Supporting Information

Design of Large Stokes Shift Fluorescent Proteins Based on Excited State Proton Transfer of an Engineered Photobase

Elizabeth M. Santos,† Wei Sheng,† Rahele Esmatpour Salmani, Setare Tahmasebi Nick, Alireza Ghanbarpour, Hadi Gholami, Chrysoula Vasileiou, James H. Geiger,* Babak Borhan*

Michigan State University, Department of Chemistry, 578 S. Shaw Ln., East Lansing, MI 48824

†These authors contributed equally to this manuscript.

Table of Contents

EXPERIMENTAL SECTION

Site-Directed Mutagenesis of hCRBPII

The pET-17b plasmid, containing hCRBPII-Q108K:K40L cloned between NdeI and XhoI, was used as a template for mutagenesis of hCRBPII.[1] Site-directed mutagenesis was conducted via polymerase chain reaction (PCR), with the specified cycling conditions shown in **Table S1**.

Table S1. PCR cycling conditions for site-directed mutagenesis.

	PCR Program	Time (min)
1x	94 °C	3:00
	94 °C	0:20
20x	3 – 5 °C below T_m	0:55
	72 °C	3:30
1x	72 °C	10:00
1x	4 °C	5:00

Reactant	Volume
DNA template	70 ng (x µL)
Forward primer	20 pmol (y µL)
Reverse primer	20 pmol (z µL)
10 mM dNTP	1 µL
50 mM $MgCl_2$	1 µL
DMSO	5 µL
10 x Cloned Pfu Reaction Buffer	5 µL
Pfu Turbo DNA polymerase (2.5 U/µL)	1 µL
Nuclease-free water	50 µL – x – y – z – 7 µL

The primers used for mutagenesis were ordered from Integrated DNA Technologies (IDT), with melting temperatures (T_m) from approximately 52 °C to 65 °C. The sequences of the forward primers (5' to 3') are listed below. It should be noted that, in all cases, the reverse primer is the reverse complement of the forward primer.

Q4E-GACGAGGGACGAGAATGGAACC

Q4F-GACGAGGGACTTCAATGGAACC

W8E-AATGGAACCGAGGAGATGGAGAGT

Y19W-CTTTGAGGGCTGGATGAAGGC

Q38E:K40L-GTACGTCTCACTGAGACGCTGGTTATTGATCAA

Q38F- GTACGTCTCACTTTTACGAAGGTTATTGAT

Q38F:K40C-GTACGTCTCACTTTTACGTGTGTTATTGAT

Q38F:K40D-GTACGTCTCACTTTTACGGACGTTATTGATCAA

Q38F:K40E-GTACGTCTCACTTTTACGGAAGTTATTGATCAA

Q38F:K40H-GTACGTCTCACTTTTACGCATGTTATTGAT

Q38F:K40L-GTACGTCTCACTTTTACGCTGGTTATTGATCAA

K40L:I42E-ACGCTGGTTGAAGATCAAGATGGT

Q38F:K40Q-GTACGTCTCACTTTTACGCAAGTTATTGAT

Q38F:K40R-GTACGTCTCACTTTTACGCGAGTTATTGAT

Q38F:K40T-GTACGTCTCACTTTTACGACAGTTATTGAT

Q38F:K40Y-GTACGTCTCACTTTTACGTACGTTATTGAT

T51E:T53A-GATAACTTCAAGGAAAAAGCAACTAGCACATTC

T53A-TTCAAGACAAAAGCCACTAGCACATTC

T53E-TTCAAGACAAAAGAGACTAGCACATTC

R58L-CTAGCACATTCCTGAACTATGATGTG

R58W-CTAGCACATTCTGGAACTATGATGTG

V62E-AACTATGATGAGGATTTCACTGTTGGAGTA

F64E-AACTATGATGTGGATGAAACTGTTGGAGTAGAG

V86E-TTAAGGCACTGGAGACCTGGGAAGG

L93E-TGGGAAGGTGATGTCGAAGTGTGTGTGCAAAAG

C95E-GATGTCCTTGTGGAGGTGCAAAAGGGG

W106E:Q108K-GAGAACCGCGGCGAGAAGAAGTGGATTGAGGGG

L115E-GGGGACAAGGAGTACCTGGAGC

L117E-CAAGCTGTACGACGAGCTGACC

L119E-GCTGTACCTGGAGGAGACCTGTGGTGAC

Q128E-CAGGTGTGCCGTGAGGTGTTCAAAAAG

F130E-TGCCGTCAAGTGGAGAAAAAGAAGTGA

The crude PCR product was then digested with 20 units DpnI enzyme (New England Biolabs) for one hour at 37 °C. The resulting solution (7 µL) was then added to *E. coli* XL-1 Blue competent cells (Novagen, 100 µL) on ice for 30 minutes. Subsequently, the cells were heat shocked for 30 seconds at 42 °C and then gently spread on a Luria broth (LB) agar plate supplemented with 100 µg/mL ampicillin and 12.5 µg/mL tetracycline. The plate was incubated at 37 °C for approximately 16 hours.

A single colony was then inoculated into 10 mL LB media supplemented with 100 µg/mL ampicillin and 12.5 µg/mL tetracycline. LB media was prepared by adding 10 g

tryptone, 7 g yeast extract and 5 g NaCl to 1 L dI water. The media was autoclaved and cooled to room temperature before use. The inoculated culture was shaken at 37 °C for 12 hours. DNA purification was performed using a Promega Wizard Plus SV miniprep DNA purification system (A1330). The concentration of the isolated plasmid was measured via Nandrop; the average concentration was 100 ng/µL. Every sample was sequenced by the Research Technology Support Facility at Michigan State University, using a primer corresponding to the T7 promoter for all samples in pET-17b plasmid.

Protein Expression and Purification of hCRBPII in pET-17b

The target gene (100 ng of DNA for 100 µL cell solution) was added to thawed BL21(DE3) pLysS competent cells (Invitrogen) on ice and incubated for 30 minutes. Subsequently, the cells were heat shocked for 30 seconds at 42 °C and then gently spread on a Luria broth (LB) agar plate supplemented with 100 µg/mL ampicillin and 27 µg/mL chloramphenicol. The plate was incubated at 37 °C for approximately 12 hours.

A single colony was then inoculated into 1 L terrific broth (TB) media supplemented with 100 µg/mL ampicillin and 27 µg/mL chloramphenicol. TB media was prepared by mixing two solutions and autoclaving them separately. The first solution consists of 12 g tryptone, 24 g yeast extract and 4 mL glycerol to 900 mL dI water. The second solution was prepared by mixing 2.31 g KH_2PO_4 and 12.54 g of K_2HPO_4 in 90 mL dI water. Before inoculation the solutions were mixed. The inoculated culture was shaken at 37 °C until optical density (OD) at 600 nm was approximately 1; this typically takes eight to nine hours. Overexpression was induced by the addition of isopropyl-β-D-

thiogalactopyranoside (IPTG, Gold Biotechnology) at a final concentration of 1 mM. The culture was then shaken at 23 °C for 20 hours.

The cells were harvested by centrifugation (5000 rpm, 10 min, 4 °C) and resuspended in Tris-binding buffer (10 mM Tris, pH=8.0, 50 mL). The cells were then lysed by sonication (Biologics, Inc, power 60%, 3 min). The solution was again centrifuged to separate the pellet and supernatant (5000 rpm, 30 min, 4 °C). All further protein purification was also conducted at 4 °C.

The supernatant was then loaded onto a FastQ anion exchange column pre equilibrated with Tris buffer (10 mM Tris, pH = 8.0). After binding of the protein to the FastQ anion exchange resin (GE Healthcare), the column was washed twice with Tris buffer (2 x 50 mL). Lastly the protein was eluted with Tris-elution buffer (10 mM Tris, 200 mM sodium chloride, pH = 8.0, 100 mL). The eluent from the FastQ anion exchange column was then desalted with Tris buffer using an ultrafiltration cell under nitrogen pressure (~20 psi) equipped with a 10 kDa molecular weight cutoff membrane (Millipore, Regenerated Cellulose membrane, diameter 63.5 mm, NMWL: 10,000). The protein was first concentrated to ~ 10 mL and then diluted to 150 mL with Tris buffer. This was again concentrated to less than 30 mL.

Purification was continued with Fast Protein Liquid Chromatography (NGC chromatography system, Biorad), equipped with a column loaded with Source 15Q (Q Sepharose Fast Flow, GE Healthcare) anion exchange resin. The method for FPLC source Q is shown in **Table S2**. The pH at all steps was set to 8.1. Percent B corresponds to the percent salt, where 100% is equivalent to 1 mM NaCl.

Table S2. FPLC Source 15Q method.

Step	Description	% B	Volume	Flow Rate
1.	Isocratic flow	0	12 mL	3 mL/min
2.	Load sample	n/a	Sample volume	2 mL/min
3.	Isocratic flow	0	10 mL	3 mL/min
4.	Gradient flow	0 to 4	12 mL	3 mL/min
5.	Isocratic flow	4	20 mL	3 mL/min
6.	Gradient flow	4 to 8	15 mL	3 mL/min
7.	Isocratic flow	8	20 mL	3 mL/min
8.	Gradient flow	8 to 15	15 mL	3 mL/min
9.	Isocratic flow	15	40 mL	3 mL/min
10.	Gradient flow	15 to 75	10 mL	3 mL/min
11.	Isocratic flow	100	20 mL	3 mL/min
12.	Isocratic flow	0	35 mL	3 mL/min

Protein was then collected from 40 mM, 80 mM or 150 mM NaCl and concentrated to 1 mL using a 10 kDa Centriprep centrifugal filter (Millipore, Regenerated Cellulose membrane, NMWL: 10,000). The concentrated sample was then loaded to the Fast Protein Liquid Chromatography (NGC chromatography system, Biorad), equipped with a column loaded with size exclusion chromatography (SEC) Superdex 75 Prep Grade resin (GE Healthcare). The method for FPLC source Q is shown in **Table S3**. At all steps, the pH was set to 8.1.

Table S3. FPLC SEC method.

Step	Description	% B	Volume	Flow Rate
1.	Load sample	0	4 mL	1 mL/min
2.	Isocratic flow	20	139 mL	1 mL/min

Determination of Extinction Coefficient

The extinction coefficients of the proteins were measured at 280 nm, following the method described by Gill and von Hippel.[2] The theoretical extinction coefficient (ε_{theor}) is calculated by the following equation:

$$\varepsilon_{theor} = a \times \varepsilon_{Trp} + b \times \varepsilon_{Tyr} + c \times \varepsilon_{Cys}$$

where a, b and c are the number of tryptophans, tyrosines and cysteine residues, respectively. The extinctions of tryptophan, tyrosine and cysteine are 5690 M^{-1} cm^{-1}, 1280 M^{-1} cm^{-1} and 120 M^{-1} cm^{-1}, respectively.

The absorbance at 280 nm of the protein was measured at the same concentration under native (2 x PBS buffer) and denaturing (final concentration 6 M guanidine HCl) conditions. The ratio of absorbance intensities under native (A_{native}) and denaturing ($A_{denaturing}$) conditions, multiplied by ε_{theor} yielded ε_{exp} as shown in the following equation.

$$\varepsilon_{exp} = \frac{A_{native}}{A_{denaturing}} \times \varepsilon_{theor}$$

A stock solution of PBS buffer (10 x) was prepared by dissolving 2.0 g KCl (26 mM), 2.4 g KH_2PO_4 (17.6 mM), 80 g NaCl (1368 mM), 11.45 g Na_2HPO_4 (80.7 mM) in 1000 mL dI water. The solution was autoclaved and then diluted to 2 x with autoclaved dI water.

Table S4. Extinction coefficients of hCRBPII mutants.

Mutant	Protein	$\varepsilon_{exp(280\ nm)}$
M1	Q108K:K40L:T53A:R58L:Q38F:Q4F	31,014
M2	Q108K:K40L:T53A:R58L:Q4F:Q38E	29,153
M3	Q108K:K40E:T53A:R58L:Q38F:Q4F	27,681
M4	Q108K:K40L:T53A:R58L:Q38F:Q4F:I42E	29,326
M5	Q108K:K40L:T53A:R58L:Q38F:Q4F:T51E	29,050
M6	Q108K:K40L:T53E:R58L:Q38F:Q4F	27,822
M7	Q108K:K40L:T53A:R58L:Q38F:Q4F:V62E	28,660
M8	Q108K:K40L:T53A:R58L:Q38F:Q4F:L117E	28,159
M9	Q108K:K40L:T53A:R58L:Q38F:Q4F:L119E	29,595
M10	Q108K:K40L:T53A:R58L:Q38F:Q4F:Q128E	30,636
M11	Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W	38,874
M12	Q108K:K40D:T53A:R58L:Q38F:Q4F	28,682
M13	Q108K:K40C:T53A:R58L:Q38F:Q4F	27,948
M14	Q108K:K40T:T53A:R58L:Q38F:Q4F	28,048
M15	Q108K:K40H:T53A:R58L:Q38F:Q4F	27,742
M16	Q108K:T53A:R58L:Q38F:Q4F	29,370
M17	Q108K:K40R:T53A:R58L:Q38F:Q4F	27,776
M18	Q108K:K40Q:T53A:R58L:Q38F:Q4F	28,759
M19	Q108K:K40Y:T53A:R58L:Q38F:Q4F	28,587

UV-Vis measurements of hCRBPII/chromophore complexes

UV-Vis spectra were recorded with a Cary 300 Bio WinUV, Varian spectrophotometer. For all experiments, 20 μM protein was incubated with ligand (10 μM, 0.5 equiv.) in PBS buffer and incubated at room temperature until Schiff base (SB) or protonated Schiff base (PSB) formation was complete. This was verified by UV-Vis.

pK_a measurements of hCRBPII/chromophores complexes

For pK_a determination, protein (20 µM in PBS) was incubated with ligand (0.5 equiv) at room temperature until Schiff base (SB) or protonated Schiff base (PSB) formation was complete. This was verified by UV-Vis. The solution was then titrated with acid (1 M HOAc, adjusted to pH 4) or base (1 M NaOH) in ~ 0.3 pH units, and the absorption spectra were recorded at each point.

Absorbance change at λ_{max} was plotted as a function of pH. A curve fit, as previously described for bacteriorhodopsin, was applied for pK_a determination:

$$\Delta A = \frac{\Delta A_0}{(1 + 10^{pH - pK_a})} + constant$$

The two parameters were: ΔA_0, the total absorbance change of the PSB and pK_a, the midpoint of titration. It should be noted that a constant is included to account for the deviation from zero absorbance intensity of the deprotonated PSB.

Fluorescence measurements

Fluorescence spectra were recorded using a Fluorolog®-3 spectrofluorometer (HORIBA, Ltd.). An entrance slit of 1 nm and exit slit of 12 nm was used for all measurements.

Kinetic measurements of hCRBPII/chromophore PSB formation

For kinetic measurements, SB formation was followed by fluorescence in PBS buffer (~ pH 7.2) at 23 °C using a Fluorolog®-3 spectrofluorometer. pH was verified before recording the spectrum. The experiment was performed with a final protein concentration of 20 µM protein and 0.5 equiv ligand. The spectra were recorded immediately after

mixing the protein and chromophore and the fluorescence intensity plotted as a function of time.

The data were fit to a second order rate equation considering multiple reagents A and B with non-equal concentrations. The reader is referred to the following description for the equations utilized to obtain the rate descriptions used here (https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Kinetics/02%3A_Reaction_Rates/2.08%3A_Second-Order_Reactions).

Assuming the protein-ligand reaction follows:

$$A + B \rightarrow P$$

where *A* is hCRBP, *B* is dye, *P* is the SB, then we already have the 2nd-order rate equation (**1**):

$$\ln \frac{[B]_t[A]_0}{[A]_t[B]_0} = k([B]_0 - [A]_0)t \quad (\mathbf{1})$$

where k is the 2nd-order rate constant, $[A]_0$ and $[B]_0$ are the initial concentrations, $[A]_t$ and $[B]_t$ are the concentrations at time t.

If the reaction can go to completion, then we define $t_{1/2}$ as the time when:

$$t = t_{1/2}, \text{ there is } [P]_{t_{1/2}} = [P]_{max}, \text{ here } [P]_{max} = [B]_0$$

$$\text{or equally, } [B]_{t_{1/2}} = \frac{1}{2}[B]_0$$

Because we always keep:

$$[A]_0 = 20\ \mu M \text{ , and } [B]_0 = 10\ \mu M$$

$$\therefore [A]_0 = 2[B]_0$$

and at time t, we have:

$$[A]_t = [A]_0 - [P]_t, \text{ and } [B]_t = [B]_0 - [P]_t$$

So, we can rewrite equation **(1)**:

$$\ln \frac{2[B]_t}{[A]_t} = -[B]_0 kt$$

$$\therefore \ln \frac{[B]_0 - [P]_t}{[A]_0 - [P]_t} = -[B]_0 kt - \ln 2$$

$$\therefore \ln\frac{2[B]_0 - [P]_t}{[B]_0 - [P]_t} = [B]_0 kt + \ln 2$$

$$\therefore \ln\left(1 + \frac{[B]_0}{[B]_0 - [P]_t}\right) = [B]_0 kt + \ln 2$$

$$\therefore 1 + \frac{[B]_0}{[B]_0 - [P]_t} = e^{[B]_0 kt + \ln 2}$$

$$\therefore 1 - \frac{[P]_t}{[B]_0} = \frac{1}{e^{[B]_0 kt + \ln 2} - 1}$$

$$\therefore [P]_t = [B]_0\left(1 - \frac{1}{e^{[B]_0 kt + \ln 2} - 1}\right) \textbf{(2)}$$

Or, we can reduce $[A]_t$ from equation **(1)** and get the same above equation (**2**), or (**3**),

$$[B]_t = \frac{[B]_0}{e^{[B]_0 kt + \ln 2} - 1} \textbf{(3)}$$

Now since $[P]_{max} = [B]_0$, at $t_{1/2}$ there is $[B]_{t_{1/2}} = \frac{1}{2}[B]_0 = [P]_{t_{1/2}}$
Use either side, we can get:

$$t_{1/2} = \frac{\ln\frac{3}{2}}{[B]_0 k}$$

As shown in Figure S14 the concentration of the complex (**TD-1V** or **FR-1V**) with **M3** versus time was potted. Non-linear fit of the data with equation (**2**) using Prism software (GraphPad, Inc) provides *k*.

Quantum yield measurements

Absolute fluorescence quantum yields (Φ) were measured on a Quantaurus-QY (model C11347-11, Hamamatsu Photonics) equipped with a xenon light source, monochromator, integration sphere and a multichannel back-thinned CCD detector at room temperature.

Cloning for mammalian expression vectors

General cloning protocol

The DNA fragment was amplified using Phusion High-Fidelity DNA Polymerase (NEB, M0530) with the appropriate primers (see below for details). PCR conditions are specified in **Table S5** using a Bio-Rad iCycler thermal cycler. Four reactions were set up for each cloning, to ensure that enough amplified product was obtained.

Table S5. PCR cycling conditions for cloning.

	PCR Program	Time (min)
1x	98 °C	0:30
40x	98 °C	0:10
	55 °C	0:30
	72 °C	Extension time
1x	72 °C	10:00
1x	4 °C	10:00

Reactant	Volume
DNA template	100 ng (x µL)
Forward primer	20 pmol (y µL)
Reverse primer	20 pmol (z µL)
10 mM dNTP	1 µL
50 mM $MgCl_2$	1 µL
DMSO	5 µL
5 x Phusion HF Reaction Buffer	10 µL
Phusion HF DNA Polymerase	0.5 µL
Nuclease-free water	50 µL – x – y – z – 17.5 µL

The PCR amplified gene was purified by Wizard® SV Gel and PCR Clean-Up System (Promega) from 1% agarose gel in amount of 20-50 ng/μL. The product was digested with the proper enzymes and ligated to a similarly prepared plasmid (50 ng/μL). Ligation between the insert fragment and plasmid was performed with 30 ng of plasmid and 90 ng of insert using T4 DNA Ligase (New England BioLabs). The ligated product was transformed into *E. coli* XL1-blue competent cells (Agilent) and grown on LB-agar plates supplemented with antibiotics (100 µg/mL ampicillin, 7.5 µg/mL tetracycline) at 37 ºC for 18 hours. Colonies were inoculated in LB medium (15 mL) with proper antibiotics (100 µg/mL ampicillin, 7.5 µg/mL tetracycline) and incubated at 37 ºC while shaking, for 10 hours. DNA purification was performed using Promega Wizard® Plus SV Miniprep DNA purification kit (A1330) following the manufacturer's protocol. The DNA sequence was verified with the corresponding sequencing primers by the MSU gene sequencing facility. Sequencing primers used are shown below.

CMV end_Seq: 5'-GGTCTATATAAGCAGAGCTGGTTTAG-3'

midGFP: 5'-CGTGCTGCTGCCCGACAACC-3'

Preparation of Plasmids

HindIII-EGFP-NotI-Q108K:K40E:T53A:R58L:Q38F:Q4F-EcoRI-3NLS-Stop, HindIII-EGFP-NotI-Q108K:K40E:T53A:R58L:Q38F:Q4F-EcoRI-NES (ELAEKLAGLDIN)-Stop and HindIII-EGFP-NotI-Q108K:K40E:T53A:R58L:Q38F:Q4F-EcoRI- CAAX (GKKKKKKSKTKCVIM)-Stop in pFlag-CMV2 was made by amplifying Q108K:K40E:T53A:R58L:Q38F:Q4F from pET-17b. The forward and reverse primers used were 5'-CGGCGGCCGCATGACGAGGGACTTCAAT-3' and 5'-

CGGAATTCGGACTTCTTTTTGAACACTTG-3' (see **Table S5** for protocol, extension time = 15 sec). The purified product was then inserted between NotI and EcoRI of the pFlag-CMV2 plasmid. The product was sequenced with the primers CMVend_Seq and midGFP.

Plasmid Sequences

HindIII-EGFP-NotI-Q108K:K40E:T53A:R58L:Q38F:Q4F-EcoRI-3NLS-Stop (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGG
ACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG
CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTG
ACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACT
TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGG
CAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG
AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACA
GCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCG
CCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACC
CCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGG
CATGGACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACTTCAAT
GGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATATTGATT
TTGCCACCCGCAAGATTGCAGTACGTCTCACTTTTACGGAAGTTATTGATCAAGATGGTGATAA
CTTCAAGACAAAAGCCACTAGCACATTCCTGAACTATGATGTGGATTTCACTGTTGGAGTAGAG
TTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGGCACTGGTCACCTGGGAAGGTG
ATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAACCGCGGCTGGAAGAAGTGGATTGAGGG
GGACAAGCTGTACCTGGAGCTGACCTGTGGTGACCAGGTGTGCCGTCAAGTGTTCAAAAAGAAG
TGCGAATTCGCTGACCCCAAGAAGAAGAGGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGGACC
CCAAGAAGAAGAGGAAGGTGTGA

HindIII-EGFP-NotI-Q108K:K40E:T53A:R58L:Q38F:Q4F-EcoRI-NES (ELAEKLAGLDIN)-Stop (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGG
ACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG
CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTG
ACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACT
TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGG
CAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG
AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACA
GCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCG

CCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACC
CCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGG
CATGGACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACTTCAAT
GGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATATTGATT
TTGCCACCCGCAAGATTGCAGTACGTCTCACTTTTACGGAAGTTATTGATCAAGATGGTGATAA
CTTCAAGACAAAAGCCACTAGCACATTCCTGAACTATGATGTGGATTTCACTGTTGGAGTAGAG
TTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGGCACTGGTCACCTGGGAAGGTG
ATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAACCGCGGCTGGAAGAAGTGGATTGAGGG
GGACAAGCTGTACCTGGAGCTGACCTGTGGTGACCAGGTGTGCCGTCAAGTGTTCAAAAAGAAG
TGCGAATTCGAGCTTGCCGAGAAACTTGCCGGGCTTGACATAAATTGA

HindIII-EGFP-NotI-Q108K:K40E:T53A:R58L:Q38F:Q4F-EcoRI-CAAX

(GKKKKKKSKTKCVIM)-Stop (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGG
ACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG
CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTG
ACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACT
TCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGG
CAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG
AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACA
GCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCG
CCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACC
CCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGG
CATGGACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACTTCAAT
GGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATATTGATT
TTGCCACCCGCAAGATTGCAGTACGTCTCACTTTTACGGAAGTTATTGATCAAGATGGTGATAA
CTTCAAGACAAAAGCCACTAGCACATTCCTGAACTATGATGTGGATTTCACTGTTGGAGTAGAG
TTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGGCACTGGTCACCTGGGAAGGTG
ATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAACCGCGGCTGGAAGAAGTGGATTGAGGG
GGACAAGCTGTACCTGGAGCTGACCTGTGGTGACCAGGTGTGCCGTCAAGTGTTCAAAAAGAAG
TGCCGAATTCGGTAAAAAGAAGAAAAAGAAGTCAAAGACAAAGTGTGTAATTATGTAG

Mammalian cell culture

HeLa cells were cultured at 37 °C with 5% CO_2 in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 1.0 mM sodium pyruvate (Life Technologies), 10% fetal bovine serum (FBS; BioWest), 1% penicillin, and 1% streptomycin (Life Technologies). Cells were seeded 1 d before transfection on an ibidi μ-Slide 8 well

coverslip (with ibiTreat). Transfection was performed at ~70% confluency with 1.0 μg of plasmid DNA using Lipofectamine® 3000 reagent (Life Technologies) according to the manufacturer's protocol. The medium was replaced after 8 h. Cells were washed with Dulbecco's phosphate-buffered saline (DPBS; Life Technologies) twice after another 36 h incubation and incubated in Phenol Red-free RPMI 1640 medium (Life Technologies). Prior to confocal imaging, **FR-1V** (in ethanol solution) was added to cells at a final concentration of 500 nM and incubated at 37 °C with 5% CO_2 for 1min, followed by three wash steps with DPBS.

General confocal imaging methods

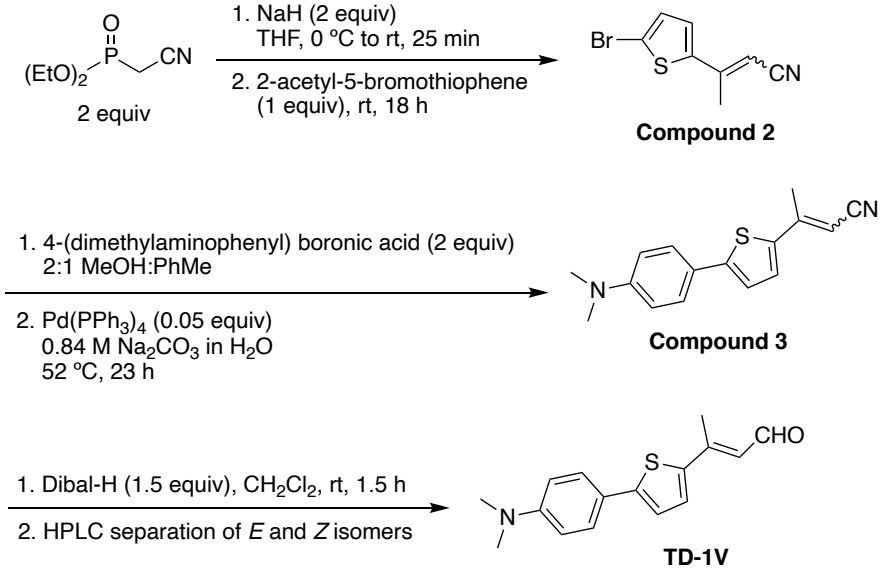
Cell microscopy was performed using an inverted laser scanning confocal microscope (LSM510Meta, Carl Zeiss, Jena, Germany) equipped with diode, argon and HeNe lasers. A 40x oil-immersed objective was used.

EGFP was imaged using 488 nm excitation, 488 nm primary dichroic, 545 nm secondary dichroic and BP 505-530 nm emission. hCRBPII/**FR-1V** was imaged using 405 nm excitation, 405 nm primary dichroic, 545 nm secondary dichroic and LP 615 nm emission filter. Kalman averaging 8 was applied in all confocal images. All images are pseudo-colored.

Synthesis of TD-1V

The synthesis of TD-1V is shown in **Scheme S1**. The synthesis of ethyl phosphonate **Compound 1** was described previously.[4] 2-Acetyl-5-bromothiophene,

(CAS# 5370-25-2) and 4-(dimethylaminophenyl) boronic acid (CAS# 28611-39-4) were purchased from Oakwood Chemical.



Scheme S1. Detailed synthesis of **TD-1V**.

*Synthesis of **Compound 2***: NaH (60% by weight, 0.38 g suspension containing 0.23 g NaH, 9.8 mmol) was placed in a dry round bottom flask and flushed with nitrogen. Anhydrous THF (40 mL) was added and then the solution was cooled to 0 °C via an ice-water bath and stirred for 10 minutes. Subsequently, diethyl(cyanomethyl)phosphonate (1.7 g, 9.8 mmol) was added dropwise and stirred for 25 minutes, gradually warming to room temperature. 2-Acetyl-5-bromothiophene (1.0 g, 4.9 mmol) in 10 mL anhydrous THF was added dropwise. The reaction was kept under nitrogen and stirred for 18 hours.

Work up: Brine solution was added to the reaction mixture and the resulting aqueous layer was extracted three times with diethyl ether (3 x 30 mL). The combined organic phase was dried over anhydrous Na_2SO_4. The solvent was then removed under reduced pressure to yield a light brown solid. The crude product was purified by silica gel

column chromatography with 20% ethyl acetate/hexane mixture to yield **Compound 2** and the starting 2-acetyl-5-bromothiophene.

*Synthesis of **Compound 3***: Toluene (7 mL) and methanol (3.6 mL) were added to a dry round bottom flask. The mixture was dearated by sonication under a nitrogen balloon. 4-(dimethylaminophenyl) boronic acid (0.33 g, 2.0 mmol) and **Compound 2** (0.23 g, 1.0 mmol) were then added and the solution was dearated again. Subsequently, sodium carbonate (0.32 g in 3.6 mL DI water) was added. After dearating the solution, tetrakis(triphenylphosphine) palladium (58 mg, 0.05 mmol) was added. The mixture was then heated at 52 °C under nitrogen for 23 hours.

Work up: The solution was extracted three times with dichloromethane (3 x 30 mL) and the combined organic phase was dried over anhydrous Na_2SO_4. The solvent was then removed under reduced pressure. The crude product was purified by silica gel column chromatography with a gradient from 100% hexane to 30% ethyl acetate/hexane. The resulting product was impure and was carried to the next step.

*Synthesis of **TD-1V:*** **Compound 3** (0.2 g, 0.75 mmol) was added to a dry round bottom flask, flushed with nitrogen and dissolved in anhydrous dichloromethane (30 mL). Then a 1 M solution of DIBAL-H in THF (1.5 mL, 1.5 mmol) was added dropwise. The reaction was stirred at room temperature, under nitrogen, for 1.5 hours.

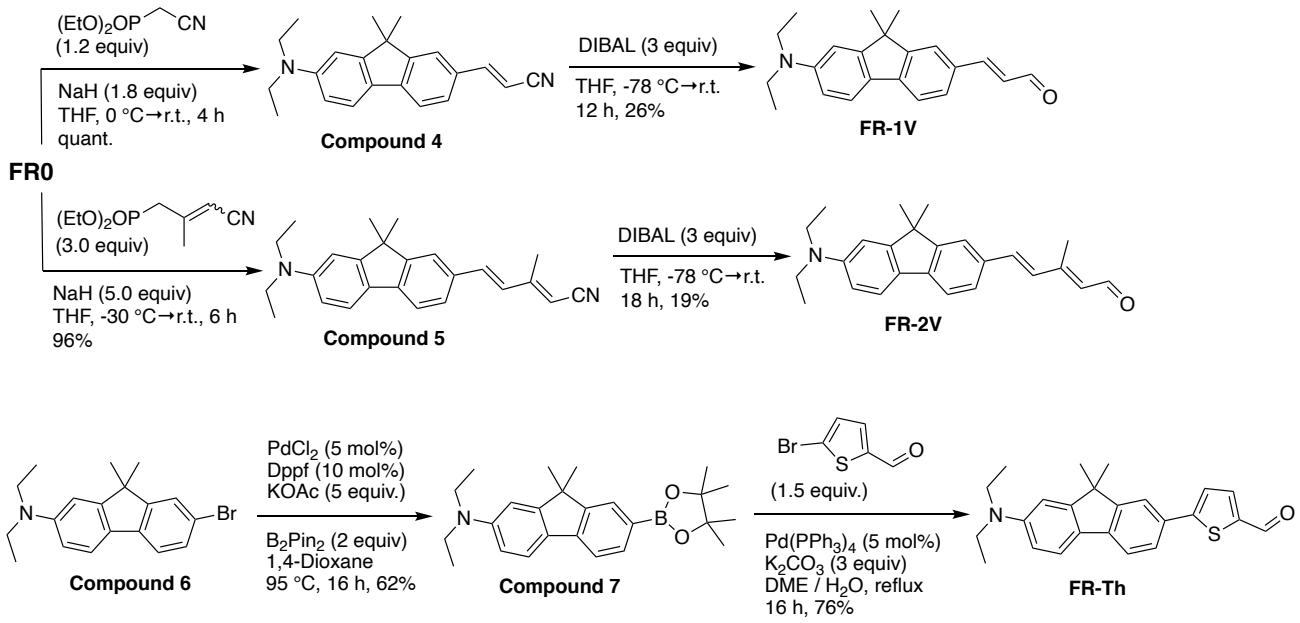
Work up: The reaction was quenched by slowly adding five drops of water, followed by five drops of 10% HCl, stirring for 15 minutes. An additional 20 mL of brine solution was added and the aqueous layer extracted with dichloromethane (3 x 40 mL). The resulting organic layer was dried over anhydrous Na_2SO_4 and the solvent removed

under reduced pressure. The crude product was initially purified by silica gel column chromatography using 10% ethyl acetate/hexane to 20% ethyl acetate/hexane. The impure product was combined, and a second column was run starting with 50% dichloromethane/hexane and then switching to 30% ethyl acetate/hexane. The crude product after the second column was dissolved in 70% ethyl acetate/hexane and the solid was filtered through cotton. HPLC was used to purify the soluble portion. The sample was separated by normal-phase HPLC (silica column, Zorbax Rx-SIL, 9.4 mm x 25 cm) after manual injection. The sample was eluted with 20% ethyl acetate/hexane at 3 mL min^{-1}. The product was detected at 400 nm. The overall yield of the *trans* isomer of **TD-1V** was 4% (11 mg).

^{1}H NMR (500 MHz, $CDCl_3$): δ 10.10 (d, J = 8.0 Hz, 1H), 7.51 (d, J = 8.5 Hz, 2H), 7.38 (d, J = 4.0 Hz, 1H), 7.15 (d, J = 3.5 Hz, 1H), 6.71 (d, J = 9.0 Hz, 2H), 6.43 (d, J = 7.5 Hz, 1H), 3.01 (s, 6H), 2.56 (s, 3H)

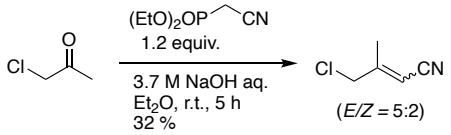
^{13}C NMR (125 MHz, $CDCl_3$): δ 190.51, 150.61, 150.13, 149.62, 140.78, 129.62, 127.02, 123.18, 121.79, 121.46, 112.23, 40.30, 15.63

Synthesis of FR-analogs



Scheme S2. Detailed synthesis of **FR**-analogs.

(2E/Z)-4-Chloro-3-methylbutyl-2-enenitrile

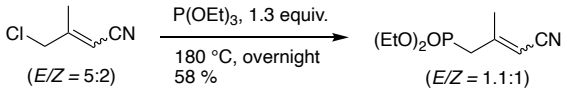


Chloroacetone (5.0 g, 54 mmol, 1.0 equiv) was added to a stirred solution of (diethylphosphono) acetonitrile (11.1 g, 62.7 mmol, 1.16 equiv) in Et_2O (15 mL in a 100 mL three-necked round bottom flask fitted with a reflux condenser) at room temperature. Solution of NaOH (3.7 M, 17 mL, 1.2 equiv.) was added dropwise over 1 h at room temperature. The reaction mixture came to an ethereal reflux and turned orange. After an additional 4 h of stirring the two-phase mixture was separated. The aqueous phase was extracted with Et_2O (3 × 10 mL) and the combined organic layer was dried over Na_2SO_4, filtered and concentrated *in vacuo*. Vacuum distillation of the residual orange oil afforded

a mixture of *cis*- and *trans*-4-chloro-3-methylbutyl-2-enenitrile (2.81 g, 32% yield, 50 ~ 60 °C at 1.5 mbar, *E/Z* = 5:2) as a colorless oil.

E/Z mixture: ^{1}H-NMR (500 MHz, $CDCl_3$): δ (ppm) = 5.52 (s, 1H), 5.29 (s, 1H), 4.27 (s, 2H), 4.07 (s, 2H), 2.15 (s, 3H), 2.06 (s, 3H). ^{13}C-NMR (125 MHz, $CDCl_3$): δ (ppm) =158.30, 158.01, 116.15, 115.36, 99.24, 99.00, 47.04, 44.41, 21.18, 19.27.

(2*E/Z*)-Diethyl (3-cyano-2-methylallyl)phosphonate



A neat mixture of triethyl phosphite (5.23 g, 31.5 mmol, 1.30 equiv.) and 4-chloro-3-methylbutyl-2-enenitrile (2.80 g, 24.2 mmol, 1.00 equiv.) in a 25 mL flask equipped with a stir bar was placed in a preheated bath at 180 °C. After stirring overnight, the mixture was allowed to cool to room temperature. Vacuum fractional distillation was employed and the mixture of geometric isomers of desired phosphonate was collected at 125 ~ 130 °C and 1.5 mbar to afforded 3.0 g (yield = 58%, *E/Z* = 1.1:1) as a colorless oil.

E/Z mixture: ^{1}H-NMR (500 MHz, $CDCl_3$): δ (ppm) = 5.25 – 5.28 (m, 2H), 4.09 – 4.19 (m, 8H), 2.96 (d, J = 23.7 Hz, 2H), 2.71 (d, J = 23.8 Hz, 2H), 2.19 (dd, J_1 = 3.3 Hz, J_2 = 0.9 Hz, 3H), 2.09 (dd, J_1 = 3.7 Hz, J_2 =1.6 Hz, 3H), 1.34 (t, J = 7.1 Hz, 6H), 1.32 (t, J = 7.1 Hz, 6H). ^{13}C-NMR (125 MHz, $CDCl_3$): δ (ppm) = 155.98 (d, J =11.2 Hz), 155.61 (d, J = 11.2 Hz), 116.50 (d, J = 5.4 Hz), 99.49 (d, J = 12.4 Hz), 99.06 (d, J = 12.4 Hz), 62.68 (d, J = 7.1 Hz), 62.62 (d, J = 7.1 Hz), 36.62 (d, J = 135.3 Hz), 34.87 (d, J = 135.3 Hz), 24.57

(d, J = 1.9 Hz), 22.62 (d, J = 3.1 Hz), 16.53 (d, J = 5.9 Hz), 16.51 (d, J = 5.7 Hz). ESI-MS calcd $[M+H]^+$: 218.09, found: 218.1.

(*E*)-3-(7-(diethylamino)-9,9-dimethyl-9*H*-fluoren-2-yl)acrylonitrile (Compound 4)

To a stirred suspension of NaH (13 mg, 0.31 mmol) in tetrahydrofuran (1.0 mL) was added a solution of (diethylphosphono)acetonitrile (36 mg, 0.21 mmol) in tetrahydrofuran (1 mL). The mixture was kept at 0 °C and stirred for 30 min. Then a solution of **FR0** (50 mg, 0.17 mmol) in dry tetrahydrofuran (3 mL) was added. The reaction mixture was stirred at ambient temperature for 4 h. The reaction mixture was then poured into cold water and extracted with ethyl acetate. The organic phase was dried and concentrated. Purification by flash chromatography (6% ethyl acetate in hexane) afforded 56 mg of a yellow solid (yield: quantitative).

^{1}H-NMR (500 MHz, $CDCl_3$), δ (ppm) = 7.55 (dd, J_1 = 15.4 Hz, J_2 = 8.1 Hz, 2H), 7.48 – 7.40 (m, 2H), 7.35 (dd, J_1 = 7.9 Hz, J_2 = 1.4 Hz, 1H), 6.73 – 6.64 (m, 2H), 5.88 – 5.79 (m, 1H), 3.45 (q, J = 7.1 Hz, 4H), 1.47 (s, 6H), 1.23 (t, J = 7.0 Hz, 6H). ^{13}C-NMR (125 MHz, $CDCl_3$): δ (ppm) = 156.45, 153.37, 151.25, 148.59, 143.76, 130.07, 127.46, 125.61, 121.81, 120.76, 119.10, 118.53, 110.81, 105.21, 93.00, 46.64, 44.71, 27.37, 12.61. ESI-HRMS: (calc.) (m/z) calcd for $C_{22}H_{24}N_2$ $[M+H]^+$ 317.2018, found 317.2019.

(*E*)-3-(7-(diethylamino)-9,9-dimethyl-9*H*-fluoren-2-yl)acrylaldehyde (FR-1V)

To a solution of compound **4** (50 mg, 0.16 mmol) in tetrahydrofuran (dry, 3 mL) was added DIBAL (1 M in THF, 0.50 mL) at -78 °C under argon. The reaction mixture was then

allowed to warm to ambient temperature gradually and was stirred for 12 h. Cold methanol was added drop wise to quench reaction. The mixture was treated with saturate solution of Rochelle's salt and extracted with dichloromethane. The combined organic phase was dried and concentrated. Purification by flash chromatography (5% ethyl acetate in hexane) afforded 13 mg of an orange solid (26% yield).

^{1}H-NMR ($CDCl_3$, 500 MHz), δ (ppm) = 9.70 (d, J = 7.8 Hz, 1H), 7.61 – 7.47 (m, 5H), 6.78 – 6.66 (m, 3H), 3.46 (q, J = 7.1 Hz, 4H), 1.49 (s, 6H), 1.24 (t, J = 7.1 Hz, 6H). ^{13}C NMR ($CDCl_3$, 125 MHz), δ (ppm) = 193.83, 156.61, 154.07, 153.38, 148.64, 144.00, 130.54, 128.81, 126.43, 125.72, 122.01, 121.89, 118.59, 110.82, 105.22, 46.63, 44.72, 27.39, 12.63. ESI-HRMS (m/z) calcd for $C_{22}H_{25}NO$ $[M+H]^+$ 320.2014, found 320.2015.

(2*E*,4*E*)-5-(7-(diethylamino)-9,9-dimethyl-9*H*-fluoren-2-yl)-3-methylpenta-2,4-dienenitrile (Compound 5)

To a stirred suspension of NaH (102 mg, 2.56 mmol) in dry tetrahydrofuran (2.0 mL) was added a solution of (2*E/Z*)-diethyl (3-cyano-2-methylallyl)phosphonate (333 mg, 1.53 mmol) in dry tetrahydrofuran (1.0 mL). The mixture was kept at -30 °C and stirred for 1 h. Then a solution of **FR0** (0.15 g, 0.51 mmol) in dry tetrahydrofuran (5.0 mL) was added. The reaction mixture was stirred at ambient temperature for 6 h. The reaction mixture was then poured into cold water and extracted with ethyl acetate. The organic phase was dried and concentrated. Purification by flash chromatography (5% ethyl acetate in hexane) afforded 174 mg of an orange solid (96% yield).

^{1}H-NMR ($CDCl_3$, 500 MHz), δ (ppm) = 7.53 (dd, J = 12.9, 8.2 Hz, 2H), 7.46 (d, J = 1.6 Hz, 1H), 7.37 (dd, J = 8.0, 1.6 Hz, 1H), 6.97 (d, J = 15.8 Hz, 1H), 6.88 – 6.81 (m, 1H), 6.72 – 6.63 (m, 2H), 5.33 – 5.29 (m, 1H), 3.44 (q, J = 7.0 Hz, 4H), 2.29 (d, J = 0.8 Hz, 3H), 1.47 (s, 6H), 1.22 (t, J = 7.1 Hz, 6H). ^{13}C NMR ($CDCl_3$, 125 MHz), δ (ppm) = 157.17, 156.15, 153.33, 148.23, 141.79, 136.90, 132.14, 127.17, 126.21, 126.13, 121.41, 120.75, 118.56, 118.28, 110.78, 105.43, 96.55, 46.61, 44.71, 27.47, 16.72, 12.63. ESI-HRMS (m/z) calcd for $C_{25}H_{29}N_2$ $[M+H]^+$ 357.2331, found 357.2349.

(2*E*,4*E*)-5-(7-(diethylamino)-9,9-dimethyl-9*H*-fluoren-2-yl)-3-methylpenta-2,4-dienal (FR-2V)

To a solution of compound **5** (0.15 g, 0.42 mmol) in dry THF (8.0 mL) was added DIBAL (1 M in THF, 1.3 mL) at -78 °C under argon. The reaction mixture was then allowed to warm to ambient temperature gradually and was stirred for 18 h. Cold methanol was added drop wise to quench the reaction. The mixture was treated with saturate solution of Rochelle's salt and extracted with dichloromethane. The combined organic phase was dried and concentrated. Purification by gravity silica chromatography and Prep-TLC (2% ethyl acetate in hexane) afforded 28 mg of a red solid (19% yield).

^{1}H-NMR ($CDCl_3$, 500 MHz), δ (ppm) = 10.17 (d, J = 8.2 Hz, 1H), 7.60 – 7.50 (m, 3H), 7.42 (dd, J_1 = 7.9 Hz, J_2 = 1.6 Hz, 1H), 7.16 (d, J = 16.0 Hz, 1H), 6.93 (dd, J_1 = 16.0 Hz, J_2 = 0.8 Hz, 1H), 6.74 – 6.65 (m, 1H), 6.15 – 6.08 (m, 1H), 3.45 (q, J = 7.1 Hz, 4H), 2.42 (d, J = 1.1 Hz, 3H), 1.50 (s, 6H), 1.23 (t, J = 7.1 Hz, 6H). ^{13}C NMR ($CDCl_3$, 125 MHz), δ (ppm) = 191.20, 156.22, 155.00, 153.36, 148.24, 141.84, 136.82, 132.51, 129.21, 129.12,

127.36, 126.25, 121.43, 120.82, 118.58, 110.78, 105.44, 46.62, 44.71, 27.48, 13.13, 12.64. ESI-HRMS (m/z) calcd for $C_{25}H_{30}NO$ $[M+H]^+$ 360.2327, found 360.2336.

***N,N*-diethyl-9,9-dimethyl-7-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-9*H*-fluoren-2-amine (Compound 7)**

To a solution of compound **6** (0.20 g, 0.58 mmol) in dry 1,4-dioxane (30 mL) was added $PdCl_2$ (6.0 mg, 0.029 mmol, 5 mol%), Dppf (33 mg, 0.058 mmol, 10 mol%), B_2Pin_2 (295 mg, 1.16 mmol, 2.0 equiv.) and potassium acetate (285 mg, 2.90 mmol, 5.0 equiv.). The resulting mixture was flushed with argon, then heated to 95 °C and kept stirring for 16 h. The reaction was quenched with water after full conversion of the starting material monitored with TLC (10% EtOAc in Hexane). The mixture was extracted with EtOAc, and run through a silica pad to remove the palladium and Dppf. The combined organic phase was dried and concentrated *in vacuo*. The crude was purified by gradient chromatography with 100% Hexane to 5% EtOAc-Hexane. Crystallization in EtOAc afforded 140 mg of a pale yellow crystal (62% yield).

^{1}H-NMR ($CDCl_3$, 500 MHz), δ (ppm) = 7.80 (s, 1H), 7.76 (d, J = 7.5 Hz, 1H), 7.57 (t, J = 8.0 Hz, 2H), 6.71 (d, J = 2.3 Hz, 1H), 6.67 (dd, J_1 = 8.4 Hz, J_2 = 2.4 Hz, 1H), 3.44 (q, J = 7.1 Hz, 4H), 1.48 (s, 6H), 1.37 (s, 12H), 1.22 (t, J = 7.1 Hz, 6H). ^{13}C NMR ($CDCl_3$, 125 MHz), δ (ppm) = 156.28, 151.82, 148.17, 143.18, 133.96, 128.28, 126.86, 121.45, 117.71, 110.61, 105.50, 83.51, 46.65, 44.69, 27.43, 24.91, 23.42, 12.64. ESI-HRMS (m/z) calcd for $C_{25}H_{35}BNO_2$ $[M+H]^+$ 392.2761, found 392.2776.

5-(7-(diethylamino)-9,9-dimethyl-9*H*-fluoren-2-yl)thiophene-2-carbaldehyde (FR-Th)

To a solution of compound **7** (50 mg, 0.13 mmol) and 5-bromo-2-thiophene-carboxaldehyde (37 mg, 0.19 mmol, 1.5 equiv.) in DME (3.0 mL) was added $Pd(PPh_3)_4$ (8.0 mg, 6.4 μmol, 5 mol%) and potassium carbonate (53 mg, 0.38 mmol, 3.0 equiv.). Water (0.4 mL) was added as a co-solvent. The mixture was heated to reflux for 15 h. The resulting crude was run through a silica pad to remove palladium and extracted with EtOAc. The organic phase was dried and concentrated. The crude was purified by flash chromatography with 5% EtOAc in Hexane and crystalized in EtOAc to afford 29 mg of a bright-red crystal (76% yield).

^{1}H-NMR ($CDCl_3$, 500 MHz), δ (ppm) = 9.88 (s, 1H), 7.75 (d, J = 4.0 Hz, 1H), 7.66 – 7.55 (m, 4H), 7.42 (d, J = 4.0 Hz, 1H), 6.74 – 6.66 (m, 2H), 3.45 (q, J = 7.1 Hz, 4H), 1.51 (s, 6H), 1.24 (q, J = 7.1, 6.1 Hz, 6H). ^{13}C NMR ($CDCl_3$, 125 MHz), δ (ppm) = 182.65, 156.10, 155.91, 153.58, 148.32, 141.88, 141.28, 137.68, 129.44, 125.95, 125.62, 123.09, 121.49, 120.27, 118.74, 110.79, 105.39, 46.79, 44.71, 27.46, 12.63. ESI-HRMS (m/z) calcd for $C_{24}H_{26}NOS$ $[M+H]^+$ 376.1735, found 376.1729.

Table S6. Spectral features of FR dyes in different organic solvents, showing the solvatochromism as their ICT characteristic.

	FR0			**FR-1V**		
	λ_{abs} (nm)	λ_{em} (nm)	Φ[a]	λ_{abs} (nm)	λ_{em} (nm)	Φ
Toluene	396	434	0.703	416	488	0.759
THF	392	470	0.802	414	537	0.796
Ethyl acetate	389	475	0.662	410	532	0.820
Dichloromethane	399	497	0.774	423	575	0.873
DMF	398	512	0.775	420	608	0.691
DMSO	402	526	0.763	425	614	0.725
Acetonitrile	393	523	0.745	415	605	0.638
Ethanol	396	556	0.664	420	633, 660	0.074
Methanol	396	573	0.445	420	669	0.024

	FR-2V			**FR-Th**		
	λ_{abs} (nm)	λ_{em} (nm)	Φ	λ_{abs} (nm)	λ_{em} (nm)	Φ
Toluene	428	559	0.675	417	503	0.879
THF	426	603	0.524	417	563	0.782
Ethyl acetate	422	600	0.555	412	561	0.770
Dichloromethane	435	634, 665	0.432	424	604	0.751
DMF	432	633, 669	0.167	424	663	0.473
DMSO	438	633, 672	0.175	429	672	0.278
Acetonitrile	425	630, 670	0.134	415	667	0.361
Ethanol	433	627, 667	0.021	419	604	0.027
Methanol	431	670	0.006	417	473	0.006

[a]Absolute fluorescence quantum yield measurement using Quantaurus-QY at room temperature.

Table S7. Spectral features of **FR0**-SB, **FR-1V**-SB and **TD-1V**-SB in EtOH.

Spectroscopic properties of **FR0-**SB, **FRV-**SB and **TD**-SB with *n*-butyl

Fluorophore	λ_{abs}	λ_{emi}	Stokes Shift	Φ[a]
FR0-SB	372	463/628	91/256	0.40
FR-1V-SB	395	507	112	0.05
TD-1V-SB	400	540	140	0.07

[a]Absolute quantum yield was measured on a Quantaurus-QY.

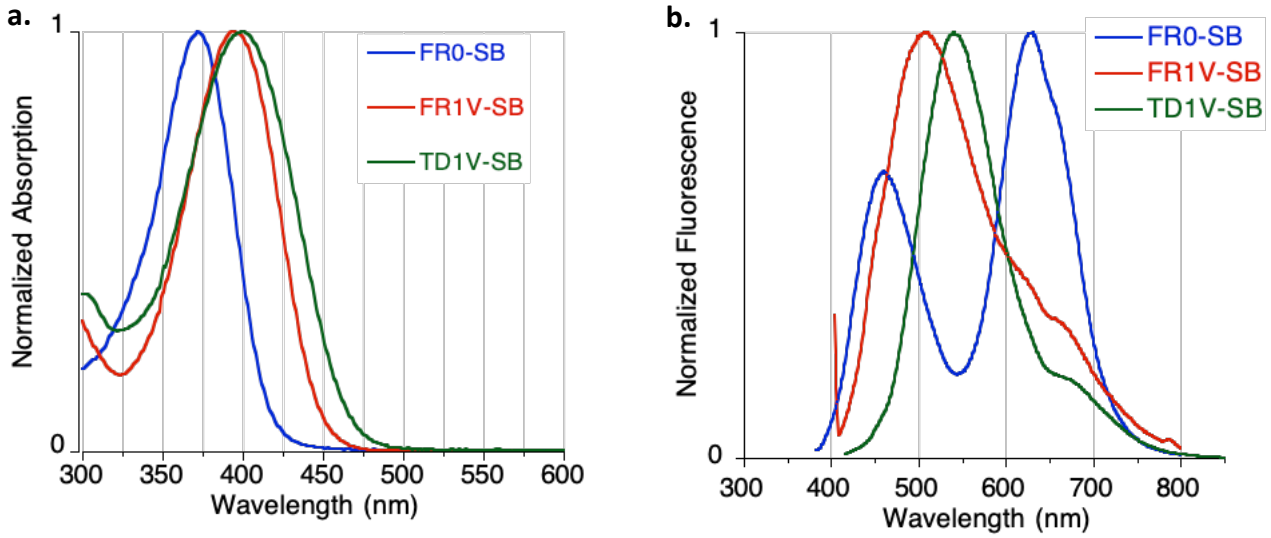


Figure S1. Comparison of **FR0**-SB**, FR1V**-SB and **TD1V**-SB absorption spectra (a) and emission spectra (b) in EtOH.

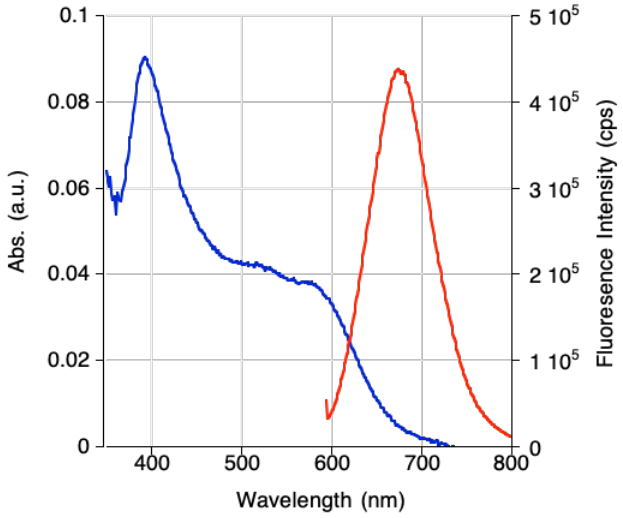


Figure S2. Absorption and emission spectra of Q108K:K40L-hCRPBPII/**TD-1V**.

STRUCTURE GUIDED DESIGN TO OBTAIN M1

Based on our previous work with retinal-bound hCRBPII, we hypothesized that the iminium was stabilized via a water mediated hydrogen bond to Gln4. Consistent with this was the crystal structure of **TD-1V**-bound Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:A33W variant, showing a *cis*-iminium stabilized via the water-mediated interaction with Gln4 (Figure S3), similar to that seen with retinal. Screening at this position resulted in the Q4F mutation, which served to both increase soluble protein expression and reduce the pK_a of the imine (pK_a = 8.3) though it did not suppress it sufficiently to completely eliminate the iminium. Additional mutations were therefore required to adjust the pK_a to favor the imine. Previous work with retinal as the ligand showed a reduction in pK_a upon Gln38 mutation. Though relatively far from the imine, Gln38 often makes water-mediated interactions in the binding cavity and it's change to a hydrophobic residue results in a more hydrophobic binding cavity more devoid of ordered solvent. However, addition of the Q38F mutation still did not give the desired imine. However, by reducing the hydrophilicity in the vicinity of the imine with the T53A mutation, and the creation of additional hydrophobicity near the mouth of the binding cavity with the R58L mutation, resulting in the Q108K:K40L:T53A:R58L:Q38F:Q4F-hCRBPII hexamutant (**M1**), finally a variant devoid of iminium formation at physiological pH was achieved. The later three (T53A, R58L, and Q38F) had the additional advantage of increased **TD-1V** binding affinity and substantially increased rate of pigment maturation. Furthermore, these mutations also ensured the expression of a stable, monomeric protein in high quantities.

It is enlightening to compare the hCRBPII **FR-IV** binding photoswitchable system previously published (Sheng *et.al. Angew. Chem.* 2018, 16083), which requires *trans-* to *cis-* imine isomerization to convert SB into longer lived PSB species, with the present system where conversion from SB to PSB involves an ESPT mechanism, where the PSB is presumably only present for nanoseconds. The photoswitchable systems all required a Gln at position 4, that stabilizes the *cis*-iminium conformation via a water mediated interaction, and also increases the pK_a of the *cis*-iminium (see Figure S3 as an example). In contrast, the ESPT systems all require mutation of Gln4 to a hydrophobic residue (typically Q4F), which destabilizes the *cis*-iminium conformation, disfavoring the isomerization pathway. Finally, as discussed above, the installation of Glu40 provides a proton donor adjacent to the *trans* imine, which is essential to the ESPT mechanism. We feel confident that the present system does not involve imine isomerization as all previous examples showed the thermal re-isomerization to occur on the minutes to hours timescale, far too slow to explain the ESPT phenomenon seen here.

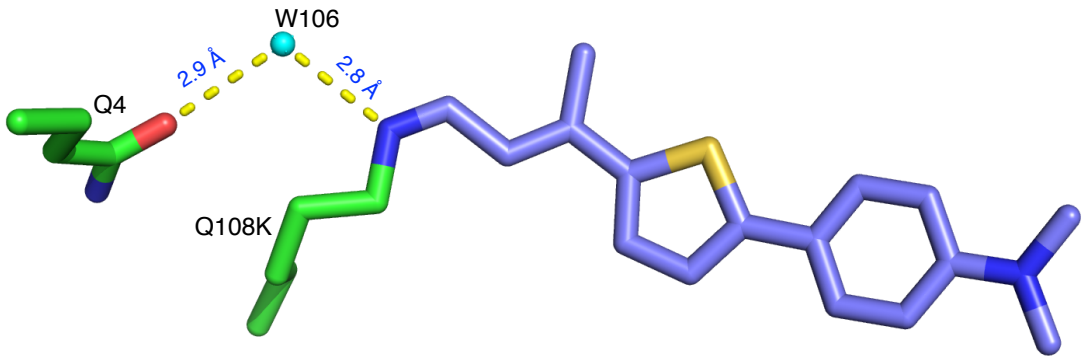


Figure S3. The crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:A33W (**M20**) co-crystalized with **TD-1V** depicts the *cis*-iminium further stabilized through a water mediated hydrogen bond between Q4 and the iminium.

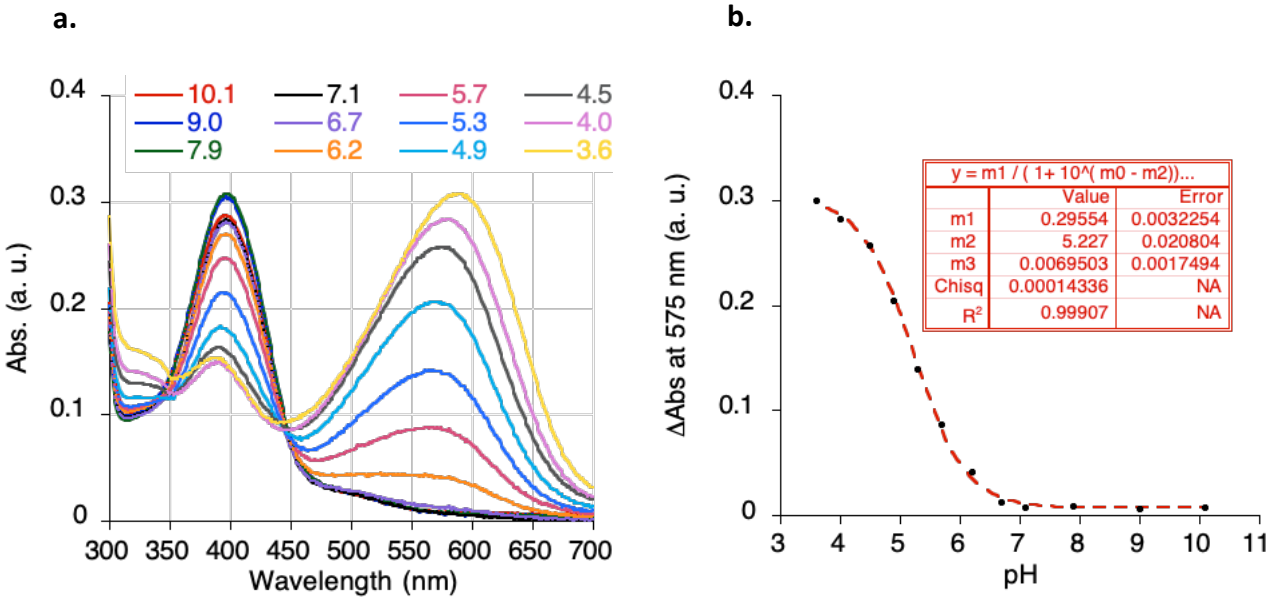


Figure S4. a. Acid/Base titration of **TD-1V/M3** complex in PBS buffer; b. The plot of pH versus absorbance derived from the data in panel a, fitted to the Henderson-Hasselbalch equation, revealing a pK_a of 5.2.

PROCEDURE FOR PEAK DECONVOLUTION:

The basic requirement for deconvolution of a peak is to find a mathematic function, which describes the shape of the peak, and to estimate the probable number of peaks in the cluster. The shape of asymmetric peaks is often described by the Exponential Modified Gauss function, and although, it is not possible to strictly determine the general number of peaks, deconvolution procedures requires the best estimate of initial parameters. In this study it was easy to recognize the number of peaks since the number of maxima were well known. The result of a deconvolution procedure is the number of notable outputs and their parameters (areas, retention times, width for peaks), which we have found in a cluster of the original peak.

The existence of two peaks in a peak envelope was assumed according to the number of species and peak maxima, SB and PSB. Thus, the two-peak concept was used in our study. The commercially available software, Origin (OriginLab) with section Peak

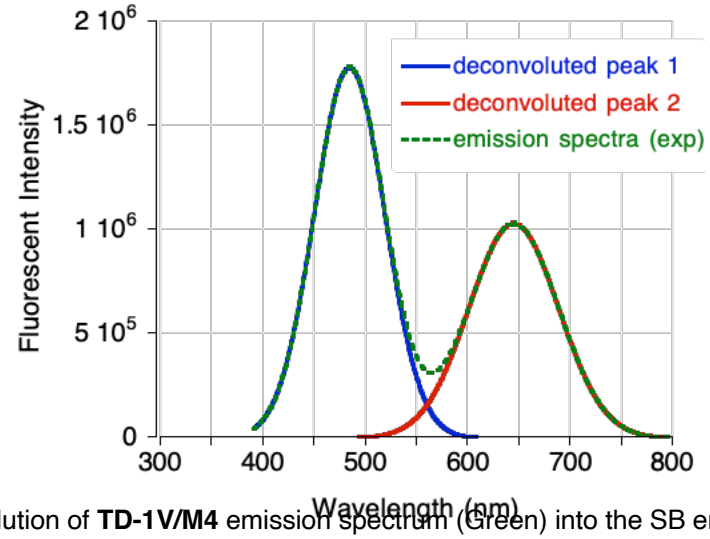


Figure S5. Deconvolution of **TD-1V/M4** emission spectrum (Green) into the SB emission (Blue) and PSB emission (Red) using Gauss Peak function.

FittingTM was employed for evaluation of experimentally obtained emission spectra. Individual emission spectra were exported to the Origin (OriginLab). The deconvolution procedure was applied, and the individual peaks areas were determined.

SUPPORTING INFORMATION FIGURES

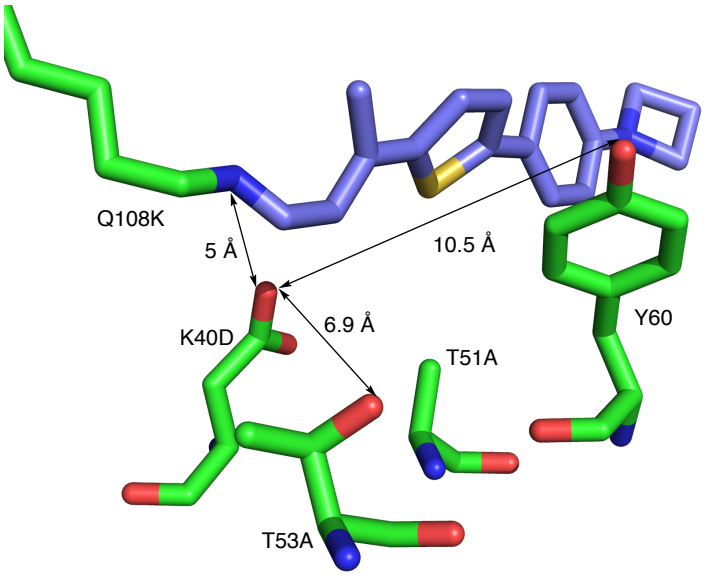


Figure S6 Crystal structure of the azetidine analog of **TD-1V** bound with a trimeric form of **M12** reveals a K40D conformation that is far away from the nitrogen of iminium. In contrast to **M11** bound with **TD-1V**, K40D does not have water mediated interactions with T51 and Y60. It should be noted that the **TD-1V**/**M12** structure did not yield a satisfactory electron density for mapping of the fluorophore within the binding pocket. Nonetheless, the azetidine analog provided excellent electron density to visualize important interactions.

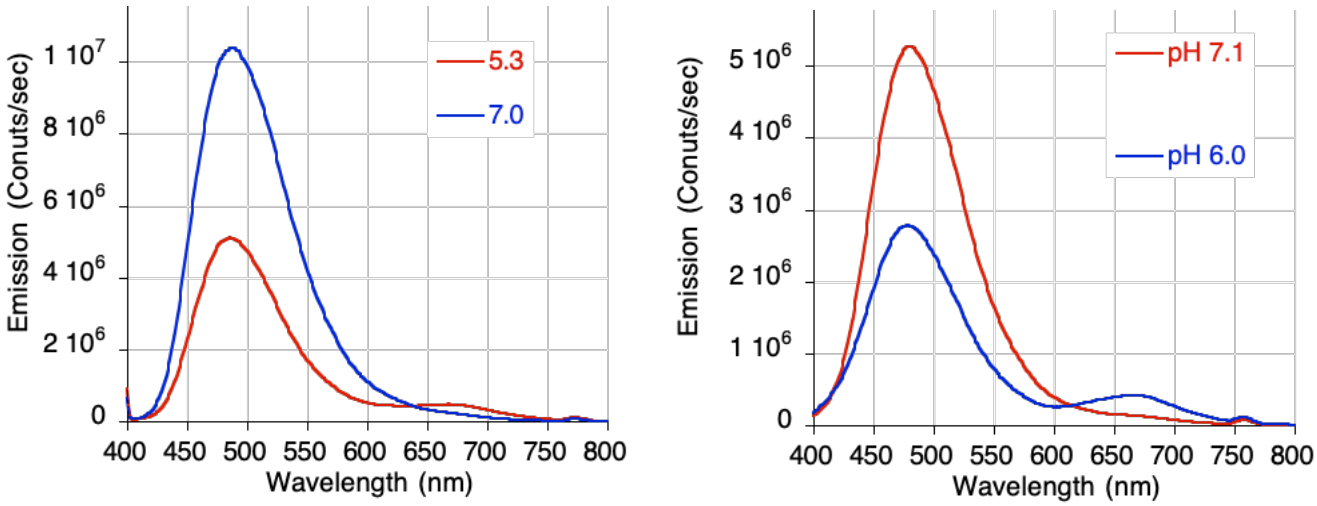


Figure S7. Emission of **TD-1V/M14** (Left) excited at 386 nm and **TD-1V/M13** (right) excited at 383 nm in neutral and acidic pH upon excitation of SB.

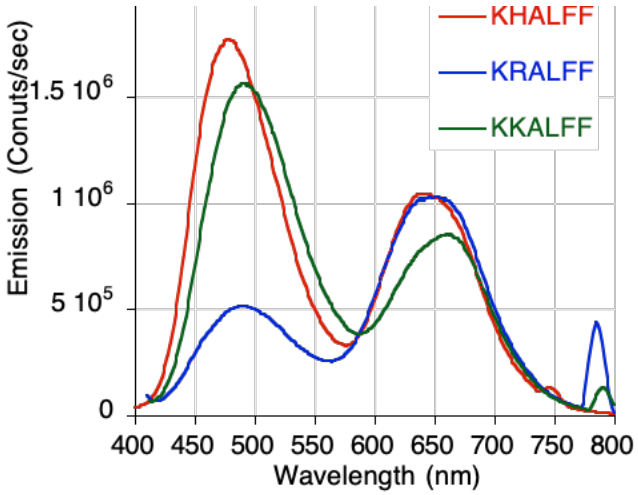


Figure S8. Emission of **TD-1V/M15** (KHALFF) excited at 393 nm, **TD-1V/M16** (KKALFF) excited at 389 nm, and **TD-1V/M17** (KRALFF) excited at 392 nm in neutral pH upon excitation of SB.

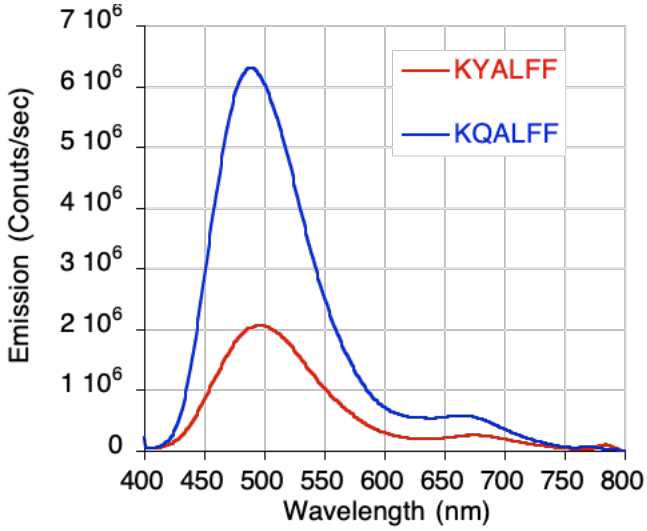


Figure S9. Emission of **TD-1V/M18** (KQALFF) excited at 388 nm, **TD-1V/M19** (KYALFF) excited at 393 nm in neutral pH upon excitation of SB.

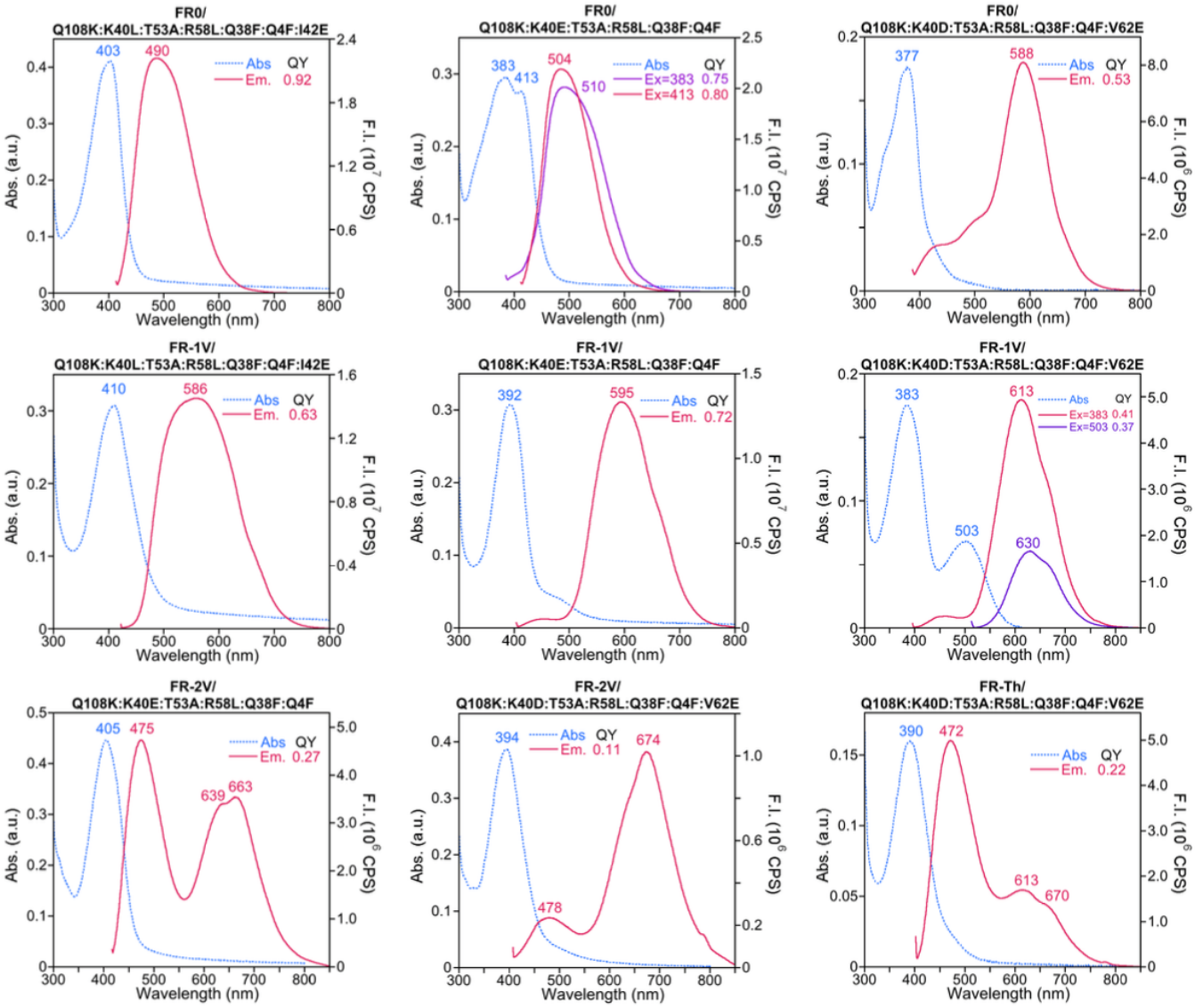


Figure S10. Emissions of **FR** series with different mutants, showing ESPT to various extents.

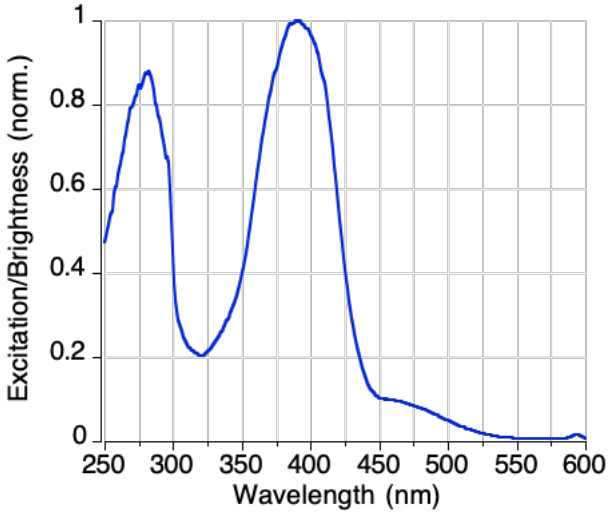


Figure S11. Excitation spectrum of **FR-1V/M3** complex (emission observed at 595 nm). The excitation source is corrected against Rhodamine-B.

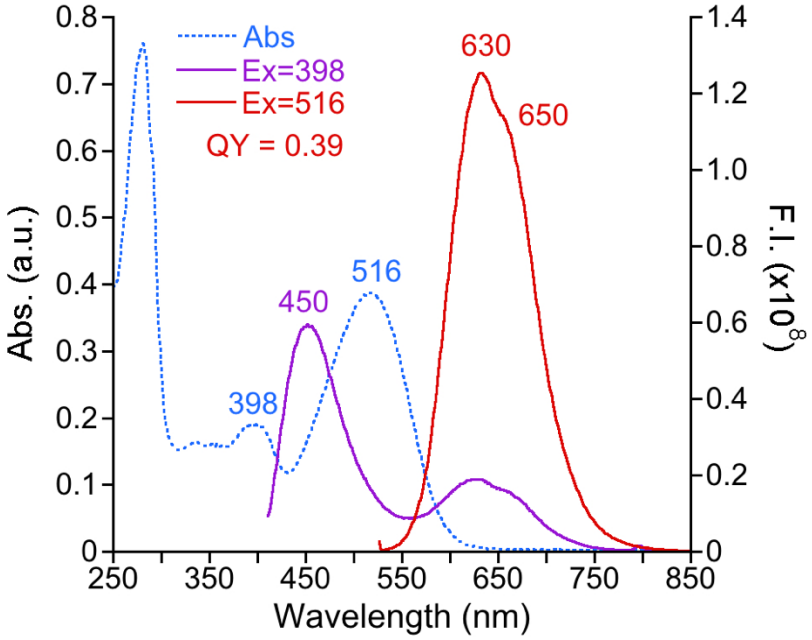


Figure S12. UV-Vis (dotted line, blue) and emission spectra (solid lines) of **FR-1V/M15.**

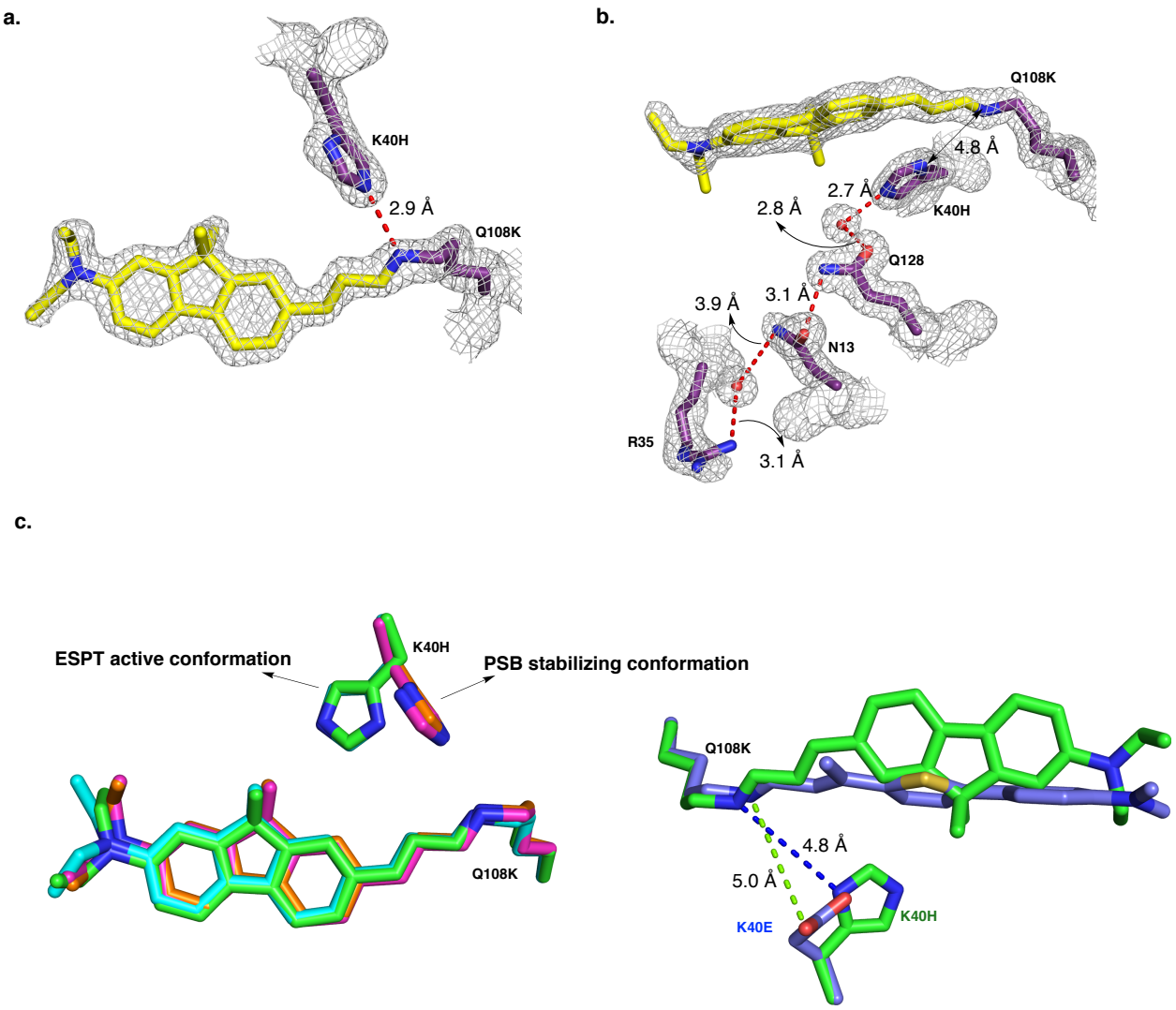


Figure S13. Crystal structure of **M15** bound with **FR-1V** shows four molecules in the asymmetric unit that are fully occupied by **FR-1V.** a. Chains A and B show the same conformation of K40H, in which it directly interacts with the nitrogen of iminium. b. Chains C and D show the same conformation of K40H, in which it does not interact with the nitrogen of iminium (4.8 Å distance). Instead, K40H gets locked into this conformation by hydrogen bond networks containing the residues highlighted in the figure. c. Left: Overlay of chain A and chain C, indicating that the only difference between the two chains is a different conformation of K40H. The K40H conformation that is closer to the nitrogen of imine would be expected to stabilize the iminium, and the other conformation would be an active ESPT ground-state conformation. Right: Overlay of **M15**/**FR-1V** (green) and **M11**/**TD-1V** (purple) crystal structures shows the similarity in the juxtaposition of proton donor and acceptor in the two ESPT-active systems.

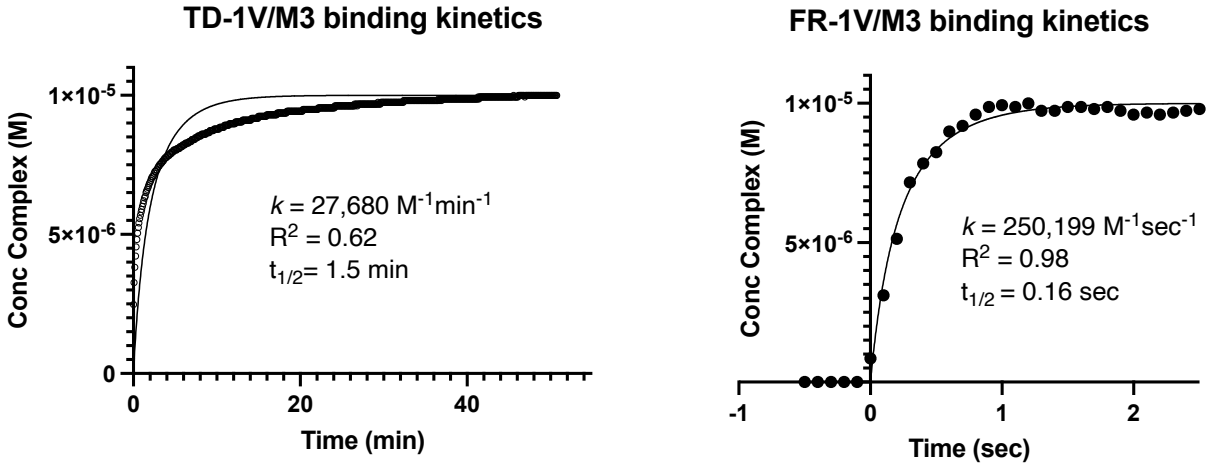


Figure S14. Rate of **TD-1V/M3 and FR-1V/M3** SB formation, fitted to 2nd order kinetics with 20 μM protein and 10 μM **TD-1V** or **FR-1V**.

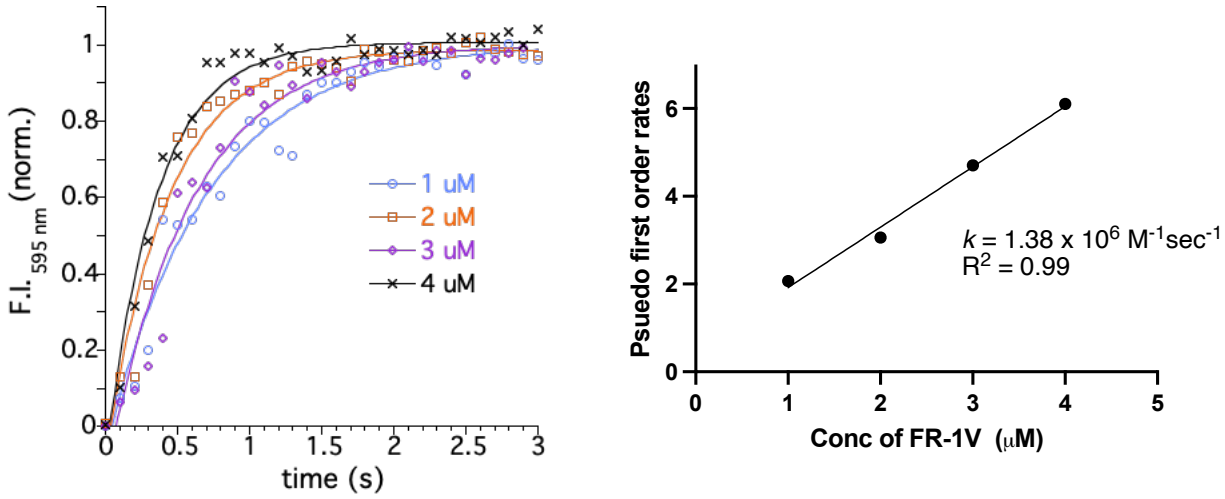


Figure S15. Binding kinetics of **FR-1V** with **M3.** a) Mixed order binding of **FR-1V** (10 μM) with **M3** (20 μM). The binding was monitored at the 595 nm ESPT emission maximum. An exponential fit gives the binding half-time as 0.18 s. b) Second-order rate constant measurement. Variable concentrations of **FR-1V** were used to bind 100 nM of **M3** under pseudo-first order conditions. Observable rate constants were derived from single exponential fits. The second-order rate constant was then derived from the linear regression of observable pseudo first order rate constants as a function of concentration. All measurements were taken in PBS buffer (pH = 7.3). Details for kinetic analysis can be found in *ChemBioChem* (**2020**), *21*, 723-729.

PROTEIN CRYSTALLIZATION AND DATA COLLECTION

The proteins were concentrated between 6 and 8 mg/mL in a buffer A (10 mM Tris, 150 mM NaCl, pH = 8). Prior to crystallization, the protein was incubated with four equivalent chromophores. Crystals were grown through the hanging drop vapor diffusion method by adding 1 µL of protein solution and 1 µL of crystallization solution in the drop and 1 mL of crystallization solution in the reservoir. All of the crystallization plates were wrapped in aluminum foil to protect the chromophore from light damage.

The crystallization solution contained 25-30% PEG 4000, 0.1 M sodium acetate, pH=4.0-4.8, 0.1 M ammonium acetate. The cryoprotectant solution (30% PEG4000, 0.1 M sodium acetate pH = 4.5, 0.1 M ammonium acetate with 15% glycerol) was employed for soaking and flash freezing crystals in loops in liquid nitrogen. The crystals were stored in a liquid nitrogen Dewar prior to data collection.

The diffraction data were collected at the Advanced Photon Source (APS) (Argonne National Laboratory IL) LS-CAT, (sector 21-ID-D, F, G) using Dectris 9M Eiger and Rayonix MX300 detectors, ~1.00Å wavelength radiation at 100K. The initial diffraction data were indexed, processed and scaled using the HKL2000 software package. The structures were solved by molecular replacement using PHASER in PHENIX and hCRBPII (PDB entry 2RCQ) as a search model. The initial electron density map was generated by Phaser-MR in PHENIX. Model rebuilding, placement of water molecules etc. were performed using COOT. The structures were refined through the PHENIX program packages. The chromophore was created using the Coot program and restraints was generated by PHENIX.

Table S8. X-Ray crystallographic data and refinement statistics.

	TD-1V/M11	**FR-1V/M15**	**TD-1V/M12 (trimer)**
Space group	*P*1	*P*1	$P22_12_1$
a(Å)	36.766	36.687	36.932
b(Å)	54.99	54.515	102.41
c (Å)	69.805	63.931	108.478
α(°)	109.338	89.996	90
β(°)	97.797	86.63	90
δ(°)	102.603	77.686	90
Molecules per Asymmetric Unit	4	4	3
Total reflection	970285	950433	1077503
Unique Reflection	54659 (5195)	55438 (5085)	14902 (1453)
Completeness (%)	96.19 (91.67)	85.90 (79.16)	99.89 (99.83)
Average I/σ	24.6(1.4)	22.50 (2.12)	18.88 (1.42)
R_{merge}	0.044 (0.51)	0.046 (0.53)	0.097 (1.251)
Resolution (Å) (Last Shell)	34.95 -1.671 (1.731 -1.671)	32.04 - 1.594 (1.651- 1.594)	34.74-2.5(2.589-2.5)
R_{work}/ R_{free} (%)	18.41/23.18	19.67/23.27	20.76/27.97
Bond Length (Å)	0.007	0.008	0.008
Bond Angle (°)	0.86	1.18	1.24
Average B factor	29.66	24.08	50.39
Number of water molecules	405	235	77
PDB ID	**7LSQ**	**7MFX**	**7MFZ**

[a] Values in the parenthesis refer to the last resolution shell.

Table S9. X-Ray crystallographic data and refinement statistics.

	TD-1V/M20
Space group	*C*21
a(Å)	28.918
b(Å)	66.617
c (Å)	63.262
α(°)	90
β(°)	90.981
δ(°)	90
Molecules per Asymmetric Unit	1
Total reflection	851159
Unique Reflection	32216 (3049)
Completeness (%)	99.10 (94.18)
Average I/σ	35.5 (2.1)
R_{merge}	0.059 (0.44)
Resolution (Å) (Last Shell)	33.31- 1.258 (1.303-1.258)
R_{work}/ R_{free} (%)	16.95/18.59
Bond Length (Å)	0.008
Bond Angle (°)	0.95
Average B factor	16.71
Number of water molecules	179
PDB ID	**7MFY**

[a] Values in the parenthesis refer to the last resolution shell.

REFERENCES

1. Wang, W., Nossoni, Z., Berbasova, T., Watson, C.T., Yapici, I., Lee, K.S., Vasileiou, C., Geiger, J.H. & Borhan, B. *Tuning the electronic absorption of protein-embedded all-trans-retinal*. Science **338**, 1340-1343 (2012).
2. Gill, S.C. & von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry* **182**, 319-326 (1989).
3. Yapici, I., Lee, K.S.S., Berbasova, T., Nosrati, M., Jia, X.F., Vasileiou, C., Wang, W.J., Santos, E.M., Geiger, J.H. & Borhan, B. "Turn-On" Protein Fluorescence: In Situ Formation of Cyanine Dyes. *Journal of the American Chemical Society* **137**, 1073-1080 (2015).
4. Lan, L., Zhan, X., Qin, W. & Liu, Z., Maoa, Z. Simple Synthesis of new 3-Substituted 4-(3-Chloro-4-fluorophenyl)-1*H*-pyrrole Derivatives and their Anticancer Activity *in vitro*. *Heterocycles* **89**, 375-397 (2014).