; Tomita, S. Chromophores of the green fluorescent protein studied in the gas phase. *Eur. Phys. J. D* **2002**, *20*, 597-600.
- [64] Treynor, T. P.; Yoshina-Ishii, C.; Boxer, S. G. Probing excited-state electron transfer by resonance Stark spectroscopy: 4. Mutants near B_L in photosynthetic reaction centers perturb multiple factors that affect B_L* → B_L⁺H_L⁻. *J. Phys. Chem. B* **2004**, *108*, 13523-13535.
- [65] 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.
- [66] 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.

- [67] Scholes, G. D. Long-range resonance energy transfer in molecular systems. *Annu. Rev. Phys. Chem.* **2003**, *54*, 57-87.
- [68] Scholes, G. D.; Fleming, G. R. Energy Transfer and Photosynthetic Light Harvesting. In *Advances in Chemical Physics, Volume 132*; Berry, R. S.; Jortner, J., Eds.; Wiley: New York, NY, 2006; pp 57-129.
- [69] Förster, Th. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **1948**, *437*, 55-75. English translation by Knox, R. S. Intermolecular Energy Migration and Fluorescence. In *Biological Physics*; Mielczarek, E. V.; Greenbaum, E.; Knox, R. S., Eds.; American Institute of Physics: New York, 1993; pp 148-160.
- [70] Parson, W. W. *Modern Optical Spectroscopy*, 2nd ed.; Springer-Verlag: New York, 2015; pp 325-352.
- [71] Krueger, B. P.; Scholes, G. D.; Fleming, G. R. Calculation of couplings and energy-transfer pathways between the pigments of LH2 by the ab initio transition density cube method. *J. Phys. Chem. B* **1998**, *102*, 5378-5386.
- [72] Krueger, B. P.; Yom, J.; Walla, P. J.; Fleming, G. R. Observation of the S₁ state of spheroidene in LH2 by two-photon fluorescence excitation. *Chem. Phys. Lett.* **1999**, *310*, 57-64.
- [73] Walla, P. J.; Linden, P. A.; Hsu, C.-P.; Scholes, G. D.; Fleming, G. R. Femtosecond dynamics of the forbidden carotenoid S₁ state in light-harvesting complexes of purple bacteria observed after two-photon excitation. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10808-10813.
- [74] Hsu, C.-P.; Walla, P. J.; Head-Gordon, M.; Fleming, G. R. The role of the S₁ state of carotenoids in photosynthetic energy transfer: the light-harvesting complex II of purple bacteria. *J. Phys. Chem. B* **2001**, *105*, 11016-11025.
- [75] Stanley, R. J.; King, B. A.; Boxer, S. G. Excited state energy transfer pathways in photosynthetic reaction centers. 1. Structural symmetry effects. *J. Phys. Chem.* **1996**, *100*, 12052-12059.
- [76] King, B. A.; McAnaney, T. B.; de Winter, A.; Boxer, S. G. Excited state energy transfer pathways in photosynthetic reaction centers. 3. Ultrafast emission from the monomeric bacteriochlorophylls. *J. Phys. Chem. B* **2000**, *104*, 8895-8902.

- [77] Scholes, G. D.; Jordanides, X. J.; Fleming, G. R. Adapting the Förster theory of energy transfer for modeling dynamics in aggregated molecular assemblies. *J. Phys. Chem. B* **2001**, *105*, 1640-1651.
- [78] Jordanides, X. J.; Scholes, G. D.; Fleming, G. R. The mechanism of energy transfer in the bacterial photosynthetic reaction center. *J. Phys. Chem. B* **2001**, *105*, 1652-1669.
- [79] London, F. On centers of van der Waals attraction. *J. Phys. Chem.* **1942**, *46*, 305-316.
- [80] Michl, J.; Bonačić-Koutecký, V. *Electronic Aspects of Organic Photochemistry*, 1st ed.; Wiley, Chichester, UK, 1990.
- [81] Shaik, S.; Hiberty, P. C. *A Chemist's Guide to Valence Bond Theory*, 1st ed.; Wiley: Hoboken, NJ, 2008.
- [82] Olsen, S. A modified resonance-theoretic framework for structure-property relationships in a halochromic oxonol dye. *J. Chem. Theory Comput.* **2010**, *6*, 1089-1103.
- [83] Fox, M. A.; Matsen, F. A. Electronic structure in π systems. Part I. Hückel theory with electronic repulsion. *J. Chem. Educ.* **1985**, *62*, 367-373.
- [84] Altland, A.; Simons, B. *Condensed Matter Field Theory*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2010.
- [85] Durand, P.; Malrieu, J.-P. Effective Hamiltonians and Pseudo-Operators as Tools for Rigorous Modeling. In *Advances in Chemical Physics, Volume 67*; Lawley, K. P., Ed.; Wiley: New York, NY, 1987; pp 321-412.
- [86] Shavitt, I.; Redmon, L. T. Quasidegenerate perturbation theories. A canonical Van Vleck formalism and its relationship to other approaches. *J. Chem. Phys.* **1980**, *73*, 5711-5717.
- [87] McConnell, H. M. Intramolecular charge transfer in aromatic free radicals. *J. Chem. Phys.* **1961**, *35*, 508-515.
- [88] Creutz, C.; Newton, M. D.; Sutin, N. Metal–ligand and metal–metal coupling elements. *J. Photochem. Photobiol. A* **1994**, *82*, 47-59.

- [89] Silverman, L. N.; Kanchanawong, P.; Treynor, T. P.; Boxer, S. G. Stark spectroscopy of mixed-valence systems. *Philos. Trans. R. Soc., A* **2008**, *366*, 33-45.
- [90] Su, W. P.; Schrieffer, J. R.; Heeger, A. J. Soliton excitations in polyacetylene. *Phys. Rev. B: Condes. Matter Mater. Phys.* **1980**, *22*, 2099-2111.
- [91] Bublitz, G. Stark Spectroscopy of Donor/Acceptor Substituted Polyenes: A Test of Structure-Function Relationships of the Nonlinear Optical Properties in Conjugated Systems. Ph.D. Dissertation; Stanford University, Stanford, CA, 1997.
- [92] Lao, K.; Moore, L. J.; Zhou, H.; Boxer, S. G. Higher-order Stark spectroscopy: polarizability of photosynthesis pigments. *J. Phys. Chem.* **1995**, *99*, 496-500.
- [93] Creemers, T. M. H.; Lock, A. J.; Subramaniam, V.; Jovin, T. M.; Völker, S. Three photoconvertible forms of green fluorescent protein identified by spectral hole-burning. *Nat. Struct. Biol.* **1999**, *6*, 557-560.
- [94] Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8362-8367.
- [95] Creemers, T. M. H.; Lock, A. J.; Subramaniam, V.; Jovin, T. M.; Völker, S. Photophysics and optical switching in green fluorescent protein mutants. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2974-2978.
- [96] Wiehler, J.; Jung, G.; Seebacher, C.; Zumbusch, A.; Steipe, B. Mutagenic stabilization of the photocycle intermediate of green fluorescent protein (GFP). *ChemBioChem* **2003**, *4*, 1164-1171.
- [97] Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormö, M.; Remington, S. J. Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2306-2311.
- [98] Craggs, T. D. Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev.* **2009**, *38*, 2865-2875.
- [99] Ward, W. W.; Prentice, H. J.; Roth, A. F.; Cody, C. W.; Reeves, S. C. Spectral perturbations of the *Aequorea* green-fluorescent protein. *Photochem. Photobiol.* **1982**, *35*, 803-808.

- [100] Pierce, D. W.; Boxer, S. G. Stark effect spectroscopy of tryptophan. *Biophys. J.* **1995**, *68*, 1583-1591.
- [101] Silverman, L. N.; Spry, D. B.; Boxer, S. G.; Fayer, M. D. Charge transfer in photoacids observed by Stark spectroscopy. *J. Phys. Chem. A* **2008**, *112*, 10244-10249.
- [102] Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T.; Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **2006**, *24*, 79-88.
- [103] 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.
- [104] 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.
- [105] 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.
- [106] Maddalo, S. L.; Zimmer, M. The role of the protein matrix in green fluorescent protein fluorescence. *Photochem. Photobiol.* **2006**, *82*, 367-372.
- [107] Olsen, S.; McKenzie, R. H. A two-state model of twisted intramolecular charge-transfer in monomethine dyes. *J. Chem. Phys.* **2012**, *137*, 164319.
- [108] De Meulenaere, E.; Nguyen Bich, N.; de Wergifosse, M.; Van Hecke, K.; Van Meervelt, L.; Vanderleyden, J.; Champagne, B.; Clays, K. Improving the second-order nonlinear optical response of fluorescent proteins: the symmetry argument. *J. Am. Chem. Soc.* **2013**, *135*, 4061-4069.
- [109] Wachter, R. M.; Elsliger, M.-A.; Kallio, K.; Hanson, G. T.; Remington, S. J. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure* **1998**, *6*, 1267-1277.
- [110] McGaughey, G. B.; Gagné, M.; Rappé, A. K. π -Stacking interactions. Alive and well in proteins. *J. Biol. Chem.* **1998**, *273*, 15458-15463.

- [111] Hunter, C. A.; Sanders, J. K. M. The nature of π - π interactions. *J. Am. Chem. Soc.* **1990**, *112*, 5525-5534.
- [112] Wheeler, S. E. Controlling the local arrangements of π -stacked polycyclic aromatic hydrocarbons through substituent effects. *CrystEngComm* **2012**, *14*, 6140-6145.
- [113] Riwar, L.-J.; Trapp, N.; Kuhn, B.; Diederich, F. Substituent effects in parallel-displaced π - π stacking interactions: distance matters. *Angew. Chem., Int. Ed.* **2017**, *56*, 11252-11257.
- [114] Kaucikas, M.; Fitzpatrick, A.; Bryan, E.; Struve, A.; Henning, R.; Kosheleva, I.; Srajer, V.; Groenhof, G.; van Thor, J. J. Room temperature crystal structure of the fast switching M159T mutant of the fluorescent protein Dronpa. *Proteins* **2015**, *83*, 397-402.
- [115] 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.
- [116] Niwa, H.; Inouye, S.; Hirano, T.; Matsuno, T.; Kojima, S.; Kubota, M.; Ohashi, M.; Tsuji, F. I. Chemical nature of the light emitter of the *Aequorea* green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13617-13622.
- [117] Fried, S. D.; Bagchi, S.; Boxer, S. G. Measuring electrostatic fields in both hydrogen-bonding and non-hydrogen-bonding environments using carbonyl vibrational probes. *J. Am. Chem. Soc.* **2013**, *135*, 11181-11192.
- [118] Chen, C.; Baranov, M. S.; Zhu, L.; Baleeva, N. S.; Smirnov, A. Y.; Zaitseva, S. O.; Yampolsky, I. V.; Solntsev, K. M.; Fang, C. Designing redder and brighter fluorophores by synergistic tuning of ground and excited states. *Chem. Commun.* **2019**, *55*, 2537-2540.
- [119] Conyard, J.; Kondo, M.; Heisler, I. A.; Jones, G.; Baldrige, A.; Tolbert, L. M.; Solntsev, K. M.; Meech, S. R. Chemically modulating the photophysics of the GFP chromophore. *J. Phys. Chem. B* **2011**, *115*, 1571-1577.
- [120] Kyndt, J. A.; Savvides, S. N.; Memmi, S.; Koh, M.; Fitch, J. C.; Meyer, T. E.; Heyn, M. P.; Van Beeumen, J. J.; Cusanovich, M. A. Structural role of tyrosine 98 in photoactive yellow protein: effects on fluorescence, gateway, and photocycle recovery. *Biochemistry* **2007**, *46*, 95-105.

- [121] Both, J. H. C. Fluorescence and Photoisomerization in Photoactive Proteins. Ph.D. Dissertation; Stanford University, Stanford, CA, 2018.
- [122] 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.
- [123] Changenet-Barret, P.; Plaza, P.; Martin, M. M.; Chosrowjan, H.; Taniguchi, S.; Mataga, N.; Imamoto, Y.; Kataoka, M. Structural effects on the ultrafast photoisomerization of photoactive yellow protein. Transient absorption spectroscopy of two point mutants. *J. Phys. Chem. C* **2009**, *113*, 11605-11613.
- [124] Mataga, N.; Chosrowjan, H.; Taniguchi, S.; Hamada, N.; Tokunaga, F.; Imamoto, Y.; Kataoka, M. Ultrafast photoreactions in protein nanospaces as revealed by fs fluorescence dynamics measurements on photoactive yellow protein and related systems. *Phys. Chem. Chem. Phys.* **2003**, *5*, 2454-2460.
- [125] 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*, 517-579.
- [126] Espagne, A.; Changenet-Barret, P.; Plaza, P.; Martin, M. M. Solvent effect on the excited-state dynamics of analogues of the photoactive yellow protein chromophore. *J. Phys. Chem. A* **2006**, *110*, 3393-3404.