DEVELOPMENT OF FLUORESCENT PROTEIN TAGS FOR LIVE-CELL IMAGING

By

Elizabeth Marie Santos

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry – Doctor of Philosophy

2017

ABSTRACT

DEVELOPMENT OF FLUORESCENT PROTEIN TAGS FOR LIVE-CELL IMAGING

By

Elizabeth Marie Santos

Our primary goal is to develop fluorescent proteins that span the entire visible spectra, to be used when conventional fluorescent proteins are inadequate. In our system, fluorescence is activated upon coupling of the protein and ligand, such that temporal control can be achieved, whereas intrinsically fluorescent proteins are constitutively on. Additionally, our system does not require oxygen and can therefore find potential uses in obligate anaerobes. Our lab has demonstrated the ability to effectively control the absorption profile of conjugated polyenes. The initial aim of this project was to regulate the emissive properties of bound fluorophores with the same degree of control. This was achieved by the coupling of the solvatochromic fluorophore **ThioFluor** to hCRBPII mutants. **ThioFluor** yielded mutants with absorption maxima varying from 501 nm to 705 nm and emission maxima from 613 nm to 744 nm. This is equivalent to regulation over 204 nm in absorption and 131 nm in emission, covering both the red and far-red fluorescence wavelength regimes. Furthermore, we have shown its utility in live-cell imaging in whole cells, and with targeting to the nucleus and extranuclear space; fortuitously, negligible background fluorescence is apparent.

In the course of optimizing binding of **ThioFluor** to hCRBPII, we discovered that it was observed that not only is the protonated Schiff base (PSB) fluorescent, but the Schiff base (SB) is as well. However, while PSB emission wavelength could be altered over 204 nm, SB emission remains nearly constant at 500 nm. Interestingly, select mutants displayed a far-red emission upon SB irradiation, similar to that obtained upon irradiation of the PSB, presumably through protonation in the excited state. This serendipitous discovery leads to more than a 200 nm Stokes shift with high quantum yield (> 60%). One such hCRBPII/**ThioFluor** complex displays ideal spectroscopic properties including fast iminium formation with a half-life of 1.7 min, low pK_a of 5.3 (rendering almost complete SB formation at physiological pH), high quantum yield (0.51) and large Stokes shift (208 nm). This fluorescent protein was successfully used to visualize whole cell fluorescence, as well as targeting to the nucleus.

The last major endeavor was to develop a protein-based pH sensor, with the ability to report pH values with high accuracy. We have previously reported an absorptive system that was capable of ratiometric sensing of pH. However, we have now developed a single protein fluorescent ratiometric pH sensor, based on the titration of an acidic residue near the iminium. Standard curves were generated based on the ratio of emissions at two excitation wavelengths, allowing for concentration independent sensing of pH. We have been able to demonstrate its applicability as an *in vivo* fluorescent pH probe, obtaining a pH value of 6.7 when hCRBPII is targeted to the nucleus of HeLa cells.

Dedicated to my family and friends for their encouragement and support.

ACKNOWLEDGMENTS

Firstly, I am thankful to my undergraduate advisor and mentor Dr. Montserrat Rabago Smith. Without her guidance I would not have pursued an education and career in chemistry.

I am very grateful to my PhD advisor, Professor Babak Borhan for instilling the knowledge needed to become a successful scientist, particularly in providing a new perspective of thinking. I also appreciate the freedom he allows in choosing and starting new projects, as well as the assistance provided in learning new techniques, and most importantly fixing instruments!

Upon joining the lab, senior graduate students were pivotal in teaching me experimental techniques, how to design projects and analyze results. For this, I would also like to thank Dr. Tetyana Berbasova, Dr. Ipek Yapici and Dr. Arvind Jaganathan. For the aforementioned, I would also like to thank Dr. Chrysoula Vasileiou. Anytime I have needed assistance, she has been more than helpful.

Many of these projects would not have been successful without collaborators, to which I would also like to express my gratitude. Special thanks to Alireza Ghanbarpour and Zahra Assar for the crystal structures, Hadi Gholami for the synthesis of chromophores and Wei Sheng for the measurement of quantum yields. Additionally, confocal imaging would not have been possible without the guidance of Dr. Melinda Frame. I want to extend my appreciation to my committee members Professors Jim Geiger, Xuefei Huang and Ned Jackson.

Lastly, I would like to thank my parents and siblings for their love and support. I have unfound gratitude to my late mom who always instilled upon me the importance of a good education.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xiv

LIST OF SCHEMES xxviii

KEY TO ABBREVIATIONS xxix

CHAPTER I: METHODS EMPLOYED TO FLUORESCENTLY LABEL PROTEINS OF INTEREST 1

I.1 Intrinsically fluorescent proteins 1

I.2 Extrinsically fluorescent proteins 8

I.2.1 Site-specific chemical labeling 11

I.2.2 Bioorthogonal reactions 20

I.2.3 Small protein tags 22

I.2.3.1 Fluorogen activating proteins 23

I.2.3.2 Photoactive yellow protein 25

I.2.3.3 *In situ* formation of a cyanine dye 30

REFERENCES 33

CHAPTER II: MODULATION OF ABSORPTION AND EMISSION VIA A PROTEIN EMBEDDED SOLVATOCHROMIC FLUOROPHORE 56

II.1 Application of merocyanine aldehyde with hCRBPII mutants for *in vivo* imaging 58

II.2 Principles of solvatochromicity 70

II.3 Choice of fluorophore 74

II.4 Spectroscopic properties of ThioFluor 77

II.5 Removal of polarity induces a red-shift in absorption and emission 81

II.6 Attempts to encapsulate the binding cavity 89

II.7 Removal of water and extensive packing of the chromophore leads to an enhanced red-shift in absorption and emission 97

II.7.1 Investigation of combinatorial residue effects on the wavelength of Q108K:K40L:T51V:T53S:R58W:Y19W 105

II.8 Localization of charge leads to a blue-shift in absorption and emission .. 107

II.8.1 Design of blue-shifted hCRBPII/ThioFluor complexes that promote iminium formation 111

II.8.2 Exploring the additive effects of acidic residues on the absorption and emission wavelength of hCRBPII/ThioFluor 117

II.9 Emission is linearly correlated to absorbance 121

II.10 Visualization of hCRBPII/ThioFluor in bacteria 124

II.11 Visualization of hCRBPII/ThioFluor in mammalian cells 126
II.11.1 Labeling of hCRBPII variants that have a propensity to oligomerize is unproductive 126
II.11.2 ThioFluor successfully labels hCRBPII in mammalian cells 131
II.12 Conclusions 133
REFERENCES 135

CHAPTER III: STRUCTURE-PROPERTY RELATIONSHIPS OF THIOFLUOR ANALOGUES BOUND TO hCRBPII MUTANTS 143
III.1 Prior work exploring substituent effects on absorption wavelength in fluorescent proteins 145
III.2 Structure of ThioFluor analogues to be explored 148
III.3 Evaluation of ThioFluor analogues with *n*-butyl amine in ethanol 151
III.4 Coupling of ThioFluor analogues to Q108K:K40L:T51V:T53S:R58W:Y19W:L117E 156
III.5 Coupling of ThioFluor analogues to Q108K:K40L:T51V:T53S:R58W:Y19W:A33W 168
III.6 Coupling of ThioFluor analogues to Q108K:K40L:R58F 172
III.7 Coupling of ThioFluor analogues to Q108K:K40L:T51V:T53S:R58W 174
III.8 Conclusions and proposed projects 176
III.8.1 Proposed design of a multi-input fluorogenic probe 178
III.8.2 Development of near infrared and large Stokes shift fluorescent probes 184
APPENDIX 186
REFERENCES 193

CHAPTER IV: ENGINEERING OF hCRBPII/THIOFLUOR INTO A LARGE STOKES SHIFT FLUORESCENT PROTEIN 202
IV.1 Preliminary observation of dual fluorescence in hCRBPII/ThioFluor 203
IV.2 Exploration of amino acid substitution at residue 40 on fluorescence 210
IV.3 Modification of active site residues to glutamic acid 222
IV.4 Effect of point mutations in Q108K:K40E:T53A:R58L:Q38F:Q4F on ESPT fluorescent properties 227
IV.5 Energy transfer from tryptophan fluorescence to the bound ligand 230
IV.6 Live cell imaging of large Stokes shift hCRBPII/ThioFluor complexes 235
IV.7 Excited state proton transfer is observed in ThioFluor analogues 238
IV.8 Conclusions 247
REFERENCES 248

CHAPTER V: DEVELOPMENT OF A SINGLE PROTEIN FLUORESCENT RATIOMETRIC PH SENSOR 255
V.1 Initial observation of pH induced wavelength shifts in a ThioFluor/hCRBPII complex 258
V.2 Optimization of hCRBPII for a ratiometric pH probe 265

V.3 L117D mutation leads to a much larger change in the iminium maxima. 272
V.4 Imaging of hCRBPII mutants displaying the pH induced wavelength shift with ThioFluor 282
V.5 ThioFluor analogues also show a pH induced wavelength shift 288
V.6 Conclusions 289
REFERENCES 291

CHAPTER VI: PROMOTING MONOMERIZATION OF hCRBPII MUTANTS .. 293
VI.1 Effect of T51 mutations on the monomer/dimer ratio during bacterial overexpression in hCRBPII mutants 295
VI.2 Spectroscopic properties of Q108K:K40L:T51X mutants 297
VI.3 Imine isomerization of Q108K:K40L:T51X mutants 299
VI.4 Thermal imine isomerization that leads to a different absorption wavelength of the iminium 302
VI.5 Different rates of iminium formation between monomer and dimer with retinal 304
VI.6 Mutation of Y60 in the triple mutant Q108K:K40L:T51F 309
VI.7 Rational mutation to promote monomerization in hCRBPII mutants 312
VI.8 Dimer formation in CRABPII, another iLBP 313
VI.9 Conclusions 315
REFERENCES 318

CHAPTER VII: MATERIALS AND METHODS 321
VII.1 Site-directed mutagenesis of hCRBPII and CRABPII 321
VII.2 Protein expression and purification of hCRBPII and CRABPII in pET-17b expression plasmid 328
VII.3 Protein extinction coefficient determination 331
VII.4 UV-Vis measurements of hCRBPII/chromophore complexes 337
VII.4.1 pK_a measurements of hCRBPII/chromophore complexes 337
VII.4.2 Kinetic measurements of hCRBPII/chromophore PSB formation 338
VII.5 Fluorescence measurements 339
VII.6 Quantum yield measurements 340
VII.7 Live cell imaging in bacteria 340
VII.8 Cloning for mammalian expression vectors 341
VII.8.1 General cloning protocol 341
VII.8.2 Preparation of plasmids for mammalian expression 343
VII.8.3 Sequences of plasmids described in this thesis 349
VII.9 Mammalian cell culture 355
VII.10 General confocal imaging methods 356
VII.11 Synthesis of chromophores 357
VII.11.1 Synthesis of ThioFluor 357
VII.11.2 Synthesis of ThioFluor-2 360
REFERENCES 364

LIST OF TABLES

Table II-1. Spectroscopic properties of hCRBPII mutants coupled with merocyanine retinal analog (**MCRA**) 60

Table II-2. Spectral properties of **KSD-1-3** in solvents of different polarities. 76

Table II-3. Spectroscopic characterization of **ThioFluor** in various solvents 78

Table II-4. Spectroscopic properties of **ThioFluor**-PSB with *n*-butyl amine 80

Table II-5. Spectroscopic change as a result of Q4F mutation 84

Table II-6. Spectroscopic change as a result of T51V mutation 85

Table II-7. Spectroscopic change as a result of mutating T53 86

Table II-8. Spectroscopic change as a result of Q38F mutation 87

Table II-9. Spectroscopic change as a result of mutating R58 88

Table II-10. Additive effects of T51 and T53 mutations 89

Table II-11. Spectroscopic changes as a result of mutating R58 in Q108K:K40L:T51V:T53S to aromatic residues. 89

Table II-12. Spectroscopic changes as a result of mutating R58 in Q108K:K40L:T51V:T53S 94

Table II-13. Spectroscopic changes as a result of Y19W mutation 98

Table II-14. Spectroscopic changes as a result of introducing bulky residues at the entrance of the binding cavity 100

Table II-15. Spectroscopic changes a result of mutating active site residues to tryptophan 102

Table II-16. Additive effect of F16Y, leading to a final bathochromic shift in wavelength 105

Table II-17. Deducing the origins of the bathochromic shift in Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor** 106

Table II-18. Spectroscopic changes observed upon the introduction of an acidic residue at L117 110

Table II-19. Spectroscopic property of protein mutants designed in order to increase iminium formation 116

Table II-20. Spectroscopic changes as a result of introducing one acidic residue near the iminium..... 117

Table II-21. Spectroscopic changes as a result of introducing two acidic residues near the iminium..... 120

Table II-22. Spectroscopic properties of Q108K:K40L:T53S:V62N/merocyanine aldehyde..... 128

Table III-1. Effect of Tyr66 replacement on spectroscopic properties of GFP .. 148

Table III-2. Spectroscopic properties of **ThioFluor** analogues (free aldehyde) in ethanol 152

Table III-3. Spectroscopic properties of **ThioFluor** analogues (SB and PSB) in ethanol 154

Table III-4. Spectroscopic properties of **ThioFluor**-PSB analogues coupled to Q108K:K40L:T51V:T53S:R58W:Y19W:L117E 159

Table III-5. Rates and half-lives of **ThioFluor**-PSB analogues coupled to Q108K:K40L:T51V:T53S:R58W:Y19W:L117E 166

Table III-6. Spectroscopic properties of **ThioFluor**-PSB analogues coupled to Q108K:K40L:T51V:T53S:R58W:Y19W:A33W..... 169

Table III-7. Rates and half-lives of **ThioFluor**-PSB analogues coupled to Q108K:K40L:T51V:T53S:R58W:Y19W:A33W..... 172

Table III-8. Spectroscopic properties of **ThioFluor**-PSB analogues coupled to Q108K:K40L:R58F..... 173

Table III-9. Spectroscopic properties of **ThioFluor**-PSB analogues coupled to Q108K:K40L:T51V:T53S:R58W 174

Table IV-1. Spectroscopic properties of large Stokes shift fluorescent proteins due to ESPT..... 208

Table IV-2. Spectroscopic properties of the SB as a result of substitution at residue 40 in the hCRBPII protein Q108K:K40L:T53A:R58L:Q38F:Q4F 213

Table IV-3. Spectroscopic properties of the PSB as a result of substitution at residue 40 in the hCRBPII protein Q108K:K40L:T53A:R58L:Q38F:Q4F 217

Table IV-4. Protein expression yields of hCRBPII mutants with glutamic acid introduced at varying locations in Q108K:K40L:T53A:R58L:Q38F:Q4F 223

Table IV-5. Spectroscopic properties of the SB as a result of glutamic acid introduction in the hCRBPII protein Q108K:K40L:T53A:R58L:Q38F:Q4F 224

Table IV-6. Fluorescence lifetimes of hCRBPII/**ThioFluor** complexes upon 1060 nm excitation (three photon excitation) 226

Table IV-7. Protein expression yields as a result of point mutations in Q108K:K40E:T53A:R58L:Q38F:Q4F 228

Table IV-8. Spectroscopic properties of the SB as a result of point mutations in Q108K:K40E:T53A:R58L:Q38F:Q4F 229

Table IV-9. Spectroscopic properties of the PSB as a result of point mutations in Q108K:K40E:T53A:R58L:Q38F:Q4F 229

Table IV-10. Protein expression yields as a result of point mutations in Q108K:K40D:T53A:R58L:Q38F:Q4F 230

Table IV-11. Spectroscopic properties of the SB with tryptophan to phenylalanine mutations in Q108K:K40E:R58L:Q38F:Q4F 232

Table IV-12. Spectroscopic properties of **ThioFluor** analogues coupled with Q108K:K40E:T53A:R58L:Q38F:Q4F 245

Table V-1. Emission ratios as a function of pH for Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**..... 262

Table V-2. Initial mutagenesis of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E for development toward a single fluorescent protein ratiometric sensor 265

Table V-3. Effect of L117D versus L117E in the development toward a single fluorescent protein ratiometric sensor 272

Table V-4. Ratio of emissions obtained upon excitation at 514 nm and 633 nm for EGFP-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-3NLS/**ThioFluor** 284

Table V-5. pH induced wavelength shift observed in Q108K:K40L:T51V:T53S:R58W:Y19W:L117E with **ThioFluor** analogues..... 289

Table VI-1. Effect of T51 mutation in KL:**T51X**:T53S:R58W:Y19W:F16Y 293

Table VI-2. hCRBPII dimerization promoted by T53S, V62N, T51A and T51N mutations..... 294

Table VI-3. Effect of T51X mutations on monomer/dimer formation in Q108K:K40L 296

Table VI-4. Spectroscopic properties of Q108K:K40L:T51X mutants bound to retinal 298

Table VI-5. Thermal isomerization half-lives with substitution of small aliphatic residues at T51 302

Table VI-6. Effect of Y60 mutation in Q108K:K40L:T51F on protein expression yields and rate of iminium formation 310

Table VI-7. Effect of T51 and T53 mutations on monomer/dimer ratio Q108K:K40D:T53A:R58L:Q38F:Q4F:V62E..... 313

Table VII-1. PCR cycling conditions for site-directed mutagenesis 321

Table VII-2. FPLC Source 15Q method..... 330

Table VII-2. FPLC SEC method..... 330

Table VII-3. Extinction coefficients of hCRBPII and CRABPII mutants 332

Table VII-4. PCR cycling conditions for cloning..... 342

LIST OF FIGURES

Figure I-1. **a**. Crystal structure of GFP. The chromophore is shown in green and is located in the center of the β-barrel. Coordinates were obtained from PDB 1EMA. **b**. Proposed scheme for the maturation of the GFP chromophore. Proper protein conformation and subsequent cyclization, oxidation and dehydration are required to form the fluorescent molecule 2

Figure I-2. Select fluorescent proteins and chromophore structures from each spectral class; the conjugated system of each chromophore is colored according to its emission 4

Figure I-3. Equilibrium between the two ionization states of GFP – the neutral phenol species and the anionic phenolate species (top). UV-Vis of GFP at physiological conditions (solid line); excitation with light at each species yields essentially identical emission spectra (dotted line) 5

Figure I-4. Water network responsible for excited state proton transfer in GFP .. 6

Figure I-5. The chemical structure of bilirubin and the crystal structure of bilirubin-bound wild-type UnaG. Coordinates were obtained from PDB 4I3B 8

Figure I-6. **a**. Structure of biliverdin IXα (BV). **b**. Schematic of the bacterial phytochrome photoreceptor (BphP) PAS and GAF domains, showing the BV binding site and reactive cysteine residue. The PHY and effector domains are omitted for clarity. **c**. Crystal structure of BV-bound BphP with BV and the reactive cysteine (Cys12) shown in green. Coordinates were obtained from PDB 3C2W 10

Figure I-7. Self-labeling protein tags: **a**. SNAP-tag. **b**. CLIP-tag. **c**. HaloTag. POI = protein of interest. The conjugated fluorophore is represented by a red star... 12

Figure I-8. Engineering a coumarin ligase. **a**. Two step probe targeting by ligation of an alkyl azide followed by ligation of a fluorophore-conjugated cyclooctyne. The fluorophore is represented by a red star. **b**. Direct coumarin ligation by an LpIA mutant. **c**. Cell-permeable coumarin analogue used for *in vivo* live-cell imaging. **d**. Coumarin labeling in mammalian cells. NES = nuclear export signal, CAAX = prenylation tag, NLS = nuclear localization signal. MAP2 = microtubule-associated protein 2. All scale bars are 10 μm 14

Figure I-9. Coumarin based fluorogenic probes for use in no-wash protein labeling. POI = protein of interest..... 17

Figure I-10. A fluorogenic probe for cell surface proteins. A Nile Red BG derivative targeting SNAP-tagged membrane receptor. The free probe is nonfluorescent, but upon reaction with SNAP-tag fluorescence is activated. The probe (2 μM) was incubated with Chinese hamster ovary (CHO) cells stably expressing SNAP-tagged HIR for 30 min at 37 °C and imaged without and with washing. All scale bars are 10 μm 18

Figure I-11. Methods of site-specific chemical labeling. The site for fluorophore incorporation is represented by a red star..... 22

Figure I-12. **a.** Chemical structure of the fluorogen malachite green. **b.** Live cell imaging of the expressed FAP-mCer3 localized to the cytosol, nucleus, mitochondria, peroxisome and endoplasmic reticulum (ER) of HEK cells. All scale bars are 10 μm..... 25

Figure I-13. Crystal structure of photoactive yellow protein (PYP) bound to 4-hydroxycinnamic acid through Cys69. Chemical structures of 4-hydoxycinnamic acid and 7-hydroxycoumarin-3-carboxylic acid, which also bind PYP, are also shown. Coordinates were obtained from PDB 2PHY 26

Figure I-14. **a**. Chemical structures of the developed cell-permeable probes for labeling PYP. **b**. Live cell imaging of MBP-PYP, PYP-NLS and PYP-EGFR with TMBDMA. The top row shows fluorescence obtained upon excitation at 473 nm, using a 490-590 nm emission filter. The controls for nontransfected cells are shown in the bottom panel. MBP = maltose binding protein, NLS = nuclear localization signal, EGFR = epidermal growth factor receptor. Scale bar = 10 μm. **c**. Time-lapse live cell imaging of PYP-NLS with TMBDMA (left). Scale bar = 5 μm and plot of average fluorescence intensity of TMBDMA against incubation time (right) 28

Figure I-15. Scheme of Yellow Fluorescence-Activating and absorption-Shifting Tag (Y-FAST). Ligand HBR and HMBR are non-fluorescent, but become fluorescent and absorption red-shifts upon interaction with Y-FAST 29

Figure I-16. Chemical structure of merocyanine retinal aldehyde (**MCRA**) and **MCRA** bound to an engineered CRABPII variant. Coordinates were obtained from PDB 3FEP. Fluorescence imaging of *E. coli* cells expressing CRABPII variant (left; overlay with brightfield on the right). Scale bar is 10 μm..... 31

Figure II-1. Reaction and spectroscopic properties of merocyanine aldehyde (**MCRA**) with hCRBPII. Coordinates were obtained from PDB 3FEP 58

Figure II-2. UV-Vis (blue curve) and emission spectra (red curve) of Q108K:K40L:T51V:R58F/**MCRA** 61

Figure II-3. Labeling of EGFP-hCRBPII$_{tetra}$ in HeLa cells. All samples were incubated with 250 nM **MCRA** for 1 h at 37 °C, washed with PBS buffer two times and then incubated at 37 °C for the indicated amount of time (left to right: 0 min, 2 h, 4 h and 8 h). The right most panel is labeling of EGFP-hCRBPII$_{tetra}^{Q108L}$ after incubating with 250 nM **MCRA** for 1 h at 37 °C, washed with PBS buffer two times and imaged immediately. All scale bars are 20 μM 62

Figure II-4. Overlay of CRABPII (R111K:R132L:Y134F:T54V:R59W, shown in magenta) and hCRBPII (Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4R, shown in green). Coordinates were obtained from PDB 4I9S and 4EEJ 65

Figure II-5. Optimization of **MCRA** concentration for live cell imaging in HeLa cells expressing EGFP-hCRBPII$_{nona}$; various concentrations (125 nM, 250 nM and 500 nM) were tested. All scale bars are 20 μM 66

Figure II-6. Labeling of EGFP-hCRBPII$_{nona}$ in Hela cells. All samples were incubated with 250 nM **MCRA** for 30 sec at room temperature, washed with PBS buffer two times and imaged immediately. Localization signals 3NLS and NES targeted hCRBPII to the nucleus and extranuclear space, respectively. NLS = nuclear localization sequence, NES = nuclear export sequence. All scale bars are 20 μM 67

Figure II-7. **a**. Simplified Jablonski diagram. **b**. Diagram of absorption and emission spectra; the difference in the band maxima is the Stokes shift 70

Figure II-8. Examples of fluorescent solvatochromic dyes. The wavy lines indicate the common point for dye conjugation 72

Figure II-9. Origins of solvatochromicity in fluorescence 73

Figure II-10. **a**. Structure of Betaine 30. **b**. Solutions of Betaine 30 in (1) methanol, (2) ethanol, (3) 1-octanol, (4) *N,N*-dimethylacetamide and (5) dichloromethane 74

Figure II-11. Fluorescence response of **A-41S** to HSA. **a.** 5 μM **A41-S** with HSA at 0, 1, 2, 3, 4, 5, 7.5, 10, 15, 20 and 30 μM. **b.** Structure of **A41-S**. **c.** Corresponding intensity changes are the ratio of intensity of **A-41-S** with/without HSA 75

Figure II-12. Structures of Dapoxyl dyes used to investigate the effect of the pi-linker 75

Figure II-13. Spectroscopic properties of **ThioFluor** in different solvents. **a**. UV-Vis and **b**. Fluorescence spectra of **ThioFluor**. **ThioFluor** under **c**. white light

and **d.** UV irradiation at 365 nm (with TLC handlamp). **e.** Stokes shift in different solvents versus the $E_T(30)$ value indicates that **ThioFluor** is solvatochromic 78

Figure II-14. Formation of the protonated Schiff base (PSB) of **ThioFluor** with *n*-butyl amine in ethanol 79

Figure II-15. Stokes shift in different solvents versus the $E_T(30)$ value indicates that **ThioFluor**-PSB is not solvatochromic. Different correlations are observed in protic and aprotic solvents 81

Figure II-16. Polar residues mutated in order to remove polarity from the binding pocket. Coordinates obtained from PDB 4EXZ (hCRBPII-Q108K:K40L/retinal). 82

Figure II-17. Water mediated hydrogen bonding between Q4 and the hCRBPII/retinal iminium. Coordinates obtained from PDB 4EFG (hCRBPII-Q108K:K40L:T51V:T53C:R58W:T29L:Y19W/retinal) 83

Figure II-18. Water mediated hydrogen bonding between T51 and T53. Coordinates obtained from PDB 4EXZ (hCRBPII-Q108K:K40L/retinal) 85

Figure II-19. Water mediated hydrogen bonding between Q38 and Q128. Coordinates obtained from PDB 4EXZ (hCRBPII-Q108K:K40L/retinal) 87

Figure II-20. Crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor** and space-filling representation. Electron density is shown at 1σ 90

Figure II-21. **a**. Crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor** (cyan) overlaid with Q108K:K40L:T51V:T53S:R58Y/**ThioFluor** (green). **b**. Water mediated hydrogen bonding between R58Y and T29 92

Figure II-22. Crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor** (cyan) overlaid with Q108K:K40L:T51V:T53S:R58Y/**ThioFluor** (green) and Q108K:K40L:T53A:R58F/**ThioFluor** (pink) 93

Figure II-23. **a**. Crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor**, highlighting the water mediated network from Y19 to **ThioFluor**. **b**. Crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor** shows this water network is abolished..... 97

Figure II-24. Residues mutated (shown in green) in an attempt to encapsulate the binding cavity. Crystal structure is of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor** 99

Figure II-25. Residues where tryptophan was introduced in an attempt to red-shift wavelength are shown in green. Crystal structure is of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor** 101

Figure II-26. Crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L117E/**ThioFluor**. Mutation of I25 to tryptophan would sterically clash with A33W 103

Figure II-27. Crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor**. The water mediated hydrogen bonding between T53S and R58W is highlighted.... 107

Figure II-28. Residues chosen for introduction of an acidic residue, in order to interact with the iminium, based on the crystal structure of Q108K:K40L/retinal. Coordinates obtained from PDB 4EXZ 108

Figure II-29. Iminium formation of Q108K:K40L:T53A and Q108K:K40L:T53S monomers and dimers. From left to right are the UV-Vis of the thermodynamic protein/**ThioFluor** complex, iminium formation as a function of time for the monomer and dimer. Data is shown in black, while curve fitting is a dotted red line (fit to pseudo first order kinetics, measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7) 111

Figure II-30. Iminium formation of KLA:R58L(R58F) and KLA:Q38A(M) monomers and dimers. From left to right are the UV-Vis of the thermodynamic protein/**ThioFluor** complex, iminium formation as a function of time for the monomer and dimer. Data is shown in black, while curve fitting is a dotted red line (fit to pseudo first order kinetics, measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7). KLA = Q108K:K40L:T53A..... 112

Figure II-31. Iminium formation of KLA:R58L:Q38F and KLA:R58L:Q38M monomer. From left to right are the UV-Vis of the thermodynamic protein/**ThioFluor** complex, iminium formation as a function of time. Data is shown in black, while curve fitting is a dotted red line (fit to pseudo first order kinetics, measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7). KLA = Q108K:K40L:T53A..... 115

Figure II-32. Positions for introducing an acidic reside to blue-shift wavelength. Based on the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor** . 118

Figure II-33. **a**. Crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor**, highlighting the acidic residues tested in this series of protein mutants. **b**. – **c**. The residues L117 and K40 (shown in grey) were mutated to glutamic acid in Pymol..... 119

Figure II-34. Range of absorption and emission achieved in this study by coupling hCRBPII monomers with **ThioFluor**. Mutants from left to right are (1) Q108K:K40L:T51D:T53A:R58Y:Q38F:Q4F:L117E (note: data was smoothed by averaging because of low absorbance intensity), (2) Q108K:K40D:T53A:R58Y:Q38F:Q4F, (3) Q108K:K40L:T53A:R58L, (4) Q108K:K40L:T51V:T53S:R58G, (5) Q108K:K40L:T51V:T53S:R58W:Y19W, (6) Q108K:K40L:T51V:T53S:R58W:Y19W:F16Y, (7) Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:Q4W 122

Figure II-35. Plot of emission versus absorbance for the approximately 80 hCRBPII/**ThioFluor** complexes examined in this study yields a linear fit, with R^2 = 0.93 122

Figure II-36. Spectroscopic properties of Q108K:K40L:T53S:V62N with **ThioFluor** including UV-Vis and fluorescence spectra, pK_a, binding affinity, and rate plot. Note: It was later found that the hCRBPII species data shown here is for the dimer 124

Figure II-37. Bacterial imaging of Q108K:K40L:T53S:V62N, labeled with **ThioFluor**..... 125

Figure II-38. Labeling of HeLa cells expressing Q108K:K40L:T53S:V62N-ECFP with **ThioFluor** at various incubation times and concentrations of **ThioFluor**.. 126

Figure II-39. Drastically different spectroscopic properties are observed upon binding **ThioFluor** to Q108K:K40L:T53S:V62N monomer and dimer (for both identical conditions are used: 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7 and 23 °C). The spectra depict the course of iminium formation (in minutes) .. 127

Figure II-40. Labeling of HeLa cells expressing Q108K:K40L:T53S:V62N-ECFP with merocyanine aldehyde..... 129

Figure II-41. Labeling of U2OS cells expressing Q108K:K40L:T53S:V62N-ECFP with **ThioFluor** and merocyanine aldehyde 130

Figure II-42. Spectroscopic properties of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E with **ThioFluor** including UV-Vis and fluorescence spectra, pK_a, and rate plot 132

Figure II-43. Labeling of HeLa cells expressing hCRBPII-EGFP, hCRBPII-EGFP-3NLS and hCRBPII-EGFP-NES with 10 μM **ThioFluor** (incubated at 37°C for 1 hour). NLS = nuclear localization sequence, NES = nuclear export signal 133

Figure III-1. Proposed scheme for the maturation of the GFP and mCherry chromophores. Proper protein conformation and subsequent cyclization, oxidation and dehydration are required to form the fluorescent state 147

Figure III-2. **ThioFluor** analogues investigated to study structure-property relationships in absorption and fluorescence. Extinction coefficients are for λ_{max} of each chromophore 149

Figure III-3. Plot of Hammett values versus absorption wavelength for **ThioFluor-5**, **ThioFluor-6**, **ThioFluor-4** and **ThioFluor** 153

Figure III-4. In the hCRBPII heptamutant Q108K:K40L:T51V:T35S:R58W:Y19W:L117E, the distance between the iminium nitrogen and L117E is 2.9 Å; it is this ionic interaction that presumably leads to the high pK_a and holds the iminium as a *trans* isomer. Density is shown at 1σ 157

Figure III-5. Schematic representation of the geometric arrangement of the donor and acceptor groups upon excitation in planarized intramolecular charge transfer (PLICT) and twisted intramolecular charge transfer (TICT) 163

Figure III-6. Emission maxima versus absorption maxima for **ThioFluor** analogues bound to hCRBPII heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E 165

Figure III-7. Double iminium pK_as observed upon coupling of hCRBPII pentamutant Q108K:K40L:T51V:T53S:R58W with **a**. **ThioFluor**, **b**. **ThioFluor-3** and **c**. **ThioFluor-4**. Absorption intensity of the iminium is plotted as a function of pH. Two curve fittings provide the two different pK_a values 175

Figure III-8. Plots of Hammett value versus absorption wavelength, rate of iminium formation and pK_a for **ThioFluor**, **ThioFluor-4** and **ThioFluor-6** coupled with the hCRBPII mutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E..... 178

Figure III-9. Proposed fluorophore and reaction scheme for an improved fluorogenic probe 180

Figure III-10. Proposed chemical steps of the *Pseudomonas sp*. strain 4-chlorobenzoyl-CoA dehalogenase-catalyzed conversion of 4-chlorobenzoyl-CoA to 4-hydroxybenzoyl-CoA..... 182

Figure III-11. Crystal structure of 4-chlorobenzoyl coenzyme A dehalogenase from *Pseudonomas sp.* Catalytic residues are shown in cyan and aromatic residues responsible for hydrophobic packing are shown in green. Coordinates obtained from PDB 1NZY..... 183

Figure III-12. Formation of **ThioFluor-PSB** with *n*-butyl amine in ethanol 187

Figure III-13. Formation of **ThioFluor-2-PSB** with *n*-butyl amine in ethanol 187

Figure III-14. Formation of **ThioFluor-3-PSB** with *n*-butyl amine in ethanol 188

Figure III-15. Formation of **ThioFluor-4-PSB** with *n*-butyl amine in ethanol 188

Figure III-16. Formation of **ThioFluor-5-PSB** with *n*-butyl amine in ethanol 189

Figure III-17. Formation of **ThioFluor-6-PSB** with *n*-butyl amine in ethanol 189

Figure III-18. Formation of **ThioFluor-7-PSB** with *n*-butyl amine in ethanol 190

Figure III-19. Formation of **ThioFluor-8-PSB** with *n*-butyl amine in ethanol 190

Figure III-20. Formation of **ThioFluor-9-PSB** with *n*-butyl amine in ethanol 191

Figure III-21. Formation of **ThioFluor-10-PSB** with *n*-butyl amine in ethanol .. 191

Figure III-22. Formation of **ThioFluor-11-PSB** with *n*-butyl amine in ethanol .. 192

Figure IV-1. Dual fluorescence observed upon irradiation of the SB of Q108K:K40D:T53A:R58L:Q38F:Q4F bound with **ThioFluor** 203

Figure IV-2. Comparison of fluorescence spectra upon irradiation of the SB and PSB of Q108K:K40D:T53A:R58L:Q38F:Q4F/**ThioFluor** 204

Figure IV-3. UV-Vis and fluorescence spectra obtained as a result of changing pH in Q108K:K40D:T53A:R58L:Q38F:Q4F/**ThioFluor** 205

Figure IV-4. UV-Vis (blue) and emission spectra (red) of Q108K:K40D:T53A:R58L:Q38F:Q4F-dimer/**ThioFluor** at pH 7.1 and pH 5.2 .. 206

Figure IV-5. Water network responsible for the excited state proton transfer in GFP 207

Figure IV-6. UV-Vis, emission and excitation spectra of Q108K:K40D:T53A:R58L:Q38F:Q4F:V62E/**ThioFluor** 209

Figure IV-7. Location of residue 40 in the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor**. L117D (shown in gray) was modeled in Pymol 211

Figure IV-8. Characterization of SB fluorescent properties of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor**. **a**. UV-Vis and emission spectra after 397 nm excitation. **b**. Excitation spectrum at 605 nm. **c**. Emission spectra upon 280 nm excitation. **d**. Overlay of 280 nm emission spectrum and UV-Vis of the hCRBPII-SB showing spectral overlap. All data was collected at pH 7.2 ... 214

Figure IV-9. **a**. pK_a titration of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor**. **b**. SB emission as a function of pH 215

Figure IV-10. Plot of fluorescence intensity at 605 nm versus time for binding of **ThioFluor** to Q108K:K40E:T53A:R58L:Q38F:Q4F at 23 °C fit to second order rate kinetics 216

Figure IV-11. **a**. Comparison of fluorescence spectra upon irradiation of the SB and PSB of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor**. **b**. PSB emission spectra as a function of pH 218

Figure IV-12. Chains A, B and C of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor**. Density is shown at 1σ in Pymol 219

Figure IV-13. Hydrogen bonding network centered around K40E and T51, based on chain A of the crystal structure Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor**. Some side chains are not shown for clarity 220

Figure IV-14. UV-Vis and emission spectra of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor** SB (pH 7.1) and PSB (pH 4.5) 221

Figure IV-15. Residues mutated to glutamic acid to test whether its location allows ESPT to occur; two views are shown. Based on the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor**. Mutation to glutamic acid of residues shown in grey did not lead to soluble protein expression..... 222

Figure IV-16. Crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor** (Chain A), showing the location of residues 40, 42, and 51 225

Figure IV-17. Four tryptophan residues are present in wild type hCRBPII. Based on the crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/ **ThioFluor**..... 231

Figure IV-18. Emission spectra of Q108K:K40E:R58L:Q38F:Q4F and subsequent mutation of the four tryptophan residues in hCRBPII 233

Figure IV-19. Live cell imaging of HeLa cells expressing EGFP-Q108K:K40E:T53A:R58L:Q38F:Q4F and EGFP-Q108K:K40E:T53A:R58L:Q38F:Q4F-3NLS, labeled with 10 μM **ThioFluor** for one minute at 37 °C. NLS = nuclear localization sequence 236

Figure IV-20. Non-transfected HeLa cells incubated with **ThioFluor** show background fluorescence when excited at 405 nm (LP 615 nm emission filter) 237

Figure IV-21. UV-Vis and emission spectra of **ThioFluor**, **ThioFluor**-SB and **ThioFluor**-PSB when bound to *n*-butyl amine 238

Figure IV-22. UV-Vis and emission spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-3** at pH 7.1 and pH 4.7 239

Figure IV-23. UV-Vis and emission spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-4** at pH 7.0 and pH 4.5 240

Figure IV-24. UV-Vis and emission spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-7** at pH 7.1 241

Figure IV-25. UV-Vis, emission spectra and pH titration of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-8** 242

Figure IV-26. UV-Vis, emission spectra and pH titration of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-9** 243

Figure IV-27. UV-Vis, emission spectra and pH titration of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-10** 244

Figure V-1. Two-protein-based ratiometric probe design. **a**. The selected proteins have distinct wavelength (UV-Vis) and pK_a values. **b**. pH titration of the fused protein. (c) Ratiometric analysis of the fused protein system at two different concentrations 256

Figure V-2. Single protein ratiometric probe design. **a**. Titration of L117E leads to a shift in wavelength. **b**. Removal of L117E does not lead to the pH induced wavelength change. **c**. A blue-shift in wavelength results due to localizing charge on the iminium through interaction with the carboxylate, while protonation of the L117E side chain leads to less localized charge 257

Figure V-3. Acid-base titration of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**..... 259

Figure V-4. Crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** highlighting the location of L117E. Density is shown at 1σ..... 260

Figure V-5. **a**. Crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor**. Density is shown at 1σ. **b**. Overlay of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor** (magenta) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** (cyan)..... 261

Figure V-6. Select standard curves obtained upon excitation of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** at 514/633 nm, 543/633 nm and 594/633 nm 263

Figure V-7. *In vitro* characterization of Q108K:K40L:T51V:T53S:R58H:Y19W:L117E/**ThioFluor** as a single protein ratiometric pH sensor. **a**. Formation of PSB, monitored at 539 nm via UV-Vis, fit to second order rate kinetics. **b**. UV-Vis titration of the hCRBPII/**ThioFluor** complex. **c**. Double pK_a curve obtained by plotting the absorbance intensity at 598 nm as a function of pH. **d**. Standard curve generated by plotting the emission ratio of 514/633 nm versus pH..... 266

Figure V-8. Overlay of Q108K:K40L:T51V:T53:R58W:Y19W:L117E/**ThioFluor** (green) and Q108K:K40L:T51V:T53:R58H:Y19W:L117E/**ThioFluor** (magenta), highlighting R58 and L117E..... 267

Figure V-9. Overlay of Q108K:K40L:T51V:T53:R58W:Y19W:L117E/**ThioFluor** (green) and Q108K:K40L:T51V:T53:R58W:Y19W:L117E:A33W/**ThioFluor** (magenta)..... 268

Figure V-10. *In vitro* characterization of Q108K:K40L:T51V:T53S:R58W:Y19W:117E:A33W/**ThioFluor** as a single protein ratiometric pH sensor **a.** UV-Vis titration of the hCRBPII/**ThioFluor** complex. **b.** pK_a curve obtained by plotting the absorbance intensity at 628 nm as a function of pH. **c.** pK_a curve obtained by plotting the absorbance intensity at 539 nm as a function of pH. **d.** Formation of PSB, monitored at 628 nm via UV-Vis, fit to second order rate kinetics..... 269

Figure V-11. *In vitro* characterization of Q108K:K40L:T51V:T53S:R58H:Y19W:L117E:A33W/**ThioFluor** as a single protein ratiometric pH sensor. **a.** Formation of PSB, monitored at 585 nm via UV-Vis, fit to second order rate kinetics. **b.** UV-Vis titration of the hCRBPII/**ThioFluor**

complex. **c.** pK_a curve obtained by plotting the absorbance intensity at 614 nm as a function of pH. **d.** pK_a curve obtained by plotting the absorbance intensity at 547 nm as a function of pH 271

Figure V-12. *In vitro* characterization of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor** as a single protein ratiometric pH sensor. **a**. UV-Vis titration of the hCRBPII/**ThioFluor** complex. **b**. pK_a curve obtained by plotting the absorbance intensity at 618 nm as a function of pH. **c**. pK_a curve obtained by plotting the absorbance intensity at 525 nm as a function of pH. **d**. Standard curve generated by plotting the emission ratio at 514/633 nm versus pH 273

Figure V-13. Overlay of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** (green) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor** (magenta) 274

Figure V-14. *In vitro* characterization of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioFluor** as a single protein ratiometric pH sensor. **a**. UV-Vis titration of the hCRBPII/**ThioFluor** complex. **b**. pK_a curve obtained by plotting the absorbance intensity at 623 nm as a function of pH. **c**. pK_a curve obtained by plotting the absorbance intensity at 539 nm as a function of pH. **d**. Standard curves generated by plotting the emission ratio of 514/633 nm, 543/633 nm and 594/633 nm versus pH 276

Figure V-15. Overlay of the crystal structures Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor** (magenta) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioFluor** (cyan) 277

Figure V-16. Overlay of crystal structures of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** (green) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioFluor** (cyan) 278

Figure V-17. **a**. Acidification of Q108K:K40L:T51V:T53S:R58H:Y19W:L117D:Q4F/**ThioFluor** complex. **b.** Basification of Q108K:K40L:T51V:T53S:R58H:Y19W:L117D:Q4F/**ThioFluor** complex 280

Figure V-18. Titration of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:A33W:Q4F/**ThioFluor** complex and pK_a obtained by plotting absorbance intensity versus pH at 649 nm 281

Figure V-19. Imaging of HeLa cells expressing EGFP-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-3NLS, labeled with **ThioFluor** for one hour at 37 °C 283

Figure V-20. Optimized imaging of HeLa cells expressing EGFP-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-3NLS upon excitation at 514 nm and 633 nm 283

Figure V-21. Linear fit to the working range of the *in vitro* generated standard curve for Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**..... 285

Figure V-22. Imaging of HeLa cells expressing Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F- 3NLS, labeled with **ThioFluor** for ten minutes at 37 °C..... 287

Figure VI-1. Formation of domain-swapped dimer in hCRBPII-Q108K:K40L:Y60W. Coordinates obtained from PDB 4ZR2 (dimer) and 4ZJ0 (monomer)..... 293

Figure VI-2. Location of T51 in Q108K:K40L/retinal. Coordinates were obtained from PDB 4EXZ..... 295

Figure VI-3. Cycle for imine isomerization in retinal bound hCRBPII and CRABPII variants 300

Figure VI-4. **a**. UV-Vis of Q108K:K40L-monomer/retinal. **b**. Absorbance intensity at SB and PSB as a function of time. **c**. UV irradiation and subsequent isomerization of the imine. **d**. Rate (pseudo-first order) of imine isomerization after UV irradiation 301

Figure VI-5. **a**. UV-Vis of Q108K:K40L:T51N-monomer/retinal. **b.** Normalized absorption intensity versus time indicates that protein unfolding occurs after iminium formation..... 303

Figure VI-6. Crystal structure of Q108K:K40L:T51N-dimer/retinal..... 304

Figure VI-7. UV-Vis of Q108K:K40L:T51F-monomer/retinal and Q108K:K40L:T51F-dimer/retinal 305

Figure VI-8. Overlay of two chains in the crystal structure of Q108K:K40L:T51F-dimer/retinal 306

Figure VI-9. Overlay of three chains in the crystal structure of Q108K:K40L:T51F-monomer/retinal..... 307

Figure VI-10. Overlay of the crystal structures of Q108K:K40L:T51F-monomer/retinal (green) and Q108K:K40L:T51F-dimer/retinal (purple) 307

Figure VI-11. Overlay of the crystal structures of Q108K:K40L:T51F-dimer/retinal (purple) and apo Q108K:K40L:T51F-dimer (green) 308

Figure VI-12. UV-Vis scanning of Q108K:K40L:T51F dimer and monomer over time with retinal 309

Figure VI-13. Absorbance intensity of PSB and SB as a function of time, followed by UV-Vis 311

Figure VI-14. Crystal structure of Q108K:K40L:T51F:Y60A-dimer/retinal 312

Figure VI-15. Size exclusion chromatography trace and denaturing protein gel for each fraction obtained from expression of I52A..... 315

Figure VI-16. Residues shown to affect the formation of domain swapped dimer in hCRBPII. Coordinates obtained from 4ZH9 316

LIST OF SCHEMES

Scheme II-1. Synthesis of **ThioFluor** 77

Scheme VII-1. Detailed synthesis of **ThioFluor** 358

Scheme VII-2. Detailed synthesis of **ThioFluor-2** 361

KEY TO ABBREVIATIONS

A	alanine, Ala
C	cysteine, Cys
D	aspartic acid, Asp
E	glutamic acid, Glu
F	phenylalanine, Phe
G	glycine, Gly
H	histidine, His
I	isoleucine, Ile
K	lysine, Lys
L	leucine, Leu
M	methionine, Met
N	asparagine, Asn
P	proline, Pro
Q	glutamine, Gln
R	arginine, Arg
S	serine, Ser
T	threonine, Thr
V	valine, Val
W	tryptophan, Trp
Y	tyrosine, Tyr

Å	angstrom
Da	dalton
nm	nanometer
µm	micrometer
cm	centimeter
M	molar
µM	micromolar
nM	nanomolar
mmol	millimole
ε	extinction coefficient
h	hour
min	minute
sec	second
t	time
$t_{1/2}$	maturation half-life
Arch	Archaerhodopsin-3
BC	O2-benzylcytosine
Betaine 30	4-(2,4,6-triphenylpyridinium)-2,6-diphenyloxide
BFP	blue fluorescent protein
BG	O6-benzylguanine
BP	band pass
BphP	Bacterial phytochromes photoreceptor

BV	biliverdin IXα
CFP	cyan fluorescent protein
CHO	Chinese hamster ovary
CRABPII	cellular retinoic acid binding protein II
CuAAC	CuI-catalyzed alkyne-azide cycloaddition
DMEM	Dulbecco's modified eagle medium
DMSO	dimethyl sulfoxide
DPBS	Dulbecco's phosphate buffered saline
eDHFR	Escherichia coli dihydrofolate reductase
EGFP	enhanced GFP
EGFR	epidermal growth factor receptor
equiv	equivalent
ER	emission ratio
ESPT	excited state proton transfer
FABP	fatty acid binding protein
FAP	fluorogen-activating proteins
FBS	fetal bovine serum
FlAsH	fluorescein arsenical hairpin binder
FP	fluorescent protein
FPLC	fast protein liquid chromatography
FRET	Förster resonance energy transfer
FRFP	far-red fluorescent protein

GAF	cGMP phosphodiesterase/adenylyl cyclase/FhlA
GFP	green fluorescent protein
hAGT	human O6-alkylguanine-DNA alkyltransferase
HBR	4-hydroxybenzylidene-rhodanine
hCRBPII	human Cellular Retinol Binding Protein II
hCRBPII$_{nona}$	Q108K:K40L:T51V:T53C:R58W:T29L:A33W:Q4F:L117E
hCRBPII$_{tetra}$	Q108K:K40L:T51V:R58F
HIR	human insulin receptor
HMBR	4-hydroxy-3-methylbenzylidene-rhodanine
HOMO	highest occupied molecular orbital
HSA	human serum albumin
HWE	Horner-Wadsworth-Emmons
ICT	intramolecular charge transfer
IDT	Integrated DNA Technologies
iLBP	intracellular lipid binding protein
IPTG	isopropyl-1-thio-D-galactopyranoside
JRA	julolidine retinal analog
KL	Q108K:K40L
KLA	Q108K:K40L:T53A
KLF	Q108K:K40L:R58F
LAP	LplA acceptor peptide
LB	Luria broth

LP	long pass
LplA	lipoic acid ligase
LSS	large Stokes shift
LUMO	lowest unoccupied molecular orbital
MAP2	microtubule-associated protein 2
mBeRFP	monomer blue light-excited RFP
MBP	maltose binding protein
MCRA	merocyanine retinal aldehyde
NBD	nitrobenzoxadiazole
nd	not determined
NES	nuclear export sequence
NIR	near-infrared
NLS	nuclear localization sequence
OD	optical density
OFP	orange fluorescent protein
p-HBI	4-(p-hydroxy-benzylidene)-5-imidazolinone
PAFP	photoactivatable FP
PALM	photoactivated localization microscopy
PAS	period/ARNT/single-minded
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein data bank

PeT	photoinduced electron transfer
PHY	phytochrome
PLICT	planarized intramolecular charge transfer
PSB	protonated Schiff base
PSFP	photoswitchable FP
PSG	penicillin-streptomycin-glutamine
PYP	photoactive yellow protein
QY	quantum yield (Φ)
ReAsH	resofurin arsenical hairpin binder
RESOLFT	reversibly saturatable optical fluorescence transitions
RFP	red fluorescent protein
rsFP	reversibly photoswitchable FP
SB	Schiff base
scFv	human single-chain antibodies
SEC	size-exclusion chromatography
SNAr	nucleophilic aromatic substitution
SPAAC	strain-promoted alkyne-azide cycloaddition
SQ	Source 15Q
TB	terrific broth
TBET	through-bond energy transfer
THF	tetrahydrofuran
TICT	twisted intramolecular charge transfer

TMP	trimethoprim
UV	ultraviolet
Vis	visible
wtGFP	wild-type green fluorescent protein
Y-FAST	Yellow Fluorescence-Activating and absorption-Shifting Tag
YFP	yellow fluorescent protein

CHAPTER I: METHODS EMPLOYED TO FLUORESCENTLY LABEL PROTEINS OF INTEREST

Fluorescence imaging is a powerful tool, typically used to analyze subcellular localization and protein dynamics.[1-12] This chapter serves to review the most commonly employed tools for fluorescence imaging to date.

I.1 Intrinsically fluorescent proteins

The most prevalent imaging tools are fluorescent proteins (FPs), which can be genetically encoded to allow for visualization of a protein of interest; fluorescent proteins have revolutionized the field of cell biology due to their ease of use and the wide array of colors that have been developed.[13-23]

Green fluorescent protein (GFP) is perhaps the canonical example of fluorescent proteins, with the Nobel Prize in Chemistry in 2008 being awarded to Osamu Shimomura, Martin Chalfie and Roger Tsien for the discovery and development of the *Aequorea victoria* GFP. GFP (27 kDa, containing 238 amino acids) was originally isolated from the jellyfish *Aequorea victoria* in the early 1960s by Shimomura.[24, 25] It was not until 1992 that Douglas Prasher was successful in cloning the gene encoding GFP.[26, 27] Soon after, Chalfie was able to express recombinant GFP in C. elegans.[28, 29] Subsequently, Tsien led the effort of protein engineering for the development of innumerable FP technologies.

The results of many studies have led to optimized folding time, protein stability and photostability. Nonetheless, all conventional fluorescent proteins share the same fold and have similar mechanisms of fluorophore formation.

Subsequent to protein folding, the fluorophore is formed auto-catalytically from three amino acids contained in the beta barrel of the protein.[18, 21, 23, 30-33]

Crystal structures of GFP show an 11-stranded beta barrel structure with a single α-helix at the center of the barrel; it is along this central helix, in the protein's hydrophobic core, that the 4-(*p*-hydroxy-benzylidene)-5-imidazolinone (*p*-HBI) chromophore forms **(Figure I-1a)**.[34-36] In GFP, the *p*-HBI chromophore is formed via the cyclization of Ser65-Tyr66-Gly67. The most widely accepted mechanism **(Figure I-1b)** begins with attack of the Gly67 amide nitrogen to the carbonyl carbon of Ser65.[37-39] Subsequently, dehydration yields the imidazolin-5-one intermediate. Lastly, oxygen catalyzes the dehydrogenation of the C_α-C_β bond of Tyr66, yielding the fully conjugated ring structure. It is important to note that upon denaturation, GFP becomes nonfluorescent, but upon refolding it regains its fluorescence.[19, 40]

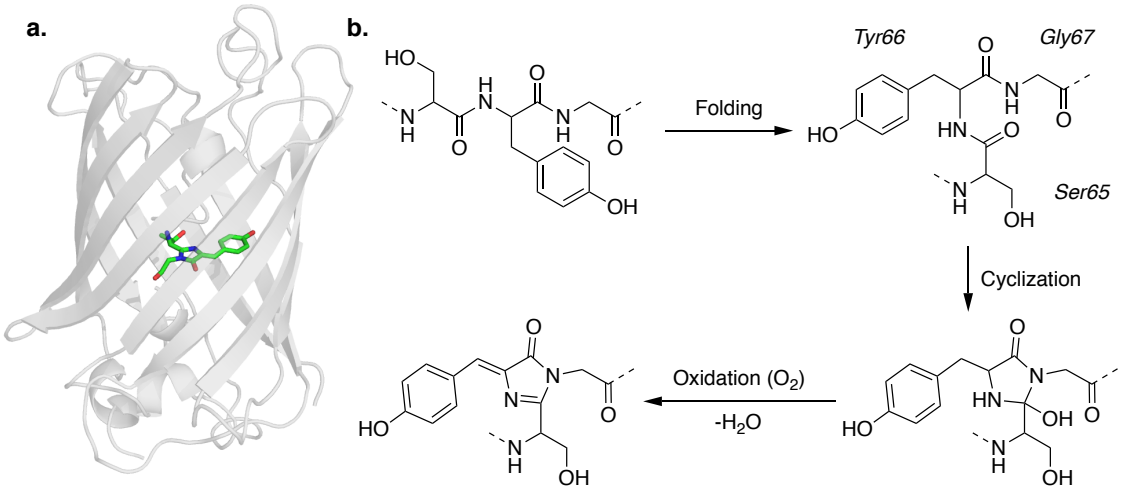


Figure I-1. **a**. Crystal structure of GFP. The chromophore is shown in green and is located in the center of the β-barrel. Coordinates were obtained from PDB 1EMA. **b**. Proposed scheme for the maturation of the GFP chromophore. Proper protein conformation and subsequent cyclization, oxidation and dehydration are required to form the fluorescent molecule.

Since its discovery nearly 60 years ago, GFP has been one of the most widely employed proteins in cell biological applications. In the past half a century, GFP has led to the development of enhanced genetic variants of different colors, spanning nearly the entire visible spectrum **(Figure I-2)**.[13, 15, 16, 33] Interestingly, GFP-like proteins have been found in numerous animal species, including jellyfishes and coral polyps,[41-44] comb jellies,[45] crustaceans, and lancelets.[46, 47]

FPs are typically categorized based upon their emission maxima: blue (BFPs: 440-470 nm), cyan (CFPs: 471-500 nm), green (GFPs: 501-520 nm), yellow (YFPs: 521-550 nm), orange (OFPs: 551-575 nm); red (RFPs: 576-610 nm) and far-red (FRFPs: 611-660 nm).[13, 15, 16, 18, 33] The optical properties of the mature chromophore depend on its specific molecular structure, which results from chemical transformations of the tripeptide maturation (including oxidation, dehydration, cyclization or protonation) and also on interactions of surrounding residues on the polypeptide moiety, such as π-stacking or hydrogen bonding. In RFPs, the conjugation of the π -system in p-HBI is extended;[18, 21, 30] for example in mCherry, the chromophore's conjugation is extended with an acylimine substituent as a result of backbone oxidation (**Figure I-2**).[33, 48]

Red FPs are of particular interest because the red emission channel has lower signal from cellular autofluorescence.[30, 33, 48-53] Furthermore, longer wavelengths allow for great tissue penetration and cause less damage to proteins and DNA. The majority of RFPs were isolated from Anthozoa species.[41-44] It should be noted that in their natural form, anthozoan FPs often form tightly

associated tetramers,[41] which has led many researchers to undertake the feat of designing monomeric variants.

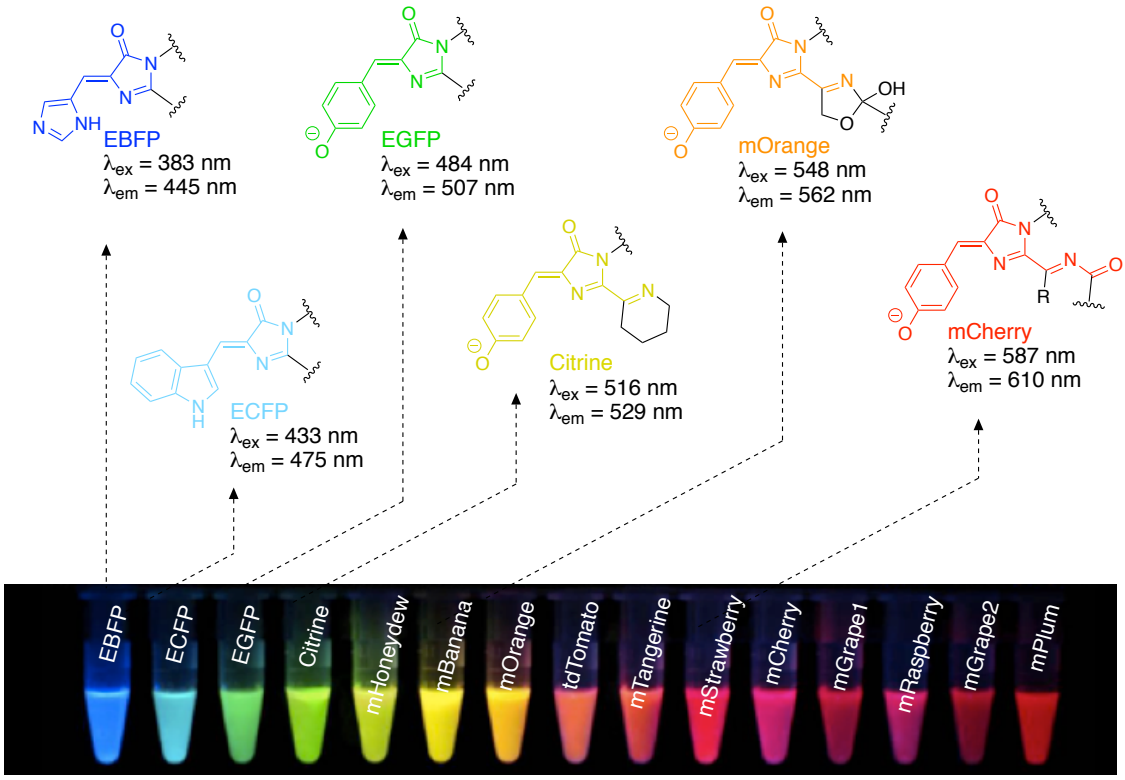


Figure I-2. Select fluorescent proteins and chromophore structures from each spectral class; the conjugated system of each chromophore is colored according to its emission.

In general, the protein microenvironment dictates the fluorescence characteristics of the mature chromophore. In wild-type GFP, the *p*-HBI species experiences two ionization states of Tyr66 – an anionic phenolate species and a neutral, protonated species. The neutral species predominates at physiological conditions, maximally absorbing at 396 nm, while the phenolate is red-shifted to 475 nm **(Figure I-3)**.[32, 34] While the absorbances are drastically different, the two species emit at virtually identical wavelengths, 508 and 503 nm, respectively. This is due to an excited state proton transfer (ESPT), occurring upon excitation

at 400 nm, [54-57] which decreases the pK_a of the tyrosine moiety of the chromophore by several units.[58] The probable mechanism is that proton transfer occurs via the hydrogen bonds of a buried water and Ser205 to Glu222. Meanwhile the side chain of Thr203 rotates to solvate and stabilize the phenolate oxyanion **(Figure I-4)**.[56, 59]

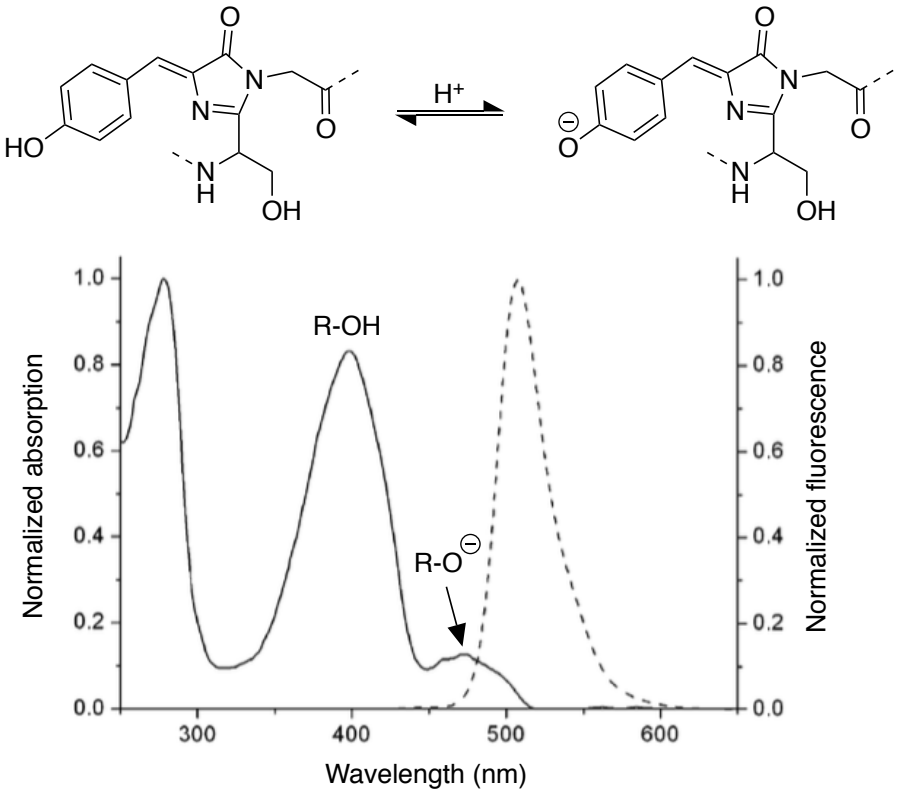


Figure I-3. Equilibrium between the two ionization states of GFP – the neutral phenol species and the anionic phenolate species (top). UV-Vis of GFP at physiological conditions (solid line); excitation with light at each species yields essentially identical emission spectra (dotted line).

The most commonly used mutation to cause ionization of the phenol of the chromophore is replacement of Ser 65 by Thr.[60, 61] In S65T, the wild-type 396 nm excitation peak due to the neutral phenol is suppressed and the 475 nm peak due to the anion is enhanced five- to six-fold in amplitude. As previously stated, S65 donates a hydrogen bond to the buried side chain of Glu222 to allow

ionization of the carboxylic acid. Threonine is too large to adopt the correct conformation in the crowded interior of the protein, forcing the carboxyl group of Glu222 to remain neutral. Consequently, the S65T mutation, along with a F64L substitution designed to improve protein folding efficiency at 37 °C, has been incorporated into an enhanced version of GFP (EGFP) and most of its color variants.

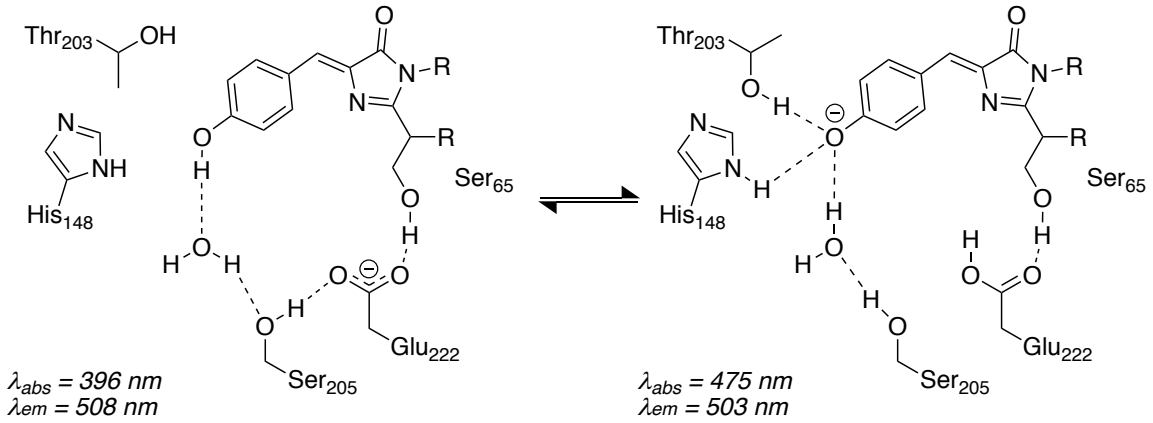


Figure I-4. Water network responsible for excited state proton transfer in GFP.

It was later shown that ESPT is also responsible for numerous large Stokes shifts (LSS) in red FPs, such as mKeima,[62-64] LssmOrange,[65-69] LSSmKates,[70-75] and mBeRFP.[76] A large Stokes shift includes a difference between the excitation and emission maxima larger than 100 nm. As an example, LSSmKate-1 and LSSmKate-2 (evolved via rational mutagenesis of monomeric mKate) possess large Stokes shifts with excitation and emission at 463/624 nm and 460/605 nm, respectively.[71, 72] The only disadvantage compared to EGFP is that their quantum yields are smaller, 0.08 and 0.17, respectively. However, these probes have been utilized simultaneously with ECFP for multicolor imaging of tumor cell migration in living mice.[69]

The broad array of variants that have been developed makes fluorescent proteins useful in a variety of applications to study the organization and function of living systems. Fluorescent proteins have become well established as *in vivo* imaging tools to study subcellular localization, visualize protein-protein interactions, and monitor gene expression, among other applications.[4, 6, 16, 22, 32, 51] Additionally, as microscopy has become more sophisticated, new FPs that are not constitutively fluorescent (as opposed to conventional FPs which are permanently fluorescent) continue to be developed. These include irreversibly photoactivatable and photoswitchable FPs which require induction by light.[6, 77, 78] Members of the first group exhibit an irreversible photoconversion from the nonfluorescent or green fluorescent state to the red fluorescent state; these are termed photoactivatable FPs (PAFPs) and photoswitchable FPs (PSFPs), respectively. Reversibly photoswitchable FPs (rsFPs) can be repeatedly photoswitched, assuming they are fatigue resistant, between fluorescent and nonfluorescent states. These systems have found utility in super-resolution techniques (techniques developed to overcome the inherent diffraction limit of light), such as reversibly saturatable optical fluorescence transition (RESOLFT) imaging[79-81] and photoactivated localization microscopy (PALM).[82, 83]

While fluorescent proteins have been employed in innumerable applications, they still have their limitations. GFP is advantageously monomeric, which allows for it to be fused to a wide variety of partners, without experiencing pitfalls such as aggregation. However, the tendency of FPs to form dimers,

tetramers or other oligomers is problematic when the FP is to be fused to another protein.[16, 43, 84] Additionally, oxygen is required to produce the mature chromophore, which limits the use of fluorescent proteins in obligate anaerobes. Thus, even after decades of research, there is still no fluorescent protein that can satisfy the necessity of all applications.

I.2 Extrinsically fluorescent proteins

Fluorescent proteins are intrinsically fluorescent; they become fluorescent after folding, rather than upon the addition of exogenous ligand. As the limitations of intrinsically fluorescent proteins become apparent, focus has been turned toward the development of fluorescent proteins that bind endogenous biomolecules or require the addition of exogenous ligand. Extrinsically fluorescent proteins that utilize endogenous biomolecules as fluorophores include UnaG as a green fluorescent protein and phytochromes as near infrared proteins.

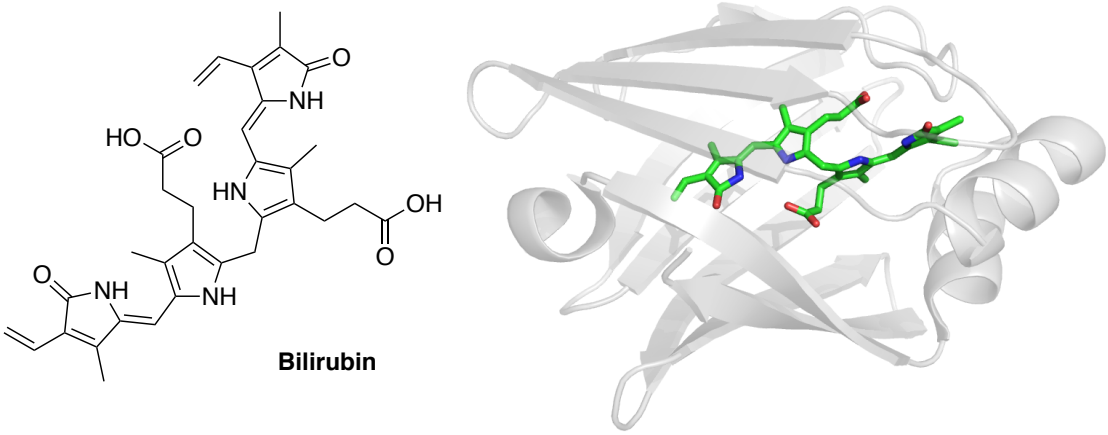


Figure I-5. The chemical structure of bilirubin and the crystal structure of bilirubin-bound wild-type UnaG. Coordinates were obtained from PDB 4I3B.

UnaG is a protein with no intrinsic fluorescence, first isolated from muscle of the Japenese eel Unagi.[85] Studies showed the isolated protein to be a

monomer with a molecular mass of approximately 17 kDa. UnaG belongs to the fatty acid binding protein (FABP) family. Its fluorescence is triggered by bilirubin **(Figure I-5)**,[86] a membrane-permeant heme metabolite, which unlike GFP occurs in an oxygen independent manner. Upon binding (non-covalently), green fluorescence at 527 is observed, with a quantum yield of 0.51.

Bacterial phytochromes photoreceptors (BphPs), which bind biliverdin IXα (BV) (**Figure I-6a**) and fluoresce weakly in the near-infrared, have certain advantages over phytochromes from plants and cyanobacteria, particularly because they bind biliverdin, which is abundant in eukaryotic cells.[87, 88] This and the fact that BphPs exhibit red-shifted NIR absorbance and fluorescence relative to other phytochromoes, renders them more practical toward development of probes for imaging. BphPs bind BV (**Figure I-6c**) in the GAF domain and subsequently a thioether bond is formed with a conserved cysteine in the PAS domain (**Figure I-6b**). Deletion of the PHY and effector domains from *Deinococcus radiodurans* and *Rhodopseudomonas palustris* BphPs, as well as extensive mutagenesis of the PAS and GAF domains yield near-infrared FPs, including IFP1.4, Wi-Phy and RFP, all of which have emission at approximately 700-720 nm, with quantum yields less than 0.10.[89, 90] They have been used for whole-body imaging, and imaging of livers and tumors in mice.[91] These engineered BphPs have also served as templates for the development of other novel fluorescent proteins.

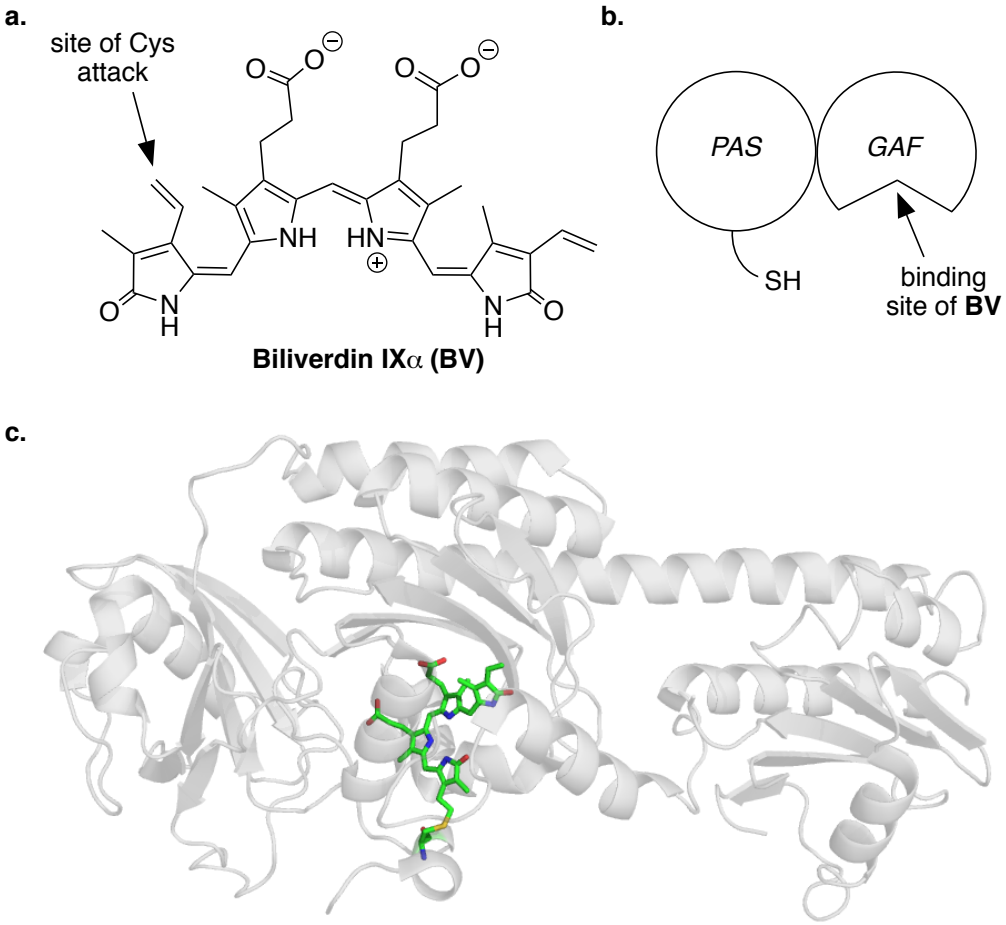


Figure I-6. **a**. Structure of biliverdin IXα (BV). **b**. Schematic of the bacterial phytochrome photoreceptor (BphP) PAS and GAF domains, showing the BV binding site and reactive cysteine residue. The PHY and effector domains are omitted for clarity. **c**. Crystal structure of BV-bound BphP with BV and the reactive cysteine (Cys12) shown in green. Coordinates were obtained from PDB 3C2W.

Recently, it was also shown that BV binds to sandercyanin; the blue colored protein ligand complex was isolated from the skin mucus of walleye.[92] However, sandercyanin differs from previously reported fluorescent BV-binding BphPs in that the chromophore does not appear to be covalently bound via a cysteine residue. BV-bound sandercyanin shows absorbance maxima at 280, 375 and 630 nm at physiological pH, and displays 675 nm red fluorescence when excited at either 375 or 630 nm, with a quantum yield of 0.016.

In terms of utility, these extrinsically fluorescent proteins that bind endogenous biomolecules are (in experimental terms) used similarly to intrinsically fluorescent proteins; however, the endogenous concentration of fluorophore has to be considered, requiring the need to supplement the cells with additional ligand.

I.2.1 Site-specific chemical labeling

Because of the aforementioned limitations of fluorescent proteins, in addition to their large size (~30 kDa), a large effort has been devoted to developing alternatives to fluorescent proteins. Perhaps the largest area of research has been focused on chemical-based methods, particularly in the development of extrinsically fluorescent proteins that become fluorescent upon the addition of an exogenous ligand to an apoprotein.

The most developed alternative to fluorescent proteins is site-specific chemical labeling systems. These methods incorporate a tag fused with a protein of interest and a fluorescent small-molecule; the tags covalently react with a small-molecule substrate containing a fluorophore. One advantage of this chemical recognition mediated labeling approach is that a wide variety of molecules, including NIR fluorescent dyes, can be incorporated into a probe by replacement of the fluorophore moiety of the probe.

The most widely used tags are SNAP-tag, CLIP-tag, HaloTag, and tetracysteine tags, to name a few. The SNAP-tag is a self-labeling protein tag that is mutated from the human suicide protein O6-alkylguanine-DNA alkyl-

transferase (hAGT), engineered to specifically bind O6-benzylguanine (BG) derivatives (**Figure I-7a**).[93-97] A second generation CLIP-tag, reacts irreversibly with O2-benzylcytosine (BC) derivatives.[93, 98] Together SNAP-tag and CLIP-tag systems have been exploited for simultaneous labeling with different probes.[99]

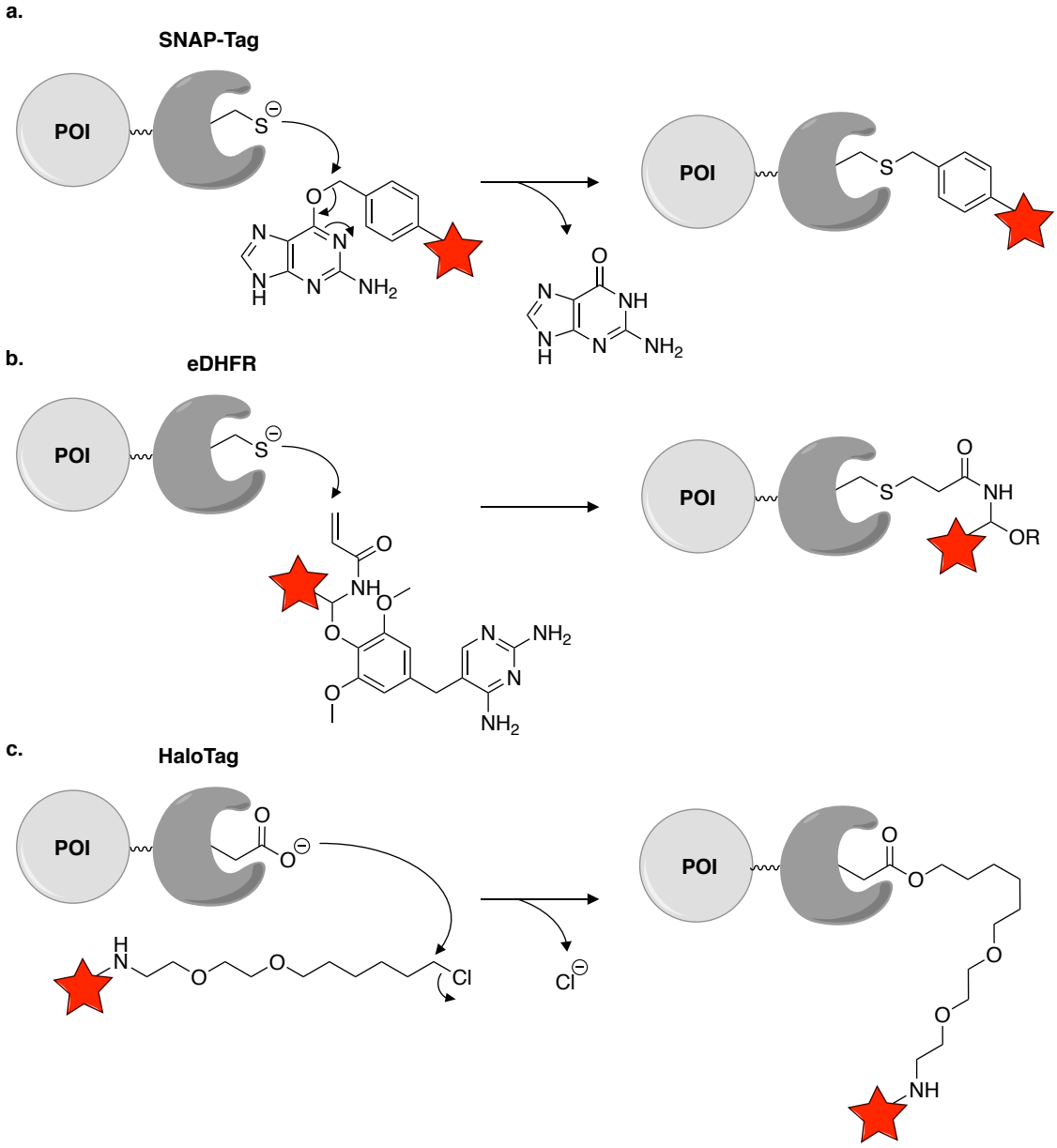


Figure I-7. Self-labeling protein tags: **a**. SNAP-tag. **b**. CLIP-tag. **c**. HaloTag. POI = protein of interest. The conjugated fluorophore is represented by a red star.

The Cornish lab introduced a similar self-labeling tag. 2,4-Diamino-5-(3,4,5-trimethoxybenzyl)pyrimidine (trimethoprim or TMP) tag uses an engineered 18 kDa, monomeric *Escherichia coli* dihydrofolate reductase (eDHFR) to react with trimethoprim-fused fluorophores (**Figure I-7b**).[100, 101] TMP-fluorophore conjugates were shown to label eDHFR fusion proteins (with nM binding affinity) in mammalian cells with low background and fast kinetics.[100, 102-104]

HaloTag (33 kDa) is derived from the bacterial *haloalkane dehalogenase* enzyme from *Rhodococcus rhodochrous*. A covalent bond is formed between the HaloTag protein and fluorophore-conjugated alkyl halide ligand (**Figure I-7c**).[105] Interaction of the enzyme and ligand generates an alkyl-enzyme intermediate via nucleophilic displacement of the alkyl chloride with Asp106 (it should be noted that a H272P mutation is included to prevent hydrolysis and release of the enzyme).[106]

As opposed to the self-labeling proteins discussed above, short peptide sequences have also been developed which prove advantageous because of their small targeting motif. Fluorescein arsenical hairpin binder (FlAsH)[107] and the more red-shifted analog resofurin arsenical hairpin binder (ReAsH)[108] are small biarsenical dyes that covalently react with a short peptide sequence containing four cysteine residues (via metal-ligand recognition). In these, the peptide sequence is Cys-Cys-Xaa-Xaa-Cys-Cys, where Xaa is any amino acid other than cysteine. As in the other systems, the sequence is genetically introduced into the

sequence of the target protein. Tsien and coworkers first described the use of biarsenical reagents for site-specific protein labeling in live cells.[107, 109] However, similar tags have also been developed based on recognition for histidine,[110-113] aspartate,[114, 115] and serine[116] among other recognition moieties.

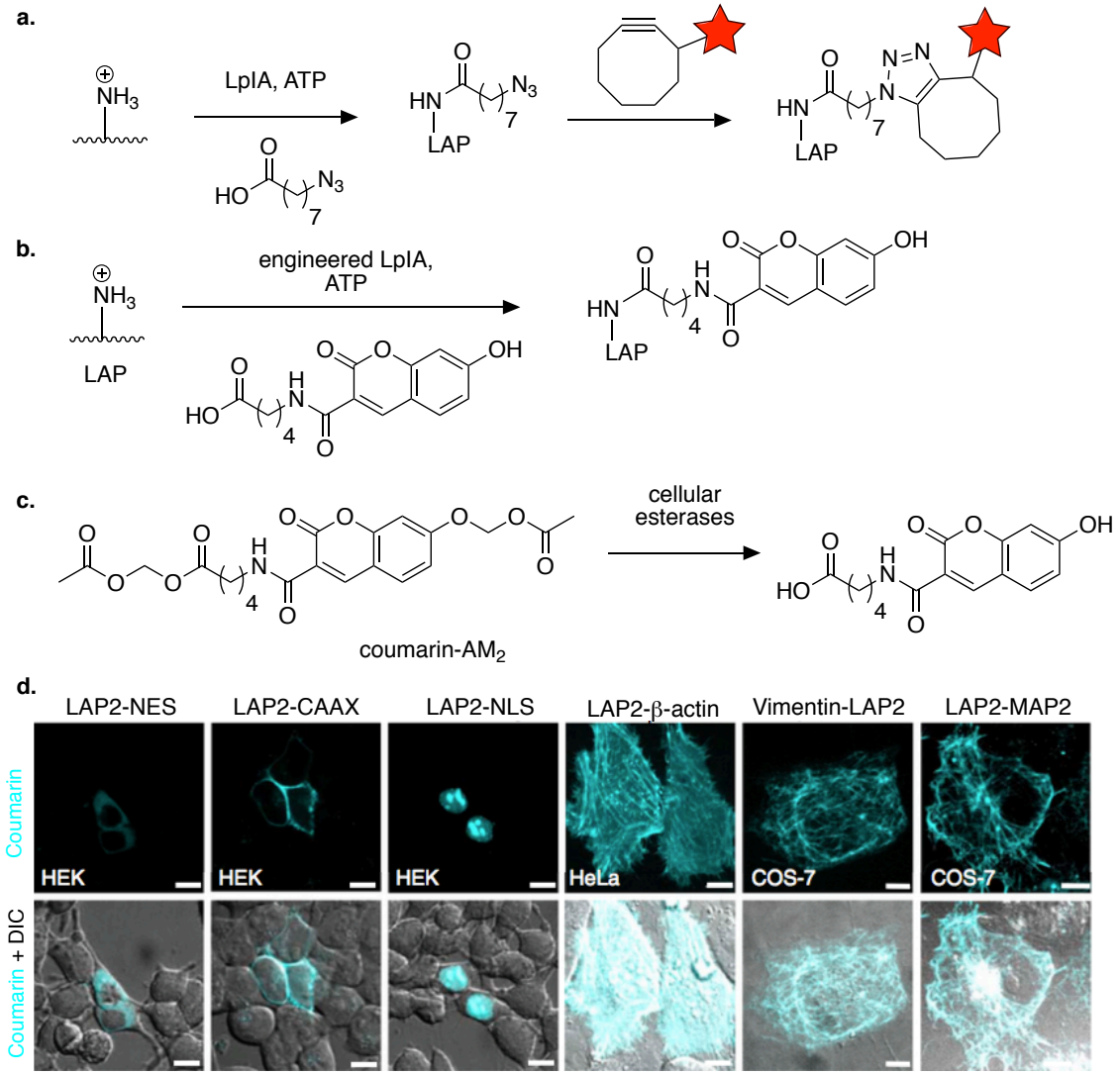


Figure I-8. Engineering a coumarin ligase. **a**. Two step probe targeting by ligation of an alkyl azide followed by ligation of a fluorophore-conjugated cyclooctyne. The fluorophore is represented by a red star. **b**. Direct coumarin ligation by an LplA mutant. **c**. Cell-permeable coumarin analogue used for *in vivo* live-cell imaging. **d**. Coumarin labeling in mammalian cells. NES = nuclear export signal, CAAX = prenylation tag, NLS = nuclear localization signal. MAP2 = microtubule-associated protein 2. All scale bars are 10 μm.

Several enzymes have also been developed to label a protein of interest with a short peptide sequence, including farnesyl transferase,[117, 118] biotin ligase,[119-122] myristyl transferase,[123-125] formylglycine-generating enzyme,[126-128] sortase,[129-132] lipoic acid ligase (LplA)[133-135] and tissue transglutaminase. As an example, LplA is based on an *Escherichia coli* enzyme whose natural function is to ligate lipoic acid onto three *E. coli* proteins involved in oxidative metabolism, essentially catalyzing the acylation of lysine residues of proteins with lipoic acid.[133] The Ting lab showed that LplA could be used for labeling of cell-surface proteins; they demonstrated that the wild-type enzyme could ligate an azidoalkanoic acid instead of lipoic acid.[134] Ligated azide could then be chemoselectively derivatized using cyclooctyne-fluorophore conjugates (**Figure I-8a**). More noteworthy, through LplA active site mutagenesis, they created a ligase capable of directly attaching the fluorophore 7-hydroxycoumarin to proteins fused to a 13-amino acid LplA acceptor peptide (LAP) (**Figure I-8b**).[135] A final analogue of the probe coumarin-AM_2, in which both the carboxylic acid and the 7-hydroxy group are protected with acetoxymethyl groups in order to allow for cell permeability was used for imaging; the groups can be cleaved to reveal reactive probe with intracellular endogenous esterases (**Figure I-8c**). The system was evaluated in mammalian cell lines, including HEK, HeLa and COS-7. Labeling of various LAP2 fusion proteins was also undertaken, where signaling peptides for localization to the nucleus (LAP2-NLS), cytosol (LAP2-NES) or plasma membrane (LAP2-CAAX) were included. Additionally, LAP2 fusions to the

cytoskeletal proteins β-actin and vimentin or microtubule-associated protein 2 (MAP2) could be specifically labeled with coumarin (**Figure I-8d**).

In all of these site-specific chemical labeling systems, the reactive group that covalently binds to the tag is independent of the attached fluorophore, allowing a wide variety of fluorophores to be attached. The capability to explicitly attach chemical probes to individual proteins represents a powerful approach to study processes in living cells. The major drawback in some of these systems is the complexity associated with adding an external substrate that can be fluorescent itself when unbound or fluorescent due to non-specific binding, which reduces the signal-to-noise ratio and detection of protein labeling. Hence, these systems typically require stringent washing to reduce background. Ultimately, this delay in acquisition of data prevents real-time monitoring of fluorescence and it is not always possible to completely remove the unbound probe. It is therefore desirable to build in both spatial and temporal control.

To address the issue of background fluorescence from unreacted probe, much effort has focused on the development of fluorogenic probes to be used in imaging.[136-139] The underlying principle in the design of a fluorogenic probe is that its intrinsic fluorescence is quenched or suppressed until a reaction abolishes the quenching effect to restore the latent fluorescence. Numerous types of quenching mechanisms exist, but they can be generally categorized as follows: environmental quenching, energy transfer, or electron transfer.[140] Environmental quenching arises from the fluorophore's inherent sensitivity to its environment

such as solvent polarity, temperature, and pH, among others.[136, 141-144] Solvatochromic fluorophores are most often used in environment-sensitive fluorogenic protein labeling, because these fluorophores typically are more fluorescent in a less polar medium or environment, such as the hydrophobic environment of a protein active site. More typically, researchers have employed probes reliant on reaction based quenching mechanisms. These include energy transfer, such as Forster resonance energy transfer (FRET),[145-147] through-bond energy transfer (TBET),[148-151] and electron transfer, as observed in photoinduced electron transfer (PeT).[152-155]

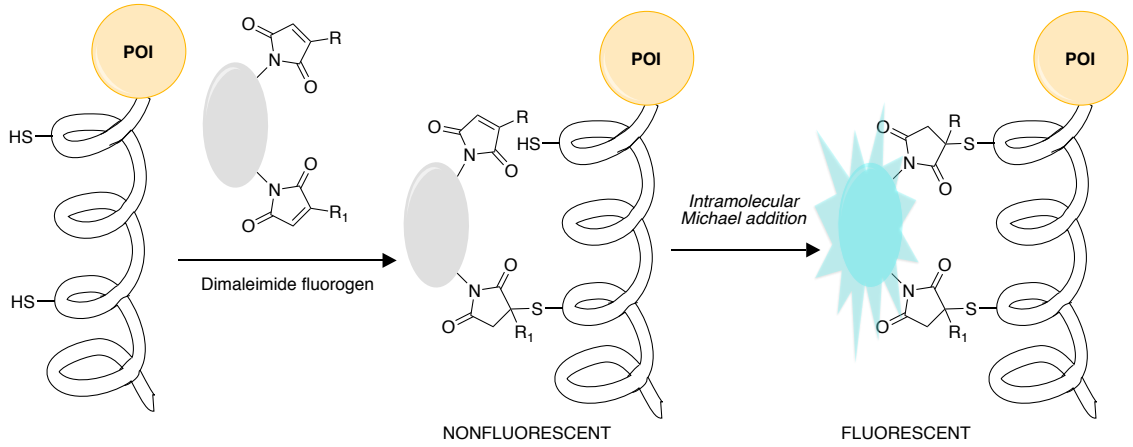


Figure I-9. Coumarin based fluorogenic probes for use in no-wash protein labeling. POI = protein of interest.

As an example, the Keillor lab has developed coumarin-based fluorogenic probes for no-wash protein labeling. In this system, a protein of interest is genetically tagged with a short peptide containing two cysteine residues that are separated by two turns of an α-helix (approximately 10 Å apart).[156-159] The probe becomes fluorescent upon selective reaction with synthetic fluorogenic ligands (**Figure I-9**). These ligands are comprised of a fluorophore and a dimaleimide

moiety, which exists to quench the fluorescence of the fluorophore via PeT until both maleimide groups undergo thiol additions. It should be noted that few native proteins contain two free cysteine residues on their surfaces at the proper distance to react with both maleimide groups. After tuning the reactivity of the probes, the developed labeling method enabled specific labeling of a protein of interest in live cells with no washing and low toxicity.[156]

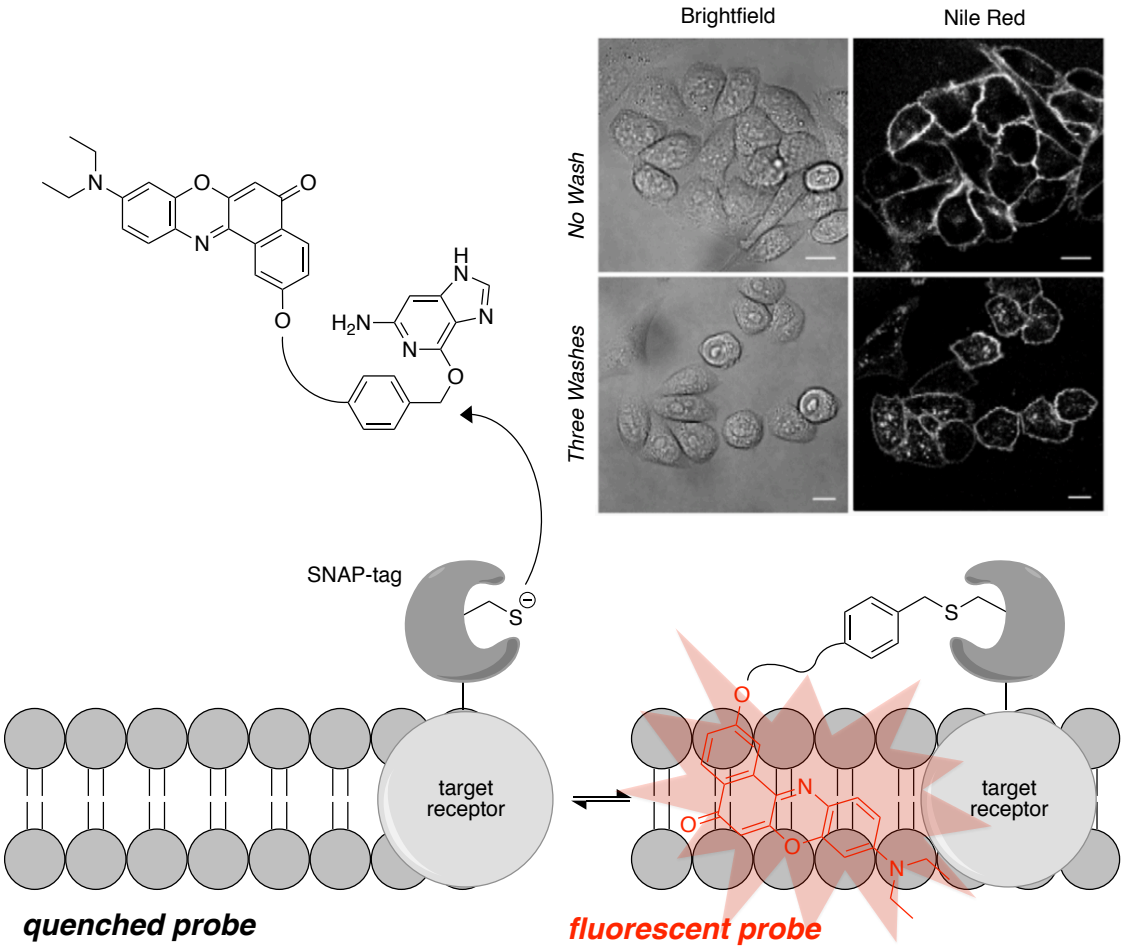


Figure I-10. A fluorogenic probe for cell surface proteins. A Nile Red BG derivative targeting SNAP-tagged membrane receptor. The free probe is nonfluorescent, but upon reaction with SNAP-tag fluorescence is activated. The probe (2 μM) was incubated with Chinese hamster ovary (CHO) cells stably expressing SNAP-tagged HIR for 30 min at 37 °C and imaged without and with washing. All scale bars are 10 μm.

Several groups have also developed FRET-based covalent fluorogenic probes. One such example is a fluorogenic probe for SNAP-tagged plasma membrane proteins based on the dye Nile Red (**Figure I-10**).[97, 160] It takes advantage of Nile Red, a solvatochromic molecule highly fluorescent in an apolar environment, such as cellular membranes, but almost dark in a polar aqueous environment. The low fluorescence properties of the dye in a polar environment are believed to be due to the formation of aggregates that causes self-quenching, and to hydrogen bonding that has been reported to incite radiationless deactivation in anthraquinone and fluorenone derivatives.[161-164] Similar to most solvatochromic probes, the absorption and emission of Nile Red-shifts strongly toward the red part of the spectrum as the solvent becomes more polar.[163]

As expected, Nile Red can spontaneously insert into membranes to become highly fluorescent due to the low polarity of these structures, and is thus a popular probe for staining membranes.[165, 166] However, it is this propensity to localize in the membrane that prevents Nile Red from being used to visualize only a subset of membranes or proteins. Johnsson and coworkers hypothesized that tuning the membrane affinity of Nile Red through chemical derivatization would prevent this spontaneous insertion.[97, 160] The probe possesses a tuned affinity for membranes allowing its Nile Red moiety to insert into the lipid bilayer of the plasma membrane, becoming fluorescent, only after its conjugation to a SNAP-tagged plasma membrane protein. CHO cells stably expressing a human insulin receptor (HIR)-SNAP-tag fusion protein were treated with the probe for 30

min at room temperature. Live cells were imaged for fluorescence after the labeling with and without wash steps. This probe proved to efficiently label SNAP-tagged plasma membrane proteins with minimal background in no-wash experiments (excitation at 572 nm and emission at 632 ± 30 nm).

I.2.2 Bioorthogonal reactions

Similar to the previously described site-specific chemical labeling approaches, other bioorthogonal reactions have been developed. Bioorthogonal reactions are between functional groups that do not react with biological entities under physiological conditions (ambient temperature, pressure, neutral pH, aqueous conditions) but selectively react with each other.[167-171] In these systems, the protein is labeled with one of the functional groups and the fluorophore is labeled with the second moiety (**Figure I-11**).

A group that is essentially absent from biological functionalities is the azide group. Azide-bearing unnatural amino acids have been incorporated into proteins and used in a variety of chemical reactions,[172-174] including reaction with phosphines in Staudinger ligations[175, 176] and reaction with terminal alkynes in [3+2] cycloadditions.[177, 178] Notably, Wang and coworkers developed the first "smart azide probes".[179] An azide in the 3-position of coumarin leads to significant quenching of its fluorescence via intramolecular charge transfer (ICT). Upon reaction with an alkyne, the lone pair of the azide is delocalized into the triazole, leading to fluorescence enhancement. 3-Azido-7-hydroxycoumarin showed the most impressive fluorescence enhancement upon click reaction, over

a hundred- fold in pH 7 phosphate buffered saline (PBS). The robust enhancement in fluorescence of this probe and its moderate water solubility and cell permeability have made this probe a popular choice for the detection of alkynes on intracellular and extracellular targets in biological systems.

The CuI-catalyzed alkyne-azide cycloaddition (CuAAC) suffers from the necessity for Cu^{+}, a species toxic to many living systems.[180, 181] Unfortunately decreasing the Cu^{+} concentration severely hampers the reaction rate.[182-184] The alkyne-azide cycloaddition is accelerated with respect to the uncatalyzed reaction by introducing ring strain into the alkyne (rather than using metal catalysis), creating a reaction dubbed the "strain-promoted alkyne-azide cycloaddition" (SPAAC).[185, 186] These reactions have shown usefulness in labeling biomolecules within complex biological systems, including live mammalian cells and animals.[187]

A light induced 1,3-dipolar cycloaddition reaction between tetrazoles and terminal alkenes has been developed for use in biological settings.[188, 189] This photoclick chemistry has been used to modify purified proteins *in vitro* and also to visualize proteins in living cells.[190-193] Because the system is inducible by UV-light, it has the advantage of spatiotemporal control.

Inverse-electron demand Diels-Alder reactions between tetrazines and strained alkenes or alkynes (including norbornenes and cyclopropenes, among others) yield dihydropyridazines or pyridiazines with nitrogen gas as the byproduct. These reactions have been explored as chemoselective reactions for

labeling in their native settings.[194, 195] Advantageously, they are extraordinarily fast and can be made fluorogenic by conjugation of the tetrazine to a fluorophore; the fluorescence is quenched by the tetrazine, but upon cycloaddition fluorescence is restored.[196]

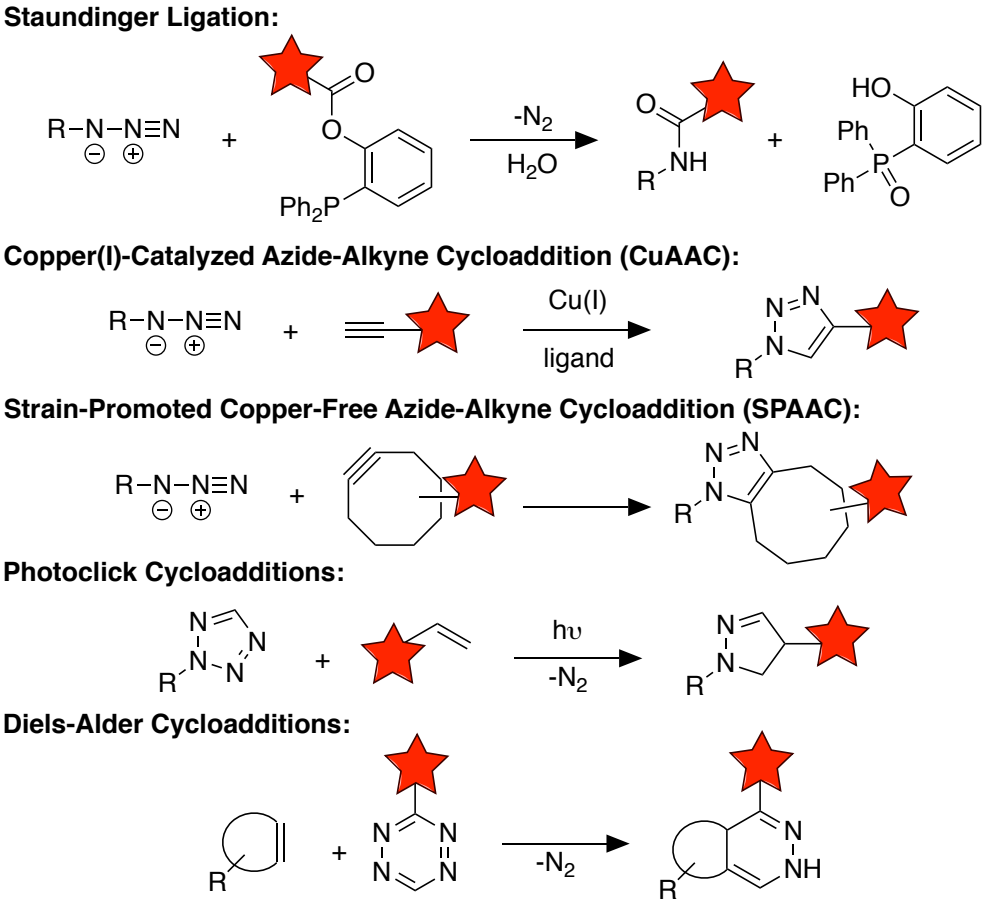


Figure I-11. Methods of site-specific chemical labeling. The site for fluorophore incorporation is represented by a red star.

I.2.3 Small protein tags

The last major class of fluorogenic proteins consists of small protein tags, which are preferable to the aforementioned intrinsically fluorescent proteins due to their smaller size, as larger protein tags might cause unfavorable steric effects on the function or intracellular trafficking of proteins of interest.

Fluorogens are fluorescence generating dyes that are dark until constrained, bound or activated, ensuring that unreacted dye produces a minimal background signal. Fluorescence is produced only when the dye is bound or modified by the target, providing temporal control of the signal based on addition of the dye. Spatial control of the signal is achieved by fusing the activating protein to a specific protein or peptide that targets an organelle. This section will describe three varieties of small protein tags and fluorogen protein/small molecule pairs developed toward no-wash live-cell imaging.

I.2.3.1 Fluorogen activating proteins

Fluorogen-activating proteins (FAPs) are single-chain antibodies that bind a nonfluorescent molecule and stabilize it in a fluorescent state. Library screening of human single-chain antibodies (scFvs) was conducted by yeast surface display to isolate proteins that noncovalently bind to fluorogens such as thiazole orange or malachite green.[197] Human scFvs were chosen because of their relatively small size of less than 30 kDa and the fact that they are easy to use as recombinant tags in fusion proteins.

Thiazole orange and malachite green (**Figure I-12a**) are essentially nonfluorescent in solution, because the free rotation of aromatic rings in the fluorogen is thought to cause the promotion of nonradiative decay and the loss of fluorescence. Rotation around the single bond of the fluorogen is constrained upon binding to the FAP, and thus, fluorescence is “activated.”

In pioneering studies conducted by Waggoner and coworkers, the developed yeast FAPs were displayed on yeast and mammalian cell surfaces. The FAPs were shown to bind fluorogens with nanomolar affinity and brightness levels typical of fluorescent proteins.[197] FAP domains provided a report of protein location and abundance with both temporal and spatial control upon addition of the fluorogen, rendering them more useful than the site-specific chemical labeling systems. The FAP-fluorogen system also allows for multicolor imaging of cells expressing different FAPs.

Fluorogens derived from the malachite green chromophore (**Figure I-12a**) are highly fluorogenic, emit in the far-red spectral region (as compared to thiazole orange that emits in the yellow region), and display low nonspecific labeling in living cells.[197] Bruchez and coworkers evaluated genetically encoded malachite green FAPs for targeting, brightness, and photostability in the cytosol, nucleus, mitochondria, peroxisomes, and endoplasmic reticulum (**Figure I-12b**).[198] Labeling was realized within 20-30 minutes for each protein upon addition of nM concentrations of dye, producing a signal that colocalized with a linked mCerulean3 fluorescent protein. At a higher dye concentration of 1 μM, in the nuclear-targeted FAP, the malachite green dye achieved its labeling plateau in 7 minutes. The FAPs and malachite green dye can be used as specific, rapid and wash-free labels for intracellular organelles in live cells with far-red excitation and emission properties. Notably, labeling was proficiently achieved even in the reducing environment of the cytosol and nucleus, in which FAPs have a

propensity to misfold due to the presence of disulfide bonds within each variable domain.

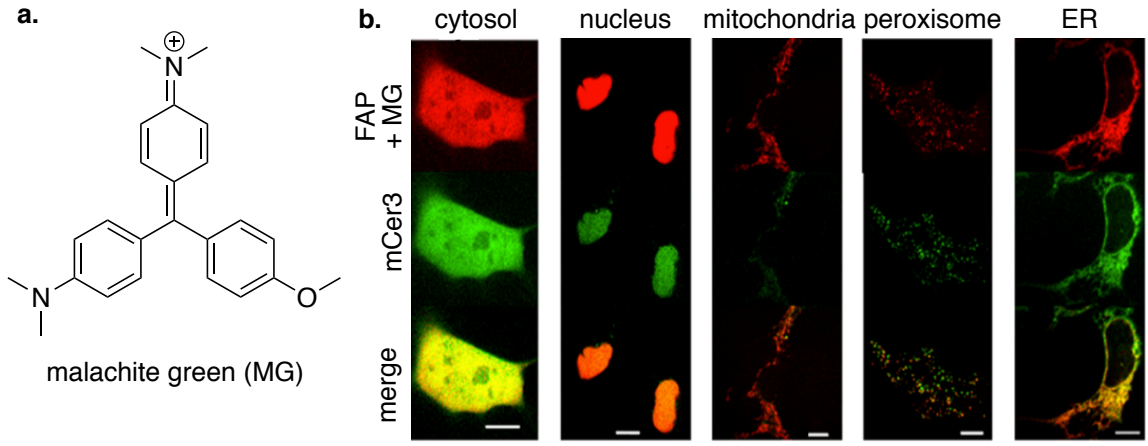


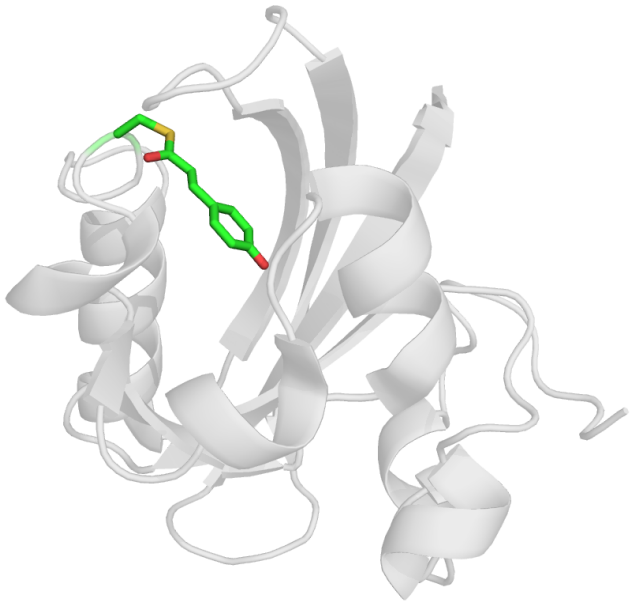
Figure I-12. **a**. Chemical structure of the fluorogen malachite green. **b**. Live cell imaging of the expressed FAP-mCer3 localized to the cytosol, nucleus, mitochondria, peroxisome and endoplasmic reticulum (ER) of HEK cells. All scale bars are 10 µm.

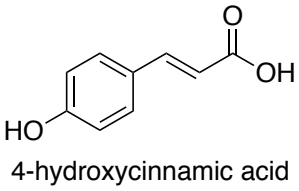
I.2.3.2 Photoactive yellow protein

Another small protein tag is photoactive yellow protein tag (PYP-tag), which consists of 125 amino acids, nearly half the size of GFP. The PYP-tag is derived from photoactive yellow protein, a small cytosolic photoreceptor originally isolated found in the purple photosynthetic bacteria *Halorhodospira halophile*.[199] PYP binds to a natural cofactor, CoA thioester of 4-hydroxycinnamic acid, through transthioesterification with Cys69, the only cysteine residue in the protein (**Figure I-13**).[200]

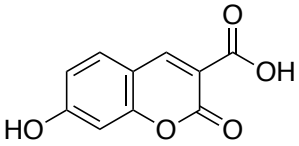
It is known that PYP-tag also binds to the thioester derivative of 7-hydroxycoumarin-3-carboxylic acid (**Figure I-13**);[201] this coumarin ligand has been utilized to design a fluorogenic probe based on the mechanism of static quenching. 7-Hydroxycoumarin-3-carboxylic acid thioester is linked to fluorescein

through an ethylene glycol linker.[202] In the absence of PYP, the probe is not fluorescent. However, upon binding to PYP, dissociation of the coumarin from fluorescein allows for fluorescence turn-on. This first generation probe required a wash step in live-cell imaging due to the extremely long incubation time required (24 h).





4-hydroxycinnamic acid



7-hydroxycoumarin-3-carboxylic acid

Figure I-13. Crystal structure of photoactive yellow protein (PYP) bound to 4-hydroxycinnamic acid through Cys69. Chemical structures of 4-hydoxycinnamic acid and 7-hydroxycoumarin-3-carboxylic acid, which also bind PYP, are also shown. Coordinates were obtained from PDB 2PHY.

Later, the same group replaced the coumarin ligand with cinnamic acid thioester, in order to reduce the intramolecular stacking interaction between the ligand and fluorophore moieties.[203] As compared to the probe containing the coumarin ligand, these probes displayed significantly improved kinetic properties, with up to 110 times acceleration in binding of the probe to PYP ($t_{1/2}$ ~ 15 min). Unfortunately, the probe is not cell-permeable, so direct imaging of cell-surface proteins was undertaken. Epidermal growth factor receptor (EGFR) fused with

the PYP tag allowed for clear fluorescence imaging along the plasma membrane in cells treated with the probe, even without a wash step.

Second generation probes were designed to be fluorogenic, cell-permeable and require a short incubation time. Designed probes were based on 7-dimethylaminocoumarin thioester derivatives.[204] Notably, 7-dialkylaminocoumarin derivatives are environment-sensitive, in that they are scarcely fluorescent in polar solvents, but become fluorescent in low polarity solvents. It was hypothesized that the designed probe would be nonfluorescent in aqueous buffer, but emit when in the protein interior. It should be noted that the trimethylammonium and carboxylic acid moieties were included to increase water solubility (**Figure I-14a**). TMBDMA labeled 50% of PYP-tag for 1.1 min. As previously accomplished, no-wash live-cell imaging of cell surface proteins on the cell membrane allowed for fluorescence observation without washing the cells (**Figure I-14b**). In contrast, non-transfected cells proved to be essentially nonfluorescent. More impressively, genes encoding PYP-tag fused to maltose-binding protein (MBP-PYP) and nuclear localization signals (PYP-NLS) were created and expressed in HEK293T cells. Live cell images were obtained with a no-wash protocol immediately after TMBDMA was incubated with cells expressing either construct for 30 min (**Figure I-14b**). Time-lapse experiments showed the amount of incubation time necessary for fluorescence detection of PYP-tag inside cells. Fluorescence appeared in as little as 2 min after addition of the probe, with the reaction being nearly completely within 6 min (**Figure I-14c**).

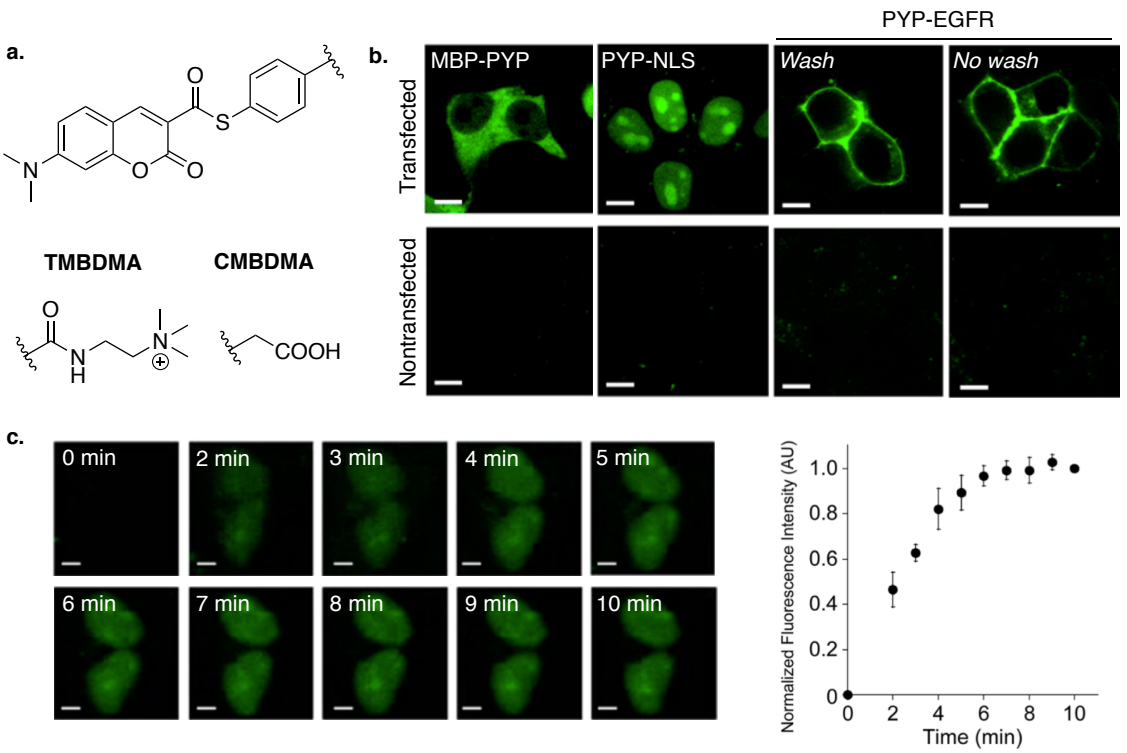


Figure I-14. **a**. Chemical structures of the developed cell-permeable probes for labeling PYP. **b**. Live cell imaging of MBP-PYP, PYP-NLS and PYP-EGFR with TMBDMA. The top row shows fluorescence obtained upon excitation at 473 nm, using a 490-590 nm emission filter. The controls for nontransfected cells are shown in the bottom panel. MBP = maltose binding protein, NLS = nuclear localization signal, EGFR = epidermal growth factor receptor. Scale bar = 10 µm. **c**. Time-lapse live cell imaging of PYP-NLS with TMBDMA (left). Scale bar = 5 µm and plot of average fluorescence intensity of TMBDMA against incubation time (right).

While TMBDMA showed better kinetic properties than CMBDMA, it was apparent that CMBDMA was brighter than TMBDMA. Thus, latest efforts focused on improving the kinetics of CMBDMA with PYP.[205, 206] The CMBDMA2/PYP3R-NLS pair enabled rapid and clear detection of fluorescence in nuclei only 1 min after the addition of the probe, showing saturation after 3 min.[205]

Most recently, Yellow Fluorescence-Activating and absorption-Shifting Tag (Y-FAST), a small protein tag enabling fluorescent labeling of proteins in living cells and multicellular organisms, was developed (**Figure I-15**).[207] Y-FAST is an

engineered variant of PYP that was evolved through yeast surface display to reversibly bind 4-hydroxybenzylidene-rhodanine (HBR) or 4-hydroxy-3-methylbenzylidene-rhodanine (HMBR).

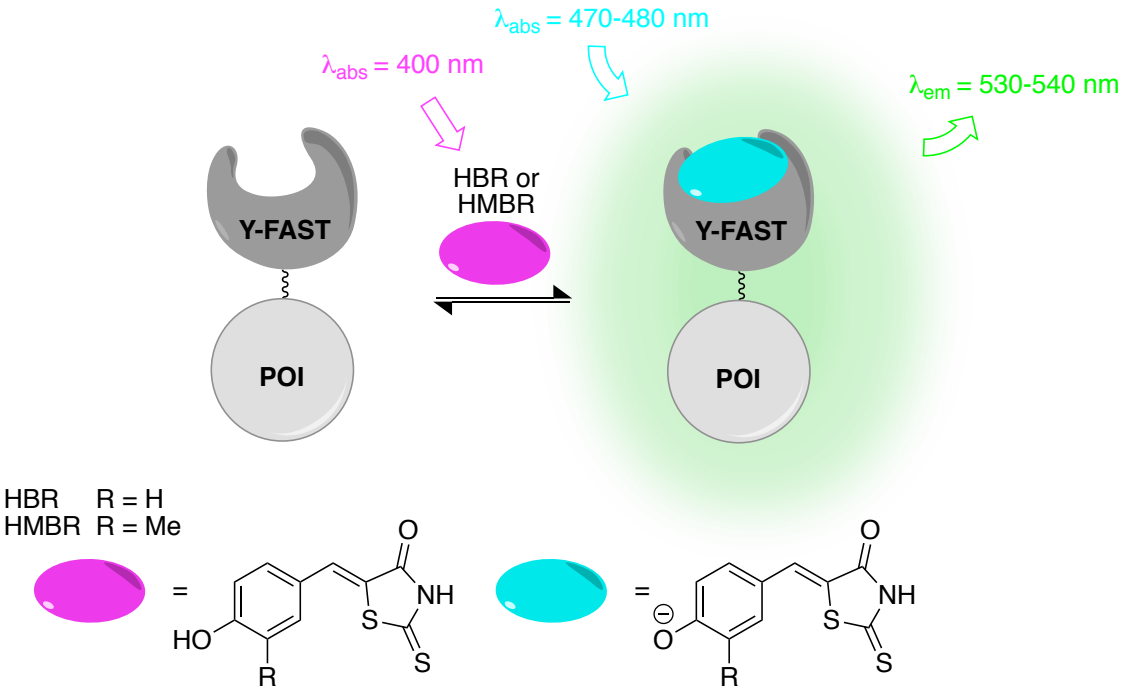


Figure I-15. Scheme of Yellow Fluorescence-Activating and absorption-Shifting Tag (Y-FAST). Ligands HBR and HMBR are non-fluorescent, but become fluorescent and their absorption red-shifts upon interaction with Y-FAST.

HBR and HMBR are themselves nonfluorescent, but fortuitously they fluoresce specifically and instantaneously when bound to Y-FAST. Its high selectivity is reliant on two spectroscopic changes for fluorogen activation: 1) binding of H(M)BR to Y-FAST results in a significant increase of fluorescence quantum yield and 2) binding induces a large absorption red-shift. Hence, unbound fluorogen does not contribute to the fluorescence signal, ensuring high signal-to-noise ratio. HeLa cells expressing various Y-FAST fusions were successfully labeled with 5 µM HMBR (*in vitro* Y-FAST/HMBR complex has a

quantum yield of 0.33), targeted to the nucleus, cell membrane, mitochondria, golgi and microtubules.

I.2.3.3 *In situ* formation of a cyanine dye

Work by our lab in using engineered variants of human cellular retinol binding protein II (hCRBPII) bound with retinal through an active site lysine to modulate absorption wavelength,[208, 209] inspired the design of a system in which the union of a protein and a nonfluorescent ligand produces a fluorescent protein.

Protein engineering of cellular retinoic acid binding protein II (CRABPII), a member of the intracellular lipid binding protein family (iLBPs), led to a genetically encoded red fluorescent protein tag. Advantageously, CRABPII is a small protein (15.6 kDa), with a relatively large binding cavity and is remarkably tolerant to mutations. CRABPII was engineered to bind a merocyanine retinal aldehyde (**MCRA**), to generate a cyanine dye *in situ*.[210] Reaction of the merocyanine aldehyde with the protein's active site lysine renders a fluorescent iminium (**Figure I-16**). Advantageously, the free aldehyde is nonfluorescent and blue-shifted relative to the iminium, rendering negligible background fluorescence from unbound merocyanine aldehyde. Formation of the iminium yields a resonating cation, leading to a bathochromic shift in absorption. Fortuitously, due to the extremely high degree of conjugation between the two nitrogen atoms, CRABPII/merocyanine complexes display high pK_a, allowing for measurements at physiological pH, and red-shifted emission up to 619 nm (upon excitation at 594 nm). Additionally, high quantum efficiencies were obtained (up to 39%). In a

proof-of-principle experiment, merocyanine aldehyde was able to instantaneously label *E. coli* cells transformed with CRABPII mutants. Unfortunately, translation to mammalian cells was not fruitful.

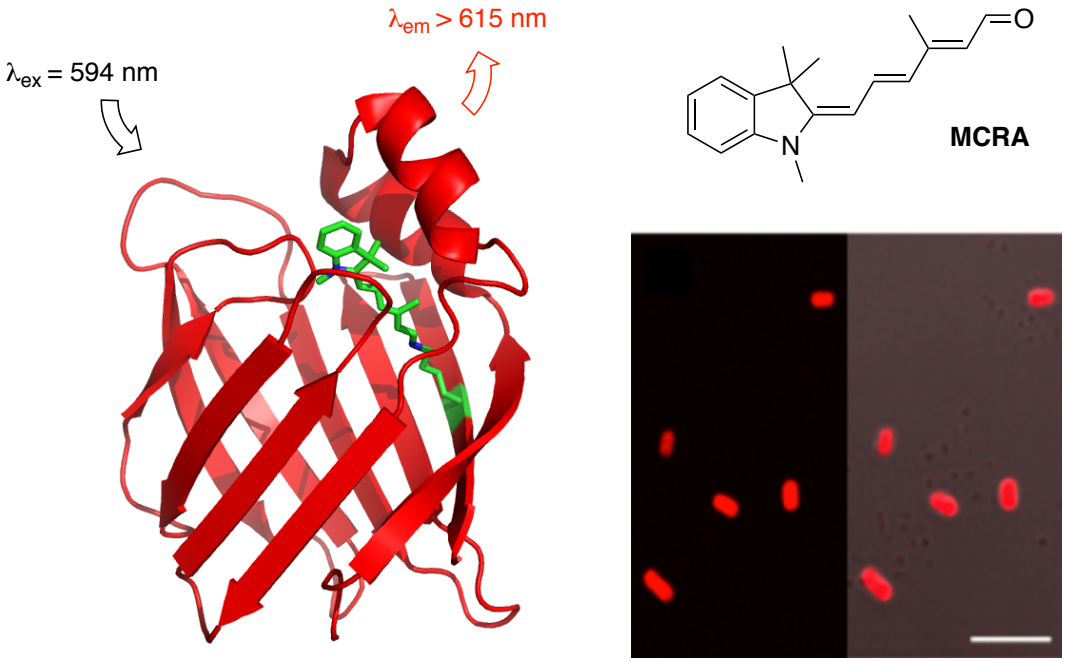


Figure I-16. Chemical structure of merocyanine retinal aldehyde (**MCRA**) and **MCRA** bound to an engineered CRABPII variant. Coordinates were obtained from PDB 3FEP. Fluorescence imaging of *E. coli* cells expressing CRABPII variant (left; overlay with brightfield on the right). Scale bar is 10 μm.

Drawing inspiration from the idea of using synthetic chromophores, the Arnold lab was able to generate a near-infrared FP based on the combination of Archaerhodopsin-3 (Arch), a proton-pumping microbial rhodopsin, and a longer merocyanine aldehyde (one double bond longer than **MCRA**).[211] Arch had also previously been engineered to bind its natural all-*trans*-retinal ligand, but brightness limits its usefulness in live cell imaging.[212-214] Because the previously described merocyanine retinal analogue was capable of binding to CRABPII, it was anticipated that the longer merocyanine aldehyde would be able to bind to

Arch; Additionally, in Arch, the natural ligand all-*trans*-retinal is covalently bound to a conserved lysine residue via a Schiff base (SB). Directed evolution of Arch around the merocyanine retinal allowed for the enhancement of chromophore affinity, as well as the augmentation of fluorescent brightness in the near-infrared. The best variant displayed excitation and emission maxima of 759 and 770 nm, with a quantum yield of 0.13. The potential of the merocyanine aldehyde bound Arch variants were shown in bacterial cells, where fluorescence is readily detected even under low power light excitation at 727 nm. Merocyanine dependent fluorescence could be detected in eukaryotic HEK cells, but the authors noted that the expression and stability of Arch should be optimized for eukaryotic imaging.

REFERENCES

REFERENCES

1. Dean, K.M. & Palmer, A.E. Advances in fluorescence labeling strategies for dynamic cellular imaging. *Nature Chemical Biology* **10**, 512-523 (2014).

2. Schaferling, M. The Art of Fluorescence Imaging with Chemical Sensors. *Angewandte Chemie-International Edition* **51**, 3532-3554 (2012).

3. Kowada, T., Maeda, H. & Kikuchi, K. BODIPY-based probes for the fluorescence imaging of biomolecules in living cells. *Chemical Society Reviews* **44**, 4953-4972 (2015).

4. Specht, E.A., Braselmann, E. & Palmer, A.E. A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging. *Annual Review of Physiology* **79**, 93-117 (2017).

5. Yuan, L., Lin, W.Y., Zheng, K.B., He, L.W. & Huang, W.M. Far-red to near infrared analyte-responsive fluorescent probes based on organic fluorophore platforms for fluorescence imaging. *Chemical Society Reviews* **42**, 622-661 (2013).

6. Nienhaus, K. & Nienhaus, G.U. Fluorescent proteins for live-cell imaging with super-resolution. *Chemical Society Reviews* **43**, 1088-1106 (2014).

7. Yang, Z.G., Cao, J.F., He, Y.X., Yang, J.H., Kim, T., Peng, X.J. & Kim, J.S. Macro-/micro-environment-sensitive chemosensing and biological imaging. *Chemical Society Reviews* **43**, 4563-4601 (2014).

8. Hilderbrand, S.A. & Weissleder, R. Near-infrared fluorescence: application to in vivo molecular imaging. *Current Opinion in Chemical Biology* **14**, 71-79 (2010).

9. Guo, Z.Q., Park, S., Yoon, J. & Shin, I. Recent progress in the development of near-infrared fluorescent probes for bioimaging applications. *Chemical Society Reviews* **43**, 16-29 (2014).

10. Luo, S.L., Zhang, E.L., Su, Y.P., Cheng, T.M. & Shi, C.M. A review of NIR dyes in cancer targeting and imaging. *Biomaterials* **32**, 7127-7138 (2011).

11. Lee, M.H., Kim, J.S. & Sessler, J.L. Small molecule-based ratiometric fluorescence probes for cations, anions, and biomolecules. *Chemical Society Reviews* **44**, 4185-4191 (2015).

12. Sevick-Muraca, E.M. Translation of Near-Infrared Fluorescence Imaging Technologies: Emerging Clinical Applications. *Annual Review of Medicine* **63**, 217-231 (2012).

13. Shaner, N.C., Steinbach, P.A. & Tsien, R.Y. A guide to choosing fluorescent proteins. *Nature Methods* **2**, 905-909 (2005).

14. Shu, X.K., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y. & Remington, S.J. Novel chromophores and buried charges control color in mFruits. *Biochemistry* **45**, 9639-9647 (2006).

15. Olenych, S.G., Claxton, N.S., Ottenberg, G.K. & Davidson, M.W. The fluorescent protein color palette. *Current Protocols in Cell Biology* **21.5**, 1-34 (2007).

16. Day, R.N. & Davidson, M.W. The fluorescent protein palette: tools for cellular imaging. *Chemical Society Reviews* **38**, 2887-2921 (2009).

17. Rizzo, M.A., Davidson, M.W. & Piston, D.W. Fluorescent protein tracking and detection: fluorescent protein structure and color variants. *Cold Spring Harbor Protocols* **2009**, 1-21 (2009).

18. Sample, V., Newman, R.H. & Zhang, J. The structure and function of fluorescent proteins. *Chemical Society Reviews* **38**, 2852-2864 (2009).

19. Remington, S.J. Green fluorescent protein: a perspective. *Protein Science* **20**, 1509-1519 (2011).

20. Wu, B., Piatkevich, K.D., Lionnet, T., Singer, R.H. & Verkhusha, V.V. Modern fluorescent proteins and imaging technologies to study gene expression, nuclear localization, and dynamics. *Current Opinion in Cell Biology* **23**, 310-317 (2011).

21. Stepanenko, O.V., Stepanenko, O.V., Kuznetsova, I.M., Verkhusha, V.V. & Turoverov, K.K. Beta-Barrel Scaffold of Fluorescent Proteins: Folding, Stability and Role in Chromophore Formation. *International Review of Cell and Molecular Biology* **302**, 221-278 (2013).

22. Mishin, A.S., Belousov, V.V., Solntsev, K.M. & Lukyanov, K.A. Novel uses of fluorescent proteins. *Current Opinion in Chemical Biology* **27**, 1-9 (2015).

23. Nienhaus, K. & Nienhaus, G.U. Chromophore photophysics and dynamics in fluorescent proteins of the GFP family. *Journal of Physics: Condensed Matter* **28**, 443001 (2016).

24. Shimomura, O. & Johnson, F.H. Properties of the bioluminescent protein aequorin. *Biochemistry* **8**, 3991-3997 (1969).

25. Shimomura, O., Johnson, F.H. & Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from Luminous Hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology* **59**, 223-& (1962).

26. Prasher, D., Mccann, R.O. & Cormier, M.J. Cloning and Expression of the Cdna Coding for Aequorin, a Bioluminescent Calcium-Binding Protein. *Biochemical and Biophysical Research Communications* **126**, 1259-1268 (1985).

27. 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* **111**, 229-233 (1992).

28. Wang, S. & Hazelrigg, T. Implications for bcd mRNA localization from spatial distribution of exu protein in Drosophila oogenesis. *Nature* **369**, 400-403 (1994).

29. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. & Prasher, D.C. Green Fluorescent Protein as a Marker for Gene-Expression. *Science* **263**, 802-805 (1994).

30. Subach, F.V. & Verkhusha, V.V. Chromophore Transformations in Red Fluorescent Proteins. *Chemical Reviews* **112**, 4308-4327 (2012).

31. Remington, S.J. Fluorescent proteins: maturation, photochemistry and photophysics. *Current Opinion in Structural Biology* **16**, 714-721 (2006).

32. Zimmer, M. GFP: from jellyfish to the Nobel prize and beyond. *Chemical Society Reviews* **38**, 2823-2832 (2009).

33. Piatkevich, K.D., Efremenko, E.N., Verkhusha, V.V. & Varfolomeev, D.D. Red fluorescent proteins and their properties. *Russian Chemical Reviews* **79**, 243-258 (2010).

34. Tsien, R.Y. The green fluorescent protein. *Annual Review of Biochemistry* **67**, 509-544 (1998).

35. Ormo, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y. & Remington, S.J. Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273**, 1392-1395 (1996).

36. Yang, F., Moss, L.G. & Phillips, G.N. The molecular structure of green fluorescent protein. *Nature Biotechnology* **14**, 1246-1251 (1996).

37. Heim, R., Prasher, D.C. & Tsien, R.Y. Wavelength Mutations and Posttranslational Autoxidation of Green Fluorescent Protein. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 12501-12504 (1994).

38. Cubitt, A.B., Heim, R., Adams, S.R., Boyd, A.E., Gross, L.A. & Tsien, R.Y. Understanding, Improving and Using Green Fluorescent Proteins. *Trends in Biochemical Sciences* **20**, 448-455 (1995).

39. Reid, B.G. & Flynn, G.C. Chromophore formation in green fluorescent protein. *Biochemistry* **36**, 6786-6791 (1997).

40. Bokman, S.H. & Ward, W.W. Renaturation of Aequorea Green-Fluorescent Protein. *Biochemical and Biophysical Research Communications* **101**, 1372-1380 (1981).

41. Baird, G.S., Zacharias, D.A. & Tsien, R.Y. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 11984-11989 (2000).

42. Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L. & Lukyanov, S.A. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotechnology* **17**, 969-973 (1999).

43. Verkhusha, V.V. & Lukyanov, K.A. The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins. *Nature Biotechnology* **22**, 289-296 (2004).

44. Gross, L.A., Baird, G.S., Hoffman, R.C., Baldridge, K.K. & Tsien, R.Y. The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 11990-11995 (2000).

45. Haddock, S.H., Mastroianni, N. & Christianson, L.M. A photoactivatable green-fluorescent protein from the phylum Ctenophora. *Proceedings of the Royal Society B: Biological Sciences* **277**, 1155-1160 (2010).

46. Deheyn, D.D., Kubokawa, K., McCarthy, J.K., Murakami, A., Porrachia, M., Rouse, G.W. & Holland, N.D. Endogenous green fluorescent protein (GFP) in amphioxus. *The Biological Bulletin* **213**, 95-100 (2007).

47. Pletnev, V.Z., Pletneva, N.V., Lukyanov, K.A., Souslova, E.A., Fradkov, A.F., Chudakov, D.M., Chepurnykh, T., Yampolsky, I.V., Wlodawer, A., Dauter, Z. & Pletnev, S. Structure of the red fluorescent protein from a lancelet (Branchiostoma lanceolatum): a novel GYG chromophore covalently bound to a nearby tyrosine. *Acta Crystallographica Section D* **69**, 1850-1860 (2013).

48. Shcherbakova, D.M., Subach, O.M. & Verkhusha, V.V. Red fluorescent proteins: advanced imaging applications and future design. *Angewandte Chemie-International Edition* **51**, 10724-10738 (2012).

49. Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., Zacharias, D.A. & Tsien, R.Y. A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7877-7882 (2002).

50. Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E. & Tsien, R.Y. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp red fluorescent protein. *Nature Biotechnology* **22**, 1567-1572 (2004).

51. Piatkevich, K.D. & Verkhusha, V.V. Advances in engineering of fluorescent proteins and photoactivatable proteins with red emission. *Current Opinion in Chemical Biology* **14**, 23-29 (2010).

52. Subach, F.V., Piatkevich, K.D. & Verkhusha, V.V. Directed molecular evolution to design advanced red fluorescent proteins. *Nature Methods* **8**, 1019-1026 (2011).

53. Shcherbakova, D.M., Baloban, M., Emelyanov, A.V., Brenowitz, M., Guo, P. & Verkhusha, V.V. Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. *Nature Communications* **7** (2016).

54. 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. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8362-8367 (1996).

55. Lossau, H., Kummer, A., Heinecke, R., PollingerDammer, F., Kompa, C., Bieser, G., Jonsson, T., Silva, C.M., Yang, M.M., Youvan, D.C. & MichelBeyerle, M.E. Time-resolved spectroscopy of wild-type and mutant Green Fluorescent Proteins reveals excited state deprotonation consistent with fluorophore-protein interactions. *Chemical Physics* **213**, 1-16 (1996).

56. Jung, G., Wiehler, J. & Zumbusch, A. The photophysics of green fluorescent protein: Influence of the key amino acids at positions 65, 203, and 222. *Biophysical Journal* **88**, 1932-1947 (2005).

57. Oltrogge, L.M., Wang, Q. & Boxer, S.G. Ground-state proton transfer kinetics in green fluorescent protein. *Biochemistry* **53**, 5947-5957 (2014).

58. Scharnagl, C. & Raupp-Kossmann, R.A. Solution pK(a) values of the green fluorescent protein chromophore from hybrid quantum-classical calculations. *Journal of Physical Chemistry B* **108**, 477-489 (2004).

59. Brejc, K., Sixma, T.K., Kitts, P.A., Kain, S.R., Tsien, R.Y., Ormo, M. & Remington, S.J. Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 2306-2311 (1997).

60. Heim, R., Cubitt, A.B. & Tsien, R.Y. Improved Green Fluorescence. *Nature* **373**, 663-664 (1995).

61. Hein, R. & Tsien, R.Y. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology* **6**, 178-182 (1996).

62. Kogure, T., Karasawa, S., Araki, T., Saito, K., Kinjo, M. & Miyawaki, A. A fluorescent variant of a protein from the stony coral Montipora facilitates dual-color single-laser fluorescence cross-correlation spectroscopy. *Nature Biotechnology* **24**, 577-581 (2006).

63. Henderson, J.N., Osborn, M.F., Koon, N., Gepshtein, R., Huppert, D. & Remington, S.J. Excited State Proton Transfer in the Red Fluorescent Protein mKeima. *Journal of the American Chemical Society* **131**, 13212-13213 (2009).

64. Violot, S., Carpentier, P., Blanchoin, L. & Bourgeois, D. Reverse pH-Dependence of Chromophore Protonation Explains the Large Stokes Shift of the Red Fluorescent Protein mKeima. *Journal of the American Chemical Society* **131**, 10356-10357 (2009).

65. Shcherbakova, D.M., Hink, M.A., Joosen, L., Gadella, T.W.J. & Verkhusha, V.V. An Orange Fluorescent Protein with a Large Stokes Shift for Single-Excitation Multicolor FCCS and FRET Imaging. *Journal of the American Chemical Society* **134**, 7913-7923 (2012).

66. Pletnev, S., Shcherbakova, D.M., Subach, O.M., Pletneva, N.V., Malashkevich, V.N., Almo, S.C., Dauter, Z. & Verkhusha, V.V. Orange Fluorescent Proteins: Structural Studies of LSSmOrange, PSmOrange and PSmOrange2. *Plos One* **9** (2014).

67. Fron, E., De Keersmaecker, H., Rocha, S., Baeten, Y., Lu, G., Uji-I, H., Van der Auweraer, M., Hofkens, J. & Mizuno, H. Mechanism Behind the Apparent Large Stokes Shift in LSSmOrange Investigated by Time-Resolved Spectroscopy. *Journal of Physical Chemistry B* **119**, 14880-14891 (2015).

68. Bergeler, M., Mizuno, H., Fron, E. & Harvey, J.N. QM/MM-Based Calculations of Absorption and Emission Spectra of LSSmOrange Variants. *Journal of Physical Chemistry B* **120**, 12454-12465 (2016).

69. Chu, J., Oh, Y., Sens, A., Ataie, N., Dana, H., Macklin, J.J., Laviv, T., Welf, E.S., Dean, K.M., Zhang, F.J., Kim, B.B., Tang, C.T., Hu, M., Baird, M.A., Davidson, M.W., Kay, M.A., Fiolka, R., Yasuda, R., Kim, D.S., Ng, H.L. & Lin, M.Z. A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo. *Nature Biotechnology* **34**, 760-767 (2016).

70. Shcherbo, D., Merzlyak, E.M., Chepurnykh, T.V., Fradkov, A.F., Ermakova, G.V., Solovieva, E.A., Lukyanov, K.A., Bogdanova, E.A., Zaraisky, A.G., Lukyanov, S. & Chudakov, D.M. Bright far-red fluorescent protein for whole-body imaging. *Nature Methods* **4**, 741-746 (2007).

71. Piatkevich, K.D., Hulit, J., Subach, O.M., Wu, B., Abdulla, A., Segall, J.E. & Verkhusha, V.V. Monomeric red fluorescent proteins with a large Stokes shift. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5369-5374 (2010).

72. Piatkevich, K.D., Malashkevich, V.N., Almo, S.C. & Verkhusha, V.V. Engineering ESPT Pathways Based on Structural Analysis of LSSmKate

Red Fluorescent Proteins with Large Stokes Shift. *Journal of the American Chemical Society* **132**, 10762-10770 (2010).

73. Randino, C., Moreno, M., Gelabert, R. & Lluch, J.M. Peek at the Potential Energy Surfaces of the LSSmKate1 and LSSmKate2 Proteins. *Journal of Physical Chemistry B* **116**, 14302-14310 (2012).

74. Chen, F.S., Zeng, Q., Zhuang, W. & Liang, W.Z. Characterizing the Structures, Spectra, and Energy Landscapes Involved in the Excited-State Proton Transfer Process of Red Fluorescent Protein LSSmKate1. *Journal of Physical Chemistry B* **120**, 9833-9842 (2016).

75. Piatkevich, K.D., English, B.P., Malashkevich, V.N., Xiao, H., Almo, S.C., Singer, R.H. & Verkhusha, V.V. Photoswitchable Red Fluorescent Protein with a Large Stokes Shift. *Chemistry & Biology* **21**, 1402-1414 (2014).

76. Yang, J., Wang, L., Yang, F., Luo, H.M., Xu, L.L., Lu, J.L., Zeng, S.Q. & Zhang, Z.H. mBeRFP, an Improved Large Stokes Shift Red Fluorescent Protein. *Plos One* **8** (2013).

77. Shcherbakova, D.M., Sengupta, P., Lippincott-Schwartz, J. & Verkhusha, V.V. Photocontrollable Fluorescent Proteins for Superresolution Imaging. *Annual Review of Biophysics* **43**, 303-329 (2014).

78. Hell, S.W., Sahl, S.J., Bates, M., Zhuang, X.W., Heintzmann, R., Booth, M.J., Bewersdorf, J., Shtengel, G., Hess, H., Tinnefeld, P., Honigmann, A., Jakobs, S., Testa, I., Cognet, L., Lounis, B., Ewers, H., Davis, S.J., Eggeling, C., Klenerman, D., Willig, K.I., Vicidomini, G., Castello, M., Diaspro, A. & Cordes, T. The 2015 super-resolution microscopy roadmap. *Journal of Physics D: Applied Physics* **48** (2015).

79. Hell, S.W. & Wichmann, J. Breaking the Diffraction Resolution Limit by Stimulated-Emission - Stimulated-Emission-Depletion Fluorescence Microscopy. *Optics Letters* **19**, 780-782 (1994).

80. Gustafsson, M.G.L. Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13081-13086 (2005).

81. Rust, M.J., Bates, M. & Zhuang, X.W. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793-795 (2006).

82. Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J. & Hess, H.F. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642-1645 (2006).

83. Hess, S.T., Girirajan, T.P.K. & Mason, M.D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical Journal* **91**, 4258-4272 (2006).

84. Shaner, N.C., Patterson, G.H. & Davidson, M.W. Advances in fluorescent protein technology. *Journal of Cell Science* **120**, 4247-4260 (2007).

85. Hayashi, S. & Toda, Y. A novel fluorescent protein purified from eel muscle. *Fisheries Science* **75**, 1461-1469 (2009).

86. Kumagai, A., Ando, R., Miyatake, H., Greimel, P., Kobayashi, T., Hirabayashi, Y., Shimogori, T. & Miyawaki, A. A Bilirubin-Inducible Fluorescent Protein from Eel Muscle. *Cell* **153**, 1602-1611 (2013).

87. Auldridge, M.E. & Forest, K.T. Bacterial phytochromes: More than meets the light. *Critical Reviews in Biochemistry and Molecular Biology* **46**, 67-88 (2011).

88. Piatkevich, K.D., Subach, F.V. & Verkhusha, V.V. Engineering of bacterial phytochromes for near-infrared imaging, sensing, and light-control in mammals. *Chemical Society Reviews* **42**, 3441-3452 (2013).

89. Shcherbakova, D.M., Baloban, M. & Verkhusha, V.V. Near-infrared fluorescent proteins engineered from bacterial phytochromes. *Current Opinion in Chemical Biology* **27**, 52-63 (2015).

90. Shcherbakova, D.M. & Verkhusha, V.V. Near-infrared fluorescent proteins for multicolor in vivo imaging. *Nature Methods* **10**, 751-754 (2013).

91. Kaberniuk, A.A., Shemetov, A.A. & Verkhusha, V.V. A bacterial phytochrome-based optogenetic system controllable with near-infrared light. *Nature Methods* **13**, 591-597 (2016).

92. Ghosh, S., Yu, C.L., Ferraro, D.J., Sudha, S., Pal, S.K., Schaefer, W.F., Gibson, D.T. & Ramaswamy, S. Blue protein with red fluorescence. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 11513-11518 (2016).

93. Gautier, A., Juillerat, A., Heinis, C., Correa, I.R., Jr., Kindermann, M., Beaufils, F. & Johnsson, K. An engineered protein tag for multiprotein labeling in living cells. *Chemistry & Biology* **15**, 128-136 (2008).

94. Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H. & Johnsson, K. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nature Biotechnology* **21**, 86-89 (2003).

95. Keppler, A., Pick, H., Arrivoli, C., Vogel, H. & Johnsson, K. Labeling of fusion proteins with synthetic fluorophores in live cells. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9955-9959 (2004).

96. Lukinavicius, G., Umezawa, K., Olivier, N., Honigmann, A., Yang, G., Plass, T., Mueller, V., Reymond, L., Correa, I.R., Jr., Luo, Z.G., Schultz, C., Lemke, E.A., Heppenstall, P., Eggeling, C., Manley, S. & Johnsson, K. A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nature Chemistry* **5**, 132-139 (2013).

97. Sun, X., Zhang, A., Baker, B., Sun, L., Howard, A., Buswell, J., Maurel, D., Masharina, A., Johnsson, K., Noren, C.J., Xu, M.Q. & Correa, I.R., Jr. Development of SNAP-tag fluorogenic probes for wash-free fluorescence imaging. *ChemBioChem* **12**, 2217-2226 (2011).

98. Hoehnel, S. & Lutolf, M.P. Capturing Cell-Cell Interactions via SNAP-tag and CLIP-tag Technology. *Bioconjugate Chemistry* **26**, 1678-1686 (2015).

99. Provost, C.R. & Sun, L. Fluorescent labeling of COS-7 expressing SNAP-tag fusion proteins for live cell imaging. *Journal of Visualized Experiments* (2010).

100. Calloway, N.T., Choob, M., Sanz, A., Sheetz, M.P., Miller, L.W. & Cornish, V.W. Optimized fluorescent trimethoprim derivatives for in vivo protein labeling. *ChemBioChem* **8**, 767-774 (2007).

101. Miller, L.W., Cai, Y.F., Sheetz, M.P. & Cornish, V.W. In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag. *Nature Methods* **2**, 255-257 (2005).

102. Liu, W., Li, F., Chen, X., Hou, J., Yi, L. & Wu, Y.W. A rapid and fluorogenic TMP-AcBOPDIPY probe for covalent labeling of proteins in live cells. *Journal of the American Chemical Society* **136**, 4468-4471 (2014).

103. Miller, L.W., Cai, Y., Sheetz, M.P. & Cornish, V.W. In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag. *Nature Methods* **2**, 255-257 (2005).

104. Jing, C. & Cornish, V.W. A fluorogenic TMP-tag for high signal-to-background intracellular live cell imaging. *ACS Chemical Biology* **8**, 1704-1712 (2013).

105. Los, G.V., Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohane, R.F., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D.H., Bulleit, R.F. & Wood, K.V. HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chemical Biology* **3**, 373-382 (2008).

106. Pries, F., Kingma, J., Krooshof, G.H., Jeronimusstratingh, C.M., Bruins, A.P. & Janssen, D.B. Histidine-289 Is Essential for Hydrolysis of the Alkyl-Enzyme Intermediate of Haloalkane Dehalogenase. *Journal of Biological Chemistry* **270**, 10405-10411 (1995).

107. Griffin, B.A., Adams, S.R. & Tsien, R.Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269-272 (1998).

108. Adams, S.R., Campbell, R.E., Gross, L.A., Martin, B.R., Walkup, G.K., Yao, Y., Llopis, J. & Tsien, R.Y. New biarsenical Ligands and tetracysteine motifs for protein labeling in vitro and in vivo: Synthesis and biological applications. *Journal of the American Chemical Society* **124**, 6063-6076 (2002).

109. Griffin, B.A., Adams, S.R., Jones, J. & Tsien, R.Y. Fluorescent labeling of recombinant proteins in living cells with FlAsH. *Method in Enzymology* **327**, 565-578 (2000).

110. Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R. & Stuber, D. Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Nature Biotechnology* **6**, 1321-1325 (1988).

111. Lata, S., Reichel, A., Brock, R., Tampe, R. & Piehler, J. High-affinity adaptors for switchable recognition of histidine-tagged proteins. *Journal of the American Chemical Society* **127**, 10205-10215 (2005).

112. Tinazli, A., Tang, J.L., Valiokas, R., Picuric, S., Lata, S., Piehler, J., Liedberg, B. & Tampe, R. High-affinity chelator thiols for switchable and

oriented immobilization of histidine-tagged proteins: A generic platform for protein chip technologies. *Chemistry-A European Journal* **11**, 5249-5259 (2005).

113. Wieneke, R., Raulf, A., Kollmannsperger, A., Heilemann, M. & Tampe, R. SLAP: Small Labeling Pair for Single-Molecule Super-Resolution Imaging. *Angewandte Chemie-International Edition* **54**, 10216-10219 (2015).

114. Nonaka, H., Tsukiji, S., Ojida, A. & Hamachi, I. Non-enzymatic covalent protein labeling using a reactive tag. *Journal of the American Chemical Society* **129**, 15777-15779 (2007).

115. Ojida, A., Honda, K., Shinmi, D., Kiyonaka, S., Mori, Y. & Hamachi, I. Oligo-Asp Tag/Zn(II) complex probe as a new pair for labeling and fluorescence imaging of proteins. *Journal of the American Chemical Society* **128**, 10452-10459 (2006).

116. Halo, T.L., Appelbaum, J., Hobert, E.M., Balkin, D.M. & Schepartz, A. Selective Recognition of Protein Tetraserine Motifs with a Cell-Permeable, Pro-fluorescent Bis-boronic Acid. *Journal of the American Chemical Society* **131**, 438-439 (2009).

117. Rashidian, M., Dozier, J.K., Lenevich, S. & Distefano, M.D. Selective labeling of polypeptides using protein farnesyltransferase via rapid oxime ligation. *Chemical Communications* **46**, 8998-9000 (2010).

118. Duckworth, B.P., Zhang, Z., Hosokawa, A. & Distefano, M.D. Selective labeling of proteins by using protein farnesyltransferase. *ChemBioChem* **8**, 98-105 (2007).

119. Chen, I., Howarth, M., Lin, W. & Ting, A.Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nature Methods* **2**, 99-104 (2005).

120. Slavoff, S.A., Chen, I., Choi, Y.A. & Ting, A.Y. Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species. *Journal of the American Chemical Society* **130**, 1160-1162 (2008).

121. Sueda, S., Yoneda, S. & Hayashi, H. Site-specific labeling of proteins by using biotin protein ligase conjugated with fluorophores. *ChemBioChem* **12**, 1367-1375 (2011).

122. Roux, K.J., Kim, D.I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of Cell Biology* **196**, 801-810 (2012).

123. Farazi, T.A., Waksman, G. & Gordon, J.I. The biology and enzymology of protein N-myristoylation. *The Journal of Biological Chemistry* **276**, 39501-39504 (2001).

124. Heal, W.P., Wickramasinghe, S.R., Bowyer, P.W., Holder, A.A., Smith, D.F., Leatherbarrow, R.J. & Tate, E.W. Site-specific N-terminal labelling of proteins in vitro and in vivo using N-myristoyl transferase and bioorthogonal ligation chemistry. *Chemical Communications*, 480-482 (2008).

125. Heal, W.P., Wickramasinghe, S.R., Leatherbarrow, R.J. & Tate, E.W. N-Myristoyl transferase-mediated protein labelling in vivo. *Organic & Biomolecular Chemistry* **6**, 2308-2315 (2008).

126. Roeser, D., Preusser-Kunze, A., Schmidt, B., Gasow, K., Wittmann, J.G., Dierks, T., von Figura, K. & Rudolph, M.G. A general binding mechanism for all human sulfatases by the formylglycine-generating enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 81-86 (2006).

127. Wu, P., Shui, W., Carlson, B.L., Hu, N., Rabuka, D., Lee, J. & Bertozzi, C.R. Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 3000-3005 (2009).

128. Rabuka, D., Rush, J.S., deHart, G.W., Wu, P. & Bertozzi, C.R. Site-specific chemical protein conjugation using genetically encoded aldehyde tags. *Nature Protocols* **7**, 1052-1067 (2012).

129. Mao, H., Hart, S.A., Schink, A. & Pollok, B.A. Sortase-mediated protein ligation: a new method for protein engineering. *Journal of the American Chemical Society* **126**, 2670-2671 (2004).

130. Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E. & Ploegh, H.L. Sortagging: a versatile method for protein labeling. *Nature Chemical Biology* **3**, 707-708 (2007).

131. Proft, T. Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation. *Biotechnology Letters* **32**, 1-10 (2010).

132. Ritzefeld, M. Sortagging: a robust and efficient chemoenzymatic ligation strategy. *Chemistry-A European Journal* **20**, 8516-8529 (2014).

133. Cronan, J.E., Zhao, X. & Jiang, Y.F. Function, attachment and synthesis of lipoic acid in Escherichia coli. *Advances in Microbial Physiology* **50**, 103-146 (2005).

134. Fernandez-Suarez, M., Baruah, H., Martinez-Hernandez, L., Xie, K.T., Baskin, J.M., Bertozzi, C.R. & Ting, A.Y. Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nature Biotechnology* **25**, 1483-1487 (2007).

135. Uttamapinant, C., White, K.A., Baruah, H., Thompson, S., Fernandez-Suarez, M., Puthenveetil, S. & Ting, A.Y. A fluorophore ligase for site-specific protein labeling inside living cells. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 10914-10919 (2010).

136. Klymchenko, A.S. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Accounts of Chemical Research* **50**, 366-375 (2017).

137. Hori, Y. & Kikuchi, K. Protein labeling with fluorogenic probes for no-wash live-cell imaging of proteins. *Current Opinion in Chemical Biology* **17**, 644-650 (2013).

138. Baranczak, A., Connelly, S., Liu, Y., Choi, S., Grimster, N.P., Powers, E.T., Wilson, I.A. & Kelly, J.W. Fluorogenic small molecules requiring reaction with a specific protein to create a fluorescent conjugate for biological imaging-what we know and what we need to learn. *Biopolymers* **101**, 484-495 (2014).

139. Bruchez, M.P. Dark dyes-bright complexes: fluorogenic protein labeling. *Current Opinion in Chemical Biology* **27**, 18-23 (2015).

140. Chen, Y.C., Tsao, K. & Keillor, J.W. Fluorogenic protein labelling: a review of photophysical quench mechanisms and principles of fluorogen design. *Canadian Journal of Chemistry* **93**, 389-398 (2015).

141. Machado, V.G., Stock, R.I. & Reichardt, C. Pyridinium N-Phenolate Betaine Dyes. *Chemical Reviews* **114**, 10429-10475 (2014).

142. Klymchenko, A.S. & Mely, Y. Fluorescent Environment-Sensitive Dyes as Reporters of Biomolecular Interactions. *Progress in Molecular Biology and Translational Science* **113**, 35-58 (2013).

143. Marini, A., Munoz-Losa, A., Biancardi, A. & Mennucci, B. What is Solvatochromism? *Journal of Physical Chemistry B* **114**, 17128-17135 (2010).

144. Loving, G.S., Sainlos, M. & Imperiali, B. Monitoring protein interactions and dynamics with solvatochromic fluorophores. *Trends in Biotechnology* **28**, 73-83 (2010).

145. Kikuchi, K., Takakusa, H. & Nagano, T. Recent advances in the design of small molecule-based FRET sensors for cell biology. *TrAC Trends in Analytical Chemistry* **23**, 407-415 (2004).

146. Yuan, L., Lin, W.Y., Zheng, K.B. & Zhu, S.S. FRET-Based Small-Molecule Fluorescent Probes: Rational Design and Bioimaging Applications. *Accounts of Chemical Research* **46**, 1462-1473 (2013).

147. Zhang, J., Campbell, R.E., Ting, A.Y. & Tsien, R.Y. Creating new fluorescent probes for cell biology. *Nature Reviews Molecular Cell Biology* **3**, 906-918 (2002).

148. Jiao, G.S., Thoresen, L.H. & Burgess, K. Fluorescent, through-bond energy transfer cassettes for Labeling multiple biological molecules in one experiment. *Journal of the American Chemical Society* **125**, 14668-14669 (2003).

149. Jiao, G.S., Thoresen, L.H., Kim, T.G., Haaland, W.C., Gao, F., Topp, M.R., Hochstrasser, R.M., Metzker, M.L. & Burgess, K. Syntheses, photophysical properties, and application of through-bond energy-transfer cassettes for biotechnology. *Chemistry-A European Journal* **12**, 7816-7826 (2006).

150. Lin, W.Y., Yuan, L., Cao, Z.M., Feng, Y.M. & Song, J.Z. Through-Bond Energy Transfer Cassettes with Minimal Spectral Overlap between the Donor Emission and Acceptor Absorption: Coumarin-Rhodamine Dyads with Large Pseudo-Stokes Shifts and Emission Shifts. *Angewandte Chemie-International Edition* **49**, 375-379 (2010).

151. Speiser, S. Photophysics and mechanisms of intramolecular electronic energy transfer in bichromophoric molecular systems: Solution and supersonic jet studies. *Chemical Reviews* **96**, 1953-1976 (1996).

152. Fox, M.A. Photoinduced Electron-Transfer. *Photochemistry and Photobiology* **52**, 617-627 (1990).

153. Shieh, P., Hangauer, M.J. & Bertozzi, C.R. Fluorogenic azidofluoresceins for biological imaging. *Journal of the American Chemical Society* **134**, 17428-17431 (2012).

154. Shieh, P., Siegrist, M.S., Cullen, A.J. & Bertozzi, C.R. Imaging bacterial peptidoglycan with near-infrared fluorogenic azide probes. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 5456-5461 (2014).

155. Ueno, T., Urano, Y., Setsukinai, K., Takakusa, H., Kojima, H., Kikuchi, K., Ohkubo, K., Fukuzumi, S. & Nagano, T. Rational principles for modulating fluorescence properties of fluorescein. *Journal of the American Chemical Society* **126**, 14079-14085 (2004).

156. Chen, Y., Clouthier, C.M., Tsao, K., Strmiskova, M., Lachance, H. & Keillor, J.W. Coumarin-based fluorogenic probes for no-wash protein labeling. *Angewandte Chemie-International Edition* **53**, 13785-13788 (2014).

157. Guy, J., Castonguay, R., Campos-Reales Pineda, N.B., Jacquier, V., Caron, K., Michnick, S.W. & Keillor, J.W. De novo helical peptides as target sequences for a specific, fluorogenic protein labelling strategy. *Molecular Biosystems* **6**, 976-987 (2010).

158. Caron, K., Lachapelle, V. & Keillor, J.W. Dramatic increase of quench efficiency in "spacerless" dimaleimide fluorogens. *Organic & Biomolecular Chemistry* **9**, 185-197 (2011).

159. Girouard, S., Houle, M.H., Grandbois, A., Keillor, J.W. & Michnick, S.W. Synthesis and characterization of dimaleimide fluorogens designed for specific labeling of proteins. *Journal of the American Chemical Society* **127**, 559-566 (2005).

160. Prifti, E., Reymond, L., Umebayashi, M., Hovius, R., Riezman, H. & Johnsson, K. A fluorogenic probe for SNAP-tagged plasma membrane proteins based on the solvatochromic molecule Nile Red. *ACS Chemical Biology* **9**, 606-612 (2014).

161. Biczok, L., Berces, T. & Inoue, H. Effects of molecular structure and hydrogen bonding on the radiationless deactivation of singlet excited fluorenone derivatives. *Journal of Physical Chemistry A* **103**, 3837-3842 (1999).

162. Cser, A., Nagy, K. & Biczok, L. Fluorescence lifetime of Nile Red as a probe for the hydrogen bonding strength with its microenvironment. *Chemical Physics Letters* **360**, 473-478 (2002).

163. Jose, J. & Burgess, K. Benzophenoxazine-based fluorescent dyes for labeling biomolecules. *Tetrahedron* **62**, 11021-11037 (2006).

164. Yatsuhashi, T. & Inoue, H. Molecular mechanism of radiationless deactivation of aminoanthraquinones through intermolecular hydrogen-bonding interaction with alcohols and hydroperoxides. *Journal of Physical Chemistry A* **101**, 8166-8173 (1997).

165. Diaz, G., Melis, M., Batetta, B., Angius, F. & Falchi, A.M. Hydrophobic characterization of intracellular lipids in situ by Nile Red red/yellow emission ratio. *Micron* **39**, 819-824 (2008).

166. Greenspan, P., Mayer, E.P. & Fowler, S.D. Nile red: a selective fluorescent stain for intracellular lipid droplets. *The Journal of Cell Biology* **100**, 965-973 (1985).

167. Gong, Y.K. & Pan, L.F. Recent advances in bioorthogonal reactions for site-specific protein labeling and engineering. *Tetrahedron Letters* **56**, 2123-2132 (2015).

168. Lang, K. & Chin, J.W. Bioorthogonal Reactions for Labeling Proteins. *ACS Chemical Biology* **9**, 16-20 (2014).

169. Li, J. & Chen, P.R. Development and application of bond cleavage reactions in bioorthogonal chemistry. *Nature Chemical Biology* **12**, 129-137 (2016).

170. Lim, R.K.V. & Lin, Q. Photoinducible Bioorthogonal Chemistry: A Spatiotemporally Controllable Tool to Visualize and Perturb Proteins in Live Cells. *Accounts of Chemical Research* **44**, 828-839 (2011).

171. Shieh, P. & Bertozzi, C.R. Design strategies for bioorthogonal smart probes. *Organic & Biomolecular Chemistry* **12**, 9307-9320 (2014).

172. Davis, L. & Chin, J.W. Designer proteins: applications of genetic code expansion in cell biology. *Nature Reviews Molecular Cell Biology* **13**, 168-182 (2012).

173. Lang, K. & Chin, J.W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chemical Reviews* **114**, 4764-4806 (2014).

174. Liu, C.C. & Schultz, P.G. Adding New Chemistries to the Genetic Code. *Annual Review of Biochemistry* **79**, 413-444 (2010).

175. Tsao, M.L., Tian, F. & Schultz, P.G. Selective staudinger modification of proteins containing p-azidophenylalanine. *ChemBioChem* **6**, 2147-2149 (2005).

176. Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K. & Yokoyama, S. Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode N(epsilon)-(o-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification. *Chemistry & Biology* **15**, 1187-1197 (2008).

177. Rostovtsev, V.V., Green, L.G., Fokin, V.V. & Sharpless, K.B. A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angewandte Chemie-International Edition* **41**, 2596-2599 (2002).

178. Tornoe, C.W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *Journal of Organic Chemistry* **67**, 3057-3064 (2002).

179. Sivakumar, K., Xie, F., Cash, B.M., Long, S., Barnhill, H.N. & Wang, Q. A fluorogenic 1,3-dipolar cycloaddition reaction of 3-azidocoumarins and acetylenes. *Organic Letters* **6**, 4603-4606 (2004).

180. Hong, V., Steinmetz, N.F., Manchester, M. & Finn, M.G. Labeling live cells by copper-catalyzed alkyne--azide click chemistry. *Bioconjugate Chemistry* **21**, 1912-1916 (2010).

181. Wolbers, F., ter Braak, P., Le Gac, S., Luttge, R., Andersson, H., Vermes, I. & van den Berg, A. Viability study of HL60 cells in contact with commonly used microchip materials. *Electrophoresis* **27**, 5073-5080 (2006).

182. Besanceney-Webler, C., Jiang, H., Zheng, T., Feng, L., Soriano del Amo, D., Wang, W., Klivansky, L.M., Marlow, F.L., Liu, Y. & Wu, P. Increasing the efficacy of bioorthogonal click reactions for bioconjugation: a comparative study. *Angewandte Chemie-International Edition* **50**, 8051-8056 (2011).

183. Presolski, S.I., Hong, V., Cho, S.H. & Finn, M.G. Tailored ligand acceleration of the Cu-catalyzed azide-alkyne cycloaddition reaction: practical and mechanistic implications. *Journal of the American Chemical Society* **132**, 14570-14576 (2010).

184. Rodionov, V.O., Presolski, S.I., Diaz, D.D., Fokin, V.V. & Finn, M.G. Ligand-accelerated Cu-catalyzed azide-alkyne cycloaddition: a mechanistic report. *Journal of the American Chemical Society* **129**, 12705-12712 (2007).

185. Agard, N.J., Baskin, J.M., Prescher, J.A., Lo, A. & Bertozzi, C.R. A comparative study of bioorthogonal reactions with azides. *ACS Chemical Biology* **1**, 644-648 (2006).

186. Agard, N.J., Prescher, J.A. & Bertozzi, C.R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *Journal of the American Chemical Society* **126**, 15046-15047 (2004).

187. Sletten, E.M. & Bertozzi, C.R. From mechanism to mouse: a tale of two bioorthogonal reactions. *Accounts of Chemical Research* **44**, 666-676 (2011).

188. Johnson, J.A., Lu, Y.Y., Van Deventer, J.A. & Tirrell, D.A. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Current Opinion in Chemical Biology* **14**, 774-780 (2010).

189. Ngo, J.T. & Tirrell, D.A. Noncanonical amino acids in the interrogation of cellular protein synthesis. *Accounts of Chemical Research* **44**, 677-685 (2011).

190. Song, W., Wang, Y., Qu, J. & Lin, Q. Selective functionalization of a genetically encoded alkene-containing protein via "photoclick chemistry" in bacterial cells. *Journal of the American Chemical Society* **130**, 9654-9655 (2008).

191. Song, W., Wang, Y., Qu, J., Madden, M.M. & Lin, Q. A photoinducible 1,3-dipolar cycloaddition reaction for rapid, selective modification of tetrazole-containing proteins. *Angewandte Chemie-International Edition* **47**, 2832-2835 (2008).

192. Yu, Z., Ohulchanskyy, T.Y., An, P., Prasad, P.N. & Lin, Q. Fluorogenic, two-photon-triggered photoclick chemistry in live mammalian cells. *Journal of the American Chemical Society* **135**, 16766-16769 (2013).

193. Yu, Z., Pan, Y., Wang, Z., Wang, J. & Lin, Q. Genetically encoded cyclopropene directs rapid, photoclick-chemistry-mediated protein labeling in mammalian cells. *Angewandte Chemie-International Edition* **51**, 10600-10604 (2012).

194. Blackman, M.L., Royzen, M. & Fox, J.M. Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *Journal of the American Chemical Society* **130**, 13518-13519 (2008).

195. Devaraj, N.K., Weissleder, R. & Hilderbrand, S.A. Tetrazine-based cycloadditions: application to pretargeted live cell imaging. *Bioconjugate Chemistry* **19**, 2297-2299 (2008).

196. Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D.J., Blackman, M.L., Fox, J.M. & Chin, J.W. Genetic Encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. *Journal of the American Chemical Society* **134**, 10317-10320 (2012).

197. Szent-Gyorgyi, C., Schmidt, B.A., Creeger, Y., Fisher, G.W., Zakel, K.L., Adler, S., Fitzpatrick, J.A.J., Woolford, C.A., Yan, Q., Vasilev, K.V., Berget, P.B., Bruchez, M.P., Jarvik, J.W. & Waggoner, A. Fluorogen-activating single-chain antibodies for imaging cell surface proteins. *Nature Biotechnology* **26**, 235-240 (2008).

198. Telmer, C.A., Verma, R., Teng, H.B., Andreko, S., Law, L. & Bruchez, M.P. Rapid, Specific, No-wash, Far-red Fluorogen Activation in Subcellular Compartments by Targeted Fluorogen Activating Proteins. *ACS Chemical Biology* **10**, 1239-1246 (2015).

199. Kumauchi, M., Hara, M.T., Stalcup, P., Xie, A.H. & Hoff, W.D. Identification of six new photoactive yellow proteins - Diversity and structure-function relationships in a bacterial blue light photoreceptor. *Photochemistry and Photobiology* **84**, 956-969 (2008).

200. Kyndt, J.A., Meyer, T.E., Cusanovich, M.A. & Van Beeumen, J.J. Characterization of a bacterial tyrosine ammonia lyase, a biosynthetic enzyme for the photoactive yellow protein. *FEBS Letters* **512**, 240-244 (2002).

201. 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. *Photochemical & Photobiological Sciences* **6**, 571-579 (2007).

202. Hori, Y., Ueno, H., Mizukami, S. & Kikuchi, K. Photoactive Yellow Protein-Based Protein Labeling System with Turn-On Fluorescence Intensity. *Journal of the American Chemical Society* **131**, 16610-16611 (2009).

203. Hori, Y., Nakaki, K., Sato, M., Mizukami, S. & Kikuchi, K. Development of Protein-Labeling Probes with a Redesigned Fluorogenic Switch Based on Intramolecular Association for No-Wash Live-Cell Imaging. *Angewandte Chemie-International Edition* **51**, 5611-5614 (2012).

204. Hori, Y., Norinobu, T., Sato, M., Arita, K., Shirakawa, M. & Kikuchi, K. Development of Fluorogenic Probes for Quick No-Wash Live-Cell Imaging of Intracellular Proteins. *Journal of the American Chemical Society* **135**, 12360-12365 (2013).

205. Hori, Y., Hirayama, S., Sato, M. & Kikuchi, K. Redesign of a Fluorogenic Labeling System To Improve Surface Charge, Brightness, and Binding Kinetics for Imaging the Functional Localization of Bromodomains. *Angewandte Chemie-International Edition* **54**, 14368-14371 (2015).

206. Kamikawa, Y., Hori, Y., Yamashita, K., Jin, L., Hirayama, S., Standley, D.M. & Kikuchi, K. Design of a protein tag and fluorogenic probe with modular structure for live-cell imaging of intracellular proteins. *Chemical Science* **7**, 308-314 (2016).

207. Plamont, M.A., Billon-Denis, E., Maurin, S., Gauron, C., Pimenta, F.M., Specht, C.G., Shi, J., Querard, J., Pan, B.Y., Rossignol, J., Morellet, N., Volovitch, M., Lescop, E., Chen, Y., Triller, A., Vriz, S., Le Saux, T., Jullien, L. & Gautier, A. Small fluorescence-activating and absorption-shifting tag for tunable protein imaging in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 497-502 (2016).

208. Wang, W.J., Geiger, J.H. & Borhan, B. The photochemical determinants of color vision. *Bioessays* **36**, 65-74 (2014).

209. Wang, W.J., Nossoni, Z., Berbasova, T., Watson, C.T., Yapici, I., Lee, K.S.S., Vasileiou, C., Geiger, J.H. & Borhan, B. Tuning the Electronic Absorption of Protein-Embedded All-trans-Retinal. *Science* **338**, 1340-1343 (2012).

210. 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).

211. Herwig, L., Rice, A.J., Bedbrook, C.N., Zhang, R.J.K., Lignell, A., Cahn, J.K.B., Renata, H., Dodani, S.C., Cho, I., Cai, L., Gradinaru, V. & Arnold, F.H. Directed Evolution of a Bright Near-Infrared Fluorescent Rhodopsin Using a Synthetic Chromophore. *Cell Chemical Biology* **24**, 415-425 (2017).

212. Engqvist, M.K.M., McIsaac, R.S., Dollinger, P., Flytzanis, N.C., Abrams, M., Schor, S. & Arnold, F.H. Directed Evolution of Gloeobacter violaceus Rhodopsin Spectral Properties. *Journal of Molecular Biology* **427**, 205-220 (2015).

213. Flytzanis, N.C., Bedbrook, C.N., Chiu, H., Engqvist, M.K.M., Xiao, C., Chan, K.Y., Sternberg, P.W., Arnold, F.H. & Gradinaru, V. Archaerhodopsin variants with enhanced voltage-sensitive fluorescence in mammalian and Caenorhabditis elegans neurons. *Nature Communications* **5** (2014).

214. McIsaac, R.S., Engqvist, M.K.M., Wannier, T., Rosenthal, A.Z., Herwig, L., Flytzanis, N.C., Imasheva, E.S., Lanyi, J.K., Balashov, S.P., Gradinaru, V. & Arnold, F.H. Directed evolution of a far-red fluorescent rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13034-13039 (2014).

CHAPTER II: MODULATION OF ABSORPTION AND EMISSION VIA A PROTEIN EMBEDDED SOLVATOCHROMIC FLUOROPHORE

Our overarching goal is to develop fluorescent proteins that span the entire visible spectra, to be used when conventional fluorescent proteins are inadequate. In our system, fluorescence is activated upon coupling of the protein and ligand, such that temporal control can be achieved, whereas intrinsically fluorescent proteins are constitutively on; it is only upon ligand addition and irradiation that our complex will become fluorescent. Additionally, our system does not require oxygen and can therefore find potential in obligate anaerobes.

Our lab has demonstrated the ability to effectively control the absorption profile of conjugated polyenes;[1-3] however, in this case we envision using solvatochromic fluorophores, in hopes that emission wavelength can be regulated with the same level of control. The term solvatochromism is used to describe the marked change in wavelength, and sometimes intensity, of an absorption and/or emission band that results from changing the polarity of the media.[4, 5] It is proposed that this effect can be mimicked inside of a protein cavity by the introduction of specific interactions between the bound fluorophore and amino acid side chains.

As seen in Chapter I, a variety of methods can be employed in developing fluorogenic (or “turn-on”) proteins. Our lab has been successful in labeling expressed variants of human cellular retinol binding protein II (hCRBPII) in living cells with merocyanine aldehyde in multiple cell lines, localized to a number of

intracellular targets (allowing for spatial control). Because fluorescence is activated upon coupling of the protein and ligand, temporal control can also be achieved. Advantageously, this system does not require oxygen and can therefore find potential in obligate anaerobes. Additionally, because the protonated Schiff base (PSB), formed as a result of protein/ligand condensation, is inherently sensitive to pH, innumerable applications can be envisioned such as multi-color imaging and *in vivo* pH sensing, among others.

Specificity is imperative in engineering a practical and efficient fluorogenic system. As will first be described, the most deleterious drawback of merocyanine aldehyde in imaging live cells is the background fluorescence that arises from non-specific labeling (i.e. from iminium formation with nonspecific proteins), due to the high pK_a and unchanging excitation and emission profiles of merocyanine aldehyde bound hCRBPII complexes.

To exclude the observation of background fluorescence due to nonspecific binding, we focused on the coupling of aldehydic derivatives of known solvatochromic fluorophores with engineered variants of hCRBPII. This offers two advantages: 1) Emission of solvatochromic fluorophores is variable based on the polarity of their environment. Thus, it should be possible to develop probes of many different colors. 2) We also should be able to alter the hCRBPII/fluorophore emission away from that of nonspecific labeling. This would result in a probe that yields little background fluorescence.

II.1 Application of merocyanine aldehyde with hCRBPII mutants for *in vivo* imaging

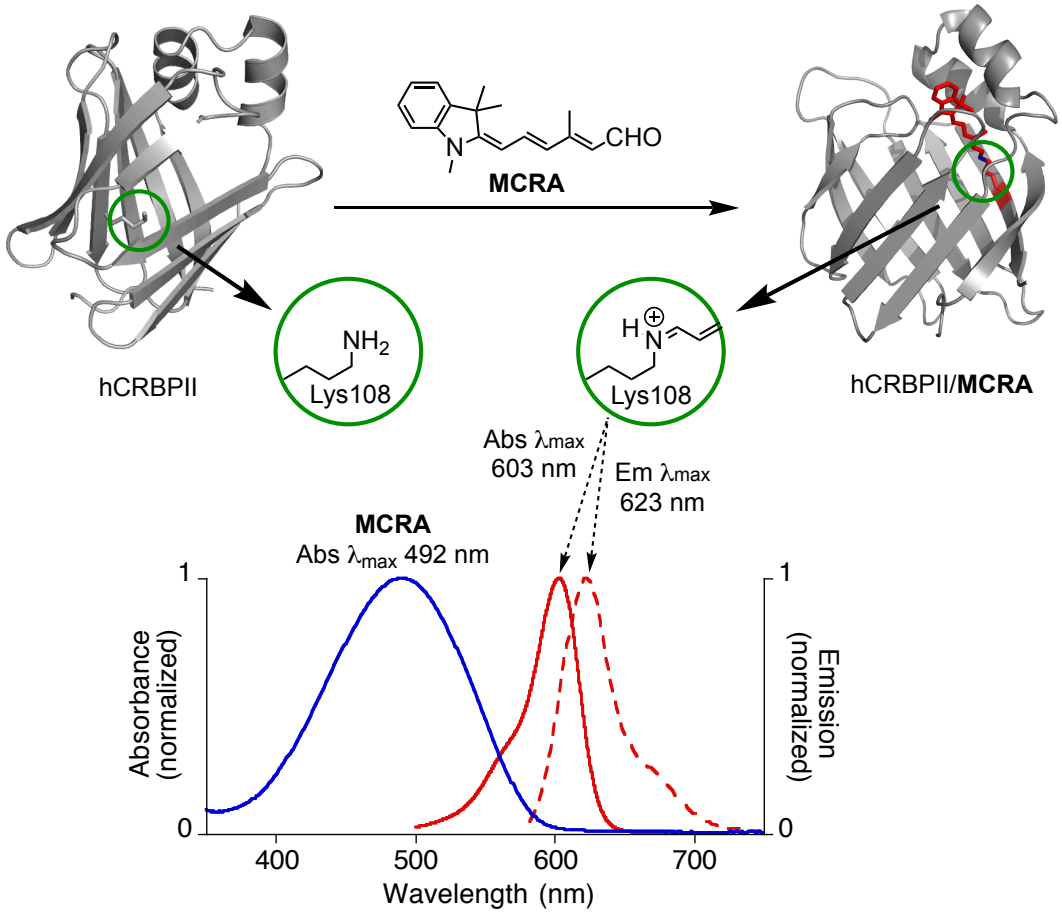


Figure II-1. Reaction and spectroscopic properties of merocyanine aldehyde (**MCRA**) with hCRBPII. Coordinates were obtained from PDB 3FEP.

Our lab has recently described the spectroscopic characterization of merocyanine retinal analog (**MCRA**), a weakly fluorescent aldehyde, which upon imine formation becomes fluorophoric (turn-on fluorescence).[6] The complex also exhibits a substantial bathochromic shift in absorption wavelength, rendering any background fluorescence from unbound reagent obsolete (**Figure II-1**). Dr. Ipek Yapici was successful in conducting preliminary studies with **MCRA** and cellular retinoic acid binding protein II (CRABPII). These studies highlighted the

advantage of using the 'turn-on' approach in bacteria. Nonetheless, expression of CRABPII in mammalian cell cultures was not fruitful, most probably as a result of protein misfolding.

On the other hand, hCRBPII variants have shown no deleterious problem in expression or folding, producing highly fluorescent pigments upon complexation with **MCRA**. Dr. Wenjing Wang and Dr. Tetyana Berbasova pioneered the efforts to couple hCRBPII with **MCRA**, to produce fast binding mutants that are complete within one minute (it should be noted that all *in vitro* data detailed in this section was completed by Dr. Wang and Dr. Berbasova). With promising *in vitro* data for hCRBPII/**MCRA** complexes, we sought to prove the system's expediency in live cells.

Preliminary screening of hCRBPII mutants coupled with **MCRA** yielded six potential candidates for further optimization (**Table II-1**, entries 3-8). Q108K:K40L (KL) and Q108K:K40L:R58F bind **MCRA** relatively fast with half-life ($t_{1/2}$) maturation times of 5.7 min and 10.0 min, respectively (**Table II-1**, entries 3 and 4). With respectable half-lives, we sought to increase the complex's quantum yield. In previous wavelength regulation studies with retinal, it was shown that R58, located at the entrance of the protein cavity, was important for encapsulating the binding cavity; upon substitution of R58 with phenylalanine, quantum efficiency did not change (18% for KL and 16% for KL:R58F). Fortuitously though, T51V doubled the quantum efficiency of the hCRBPII/**MCRA** complex (**Table II-1**, entries 5 and 6), yielding a two-fold increase in brightness

relative to KL:R58F. Unfortunately, the T51V addition led to substantially slower formation of the iminium.

Table II-1. Spectroscopic properties of hCRBPII mutants coupled with merocyanine retinal analog (**MCRA**).

Entry	Mutant	λ_{max} (abs/em)	Φ^a	ε (10^{-3} $M^{-1}cm^{-1}$)	$t_{1/2}{}^b$ (min)
1	mRFP	584/607	0.25	50	-
2	mRaspberry	598/625	0.15	86	-
3	Q108K:K40L (KL)	600/619	0.18	162	5.7
4	KL:R58F	600/619	0.16	165	10.0
5	KL:T51V (KLV)	598/616	0.37	166	140.3
6	KLV:R58F	603/623	0.36	161	76.6
7	KLV:T53C:R58W:T29L:Y19W:A33W:Q4F	600/619	0.19	190	10.4
8	KLV:T53C:R58W:T29L:A33W:Q4F	605/625	0.19	170	2.7
9	KLV:R58F:L117E	593/615	0.36	157	16.1
10	KLV:R58F:L117E:Q4F	594/615	0.32	161	14.4
11	KLV:T53C:R58W:T29L:A33W:Q4F:L117E	590/611	0.37	166	0.15
12	KLV:T53C:R58W:T29L:A33W:Q4F:L117D	590/613	0.30	197	0.16

[a]Quantum yields were determined based on two fluorescent standards (Oxazine-1 and Oxazine-170). [b]Half-lives are based on rate constants obtained from fitting the data to second order kinetics.

We then screened hCRBPII mutants previously, produced in our lab, for faster kinetics of iminium formation with **MCRA**, retaining the T51V mutation. Iminium formation was faster in a few mutants (**Table II-1**, entries 7 and 8). Unfortunately, the quantum yield was significantly decreased, as compared to KLV:R58F. Regardless of the low quantum yield, the hCRBPII/**MCRA** complexes display brightness (the product of quantum yield and extinction coefficient) higher than FPs emitting in the same wavelength regime (**Table II-1**, entries 1 and 2).

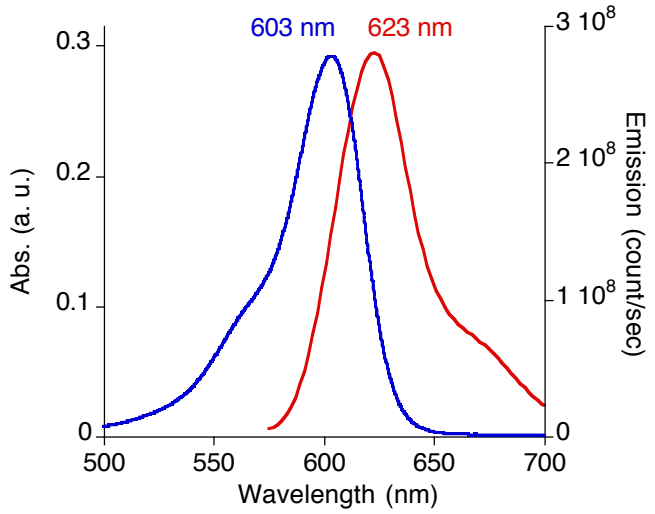


Figure II-2. UV-Vis (blue curve) and emission spectra (red curve) of Q108K:K40L:T51V:R58F/**MCRA**.

Because KL:T51V:R58F/**MCRA** is an acceptably bright complex, we chose to test its utility as a fluorescent tag in mammalian cells. Additionally, its absorption and emission spectra (**Figure II-2**) are well aligned with available lasers on the confocal microscope (594 nm excitation and 615 nm emission long pass filter). Labeling of KL:T51V:R58F (hCRBPII$_{tetra}$) with **MCRA** was evaluated in cervical carcinoma (HeLa) cells. hCRBPII was cloned into a pFlag-CMV2 vector containing EGFP, which serves as a positive control of transfection. Cells expressing EGFP-hCRBPII$_{tetra}$, were incubated with 250 nM **MCRA** for 1 hour at 37 °C. Whole cell fluorescence is observed upon excitation of EGFP at 488 nm (BP 505-530 emission filter). To our delight, upon excitation at 594 nm (LP 615 nm emission filter), whole cell fluorescence can be detected. While it was pleasing to see that **MCRA** can efficiently penetrate the cell membrane, significant background fluorescence was observed (**Figure II-3**, EGFP-hCRBPII$_{tetra}$ at 0 min) when cells were imaged immediately after **MCRA**

incubation. As hCRBPII/**MCRA** complexes do not show wavelength regulation, background fluorescence is presumably due to off-target iminium formation that absorbs and emits in the same wavelength regime as our hCRBPII/**MCRA** complex.

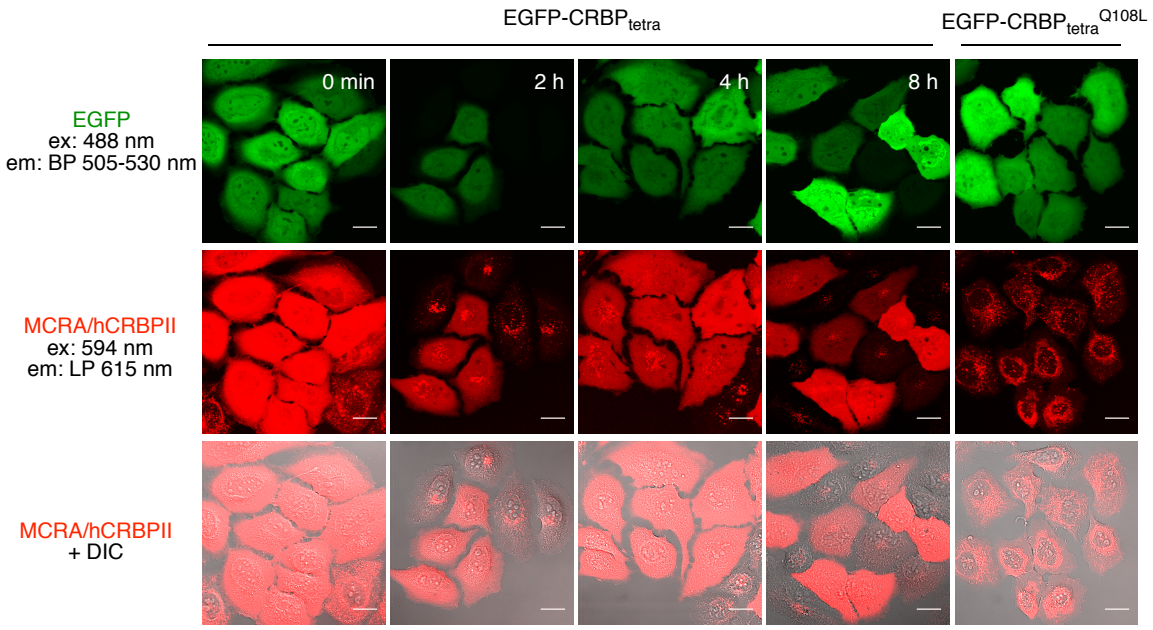


Figure II-3. Labeling of EGFP-hCRBPII$_{tetra}$ in HeLa cells. All samples were incubated with 250 nM **MCRA** for 1 h at 37 °C, washed with PBS buffer two times and then incubated at 37 °C for the indicated amount of time (left to right: 0 min, 2 h, 4 h and 8 h). The right most panel is labeling of EGFP-hCRBPII$_{tetra}^{Q108L}$ after incubating with 250 nM **MCRA** for 1 h at 37 °C, washed with PBS buffer two times and imaged immediately. All scale bars are 20 μM.

To this end, we sought to test whether an incubation period after **MCRA** staining would allow for nonspecific staining to be degraded. HeLa cells expressing the protein fusion were labeled with **MCRA** for 1 hour at 37 °C and imaged after various post **MCRA** incubation times. Cells were imaged 0 min, 2 hours, 4 hours and 8 hours after the initial 1 hour **MCRA** incubation period (**Figure II-3**, EGFP-hCRBPII$_{tetra}$). As expected, the EGFP channel is essentially the same in each experiment. However, in the red channel indicative of

hCRBPII/**MCRA**, more background fluorescence is observed with no post **MCRA** incubation period (cells were imaged directly after **MCRA** incubation and washout), while background fluorescence is dimmest 8 hours post **MCRA** staining. This indicates that a 'degradation' period is necessary after the 1 hour **MCRA** incubation, to allow for the abatement of nonspecific fluorescence.

To irrefutably show that **MCRA** is labeling hCRBPII in the expressed EGFP-hCRBPII$_{tetra}$ fusion, we made the active site lysine free variant, EGFP-hCRBPII$_{tetra}^{Q108L}$, in which Q108 is mutated to leucine. It should not form an iminium with **MCRA**, and thus will not lead to a turn-on response by the addition of **MCRA**. HeLa cells expressing EGFP-hCRBPII$_{tetra}^{Q108L}$ were incubated with 250 nM **MCRA** for 1 hour at 37 °C, washed two times with PBS buffer and imaged immediately (**Figure II-3**, EGFP-hCRBPII$_{tetra}^{Q108L}$). Whole cell fluorescence is observed upon excitation of EGFP at 488 nm (BP 505-530 emission filter). However, upon excitation at 594 nm, no whole cell fluorescence is detected; only significant background fluorescence is observed. Colocalization analyses of **MCRA** in non-transfected cells indicate that nonspecific iminium formation proceeds in the lysosomes (see Dr. Tetyana Berbasova's thesis, **Figure IV-26**). Because the fluorescence from this nonspecific iminium formation is in the same channel as specific hCRBPII/**MCRA** fluorescence, imaging will be encumbered due to low signal to noise ratio. It was hypothesized that with a lower incubation time (requiring a faster reaction between hCRBPII and **MCRA**), hCRBPII/**MCRA** iminium formation would be able to outcompete nonspecific binding.

Having shown that **MCRA** can label hCRBPII$_{tetra}$ in live cells, we focused our efforts on the generation of a fast reacting hCRBPII mutant. This was desirable for a number of reasons, but most notably, to avoid the reaction of **MCRA** with off-target lysines that produce complexes fluorescing at the same wavelength as hCRBPII/**MCRA**. As aforementioned, the long incubation time of 1 hour required for labeling hCRBPII$_{tetra}$, due to the slow rate of iminium formation, leads to substantial nonspecific iminium formation. Additionally, a faster reacting hCRBPII mutant would enable quick imaging and visualization; the 8 h incubation post-addition of **MCRA**, necessary for high contrast imaging, would hamper visualization of processes within this time frame.

To engineer a fast binding hCRBPII mutant for **MCRA**, we took inspiration from the PLP-dependent enzymes and aldolases.[7] In these enzymes, iminium formation is optimized as a necessary part of their function. To aid in the first step of iminium formation, these enzymes employ an acidic residue to activate the substrate's carbonyl for nucleophilic attack by an active-site lysine. We have also observed similar acid catalyzed iminium formation in our lab.

Previous studies in CRABPII showed that L121E was successful at enhancing the rate of iminium formation.[6] An overlay of CRABPII and hCRBPII identified hCRBPII's L117 as an analogous residue to L121 in CRABPII (**Figure II-4**). Gratifyingly, when L117E was introduced into hCRBPII$_{tetra}$, the rate of iminium formation increased by nearly four-fold (**Table II-1**, entry 9). Fortuitously, the quantum yield and brightness of the fluorescent complex were not affected.

Thus, we chose to retain the L117E mutation, which led to a number of fast binding hCRBPII mutants. A Q4F mutation was also included as it has been shown to lead to higher protein expression yields without negatively affecting the hCRBPII/**MCRA** complex's quantum yield or rate of iminium formation (**Table II-1**, entry 10).

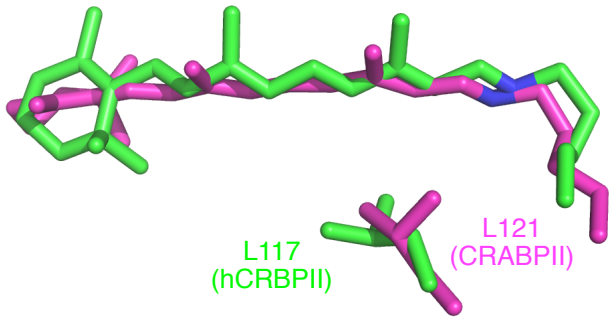


Figure II-4. Overlay of CRABPII (R111K:R132L:Y134F:T54V:R59W, shown in magenta) and hCRBPII (Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4R, shown in green). Coordinates were obtained from PDB 4I9S and 4EEJ.

Iminium formation was fastest when L117E was incorporated into the Q4F containing octa mutant shown in entry 8 (**Table II-1**). The resulting Q108K:K40L:T51V:T53C:R58W:T29L:A33W:Q4F:L117E-hCRPBII nona mutant ($hCRBPII_{nona}$) shows extraordinarily fast reaction kinetics of iminium formation, with a half-life for maturation of 10 seconds (based on fitting to second order kinetics; measured at 23 °C with 20 μM protein and 0.5 equiv **MCRA** at pH ~ 7; the rate of iminium formation was too fast at 37 °C). The introduction of the shorter aspartic acid residue (L117D) led to similar results (**Table II-1**, entry 12). Fortuitously, the quantum efficiency of the $hCRBPII_{nona}$/**MCRA** complex was equivalent to that of $hCRBPII_{tetra}$/**MCRA**, presumably due to the strong electrostatic interaction between the negatively charged carboxylate and the

positively charged iminium, confining the conformation of the generated cyanine dye.

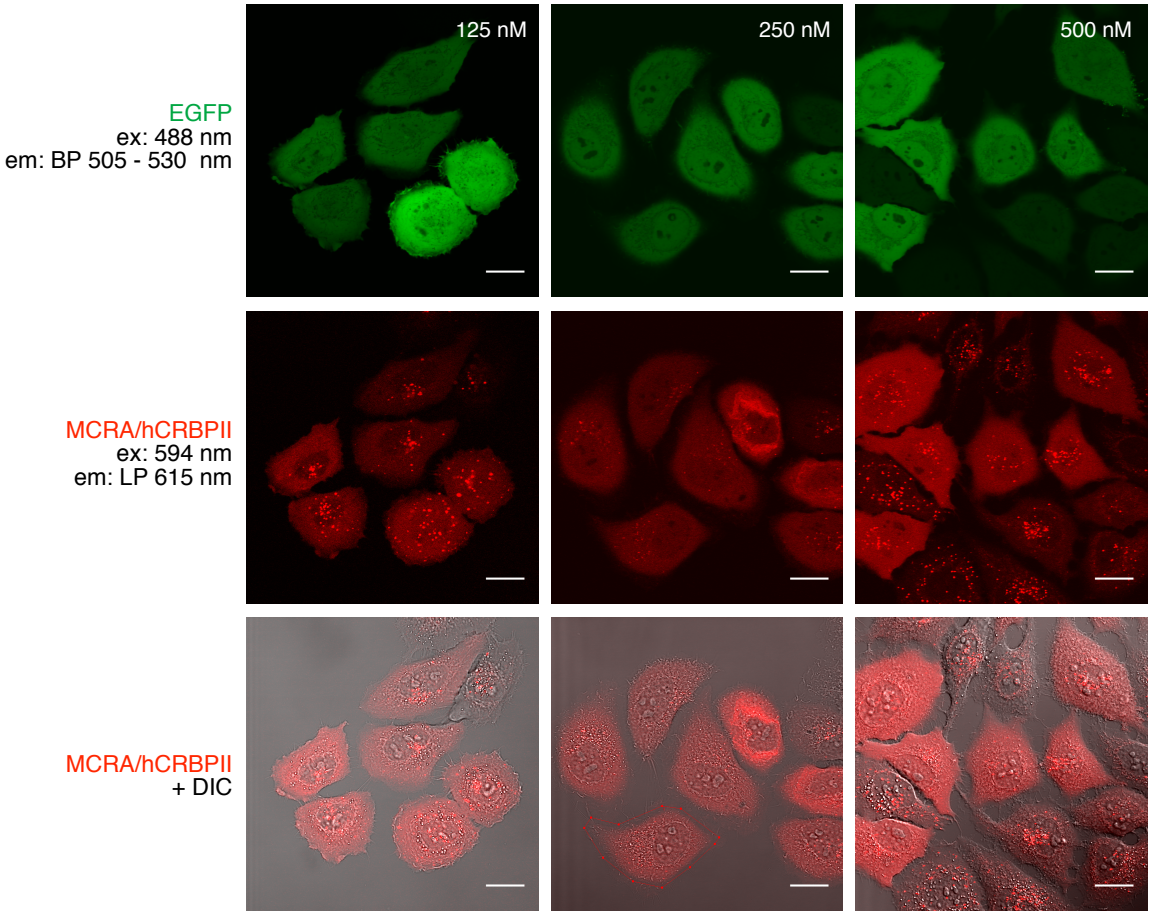


Figure II-5. Optimization of **MCRA** concentration for live cell imaging in HeLa cells expressing EGFP-hCRBPII$_{nona}$; various concentrations (125 nM, 250 nM and 500 nM) were tested. All scale bars are 20 µM.

With a fast binding hCRBPII variant (hCRBPII$_{nona}$) in hand, we sought to test its applicability in live cell imaging. First, we chose to optimize the **MCRA** concentration necessary for effective labeling of hCRBPII; in these experiments, we incubated the cells with **MCRA** for 30 sec at room temperature following by washing with PBS buffer two times before imaging. Three concentrations of **MCRA** of 125 nM, 250 nM and 500 nM were tested in HeLa cells for hCRBPII$_{nona}$ labeling (**Figure II-5**). The best results were obtained when using a **MCRA**

concentration of 250 nM. A **MCRA** concentration of 500 nM led to substantially more background fluorescence, as evidenced by the red fluorescence from non-transfected cells (transfected cells are indicated by the EGFP channel). On the other hand, the intensity of red fluorescence from hCRBPII/**MCRA**, when using a **MCRA** concentration of 125 nM was more dim than when 250 nM **MCRA** was used. Therefore, we chose to proceed with a **MCRA** concentration of 250 nM, as the best signal-to-noise ratio is obtained after 30 sec incubation.

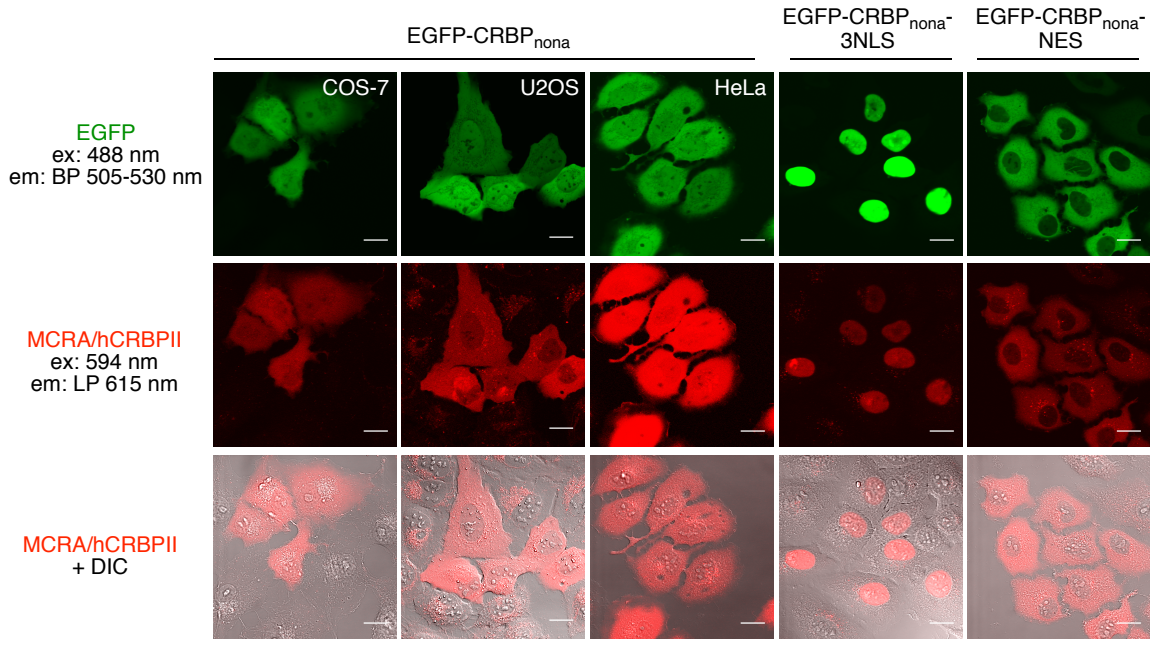


Figure II-6. Labeling of EGFP-hCRBPII$_{nona}$ in Hela cells. All samples were incubated with 250 nM **MCRA** for 30 sec at room temperature, washed with PBS buffer two times and imaged immediately. Localization signals 3NLS and NES targeted hCRBPII to the nucleus and extranuclear space, respectively. NLS = nuclear localization sequence, NES = nuclear export sequence. All scale bars are 20 μM.

For subsequent experiments, routine imaging was performed by incubating the cells with 250 nM **MCRA** for 30 seconds at room temperature. After washing with buffer two times, confocal images were acquired immediately,

demonstrating that the hCRBPII$_{nona}$/**MCRA** complex is formed almost instantaneously *in vivo* and can be used for imaging soon after addition of ligand.

EGFP-hCRBPII$_{nona}$ was successfully expressed and labeled with **MCRA** in multiple cell lines including COS7, U2OS and HeLa cells (**Figure II-6**). As before, excitation at 488 nm (BP 505-530 emission filter) leads to whole cell fluorescence from EGFP. Similarly, excitation at 594 nm (LP 615 nm emission filter) shows whole cell fluorescence as a result of hCRBPII/**MCRA** complex. Minimal background in the red channel is observed in both COS7 and HeLa cells; moderate background fluorescence is apparent in U2OS cells.

Subsequently, targeting of EGFP-hCRBPII$_{nona}$ was investigated by the inclusion of a nuclear localization sequence (3NLS) and nuclear export sequence (NES).[8] Fortuitously, the nucleus fluoresces as efficiently as the cytosol localization under identical conditions, even though the chromophore needs to penetrate not only the cellular membrane, but also enter the nucleus before binding to hCRBPII$_{nona}$ (**Figure II-6**). EGFP-hCRBPII$_{nona}$-NES showed fluorescence in the extranuclear space in both the EGFP and hCRBPII/**MCRA** (red) channels.

Gratifyingly, nonspecific fluorescence is not observable in any of the hCRBPII$_{nona}$ fusion constructs in HeLa cells. It appears that labeling of hCRBPII$_{nona}$ with **MCRA** occurs within 30 seconds and can outcompete nonspecific iminium formation in the lysosomes. Hence, long incubation times

following the buffer exchange necessary for $hCRBPII_{tetra}$ were not needed for $hCRBPII_{nona}$, due to its fast iminium formation with **MCRA**.

As has been shown, hCRBPII/**MCRA** can successfully label mammalian cells, targeting to intracellular organelles, with an unprecedented half-life of 10 seconds. Advantageously, hCRBPII/**MCRA** complexes also exhibit high quantum yields and extinction coefficients. The only pitfall with the system is that background fluorescence is observable from nonspecific iminium formation. Presumably, this is due to the high pK_a of the *in situ* generated cyanine dye. Additionally, the absorption wavelength is essentially constant in different hCRBPII variants when coupled with **MCRA** due to the high degree of conjugation between the iminium and terminal nitrogen.

Thus, we can imagine developing fluorogenic probes in two ways. Firstly, we can prevent nonspecific iminium formation. It is envisioned that this can be done by using an initially nonfluorescent probe that becomes fluorescent upon iminium formation with hCRBPII. However, it would require a less reactive fluorophore than merocyanine aldehyde (namely the methyl ketone variant), such that the probe does not react nonspecifically, rendering it only fluorescent upon reaction inside an activated hCRBPII variant. A second option would involve utilizing environmentally sensitive probes to tailor the specific hCRBPII/fluorophore emission, such that the fluorescence from the desired reaction is red-shifted as compared to that of nonspecific iminium formation. Presumably, this can be achieved by changing the electrostatic environment

around the fluorophore inside of the protein cavity. This chapter will focus on the latter.

II.2 Principles of solvatochromicity

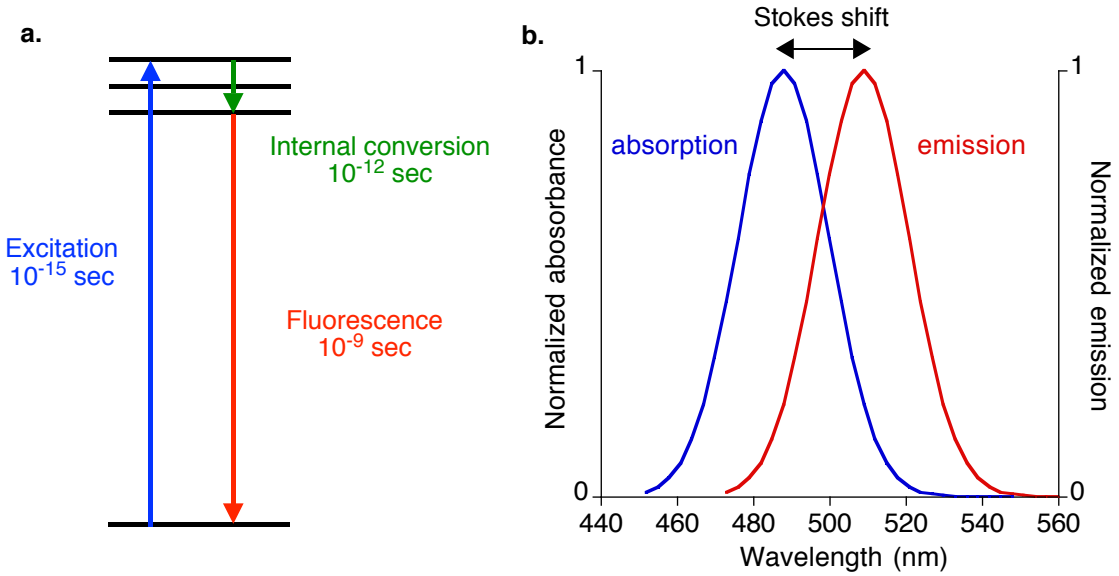


Figure II-7. **a**. Simplified Jablonski diagram. **b**. Diagram of absorption and emission spectra; the difference in the band maxima is the Stokes shift.

Fluorescence generally involves three processes; these are excitation, nonradiative transition, and fluorescence emission as is shown by the simplified Jablonski diagram in **Figure II-7a**.[9-11] First, a photon is supplied and the fluorophore is excited to a vibrational level within the first excited state (S_1). This process occurs on a 10^{-15} seconds time scale. The excited state exists for a finite time (typically nanoseconds), after which the fluorophore undergoes conformational changes such as internal conversion (10^{-12} sec) or intersystem crossing (this produces the triplet state which leads to phosphorescence). Upon returning to the ground state, a photon is emitted (10^{-9} seconds).

Emission from fluorophores occurs at wavelengths bathochromically shifted from the absorption due to the dissipation of energy in the excited state. This difference in energy of the emission maxima and absorption maxima is referred to as the Stokes shift (**Figure II-7b**). Importantly, a larger Stokes shift allows for emission photons to be isolated from excitation photons.

Because the time required to electronically excite a fluorophore is much shorter than the time required to undergo vibrational motion, the position of nuclei of the absorbing molecule do not change between the ground state and excited state.[12] However, if the lifetime of the excited state is long enough, then reorientation of the surrounding solvent molecules can take place, after which fluorescence occurs from this state. Thus, solvent and environment can affect a fluorophore's emission spectra and quantum yields.[13] Examples include solvent polarity, viscosity, rate of solvent relaxation, fluorophore conformational changes, rigidity, charge transfer, excited state reactions, and changes in radiative and non-radiative decay rates.[13] These factors have all led to the development of a variety of environment-sensitive dyes.[14-16]

One such class of dyes is solvatochromic fluorophores.[14-16] Solvatochromic fluorophores display emission properties, including emission wavelengths, quantum yields and fluorescence lifetimes, that are highly sensitive to their environment. The structures of some of the most common solvatochromic fluorophores are shown in **Figure II-8**. Typically, only fluorophores that are themselves polar exhibit substantial sensitivity to solvent polarity. These include

fluorophores derived from a variety of molecules, including dansyl,[17, 18] prodan and derivatives,[19-23] anthradan,[24] fluorene,[25-28] dapoxyl,[29-31] nitrobenzoxadiazole (NBD),[32-35] dimethylaminophthalimide,[36-39] coumarin,[40-42] phenoxazine,[43-47] among others. The similarity amongst these dyes is that they bear a conjugated system with an electron-donor and electron-acceptor group, generating a push-pull system. Most solvatochromic dyes show relatively low fluorescence quantum yields in polar solvents and high fluorescence quantum yields in apolar solvents.[13-16]

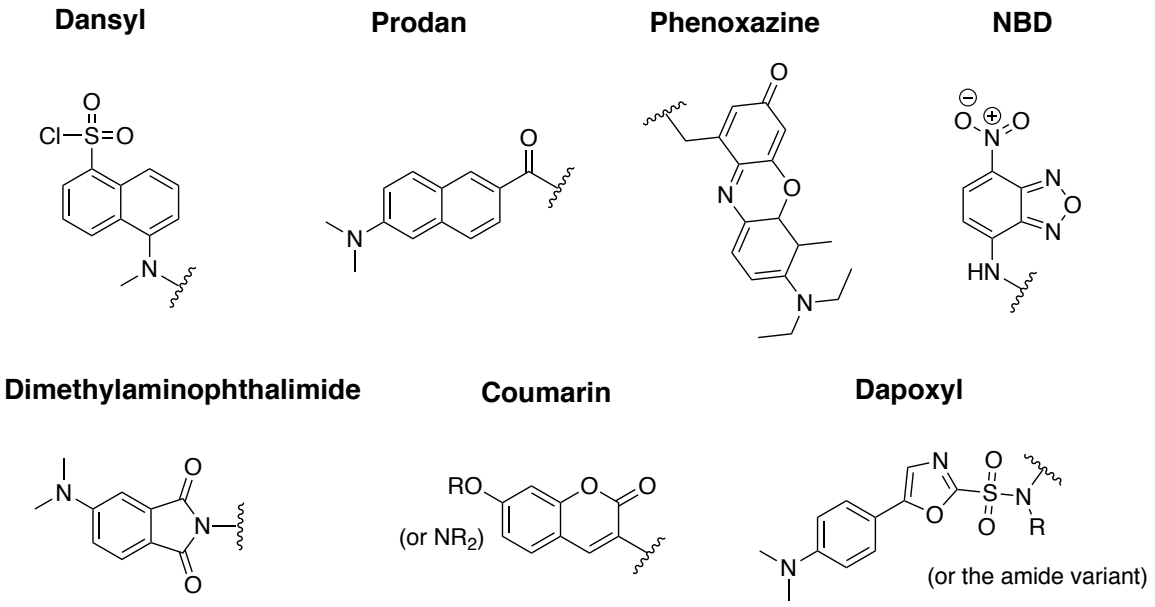


Figure II-8. Examples of fluorescent solvatochromic dyes. The wavy lines indicate the common point for dye conjugation.

Typically, these fluorophores have a larger dipole moment in the excited state than in the ground state. If the dipole moment of the molecule increases upon excitation, then a more polar solvent will serve to stabilize the excited state. As a result, the separation between the ground state and excited state energies is decreased. Hence, the emission is red-shifted; this effect is called positive

solvatochromism. Conversely, a hypsochromic shift of the emission band with increasing solvent polarity is termed negative solvatochromism. The magnitude of the shift is dependent on the change in the dipole moment of the probe. The Jablonski diagram depicting the energies of the different electronic states of the system is shown in **Figure II-9**.

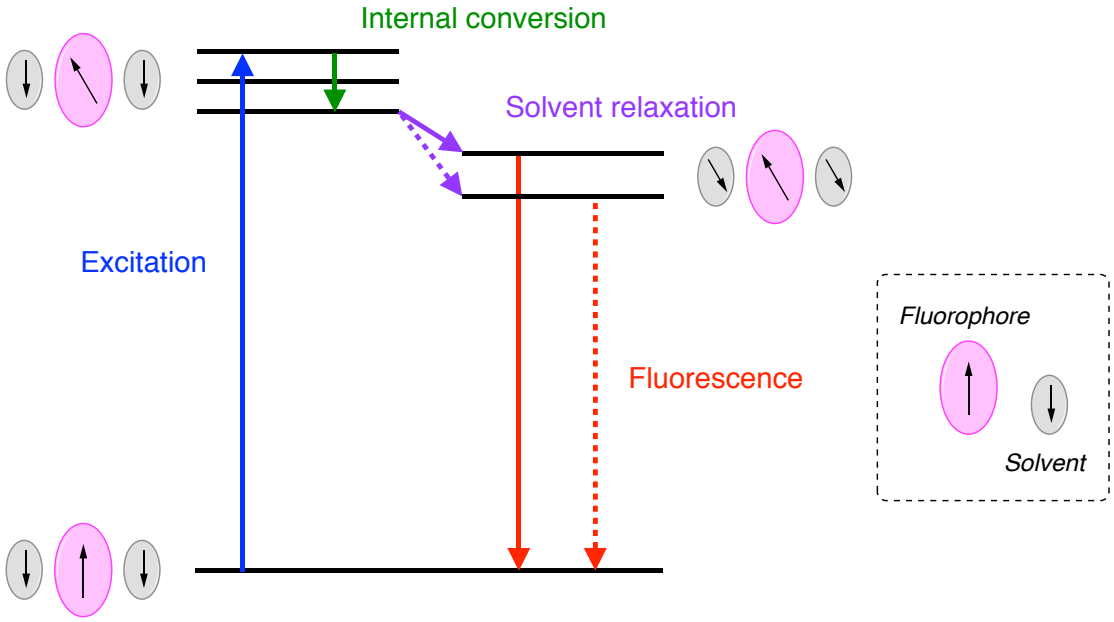


Figure II-9. Origins of solvatochromicity in fluorescence.

Numerous scales of solvent polarity have been developed; many of these scales are based on the spectral data of a single standard probe.[4, 48-57] Reichardt developed one such scale, the $E_T(30)$ scale, which is based on the spectroscopic behavior of the negatively solvatochromic betaine indicator dye 2,6-diphenyl-4-(2,4,6-triphenyl-1-pyridinio)phenolate (Betaine 30).[57-59] Betaine 30 shows solvent dependent spectral shifts (**Figure II-10**), in which absorption wavelength is higher (i.e. lower in energy) in nonpolar solvents.[59] Since its introduction, the $E_T(30)$ parameter has been used for empirically measuring the polarity of many systems.

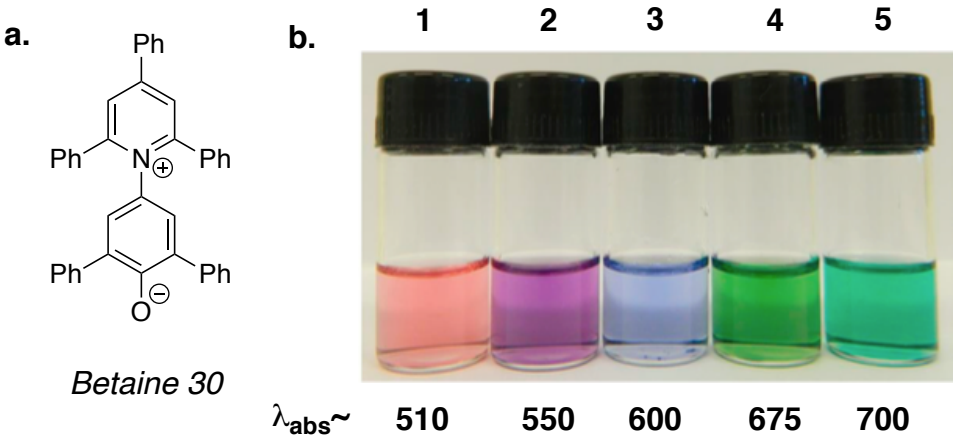


Figure II-10. **a**. Structure of Betaine 30. **b**. Solutions of Betaine 30 in (1) methanol, (2) ethanol, (3) 1-octanol, (4) *N,N*-dimethylacetamide and (5) dichloromethane.

II.3 Choice of fluorophore

In the late 1990s, Diwu and coworkers developed the Dapoxyl dyes (**Figure II-8**); these strongly solvatochromic compounds exhibit distinctive variations in quantum yields, fluorescence wavelength maxima (i.e. larger Stokes shifts) and extinction coefficients with changes in polarity.[29] Fortuitously, they also display a much lower quantum yield in polar solvents than nonpolar solvents (Φ_{hexane}=0.91 and $\Phi_{1:4\ acetonitrile:water}$=0.04). Their characteristic of being almost nonfluorescent in aqueous solutions should translate to almost no fluorescence being observed from the unbound probe in an aqueous solution, a feature beneficial in live cell imaging.

Since their discovery, many derivatives of the Dapoxyl dyes have been synthesized. Min and coworkers were successful in generating a Dapoxyl dye library of 80 unique structures.[30] Through a fluorescence-based high throughput screening of their dapoxyl library, one Dapoxyl derivative was found to be highly sensitive toward human serum albumin (HSA). Under physiological conditions

(50 mM HEPES, pH 7.4), **A-41S** displayed a fluorescence increase of 55-fold and a blue-shift of the fluorescence emission (520 to 473 nm) (**Figure II-11**).

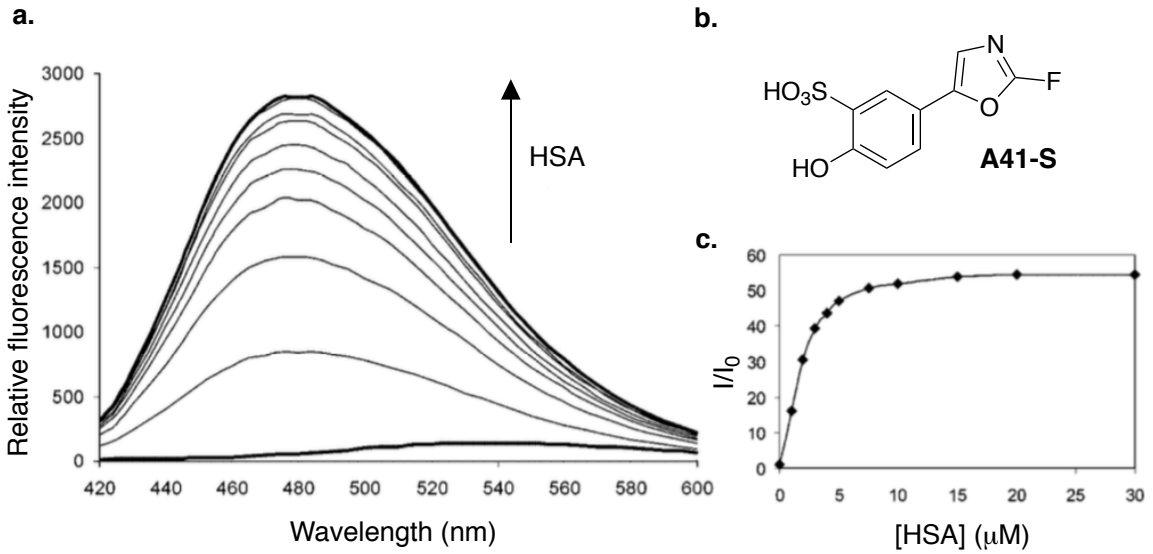


Figure II-11. Fluorescence response of **A-41S** to HSA. **a**. 5 μM **A41-S** with HSA at 0, 1, 2, 3, 4, 5, 7.5, 10, 15, 20 and 30 μM. **b**. Structure of **A41-S**. **c**. Corresponding intensity changes are the ratio of intensity of **A-41-S** with/without HSA.

As aforementioned, Dapoxyl dyes exhibit a typical push pull system, conjugated through a pi-system, where the effect of donor and acceptor substitutions has already been evaluated. To this end, Koji Suzuki and coworkers investigated the effect of the central heteroaromatic ring system (**Figure I-12**), which serves as a pi-linker between the donor and acceptor moieties of the fluorophore, replacing the oxazole with thiophene (**KSD-3**), furan (**KSD-2**) and pyrrole (**KSD-1**).[31]

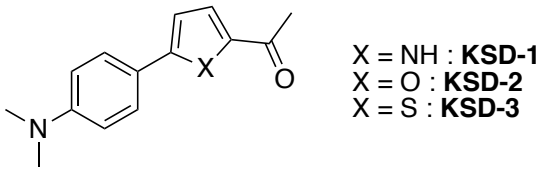


Figure II-12. Structures of Dapoxyl dyes used to investigate the effect of the pi-linker.

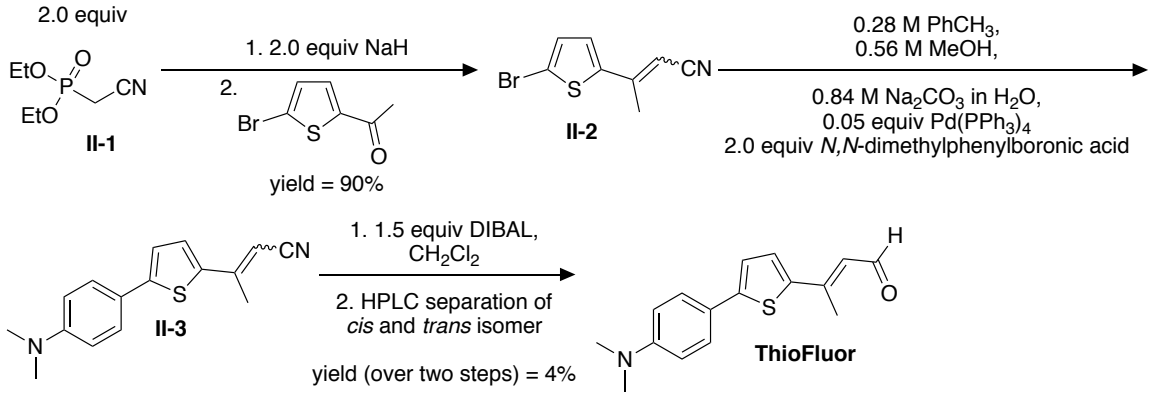
Some spectroscopic properties of the chromophores are detailed in **Table II-2**.[31] The absorption maxima do not change significantly upon changing polarity of the solvent, indicating that the solvent does not stabilize the ground state conformations of these molecules. Conversely, strong positive solvatochromic shifts are observed when the solvent polarity is increased for all three dyes. As described previously, this indicates a solvent stabilized excited state, presumably due to the difference in dipole moments of the ground state and excited state.

Table II-2. Spectral properties of **KSD-1-3** in solvents of different polarities.

Solvent	**KSD-1**			**KSD-2**			**KSD-3**		
	λ_{abs}	λ_{em}	Φ	λ_{abs}	λ_{em}	Φ	λ_{abs}	λ_{em}	Φ
Toluene	358	413	0.32	368	428	0.60	386	455	0.67
1,4-Dioxane	352	415	0.59	366	434	0.82	385	466	0.76
Ethyl acetate	352	423	0.64	364	446	0.72	384	479	0.61
Chloroform	365	443	0.33	378	465	0.62	395	490	0.54
Acetone	355	449	0.77	369	474	0.78	389	506	0.60
DMSO	363	462	0.82	380	499	0.74	400	531	0.49
Acetonitrile	355	458	0.77	371	491	0.80	390	519	0.57
i-Propanol	366	479	0.34	379	512	0.60	394	535	0.46
Ethanol	365	495	0.20	380	528	0.40	395	548	0.18
Methanol	363	509	0.11	381	541	0.13	396	559	0.02

The shifts in emission spectra for the dyes from toluene to methanol are 96 nm, 113 nm and 104 nm (indicating a polarity dependent intramolecular charge transfer from the electron donating *N,N*-dimethyl amino phenyl moiety to the electron withdrawing acetyl unit), for **KSD-1**, **KSD-2** and **KSD-3**, respectively. It is also observed that when thiophene is used as the pi-linker, the longest

emission wavelengths are obtained. This is likely due to better effective conjugation of the thiophene moiety.



Scheme II-1. Synthesis of **ThioFluor**.

Based on the previous data, we chose to start from the aldehydic analog of **KSD-3**. **ThioFluor** was synthesized in three steps from the commercially available 2-acetyl-3-bromothiophene (**Scheme II-1**). Horner-Wadsworth-Emmons (HWE) reaction with ylide **II-1** yields intermediate **II-2**. **II-2** was coupled with *N,N*-dimethylphenylboronic acid, resulting in **II-3**. Final reduction of the nitrile resulted in the desired fluorophore ThioFlour.

II.4 Spectroscopic properties of ThioFluor

Spectroscopic characterization of the free aldehyde was first undertaken. Results are summarized in **Table II-3**. As is the case for the known solvatochromic fluorophores, **ThioFluor**'s absorbance does not change significantly in different solvents (**Figure II-13a**), but its emission varies over 140 nm from nonpolar toluene to the more polar ethanol, with a larger Stokes shift in polar solvents (**Figure II-13b**). Thus, under white light, all solutions appear colorless to faint yellow (**Figure II-3c**), while under UV irradiation (with a TLC

headlamp, 365 nm), the solutions show radically different colors (**Figure II-13d**). As expected the free aldehyde's absorbance is nearly constant while its emission changes as a function of the solvent's $E_T(30)$ value (**Figure II-13e**).[4]

Table II-3. Spectroscopic characterization of **ThioFluor** in various solvents.

Solvent	λ_{abs}	λ_{em}	Stokes shift	ε ($M^{-1}cm^{-1}$)	Φ^a
Toluene	422	516	94	32,149	0.02
Ethyl acetate	414	558	144	28,939	0.11
Dimethyl sulfoxide	437	623	186	26,891	0.53
Ethanol	430	654	224	27,021	0.06
PBS buffer	393	-	-	8,842	0.00

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

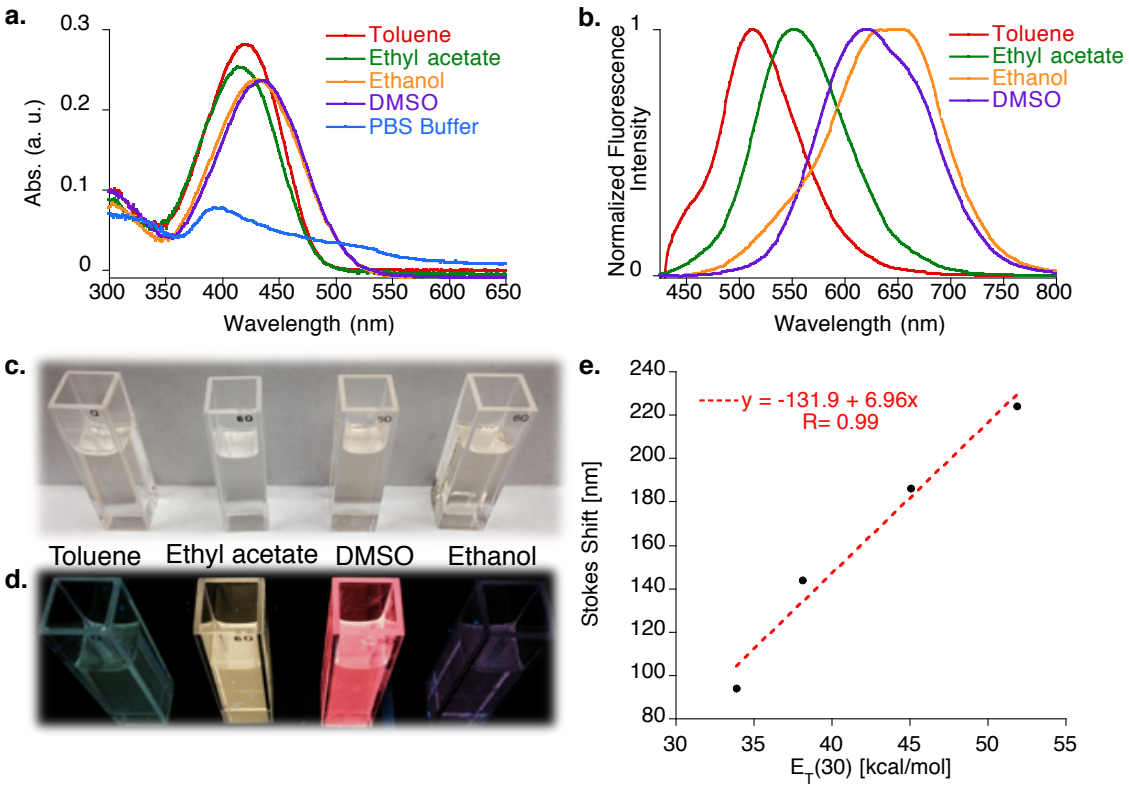
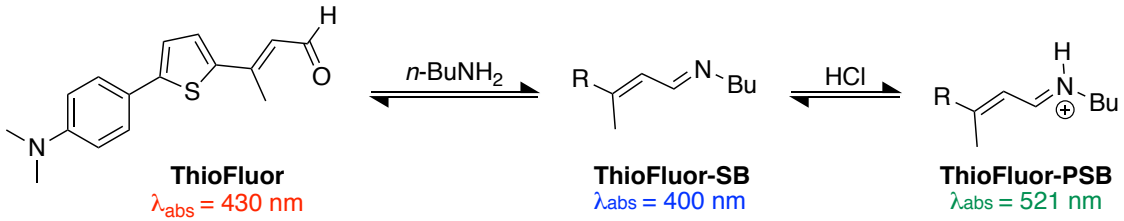


Figure II-13. Spectroscopic properties of **ThioFluor** in different solvents. **a**. UV-Vis and **b**. Fluorescence spectra of **ThioFluor**. **ThioFluor** under **c**. white light and **d**. UV irradiation at 365 nm (with TLC handlamp). **e**. Stokes shift in different solvents versus the $E_T(30)$ value indicates that **ThioFluor** is solvatochromic.

The extinction coefficients of **ThioFluor** in toluene, ethyl acetate, dimethyl sulfoxide and ethanol do not drastically differ; they are all approximately 30,000 $M^{-1}cm^{-1}$. However, the formation of aggregates is quite apparent from the UV-Vis spectrum of **ThioFluor** in PBS buffer.[60] Fortuitously, this leads to essentially a non-fluorescent species in PBS buffer, which may indicate that when applied to *in vivo* applications the free aldehyde should not be fluorescent, acting as a fluorogenic probe. Most solvatochromic fluorophores are highly fluorescent in nonpolar solvents; however, **ThioFluor** is only slightly emissive in toluene. Interestingly, **ThioFluor** is only highly fluorescent in dimethyl sulfoxide, with a quantum yield 0f 0.53.



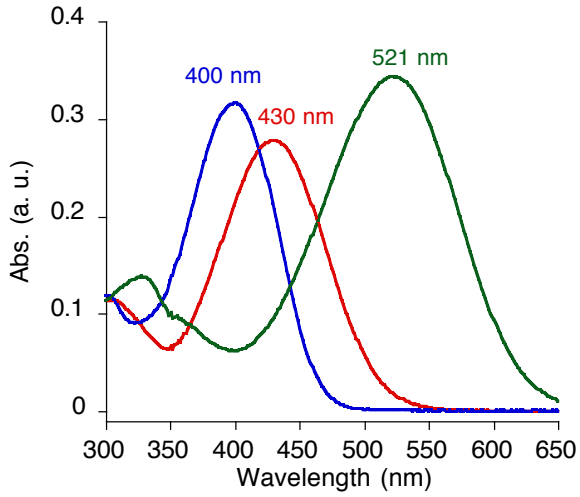


Figure II-14. Formation of the protonated Schiff base (PSB) of **ThioFluor** with *n*-butyl amine in ethanol.

The aldehyde was then coupled with *n*-butyl amine in ethanol, in order to mimic the product that will be obtained upon condensation with an active site lysine in the protein pocket. Reaction of the aldehyde with *n*-butyl amine produces a blue-shifted imine, or Schiff base (SB), with absorption maxima of 400 nm (**Figure II-14**). Subsequent protonation of the **ThioFluor**-SB yields a resonating cation, resulting in a large red-shift in UV-Vis, with an absorbance maximum of 521 nm. **ThioFluor**-PSB was formed in a variety of solvents. Results are summarized in **Table II-4.**

Table II-4. Spectroscopic properties of **ThioFluor**-PSB with *n*-butyl amine.

Solvent	λ_{abs}	λ_{em}	Stokes Shift	Φ^a
PBS buffer	464	-	-	-
Ethanol	521	689	168	0.02
Tetrahydrofuran	494	674	180	0.17
Formamide	525	707	182	0.01
Acetone	493	687	194	0.03
Dimethyl sulfoxide	512	716	204	0.02

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

Neither the absorption maxima nor the emission maxima of the **ThioFluor**-PSB vary significantly in different solvents. It should be noted that THF was the least polar solvent employed, because acidification was carried out using HCl, which requires the solvents to be miscible. Nonetheless, it is seen that the **ThioFluor**-PSB is much less solvatochromic than the **ThioFluor** aldehyde. Additionally, the emission wavelength of the PSB is different in aprotic and protic solvents, indicating that PSB emission does not exactly follow a purely solvatochromic trend (as a result of a purely polar effect) based on just solvent

polarity (**Figure II-15**).[4] This suggests that some other factor, such as hydrogen bonding, is affecting the emission wavelength of **ThioFluor**-PSB; this is not surprising given that **ThioFluor** possess a positively charged nitrogen, which can hydrogen bond with polar solvents. **ThioFluor**-PSB was much less solvatochromic than **ThioFluor**; however, we chose to start coupling the chromophore to hCRBPII protein mutants, in order to deduce whether **ThioFluor**'s emission could be altered by the introduction of point mutations in hCRBPII.

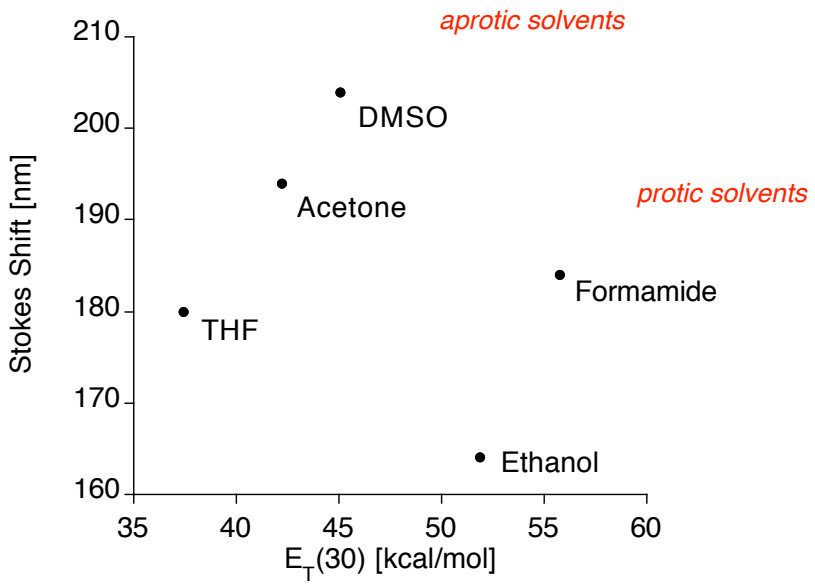


Figure II-15. Stokes shift in different solvents versus the $E_T(30)$ value indicates that **ThioFluor**-PSB is not solvatochromic. Different correlations are observed in protic and aprotic solvents.

II.5 Removal of polarity induces a red-shift in absorption and emission

It was quickly realized that global polarity changes of the binding pocket led to little to no soluble protein expression, and the application of solvatochromicity purely in terms of its definition would not be possible. In other

words, it is not possible to introduce enough polar residues into the binding pocket to induce a global polarity change, while still obtaining soluble protein after expression. It is hypothesized that the introduction of hydrophilic residues in the protein active site leads to protein destabilization, causing protein unfolding so that the polar residues are exposed to the solvent exterior. Hence, it was decided to introduce interactions to the chromophore via point mutations, which could alter its absorption.

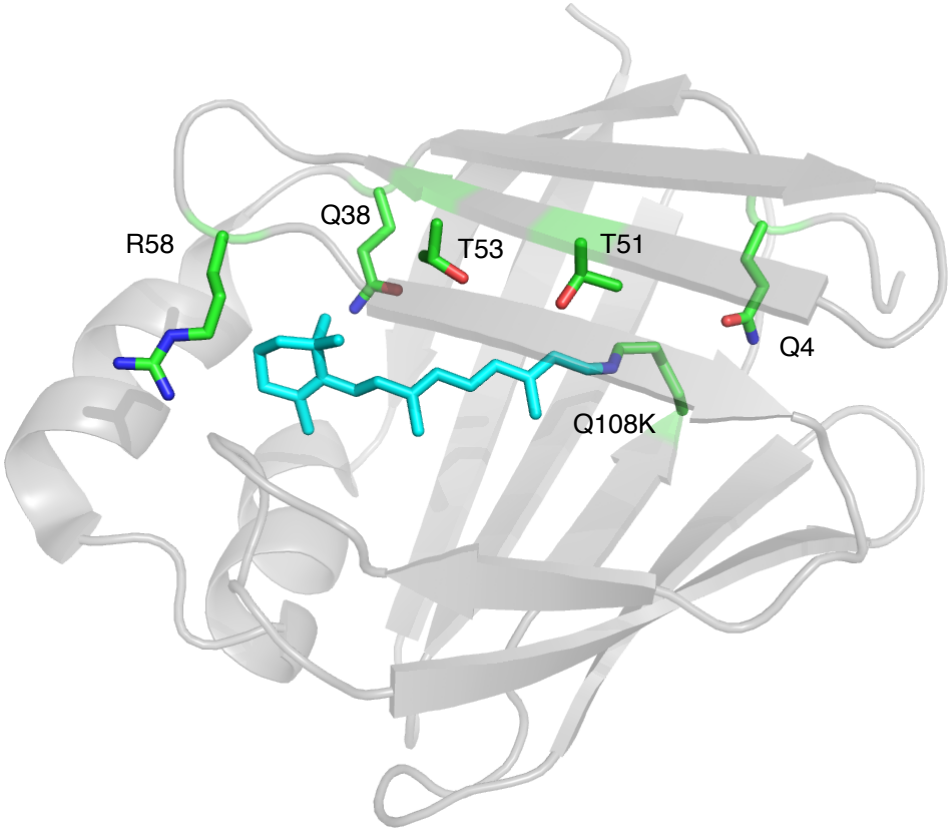


Figure II-16. Polar residues mutated in order to remove polarity from the binding pocket. Coordinates obtained from PDB 4EXZ (hCRBPII-Q108K:K40L/retinal).

Our group has previously shown, utilizing retinal as a ligand, that an even distribution of electrostatic potential across the entire chromophore is essential for maximal bathochromic shifting.[1-3] To this end, we tried to remove polarity in

two ways: 1) removal of the more polar residues in the binding pocket nearest the chromophore and 2) removal of the water molecule directly interacting with the iminium. Removal of these negative dipoles in the PSB region should encourage positive charge delocalization.

Generally, hCRBPII has a hydrophobic binding pocket, with minimal polar residues. To this end, mutagenesis was carried out on T51, T53, Q4, Q38 and R58 residues (**Figure II-16**), mutating them to similarly sized hydrophobic residues in the parent double mutant Q108K:K40L (KL). KL was used, as PSB could not be observed for the Q108K mutant, as acidification of the complex only led to precipitation. Presumably K40L increases the pK_a slightly, such that some PSB is formed.

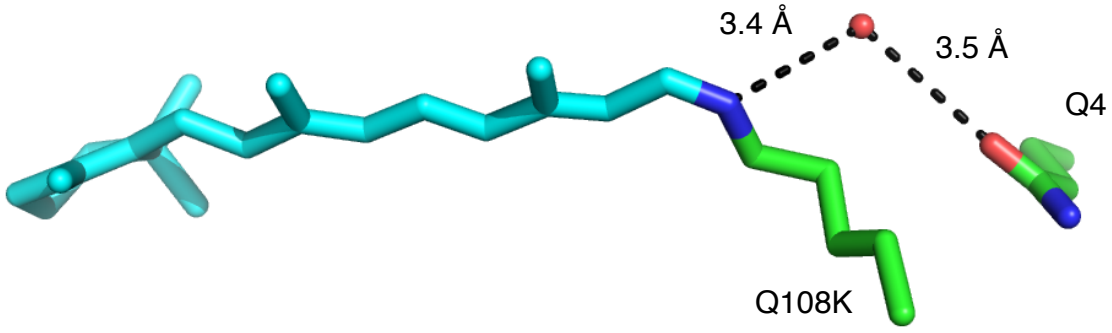


Figure II-17. Water mediated hydrogen bonding between Q4 and the hCRBPII/retinal iminium. Coordinates obtained from PDB 4EFG (hCRBPII-Q108K:K40L:T51V:T53C:R58W:T29L:Y19W/retinal).

We chose to start with mutation of Q4 because in crystal structures of hCRBPII/retinal with Q4, there is a water mediated hydrogen bond from the iminium to Q4, capable of stabilizing the protonated state of the PSB (**Figure II-17**).[3] To disrupt this presumed water network, Q4 was mutated to hydrophobic

phenylalanine in KL (**Table II-5**, entry 1). As expected, the removal of polarity leads to a red-shift in both absorption and fortuitously emission, a change of 27 nm and 17 nm for absorption and emission, respectively from KL (**Table II-5**, entry 2). From this initial hit, it was suspected that a bathochromic shift in absorption should result in a bathochromic shift in emission; hence, studies were continued by attempting to red-shift the absorption (and in effect emission).

Table II-5. Spectroscopic change as a result of Q4F mutation.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L	580	674	nd
2	Q108K:K40L:Q4F	607	691	0.12

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

Hydrophobic valine was substituted for the more polar T51 as an isosteric replacement. Gratifyingly, the Q108K:K40L:T51V hCRBPII triple mutant led to a 28 nm red-shift in absorption and 16 nm red-shift in emission from the parent Q108K:K40L (**Table II-6**). This red-shift is in agreement with the prior conclusion that removing negative polarity in the Schiff base region leads to more delocalization along the polyene, in the case of hCRBPII/retinal. Beyond the T51V mutation, other residues were not tested at position 51. The T51V mutation served an even larger purpose than red-shifting the wavelength; fortuitously, T51V is also capable of monomerizing hCRBPII variants (see **Chapter VI** for a full discussion) and it was thus retained in most mutants tested. As will be discussed later (**Chapter II.8.1**), complexes of hCRBPII with **ThioFluor** displayed significantly different spectroscopic properties of the monomer and domain-

swapped dimer. Only spectroscopic properties of the monomer will be discussed here.

Table II-6. Spectroscopic change as a result of T51V mutation.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L	580	674	nd
2	Q108K:K40L:T51V	608	690	0.12

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

As shown in **Figure II-18**, T53 is situated near the middle of the polyene in the crystal structure of KL/retinal, providing a hydrogen bond to T51 through a water molecule. In order to alter its polarity, T53 was replaced with various residues including alanine, serine and valine in the parent template KL. Interestingly, all three substitutions led to at least a slight red-shift in absorption. From KL the change in absorption was 5 nm, 7 nm, and 17 nm for T53A, T53S and T53V, respectively (**Table II-7**).

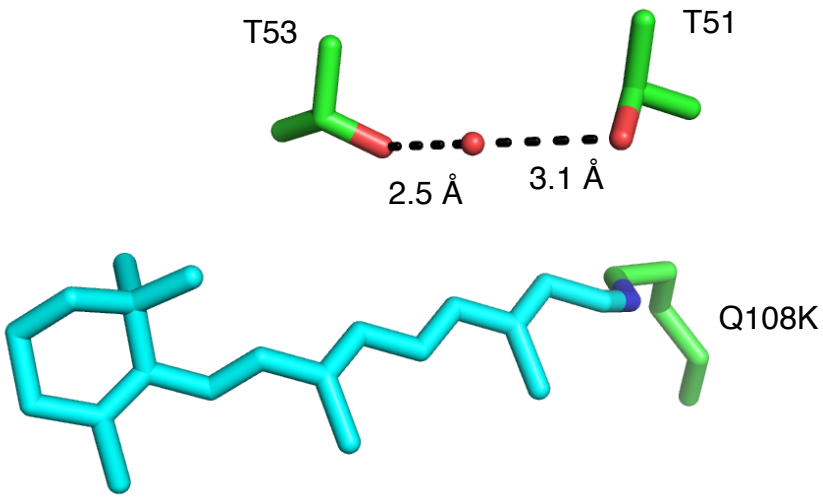


Figure II-18. Water mediated hydrogen bonding between T51 and T53. Coordinates obtained from PDB 4EXZ (hCRBPII-Q108K:K40L/retinal).

There seems to be no trend when looking at the absorption wavelength versus the polarity. Mutations to nonpolar alanine and valine red-shifted absorption, but did not significantly affect the emission wavelength. T53S moderately affected the absorption wavelength, yet had the largest effect on emission (674 nm in KL to 682 nm). Threonine and serine are of approximately the same polarity; it is the volume of the side chain that is different. Presumably, in these cases, polarity is not the only determining factor since no correlation is found between the electrostatic substitution and wavelength. It is possible that chromophore orientation is not preserved through the course of these mutations, or the conformation of residues in close contact with T53 have changed.

Table II-7. Spectroscopic change as a result of mutating T53.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L	580	674	nd
2	Q108K:K40L:T53A	585	677	0.11
3	Q108K:K40L:T53S	587	682	0.08
4	Q108K:K40L:T53V	597	672	0.13

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

As seen in the crystal structure of KL/retinal (**Figure II-18**), T51 and T53 interact via a water-mediated hydrogen bond. Mutation of T51 to valine would abolish this interaction, leading to a red-shift in wavelength, as discussed above. T53S could potentially incite a similar effect. From the KL/retinal structure, a water-mediated hydrogen bonding interaction between Q38 and Q128 is observed (**Figure II-19**). If T53S could form a tight hydrogen bond with Q38 (meaning that T53S has a higher propensity to interact with Q38 rather than

T51), it would disrupt the conserved water-mediated interaction between Q38 and Q128 that is presumably apparent in the KL/**ThioFluor** structure. The overall effect is the ejection of water, and thus reduced polarity, near the chromophore.

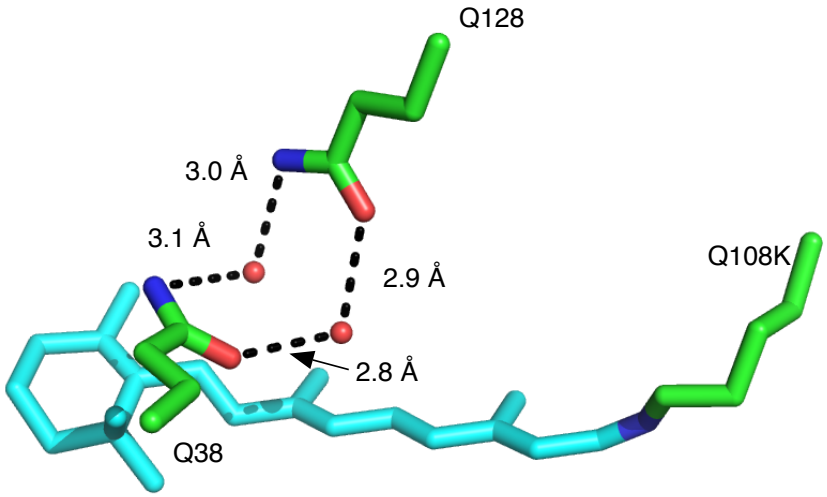


Figure II-19. Water mediated hydrogen bonding between Q38 and Q128. Coordinates obtained from PDB 4EXZ (hCRBPII-Q108K:K40L/retinal).

If interrupting the water mediated hydrogen bonding between Q38 and Q128 leads to a red-shift in wavelength, one would hypothesize that a Q38 to nonpolar mutation should also result in a red-shift in wavelength. To this end, Q38 was mutated to phenylalanine in KL. Gratifyingly, a red-shift of 8 nm was observed in absorption (**Table II-8**). This further supports the theory that T53S mutation must be removing polarity in the pocket by way of interfering with the water mediated interaction between Q38 and Q128.

Table II-8. Spectroscopic change as a result of Q38F mutation.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L	580	674	nd
2	Q108K:K40L:Q38F	588	680	0.13

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

Although located at the entrance of the binding pocket, R58 is the last of the more polar residues that is within interacting distance of the chromophore. Replacement of polar charged arginine with aromatic tyrosine and phenylalanine in KL both led to a bathochromic shift in absorption and emission with **ThioFluor** (**Table II-9**). This is in line with previous studies that showed that encapsulation of the retinal chromophore in hCRBPII resulted in a red-shift in wavelength due to more effective shielding of the binding pocket.[3] A wider screening of R58 mutations is detailed in **Section II.6**.

Table II-9. Spectroscopic change as a result of mutating R58.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L	580	674	nd
2	Q108K:K40L:R58Y	591	681	0.09
3	Q108K:K40L:R58F	597	678	0.18

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

In summary, it should be noted that residues closer to the iminium, Q4 and T51, led to a larger red-shift than the T53 and Q38 residues farther away. As aforementioned and reiterated in **Table II-10**, both T51V and T53S red-shift the absorption and emission. As previously stated, T51V is also capable of monomerizing hCRBPII variants (see **Chapter VI** for a full discussion). To look at the combinatorial effect of T51V and T53S, the tetramutant Q108K:K40L:T51V:T53S was made. The additive effect from the addition of T51V and T53S to KL was quite apparent. Gratifyingly, red-shifts in absorption and emission from KL were substantial, at 47 nm and 23 nm, respectively (**Table II-10**, entry 4). Thus, the template Q108K:K40L:T51V:T53S was retained for

further protein engineering. Q4 mutations were not included, due to prior studies showing that the pK_a of hCRBPII/retinal complexes are severely depressed when Q4 was substituted with phenylalanine.[61, 62]

Table II-10. Additive effects of T51 and T53 mutations.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L	580	674	nd
2	Q108K:K40L:T53S	587	682	0.08
3	Q108K:K40L:T51V	608	690	0.12
4	Q108K:K40L:T51V:T53S	627	697	0.15

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

II.6 Attempts to encapsulate the binding cavity

Studies conducted by Dr. Wenjing Wang previously showed that an aromatic residue at position 58 leads to a bathochromic shift in absorption when retinal is used as a ligand,[3] presumably due to more effective shielding of the binding pocket. To this end, R58F, R58Y, R58W and R58H mutations were introduced into the parent mutant Q108K:K40L:T51V:T53S (**Table II-11**).

Table II-11. Spectroscopic changes as a result of mutating R58 in Q108K:K40L:T51V:T53S to aromatic residues.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a	$t_{1/2}{}^b$ (min)
1	Q108K:K40L:T51V:T53S:R58F	617	691	0.12	184
2	Q108K:K40L:T51V:T53S:R58Y	622	691	0.11	226
3	Q108K:K40L:T51V:T53S:R58W	623	697	0.08	320
4	Q108K:K40L:T51V:T53S	627	697	0.15	146
5	Q108K:K40L:T51V:T53S:R58H	643	705	0.14	176

[a]Absolute quantum yield was measured on a Quantaurus-QY. [b]Half-life based on the rate constant obtained from pseudo first order fitting; measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7.

Unexpectedly, of the four aromatic residues, only R58H led to a bathochromic shift (**Table II-11**, entry 5), while the other three aromatic mutations actually led to a minor blue-shift in absorption. However, unlike retinal, **ThioFluor** contains heteroatoms that can interact with surrounding residues through hydrogen bonding, hydrophobic interactions, pi-pi stacking, pi-cation interactions, ionic interactions, etc.

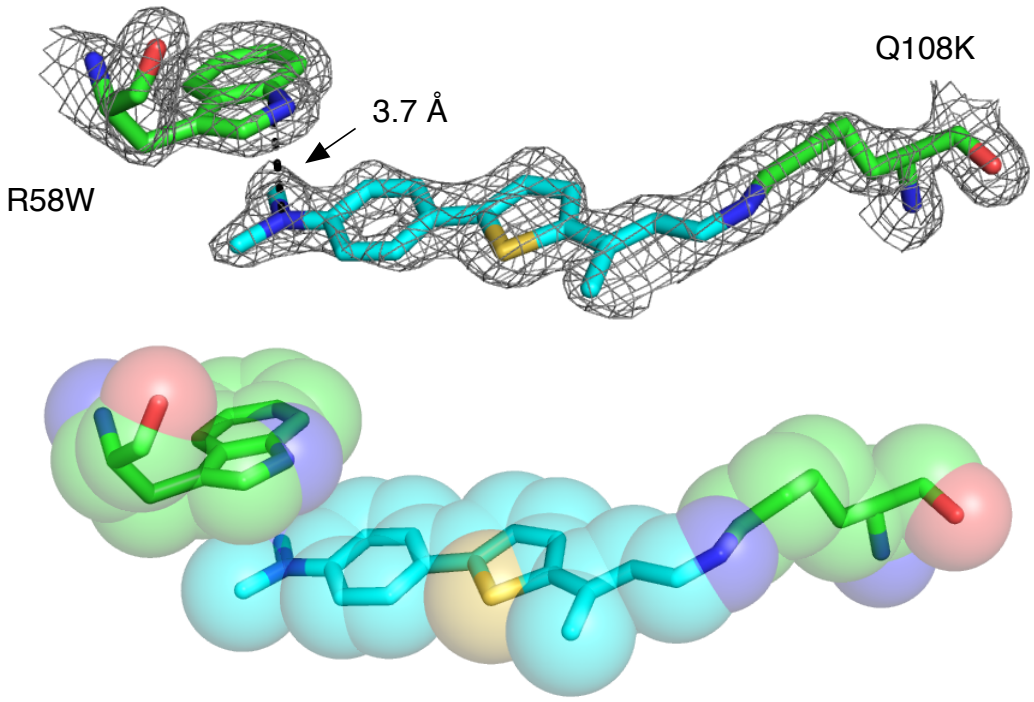


Figure II-20. Crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor** and space-filling representation. Electron density is shown at 1σ.

From the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor**, it is quite obvious that the tryptophan at position 58 does not serve to sequester the binding pocket, in contrast to when retinal is used as a ligand. R58W sits directly on top of the *N,N*-dimethyl amino moiety of the bound **ThioFluor**. As shown in **Figure II-20**, the closest distance between the tryptophan and nitrogen is 3.7 Å,

well within van der Waals' contact. This could suggest that **ThioFluor**'s amino group possesses a partial positive charge that can interact with R58W through an amino-aromatic stacking interaction, in which R58W serves as the hydrogen bonding acceptor. Similar interactions have been observed between the side chains of asparagine and glutamine with aromatic residues, where the amino group hydrogen atoms favor polar interactions with the pi-electrons of the aromatic. However, this interaction does not significantly affect wavelength, only a 6 nm hypsochromic shift from Q108K:K40L:T51V:T53S is obtained.

It was hypothesized that substitution of R58 for histidine, could result in a red-shift if a stronger hydrogen bonding interaction between the positively charged histidine side chain and *N,N*-dimethyl amino moiety of the ligand was introduced. Fortuitously, Q108K:K40L:T51V:T53S:R58H shows a 16 nm shift in absorption and 8 nm shift in emission from Q108K:K40L:T51V:T53S. Unfortunately, crystallization of Q108K:K40L:T51V:T53S:R58H was not fruitful. Based on *in silico* modeling of R58H in the crystal structure of Q108K:K40L:T51V:T53S:R58W with **ThioFluor**, it is expected that histidine adopts a similar conformation as for tryptophan. Histidine could potentially act as a hydrogen bond donor to the *N,N*-dimethyl amino moiety. Localization of charge on the tail end of the bound chromophore would promote positive charge delocalization from the iminium, resulting in a red-shift in wavelength, as was observed upon the introduction of R58H to Q108K:K40L:T51V:T53S.

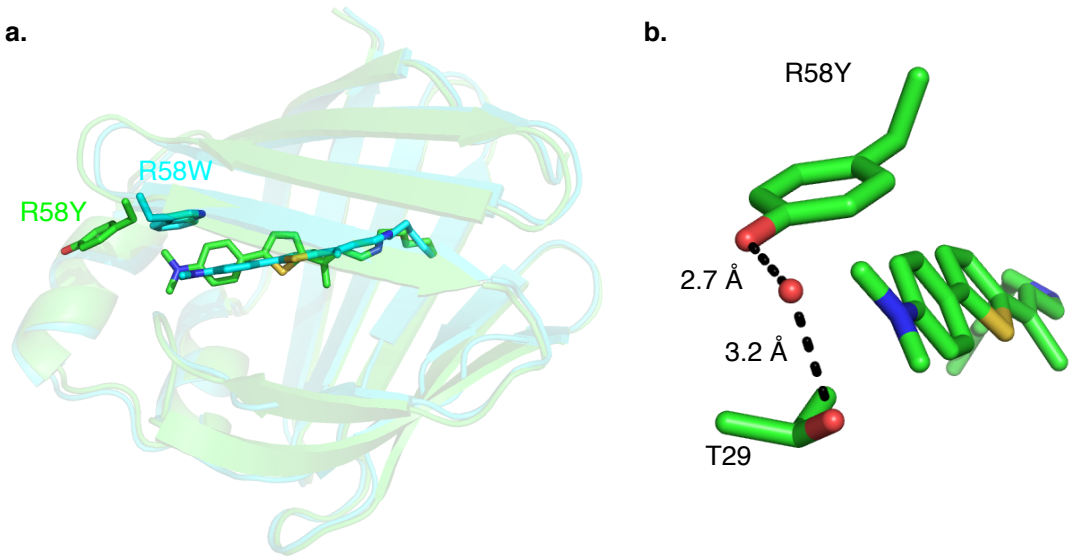


Figure II-21. **a**. Crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor** (cyan) overlaid with Q108K:K40L:T51V:T53S:R58Y/**ThioFluor** (green). **b**. Water mediated hydrogen bonding between R58Y and T29.

Next, R58F and R58Y were introduced into the same parent template Q108K:K40L:T51V:T53S. Both mutations actually led to a blue-shift in the absorption profile of the hCRBPII/**ThioFluor** complex (**Table II-11**). The crystal structure of Q108K:K40L:T51V:T53S:R58Y/**ThioFluor** shows a vastly different trajectory of residue 58 and chromophore orientation than that observed for Q108K:K40L:T51V:T53S:R58W/**ThioFluor** (**Figure II-21a**). As aforementioned, R58W flips in to stack with the *N,N*-dimethyl moiety of the ligand. However, tyrosine does not participate in aromatic interactions as often as tryptophan, and that is the case here. Tyrosine is flipped out of the pocket, providing no stacking interaction with the ligand. Alternatively, the phenolic group of R58Y participates in a water-mediated hydrogen bond with the hydroxyl side chain of T29 (**Figure II-21b**). As the ligand is not 'held' in place by the R58 mutation, this presumably causes the translocation and rotation of the chromophore. These changes, and

the fact that the binding pocket is no more sequestered, may be responsible for the negligible effect on wavelength.

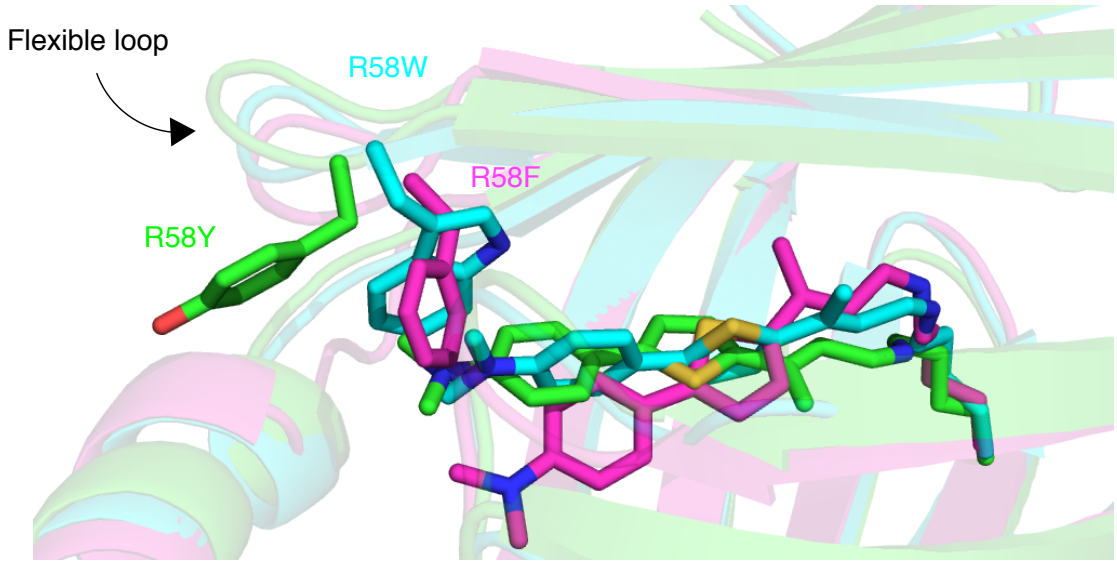


Figure II-22. Crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor** (cyan) overlaid with Q108K:K40L:T51V:T53S:R58Y/**ThioFluor** (green) and Q108K:K40L:T53A:R58F/**ThioFluor** (pink).

The crystal structure of Q108K:K40L:T51V:T53S:R58F/**ThioFluor** was not obtained, but Ms. Zahra Assar provided the structure of the similar variant Q108K:K40L:T53A:R58F with **ThioFluor**. Most notably, the flexibility of the loop at the entrance of the binding pocket was revealed, upon overlay of the structure with both Q108K:K40L:T51V:T53S:R58W/**ThioFluor** and Q108K:K40L:T51V:T53S:R58Y/**ThioFluor** (**Figure II-22**). In the R58F mutation, the loop moves such that the phenylalanine is within van der Waals contact of the ligand. A slight bathochromic shift (4 nm) in absorption is observed for Q108K:K40L:T53A:R58F from Q108K:K40L:T53A, suggesting again that R58F does not sequester the binding cavity. This suggestion is further supported by looking at the surface of the protein, which again shows a large hole at the binding cavity's entrance.

After exhaustive screening of R58 for aromatic residues, we chose to explore the effect of other residues including mutation to glycine, leucine, proline, glutamine, and glutamic acid in the same parent template Q108K:K40L:T51V:T53S. As shown in **Table II-12**, mutation of R58 to the nonpolar aliphatic glycine and leucine has essentially no effect on the wavelength of absorption (**Table II-12**, entries 1 and 3). Because arginine is polar and charged, it is likely that R58 is exposed to the aqueous exterior of the protein, and projects minimal electrostatic potential onto the chromophore. It would then be expected that the removal of the arginine would not significantly affect wavelength, as is observed for mutation to leucine and glycine.

Table II-12. Spectroscopic changes as a result of mutating R58 in Q108K:K40L:T51V:T53S.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a	$t_{1/2}{}^b$ (min)
1	Q108K:K40L:T51V:T53S:R58G	621	699	0.09	249
2	Q108K:K40L:T51V:T53S	627	697	0.15	146
3	Q108K:K40L:T51V:T53S:R58L	629	697	0.14	205
4	Q108K:K40L:T51V:T53S:R58Q	632	698	0.13	225
5	Q108K:K40L:T51V:T53S:R58P	636	703	0.14	168
6	Q108K:K40L:T51V:T53S:R58E	636	700	0.14	240

[a]Absolute quantum yield was measured on a Quantaurus-QY. [b]Half-lives are based on the rate constant obtained from pseudo first order fitting; measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7.

On the other hand, in the case of mutating R58 to the shorter polar uncharged glutamine and polar charged glutamic acid residues, bathochromic shifts in wavelength are observed from the parent Q108K:K40L:T51V:T53S.

R58Q and R58E mutations result in 5 nm and 9 nm, respectively (**Table II-12**, entries 4 and 6). It is possible that because these residues are shorter than arginine, and thus can flip into the pocket, they may have a higher propensity to interact with the ligand. In this case, polar residues could result in a similar orientation of the chromophore as that in Q108K:K40L:T51V:T53S:R58W, in which residue 58 is set up to interact with the *N,N*-dimethyl amino group of **ThioFluor**. The hydrogen bonding interaction of glutamine's amino hydrogen atoms or glutamic acid's carboxylic acid hydrogen atom (assuming the side chain is protonated) with the *N,N*-dimethyl amino group of **ThioFluor** would again promote positive charge delocalization along the chromophore. Nonetheless, further crystallographic analyses are required in order to test these hypotheses.

Similar to R58E, mutation of R58 to proline results in a 9 nm red-shift in wavelength (**Table II-12**, entries 5), even though glutamic acid is much more polar than proline. It was previously shown in the crystal structure of Q108K:K40L:T53A:R58F/**ThioFluor** that the loop at the entrance of the binding cavity is flexible; it is likely that introduction of the rigid proline alters the loop's location and potentially results in a different orientation of the chromophore. It is known that the distinctive cyclic structure of proline's side chain provides conformational rigidity compared to other amino acids. In these studies, this theory may be supported by the rate of iminium formation in the Q108K:K40L:T51V:T53S:R58X series (fitting was based on pseudo first order rate kinetics as the product concentration cannot be accurately determined;

measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7). While many factors can influence the rate of iminium formation, it is fastest in the case of the rigidifying R58P mutation and slowest when the bulky tryptophan residue is introduced at the mouth of the binding cavity (**Table II-11** and **Table II-12**). Presumably, in the case of R58P, the loop is permanently out of the way, leading to a more accessible binding cavity, and thus a faster rate of iminium formation.

It can be concluded that R58 mutations do not significantly affect wavelength, in contrast to that seen for hCRBPII/retinal complexes; it appears that a residue at this position may have the propensity to interact with the ligand, due to the presence of an *N,N*-dimethyl amino moiety in its vicinity. As aforementioned, as observed in the crystal structure of Q108K:K40L:T51V:T53S:R58W, the tryptophan does not serve to further sequester the binding pocket, but appears to interact with the nitrogen atom of the chromophore, through a stacking amino-aromatic interaction. In the same context, R58Y again does not sufficiently enclose the binding pocket. However, there is a rotation of the chromophore, which could also affect the complexes absorption and emission profiles. Further crystallographic analysis would need to be conducted in order to determine the effect of other R58 mutations on the spectroscopic properties of the complexes, as they are shown to dramatically alter the chromophore orientation. Nonetheless, it was hypothesized that no mutation at R58 will be capable of sequestering the binding cavity, in order to

bathochromically shift the wavelength. Thus, we chose alternate positions for mutagenesis, in hopes to achieve the same effect.

II.7 Removal of water and extensive packing of the chromophore leads to an enhanced red-shift in absorption and emission

As shown in the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor**, there is a conserved water network from the hydroxyl group of Y19 to the thiophene sulfur atom of **ThioFluor** (**Figure II-23a**). As removal of polarity was previously shown to red-shift wavelength it was predicted that mutation of Y19, to a residue less likely to provide this hydrogen bonding would result in removal of this water network. Tryptophan was introduced at Y19 for this purpose. A red-shift of 30 nm is observed with the introduction of Y19W to Q108K:K40L:T51V:T53S:R58W (**Table II-13**, entry 2).

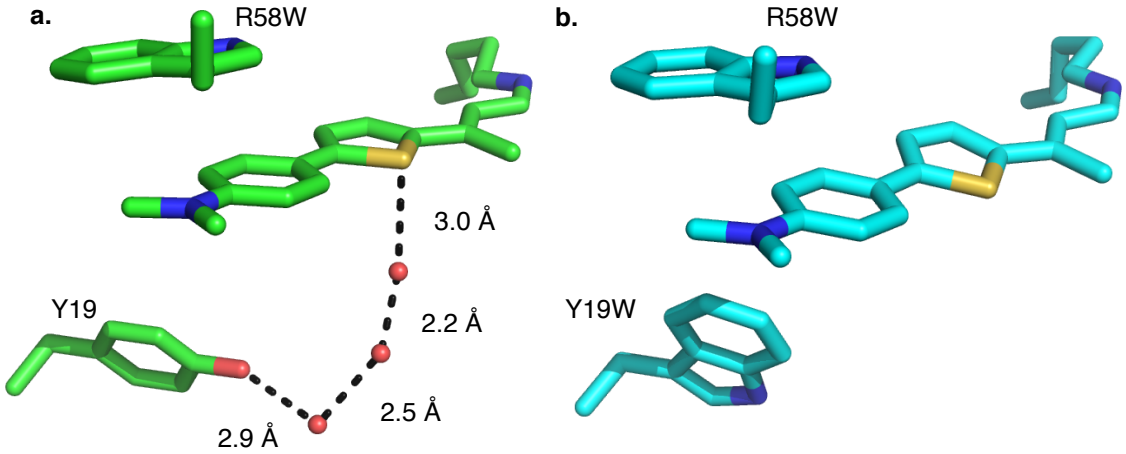


Figure II-23. **a**. Crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor**, highlighting the water mediated network from Y19 to **ThioFluor**. **b**. Crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor** shows this water network is abolished.

The crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor** (obtained by Ms. Zahra Assar) showed that substitution of tryptophan at Y19 did indeed abolish the water network present in Q108K:K40L:T51V:T53S:R58W/**ThioFluor** (**Figure II-23b**), with no other significant changes of the chromophore's orientation or any of the other residues. However, there could be two reasons for this red-shift in wavelength. It could be solely due to the removal of the water network localizing polarity on the thiophene core or to the replacement of this water network with a polarizable residue, which would more efficiently delocalize charge along the ligand.

Table II-13. Spectroscopic changes as a result of Y19W mutation.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L:T51V:T53S:R58W	623	697	0.08
2	Q108K:K40L:T51V:T53S:R58W:Y19W	653	719	0.10
3	Q108K:K40L:T51V:T53S:R58Y	622	691	0.11
4	Q108K:K40L:T51V:T53S:R58Y:Y19W	635	698	0.16
5	Q108K:K40L:T51V:T53S:R58H	643	705	0.14
6	Q108K:K40L:T51V:T53S:R58H:Y19W	642	705	0.15

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

The effect of the Y19W mutation was also tested in the templates with R58Y and R58H (**Table II-13**, entries 4 and 6). A significantly smaller red-shift of 13 nm was observed in adding Y19W to Q108K:K40L:T51V:T53S:R58Y, and addition of Y19W to Q108K:K40L:T51V:T53S:R58H unfortunately led to no change. For this reason, Q108K:K40L:T51V:T53S:R58W:Y19W was utilized for further protein engineering in an attempt to red-shift absorption and emission wavelengths. Introduction of four mutations to the parent KL protein

(Q108K:K40L:T51V:T53S:R58W:Y19W, λ_{abs} = 653 nm, λ_{em} = 719 nm) leads to the largest change in absorption (Δ_{abs} = 73 nm) and emission (Δ_{em} = 45 nm) from Q108K:K40L (λ_{abs} = 580 nm, λ_{em} = 674 nm) thus far.

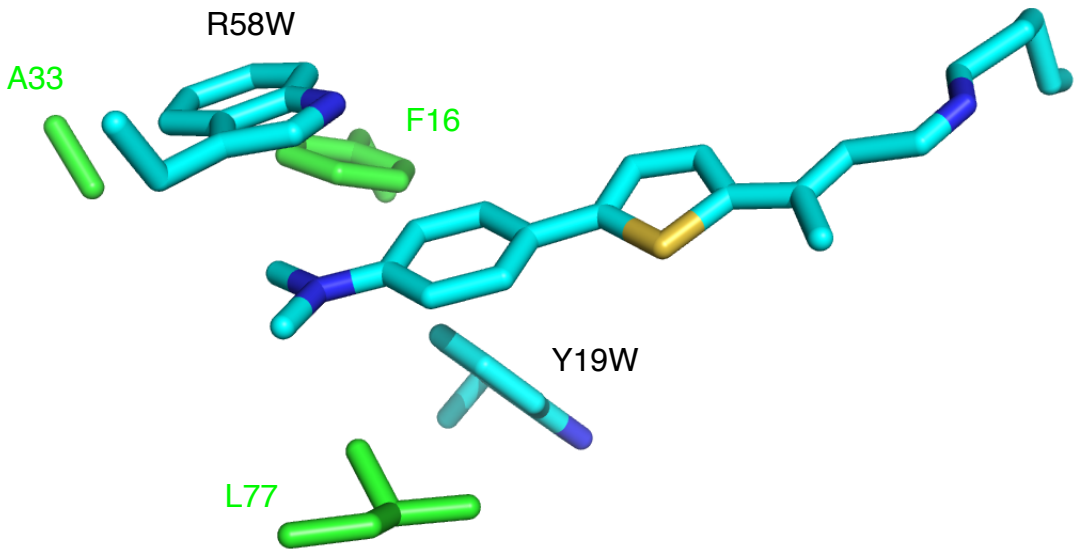


Figure II-24. Residues mutated (shown in green) in an attempt to encapsulate the binding cavity. Crystal structure is of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor**.

With the red-shifted hexamutant in hand, we sought to encapsulate the binding cavity via the introduction of a bulky tryptophan residue at L77, A33 and F16 to Q108K:K40L:T51V:T53S:R58W:Y19W (**Figure II-24**). Fortuitously, both L77W and A33W mutations red-shift the wavelength, 5 nm and 13 nm, respectively (**Table II-14**, entries 3 and 4). Unfortunately, crystallization of Q108K:K40L:T51V:T53S:R58W:Y19W:A33W was not fruitful. Presumably though, A33W sufficiently covers the binding cavity (see **Figure II-26** for a similar mutant crystallized with **ThioFluor**).

While direct encapsulation of the pocket was presumably achieved through the introduction of tryptophans at L77 and A33, located at the interface between the aqueous media and protein cavity, it was also suspected that introducing a bulky residue near R58W could force this residue to flip out and result in a similar effect. F16 is located at the interior pocket, but it can be envisioned that mutation of F16 to tryptophan would lead to steric clash of F16W with R58W, and may be able to flip R58W outward, in effect sequestering the binding pocket. Unfortunately, room temperature expression of Q108K:K40L:T51V:T53S:R58W:Y19W:F16W did not result in soluble protein. Conversely, Q108K:K40L:T51V:T53S:R58W:F16Y did provide soluble protein, perhaps because tyrosine is more similar to the wild-type phenylalanine than tryptophan. A red-shift from Q108K:K40L:T51V:T53S:R58W:Y19W of 11 nm was observed via the introduction of F16Y (**Table II-14**, entry 2). The possibilities for this red-shift are two-fold. It is possible that R58W is flipped outward as a result of introducing the longer tyrosine at F16, or the tyrosine phenol may serve as a hydrogen bond acceptor to the *N,N*-dimethyl amino moiety of **ThioFluor**.

Table II-14. Spectroscopic changes as a result of introducing bulky residues at the entrance of the binding cavity.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L:T51V:T53S:R58W:Y19W	653	719	0.10
2	Q108K:K40L:T51V:T53S:R58W:Y19W:F16Y	664	726	0.05
3	Q108K:K40L:T51V:T53S:R58W:Y19W:L77W	658	711	0.10
4	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W	666	724	0.14

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

Because the introduction of polarizable tryptophan residues at R58 and Y19 lead to a red-shift, from Q108K:K40L:T51V:T53S (λ_{abs} = 627 nm, λ_{em} = 697 nm) to Q108K:K40L:T51V:T53S:R58W:Y19W (λ_{abs} = 653 nm, λ_{em} = 719 nm), we chose to investigate the effect of tryptophan in other positions. It was also predicted that these residues might provide tight packing of the chromophore, which should lead to chromophore rigidification and restriction. To this end, tryptophan mutations were singly made at each residue in the protein cavity using the template Q108K:K40L:T51V:T53S:R58W:Y19W:A33W, along the entire chromophore's length.

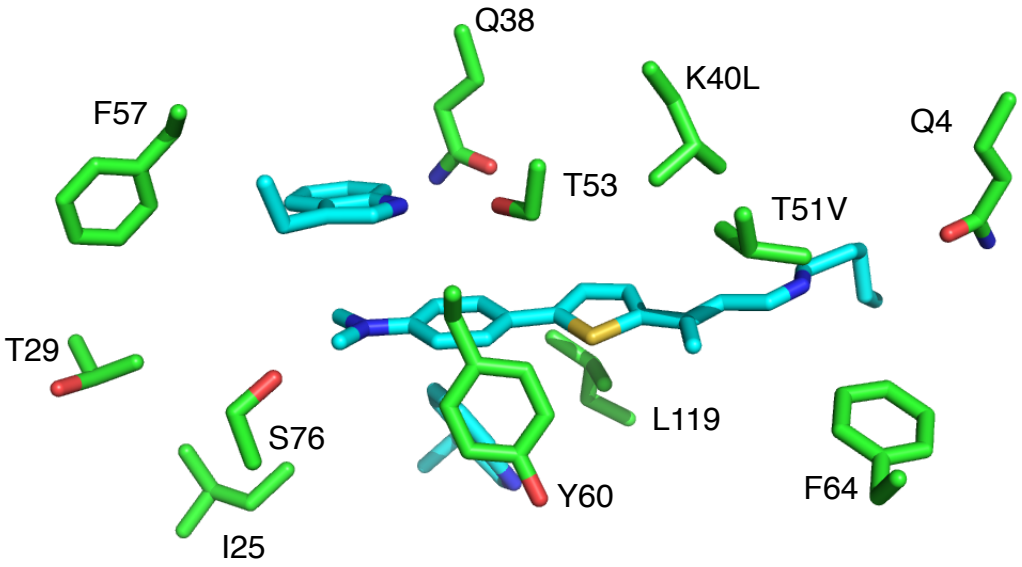


Figure II-25. Residues where tryptophan was introduced in an attempt to red-shift wavelength are shown in green. Crystal structure is of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor**.

Residues mutated in this study (**Figure II-25**) were T51, K40, Q38, I25, Y60, S76, L119, T53, T29, F57, F64 and Q4. Mutation of residues located at the mouth of the binding cavity, including S76, T29 and F57 to tryptophan led to negligible changes in the wavelength of the iminium (**Table II-15,** entries 7, 10

and 11). Presumably, the binding cavity in the parent mutant Q108K:K40L:T51V:T53S:R58W:Y19W:A33W is completely sequestered, rendering the introduction of new tryptophans at the entrance to the binding cavity as obsolete.

Table II-15. Spectroscopic changes a result of mutating active site residues to tryptophan.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L:**T51W**:T53S:R58W:Y19W:A33W	603	689	nd
2	Q108K:**K40H**:T51V:T53S:R58W:Y19W:A33W	612	694	0.19
3	Q108K:**K40W**:T51V:T53S:R58W:Y19W:A33W	627	701	0.17
4	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:**Q38W**	641	712	0.10
5	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:**I25W**	646	711	0.09
6	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:**Y60W**	661	718	0.10
7	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:**S76W**	665	721	0.14
8	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:**L119W**	665	729	nd
9	Q108K:K40L:T51V:**T53W**:R58W:Y19W:A33W	666	724	nd
10	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:**T29W**	666	726	0.12
11	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:**F57W**	667	722	0.12
12	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:**F64W**	669	727	0.12
13	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:**Q4W**	705	744	nd

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

Only I25W has an effect on the wavelength; Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:I25W is 20 nm blue-shifted in comparison to Q108K:K40L:T51V:T53S:R58W:Y19W:A33W (**Table II-15,** entry 5). While efforts to crystallize Q108K:K40L:T51V:T53S:R58W:Y19W:A33W/**ThioFluor** were not successful, a crystal structure of

Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L117E/**ThioFluor** was achieved. Assuming, the two structures are similar, it is apparent that the introduction of a large tryptophan residue at residue 25 would force the conformational change of a neighboring residue to avoid steric clashing (**Figure II-26**). Indeed, upon *in silico* mutation of I25 to tryptophan using the crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L117E/**ThioFluor** as a template, all rotamers of I25W sterically clash with A33W. This may force A33W to adopt a different conformation, negating the sequestering effect it delivered before.

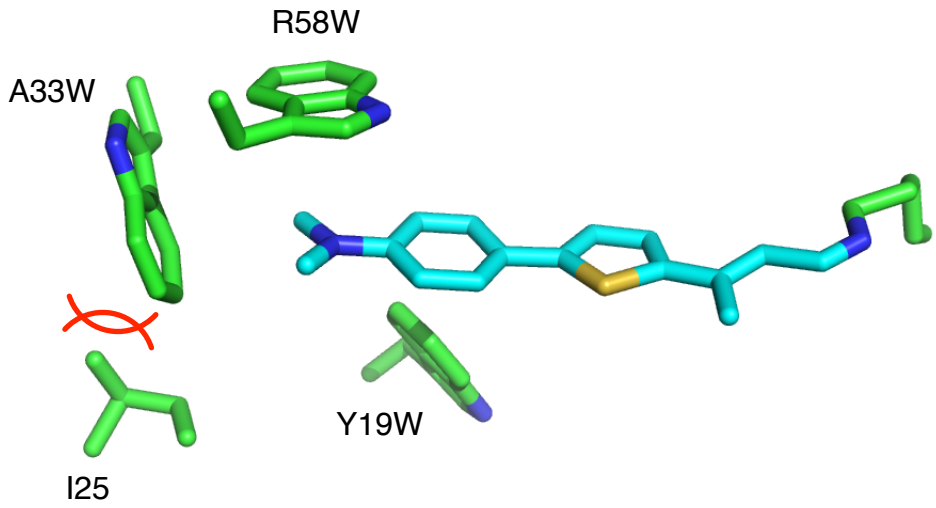


Figure II-26. Crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L117E/**ThioFluor**. Mutation of I25 to tryptophan would sterically clash with A33W.

Mutation of Q38, Y60, K40 and T51 to tryptophan each led to a blue-shift in absorption wavelength of 25 nm, 5 nm, 39 nm and 63 nm from the parent mutant Q108K:K40L:T51V:T53S:R58W:Y19W:A33W, respectively (**Table II-15**, entries 4, 6, 3 and 1). Interestingly, the most hypsochromic shifts were observed via introduction of tryptophan at the residues closest to the iminium, at positions

K40 and T51 (the same result was also achieved via introduction of histidine at K40; **Table II-15**, entry 2). Presumably, the introduction of a polarizable tryptophan in close proximity to the iminium localizes the charge, leading to a blue-shift in wavelength.

While some exchanges led to significant hypsochromic shifts in wavelength, numerous substitutions including tryptophan substitution at L119, T53 and F64, had a negligible effect on the wavelength (**Table II-15**, entries 8, 9 and 12). As these residues are at the interior of the protein pocket and size was significantly altered for each point mutation, without crystal structures it is not possible to explain the null effect on wavelength. As a result of increased steric bulk, it is possible that the chromophore has changed conformation, or the changes have impeded the interaction of other residues with the chromophore as previously described.

The only substitution of tryptophan that led to a significant bathochromic shift in wavelength was at Q4 (**Table II-15**, entry 13). Intuitively, the red-shift would be attributable to the removal of polarity by removing the water mediated network between Q4 and the iminium. However, there is another possibility. From previous studies with hCRBPII/retinal it was observed that the removal of the water mediated hydrogen bond from Q4 to a residue with a non-hydrogen bonding side chain leads to a thermodynamically stable *trans* isomer of the iminium, rather than the *cis* iminium. It is quite possible that the addition of Q4W to Q108K:K40L:T51V:T53S:R58W:Y19W:A33W also leads to a *trans* isomer of

the iminium. This is supported by the fact that Q4W dramatically alters the pK_a of the iminium. Q108K:K40L:T51V:T53S:R58W:Y19W:A33W/**ThioFluor** exhibits a high pK_a of 9.3. Conversely, at a neutral pH of 7, no iminium species is apparent by UV-Vis for Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:Q4W/**ThioFluor**.

After exhaustive tryptophan screening, one last effort was undertaken to red-shift the absorption and emission profiles of the hCRBPII/**ThioFluor** mutants through combining residues that red-shifted wavelength individually. To this end, a few mutants were made, as detailed in **Table II-16**. The most red-shifted hCRBPII mutant coupled with **ThioFluor** was found to be Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:Q4W:F16Y with an absorption maxima of 705 and emission maxima of 744 nm. From Q108K:K40L (λ_{abs} = 580 nm, λ_{em} = 674 nm), a change of 125 nm is obtained in absorption and 70 nm in emission.

Table II-16. Additive effect of F16Y, leading to a final bathochromic shift in wavelength.

Entry	Mutant[a]	λ_{abs}	λ_{em}	Φ[b]
1	KL:T51V:T53S:R58W:Y19W:A33W:F16Y	670	735	0.06
2	KL:T51V:T53S:R58W:Y19W:L77W:F16Y	672	723	0.06
3	KL:T51V:T53S:R58W:Y19W:A33W:Q4W:F16Y	705	744	nd

[a]KL = Q108K:K40L. [b]Absolute quantum yield was measured on a Quantaurus-QY.

II.7.1 Investigation of combinatorial residue effects on the wavelength of Q108K:K40L:T51V:T53S:R58W:Y19W

As previously discussed, the introduction of four mutations to the parent KL protein (Q108K:K40L:T51V:T53S:R58W:Y19W, λ_{abs} = 653 nm, λ_{em} = 719 nm)

leads to a red-shift in absorption (Δ_{abs} = 73 nm) and emission (Δ_{em} = 45 nm). By mutating residues back to their wild type, we sought to investigate which residues are the most necessary to red-shift wavelength.

Interestingly, when introduced to Q108K:K40L, the addition of T51V and T53S (**Table II-17**, entry 4) yields a red-shift of 47 nm, while the addition R58W and Y19W (**Table II-17**, entry 1) leads to only a 25 nm red-shift in absorption. Introduction of R58W or Y19W individually to Q108K:K40L:T51V:T53S do not substantially alter the wavelength of absorption (**Table II-17**, entries 3 and 5). However, their combination leads to the red-shifted hexamutant Q108K:K40L:T51V:T53S:R58W:Y19W (**Table II-17,** entry 7).

Table II-17. Deducing the origins of the bathochromic shift in Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor**.

Entry	Mutant						λ_{abs}	λ_{em}	Φ^a
1	Q108K:	K40L:			R58W:	Y19W	605	691	0.14
2	Q108K:	K40L:	T51V:		R58W:	Y19W	612	691	0.07
3	Q108K:	K40L:	T51V:	T53S:	R58W		623	697	0.08
4	Q108K:	K40L:	T51V:	T53S			627	697	0.15
5	Q108K:	K40L:	T51V:	T53S:		Y19W	628	699	0.12
6	Q108K:	K40L:		T53S:	R58W:	Y19W	635	704	0.16
7	Q108K:	K40L:	T51V:	T53S:	R58W:	Y19W	653	719	0.10

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

The crystal structures of Q108K:K40L:T51V:T53S:R58W/**ThioFluor** and Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor** may provide insight into the origin of this synergistic combination that leads to a red-shift in wavelength. In the crystal structure, two conserved hydrogen bonding water mediated networks are

apparent. First, the hydroxyl of T53S hydrogen bonds with both the indole side chain of R58W and the amide moiety of Q38 through a water mediated hydrogen bond (**Figure II-27**). Secondly, Y19 interacts with the sulfur atom of **ThioFluor** through an extensive water network (**Figure II-23a**). These interactions presumably dictate the chromophore's orientation.

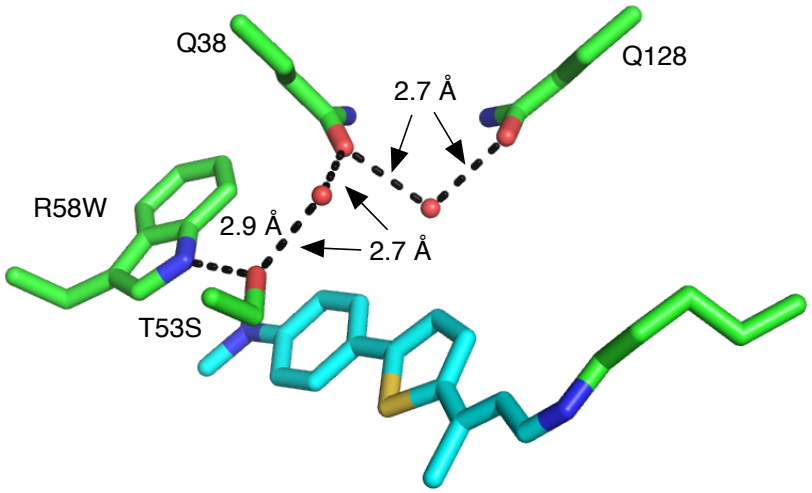


Figure II-27. Crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor**. The water mediated hydrogen bonding between T53S and R58W is highlighted.

As seen in the crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W/ **ThioFluor** and described previously (**Section II.7**, **Figure II-23**), the introduction of Y19W abolishes the water network from Y19 to the sulfur atom of **ThioFluor**, presumably leading to a red-shift in wavelength. It is predicted that in the absence of the R58W mutation, **ThioFluor** cannot adopt an orientation in which the Y19W mutation is able to eliminate this water network.

II.8 Localization of charge leads to a blue-shift in absorption and emission

In studies with hCRBPII/retinal, Dr. Wenjing Wang previously demonstrated that while an even distribution of electrostatic potential across the

polyene led to a bathochromic shift in absorption, it was localizing the cation on the iminium nitrogen that led to a blue-shift.[3] Similarly, we sought to introduce negatively polarized residues near the iminium.

It is widely accepted that hydrophobic interactions can influence protein folding and stability.[63-65] While tryptophans were introduced with relative ease into the hCRBPII sequence, it was more of a challenge to substitute acidic residues for otherwise non-charged or less polar residues. The primary challenges observed in introducing acidic residues were the ability to obtain soluble protein during protein expression, protein stability and the formation of domain-swapped dimers.

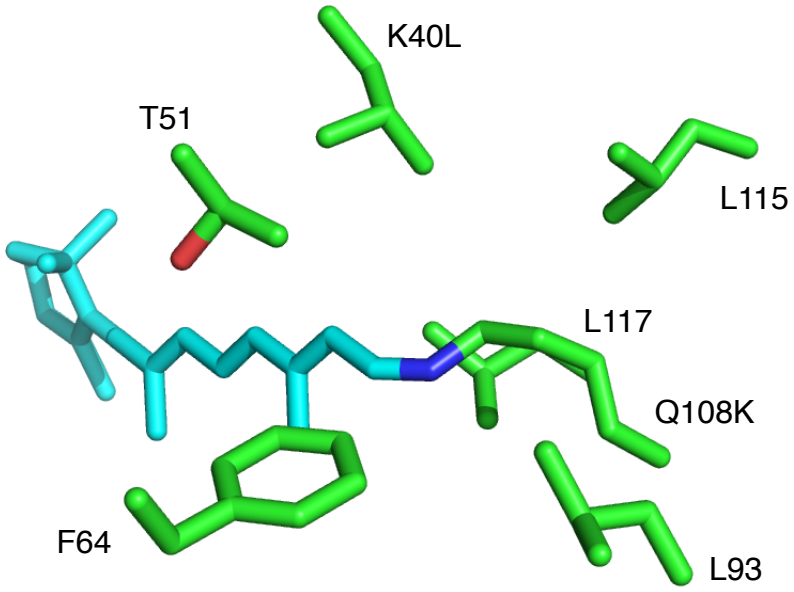


Figure II-28. Residues chosen for introduction of an acidic residue, in order to interact with the iminium, based on the crystal structure of Q108K:K40L/retinal. Coordinates obtained from PDB 4EXZ.

We chose to start again from Q108K:K40L, introducing acidic residues which should reside in close proximity to the iminium, namely at residues K40,

T51, L117, L115, F64 and L93 (**Figure II-28**). Q108K:K40L:L115E, Q108K:K40L:F64E and Q108K:K40L:L93E led to no soluble protein expression, so exploration at these positions were not followed. Incubation of **ThioFluor** with Q108K:K40L:L117E, Q108K:K40E and Q108K:K40D led to no substantial change when monitored by UV-Vis; only precipitation was apparent. Only Q108K:K40L:T51E yielded an observable iminium species(λ_{abs} = 592 nm, λ_{em} = 685 nm). Interestingly, a red-shift of 12 nm from Q108K:K40L (λ_{abs} = 580 nm, λ_{em} = 674 nm) was observed. This proves counterintuitive, as the introduction of polarity near the iminium is expected to result in a blue-shift in wavelength.

It was hypothesized that the introduction of an acidic residue near the basic active site lysine may cause the formation of a salt bridge, preventing attack of the lysine to ligand to form an iminium. To test this, Q108K:K40E was incubated with retinal. PSB forms quickly, and further acidification results in a clear iminium species. As retinal is capable of forming the iminium with Q108K:K40E, it was assumed that **ThioFluor** should be able to also. Therefore, it was hypothesized that the pK_a of the protein/**ThioFluor** complexes are at least two units lower than physiological pH. However, iminium cannot be observed upon lowering the pH of the protein/**ThioFluor** complexes, due to the sensitivity of the protein to acid (only precipitation is observed in these cases).

As discussed above, we were not able to conclude anything based on the introduction of an acidic residue in the Q108K template. Thus, we chose to introduce an acidic residue into one of the red-shifted mutants, in order to test

whether a carboxylate can localize charge on the iminium, leading to a blue-shift in wavelength. To this end, L117 was mutated to both aspartic acid and glutamic acid in Q108K:K40L:T51V:T53S:R58W:Y19W and Q108K:K40L:T51V:T53S:R58W:Y19W:A33W, both of which when bound with **ThioFluor** produced a clearly observable red-shifted iminium. Fortuitously, the complexes were stable, with the complex of Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L117E surviving to a pH as low as 1. Noteworthy, large hypsochromic shifts were observed under circumstances when the carboxylate side chain is either protonated or deprotonated (these mutants will be discussed in more detail in **Chapter IV** for their application toward the development of a ratiometric pH sensor). This means that the introduction of a counter anion blue-shifts the wavelength more than simply just introducing polarity (consider L117E-COO^- versus L117E-COOH).

Table II-18. Spectroscopic changes observed upon the introduction of an acidic residue at L117.

Entry	Mutant	λ_{abs1}	λ_{abs2}[a]	Φ[b]
1	Q108K:K40L:T51V:T53S:R58W:Y19W	653		0.10
2	Q108K:K40L:T51V:T53S:R58W:Y19W:L117E	558	614	0.31
3	Q108K:K40L:T51V:T53S:R58W:Y19W:L117D	525	618	0.30
4	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W	666		0.14
5	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L117E	539	628	0.28
6	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L117D	510	645	0.16

[a]λ_{abs1} is absorbance maxima at basic pH and λ_{abs2} is absorbance maxima at acidic pH. [b]Absolute quantum yield was measured on a Quantaurus-QY.

With the knowledge that a blue-shift in wavelength can be achieved via the localization of charge on the iminium, we sought to design a protein template that

is similar in wavelength (or at least not significantly red-shifted) to Q108K:K40L and produces an iminium at physiological pH or is acid stable enough to observe the iminium by lowering the pH of the solution.

II.8.1 Design of blue-shifted hCRBPII/ThioFluor complexes that promote iminium formation

From the UV-vis of Q108K:K40L/**ThioFluor** it is quite obvious that the complex suffers from either a severely suppressed pK_a or extinction coefficient of the iminium. In preliminary studies, introduction of T53A or T53S led to a mix of hCRBPII monomer and dimer. Interestingly, the spectroscopic properties of these

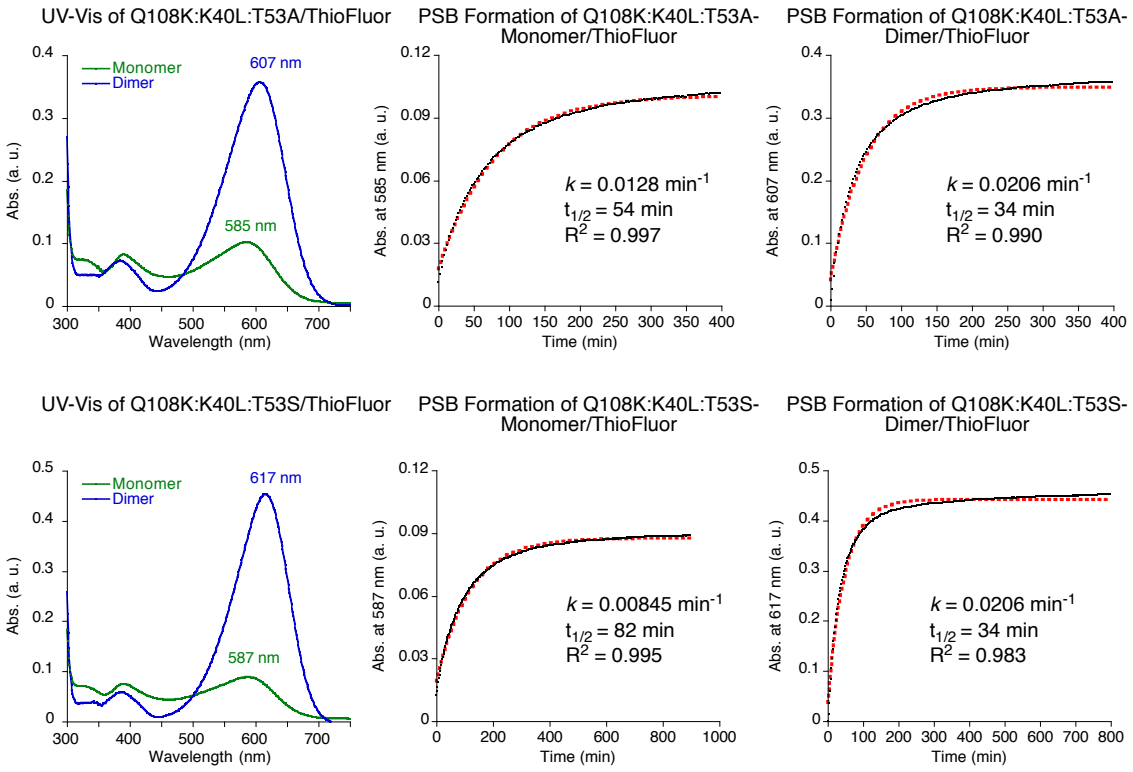


Figure II-29. Iminium formation of Q108K:K40L:T53A and Q108K:K40L:T53S monomers and dimers. From left to right are the UV-Vis of the thermodynamic protein/**ThioFluor** complex, iminium formation as a function of time for the monomer and dimer. Data is shown in black, while curve fitting is a dotted red line (fit to pseudo first order kinetics, measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7).

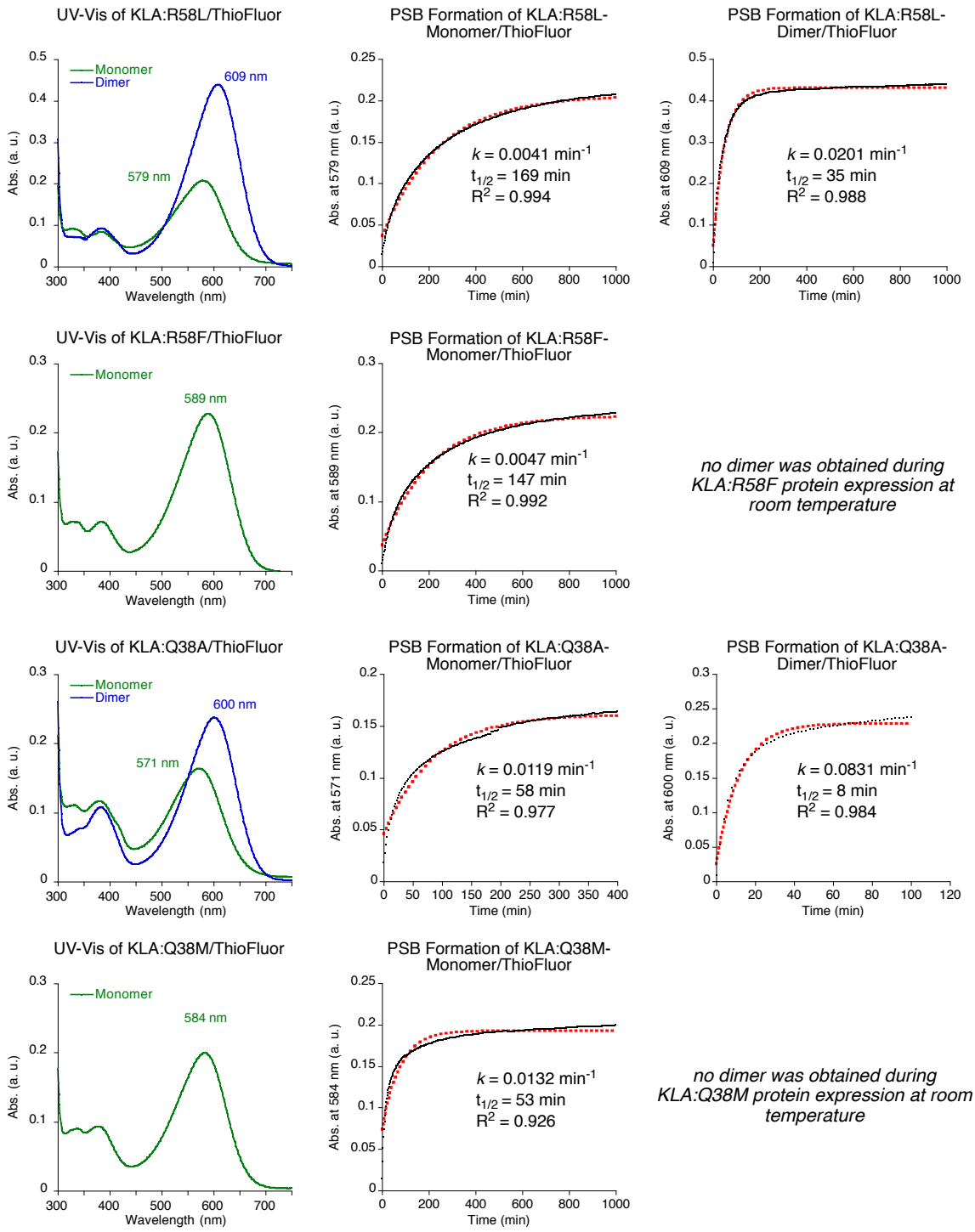


Figure II-30. Iminium formation of KLA:R58L(R58F) and KLA:Q38A(M) monomers and dimers. From left to right are the UV-Vis of the thermodynamic protein/**ThioFluor** complex, iminium formation as a function of time for the monomer and dimer. Data is shown in black, while curve fitting is a dotted red line (fit to pseudo first order kinetics, measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7). KLA = Q108K:K40L:T53A.

are vastly different. Under identical conditions, the dimers of Q108K:K40L:T53A and Q108K:K40L:T53S both led to a significantly larger amount of iminium formation than their associated monomers (**Figure II-29**); additionally, the absorption wavelength of the dimer is more red than that of the monomer for both proteins. Interestingly, a difference in the rates of iminium formation is apparent (**Figure II-29**), in which the dimer forms iminium substantially faster with **ThioFluor** for both Q108K:K40L:T53A and Q108K:K40L:T53S.

While these structures have not been obtained, it is likely that the chromophore's iminium experiences different local environments, leading to either a difference in extinction coefficient or pK_a. It should be noted that we did not measure pK_as or extinction coefficients, so the origin of difference between monomer and dimer cannot be determined at this time. This information may prove invaluable in future studies to develop differentially binding probes to hCRBPII monomers and dimers *in vivo*. It is hypothesized that the extent of iminium formation may be due to different conformations of the iminium (*cis* or *trans*), which has previously been shown to alter the iminium pK_a in both CRABPII and hCRBPII when bound to retinal. However, further crystallographic studies should be undertaken to understand the true difference between the holo-hCRBPII monomers and dimers.

From the Q108K:K40L:T53A (KLA) mutant, it was observed that further removal of polarity at R58 and Q38 led to a further increase in the extent of iminium formation, although the reason is not apparent at this time. We chose to

mutate R58 to the nonpolar residues leucine and phenylalanine. At the same time, Q38 was replaced with alanine, methionine and phenylalanine (unfortunately, introduction of Q38F led to low soluble protein expression). As seen in **Figure II-30**, both Q108K:K40L:T53A:R58L-monomer and Q108K:K40L:T53A:R58F-monomer lead to a more intense iminium species, as did Q108K:K40L:T53A:Q38M-monomer and Q108K:K40L:T53A:Q38A-monomer. It should be noted that the mutation of Q38 to both alanine and methionine led to a less stable protein, as the complex begins to precipitate after formation of the iminium with **ThioFluor**. Dimer/**ThioFluor** is also presented for comparison purposes. In all cases, the dimer/**ThioFluor** iminium is more red-shifted than the monomer/**ThioFluor** iminium. Additionally, the rate of iminium rate formation for the dimer is faster than the monomer.

Final combination of iminium promoting mutations (at residues 53, 58, and Q38) were introduced to yield Q108K:K40L:T53A:R58L:Q38M and Q108K:K40L:T53A:R58L:Q38F. Only monomer was obtained during protein expression at room temperature. R58L was retained in lieu of R58F because the complex of Q108K:K40L:T53A:R58L was more blue-shifted than Q108K:K40L:T53A:R58F by 10 nm. As seen in **Figure II-31**, both pentamutants produced a clearly observable iminium species for the monomer-bound **ThioFluor**. Fortuitously, Q108K:K40L:T53A:R58L:Q38F (λ_{abs} = 577 nm, λ_{em} = 671 nm). is of approximately the same wavelength as Q108K:K40L (λ_{abs} = 580

nm, λ_{em} = 674 nm), rendering the effect of the installed substitutions on the wavelength negligible as was desired (**Table II-19**).

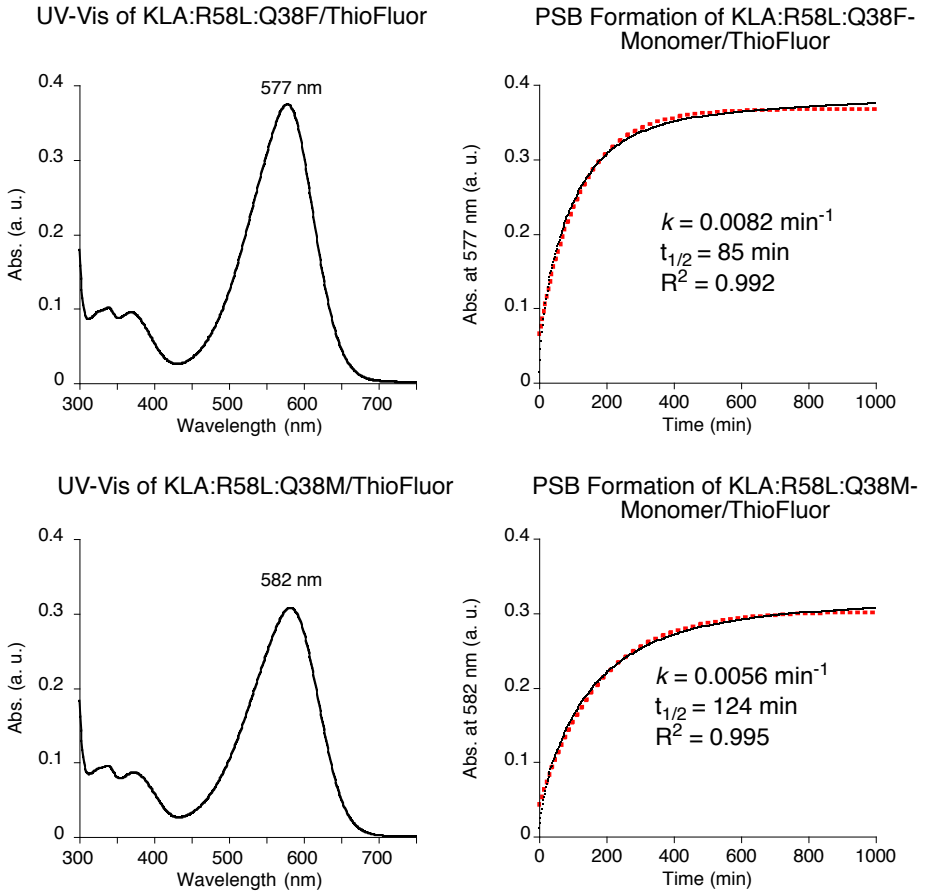


Figure II-31. Iminium formation of KLA:R58L:Q38F and KLA:R58L:Q38M monomer. From left to right are the UV-Vis of the thermodynamic protein/**ThioFluor** complex, iminium formation as a function of time. Data is shown in black, while curve fitting is a dotted red line (fit to pseudo first order kinetics, measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7). KLA = Q108K:K40L:T53A.

It should be noted that the introduction of Q128L, yielding Q108K:K40L:T53A:R58L:Q128L, provided the same effect, however, the protein was obtained with a lower expression yield. Presumably mutation of either Q38 or Q128 to a nonpolar residue removes the conserved water mediated network between the two residues. A large set of other mutations were tested, however,

for simplicity and with the goal of blue-shifting wavelength, we continued with the template Q108K:K40L:T53A:R58L:Q38F for further protein engineering. The spectroscopic properties of these mutants are reported in **Table II-19**. It should be reiterated that in order to understand the effect of these mutations, it would be essential to collect pK_a and extinction data on these mutants.

Table II-19. Spectroscopic property of protein mutants designed in order to increase iminium formation.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L:T53A:Q38A	571	667	0.14
2	Q108K:K40L:T53A:R58L:Q38A	575	669	0.13
3	Q108K:K40L:T53A:R58L:Q38F	577	671	0.08
4	Q108K:K40L:T53A:R58L:Q38F:Y19W	577	667	0.08
5	Q108K:K40L:T53A:R58L:Y19F	578	672	0.10
6	Q108K:K40L:T53A:R58L:Q38F:S76A	578	670	nd
7	Q108K:K40L:T53A:R58L:Q128L	579	670	0.10
8	Q108K:K40L:T53A:R58L	579	672	0.09
9	Q108K:K40L	580	674	nd
10	Q108K:K40L:T53A:R58L:Q38M	582	674	0.11
11	Q108K:K40L:T53A:R58L:A33W	584	672	0.06
12	Q108K:K40L:T53A:Q38M	584	677	nd
13	Q108K:K40L:T53A	585	677	0.11
14	Q108K:K40L:T53A:R58W	586	678	0.11
15	Q108K:K40L:T53V:R58L	587	678	0.16
16	Q108K:K40L:T53A:R58F	589	681	0.10
17	Q108K:K40L:T53A:R58L:Q38F:Y60F	590	675	0.19
18	Q108K:K40L:T53A:R58L:Q38F:F16Y	594	677	0.13
19	Q108K:K40L:T51C:T53A:R58L:Q38F	594	677	0.12
20	Q108K:K40L:T53A:R58L:Q38F:L77W	597	678	0.15

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

II.8.2 Exploring the additive effects of acidic residues on the absorption and emission wavelength of hCRBPII/ThioFluor

Having in hand a blue-shifted hCRBPII mutant capable of forming an iminium with **ThioFluor** at physiological pH, we chose to introduce acidic residues in an attempt to blue-shift its wavelength. It was quickly realized that the introduction of an acidic residue at residue 40 to Q108K:K40L:T53A:R58L:Q38F would not prove successful, as soluble protein was not obtained from either Q108K:K40D:T53A:R58L:Q38F or Q108K:K40E:T53A:R58L:Q38F.

Table II-20. Spectroscopic changes as a result of introducing one acidic residue near the iminium.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L:T53A:R58Y:Q38F:Q4F:**L117E**	510	653	0.20
2	Q108K:**K40D**:T53A:R58L:Q38F:Q4F	546	649	0.21
3	Q108K:**K40D**:T53A:R58Y:Q38F:Q4F	551	641	0.26
4	Q108K:**K40E**:T53A:R58L:Q38F:Q4F	575	671	0.15
5	Q108K:K40L:**T51D**:T53A:R58Y:Q38F:Q4F	576	665	0.25
6	Q108K:K40L:T53A:R58L:Q38F:Q4F:**V62E**	584	656	0.15
7	*Q108K:K40L:T53A:R58L:Q38F:Q4F*	*616*	*691*	*0.10*

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

Fortuitously though, our lab has observed the general trend that the Q4F mutation is capable of increasing protein expression yield while also resulting in a higher level of soluble protein. Thus, acidic residues were introduced to the Q108K:K40L:T53A:R58L:Q38F:Q4F hCRBPII hexamutant. To our delight soluble protein was obtained in most of the desired proteins. It should be noted that in some cases R58Y was included instead of R58L, as R58Y led to better soluble

protein expression. However, in comparing Q108K:K40D:T53A:R58L:Q38F:Q4F (**Table II-20**, entry 2) and Q108K:K40D:T53A:R58Y:Q38F:Q4F (**Table II-20**, entry 3), it is obvious that the substitution at R58 does not drastically alter the wavelength in these mutants, rendering them comparable in terms of spectroscopic properties.

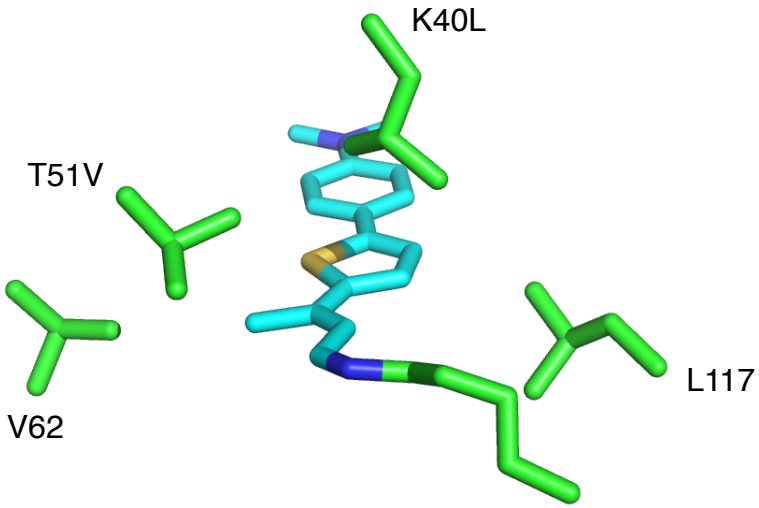


Figure II-32. Positions for introducing an acidic reside to blue-shift wavelength. Based on the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor**.

Acidic residues were introduced at K40, T51, V62 and L117 (**Figure II-32**). Results are detailed **Table II-20**. The introduction of L117E provides the largest hypsochromic shift from Q108K:K40L:T53A:R58L:Q38F:Q4F, an astounding 106 nm in absorption maxima and 38 nm in emission maxima.

The crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor** (R58W and Y19W seem to rigidify the chromophore, such that ligand density is well-ordered; a structure was not successfully obtained without these two tryptophans) was obtained and solved by Mr. Alireza Ghanbarpour. If the chromophore adopts the same

orientation in each mutant (and exists as the same isomer of the iminium), L117 would be closest to the iminium (**Figure II-33**); and V62 would be furthest from the iminium when mutated to glutamic acid (mutagenesis suggests the glutamic acid must be oriented away from the iminium to avoid steric clash with the methyl group). It would appear that the distance from the glutamic acid residue to the iminium is roughly correlated with how much of a hypsochromic shift is observed with its introduction. While crystallographic data would provide concrete evidence, it is hypothesized the larger shift is observed as the residue resides closer to the iminium or is involved in a tighter hydrogen bonding network with the iminium.

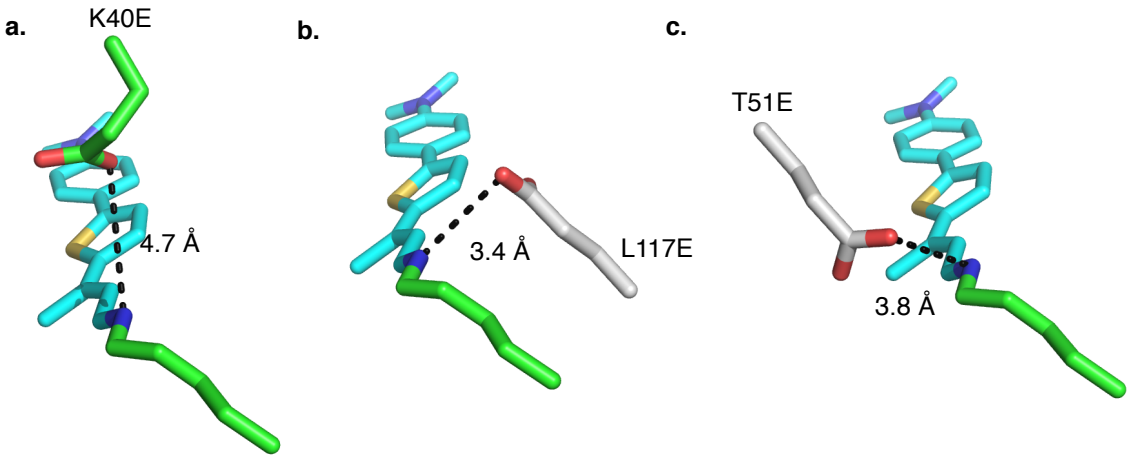


Figure II-33. **a**. Crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor**, highlighting the acidic residues tested in this series of protein mutants. **b**. – **c**. The residues L117 and K40 (shown in grey) were mutated to glutamic acid in pymol.

Additionally, higher quantum yields are obtained with the introduction of an acidic residue; every mutation has a higher quantum yield than that of the parent template. It could be that the quantum yield is higher solely based on the fact that these mutants are more blue-shifted than their parent protein. From the crystal

structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor**, it is also apparent that the residues are within distance from the iminium to interact via a water mediated network. It is possible that the presence of an acidic residue rigidifies **ThioFluor**, which could lead to an enhancement of quantum yield.

Table II-21. Spectroscopic changes as a result of introducing two acidic residues near the iminium.

Entry	Mutant	λ_{abs}	λ_{em}	Φ^a
1	Q108K:K40L:**T51D**:T53A:R58Y:Q38F:Q4F:**L117E**	501	613	0.49
2	Q108K:**K40D**:T53A:R58Y:Q38F:Q4F:**L117E**	505	653	0.13
3	Q108K:**K40E**:T53A:R58L:Q38F:Q4F:**V62E**	521	656	0.24
4	Q108K:**K40D**:T51V:R58L:Q38F:Q4F:**V62E**	528	621	0.65
5	Q108K:**K40D**:T51V:T53A:R58L:Q38F:Q4F:**V62E**	571	670	0.21
6	*Q108K:K40L:T53A:R58L:Q38F:Q4F*	*616*	*691*	*0.10*

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

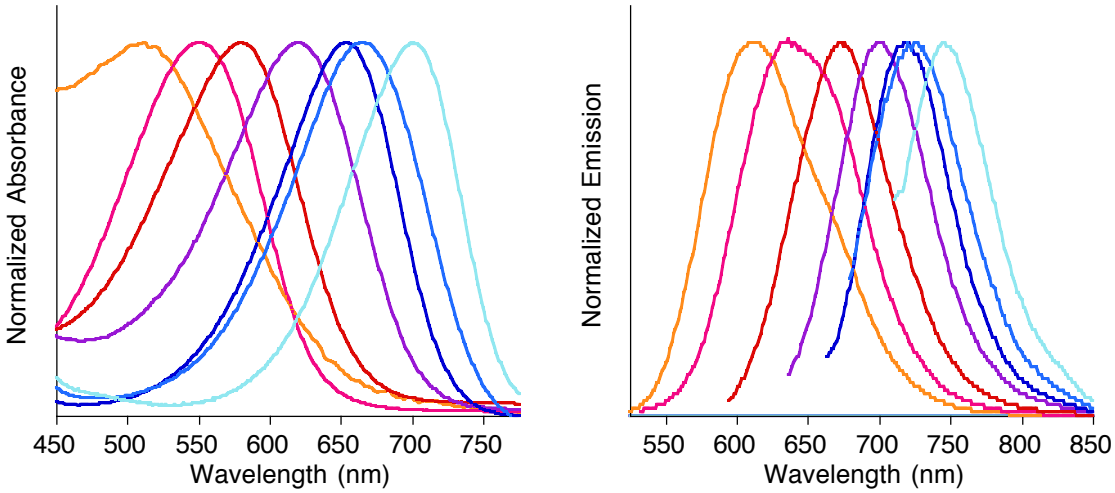
We next sought to investigate whether the introduction of two acidic residues concurrently could further shift the wavelength or increase the quantum yield. Results are shown in **Table II-21**. It should be noted that T51V was included in some mutants in order to obtain soluble monomer during protein expression (see **Chapter VI** for a more detailed description); Q108K:K40D:T53A:R58L:Q38F:Q4F:V62E produced only dimer at both room temperature expression, as well as at 16 °C (we know from previous studies that dimer formation is temperature dependent).[66] Similarly, the combination of K40D and T51D also led to mainly dimer formation during expression of Q108K:K40D:T51D:T53A:R58Y:Q38F:Q4F.

Again, inclusion of L117E led to the most blue-shifted mutants (**Table II-21**, entries 1 and 2). Interestingly, while absorption was not affected substantially, Q108K:K40L:T51D:T53A:R58Y:Q38F:Q4F:L117E was 40 nm blue-shifted in emission as compared to the mutant without T51D (**Table II-20**, entry 1). It is possible that we have reached the threshold at which the absorption can be varied.

Fortuitously, Q108K:K40L:T51D:T53A:R58Y:Q38F:Q4F:L117E (**Table II-21**, entry 1) and Q108K:K40D:T51V:R58L:Q38F:Q4F:V62E (**Table II-21**, entry 4) both show high quantum yields of 0.49 and 0.65, respectively, when coupled with **ThioFluor**. Interestingly, these mutants do not contain acidic residues in common. If it is the presence of the acidic residue that is responsible for increasing the quantum yield of the complex, one would venture to hypothesize that the chromophore binding in the two proteins is dissimilar, such that the iminium experiences different microenvironments. It is also possible that there is a purely synergistic combination (i.e. orientation and distance to the iminium). However, the origin of the large quantum yields cannot be deduced at this time; crystallization of these hCRBPII mutants with **ThioFluor** may provide this insight.

II.9 Emission is linearly correlated to absorbance

As was first stated at the opening of this chapter, our overall goal is to develop fluorescent proteins that span the entire visible spectra. **Figure II-34** depicts select hCRBPII mutants coupled with **ThioFluor**, clearly shows that we have achieved this goal. hCRBPII mutants bound with **ThioFluor** yielded mutants



$\lambda_{abs,max}$ = 501 - 705 nm
Δ_{abs} = 204 nm

$\lambda_{em,max}$ = 613 - 744 nm
Δ_{em} = 131 nm

Figure II-34. Range of absorption and emission achieved in this study by coupling hCRBPII monomers with **ThioFluor**. Mutants from left to right are (1) Q108K:K40L:T51D:T53A:R58Y:Q38F:Q4F:L117E (note: data was smoothed by averaging because of low absorbance intensity), (2) Q108K:K40D:T53A:R58Y:Q38F:Q4F, (3) Q108K:K40L:T53A:R58L, (4) Q108K:K40L:T51V:T53S:R58G, (5) Q108K:K40L:T51V:T53S:R58W:Y19W, (6) Q108K:K40L:T51V:T53S:R58W:Y19W:F16Y, (7) Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:Q4W.

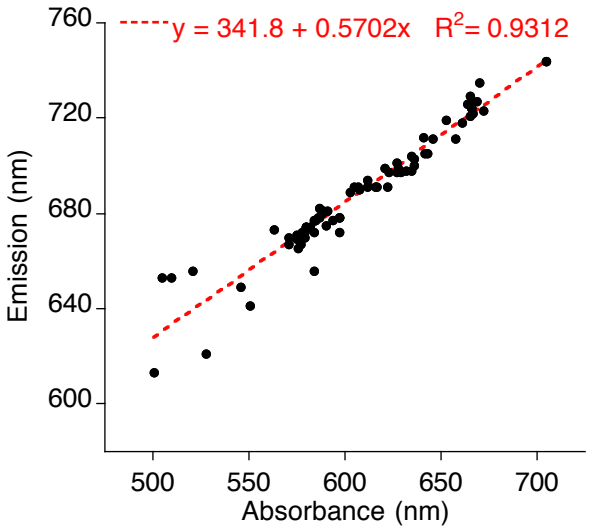


Figure II-35. Plot of emission versus absorbance for the approximately 80 hCRBPII/**ThioFluor** complexes examined in this study yields a linear fit, with R^2 = 0.93.

with absorption maxima varying from 501 nm to 705 nm and emission maxima from 613 nm to 744 nm. This is equivalent to regulation over 204 nm in absorption and 131 nm in emission, covering both the red and far-red fluorescence wavelength regimes. A plot of emission versus absorbance shows a linear response (R^2 = 0.93), with a slope of 0.57 (**Figure II-35**) indicating that approximately a 1 nm red-shift is gained in emission for a 2 nm change in absorption.

The ultimate goal would now be to demonstrate success in multi-color imaging by using two different hCRBPII variants targeted to different intracellular organelles (i.e. one variant targeted to the nucleus and the second targeted to the extra nuclear space), labeled with the same **ThioFluor** dye, to allow for multi-organelle imaging simultaneously. However, the debilitating factor here is that the developed blue-shifted hCRBPII/**ThioFluor** complexes exhibit little iminium at physiological pH due to suppressed pK_a. While this study has shown that emission wavelength can be regulated with absorption wavelength, we will require a new fluorophore that can shift wavelength in the same manner, while also possessing a high pK_a in these mutants at physiological pH. Nonetheless, the rest of this chapter will prove the probe's usefulness in live cell imaging, most prominently showing that **ThioFluor** is cell permeable and displays no background fluorescence when excited in the red, even when long incubation periods with **ThioFluor** are required, unlike merocyanine aldehyde.

II.10 Visualization of hCRBPII/ThioFluor in bacteria

We next sought to demonstrate the engineered hCRBPII/**ThioFluor** system in cells. The mutant Q108K:K40L:T53S:V62N was chosen for these preliminary studies, for imaging in *E. coli* because of it's moderate quantum yield (0.23), sufficient pK_a of 8.3 and nanomolar binding affinity with the Q108L variant (K_d = 3.2 nM) with **ThioFluor** (**Figure II-36**). Additionally, binding was relatively fast; the pseudo first order half-life of iminium formation is 30 min at 23 °C (20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7).

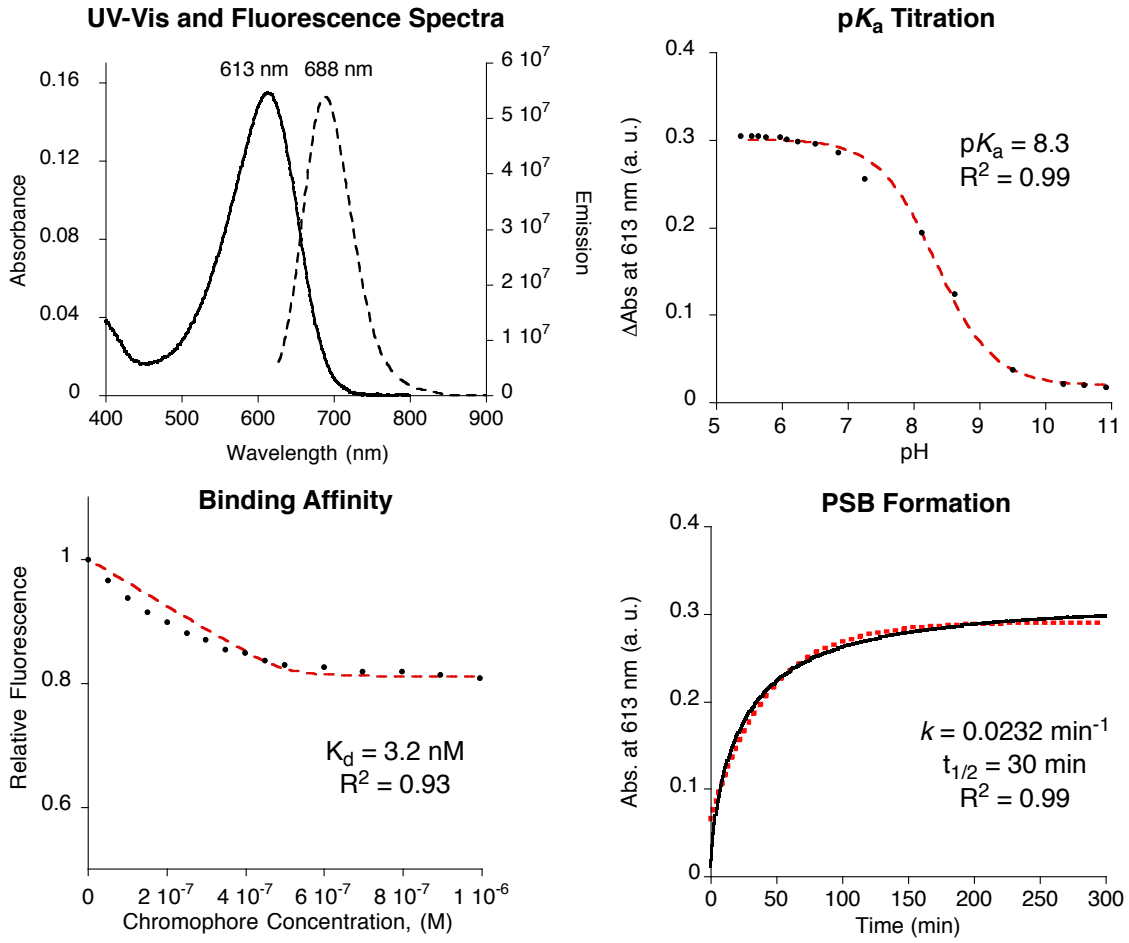


Figure II-36. Spectroscopic properties of Q108K:K40L:T53S:V62N with **ThioFluor** including UV-Vis and fluorescence spectra, pK_a, binding affinity, and rate plot. Note: It was later found that the hCRBPII species data shown here is for the dimer.

BL21 cells, expressing Q108K:K40L:T53S:V62N, were incubated with 10 μM **ThioFluor** for 1 hour at 37°C and washed with PBS buffer. As evidenced from confocal imaging (**Figure II-37**), the bacterial cells readily uptake the chromophore. Fluorescence is observed upon excitation at 594 nm, and collecting emission with a long pass (LP) 615 nm filter (fluorescence greater than 615 nm is collected). The control cells, which were not transformed with the hCRBPII plasmid, lack fluorescence, demonstrating that background fluorescence due to unbound chromophore is minimal.

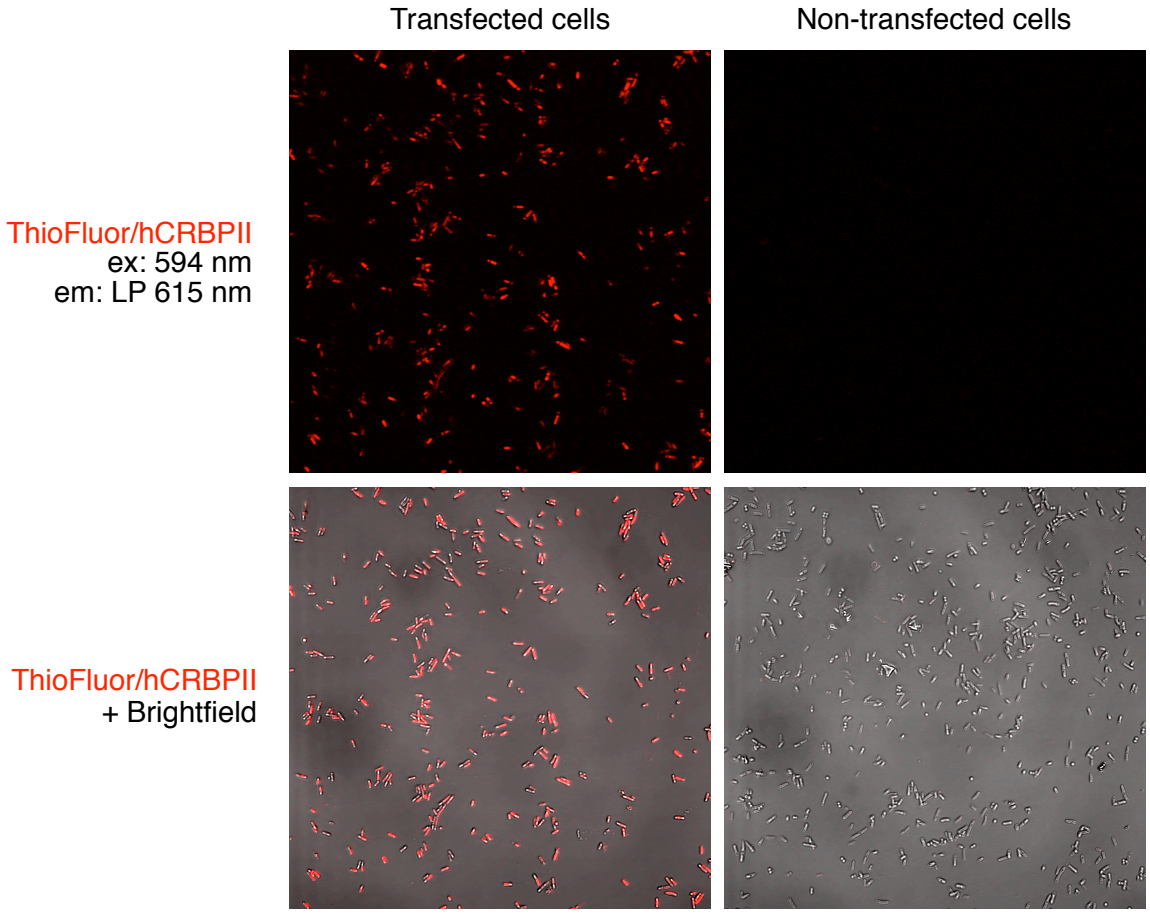


Figure II-37. Bacterial imaging of Q108K:K40L:T53S:V62N, labeled with **ThioFluor**.

II.11 Visualization of hCRBPII/ThioFluor in mammalian cells

II.11.1 Labeling of hCRBPII variants that have a propensity to oligomerize is unproductive

To further highlight its application as a fluorescent protein tag, Q108K:K40L:T53S:V62N was cloned into a pECFP-N1 vector containing ECFP to serve as a positive control for transfection. The final product of cloning was XhoI-hCRBPII-BamHI-AgeI-ECFP-Stop-NotI. No localization signal is included, so fluorescence is expected in the entire cell.

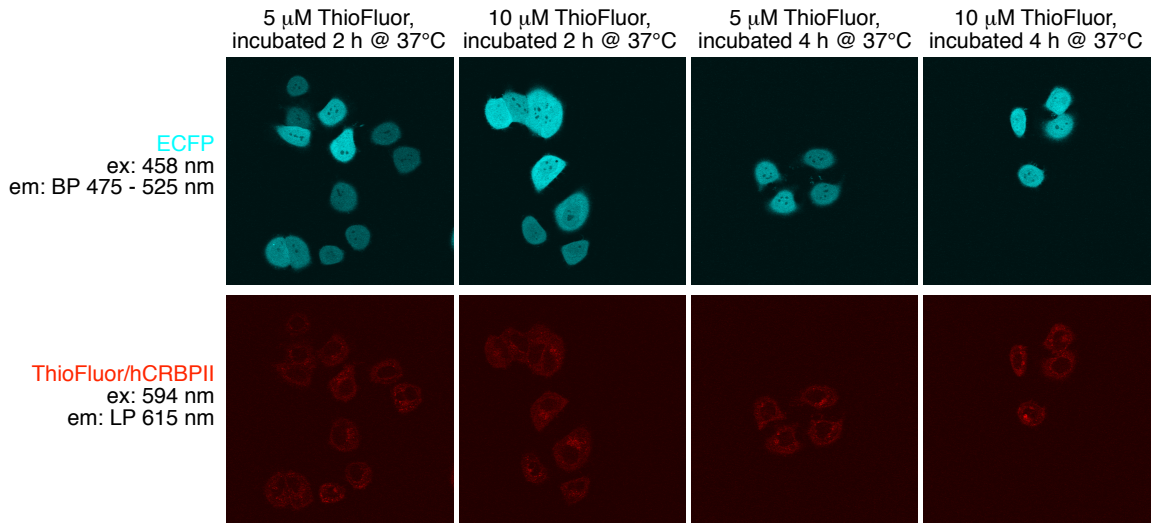


Figure II-38. Labeling of HeLa cells expressing Q108K:K40L:T53S:V62N-ECFP with **ThioFluor** at various incubation times and concentrations of **ThioFluor**.

HeLa cells were incubated with **ThioFluor** at 37 °C with various concentrations (1 μM, 5 μM, 10 μM, 50 μM, 100 μM, 200 μM) and incubation times (1 h, 2 h, 4 h). Fluorescence from ECFP is observable upon excitation at 458 nm (band pass (BP) 475-535 emission filter). Excitation at 594 nm (LP 615 nm emission filter) did not provide images corresponding to that of ECFP (**Figure II-38**). Unfortunately, in all cases (all data not shown), no whole cell fluorescence

from **ThioFluor** is visible when hCRBPII-ECFP is expressed and incubated with **ThioFluor**. Faint fluorescence is observable in the red channel, presumably due to nonspecific labeling, which indicates **ThioFluor** is cell permeable, but unfortunately is unable to bind to the expressed hCRBPII-ECFP fusion protein.

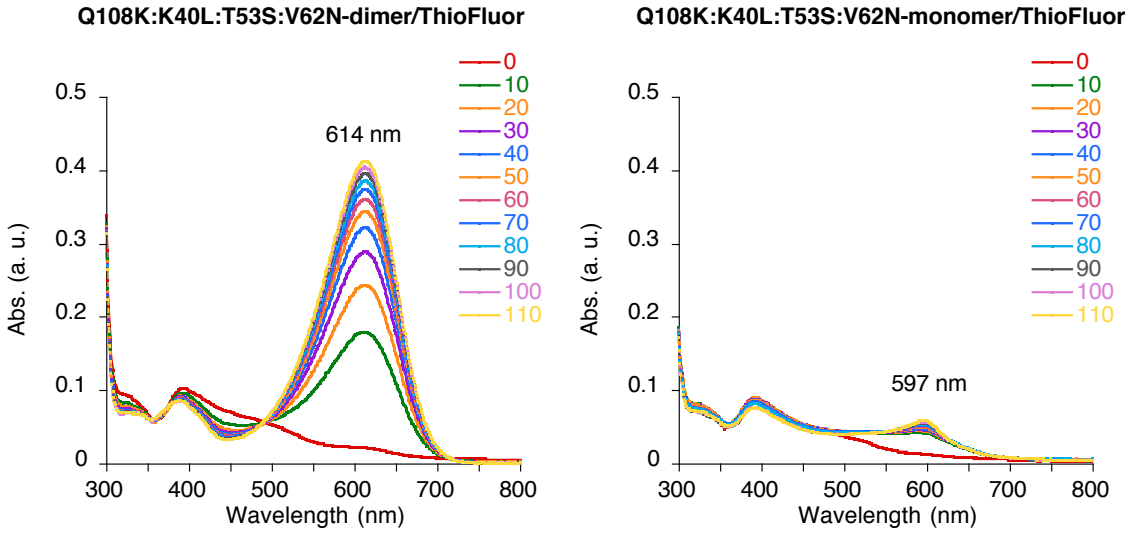


Figure II-39. Drastically different spectroscopic properties are observed upon binding **ThioFluor** to Q108K:K40L:T53S:V62N monomer and dimer (for both identical conditions are used: 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7 and 23 °C). The spectra depict the course of iminium formation (in minutes).

These preliminary *in vivo* studies were disheartening. However, it was soon realized that this protein makes both monomer and dimer during protein expression at room temperature, 8.0 mg/L and 7.1 mg/L, respectively. After separation of the monomer and dimer via size exclusion chromatography, we were able to look at the spectroscopic properties of the monomer and dimer independently. Data presented in the previous section (**Chapter II.10**) is equivalent to that obtained for the dimer. While the spectroscopic properties of Q108K:K40L:T53S:V62N-dimer/**ThioFluor** should be suitable for imaging, there appears to be little iminium formation of **ThioFluor** with the monomer of

Q108K:K40L:T53S:V62N (**Figure II-39**) under identical conditions. Disadvantageously, the monomer is also not stable to acid.

Because **ThioFluor** has not been previously used in live cell imaging, we chose to test merocyanine aldehyde with the same hCRBPII mutant, as merocyanine aldehyde has shown success in labeling live cells (although with other hCRBPII mutants). Binding to the monomer and dimer independently was verified *in vitro* under identical conditions. Spectroscopic properties, including the wavelength of absorption, extinction coefficient and rate of iminium formation are essentially the same for both monomer and dimer (**Table II-22**). It should be noted that brightness is almost twelve times higher for the hCRBPII monomer complexed with merocyanine aldehyde than for **ThioFluor**, and its brightness is comparable to the previously reported hCRBPII/merocyanine aldehyde complexes.

Table II-22. Spectroscopic properties of Q108K:K40L:T53S:V62N/merocyanine aldehyde.

hCRBPII	λ_{abs}	λ_{em}	Φ^a	ε ($M^{-1}cm^{-1}$)	k (min^{-1})	$t_{1/2}{}^b$ (min)
monomer	598	nd	nd	212,000	0.3148	2.2
dimer	595	613	0.37	207,000	0.3475	2.0

[a]Absolute quantum yield was measured on a Quantaurus-QY. [b]Half-lives based on the rate constant obtained from pseudo first order fitting of the data; measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7.

Next, we attempted to label the same hCRBPII-ECFP fusion *in vivo* with merocyanine aldehyde (**Figure II-40**). The same hCRBPII-ECFP construct was expressed in HeLa cells and subsequently incubated with 250 nM merocyanine

aldehyde for 1 h at 37 °C. Again, fluorescence from the ECFP channel is visible, denoting that transfection of the DNA was successful. However, as was the case for **ThioFluor**, no whole cell fluorescence is visible upon excitation at 594 nm (LP 615 nm emission filter). The fluorescence from the red channel seems to be consistent with background fluorescence due to nonspecific labeling.

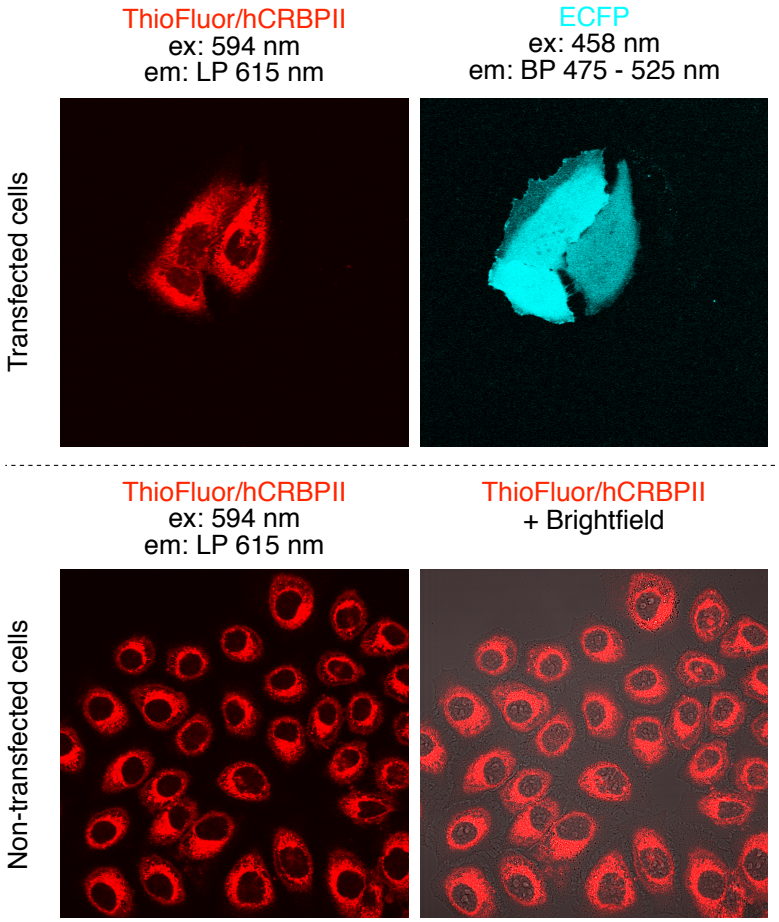


Figure II-40. Labeling of HeLa cells expressing Q108K:K40L:T53S:V62N-ECFP with merocyanine aldehyde.

Interestingly labeling in U2OS cells shows a bit more hope. The same hCRBPII-ECFP fusion was expressed in U2OS cells and subsequently labeled with **ThioFluor** (incubated with 50 µM for 1 h at 37 °C) or merocyanine aldehyde (incubated with 250 nM for 5 min at 37 °C) (**Figure II-41**). As expected,

fluorescence from the ECFP channel is visible, showing that transfection of the DNA was successful. It should be noted that the ECFP channel of cells labeled with **ThioFluor** shows background, which is presumably due to the excitation of free chromophore. Upon excitation at 594 nm (LP 615 nm emission filter), both experiments show whole cell labeling, though more dim than expected with a substantial amount of background fluorescence. Perhaps, different cell lines lead to differences in monomer/dimer ratios or protein folding. However, further experiments were not conducted in order to optimize this system.

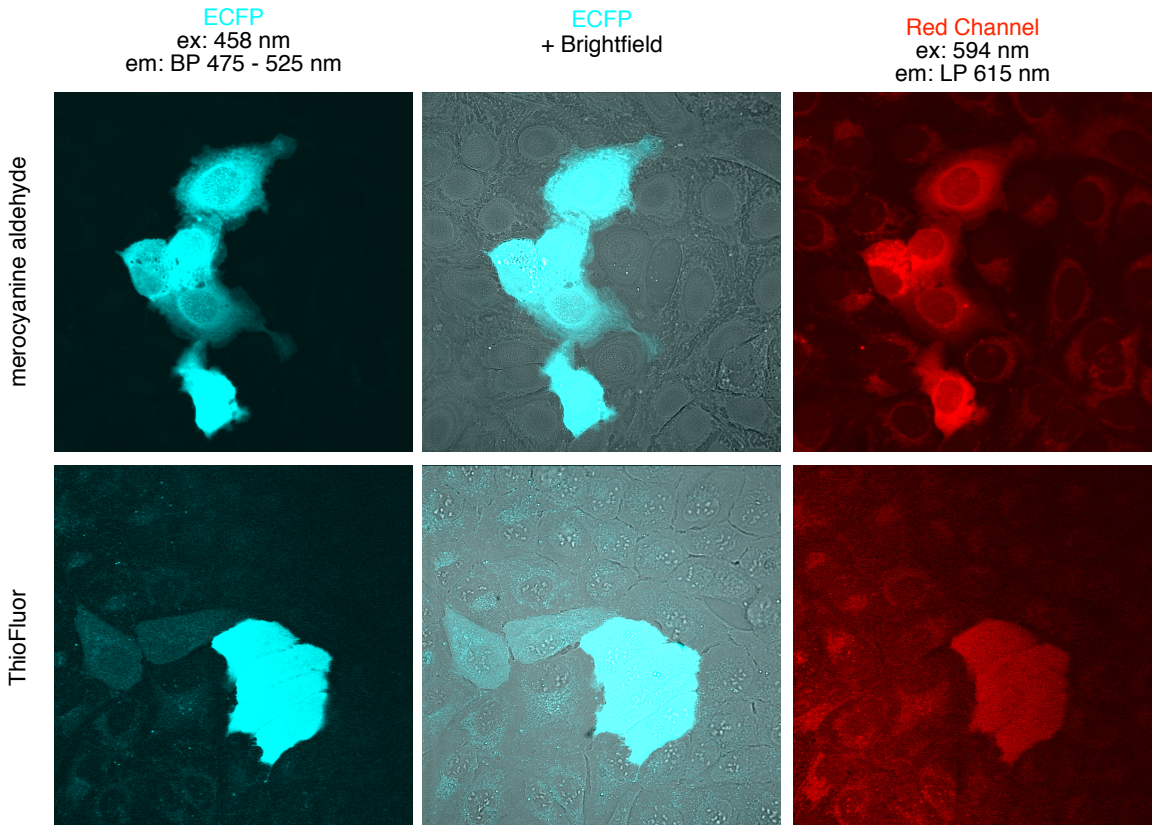


Figure II-41. Labeling of U2OS cells expressing Q108K:K40L:T53S:V62N-ECFP with **ThioFluor** and merocyanine aldehyde.

From the data presented in this section, it was hypothesized that *in vivo* labeling was unsuccessful due to the propensity of Q108K:K40L:T53S:V62N to

oligomerize (which is due to both the presence of T53S and V62N). Hence, mutants that produced a majority of monomer during protein expression were utilized for future imaging of mammalian cells.

II.11.2 ThioFluor successfully labels hCRBPII in mammalian cells

In the previous section, it was observed that many mutations make hCRBPII prone to domain-swapped dimerization (see **Chapter VI** for a full discussion), and that one such mutant (Q108K:K40L:T53S:V62N) displays vastly different spectroscopic properties when coupled with **ThioFluor**. Additionally, **ThioFluor** was not able to label the fusion protein when expressed in mammalian cells. It was predicted that it is the propensity to form oligomers that renders this mutant unsuccessful when moving from labeling *in vitro to in vivo*.

The mutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E was chosen for application toward *in vivo* imaging, due to its high quantum yield (0.31), high pKa (11.2) and reasonable half-life of 82 min (based on fitting to second order kinetics; measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7) with **ThioFluor** (**Figure II-42**). This absorption and emission wavelengths are also well aligned with lasers available on the confocal microscope (*in vitro* this mutant displayed an absorption maxima of 563 nm and emission maxima of 673 nm at ~ pH 7). Additionally, it is solely expressed in the monomeric form. hCRBPII was cloned into the pFlag-CMV2 vector containing EGFP, both with and without localization peptides, for labeling the whole cell (hCRBPII-EGFP), nucleus (hCRBPII-EGFP-3NLS) and the extranuclear space (hCRBPII-EGFP-

NES). Imaging was performed by incubating HeLa cells with 10 μM **ThioFluor** for 1 h at 37 °C. The cells were then washed with PBS buffer two times before imaging.

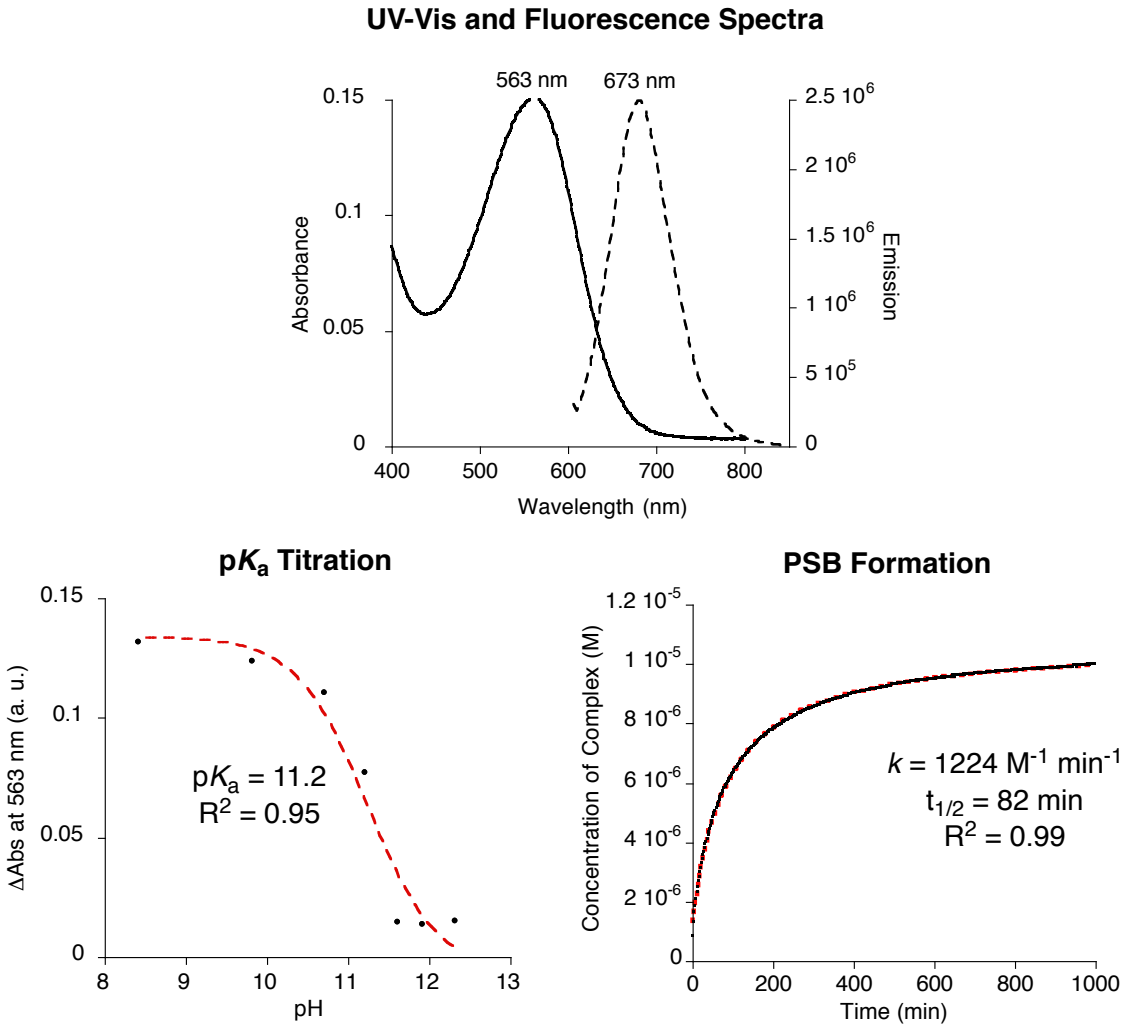


Figure II-42. Spectroscopic properties of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E with **ThioFluor** including UV-Vis and fluorescence spectra, pK_a, and rate plot.

In all cases, green fluorescence was observed when excited at 488 nm, indicating that transfection was successful and the fusion protein is being expressed (**Figure II-43**, green channel). Excitation at 594 nm of the hCRBPII/**ThioFluor** complex provided almost identical images (**Figure II-43**, red

channel). Noteworthy, background fluorescence is not apparent in the red channel, indicating that **ThioFluor** does not label off target lysines. This proves the utility of labeling Q108K:K40L:T51V:T53S:R58W:Y19W:L117E with **ThioFluor** in mammalian cells.

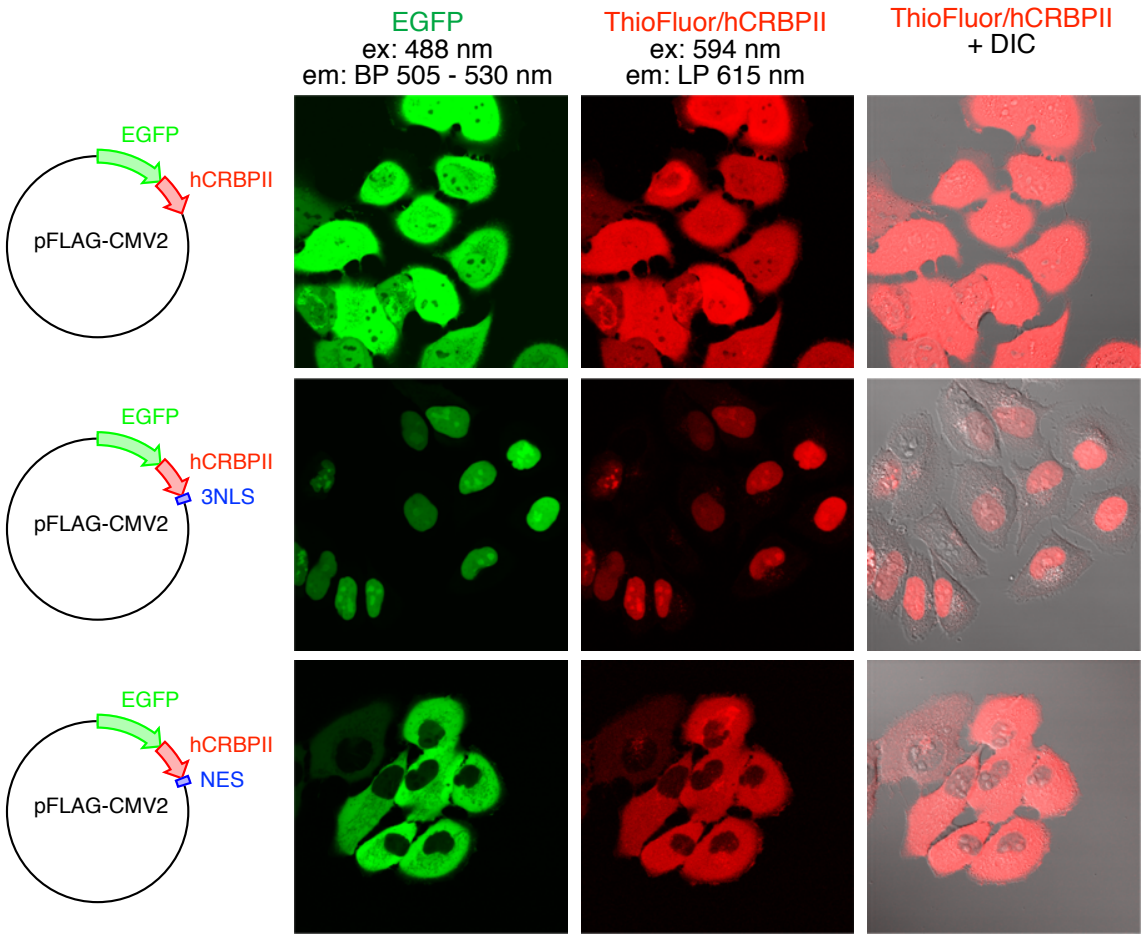


Figure II-43. Labeling of HeLa cells expressing hCRBPII-EGFP, hCRBPII-EGFP-3NLS and hCRBPII-EGFP-NES with 10 μM **ThioFluor** (incubated at 37°C for 1 hour). NLS = nuclear localization sequence, NES = nuclear export signal.

II.12 Conclusions

As stated in the introduction of this chapter, our overarching goal is to develop fluorescent proteins that span the entire visible spectra. This was achieved by the coupling of the solvatochromic fluorophore ThioFluor to various

hCRBPII mutants. **ThioFluor** yielded mutants with absorption maxima varying from 501 nm to 705 nm and emission maxima from 613 nm to 744 nm. This is equivalent to regulation over 204 nm in absorption and 131 nm in emission, covering both the red and far-red fluorescence wavelength regimes. A plot of emission versus absorbance shows a linear response ($R^2 = 0.93$), with a slope of 0.57 (**Figure II-35**) indicating that approximately a 1 nm red-shift is gained in emission for a 2 nm change in absorption. Furthermore, we have shown its utility in live-cell imaging in whole cells, and with targeting to the nucleus and extranuclear space; fortuitously, negligible background fluorescence is apparent.

The ultimate goal would now be to demonstrate success in multi-color imaging by using two different hCRBPII variants targeted to different intracellular organelles (i.e. one variant targeted to the nucleus and the second targeted to the extra nuclear space), labeled with the same **ThioFluor** dye, to allow for multi-organelle imaging simultaneously. However, blue-shifted hCRBPII/**ThioFluor** complexes exhibit little iminium at physiological pH due to suppressed pK_a. We will require a new fluorophore that can shift wavelength in the same manner, while also possessing a high pK_a in these mutants at physiological pH in order to show success in multi-color imaging.

REFERENCES

REFERENCES

1. Wang, W., Geiger, J.H. & Borhan, B. The photochemical determinants of color vision: revealing how opsins tune their chromophore's absorption wavelength. *Bioessays* **36**, 65-74 (2014).

2. Berbasova, T., Nosrati, M., Vasileiou, C., Wang, W., Lee, K.S., Yapici, I., Geiger, J.H. & Borhan, B. Rational design of a colorimetric pH sensor from a soluble retinoic acid chaperone. *Journal of the American Chemical Society* **135**, 16111-16119 (2013).

3. 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).

4. Reichardt, C. Solvatochromic Dyes as Solvent Polarity Indicators. *Chemical Reviews* **94**, 2319-2358 (1994).

5. Marini, A., Munoz-Losa, A., Biancardi, A. & Mennucci, B. What is Solvatochromism? *Journal of Physical Chemistry B* **114**, 17128-17135 (2010).

6. 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).

7. Richard, J.P., Amyes, T.L., Crugeiras, J. & Rios, A. Pyridoxal 5 '-phosphate: electrophilic catalyst extraordinaire. *Current Opinion in Chemical Biology* **13**, 475-483 (2009).

8. Uttamapinant, C., White, K.A., Baruah, H., Thompson, S., Fernandez-Suarez, M., Puthenveetil, S. & Ting, A.Y. A fluorophore ligase for site-specific protein labeling inside living cells. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 10914-10919 (2010).

9. Lakowicz, J.R. Principles of Fluorescence Spectroscopy, Edn. 3rd. (Springer, 2006).

10. Engelborghs, Y. & Visser, A.J.W.G. Fluorescence Spectroscopy and Microscopy: Methods and Protocols. (Springer, 2014).

11. Klan, P. & Wirz, J. in Photochemistry of Organic Compounds: From Concepts to Practice 25-72 (Wiley-Blackwell, 2009).

12. Ahmed, F. A Good Example of the Franck-Condon Principle. *Journal of Chemical Education* **64**, 427-428 (1987).

13. Lakowicz, J.R. in Principles of Fluorescence Spectroscopy, Edn. 3rd 205-235 (Springer, 2006).

14. Klymchenko, A.S. & Mely, Y. Fluorescent Environment-Sensitive Dyes as Reporters of Biomolecular Interactions. *Progress in Molecular Biology and Translational Science* **113**, 35-58 (2013).

15. Loving, G.S., Sainlos, M. & Imperiali, B. Monitoring protein interactions and dynamics with solvatochromic fluorophores. *Trends in Biotechnology* **28**, 73-83 (2010).

16. Klymchenko, A.S. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Accounts of Chemical Research* **50**, 366-375 (2017).

17. Holmesfarley, S.R. & Whitesides, G.M. Fluorescence Properties of Dansyl Groups Covalently Bonded to the Surface of Oxidatively Functionalized Low-Density Polyethylene Film. *Langmuir* **2**, 266-281 (1986).

18. Sameiro, M. & Goncalves, T. Fluorescent Labeling of Biomolecules with Organic Probes. *Chemical Reviews* **109**, 190-212 (2009).

19. Weber, G. & Farris, F.J. Synthesis and Spectral Properties of a Hydrophobic Fluorescent-Probe - 6-Propionyl-2-(Dimethylamino)Naphthalene. *Biochemistry* **18**, 3075-3078 (1979).

20. Lehrer, S.S. & Ishii, Y. Fluorescence Properties of Acrylodan-Labeled Tropomyosin and Tropomyosin Actin - Evidence for Myosin Subfragment-1 Induced Changes in Geometry between Tropomyosin and Actin. *Biochemistry* **27**, 5899-5906 (1988).

21. Hiratsuka, T. ATP-induced opposite changes in the local environments around Cys(697) (SH2) and Cys(707) (SH1) of the myosin motor domain revealed by the prodan fluorescence. *Journal of Biological Chemistry* **274**, 29156-29163 (1999).

22. Shi, J.X., Boyd, A.E., Radic, Z. & Taylor, P. Reversibly bound and covalently attached ligands induce conformational changes in the omega

loop, Cys(69)-Cys(96), of mouse acetylcholinesterase. *Journal of Biological Chemistry* **276**, 42196-42204 (2001).

23. Harikumar, K.G., Pinon, D.L., Wessels, W.S., Prendergast, F.G. & Miller, L.J. Environment and mobility of a series of fluorescent reporters at the amino terminus of structurally related peptide agonists and antagonists bound to the cholecystokinin receptor. *Journal of Biological Chemistry* **277**, 18552-18560 (2002).

24. Lu, Z.K., Lord, S.J., Wang, H., Moerner, W.E. & Twieg, R.J. Long-wavelength analogue of PRODAN: Synthesis and properties of Anthradan, a fluorophore with a 2,6-donor-acceptor anthracene structure. *Journal of Organic Chemistry* **71**, 9651-9657 (2006).

25. Kucherak, O.A., Didier, P., Mely, Y. & Klymchenko, A.S. Fluorene Analogues of Prodan with Superior Fluorescence Brightness and Solvatochromism. *Journal of Physical Chemistry Letters* **1**, 616-620 (2010).

26. Carlotti, B., Flamini, R., Kikas, I., Mazzucato, U. & Spalletti, A. Intramolecular charge transfer, solvatochromism and hyperpolarizability of compounds bearing ethenylene or ethynylene bridges. *Chemical Physics* **407**, 9-19 (2012).

27. Singh, P., Baheti, A., Thomas, K.R.J., Lee, C.P. & Ho, K.C. Fluorene-based organic dyes containing acetylene linkage for dye-sensitized solar cells. *Dyes and Pigments* **95**, 523-533 (2012).

28. Tigreros, A., Ortiz, A. & Insuasty, B. Effect of pi-conjugated linkage on photophysical properties: Acetylene linker as the better connection group for highly solvatochromic probes. *Dyes and Pigments* **111**, 45-51 (2014).

29. Diwu, Z., Lu, Y.X., Zhang, C.L., Klaubert, D.H. & Haugland, R.P. Fluorescent molecular probes II. The synthesis, spectral properties and use of fluorescent solvatochromic Dapoxyl(TM) dyes. *Photochemistry and Photobiology* **66**, 424-431 (1997).

30. Min, J., Lee, J.W., Ahn, Y.H. & Chang, Y.T. Combinatorial dapoxyl dye library and its application to site selective probe for human serum albumin. *Journal of Combinatorial Chemistry* **9**, 1079-1083 (2007).

31. Ando, Y., Homma, Y., Hiruta, Y., Citterio, D. & Suzuki, K. Structural characteristics and optical properties of a series of solvatochromic

fluorescent dyes displaying long-wavelength emission. *Dyes and Pigments* **83**, 198-206 (2009).

32. Dattelbaum, J.D., Looger, L.L., Benson, D.E., Sali, K.M., Thompson, R.B. & Hellinga, H.W. Analysis of allosteric signal transduction mechanisms in an engineered fluorescent maltose biosensor. *Protein Science* **14**, 284-291 (2005).

33. Shi, Z.D., Karki, R.G., Oishi, S., Worthy, K.M., Bindu, L.K., Dharmawardana, P.G., Nicklaus, M.C., Bottaro, D.P., Fisher, R.J. & Burke, T.R. Utilization of a nitrobenzoxadiazole (NBD) fluorophore in the design of a Grb2 SH2 domain-binding peptide mimetic. *Bioorganic & Medicinal Chemistry Letters* **15**, 1385-1388 (2005).

34. Jiang, W., Fu, Q.Q., Fan, H.Y. & Wang, W. An NBD fluorophore-based sensitive and selective fluorescent probe for zinc ion. *Chemical Communications*, 259-261 (2008).

35. Zhou, G.D., Wang, H.L., Ma, Y. & Chen, X.Q. An NBD fluorophore-based colorimetric and fluorescent chemosensor for hydrogen sulfide and its application for bioimaging. *Tetrahedron* **69**, 867-870 (2013).

36. Soujanya, T., Fessenden, R.W. & Samanta, A. Role of nonfluorescent twisted intramolecular charge transfer state on the photophysical behavior of aminophthalimide dyes. *Journal of Physical Chemistry* **100**, 3507-3512 (1996).

37. Vazquez, M.E., Rothman, D.M. & Imperiali, B. A new environment-sensitive fluorescent amino acid for Fmoc-based solid phase peptide synthesis. *Organic & Biomolecular Chemistry* **2**, 1965-1966 (2004).

38. Vazquez, M.E., Blanco, J.B. & Imperiali, B. Photophysics and biological applications of the environment-sensitive fluorophore 6-N,N-Dimethylamino-2,3-naphthalimide. *Journal of the American Chemical Society* **127**, 1300-1306 (2005).

39. Loving, G. & Imperiali, B. A versatile amino acid analogue of the solvatochromic fluorophore 4-N,N-dimethylamino-1,8-naphthalimide: A powerful tool for the study of dynamic protein interactions. *Journal of the American Chemical Society* **130**, 13630-13638 (2008).

40. Liu, X., Cole, J.M. & Low, K.S. Solvent Effects on the UV-vis Absorption and Emission of Optoelectronic Coumarins: a Comparison of Three

Empirical Solvatochromic Models. *Journal of Physical Chemistry C* **117**, 14731-14741 (2013).

41. Maity, B., Chatterjee, A. & Seth, D. Photophysics of a Coumarin in Different Solvents: Use of Different Solvatochromic Models. *Photochemistry and Photobiology* **90**, 734-746 (2014).

42. Lanke, S.K. & Sekar, N. Coumarin Push-Pull NLOphores with Red Emission: Solvatochromic and Theoretical Approach. *Journal of Fluorescence* **26**, 949-962 (2016).

43. Greenspan, P., Mayer, E.P. & Fowler, S.D. Nile Red - a Selective Fluorescent Stain for Intracellular Lipid Droplets. *Journal of Cell Biology* **100**, 965-973 (1985).

44. Jose, J. & Burgess, K. Benzophenoxazine-based fluorescent dyes for labeling biomolecules. *Tetrahedron* **62**, 11021-11037 (2006).

45. Sherman, D.B., Pitner, J.B., Ambroise, A. & Thomas, K.J. Synthesis of thiol-reactive, long-wavelength fluorescent phenoxazine derivatives for biosensor applications. *Bioconjugate Chemistry* **17**, 387-392 (2006).

46. Thomas, K.J., Sherman, D.B., Amiss, T.J., Andaluz, S.A. & Pitner, J.B. A long-wavelength fluorescent glucose biosensor based on bioconjugates of galactose/glucose binding protein and Nile Red derivatives. *Diabetes Technology & Therapeutics* **8**, 261-268 (2006).

47. Kawski, A., Bojarski, P. & Kuklinski, B. Estimation of ground- and excited-state dipole moments of Nile Red dye from solvatochromic effect on absorption and fluorescence spectra. *Chemical Physics Letters* **463**, 410-412 (2008).

48. Kamlet, M.J. & Taft, R.W. Solvatochromic Comparison Method 1. Beta-Scale of Solvent Hydrogen-Bond Acceptor (Hba) Basicities. *Journal of the American Chemical Society* **98**, 377-383 (1976).

49. Taft, R.W. & Kamlet, M.J. Solvatochromic Comparison Method 2. Alpha-Scale of Solvent Hydrogen-Bond Donor (Hbd) Acidities. *Journal of the American Chemical Society* **98**, 2886-2894 (1976).

50. Kamlet, M.J., Abboud, J.L.M., Abraham, M.H. & Taft, R.W. Linear Solvation Energy Relationships 23. A Comprehensive Collection of the Solvatochromic Parameters, Pi-Star, Alpha and Beta, and Some Methods

for Simplifying the Generalized Solvatochromic Equation. *Journal of Organic Chemistry* **48**, 2877-2887 (1983).

51. Abraham, M.H., Grellier, P.L., Abboud, J.L.M., Doherty, R.M. & Taft, R.W. Solvent Effects in Organic-Chemistry - Recent Developments. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **66**, 2673-2686 (1988).

52. Catalan, J., Lopez, V., Perez, P., Martinvillamil, R. & Rodriguez, J.G. Progress Towards a Generalized Solvent Polarity Scale - the Solvatochromism of 2-(Dimethylamino)-7-Nitrofluorene and Its Homomorph 2-Fluoro-7-Nitrofluorene. *Liebigs Annalen*, 241-252 (1995).

53. Catalan, J., Diaz, C., Lopez, V., Perez, P., dePaz, J.L.G. & Rodriguez, J.G. A generalized solvent basicity scale: The solvatochromism of 5-nitroindoline and its homomorph 1-methyl-5-nitroindoline. *Liebigs Annalen*, 1785-1794 (1996).

54. Catalan, J. & Diaz, C. A generalized solvent acidity scale: The solvatochromism of o-tert-butylstilbazolium betaine dye and its homomorph o,o'-di-tert-butylstilbazolium betaine dye. *Liebigs Annalen*, 1941-1949 (1997).

55. Zou, J.W., Yu, Q.S. & Shang, Z.C. Correlation between empirical solvent polarity scales and computed quantities derived from molecular surface electrostatic potentials. *Journal of the Chemical Society-Perkin Transactions 2*, 1439-1443 (2001).

56. Catalan, J. Toward a Generalized Treatment of the Solvent Effect Based on Four Empirical Scales: Dipolarity (SdP, a New Scale), Polarizability (SP), Acidity (SA), and Basicity (SB) of the Medium. *Journal of Physical Chemistry B* **113**, 5951-5960 (2009).

57. Ceron-Carrasco, J.P., Jacquemin, D., Laurence, C., Planchat, A., Reichardt, C. & Sraidi, K. Solvent polarity scales: determination of new E-T(30) values for 84 organic solvents. *Journal of Physical Organic Chemistry* **27**, 512-518 (2014).

58. Reichardt, C. Pyridinium N-phenolate betaine dyes as empirical indicators of solvent polarity: Some new findings. *Pure and Applied Chemistry* **76**, 1903-1919 (2004).

59. Machado, V.G., Stock, R.I. & Reichardt, C. Pyridinium N-Phenolate Betaine Dyes. *Chemical Reviews* **114**, 10429-10475 (2014).

60. Eisfeld, A. & Briggs, J.S. The J- and H-bands of organic dye aggregates. *Chemical Physics* **324**, 376-384 (2006).

61. Wang, W. in PhD Thesis, Michigan State University (2012).

62. Yapici, I. in PhD Thesis, Michigan State University (2015).

63. Yutani, K., Ogasahara, K., Tsujita, T. & Sugino, Y. Dependence of Conformational Stability on Hydrophobicity of the Amino-Acid Residue in a Series of Variant Proteins Substituted at a Unique Position of Tryptophan Synthase Alpha-Subunit. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 4441-4444 (1987).

64. Pace, C.N., Trevino, S., Prabhakaran, E. & Scholtz, J.M. Protein structure, stability and solubility in water and other solvents. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **359**, 1225-1234 (2004).

65. Pace, C.N., Grimsley, G.R. & Scholtz, J.M. Protein Ionizable Groups: pK Values and Their Contribution to Protein Stability and Solubility. *Journal of Biological Chemistry* **284**, 13285-13289 (2009).

66. Assar, Z., Nossoni, Z., Wang, W., Santos, E.M., Kramer, K., McCornack, C., Vasileiou, C., Borhan, B. & Geiger, J.H. Domain-Swapped Dimers of Intracellular Lipid-Binding Proteins: Evidence for Ordered Folding Intermediates. *Structure* **24**, 1590-1598 (2016).

CHAPTER III: STRUCTURE-PROPERTY RELATIONSHIPS OF THIOFLUOR ANALOGUES BOUND TO hCRBPII MUTANTS

In developing novel fluorescent proteins and small molecule tags, there are numerous desirable qualities of the probe that should be achieved. It is these requirements that present a challenge in the design of new synthetic small molecule probes. Perhaps the most imperative characteristics are that the probe is fluorogenic (ideally in the red or near-infrared wavelength regime), cell permeable and rapid (i.e. short incubation time). Fluorogenic probes should allow for temporal control, in which fluorescence is produced only when the dye is bound to the target.[1-4] Nonspecific labeling of a dye may yield background fluorescence that decreases the detection limit or signal to noise ratio of the dye. Undesired fluorescence may also originate from the autofluroescence of endogenous compounds in mammalian cells, such as riboflavin and flavin coenzymes, among others.[5-8] When using green or yellow fluorescent proteins (FPs), whose emission overlays cellular autofluorescence, detection of the FP may be hampered. Choosing red or far red emitting fluorescent dyes can help to increase the signal to noise ratio. Red-shifted emission can also provide the added benefit of reduced scattering and deeper tissue penetration, due to the longer wavelength of light.[9, 10] In line with fluorogenicity, it is advantageous for the system to require a short incubation time with the dye. Not only does this decrease the amount of background fluorescence in some cases (as was seen for merocyanine aldehyde), but a faster reaction time also allows for performing

real time measurements and the monitoring of molecular events directly after labeling.[11-13] Cell permeable probes again extend the applicability of the dye, as they allow for the imaging of intracellular organelles such as the nucleus, while impermeable dyes limit imaging to the cell membrane,[14, 15] which limits their usefulness.

Other desirable properties of dyes in live cell imaging are brightness and photostability, and their ability to be targeted to allow for spatial control. Brightness is the product of the extinction coefficient (at the relevant excitation wavelength) and quantum yield. Extinction coefficient is a measure of the ability of a dye to absorb light, while quantum yield is the number of photons emitted per the number of photons absorbed. Thus, both a higher extinction coefficient and higher quantum yield can lead to increased brightness.[16-19] Along the same lines, it is desirable that the dye be photostable. During excitation of a fluorophore, the dye can be photobleached, which can lead to its permanent loss of fluorescence due to photon-induced damage.[20-22] Some fluorophores bleach quickly, while others are robust and can undergo cycling between excitation and emission many times before bleaching. One of the ways to minimize photobleaching of the fluorophore is to limit the amount of light it is exposed to, which can be done by reducing light intensity (which also reduces signal intensity); it is in such a case that brighter fluorophores are desirable.

Chapter II detailed the design of fluorescent proteins based on hCRBPII coupled with the solvatochromic fluorophore **ThioFluor**. Significant achievements

include the wavelength and emission regulation, via site directed mutagenesis, with absorptions spanning 501 nm to 705 nm and emissions 613 nm to 744 nm. The system displayed many desirable properties, such as being fluorogenic, which allowed for minimal background staining and spatial control (as the tags can be localized to organelles via sequencing peptides), and decent quantum yields. However, properties to be improved are reaction kinetics and pK_a of the *in situ* formed iminium. To this end, we sought to generate a small library of **ThioFluor** analogues, in order to understand how the spectroscopic properties of hCRBPII and an aldehydic ligand pair can be manipulated, such as the absorption wavelength, emission wavelength, the rate of iminium formation, pK_a and quantum yield.

Herein, the goal is to make rational, meticulous modifications on **ThioFluor**, and individually look at the effect of the substitution. Better understanding of the structure-property relationships should establish design parameters for improved dyes. We used **ThioFluor** as a starting point for modification to develop analogues, generally based on substitution of the electron donor and acceptor substituent's and the π linker between the electron donor and electron acceptor moieties.

III.1 Prior work exploring substituent effects on absorption wavelength in fluorescent proteins

Excitation wavelength tuning in fluorescent proteins is dictated by the covalent structure of the chromophore, while also being influenced by hydrogen

bonding and electrostatic interactions with surrounding residues.[20, 23-26] The chromophore of green fluorescent protein (GFP) is comprised of phenol and imidazolinone rings derived from the inherent tyrosine 66 residue and backbone atoms of adjacent residues (serine 65 and glycine 67).[27-29] An olefin group bridges the two rings to create an extended conjugated π system (**Figure III-1**). Additionally, in most red fluorescent proteins (RFPs), oxidation of the *N*-C_α bond immediately *N*-terminal to the chromophore extends conjugation through the acylimine group.[26, 30] The chromophore of mCherry is shown in **Figure III-1** as an example. In this case, the extra conjugation from the acylimine group is partially responsible for the 100 nm red-shift in absorption wavelength.[26 31]

The approach of systematically exploring a chromophore's donor and acceptor abilities is not an unprecedented one. The Schultz lab used unnatural amino acid mutagenesis to selectively substitute tyrosine 66 of GFP with tyrosine analogues bearing different substituents at the para position of the phenyl ring including amino, methoxy, iodo, and bromo groups (**Table III-1**, entries 1 – 5).[32] When the unnatural amino acids were introduced, single maxima were observed for both absorbance and emission spectra, unlike wild-type GFP which displays two absorbance maxima due to the presence of both the protonated and unprotonated forms of the tyrosine's hydroxyl group (**Table III-1**, entry 2). The wavelengths for both absorbance and emission peaks increase in the order of electron donating ability: *p*-bromo, *p*-iodo, *p*-methoxy, *p*-hydroxyl, *p*-amino and *p*-hydroxide (**Table III-1**). The quantum yields of the modified GFPs with *p*-amino

and *p*-methoxy are lower than wild-type GFP, but their extinction coefficients are slightly higher.

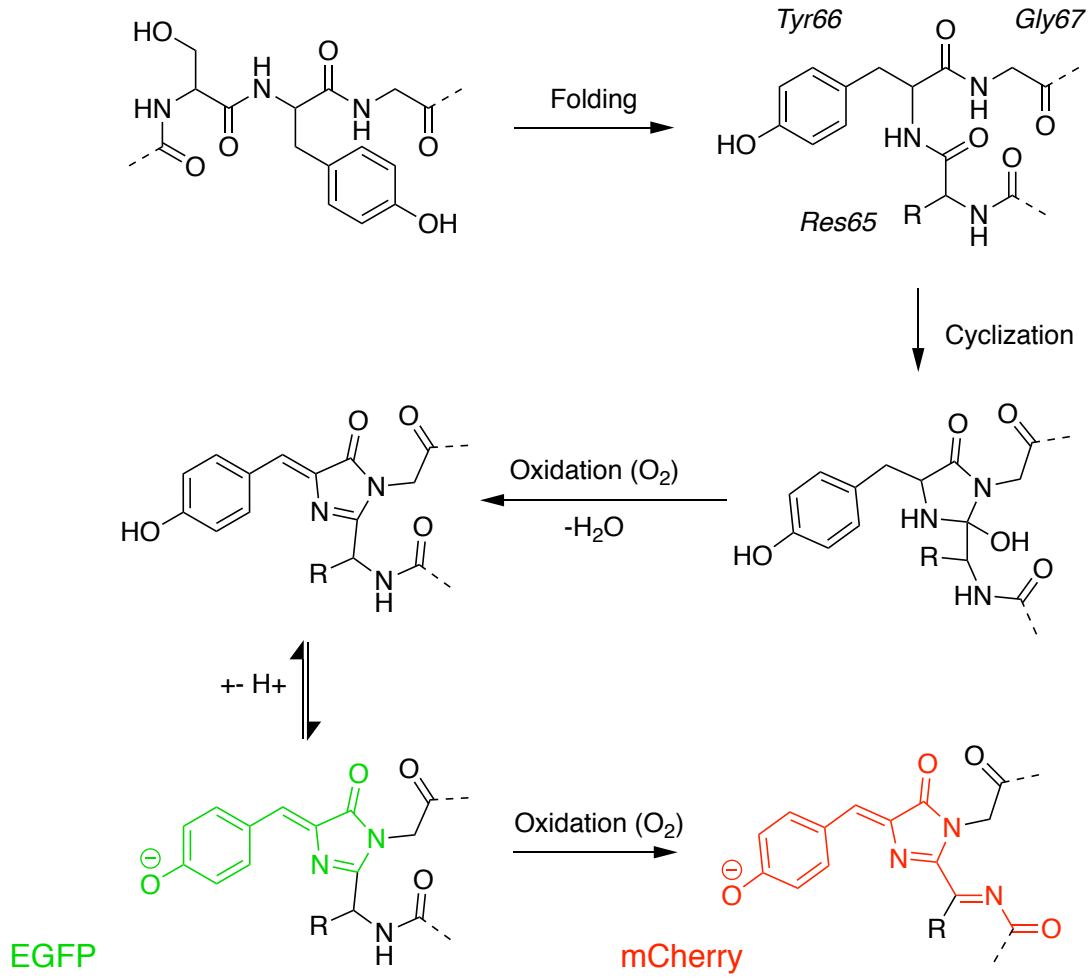


Figure III-1. Proposed scheme for the maturation of the GFP and mCherry chromophores. Proper protein conformation and subsequent cyclization, oxidation and dehydration are required to form the fluorescent state.

The same group later introduced numerous other electron substituents via the site-specific incorporation of unnatural amino acids at Tyr66. While absorption wavelength data was not provided for the other substituents (including nitrile, nitro, acetyl, azido, o-allyl and phenyl) it is observed that the emission maxima are all blue-shifted as compared to the wild-type tyrosine (**Table III-1**,

entries 6 – 11).[33, 34] The blue-shift in these wavelengths is because the electron donating ability of these functional groups is lower than the hydroxyl group of tyrosine. Additionally, the replacement of the hydroxyl group of tyrosine reduces the fluorescence quantum yields substantially in all cases, presumably because the electron-withdrawing group prevents the formation of a fluorescent push-pull system. Alternatively, it has been shown that the incorporation of halogens in porphyrins leads to a higher formation of the triplet state upon excitation,[35] which leads to a decrease in fluorescence quantum yield.

Table III-1. Effect of Tyr66 replacement on spectroscopic properties of GFP.

Entry	Substituent	λ_{abs}	λ_{em}	ε ($M^{-1}cm^{-1}$)	Φ
1	*p*-bromo	375	428	20,000	0.01
2	*p*-hydroxy (wildtype)	397,475	506	25,000	0.76
3	*p*-iodo	381	438	16,000	0.01
4	*p*-methoxy	394	460	27,000	0.37
5	*p*-amino	435	498	31,000	0.43
6	*p*-nitro	nd	-	-	-
7	*p*-acetyl	nd	435	nd	0.04
8	*p*-cyano	nd	442	nd	0.04
9	*p*-phenyl	nd	459	nd	0.07
10	*p*-vinyloxy	nd	461	nd	0.04
11	*p*-azido	nd	501	nd	0.11

III.2 Structure of ThioFluor analogues to be explored

As aforementioned, the goal is to develop a small library of **ThioFluor** analogues, in order to gain a better understanding of the structure-property relationships apparent in this system. It is hoped this would establish design

parameters for the rational design of dyes with improved spectroscopic properties. We chose to start with **ThioFluor**, as it has already shown success in live cell imaging. Mr. Hadi Gholami was successful in synthesizing a variety of analogues, which differ based on their electron donating or acceptor groups and the aromatic linker between the two groups (**Figure III-2**).

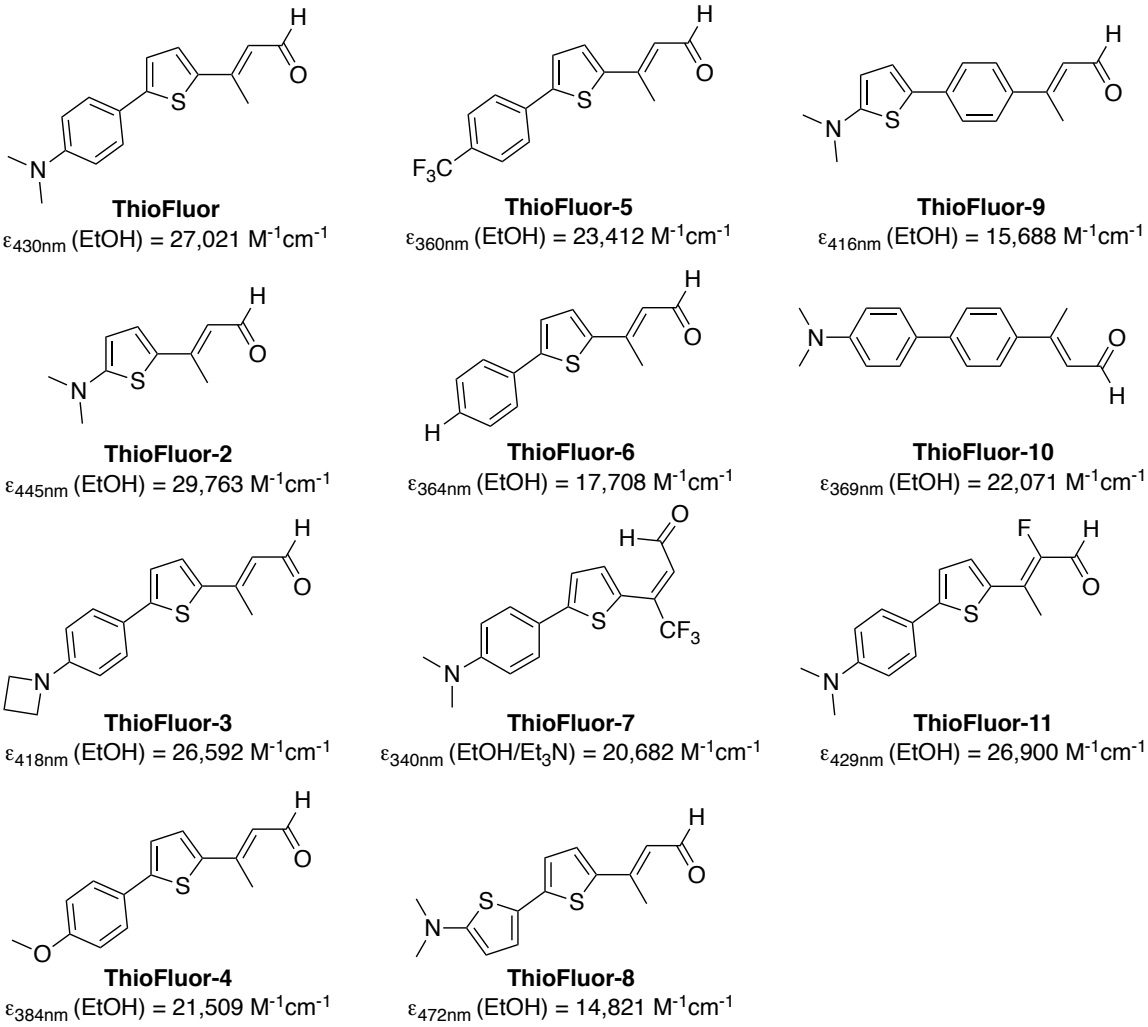


Figure III-2. **ThioFluor** analogues investigated to study structure-property relationships in absorption and fluorescence. Extinction coefficients are for λ_{max} of each chromophore.

The first set of analogues was made in order to explore the role of substituent effects on **ThioFluor**. For this reason, we tested the substitution of

electron donating groups on **ThioFluor**, replacing its para *N,N*-dimethyl aniline group with azetidine (**ThioFluor-3**), methoxy (**ThioFluor-4**), trifluoromethyl (**ThioFluor-5**), and hydrogen (**ThioFluor-6**). To alter the electron acceptor moiety, while keeping the same **ThioFluor** scaffold, we are more limited as iminium formation requires an aldehydic moiety. To this end we made **ThioFluor-7** and **ThioFluor-11**, in which a trifluoromethyl group is introduced beta to the aldehyde and a fluoro group is introduced alpha to the aldehyde, respectively.

The second set of analogues was synthesized in order to test the effect of conjugation across the pi system. For this reason, we kept the electron donating and electron acceptor groups equivalent to **ThioFluor**, and chose to alter the aromatic rings between the two. Both the bithiophene and biphenyl variants, **ThioFluor-8** and **ThioFluor-10**, were synthesized as well as **ThioFluor-9**, in which the order of the two aromatic rings is reversed. It should be noted that the shortest chromophore **ThioFluor-2** was made in order to create a dye that is more cyanine like, in that it should remain fully conjugated. Cyanine dyes have been a focus in recent research in the field of fluorescence,[36-42] as well as in our lab.[43] They satisfy many of the aforementioned requirements for use in imaging such as exhibiting fluorogenic behavior, high molar extinction coefficient and high quantum yields. Additionally, as shown in **Chapter II.1**, merocyanine aldehyde can bind optimized hCRBPII mutants, with half-lives as fast as ten seconds.

With the synthesized analogues in hand, we first chose to test the spectroscopic properties of the free aldehyde, and the model Schiff bases (SB)

and protonated Schiff bases (PSB). This data will provide insight into the relative wavelengths of both absorption and emission for the **ThioFluor** analogues.

III.3 Evaluation of ThioFluor analogues with *n*-butyl amine in ethanol

Spectroscopic properties of the ligands were first measured. Results are summarized in **Table III-2**, shown in increasing order of absorption wavelength. All spectroscopic data was recorded in ethanol. It should be noted that **ThioFluor-7**, with the trifluoromethyl group at the beta position of the aldehyde, exists in equilibrium with its hemiacetal form in ethanol. The extinction coefficient was measured by adding triethylamine to the solution to force the equilibrium almost solely to the hemiacetal form (λ_{abs} = 340 nm). **ThioFluor-10** also shows two bands in absorption; presumably the blue-shifted absorption band is due to a twisted ground state, which does not allow the chromophore to be coplanar.

A plot of Hammett value versus absorption wavelength is shown in **Figure III-3**.[44] An increase in electron donating ability of the group at the para position of the phenyl ring leads to a red-shift in absorption, increasing in order for trifluoromethyl, hydrogen, methoxy and *N,N*-dimethyl. Changing the aromatic rings also results in regulation of wavelength; interestingly, reversal of the thiophene and phenyl rings in **ThioFluor-9** leads to no significant change in absorption wavelength from the parent **ThioFluor**. However, the biphenyl **ThioFluor-10** is more blue-shifted than **ThioFluor**, while the bithiophene **ThioFluor-8** is more red-shifted. It stands to reason that **ThioFluor-8** is more conjugated. This is supported by the resonance energies of benzene and

thiophene. The resonance energy of benzene is 36 kcal/mol, while the resonance energy of thiophene is 29 kcal/mol;[45] thus it should be easier to break aromaticity and resonate through a thiophene pi linker, as opposed to a phenyl linker. The overall effect is better delocalization over the whole molecule, which leads to a smaller energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO).

Table III-2. Spectroscopic properties of **ThioFluor** analogues (free aldehyde) in ethanol.

Ligand	λ_{abs}	λ_{em}[a]	ε ($M^{-1}cm^{-1}$)
ThioFluor-5	360	402	23,412
ThioFluor-6	364	396	17,708
ThioFluor-10	271, 369	556	22,071
ThioFluor-4	384	447	21,509
ThioFluor-9	416	592	15,688
ThioFluor-3	418	562, 626, 660	26,592
ThioFluor-11	429	~ 556, 630 653	26,900
ThioFluor	430	634	27,021
ThioFluor-7	320, 438	601	20,682
ThioFluor-2	445	498	29,763
ThioFluor-8	472	669	14,821

[a]Emission maxima is recorded as a result of exciting the absorption peak that is expected to correspond to the fully conjugated aldehyde.

Unless otherwise noted, the emission peaks are singular, implying that emission occurs from only one excited state. The outliers are seen in **ThioFluor-3** and **ThioFluor-11**. Azetidine has been hypothesized to minimize formation of the twisted intramolecular charge transfer (TICT) in numerous fluorophore

families,[46, 47] and perhaps that is the reason for the multiple emission bands observed upon excitation of the aldehyde in solution. Presumably emission occurs from both the locally excited state and the intramolecular charge transfer state, which should correspond to the blue-shifted and red-shifted bands, respectively. **ThioFluor**, possessing a para *N,N*-dimethyl substituent displays only a red-shifted emission band (which is close in wavelength to the red-shifted emission band in **ThioFluor-3**). **ThioFluor-5** and **ThioFluor-6**, which do not have electron donating groups in the chromophore, are not capable of charge transfer, and thus display small Stokes shift upon excitation of the aldehyde.

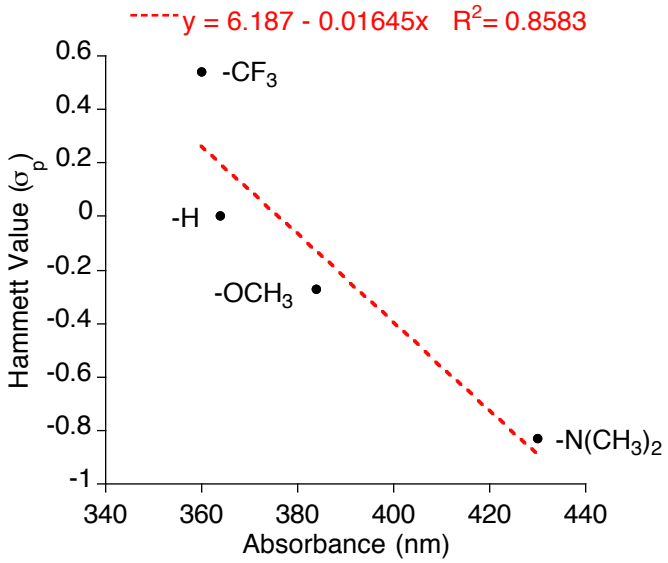


Figure III-3. Plot of Hammett values versus absorption wavelength for **ThioFluor-5**, **ThioFluor-6**, **ThioFluor-4** and **ThioFluor**.

Model Schiff bases were then prepared by reacting each **ThioFluor** analogue with *n*-butyl amine in ethanol. In all cases, the SB displays an absorption maxima blue-shifted than the free aldehyde. Subsequently the protonated Schiff base (PSB) was formed by acidification of the SB.

Spectroscopic properties are detailed in **Table III-3**. Not unexpectedly, the most red-shifted PSB belongs to the most conjugated aldehyde (**ThioFluor-8**). Interestingly though, the iminiums formed with **ThioFluor-11** and **ThioFluor-7** are more red-shifted than with **ThioFluor**, even though the absorption wavelengths of the free aldehyde were approximately the same.

Upon excitation of the PSB, singular peaks are observed. With the exception of **ThioFluor-5**-PSB, **ThioFluor-6**-PSB and **ThioFluor-4**-PSB (with a trifluoromethyl group, hydrogen, and methoxy group at the para position of the phenyl, respectively), the emissions fall in the red emission range. Interestingly the largest Stokes shift (238 nm) is observed from **ThioFluor-10**-PSB. Similarly reported biphenyl push-pull probes also display large Stokes shifts (LSS).[48]

Table III-3. Spectroscopic properties of **ThioFluor** analogues (SB and PSB) in ethanol.

Ligand	SB		PSB		
	λ_{abs}	λ_{em}[a]	λ_{abs}	λ_{em}	Stokes Shift
ThioFluor-5	352	399	403	489	86
ThioFluor-6	347	395	418	503	85
ThioFluor-10	349	488	443	681	238
ThioFluor-4	358	440	440	556	116
ThioFluor-9	390	535	506	690	184
ThioFluor-3	392	550	508	690	182
ThioFluor-11	397	552	543	698	155
ThioFluor	400	540	521	688	167
ThioFluor-7	317, 397	~ 487, 585	336, 568	-	-
ThioFluor-2	400	485	510	531	21
ThioFluor-8	432	584	600	709	109

[a]Emission recorded as a result of exciting the most red-shifted absorption peak.

After testing the **ThioFluor** analogues with *n*-butyl amine in solution, we chose to test them with four hCRBPII protein mutants. As stated previously, the problem with the hCRBPII/**ThioFluor** protein series (i.e. generally when acidic residues are present) is that many of the blue-shifted mutants suffer from a severely suppressed pK_a. For this reason, we chose to start with mutants that were expected to exhibit a pK_a higher than physiological pH, as evidenced by the formation of iminium at a pH of approximately 7. We chose to test the following mutants – Q108K:K40L:T51V:T53S:R58W:Y19W:L117E, Q108KK40L:T51V:T53S:R58W:Y19W:A33W, Q108K:K40L:R58F, and Q108K:K40L:T51V:T53S:R58W.

Firstly, Q108K:K40L:T51V:T53S:R58W:Y19W:L117E is an ideal mutant when coupled with **ThioFluor** because it has a high quantum yield (0.31), high pK_a (11.2), and this hCRBPII/**ThioFluor** complex was successfully utilized in live cell imaging. Additionally, we predicted that **ThioFluor** analogues would bind in the same conformation as ThioFluor, due to the stabilization of the iminium by L117E. Thus, we predicted that we could make a direct correlation between the electron donating ability of the para-substituent or overall conjugation and the spectroscopic properties of the resultant complex, such as absorption wavelength, emission wavelength, quantum yield and the rate of iminium formation.

The other three hCRBPII mutants chosen also showed substantial iminium at physiological pH when coupled to **ThioFluor**.

Q108K:K40L:T51V:T53S:R58W:Y19W:L117E was the most blue-shifted mutant (λ_{abs} = 558 nm and λ_{em} = 663 nm with **ThioFluor**) tested with **ThioFluor** analogues. It should be noted that we did attempt to test the more blue-shifted mutant Q108K:K40D:T51V:R58L:Q38F:Q4F:V62E with **ThioFluor** analogues, but the protein was not acid stable enough to form an iminium with many of the analogues, due to their extremely low pK_as. spanning to the red wavelength regime we tested Q108K:K40L:R58F (λ_{abs} = 597 nm and λ_{em} = 678 nm with **ThioFluor**), Q108K:K40L:T51V:T53S:R58W (λ_{abs} = 623 nm and λ_{em} = 697 nm with **ThioFluor**) and Q108K:K40L:T51V:T53S:R58W:Y19W:A33W (λ_{abs} = 666 nm and λ_{em} = 724 nm with **ThioFluor**).

III.4 Coupling of ThioFluor analogues to Q108K:K40L:T51V:T53S:R58W:Y19W:L117E

In **Chapter II.11.2**, it was shown that **ThioFluor** could successfully label expressed Q108K:K40L:T51V:T53S:R58W:Y19W:L117E in mammalian cells. It is an ideal mutant when coupled with **ThioFluor** for live cell imaging due to its high quantum yield (0.31), high pK_a (11.2) and reasonable half-life of 82 min (based on fitting to second order kinetics; measured at 23 °C with 20 μM protein and 0.5 equiv **ThioFluor** at pH ~ 7). With the use of **ThioFluor** analogues, it would be ideal to retain the high iminium pK_a and quantum yield, while decreasing the rate of iminium formation. The longer half-life, requiring an hour of **ThioFluor** incubation, hampers visualization of fluorescence within this time domain and prohibits the collection of real time measurements.

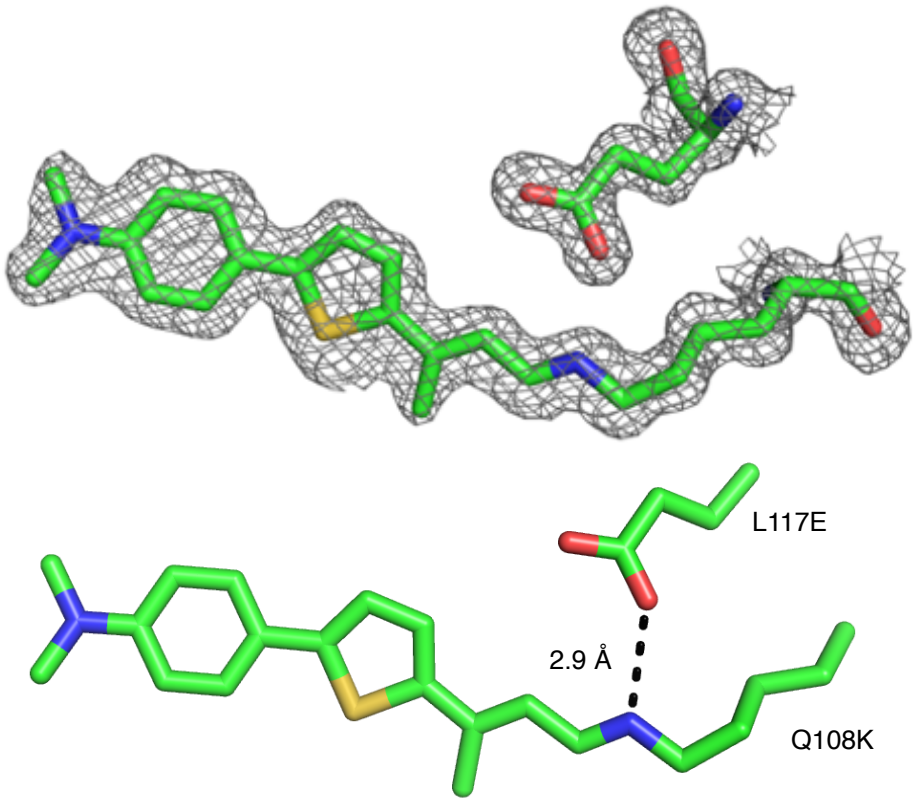


Figure III-4. In the hCRBPII heptamutant Q108K:K40L:T51V:T35S:R58W:Y19W:L117E, the distance between the iminium nitrogen and L117E is 2.9 Å; it is this ionic interaction that presumably leads to the high pK_a and holds the iminium as a *trans* isomer. Density is shown at 1σ.

From the crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**, a *trans* iminium of **ThioFluor** is observed, which puts the iminium nitrogen just 2.9 Å away from the neighboring L117E residue (**Figure III-4**). Presumably, the ionic interaction between the positively charged iminium and the negatively charged carboxylate side chain of L117E leads to the high observed pK_a of 11.2. Indeed, upon acid titration of the Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** complex from a neutral pH, a red-shift in absorption is observed, presumably due to protonation of L117E. The pK_a of L117E was found to be 6.6, rendering the

carboxylate species predominate at pH 7. This result supports the theory that the elevated iminium pK_a is due to an ionic interaction with the carboxylate side chain of the nearby glutamate residue.

As aforementioned, we predicted that the pK_a of the iminium should be correlated with the degree of conjugation along the chromophore. To this end, the electron donating ability of the para substituent should affect iminium pK_a, in which a stronger electron-donating group can stabilize the positive charge of the iminium nitrogen better. However, it was hypothesized that **ThioFluor** analogues, if bound to Q108K:K40L:T51V:T53S:R58W:Y19W:L117E in the same manner, could all adopt a similar *trans* iminium and be stabilized by the carboxylate present at residue 117, leading to the iminium possessing a high pK_a, rendering the effect of chromophore conjugation insignificant.

Assuming binding of the chromophore to be comparable (in terms of chromophore trajectory and isomer of the resultant iminium), we should be able to make a direct correlation between the electron donating ability of the para-substituent or overall conjugation and the spectroscopic properties of the resultant complex, such as absorption wavelength, emission wavelength, quantum yield and the rate of iminium formation.

Of the eleven chromophores tested with the hCRBPII heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E, formation of iminium was observed for ten; it appears that iminium formation does not occur for **ThioFluor-11**. On steric grounds, fluorine is not significantly larger than hydrogen, so this

result is unexpected. A summary of the spectroscopic properties of the **ThioFluor** analogues, coupled with Q108K:K40L:T51V:T53S:R58W:Y19W:L117E is provided in **Table III-4**.

Table III-4. Spectroscopic properties of **ThioFluor**-PSB analogues coupled to Q108K:K40L:T51V:T53S:R58W:Y19W:L117E.

Ligand	λ_{abs}	λ_{em}	Φ	ε ($M^{-1}cm^{-1}$)	pK_a
ThioFluor-5	423	-	-	23,530	10.9
ThioFluor-6	424	502	-	13,820	10.6
ThioFluor-4	466	543	-	29,878	10.9
ThioFluor-10	489	675	0.06	7,050	11.3
ThioFluor-2	515	536	0.13	50,578	11.3
ThioFluor-3	548	679	0.27	21,223	10.9
ThioFluor	558	673	0.31	15,091	11.2
ThioFluor-9	562	700	0.02	15,406	11.1
ThioFluor-8	641	715	0.09	22,948	11.2
ThioFluor-7	714	780	-	22,517	4.3

Generally the absorption wavelength seems to be correlated to the degree of conjugation between the electron donor and electron acceptor groups. The most blue-shifted analogues when coupled with Q108K:K40L:T51V:T53S:R58W:Y19W:L117E are **ThioFluor-5** (λ_{abs} = 423 nm) and **ThioFluor-6** (λ_{abs} = 424 nm), which possess a trifluoromethyl group and hydrogen at the para-phenyl position, respectively. The pigments formed are essentially nonfluorescent, but their iminiums do exhibit high pK_as. **ThioFluor-4** contains the electron donating methoxy group at the para position, and as expected, is more red-shifted than with hydrogen or trifluoromethyl groups (λ_{abs} =

466 nm). Again, the iminium is essentially nonfluorescent, suggesting that methoxy is not electron donating enough to generate a strong enough push-pull system.

Further, *N,N*-dimethylamino and azetidine groups provide the largest red-shifts amongst those tested at the para position (λ_{abs} = 558 nm and 548 nm, respectively). Notably, a bathochromic shift of 134 nm is achieved in absorption wavelength via the introduction of a para *N,N*-dimethyl amino group, as compared to **ThioFluor-6** which simply contains an electronically neutral hydrogen atom. Interestingly, the quantum yields of the para substituted *N,N*-dimethylamino and azetidine groups are similar. Previous studies by the Lavis lab showed that replacement of the *N,N*-dimethyl group with azetidine (and later aziridine) led to a large increase in quantum yield relative to that of the parent fluorophore in numerous fluorophores such as coumarin, napthalimide, acridine, rhodol, carborhodamine and oxazine.[46, 47] It was predicted that azetidine and aziridine might inhibit TICT in these cases. With this knowledge, we are compelled to hypothesize that with **ThioFluor**, fluorescence is not lost due to nonradiative decay from the excited state as a result of TICT in the case of **ThioFluor**. Interestingly, the crystal structures of **ThioFluor** and **ThioFluor-3** show that both the *N,N*-dimethyl and azetidine groups are planar with the rest of the chromophore, indicating that they may exhibit high TICT resistance.

Because **ThioFluor-11** did not form an iminium with Q108K:K40L:T51V:T53S:R58W:Y19W:L117E, only one modification of the

electron acceptor moiety was tested. **ThioFluor-7**, which contains a trifluoromethyl group in the beta position to the aldehyde led to the most red-shifted absorption wavelength obtained in this hCRBPII mutant. From the parent **ThioFluor**, a remarkable red-shift of 156 nm is obtained from the structurally small modification (λ_{abs} = 714 nm).

Just as a stronger electron donating group in the para position leads to a bathochromic shift, a red-shift can also be obtained by decreasing the aromaticity of the central rings, such that it is easier for electrons to delocalize across the entire chromophore. As previously described, the resonance energy of benzene (36 kcal/mol) is higher than that of thiophene (29 kcal/mol). For this reason, it should be easier to break aromaticity and resonate through a thiophene π linker, as opposed to a phenyl linker. The overall effect is better delocalization over the whole molecule, leading to a smaller gap between the HOMO and LUMO. To our satisfaction, substitution of the phenyl ring in **ThioFluor**, to yield the bithiophene scaffold in **ThioFluor-8**, results in a red-shift of almost 100 nm from **ThioFluor** (λ_{abs} = 641 nm), which is indicative of a more conjugated π system. It is hypothesized that the two aromatic rings in **ThioFluor-8** are coplanar when bound to the protein, because **ThioFluor** is essentially planar in hCRBPII crystal structures.

Interestingly, replacement of the thiophene ring in **ThioFluor** with a second phenyl ring yields an approximately 70 nm hypsochromic shift (λ_{abs} = 489 nm), yet with a remarkable Stokes shift of 186 nm. Previously it has been shown

that the torsion on push-pull biphenyl analogs enhances their fluorescence solvatochromism. The observed hypsochromic shift in absorption wavelength could occur for various reasons, but we will make two suggestions here. It is possible that steric hindrance forces the biphenyl **ThioFluor-10** to remain in a twisted ground state, such that only the two aromatic phenyl rings are coplanar, and the donating *N,N*-dimethyl is rotated out of the plane (**Figure III-5a**). This would account for the large blue-shift observed in absorption wavelength. Subsequently, a planarized intramolecular charge transfer (PLICT) could explain the large Stokes shift.[49, 50] In this scenario, upon excitation of the molecule, a charge transfer first occurs from the donor to acceptor. Then the single bond between the *N,N*-dimethyl group and phenyl ring can rotate around 90° in order to stabilize the positive charge on the donor group. Due to the geometric relaxation in the excited state, a large Stokes shift and quantum yield should be observed.

An alternative explanation would be almost the exact opposite. If **ThioFluor-10** is planar in the ground state when bound to the protein (**Figure III-5b**), it would stand to reason that the large hypsochromic shift in absorption wavelength is solely due to the decrease in delocalization along the molecule, due to the increasing aromaticity of the second ring (from thiophene to phenyl). Assuming, the donor and acceptor moieties are in the same plane in the ground state, excitation could result in an intramolecular charge transfer (ICT) from the donor to acceptor.[49-51] Subsequent relaxation in the excited state is achieved by

rotation around the single bond connecting the donor and acceptor, yielding a perpendicular conformation. This twisted intramolecular charge transfer (TICT) state typically emits dual fluorescence in emission spectra and results in large Stokes shift.

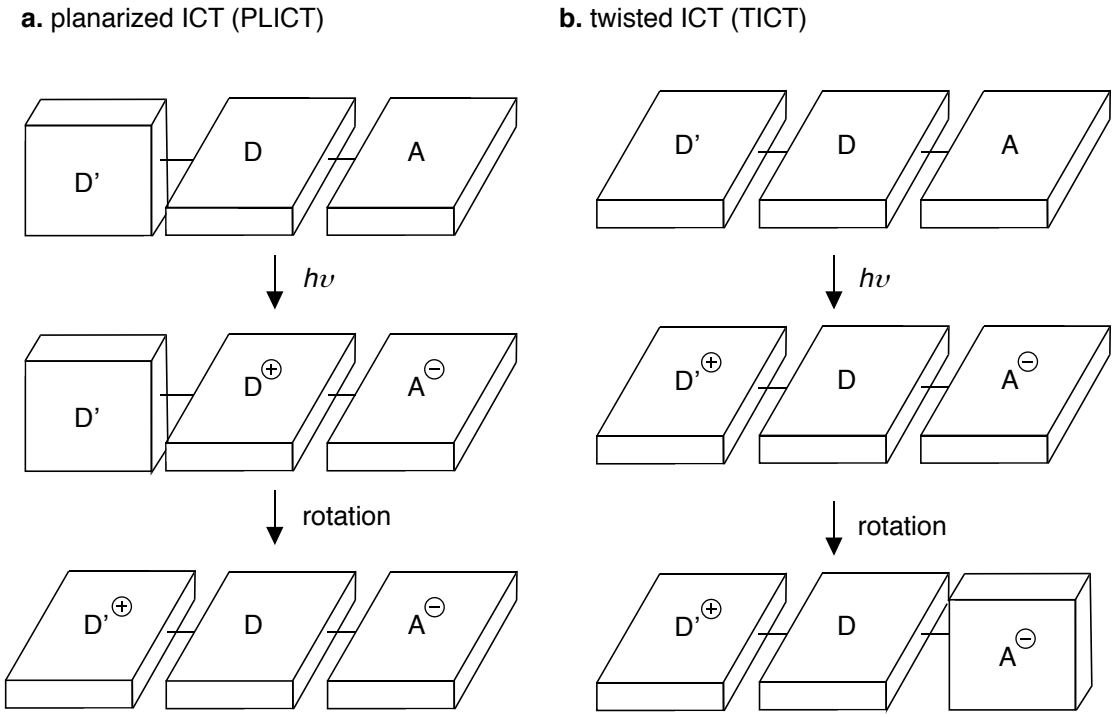


Figure III-5. Schematic representation of the geometric arrangement of the donor and acceptor groups upon excitation in planarized intramolecular charge transfer (PLICT) and twisted intramolecular charge transfer (TICT).

The emission of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor-10** appears to follow a normal distribution, and clearly no conclusion can be drawn as to the mechanism which leads to the LSS. One subtle factor that may provide information is the rate of iminium formation. Iminium formation of **ThioFluor-10** with Q108K:K40L:T51V:T53S:R58W:Y19W:L117E is not significantly different than with **ThioFluor**. This suggests that the aldehyde is of comparable electrophilicity. As seen from the crystal structure of

Q108K:K40L:T51V:T53S:R58W:Y19W:L117E with **ThioFluor**, **ThioFluor** is planar in the ground state; this would lead one to hypothesize that **ThioFluor-10** possesses a completely planar configuration of the chromophore in the ground state. Nonetheless, crystal structures may provide more information, as they could provide details of the ground state chromophore configuration. It would also be interesting to investigate the spectroscopic properties of azetidine substituted **ThioFluor-10**. If fluorescence does involve a TICT mechanism, the inclusion of azetidine may decrease or eliminate TICT and improve fluorescence properties.

Changing the order of the phenyl and thiophene rings does not seem to have a significant effect on the absorption wavelength. However, the quantum yield is drastically decreased in comparison to **ThioFluor**. They are 0.31 and 0.02 for **ThioFluor** and **ThioFluor-9**, respectively. Interestingly, removal of the phenyl ring altogether leads to an absorption maxima of 515 nm, which will be shown to remain constant in all proteins tested. **ThioFluor-2** was evaluated because it was expected to lead to a more cyanine-like dye. Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor-2** exhibits a high iminium pK_a, high extinction coefficient and extremely small Stokes shift. Both of these characteristics are reminiscent of typical cyanine dyes.

A plot of absorption wavelength versus emission wavelength is given in **Figure III-6a**. In general a linear trend is observed. However, two outliers are clearly observed. These belong to Q108K:K40L:T51V:T53S:R58W:Y19W:L117E coupled with **ThioFluor-2** and **ThioFluor-10**. As **ThioFluor-2** is cyanine-like and

the wavelength is not regulatable (to be shown later) due to its more permanent resonance, it is acceptable to remove this data from the series. On the other hand, the emission of **ThioFluor-10** is much more red-shifted than would have been predicted; Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor-10** exhibits a Stokes shift of 186 nm. This is likely due to an intramolecular charge transfer from the electron donating *N,N*-dimethyl amino group to the electron accepting iminium. A plot of absorption wavelength versus emission wavelength shows a linear correlation (**Figure III-6b**).

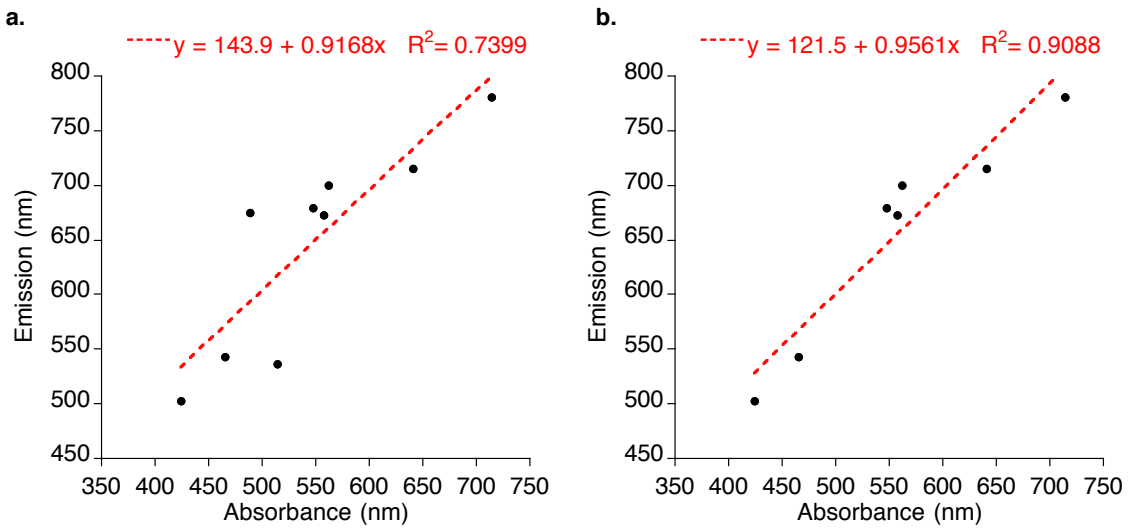


Figure III-6. Emission maxima versus absorption maxima for **ThioFluor** analogues bound to hCRBPII heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E.

Rate constants and half-lives are given in **Table III-5**. All rates were plotted according to second order kinetics (measured by UV-Vis at 23 °C with 20 μM protein and 0.5 equiv ligand at pH ~ 7). Drastically different rates of iminium formation are observed. The fastest reactions are those that possess the weakest push-pull system. Presumably the trifluoromethyl group and hydrogen in the para position, in **ThioFluor-5** and **ThioFluor-6**, leads to a more electrophilic

aldehyde. Following the same trend, a methoxy substituent results in almost a five-fold enhancement of rate. Interestingly, reversal of the thiophene and phenyl rings leads to a six-fold increase in reaction rate, suggesting that donation of the *N,N*-dimethylamino group to the aldehyde is much less when aromaticity has to be broken further from the electron donating site. In other words, the aldehyde is more electrophilic in **ThioFluor-9** than **ThioFluor**. This is further supported by the extremely low quantum yield of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor-9** as compared to Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**. Unfortunately, the rate of iminium formation cannot be measured with **ThioFluor-7** under analogous conditions, as the pK_a is several units lower than necessary to observe iminium formation.

Table III-5. Rates and half-lives of **ThioFluor**-PSB analogues coupled to Q108K:K40L:T51V:T53S:R58W:Y19W:L117E.

Ligand	k ($min^{-1}M^{-1}$)	$t_{1/2}$ (min)
ThioFluor-6	31,151	3
ThioFluor-5	10,699	9
ThioFluor-9	7,624	13
ThioFluor-4	6,001	17
ThioFluor-3	3,659	27
ThioFluor	1,224	82
ThioFluor-10	1,082	92
ThioFluor-8	284	352
ThioFluor-2	89	1124

Conversely, the most conjugated systems lead to much slower iminium formation. Both the cyanine like **ThioFluor-2** and bithiophene analog **ThioFluor-8** are substantially slower than **ThioFluor**. These slow reaction rates would find essentially no use in live cell imaging experiments.

Upon analysis of **ThioFluor** analogues with Q108K:K40L:T51V:T53S:R58W:Y19W:L117E, the following general trends are concluded, which will be discussed in greater detail below.

- As expected, absorption wavelength seems to be correlated to the degree of conjugation between the electron donor and electron acceptor groups.
- Emission wavelength is generally linearly correlated with absorption wavelength, albeit no significant environmental factors are in play.
- The reaction rate is faster for those chromophores containing electron-withdrawing substituents in the para position, which is not surprising, as less electron donation should result in a more electrophilic aldehyde center.
- The iminium pK_a of all the hCRBPII/**ThioFluor** analog complexes, with the exception of **ThioFluor-7**, are high. This indicates that the ligand experiences the same iminium environment, presumably due to an ionic interaction between the protonated nitrogen of the iminium and the negatively charge carboxylate of the neighboring L117E residue. It is predicted that the pK_a would be variable were this interaction was not present.

- Quantum yield is hard to assess. However, it is quite apparent that a push-pull system needs to be present in order to show fluorescence.

Three other hCRBPII mutants were chosen for binding with the **ThioFluor** analogues to verify the trends observed.

III.5 Coupling of ThioFluor analogues to Q108K:K40L:T51V:T53S:R58W:Y19W:A33W

As previously mentioned, the iminium pK_a of **ThioFluor** analogs with the heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E were high, presumably due to stabilization via the carboxylate installed at residue 117. We now turned our attention to a protein of similar sequence, but with no stabilizing ionic interaction at the iminium. It is hypothesized that without the interaction, the pK_a of the iminium will be variable, dependent on chromophore structure. The heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:A33W was chosen because it is one of the more red-shifted mutants when coupled with **ThioFluor** (λ_{abs} = 666 nm), and it was desirable to see if the analogues of **ThioFluor** could also regulate wavelength, as was shown in **Chapter II**. Spectroscopic properties of the iminium formed between **ThioFluor** analogs and Q108K:K40L:T51V:T53S:R58W:Y19W:A33W are given in **Table III-6**.

It is apparent that the absorption wavelength is more red-shifted as conjugation is extended across the molecule. Para substitution of trifluoromethyl (**ThioFluor-5**, λ_{abs} = 484 nm), hydrogen (**ThioFluor-6**, λ_{abs} = 495 nm), methoxy (**ThioFluor-4**, λ_{abs} = 530 nm) and *N,N*-dimethyl (**ThioFluor**, λ_{abs} = 666 nm)

becomes increasingly more red-shifted with the increased electron donating ability of the para substituent. It appears that the simple introduction of *N,N*-dimethyl at the para position is responsible for a 171 nm red-shift in absorption wavelength, five unit increase in iminium pK_a and quantum yield of 0.14.

Table III-6. Spectroscopic properties of **ThioFluor**-PSB analogues coupled to Q108K:K40L:T51V:T53S:R58W:Y19W:A33W.

Ligand	λ_{abs}	λ_{em}	Φ	ε ($M^{-1}cm^{-1}$)	pK_a
ThioFluor-5	484	-	-	39,228	3.7
ThioFluor-6	495	-	-	36,566	4.0
ThioFluor-2	519	536	-	36,594	11.1
ThioFluor-4	530	577	-	24,875	9.1
ThioFluor-10	603	733	-	15,547	8.7
ThioFluor-3	658	732	0.10	40,965	9.2
ThioFluor	666	724	0.14	57,316	9.3
ThioFluor-9	693	769	-	31,531	9.5
ThioFluor-11	694	744	-	-	-
ThioFluor-8	722	751	0.02	62,382	9.1
ThioFluor-7	797	-	-	-	-

It should be noted that the iminium pK_a in both complexes of **ThioFluor-5** and **ThioFluor-6** with Q108K:K40L:T51V:T53S:R58W:Y19W:A33W are approximately five units lower than **ThioFluor**. This can be due to the fact that there is no stabilization of the iminium via an electron donating para substituent; it is also possible that the opposite isomer of the iminium is formed, which exhibits a decreased pK_a. As opposed to the heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E coupled with **ThioFluor** analogues, the quantum yields of

Q108K:K40L:T51V:T53S:R58W:Y19W:A33W/**ThioFluor** analogues are much smaller or nonexistent. For example, the quantum yield of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** (excited at 558 nm) is 0.31, while Q108K:K40L:T51V:T53S:R58W:Y19W:A33W/**ThioFluor** is 0.14 (excited at 666 nm). The quantum yield can be decreased as a result of the bathochromic shift of over 100 nm, or due to the removal of L117E, which may confer restriction on the chromophore.

Modification of the electron acceptor moiety with alpha fluoro (**ThioFluor-11**) and beta trifluoromethyl (**ThioFluor-7**) groups leads to red-shifts of 28 nm and 131 nm from the parent **ThioFluor**, respectively. Q108K:K40L:T51V:T53S:R58W:Y19W:A33W/**ThioFluor-7** absorbs in the infrared at an astonishing 797 nm, but does suffer from a severely depressed pK_a (it cannot be measured because it is too low) and fluorescence is not observed upon excitation of the iminium at 797 nm. It should be noted that our fluorometer does restrict fluorescence observation to below 900 nm; although we do not think the complex will emit higher than this cutoff, it has to be considered as a possibility.

As previously suggested, the aromaticity of the central hetero rings can significantly alter the spectroscopic properties of the hCRBPII bound ligand. Better delocalization across the molecule is observed for the bithiophene **ThioFluor-8** (λ_{abs} = 722 nm) as compared to **ThioFluor** (λ_{abs} = 666 nm). The biphenyl **ThioFluor-10** (λ_{abs} = 603 nm) is again more blue-shifted. Removal of

the phenyl ring results in an iminium with absorption maxima of 519 nm, which is consistent with other hCRBPII mutants tested.

Rate constants and half-lives are given in **Table III-7**. All rates were plotted according to second order kinetics (measured by UV-Vis at 23 °C with 20 μM protein and 0.5 equiv ligand at pH ~ 7). Varying rates of iminium formation are observed with the different **ThioFluor** analogues. Rate constants for many of the ligands could not be measured by UV-Vis because of their extremely low iminium pK_as; additionally, the ligands were not fluorescent when bound to Q108K:K40L:T51V:T53S:R58W:Y19W:A33W, negating the measurement of rates by fluorescence. Generally, the rate of iminium formation is slower than those observed with Q108K:K40L:T51V:T53S:R58W:Y19W:L117E (**Table III-5**). Presumably, the carboxylate at residue 117 provides some means of acid catalyzed iminium formation. It is also possible that a different isomer of the iminium is formed, yet there is not direct evidence to support this. As similarly reported for the previous hCRBPII heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E, the rate of iminium formation is faster with the less electron donating methoxy group at the para position (**ThioFluor-4**) as compared to the *N,N*-dimethyl group. The fastest rate is observed for **ThioFluor-9**, which is presumably due to its more electrophilic aldehyde. The only pitfall is that the hCRBPII/**ThioFluor-9** complex is not fluorescent.

Table III-7. Rates and half-lives of **ThioFluor**-PSB analogues coupled to Q108K:K40L:T51V:T53S:R58W:Y19W:A33W.

Ligand	k ($min^{-1}M^{-1}$)	$t_{1/2}$ (min)
ThioFluor-9	6252	16
ThioFluor-4	2903	34
ThioFluor-10	503	199
ThioFluor-3	452	221
ThioFluor-8	284	352
ThioFluor	238	420

III.6 Coupling of ThioFluor analogues to Q108K:K40L:R58F

Similar trends in absorption wavelength were obtained for the hCRBPII triple mutant Q108K:K40L:R58F as for the heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:A33W. Spectroscopic properties for the iminium formed upon condensation of Q108K:K40L:R58F and the **ThioFluor** analogues are detailed in **Table III-8**.

Para substitution of trifluoromethyl (**ThioFluor-5**, λ_{abs} = 442 nm), hydrogen (**ThioFluor-6**, λ_{abs} = 458 nm), methoxy (**ThioFluor-4**, λ_{abs} = 496 nm) and *N,N*-dimethyl (**ThioFluor**, λ_{abs} = 597 nm) becomes increasingly more red-shifted with the increased electron donating ability of the para substituent. As observed in the previous two hCRBPII mutants, only the iminium with **ThioFluor** is fluorescent, exhibiting a quantum yield of 0.18.

It should be noted that the iminium pK_as for the ligands are much more similar to each other. This is in contrast to complexes with Q108K:K40L:T51V:T53S:R58W:Y19W:A33W, in which **ThioFluor-5** and **ThioFluor-6** had iminium pK_a values five units below that of **ThioFluor**. It is

predicted that the large change in pK_a is due to isomerization of the iminium (this will be further supported by coupling of the ligands to Q108K:K40L:T51V:T53S:R58W in **Section III.7**). Our lab has previously observed the trend that CRABPII/retinal *cis* imine showed a higher pK_a than the *trans* imine;[52, 53] presumably, a similar phenomena is occurring here.

Table III-8. Spectroscopic properties of **ThioFluor**-PSB analogues coupled to Q108K:K40L:R58F.

Ligand	λ_{abs}	λ_{em}	Φ	ε ($M^{-1}cm^{-1}$)	pK_a
ThioFluor-5	442	-	-	13,721	8.0
ThioFluor-6	458	509	-	12,634	8.4
ThioFluor-10	487	675	0.02	4,632	8.2
ThioFluor-4	496	554	-	31,580	8.8
ThioFluor-2	518	534	-	34,188	10.9
ThioFluor-3	581	683	0.13	24,105	8.7
ThioFluor	597	678	0.18	18,136	9.3
ThioFluor-9	602	709	-	14,235	8.2
ThioFluor-11	619	695	-	-	-
ThioFluor-8	673	712	0.06	20,147	9.1

Absorption wavelength is altered via substitution of the central aromatic rings. In the case of the bithiophene containing **ThioFluor-8**, the absorption maxima for the iminium is 673 nm, while the biphenyl **ThioFluor-10** shows a hypsochromic shift, with absorption maxima of 487 nm. Removal of the phenyl ring (**ThioFluor-2**) results in an iminium with absorption maxima of 518 nm, which is shown to be static in all hCRBPII mutants tested.

III.7 Coupling of ThioFluor analogues to Q108K:K40L:T51V:T53S:R58W

The general trends already discussed for the hCRBPII/**ThioFluor** analogues complexes are applicable here, including absorption regulation and the necessity of a push-pull system to generate fluorescence. Spectroscopic data for the iminium complexes are presented in **Table III-9**.

Table III-9. Spectroscopic properties of **ThioFluor**-PSB analogues coupled to Q108K:K40L:T51V:T53S:R58W.

Ligand	λ_{abs}	λ_{em}	Φ	pK_a[a]
ThioFluor-5	446	503	-	-
ThioFluor-6	469	517	-	nd, 7.1
ThioFluor-4	502	556	-	4.8, 8.4
ThioFluor-2	516	531	-	10.9
ThioFluor-10	517	684	-	7.7
ThioFluor-3	610	699	0.08	6.4, 8.7
ThioFluor	623	697	0.08	6.4, 8.4
ThioFluor-9	639	745	-	nd, 8.4
ThioFluor-11	661	723	-	4.3
ThioFluor-8	706	736	0.02	8.1

[a]Presumably the two pK_as correspond to the *trans* and *cis* iminiums.

Only one data appears to be an outlier from the trends observed thus far. In the cases of hCRBPII mutants Q108K:K40L:T51V:T53S:R58W:Y19W:A33W and Q108K:K40L:R58F, the iminium formed with **ThioFluor-8** (λ_{abs} = 722 nm and λ_{abs} = 673 nm, respectively) is more red-shifted than for **ThioFluor-11** (λ_{abs} = 694 nm and λ_{abs} = 619 nm, respectively). In this case, the bithiophene **ThioFluor-8** exhibits a more red-shifted iminium than **ThioFluor-11**.

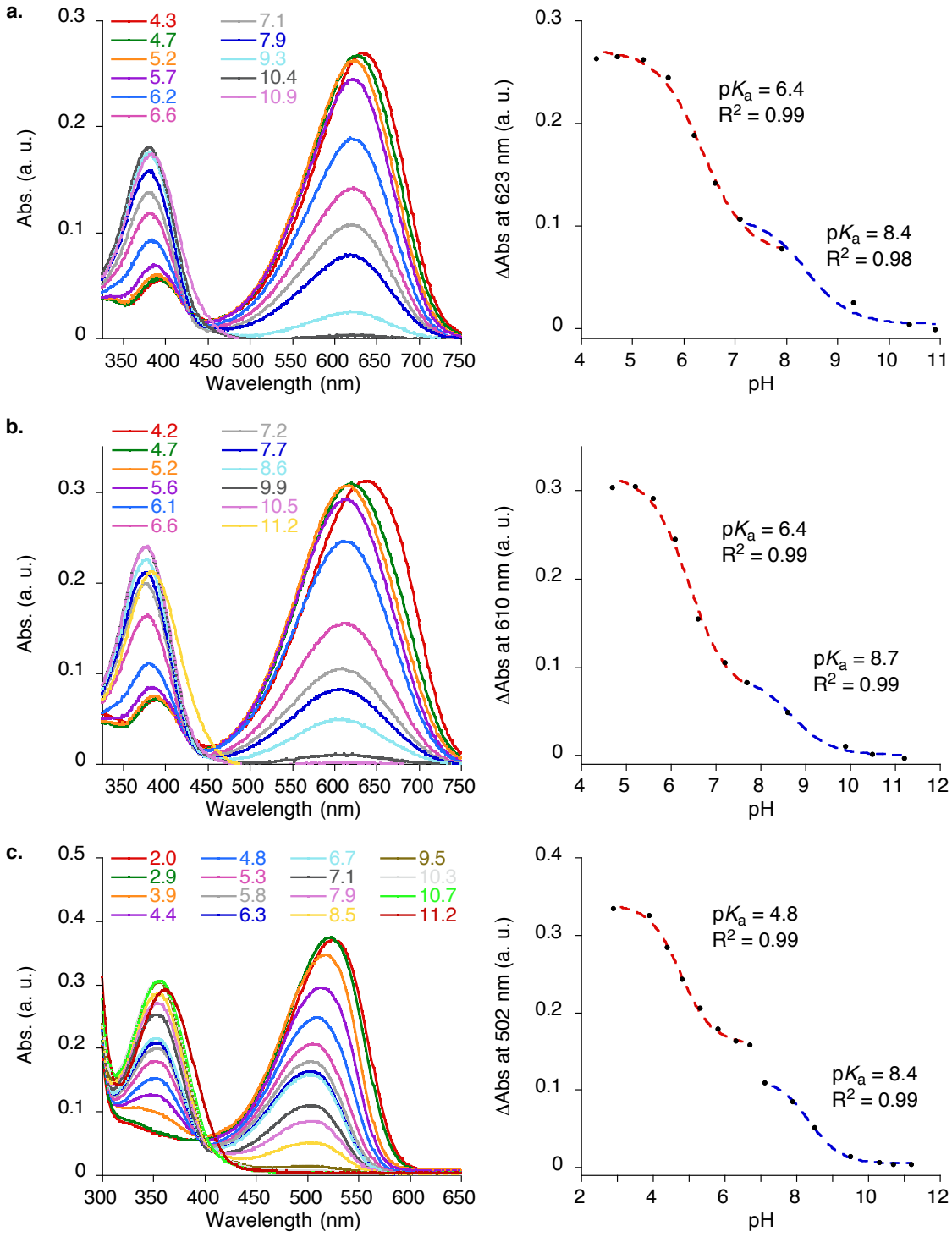


Figure III-7. Double iminium pK_as observed upon coupling of hCRBPII pentamutant Q108K:K40L:T51V:T53S:R58W with **a**. **ThioFluor**, **b**. **ThioFluor-3** and **c**. **ThioFluor-4**. Absorption intensity of the iminium is plotted as a function of pH. Two curve fittings provide the two different pK_a values.

While absorption wavelength regulation supports the general trends already observed in the other hCRBPII mutants, one interesting new observation was made with regard to the iminium pK_a. The striking difference in this hCRBPII mutant is the observation of two independent pK_as for the iminium. For this reason, extinction coefficients are not presented, as the concentration of the iminium complex cannot be accurately determined.

The pK_a curves are presented for those hCRBPII/ligand complexes which clearly show two pK_a curves (**Figure III-7**). These two pK_as presumably correspond to the *cis* and *trans* isomers of the iminium; Dr. Meisam Nosrati has previously shown that these two species to be present crystallographically when retinal was used as a ligand in CRABPII mutants.[53] It should be noted that the iminium with the lower pK_a is also observed with **ThioFluor-9** and **ThioFluor-6**, but the pK_a is too low to measure in this hCRBPII mutant, due to protein stability in acidic conditions. With this knowledge in hand, all pK_a data should be considered cautiously, as the values are not only dependent on the chromophore electronics, but also the isomer of the iminium, which can lead to drastically different pK_a values. Presumably, the chromophore structure dictates the isomer of the iminium formed. However, at this time, no conclusions can be drawn as to the exact origin; crystal structures may provide further insight.

III.8 Conclusions and proposed projects

While none of the analogues investigated in this chapter will be deemed suitable for use for fluorescent live cell imaging (due to their low quantum yields),

the data presented in this chapter is no less pertinent. The structure-property relationships revealed should establish design parameters for improved dyes. To conclude, the following relationships have been determined:

1. Extended conjugation leads to a bathochromic shift in absorption wavelength. Emission wavelength is linearly correlated to absorption wavelength, when no other environmental factors (i.e. intramolecular charge transfer) are present.
2. A more electrophilic aldehyde (through either less donation from an electron donating group or increased aromaticity of the hetero rings) is correlated with an increased rate of iminium formation.
3. A push-pull system is necessary for fluorescence. When electron withdrawing or weakly electron donating groups are substituted at the para position of the phenyl ring of the chromophore, a small peak in fluorescence is detectable, which equates to a less than 0.01 quantum yield.
4. The iminium pK_a can be manipulated by protein environment (such as the introduction of a counter anion near the iminium), but in general a more conjugated system provides a higher pK_a. This is most apparent by the cyanine-like dye **ThioFluor-2**, which in all proteins tested retained a high pK_a.

As an example of the previous four conclusions, plots of Hammett value versus absorption wavelength, rate of iminium formation and pK_a for select **ThioFluor**

analogues (**ThioFluor**, **ThioFluor-4**, **ThioFluor-6**) coupled with the hCRBPII mutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E are provided in **Figure III-8**.

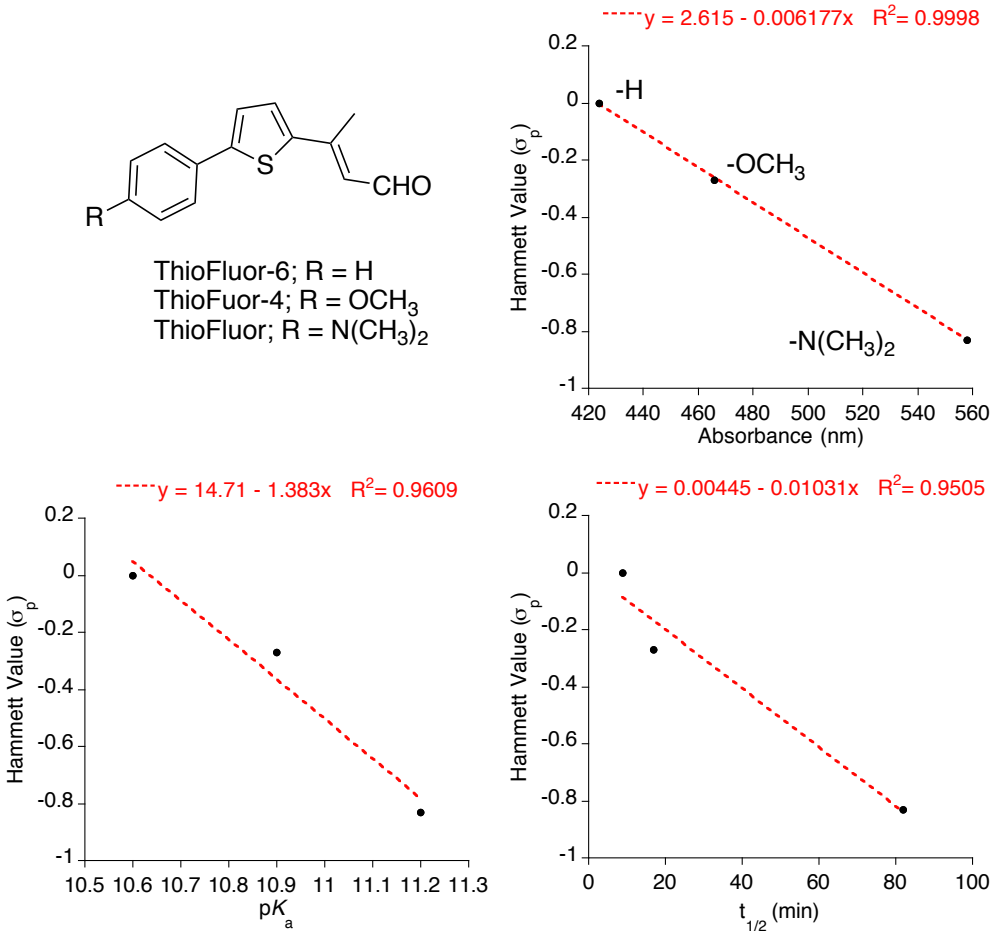


Figure III-8. Plots of Hammett value versus absorption wavelength, rate of iminium formation and pK_a for **ThioFluor**, **ThioFluor-4** and **ThioFluor-6** coupled with the hCRBPII mutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E.

III.8.1 Proposed design of a multi-input fluorogenic probe

The structure-property relationships described in this chapter can be exploited to generate a truly fluorogenic (turn-on fluorescence) system, with fast iminium formation. As was shown in **Chapter II.4**, **ThioFluor** is fluorescent in numerous solvents, yet it can be used in live cell imaging. This is because the

hCRBPII/**ThioFluor** complex is much more red-shifted than the free aldehyde in solution, and can thus be specifically excited as long as iminium formation does not occur with off target lysines. However, it would be ideal to use a probe that is always nonfluorescent. Free aldehyde will then provide no background fluorescence, whether in aqueous media or embedded in the hydrophobic cellular membranes (recall that most solvatochromic fluorophores are fluorescent in hydrophobic environments). Additionally, even if the aldehyde reacts with off target lysines, fluorescence from these nonspecific reactions will not be apparent, because both the imine and iminium are still not fluorescent. This should hold true for fluorophores that contain an electron withdrawing or weakly electron donating group at the para position of the phenyl substituent.

The proposed fluorophore and reaction scheme when bound to hCRBPII is shown in **Figure III-9**. Hammett values are also provided as a reference. Turn on fluorescence would require four inputs including: 1) iminium formation with hCRBPII, 2) nucleophilic substitution reaction with an active site glutamic acid or aspartic acid residue, 3) hydrolysis of the resultant ester, and 4) deprotonation of the hydroxyl group. The concept of multi-input fluorescent probes, which require two or more parameters to turn on fluorescence, is not a new one, though it does appear to be a severely underdeveloped field. Probes have been developed that require chromophore binding to a target, followed by a change in pH, chelation to a metal or enzymatic reaction to produce a fluorescent probe.[54-60]

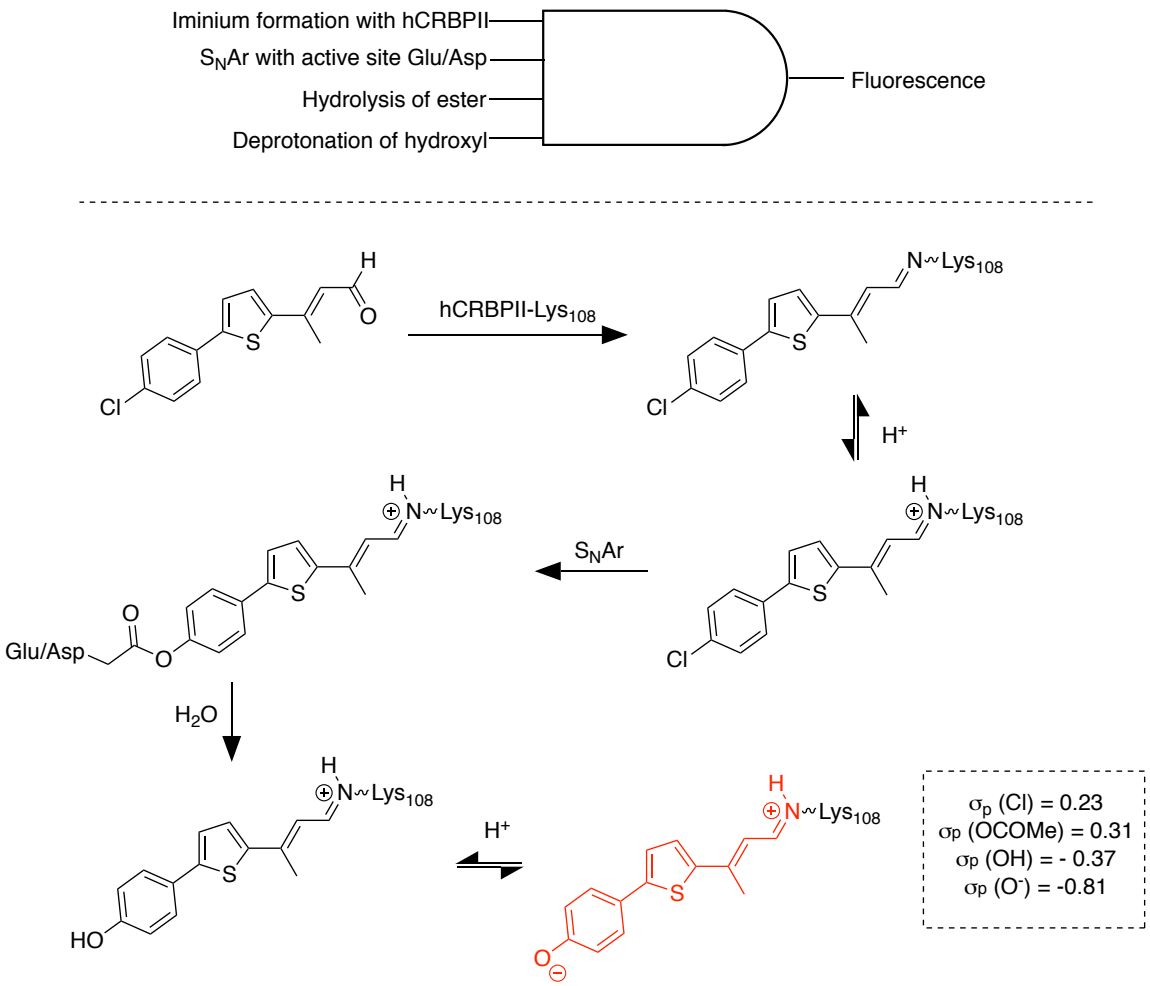


Figure III-9. Proposed fluorophore and reaction scheme for an improved fluorogenic probe.

It is suggested that a *para*-chloro substituent be used in the same scaffold as **ThioFluor**, as the system has now been well developed and the probe shown to be cell permeable. In the aldehydic, imine and iminium states, quantum yield is expected to be less than 0.01. Additionally, iminium formation should proceed swiftly due to the electron withdrawing nature of the chlorine group in conjugation with the aldehyde. Upon iminium formation, the aromatic chlorine is set up for nucleophilic aromatic substitution (a similar mechanism is suggested for the dehalogenation in the enzyme 4-chlorobenzoyl-CoA dehalogenase), being

activated by the electron withdrawing iminium. This is envisioned by attack of a nearby aspartic acid or glutamic acid residue. This intermediate still would not be fluorescent, predicted by its high Hammett value, even if off target reactions occur. Further hydrolysis of the ester should result in the unmasking of a hydroxyl group, which can be deprotonated by a nearby basic residue such as histidine to turn on fluorescence. It is predicted that fluorescent properties will be similar to **ThioFluor**, as the Hammett values of *N,N*-dimethyl and phenoxide are similar.

The proposal outlined is similar to reactions catalyzed by hydrolytic aromatic dehalogenases, including 4-chlorobenzoyl-CoA dehalogenase, chlorothalonil dehalogenase, atrazine dehalogenase,[61] as well as the nonhydrolytic dehalogenase glutathione-*S* transferase.[62] The best studied hydrolytic aromatic dehalogenase is 4-chlorobenzoyl-CoA dehalogenase.[61, 63-71] This enzyme is found in several 4-chlorobenzoate degrading bacteria, and is responsible for catalyzing the conversion of 4-chlorobenzoyl-CoA to 4-hydroxybenzoyl-CoA (**Figure III-10**). The steady state k_{cat} and K_m values measured for wild-type enzyme are 0.60 s^{-1} and 3.7 μM, respectively.[67]

The first step is the ligation of CoA to the substrate (4-chlorobenzoate) via 4-chlorobenzoate CoA ligase. The CoA ligated product is then bound by 4-chlorobenzoyl-CoA dehalogenase (**ES**). The mechanism proceeds by nucleophilic aromatic substitution (S_NAr) through an enzyme-Meisenheimer (**EMc**) intermediate, via Asp145.[68] The collapse of this intermediate, with the expulsion of a chloride ion, leads to the arylated enzyme (**EAr**) complex. This

product is attacked by a water molecule, which is activated by His90, to form the 4-hydroxybenzoyl-CoA product (**EP**).[70]

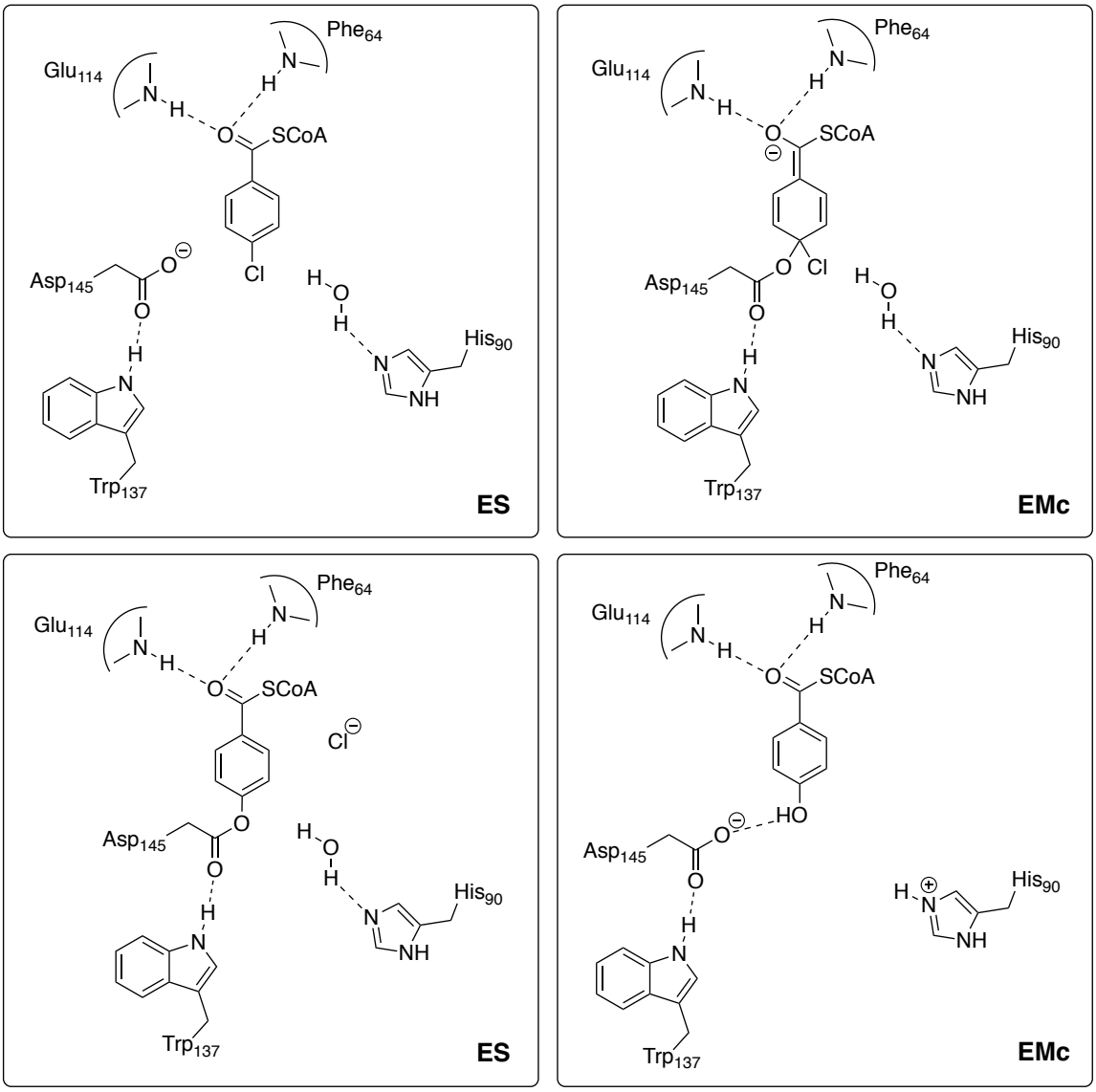


Figure III-10. Proposed chemical steps of the *Pseudomonas sp*. strain 4-chlorobenzoyl-CoA dehalogenase-catalyzed conversion of 4-chlorobenzoyl-CoA to 4-hydroxybenzoyl-CoA.

The crystal structure provides insight into the favorable interactions between the enzyme and product ligand (**Figure III-11**).[72] 4-Chlorobenzoyl-CoA is bound in a primarily hydrophobic environment, except for the catalytic Asp145 and His90 residues. The aromatic groups (Phe64, Phe82, Trp89 and Trp137)

around the benzoyl ring assist in the polarization of the π-electrons in the benzoyl ring during the reaction.[65, 73] It was shown that Trp137 (which forms a hydrogen bond with Asp145) has a significant effect on the formation of the Meisenheimer intermediate; the mutation W137F causes a 20,000-fold reduction in the rate of formation of the Meisenheimer intermediate,[66] but does not affect binding of the substrate;[65] presumably Trp137 positions Asp145 for catalysis.

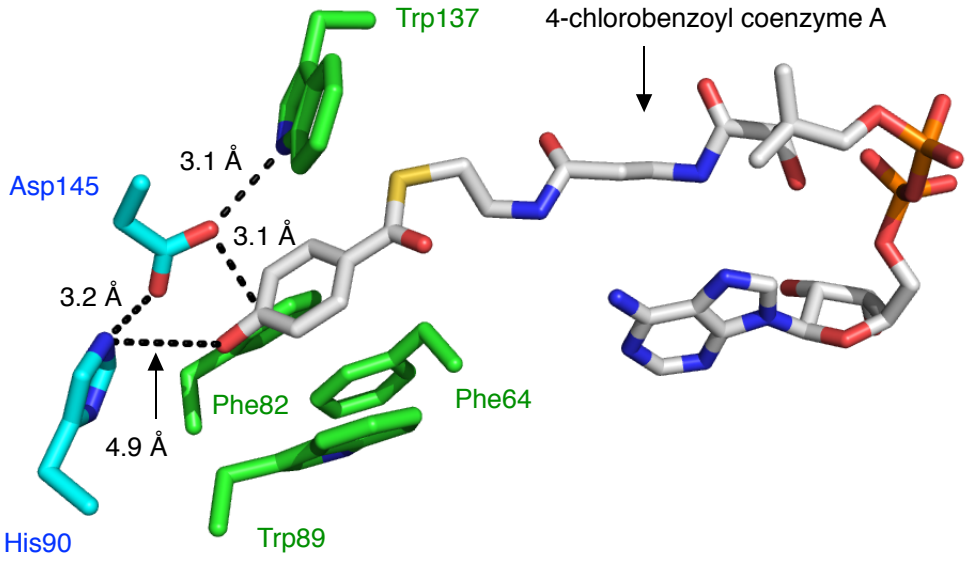


Figure III-11. Crystal structure of 4-chlorobenzoyl coenzyme A dehalogenase from *Pseudonomas sp.* Catalytic residues are shown in cyan and aromatic residues responsible for hydrophobic packing are shown in green. Coordinates obtained from PDB 1NZY.

From the crystal structure, it is also evident that the thioester carbonyl of the ligand is within hydrogen-bonding distance of the backbone amide hydrogen atoms of Gly114 and Phe64 (**Figure III-10**).[72] Catalysis in F64P, F64A and G114A dehalogenase mutants were shown to be less active than the wild-type enzyme, while the binding affinities of the ligands are unchanged.[67]

The proposed system should meet most of the requirements outlined at the beginning of this chapter. Fluorescence should only be turned on when coupled to the target hCRBPII (with multiple inputs). This would allow for high specificity, in which fluorescence should never be observed from free probe or if it does bind to off target lysines. Additionally, upon reaction of the probe with hCRBPII, the fluorophore should be red-shifted (due to the presence of the electron donating phenoxide). Lastly the probe should prove to be cell permeable, as the analogue **ThioFluor** has already shown utility in live cell imaging. It should be noted that this scheme requires three reactions – iminium formation, S_NAr and hydrolysis of the ester; it is hard predict the reaction kinetics under such rigorous requirements.

III.8.2 Development of near infrared and large Stokes shift fluorescent probes

Other noteworthy observations include the LSS observed with **ThioFluor-10** when coupled with select hCRBPII mutants. Preliminary results show a Stokes shift of 186 nm and a quantum yield of 0.06 in the hCRBPII heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E. It is suggested that hCRBPII mutants be optimized for binding to **ThioFluor-10**, in order to improve brightness and be applicable as a LSS fluorescent probe.

ThioFluor-7, or similar analogues, with a beta trifluoromethyl group (or perhaps other electron withdrawing groups) to the aldehyde can be developed into near infrared probes. The current pitfall with **ThioFluor-7** is the severely

depressed pK_a and the lack of fluorescence upon excitation of the PSB. It is suggested that other fluorophore scaffolds that are inherently brighter than **ThioFluor**, be screened for electron withdrawing groups near the aldehyde. It is possible that halogens such as the trifluoromethyl group in **ThioFluor-7** and the fluorine in **ThioFluor-11** are responsible for the fluorescence quenching observed.

APPENDIX

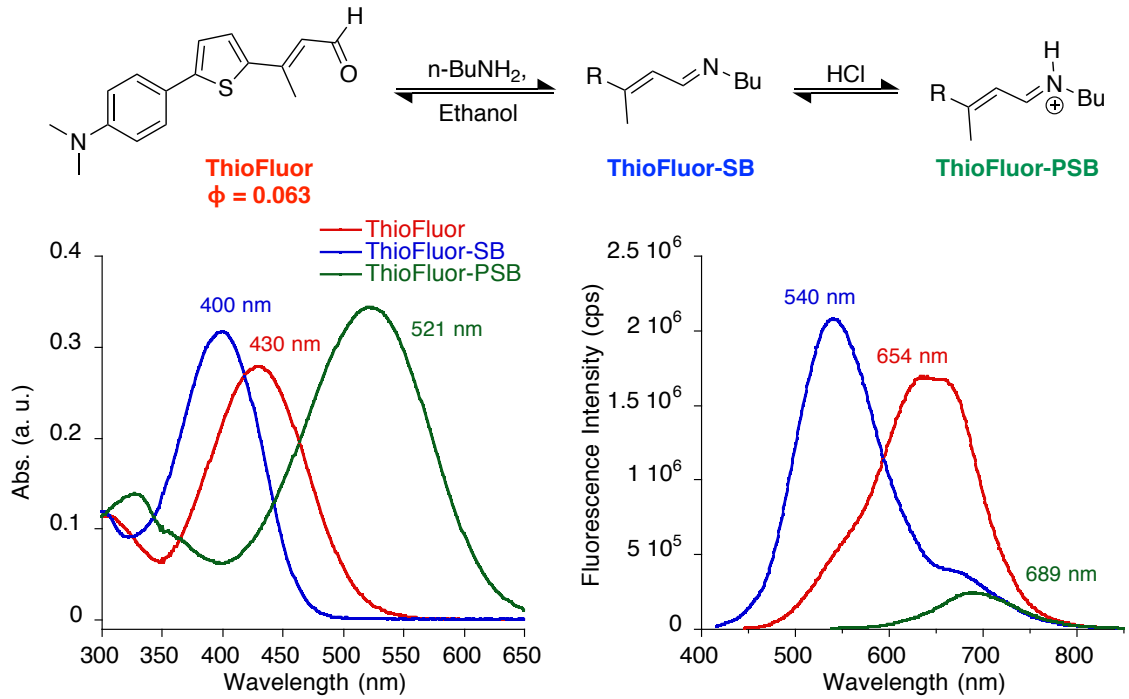


Figure III-12. Formation of **ThioFluor-PSB** with *n*-butyl amine in ethanol.

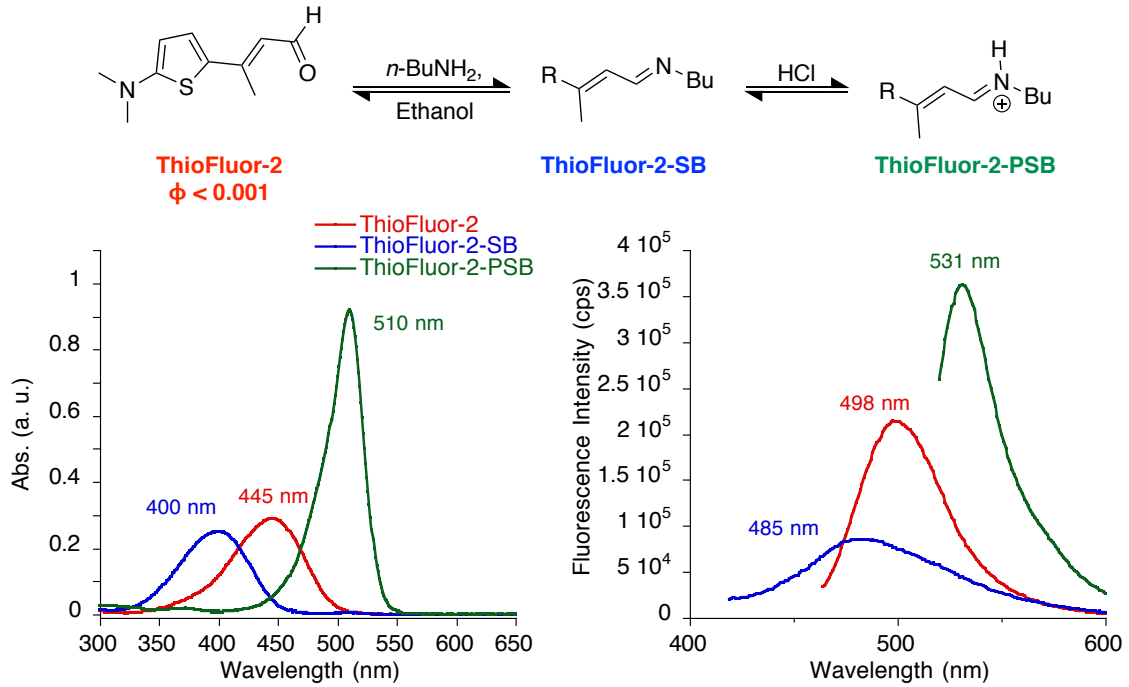


Figure III-13. Formation of **ThioFluor-2-PSB** with *n*-butyl amine in ethanol.

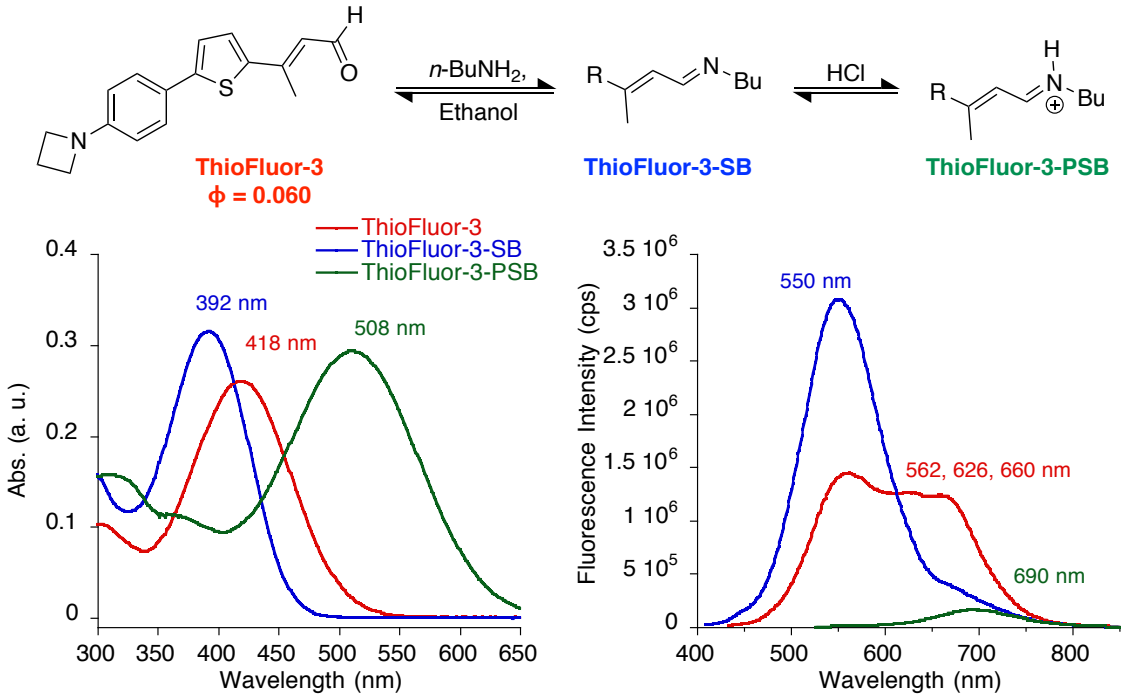


Figure III-14. Formation of **ThioFluor-3-PSB** with *n*-butyl amine in ethanol.

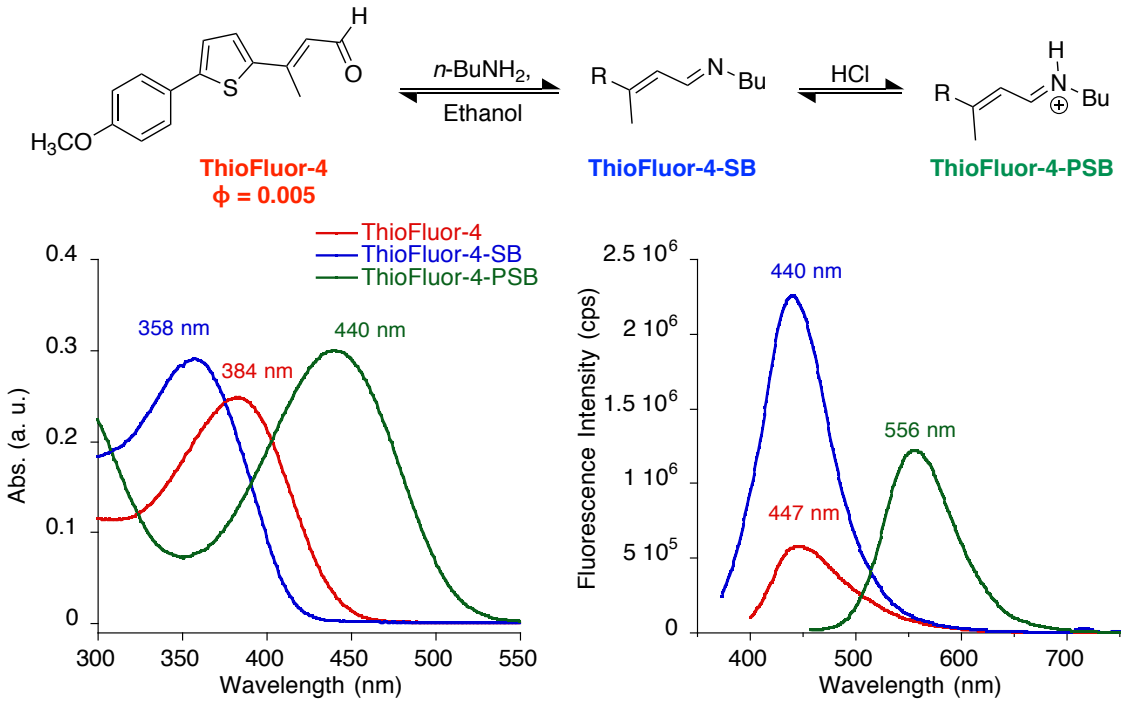


Figure III-15. Formation of **ThioFluor-4-PSB** with *n*-butyl amine in ethanol.

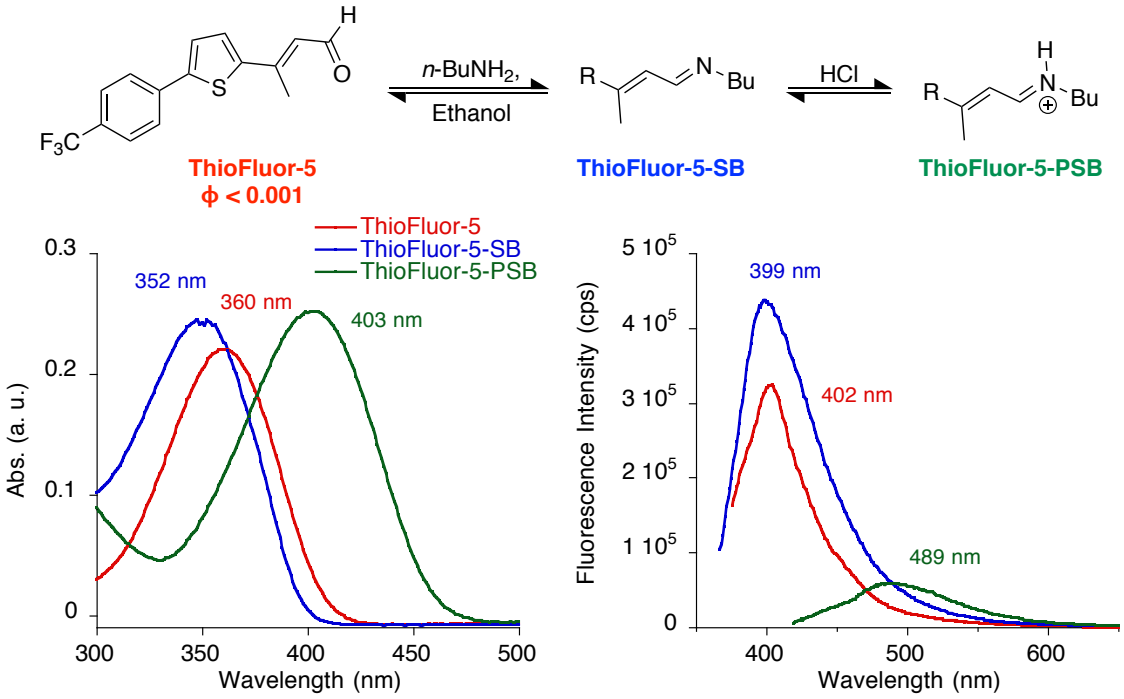


Figure III-16. Formation of **ThioFluor-5-PSB** with *n*-butyl amine in ethanol.

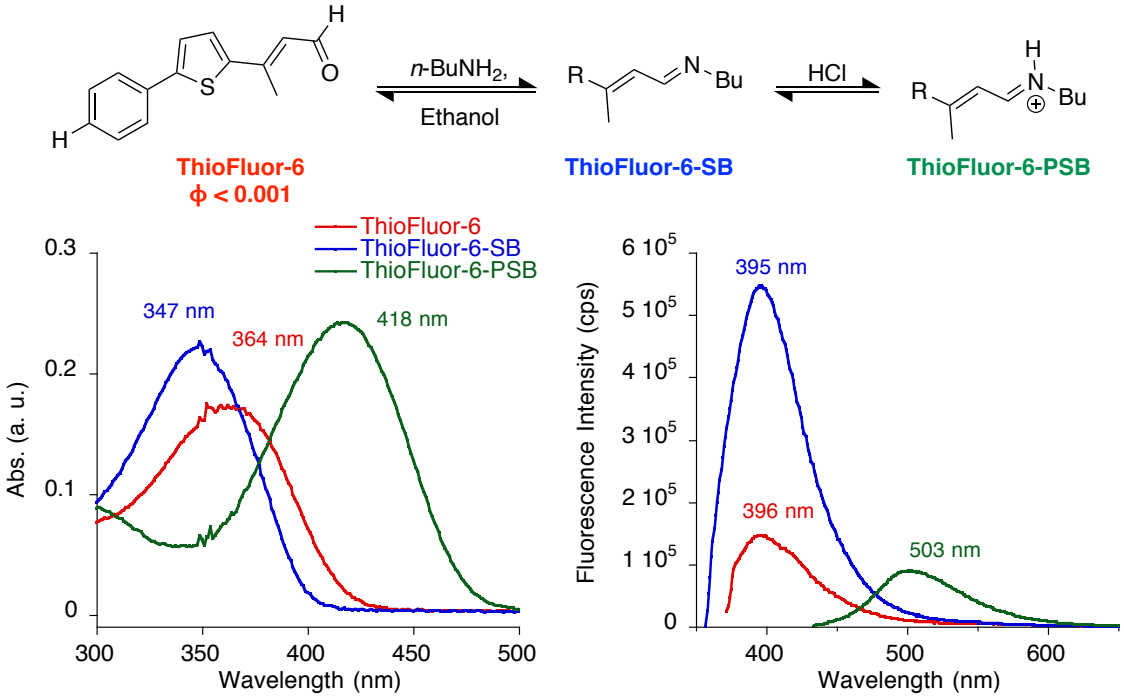


Figure III-17. Formation of **ThioFluor-6-PSB** with *n*-butyl amine in ethanol.

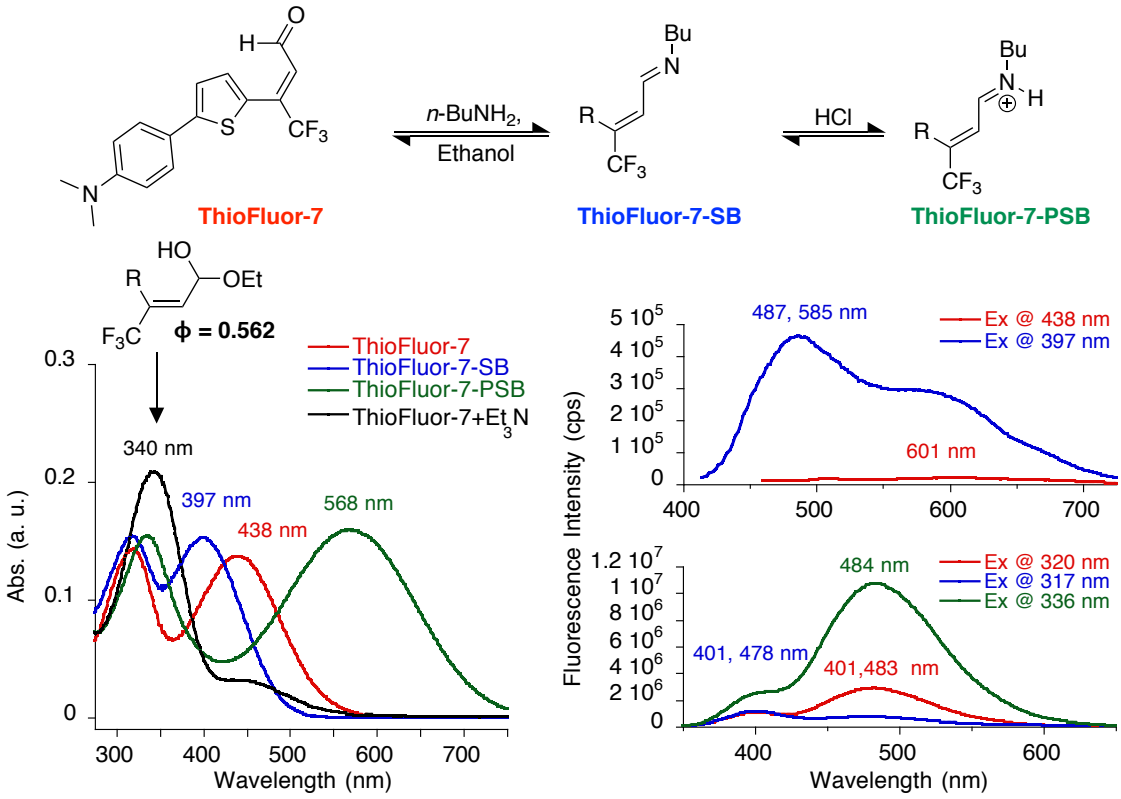


Figure III-18. Formation of **ThioFluor-7-PSB** with *n*-butyl amine in ethanol.

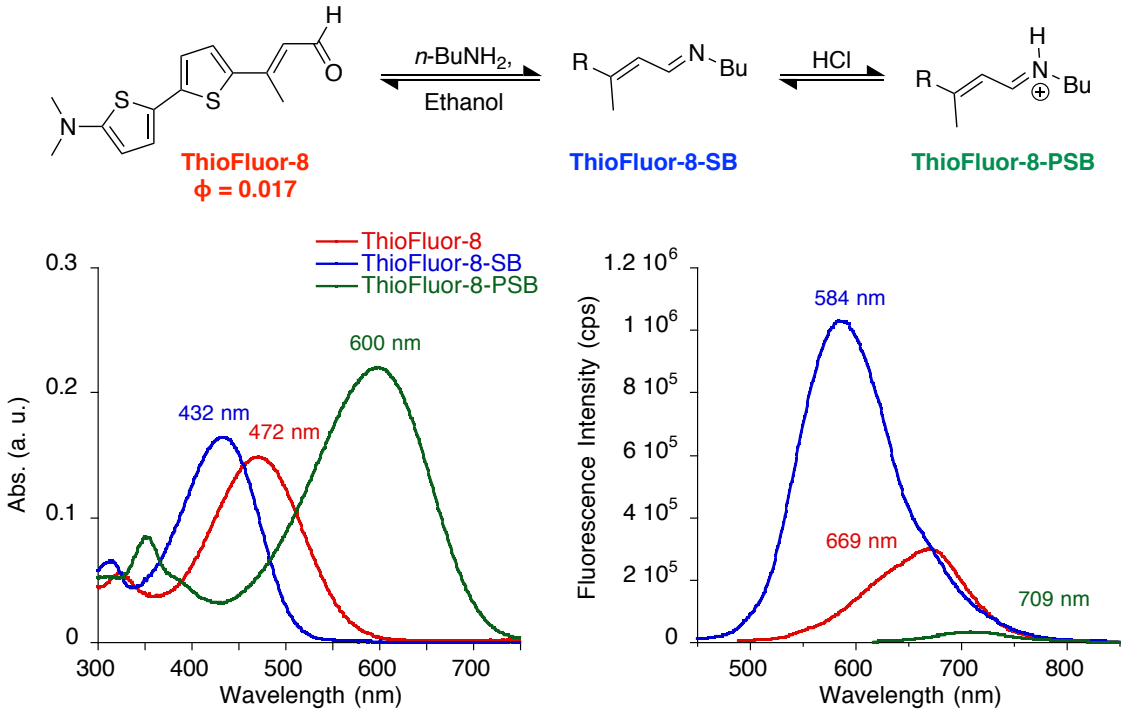


Figure III-19. Formation of **ThioFluor-8-PSB** with *n*-butyl amine in ethanol.

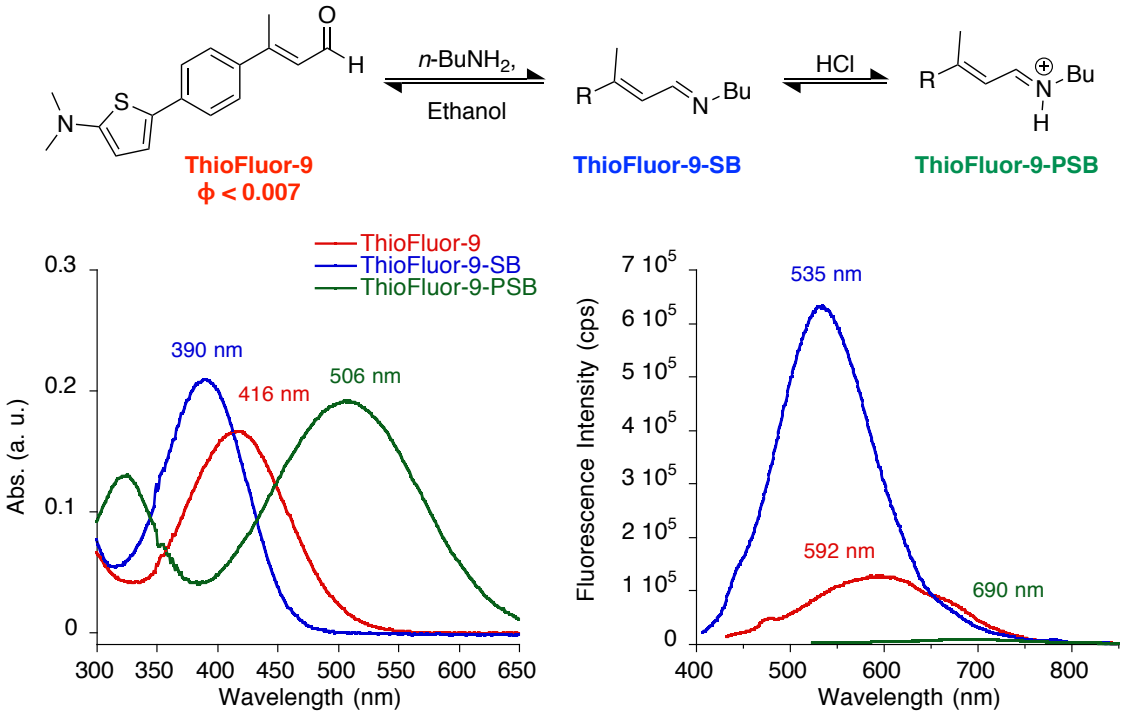


Figure III-20. Formation of **ThioFluor-9-PSB** with *n*-butyl amine in ethanol.

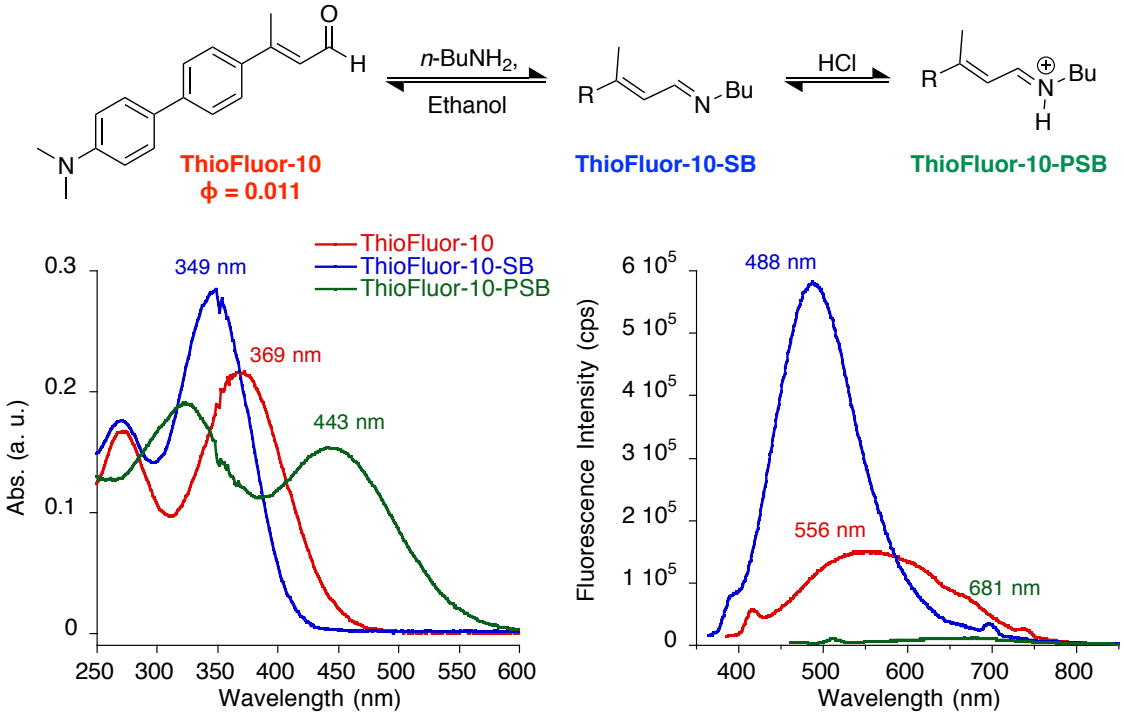


Figure III-21. Formation of **ThioFluor-10-PSB** with *n*-butyl amine in ethanol.

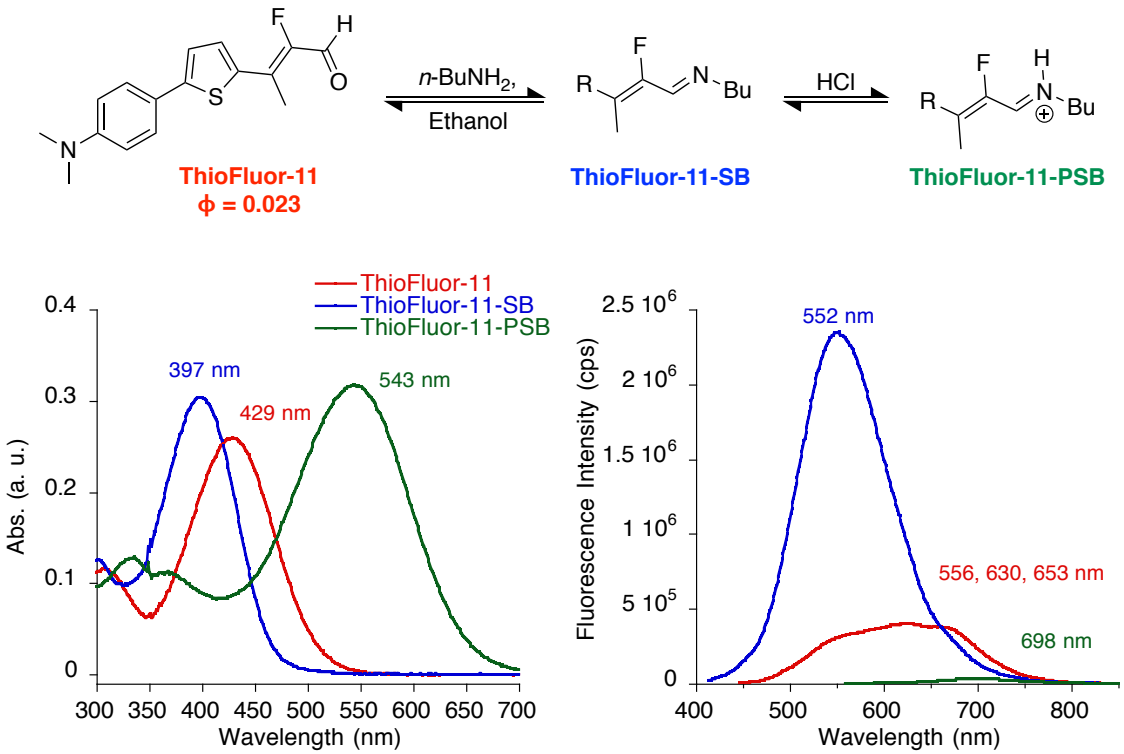


Figure III-22. Formation of **ThioFluor-11-PSB** with *n*-butyl amine in ethanol.

REFERENCES

REFERENCES

1. Hori, Y. & Kikuchi, K. Protein labeling with fluorogenic probes for no-wash live-cell imaging of proteins. *Current Opinion in Chemical Biology* **17**, 644-650 (2013).

2. Baranczak, A., Connelly, S., Liu, Y., Choi, S., Grimster, N.P., Powers, E.T., Wilson, I.A. & Kelly, J.W. Fluorogenic small molecules requiring reaction with a specific protein to create a fluorescent conjugate for biological imaging-what we know and what we need to learn. *Biopolymers* **101**, 484-495 (2014).

3. Bruchez, M.P. Dark dyes-bright complexes: fluorogenic protein labeling. *Current Opinion in Chemical Biology* **27**, 18-23 (2015).

4. Chen, Y.C., Tsao, K. & Keillor, J.W. Fluorogenic protein labelling: a review of photophysical quench mechanisms and principles of fluorogen design. *Canadian Journal of Chemistry* **93**, 389-398 (2015).

5. Andersson, H., Baechi, T., Hoechl, M. & Richter, C. Autofluorescence of living cells. *Journal of Microscopy-Oxford* **191**, 1-7 (1998).

6. Aubin, J.E. Autofluorescence of Viable Cultured Mammalian-Cells. *Journal of Histochemistry & Cytochemistry* **27**, 36-43 (1979).

7. Benson, R.C., Meyer, R.A., Zaruba, M.E. & Mckhann, G.M. Cellular Autofluorescence - Is It Due to Flavins. *Journal of Histochemistry & Cytochemistry* **27**, 44-48 (1979).

8. Muller, M.G., Georgakoudi, I., Zhang, Q.G., Wu, J. & Feld, M.S. Intrinsic fluorescence spectroscopy in turbid media: disentangling effects of scattering and absorption. *Applied Optics* **40**, 4633-4646 (2001).

9. Weissleder, R. A clearer vision for in vivo imaging. *Nature Biotechnology* **19**, 316-317 (2001).

10. Weissleder, R. & Ntziachristos, V. Shedding light onto live molecular targets. *Nature Medicine* **9**, 123-128 (2003).

11. Venkatraman, P., Nguyen, T.T., Sainlos, M., Bilsel, O., Chitta, S., Imperiali, B. & Stern, L.J. Fluorogenic probes for monitoring peptide

binding to class II MHC proteins in living cells. *Nature Chemical Biology* **3**, 222-228 (2007).

12. Fujikawa, Y., Urano, Y., Komatsu, T., Hanaoka, K., Kojima, H., Terai, T., Inoue, H. & Nagano, T. Design and Synthesis of Highly Sensitive Fluorogenic Substrates for Glutathione S-Transferase and Application for Activity Imaging in Living Cells. *Journal of the American Chemical Society* **130**, 14533-14543 (2008).

13. Barnett, E.M., Zhang, X., Maxwell, D., Chang, Q. & Piwnica-Worms, D. Single-cell imaging of retinal ganglion cell apoptosis with a cell-penetrating, activatable peptide probe in an in vivo glaucoma model. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9391-9396 (2009).

14. Hori, Y., Nakaki, K., Sato, M., Mizukami, S. & Kikuchi, K. Development of Protein-Labeling Probes with a Redesigned Fluorogenic Switch Based on Intramolecular Association for No-Wash Live-Cell Imaging. *Angewandte Chemie-International Edition* **51**, 5611-5614 (2012).

15. Hori, Y., Norinobu, T., Sato, M., Arita, K., Shirakawa, M. & Kikuchi, K. Development of Fluorogenic Probes for Quick No-Wash Live-Cell Imaging of Intracellular Proteins. *Journal of the American Chemical Society* **135**, 12360-12365 (2013).

16. Davidson, M.W. & Campbell, R.E. Engineered fluorescent proteins: innovations and applications. *Nature Methods* **6**, 713-717 (2009).

17. Marx, V. Probes: paths to photostability. *Nature Methods* **12**, 187-190 (2015).

18. Cranfill, P.J., Sell, B.R., Baird, M.A., Allen, J.R., Lavagnino, Z., de Gruiter, H.M., Kremers, G.J., Davidson, M.W., Ustione, A. & Piston, D.W. Quantitative assessment of fluorescent proteins. *Nature Methods* **13**, 557-562 (2016).

19. Chudakov, D.M., Lukyanov, S. & Lukyanov, K.A. Fluorescent proteins as a toolkit for in vivo imaging. *Trends in Biotechnology* **23**, 605-613 (2005).

20. Shaner, N.C., Steinbach, P.A. & Tsien, R.Y. A guide to choosing fluorescent proteins. *Nature Methods* **2**, 905-909 (2005).

21. Johnson, I.D. in Handbook Of Biological Confocal Microscopy, Edn. 3rd. (ed. J.B. Pawley) 353-367 (Springer, 2006).

22. Remington, S.J. Fluorescent proteins: maturation, photochemistry and photophysics. *Current Opinion in Structural Biology* **16**, 714-721 (2006).

23. Olenych, S.G., Claxton, N.S., Ottenberg, G.K. & Davidson, M.W. The fluorescent protein color palette. *Current Protocols in Cell Biology* **21.5**, 1-34 (2007).

24. Day, R.N. & Davidson, M.W. The fluorescent protein palette: tools for cellular imaging. *Chemical Society Reviews* **38**, 2887-2921 (2009).

25. Sample, V., Newman, R.H. & Zhang, J. The structure and function of fluorescent proteins. *Chemical Society Reviews* **38**, 2852-2864 (2009).

26. Piatkevich, K.D., Efremenko, E.N., Verkhusha, V.V. & Varfolomeev, D.D. Red fluorescent proteins and their properties. *Russian Chemical Reviews* **79**, 243-258 (2010).

27. Heim, R., Prasher, D.C. & Tsien, R.Y. Wavelength Mutations and Posttranslational Autoxidation of Green Fluorescent Protein. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 12501-12504 (1994).

28. Cubitt, A.B., Heim, R., Adams, S.R., Boyd, A.E., Gross, L.A. & Tsien, R.Y. Understanding, Improving and Using Green Fluorescent Proteins. *Trends in Biochemical Sciences* **20**, 448-455 (1995).

29. Reid, B.G. & Flynn, G.C. Chromophore formation in green fluorescent protein. *Biochemistry* **36**, 6786-6791 (1997).

30. Shcherbakova, D.M., Subach, O.M. & Verkhusha, V.V. Red fluorescent proteins: advanced imaging applications and future design. *Angewandte Chemie-International Edition* **51**, 10724-10738 (2012).

31. Shu, X., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y. & Remington, S.J. Novel chromophores and buried charges control color in mFruits. *Biochemistry* **45**, 9639-9647 (2006).

32. Wang, L., Xie, J.M., Deniz, A.A. & Schultz, P.G. Unnatural amino acid mutagenesis of green fluorescent protein. *Journal of Organic Chemistry* **68**, 174-176 (2003).

33. Wang, F., Niu, W., Guo, J. & Schultz, P.G. Unnatural amino acid mutagenesis of fluorescent proteins. *Angewandte Chemie-International Edition* **51**, 10132-10135 (2012).

34. Young, D.D., Jockush, S., Turro, N.J. & Schultz, P.G. Synthetase polyspecificity as a tool to modulate protein function. *Bioorganic & Medicinal Chemistry Letters* **21**, 7502-7504 (2011).

35. Bonnett, R., Harriman, A. & Kozyrev, A.N. Photophysics of Halogenated Porphyrins. *Journal of the Chemical Society-Faraday Transactions* **88**, 763-769 (1992).

36. Ozhalici-Unal, H., Pow, C.L., Marks, S.A., Jesper, L.D., Silva, G.L., Shank, N.I., Jones, E.W., Burnette, J.M., Berget, P.B. & Armitage, B.A. A rainbow of fluoromodules: A promiscuous scFv protein binds to and activates a diverse set of fluorogenic cyanine dyes. *Journal of the American Chemical Society* **130**, 12620-12621 (2008).

37. Ptaszek, M. Rational Design of Fluorophores for In Vivo Applications. *Progress in Molecular Biology and Translational Science* **113**, 59-108 (2013).

38. Yuan, L., Lin, W.Y., Zheng, K.B., He, L.W. & Huang, W.M. Far-red to near infrared analyte-responsive fluorescent probes based on organic fluorophore platforms for fluorescence imaging. *Chemical Society Reviews* **42**, 622-661 (2013).

39. Guo, Z.Q., Park, S., Yoon, J. & Shin, I. Recent progress in the development of near-infrared fluorescent probes for bioimaging applications. *Chemical Society Reviews* **43**, 16-29 (2014).

40. Juette, M.F., Terry, D.S., Wasserman, M.R., Zhou, Z., Altman, R.B., Zheng, Q. & Blanchard, S.C. The bright future of single-molecule fluorescence imaging. *Current Opinion in Chemical Biology* **20**, 103-111 (2014).

41. Gorka, A.P., Nani, R.R. & Schnermann, M.J. Cyanine polyene reactivity: scope and biomedical applications. *Organic & Biomolecular Chemistry* **13**, 7584-7598 (2015).

42. Hague, A., Faizi, M.S.H., Rather, J.A. & Khan, M.S. Next generation NIR fluorophores for tumor imaging and fluorescence-guided surgery: A review. *Bioorganic & Medicinal Chemistry* **25**, 2017-2034 (2017).

43. 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).

44. Hansch, C., Leo, A. & Taft, R.W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chemical Reviews* **91**, 165-195 (1991).

45. Chen, W., Li, H., Widawsky, J.R., Appayee, C., Venkataraman, L. & Breslow, R. Aromaticity decreases single-molecule junction conductance. *Journal of the American Chemical Society* **136**, 918-920 (2014).

46. Grimm, J.B., English, B.P., Chen, J., Slaughter, J.P., Zhang, Z., Revyakin, A., Patel, R., Macklin, J.J., Normanno, D., Singer, R.H., Lionnet, T. & Lavis, L.D. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nature Methods* **12**, 244-250 (2015).

47. Liu, X., Qiao, Q., Tian, W., Liu, W., Chen, J., Lang, M.J. & Xu, Z. Aziridinyl Fluorophores Demonstrate Bright Fluorescence and Superior Photostability by Effectively Inhibiting Twisted Intramolecular Charge Transfer. *Journal of the American Chemical Society* **138**, 6960-6963 (2016).

48. Sasaki, S., Niko, Y., Klymchenko, A.S. & Konishi, G.I. Design of donor-acceptor geometry for tuning excited-state polarization: fluorescence solvatochromism of push-pull biphenyls with various torsional restrictions on their aryl-aryl bonds. *Tetrahedron* **70**, 7551-7559 (2014).

49. Haberhauer, G., Gleiter, R. & Burkhart, C. Planarized Intramolecular Charge Transfer: A Concept for Fluorophores with both Large Stokes Shifts and High Fluorescence Quantum Yields. *Chemistry-A European Journal* **22**, 971-978 (2016).

50. Zhang, Z.Y., Zhang, G.Q., Wang, J.X., Sun, S.S. & Zhang, Z.Z. The mechanisms of Large Stokes Shift and Fluorescence Quantum Yields in anilino substituted Rhodamine analogue: TICT and PICT. *Computational and Theoretical Chemistry* **1095**, 44-53 (2016).

51. Sasaki, S., Drummen, G.P.C. & Konishi, G. Recent advances in twisted intramolecular charge transfer (TICT) fluorescence and related phenomena in materials chemistry. *Journal of Materials Chemistry C* **4**, 2731-2743 (2016).

52. Berbasova, T., Santos, E.M., Nosrati, M., Vasileiou, C., Geiger, J.H. & Borhan, B. Light-Activated Reversible Imine Isomerization: Towards a Photochromic Protein Switch. *ChemBioChem* **17**, 407-414 (2016).

53. Nosrati, M., Berbasova, T., Vasileiou, C., Borhan, B. & Geiger, J.H. A Photoisomerizing Rhodopsin Mimic Observed at Atomic Resolution. *Journal of the American Chemical Society* **138**, 8802-8808 (2016).

54. Hettie, K.S., Klockow, J.L. & Glass, T.E. Three-input logic gates with potential applications for neuronal imaging. *Journal of the American Chemical Society* **136**, 4877-4880 (2014).

55. Debieu, S. & Romieu, A. Dual enzyme-responsive "turn-on" fluorescence sensing systems based on in situ formation of 7-hydroxy-2-iminocoumarin scaffolds. *Organic & Biomolecular Chemistry* **13**, 10348-10361 (2015).

56. Klockow, J.L., Hettie, K.S., Secor, K.E., Barman, D.N. & Glass, T.E. Tunable Molecular Logic Gates Designed for Imaging Released Neurotransmitters. *Chemistry-A European Journal* **21**, 11446-11451 (2015).

57. Romieu, A. "AND" luminescent "reactive" molecular logic gates: a gateway to multi-analyte bioimaging and biosensing. *Organic & Biomolecular Chemistry* **13**, 1294-1306 (2015).

58. Dai, C.G., Liu, X.L., Du, X.J., Zhang, Y. & Song, Q.H. Two-Input Fluorescent Probe for Thiols and Hydrogen Sulfide Chemosensing and Live Cell Imaging. *ACS Sensors* **1**, 888-895 (2016).

59. Grimm, J.B., Gruber, T.D., Ortiz, G., Brown, T.A. & Lavis, L.D. Virginia Orange: A Versatile, Red-Shifted Fluorescein Scaffold for Single- and Dual-Input Fluorogenic Probes. *Bioconjugate Chemistry* **27**, 474-480 (2016).

60. Debieu, S. & Romieu, A. In situ formation of pyronin dyes for fluorescence protease sensing. *Organic & Biomolecular Chemistry* **15**, 2575-2584 (2017).

61. Arora, P.K. & Bae, H. Role of Dehalogenases in Aerobic Bacterial Degradation of Chlorinated Aromatic Compounds. *Journal of Chemistry* (2014).

62. Graminski, G.F., Zhang, P.H., Sesay, M.A., Ammon, H.L. & Armstrong, R.N. Formation of the 1-(S-Glutathionyl)-2,4,6-Trinitrocyclohexadienate Anion at the Active-Site of Glutathione S-Transferase - Evidence for Enzymic Stabilization of Sigma-Complex Intermediates in Nucleophilic Aromatic-Substitution Reactions. *Biochemistry* **28**, 6252-6258 (1989).

63. Crooks, G.P. & Copley, S.D. A Surprising Effect of Leaving Group on the Nucleophilic Aromatic-Substitution Reaction Catalyzed by 4-Chlorobenzoyl-Coa Dehalogenase. *Journal of the American Chemical Society* **115**, 6422-6423 (1993).

64. Taylor, K.L., Liu, R.Q., Liang, P.H., Price, J., Dunawaymariano, D., Tonge, P.J., Clarkson, J. & Carey, P.R. Evidence for Electrophilic Catalysis in the 4-Chlorobenzoyl-Coa Dehalogenase Reaction - Uv, Raman, and C-13-Nmr Spectral Studies of Dehalogenase Complexes of Benzoyl-Coa Adducts. *Biochemistry* **34**, 13881-13888 (1995).

65. Yang, G.A., Liu, R.Q., Taylor, K.L., Xiang, H., Price, J. & DunawayMariano, D. Identification of active site residues essential to 4-chlorobenzoyl-coenzyme A dehalogenase catalysis by chemical modification and site directed mutagenesis. *Biochemistry* **35**, 10879-10885 (1996).

66. Lau, E.Y. & Bruice, T.C. The active site dynamics of 4-chlorobenzoyl-CoA dehalogenase. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 9527-9532 (2001).

67. Luo, L.S., Taylor, K.L., Xiang, H., Wei, Y.S., Zhang, W.H. & Dunaway-Mariano, D. Role of active site binding interactions in 4-chlorobenzoyl-coenzyme A dehalogenase catalysis. *Biochemistry* **40**, 15684-15692 (2001).

68. Dong, J., Carey, P.R., Wei, Y.S., Luo, L.S., Lu, X.F., Liu, R.Q. & Dunaway-Mariano, D. Raman evidence for Meisenheimer complex formation in the hydrolysis reactions of 4-fluorobenzoyl- and 4-nitrobenzoyl-coenzyme a catalyzed by 4-chlorobenzoyl-coenzyme A dehalogenase. *Biochemistry* **41**, 7453-7463 (2002).

69. de Jong, R.M. & Dijkstra, B.W. Structure and mechanism of bacterial dehalogenases: different ways to cleave a carbon-halogen bond. *Current Opinion in Structural Biology* **13**, 722-730 (2003).

70. Xu, D.G. & Guo, H. Electrostatic influence of active-site waters on the nucleophilic aromatic substitution catalyzed by 4-chlorobenzoyl-CoA dehalogenase. *FEBS Letters* **579**, 4249-4253 (2005).

71. Wu, R., Cao, J., Lu, X.F., Reger, A.S., Gulick, A.M. & Dunaway-Mariano, D. Mechanism of 4-chlorobenzoate: Coenzyme a ligase catalysis. *Biochemistry* **47**, 8026-8039 (2008).

72. Benning, M.M., Taylor, K.L., Liu, R.Q., Yang, G., Xiang, H., Wesenberg, G., DunawayMariano, D. & Holden, H.M. Structure of 4-chlorobenzoyl coenzyme A dehalogenase determined to 1.8 angstrom resolution: An enzyme catalyst generated via adaptive mutation. *Biochemistry* **35**, 8103-8109 (1996).

73. Clarkson, J., Tonge, P.J., Taylor, K.L., DunawayMariano, D. & Carey, P.R. Raman study of the polarizing forces promoting catalysis in 4-chlorobenzoate-CoA dehalogenase. *Biochemistry* **36**, 10192-10199 (1997).

CHAPTER IV: ENGINEERING OF hCRBPII/THIOFLUOR INTO A LARGE STOKES SHIFT FLUORESCENT PROTEIN

Our efforts have been focused on developing protein fusion tags based on the reengineering of human Cellular Retinol Binding Protein II (hCRBPII) to bind aldehydes, which become fluorescent upon iminium formation. During the course of reengineering hCRBPII to bind the solvatochromic fluorophore **ThioFluor**, it was observed that not only the protonated Schiff base (PSB), but also the Schiff base (SB) is fluorescent. However, while the PSB emission wavelength can be regulated over a range of 130 nm (613 nm to 744 nm), the emission of the SB remains nearly consistent at approximately 500 nm. Interestingly though, we observed that when specific mutants were irradiated at the Schiff base (SB), they displayed dual fluorescence. These two bands correspond to the expected SB emission and to a large Stokes shift (LSS) species, emitting in the far-red wavelength regime, more reminiscent of PSB emission.

An increased Stokes shift, and hence better spectral separation, has numerous advantages, namely to diminish self-absorption and light scattering.[1, 2] However, in most cases, high fluorescence efficiency and large Stokes shift are not compatible, particularly in synthetic probes.[3-8] Fluorophores with large Stokes shift can be achieved by a variety of mechanisms: intramolecular charge transfer (ICT),[9-12] solvatochromism,[12-14] excimer/exciplex formation,[15-17] or excited state intramolecular proton transfer.[18-21] Unfortunately, the precise structure-property

relationships are difficult to predict, and in many cases these phenomena lead to fluorescence quenching.

IV.1 Preliminary observation of dual fluorescence in hCRBPII/ThioFluor

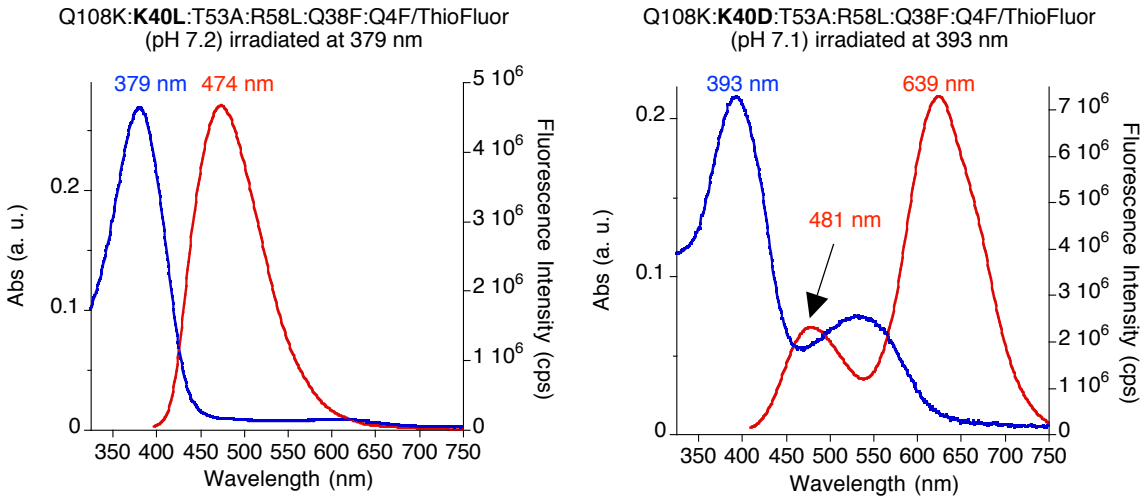


Figure IV-1. Dual fluorescence observed upon irradiation of the SB of Q108K:K40D:T53A:R58L:Q38F:Q4F bound with **ThioFluor**.

Preliminary analysis of sequences comparing the initial hit of the hCRBPII/**ThioFluor** LSS protein versus those displaying normal SB emission, revealed that an acidic residue had been introduced in close proximity to the iminium. As shown in **Figure IV-1**, upon irradiation of the SB at 379 nm of the hCRBPII hexamutant Q108K:K40L:T53A:R58L:Q38F:Q4F bound with **ThioFluor**, an emission maxima of 474 nm is observed. Under identical conditions, irradiation of the SB (at 393 nm) of Q108K:K40D:T53A:R58L:Q38F:Q4F bound with **ThioFluor** results in fluorescence with emission maxima of 481 nm and 639 nm. The first band corresponds to the expected SB emission, while the second band displays an emission maximum more indicative of PSB emission. The quantum yields of these two mutants are quite different; they are 0.03 and 0.34

for Q108K:K40L:T53A:R58L:Q38F:Q4F and Q108K:K40D:T53A:R58L:Q38F:Q4F, respectively.

Next, we sought to investigate the spectroscopic properties of the PSB, to deduce whether its emission maxima is in agreement with that obtained from excitation at the SB. **Figure IV-2** shows a plot of the emissions obtained upon excitation of the SB and PSB. As observed, the red-shifted band of the emission as a result of irradiating the SB is not normally distributed, possessing a red-shifted shoulder. Moreover, from the overlay with normalized emission intensity, it appears that the emission maximum from PSB excitation (649 nm) is equivalent to this red shoulder. The other notable difference is that the quantum yield of SB emission is 0.34, while the quantum yield of the PSB is 0.21.

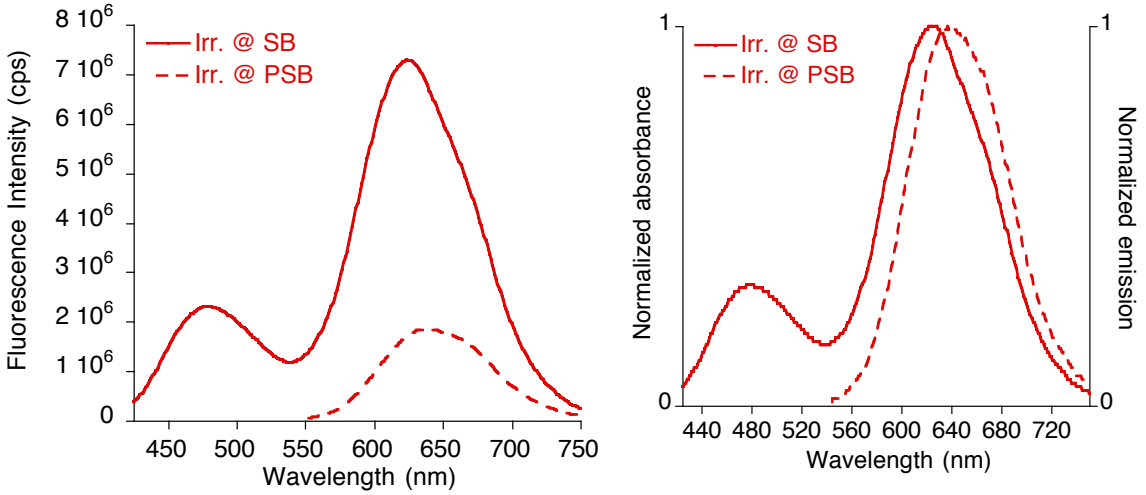


Figure IV-2. Comparison of fluorescence spectra upon irradiation of the SB and PSB of Q108K:K40D:T53A:R58L:Q38F:Q4F/**ThioFluor**.

It should be noted that the emission maxima of both SB and PSB emissions do not change significantly with pH (**Figure IV-3**). This phenomena will be discuss further in depth later when a larger change in emission maxima is

observed upon excitation of the SB and PSB (section **IV.2**). Additionally, the ratio of the two emission bands observed upon excitation of the SB remains nearly constant at pH values of 7.1, 8.0 and 9.0. Presumably, the excited state pK_a is far enough from this pH regime, and thus the protonation state of the imine is unaffected.

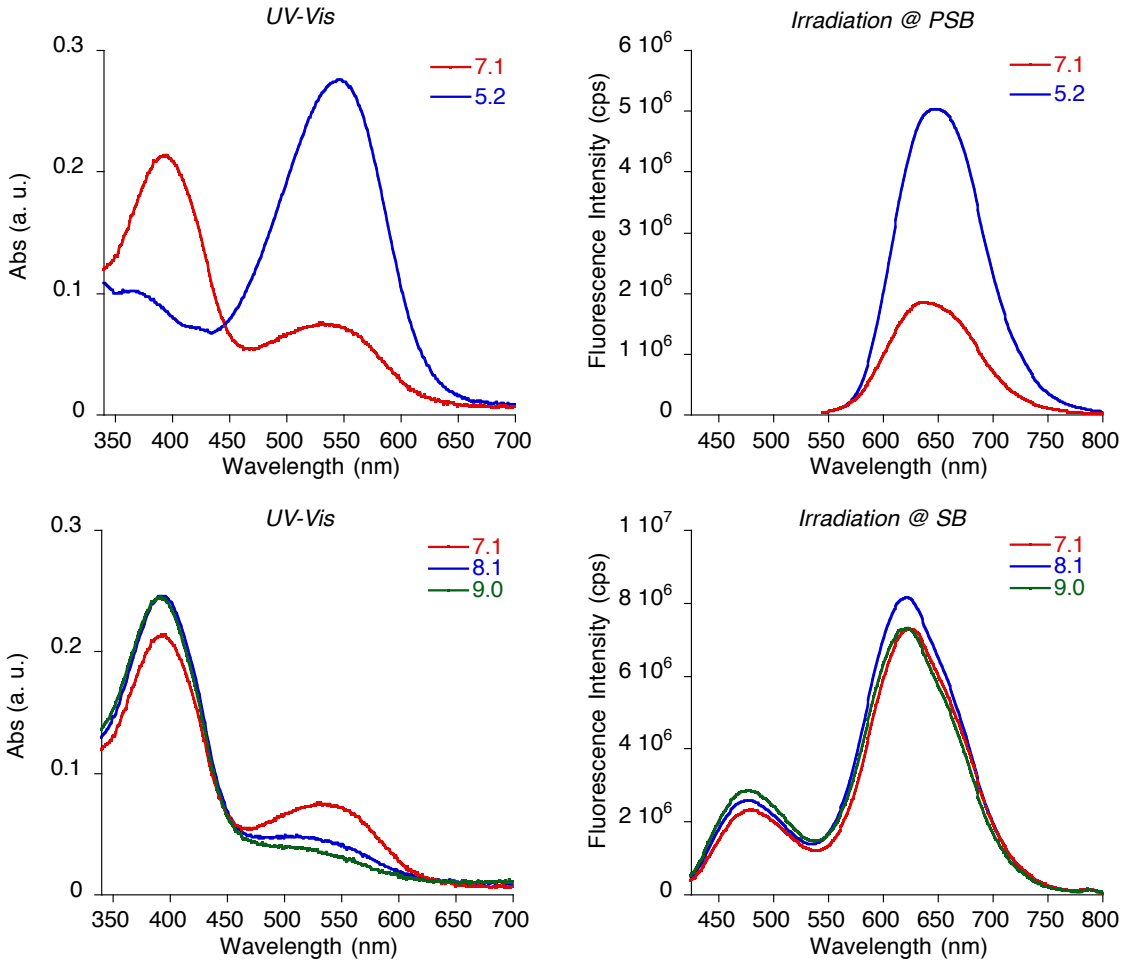


Figure IV-3. UV-Vis and fluorescence spectra obtained as a result of changing pH in Q108K:K40D:T53A:R58L:Q38F:Q4F/**ThioFluor**.

During protein expression at room temperature, both monomer and dimer were obtained due to the presence of K40D (see **Chapter VI** for a full discussion). Dimer was favored during expression, which led to only 35 mol%

monomer (8.3 mg/L monomer and 30.8 mg/L dimer). Interestingly, the dimer of the same hexamutant displays essentially the same fluorescent properties. Upon excitation of the SB at 395 nm, dual fluorescence is observed (**Figure IV-4**), with a quantum yield of 0.35. Similar to the monomer, the red-shifted emission band appears to be composed of two species, of which the major species has an emission maximum of 623 nm. Additionally, excitation of the PSB leads to a more red-shifted emission band at 658 nm with a lower quantum yield (0.17).

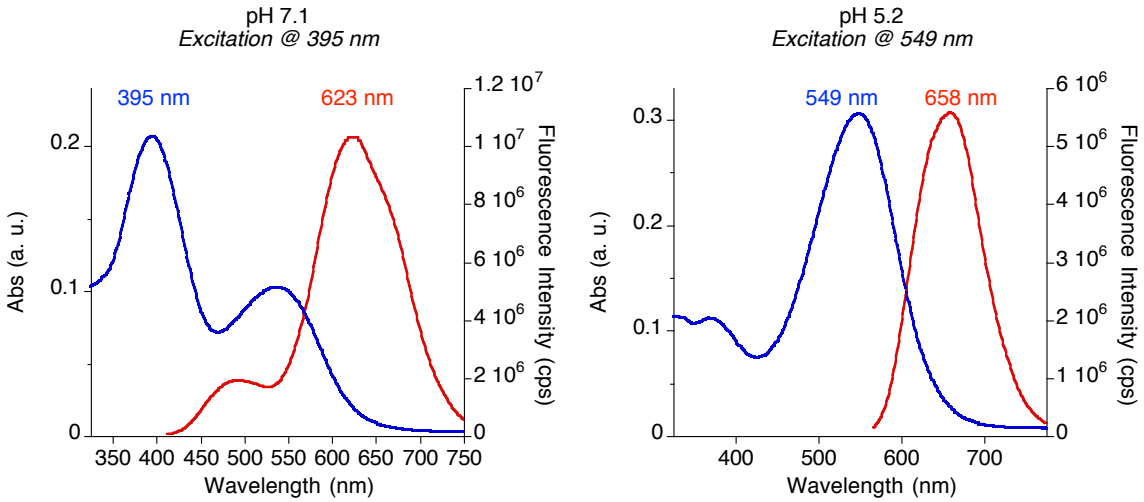


Figure IV-4. UV-Vis (blue) and emission spectra (red) of Q108K:K40D:T53A:R58L:Q38F:Q4F-dimer/**ThioFluor** at pH 7.1 and pH 5.2.

From the aforementioned data, we hypothesized that a proton transfer occurs in the excited state. This would, simply put, produce a PSB in the excited state, which would explain the red-shifted emission band. However, efficiency of the excited state proton transfer (ESPT) dictates that some emission from the SB still occurs. It is further hypothesized that the ESPT emission band is composed of two species, as indicated by the red-shifted shoulder of 623 nm (**Figure IV-4**).

This process is similar to the ESPT observed in wild-type green fluorescent protein (wtGFP), in which the two absorption bands of wtGFP with maxima at 396 and 476 correspond to the neutral and anionic ground-state forms of the chromophore's tyrosyl group (**Figure IV-5**).[22, 23] Excitation at either absorption leads to green fluorescence. Upon excitation of the neutral chromophore, the pK_a is decreased by several units and proton transfer from the hydroxyl group takes place to generate the anionic species that emits green fluorescence.[24-27] Excitation of the anionic species directly leads to green fluorescence. Crystal structures show a proton wire connecting the hydroxyl group through a water molecule and Ser205 to the proton acceptor Glu222.[25, 28]

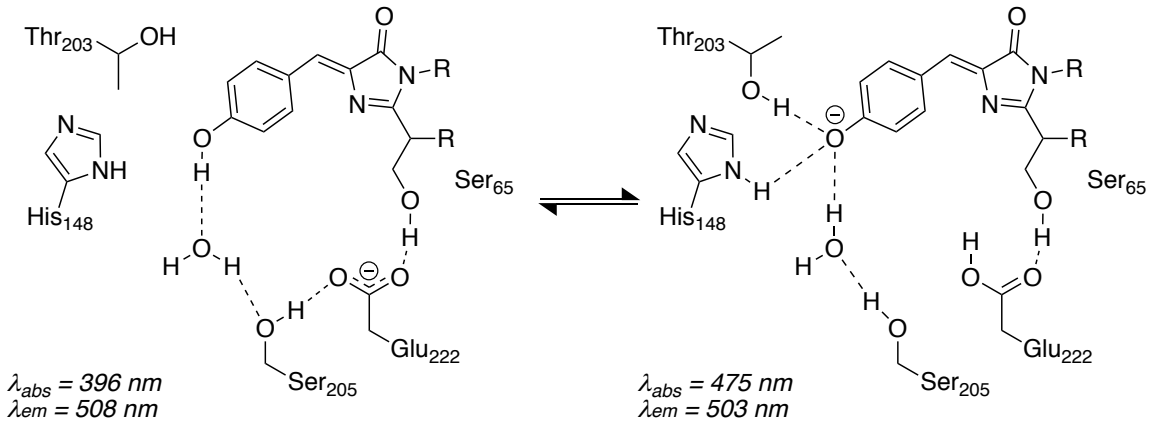


Figure IV-5. Water network responsible for the excited state proton transfer in GFP.

It has been shown that excited state proton transfer (ESPT) is also responsible for the large Stokes shift (LSS) in red FPs, including mKeima,[29-31] LSSmOrange[32-36] and LSSmKates[37-42] (**Table IV-1**). mBeRFP (monomer Blue light-excited RFP) is a monomer that was developed via site-directed mutagenesis from mKate (mKate is majorly monomeric but there is weak dimer

formation).[43] It displays improved brightness and a faster maturation time than the initially developed mKeima and LSS-mKates. Additionally, even though it requires blue light excitation, it displays higher photostability than its predecessors; it takes 60 minutes to photobleach in *E. coli* cells. Nonetheless, each of these variants has proven suitable for imaging in cells.

Table IV-1. Spectroscopic properties of large Stokes shift fluorescent proteins due to ESPT.

Protein	λ_{abs}	λ_{em}	ε ($M^{-1}cm^{-1}$)	Φ	Brightness	Maturation $t_{1/2}$ at 37 °C (h)
LSSmOrange	437	572	52,000	0.45	23,400	2.3
mKeima	440	620	13,400	0.24	3,216	4.4
mBeRFP	446	611	65,000	0.45	29,250	1.0
LSSmKate2	460	605	26,000	0.17	4,420	2.5
LSSmKate1	463	624	31,200	0.08	2,496	1.8

In this class of LSS fluorescent proteins, the chromophore environments exhibit high structural similarity; additionally, the interactions between the chromophores and neighboring residues are similar. The active site residues have been optimized such that the chromophore is neutral in the ground state (stabilized by an acidic residue through hydrogen bonding), and the pK_a is decreased several units in the excited state, such that the acidic residue now acts as a proton acceptor. Notably, Piatkevich and coworkers showed that several conventional red-shifted FPs, including mNeptune, mCherry, mStrawberry, mOrange and mKO, could be developed in LSS fluorescent proteins by modification of neighboring residues to the chromophore's hydroxyl group to Asp or Glu.[40]

As shown in **Figure IV-1**, excitation of the SB of Q108K:K40D:T53A:R58L:Q38F:Q4F/**ThioFluor** does not lead to complete ESPT. We hypothesized that the presence of two proton acceptors would lead to complete excited state protonation of the SB, leading to only the LSS fluorescent species. To this end, we chose to introduce another acidic residue in the vicinity of the iminium. We added V62E to the same hCRBPII mutant Q108K:K40D:T53A:R58L:Q38F:Q4F. Unfortunately, this mutant is expressed solely in the dimeric form at both room temperature and at 16 °C.

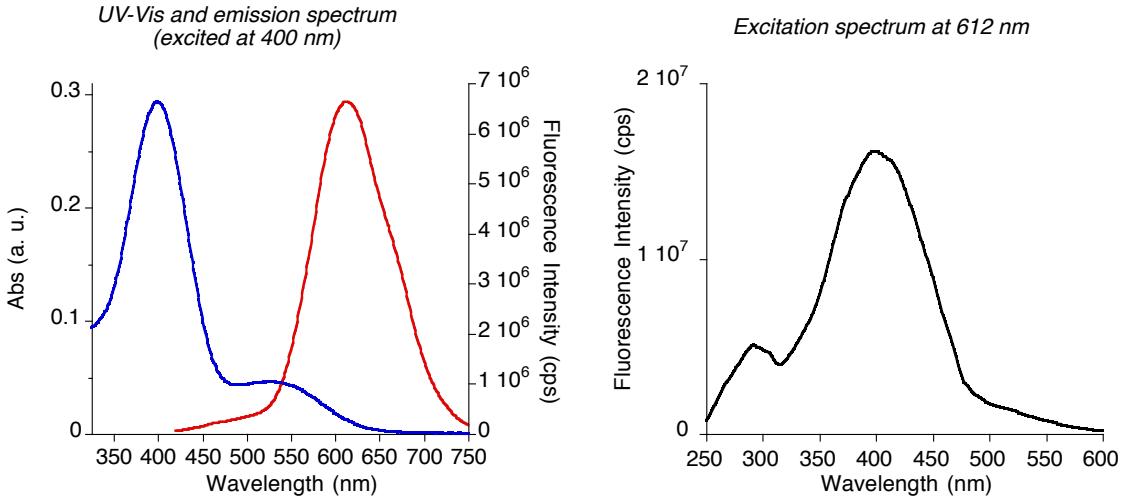


Figure IV-6. UV-Vis, emission and excitation spectra of Q108K:K40D:T53A:R58L:Q38F:Q4F:V62E/**ThioFluor**.

Fortuitously though, upon excitation of the SB at 400 nm, a major emission peak with maxima at 612 nm is observed (**Figure IV-6**); only a small blue-shifted shoulder corresponding to normal SB emission is observed. Additionally, a quantum yield of 0.50 is obtained; this is a substantial increase than that observed in Q108K:K40D:T53A:R58L:Q38F:Q4F (0.34 for the monomer and 0.35 for the dimer). Presumably, the ionic interactions between the acidic residues and

the iminium leads to chromophore rigidification, leading to an increase in quantum yield.

To irrefutably show that the LSS fluorescent species is coming from the SB, an excitation spectrum was collected (**Figure IV-6**). In this experiment, the emission wavelength is held constant while the excitation wavelength is scanned, indicating which wavelengths produced its fluorescence. This clearly shows that the fluorescence at 612 nm originates from the SB and tryptophans inherent to the protein (to be discussed in Section **IV.5**).

While the dimeric species of hCRBPII is not desirable, these results tell us that in our system, changing the environment around the iminium can alter ESPT. Next we chose to investigate both the introduction of different ionizable residues at position 40 and how the location of a glutamic acid from the iminium affects the ESPT pathway and fluorescent properties.

IV.2 Exploration of amino acid substitution at residue 40 on fluorescence

The two protein sequences (Q108K:K40L:T53A:R58L:Q38F:Q4F and Q108K:K40D:T53A:R58L:Q38F:Q4F) just discussed only vary by the substitution at residue 40. The position of residue 40 is highlighted in the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor** (**Figure IV-7**). Presumably, the introduction of aspartic acid is responsible for the ESPT observed (L117 has been mutated to aspartic acid via mutagenesis in Pymol in **Figure IV-7**). To this end, we introduced other ionizable amino acids at position 40 to see if they also produce the red-shifted emission. We wanted to limit the number of variables that

are changed at one time, so we chose to explore modifications retaining Q108K:K40L:T53A:R58L:Q38F:Q4F as the parent hCRBPII mutant. This mutant was chosen as introduction of an acidic residue near the iminium was already shown to lead to ESPT. Additionally, when bound to **ThioFluor**, it possesses a low pK_a (5.1), such that almost no iminium is present at physiological pH. Fortuitously, **ThioFluor** also displays an acceptable half-life for iminium formation (16.0 minutes; binding of 20 μM protein with 10 μM **ThioFluor** was measured via fluorescence at 23 °C and plotted to pseudo-first order rate kinetics), and *in vitro* protein expression yields only monomeric hCRBPII (10.7 mg/L).

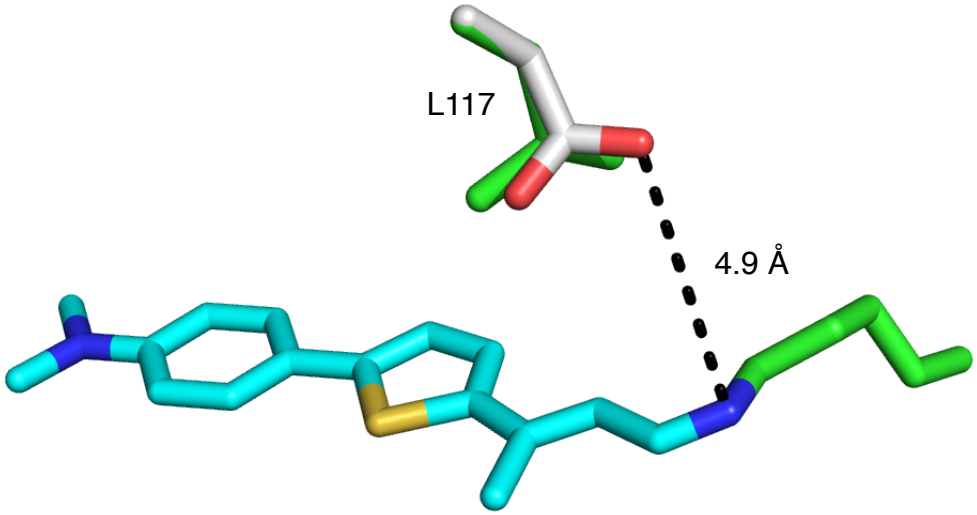


Figure IV-7. Location of residue 40 in the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor**. L117D (shown in gray) was modeled in Pymol.

Numerous amino acids, including cysteine, glutamic acid, histidine, glutamine, arginine, threonine, tyrosine and the wild-type lysine, were introduced at position 40 of Q108K:K40L:T53A:R58L:Q38F:Q4F. Interestingly, the iminium pK_a of these mutants coupled with **ThioFluor** varied over a range of about two pH units (**Table IV-2**). The most perplexing result is that introduction of the basic

residues histidine and lysine at position 40 leads to higher pK_as (7.8 and 7.4, respectively) than the introduction of the acidic residues aspartic acid and glutamic acid (6.4 and 5.2, respectively). This is counterintuitive, as it would be expected that a carboxylate would more easily stabilize a positive charge on the nitrogen atom of the iminium, while a basic residue would destabilize this state. Presumably, the imine exists as a different isomer in each of these cases, such that the dissimilar environments around the iminium dictate the pK_a. Unfortunately, crystal structures have not been determined for any of these protein/**ThioFluor** complexes.

Fluorescent properties upon irradiation of the SB were analyzed (all data shown is for monomer) at pH 7.0 – 7.2. We defined Φ_{ESPT} as the fraction of photons absorbed at the absorption maximum that leads to photochemical protonation of the chromophore's imine. Φ_{ESPT} was derived by integrating the area of the emission spectra of both the deprotonated and protonated forms obtained upon excitation of the SB; Φ_{ESPT} was then obtained as a ratio the red-shifted emission band area over total fluorescence area.

Both cysteine and threonine do not lead to ESPT upon excitation of the SB at physiological pH, as was similarly observed for the introduction of nonpolar leucine at residue 40. However, upon acidification of the solution, a very minor amount of the LSS species can be observed. Basic residues histidine, lysine and arginine all results in partial ESPT. Presumably, they are capable of directly protonating the imine in the excited state, or they polarize a water molecule,

which can act as a proton donor in the excited state. However, quantum yields were much lower than that observed with K40D (0.34). Similarly, glutamine and tyrosine showed dual fluorescence upon excitation of the SB, with suppressed quantum yields.

Table IV-2. Spectroscopic properties of the SB as a result of substitution at residue 40 in the hCRBPII protein Q108K:K40L:T53A:R58L:Q38F:Q4F.

Residue 40	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	Φ_{ESPT}	Φ	pK_a	$t_{1/2}$ (min)[a]
K40C	383	480	-	< 0.02	0.07	6.5	12.8
K40D	393	481	639	0.80	0.34	6.4	13.8
K40E	397	-	605	> 0.98	0.51	5.2	3.6
K40H	393	477	642	0.40	0.09	7.8	15.4
K40K	389	491	661	0.37	0.08	7.4	28.8
K40L	379	474	-	< 0.02	0.03	5.1	16.0
K40Q	388	488	663	0.07	0.11	7.5	40.7
K40R	392	492	653	0.69	-	-	nd
K40T	386	487	-	< 0.02	0.13	6.0	nd
K40Y	393	495	674	0.09	0.04	6.8	12.9

[a]Half-lives were computed from the rate constant that was obtained from plotting data to pseudo-first order kinetics, to ensure the numbers are comparable (note, they should only be compared in relative terms). Second order rates are not given, as concentration is difficult to determine when both SB and PSB are formed in the course of binding. Mutants with K40E and K40L were measured via fluorescence monitoring SB emission while others were measured via UV-Vis following PSB formation.

Fortuitously, the introduction of glutamic acid at position 40 resulted in complete ESPT upon excitation of Q108K:K40E:T53A:R58L:Q38F:Q4F/ **ThioFluor** at its SB absorption maxima of 397 nm. A single emission band with maximum of 605 nm was obtained (**Figure IV-8a**), which is more blue-shifted than the other mutants. An excitation spectrum, collected at 605 nm showed that

the fluorescence originates from the SB and tryptophans in hCRBPII (**Figure IV-8b**). The emission spectrum obtained upon excitation of tryptophan of the hCRBPII protein at 280 nm suggests the existence of an energy transfer process from tryptophan to the bound **ThioFluor**-SB (**Figure IV-8c**). The emission spectrum obtained upon excitation of 280 nm overlays with the emission spectrum obtained upon excitation of the SB at 397 nm, indicating that fluorescence results form the same excited state species.

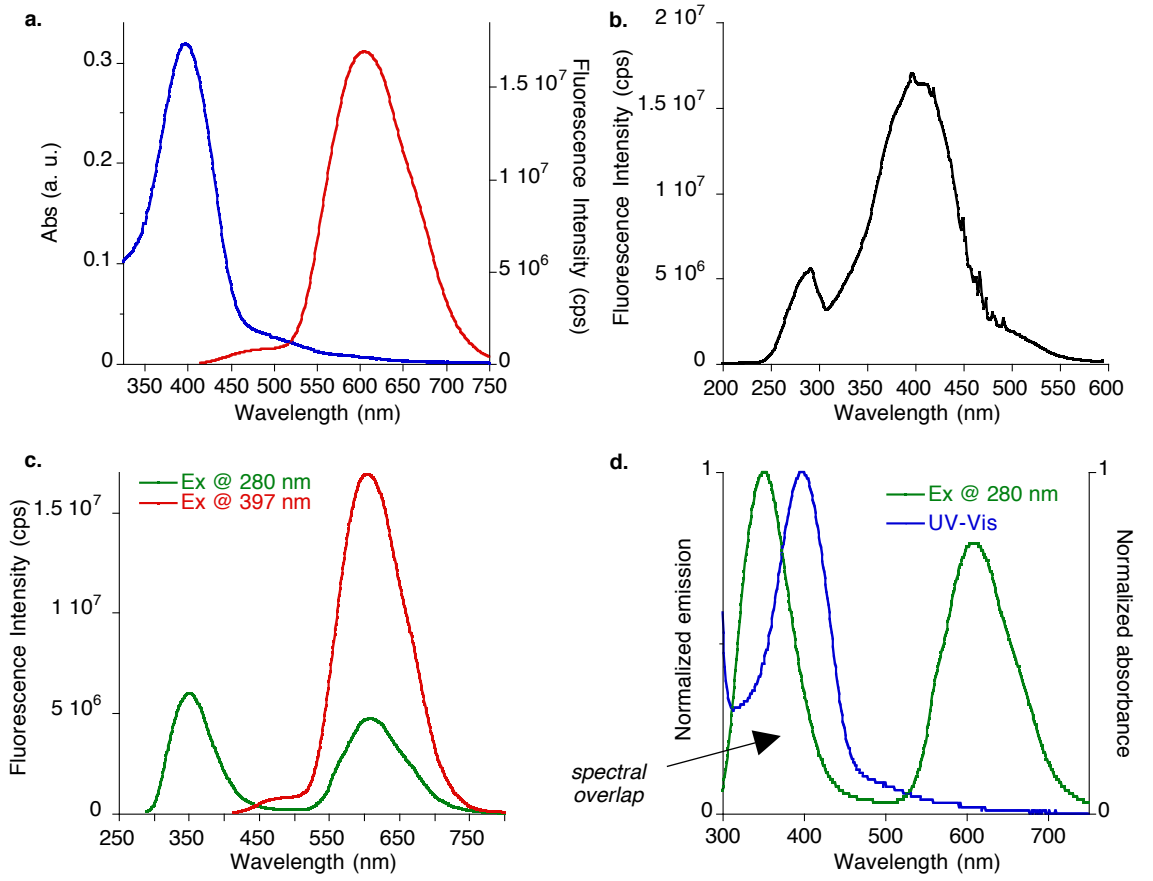


Figure IV-8. Characterization of SB fluorescent properties of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor**. **a**. UV-Vis and emission spectra after 397 nm excitation. **b**. Excitation spectrum at 605 nm. **c**. Emission spectra upon 280 nm excitation. **d**. Overlay of 280 nm emission spectrum and UV-Vis of the hCRBPII-SB showing spectral overlap. All data was collected at pH 7.2.

Figure IV-8d shows the spectral overlap between the normalized protein emission and the **ThioFluor**-SB absorption spectra. This spectral overlap indicates the possibility of energy transfer from the excited tryptophan(s) in hCRBPII to the bound SB, which then undergoes ESPT.

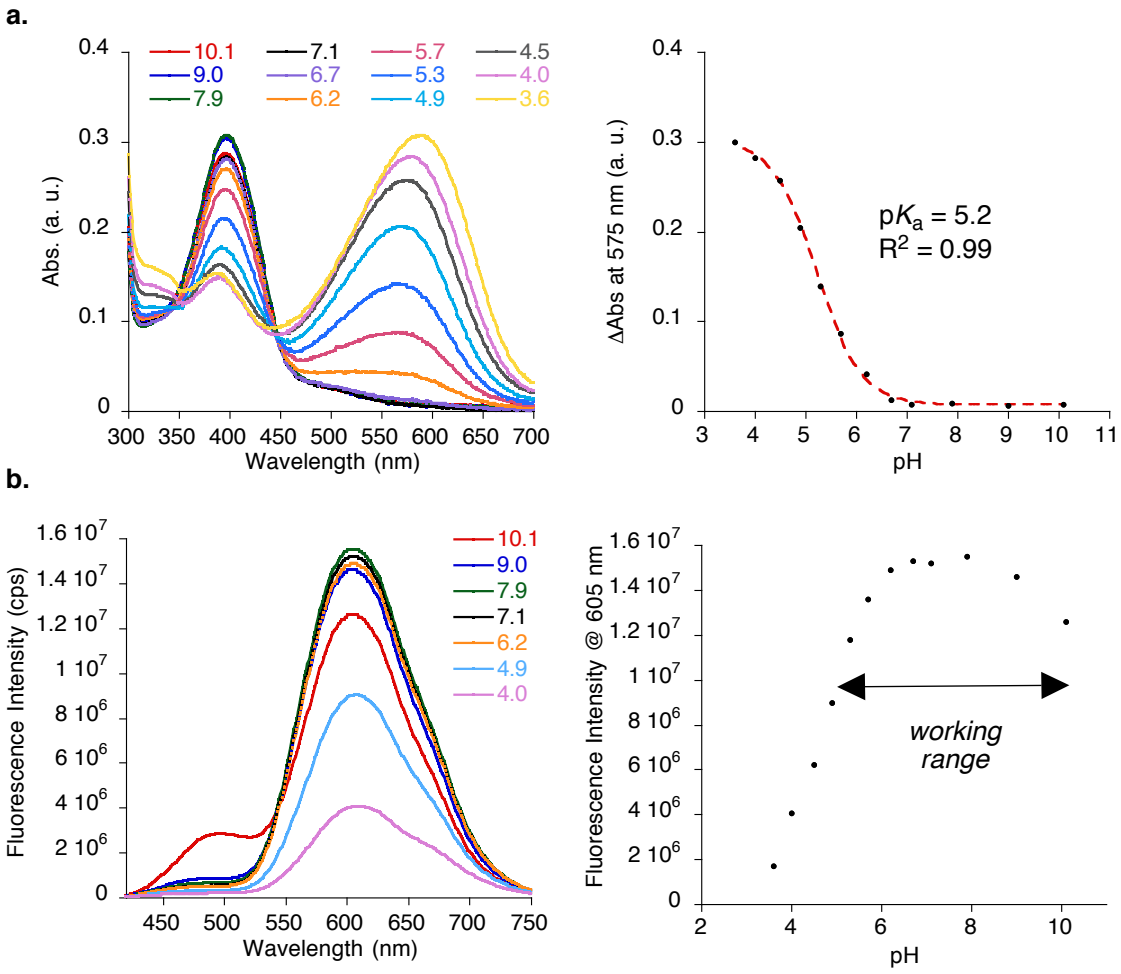


Figure IV-9. **a**. pK_a titration of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor**. **b**. SB emission as a function of pH.

To our delight, a high quantum yield of 0.51 and low pK_a of 5.2 (**Figure IV-9a**) were observed. This is desirable since at physiological pH the system will exist mostly as the SB form, ready to undergo ESPT. This should lead to background free fluorescence live cell imaging, as fluorescence will originate

from the ESPT emitting state, which is not expected to occur with nonspecific entities. Additionally, upon excitation of the SB, ESPT is observed over a wide range of pH values. Select emission curves obtained upon SB excitation at varying pH values are shown in **Figure IV-9b**. Typical SB emission only becomes apparent at pH 10.1.

Notably, the reaction between Q108K:K40E:T53A:R58L:Q38F:Q4F and **ThioFluor** is complete within 10 minutes, with a half-life of 1.7 minutes at 23 °C (measured via fluorescence with 20 μM protein with 10 μM **ThioFluor** and plotted to second order rate kinetics; concentration should be accurate as the pK_a is low enough that only SB is formed at pH 7.2) (**Figure IV-10**). This fast reaction time should allow for almost instantaneous visualization of hCRBPII/**ThioFluor** after dye addition.

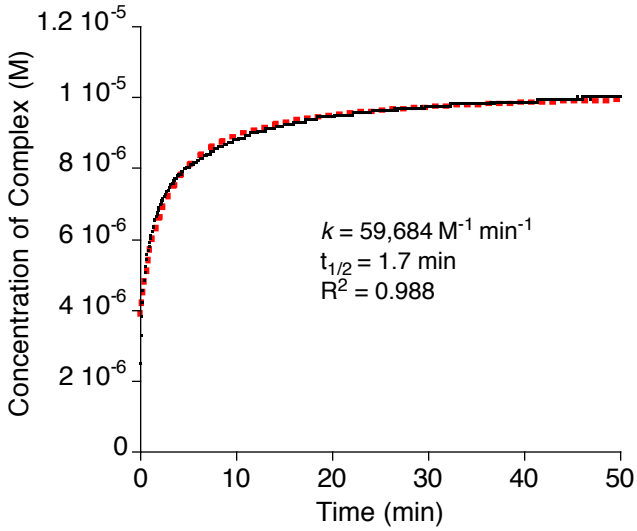


Figure IV-10. Plot of fluorescence intensity at 605 nm versus time for binding of **ThioFluor** to Q108K:K40E:T53A:R58L:Q38F:Q4F at 23 °C fit to second order rate kinetics.

Fluorescent properties obtained upon excitation of the PSB were also collected (**Table IV-3**). In most cases, the emission maxima upon excitation of the PSB was within a few nanometers of the emission maxima of the LSS fluorescent species observed upon irradiation of the SB. The introduction of K40E, yielding Q108K:K40E:T53A:R58L:Q38F:Q4F, was the exception (and to a lesser extent K40D, shown in **Figure IV-2**).

Table IV-3. Spectroscopic properties of the PSB as a result of substitution at residue 40 in the hCRBPII protein Q108K:K40L:T53A:R58L:Q38F:Q4F.

Residue 40	λ_{abs}	λ_{em}	Φ
K40C	558	663	0.17
K40D	546	649	0.21
K40E	575	671	0.15
K40H	544	648	0.23
K40K	557	664	0.20
K40L	613	690	0.10
K40Q	558	664	0.17
K40R	-	657	-
K40T	550	666	0.12
K40Y	595	677	0.16

Figure VI-11a shows an overlay of the normalized emission spectra obtained upon excitation of the SB and PSB (at pH 7.1 and 4.0, respectively; the UV-Vis is shown in **Figure IV-9a**) of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor**. It is quite apparent that the emission maxima are dramatically different – 605 nm and 671 nm for the SB and PSB, respectively. Emission spectra were collected at different pHs. Similarly to that observed for the SB, the PSB emission maxima does not change

significantly with pH (**Figure IV-11b**). This would suggest that hydrogen bonding does not affect the wavelength of emission (as the carboxylate of K40E would be expected to exist at basic pH and the protonated form at lower pH). It would stand to reason that the observed difference between the red-shifted SB emission and PSB emission is not due to charged state interactions between the imine and surrounding residues. It is hypothesized that the difference is due to the presence of different isomers at the imine (*cis* versus *trans*).

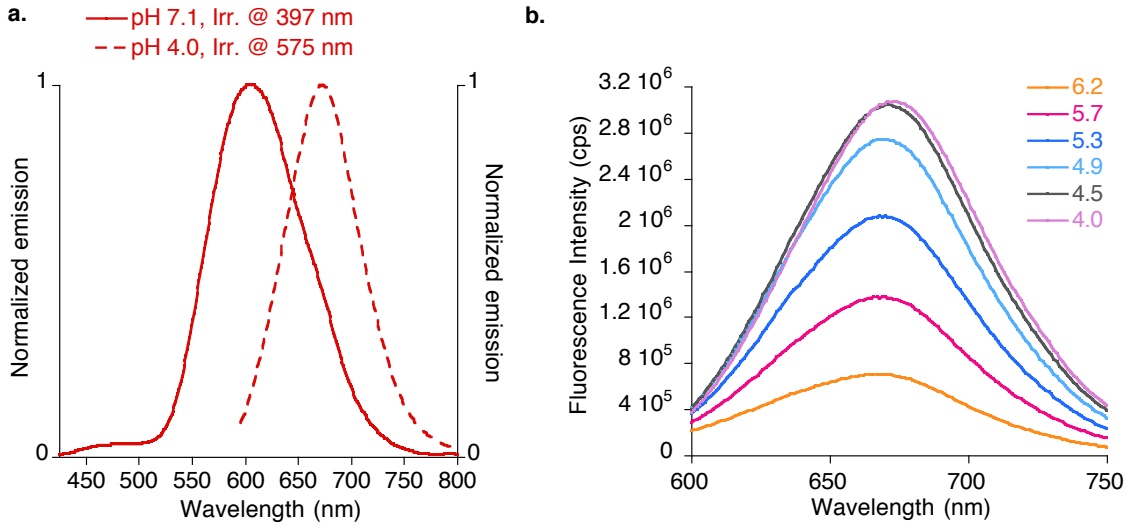


Figure IV-11. **a**. Comparison of fluorescence spectra upon irradiation of the SB and PSB of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor**. **b**. PSB emission spectra as a function of pH.

Unfortunately, attempts to obtain the crystal structure of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor** were not fruitful. However, Mr. Alireza Ghanbarpour was able to obtain the crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor** (it should be noted that the crystal was frozen after approximately 2 months at room temperature), in which tryptophans are introduced at positions 58 and 19 in an attempt to rigidify

the chromophore. Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor** is crystallized with four molecules in the asymmetric unit. Three of the chains show good electron density for the chromophore. Interestingly, two distinct orientations of the active site residues Q38F, R58W and F57 are apparent, while K40E conformation and chromophore trajectory remain the same (**Figure IV-12**). An overlay of Chains A and B is shown in **Figure IV-12**. It should be noted that both *cis* and *trans* isomers of the imine are observed, indicating that both isomers are present in the thermodynamic ground state. Interestingly, in the *trans* isomer, the imine nitrogen atom is pointing towards K40E. However, in the *cis* isomer, the nitrogen is pointing toward tryptophan 106, stabilized through a π-cation interaction (these crystals were obtained at pH 4.5).

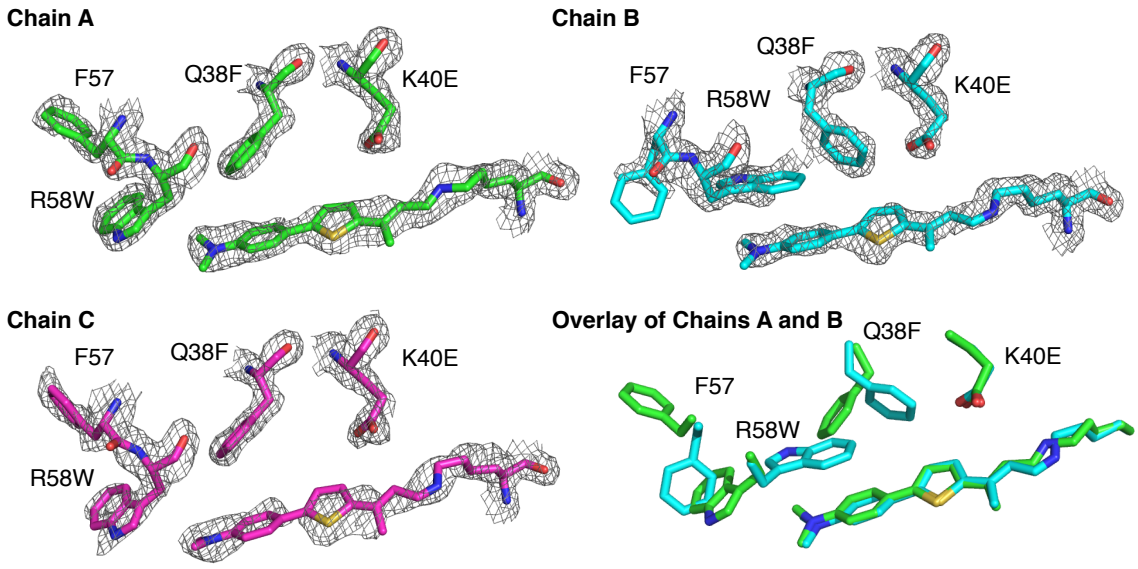


Figure IV-12. Chains A, B and C of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor**. Density is shown at 1σ in Pymol.

It is interesting that the K40E rotamer is constant, while the surrounding aromatic residues exist in two conformations. The distances from K40E to the

imine nitrogen atom are 4.7 Å, 5.2 Å and 5.2 Å for chains A, B and C, respectively. In Chain A of the Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor** crystal structure, an extensive hydrogen-bonding network is observed (**Figure IV-13**); however, it should be noted that the water molecule between K40E and T51 is not seen in the other two chains. Perhaps, this is the reason that K40E does not change orientation even when Q38F, R58W and F57 do adopt different conformations. In this network, a conserved water molecule lies between the side chains of K40E and T51E and the main chain carbonyl of Y60. Side chain interactions of other neighboring residues presumably hold T51 in the orientation shown.

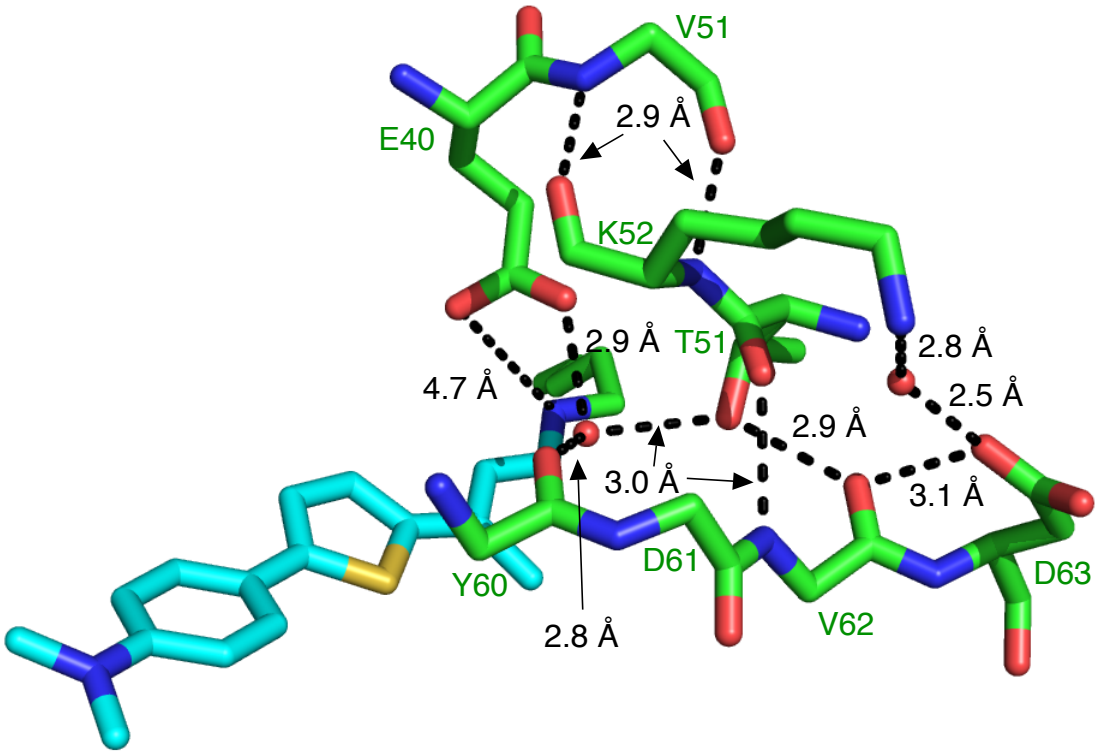


Figure IV-13. Hydrogen bonding network centered around K40E and T51, based on chain A of the crystal structure Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor**. Some side chains are not shown for clarity.

Spectroscopic properties of this hCRBPII/**ThioFluor** complex are shown in **Figure IV-14**. Similar to Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor**, the pK_a of the iminium is low and upon excitation of the SB at 392 nm at pH 7.1, a single emission band at 621 nm is observed, indicating full ESPT. However, the quantum yield is 0.30, which is much lower than Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor** (0.51). Additionally, upon excitation of the ground state PSB at 612 nm (pH 4.5), a much more red-shifted emission band is observed, with a maximum of 687 nm. It should be noted that introduction of the tryptophans at positions 58 and 19 individually did not change the quantum yield substantially upon excitation of the SB, both of which showed complete ESPT (Q108K:K40E:T53A:R58W:Q38F:Q4F was 0.44 and Q108K:K40E:T53A:R58L:Q38F:Q4F:Y19W was 0.43).

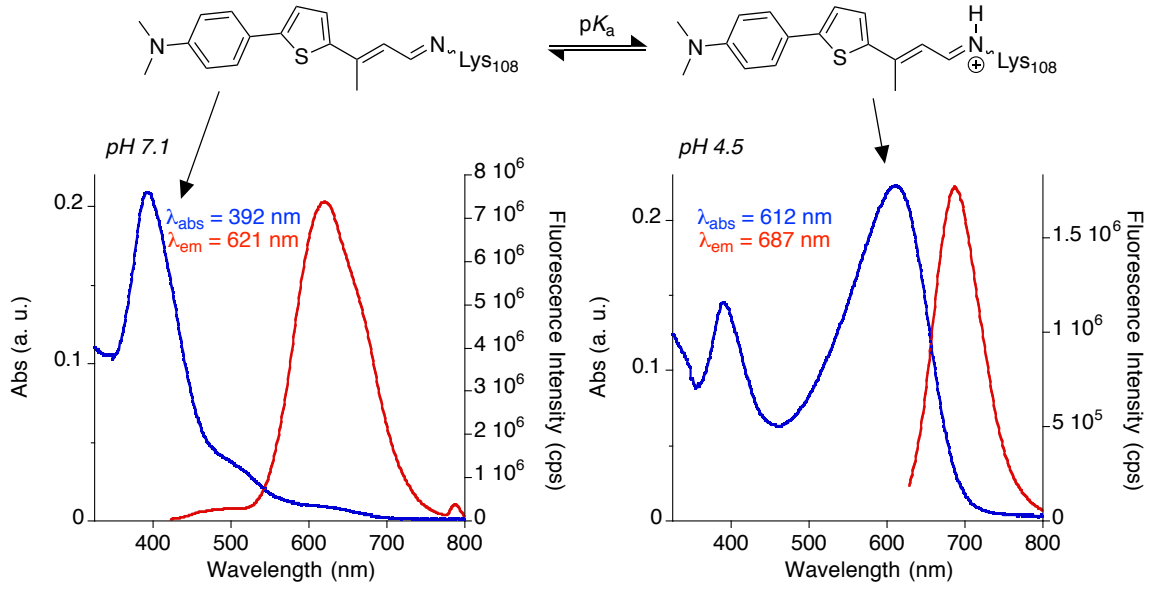


Figure IV-14. UV-Vis and emission spectra of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor** SB (pH 7.1) and PSB (pH 4.5).

IV.3 Modification of active site residues to glutamic acid

As seen in the previous section, modification of residue 40 to glutamic acid provided the most striking results. Thus, we next sought to investigate how location of the proton donor affects ESPT and fluorescent properties of the SB. To this end we chose to modify other residues along the entire chromophores' length. Glutamic acid was utilized for this purpose; however, it should be noted that aspartic acid might have produced better results at different positions, as the length of the side chain ostensibly affects the proton wire that is responsible for the ESPT pathway.

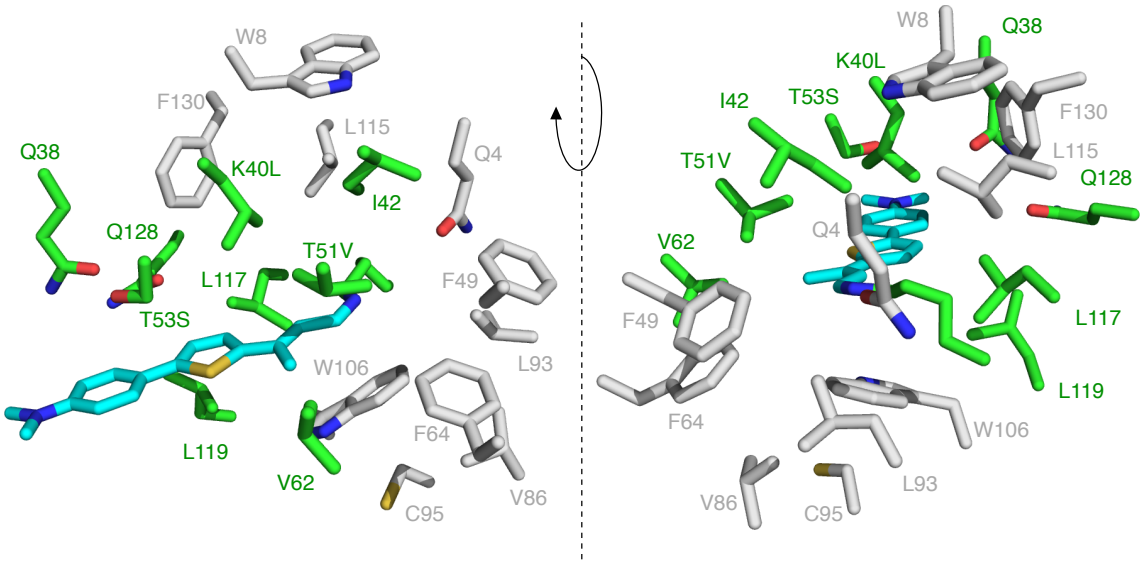


Figure IV-15. Residues mutated to glutamic acid to test whether its location allows ESPT to occur; two views are shown. Based on the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioFluor**. Mutation to glutamic acid of residues shown in grey did not lead to soluble protein expression.

We introduced glutamic acid in the same parent hCRBPII hexamutant (Q108K:K40L:T53A:R58L:Q38F:Q4F) at the following 19 residues: Q4, W8, Q38, K40, I42, F49, T51, T53, V62, F64, V86, L93, C95, W106, L115, L117, L119, Q128 and F130 (**Figure IV-15**). Residues as far as 10 Å from the imine were

tested, as it was predicted that ESPT could occur as a result of hydrogen bonding water networks. Unfortunately, many of the mutants led to insoluble protein expression, as evidenced by the presence of hCRBPII in the pellet of lysed cells via gel electrophoresis. Soluble protein was obtained after substitution of glutamic acid at nine positions, including at residues K40, T51, V62, I42, L117, T53, L119, Q38 and Q128. Interestingly, mutation of the conserved isoleucine at residue 42 to glutamic acid led to soluble protein expression, while previous attempts to insert an I42E mutation in other templates was unsuccessful. Presumably it is the presence of Q4F that allows for soluble protein expression, although the reason is unknown at this time. It should also be noted that the addition of F49E led to soluble protein expression, but only one time, and is thus not included because the results could not be reproduced.

Table IV-4. Protein expression yields of hCRBPII mutants with glutamic acid introduced at varying locations in Q108K:K40L:T53A:R58L:Q38F:Q4F.

hCRBPII Mutant	Monomer (mg/L)	Dimer (mg/L)	Mol% monomer
Q108K:K40L:T53A:R58L:Q4F:**Q38E**	16.3	-	> 95
Q108K:**K40E**:T53A:R58L:Q38F:Q4F	22.6	-	> 95
Q108K:K40L:T53A:R58L:Q38F:Q4F:**I42E**	26.9	-	> 95
Q108K:K40L:T53A:R58L:Q38F:Q4F:**T51E**	35.4	2.0	> 95
Q108K:K40L:**T53E**:R58L:Q38F:Q4F	22.7	-	> 95
Q108K:K40L:T53A:R58L:Q38F:Q4F:**V62E**	16.0	-	> 95
Q108K:K40L:T53A:R58L:Q38F:Q4F:**L117E**	32.6	1.5	98
Q108K:K40L:T53A:R58L:Q38F:Q4F:**L119E**	17.2	-	> 95
Q108K:K40L:T53A:R58L:Q38F:Q4F:**Q128E**	19.6	-	> 95

Fortuitously, only monomeric protein was isolated during protein expression except for those mutants when T51E and L117E were introduced (**Table IV-4**). The spectroscopic properties upon excitation of the SB were explored. Results, including absorption and emission maxima, quantum yield and select half-lives are detailed in **Table IV-4**.

Table IV-5. Spectroscopic properties of the SB as a result of glutamic acid introduction in the hCRBPII protein Q108K:K40L:T53A:R58L:Q38F:Q4F.

hCRBPII Mutant	λ_{abs}	λ_{em}	Φ	Φ_{ESPT}	$t_{1/2}$ (min)[a]
Q108K:K40L:T53A:R58L:Q4F:**Q38E**	389	485	0.05	< 0.02	nd
Q108K:**K40E**:T53A:R58L:Q38F:Q4F	397	605	0.51	> 0.98	3.6
Q108K:K40L:T53A:R58L:Q38F:Q4F:**I42E**	406	603	0.41	0.95	11.5
Q108K:K40L:T53A:R58L:Q38F:Q4F:**T51E**	399	602	0.12	0.94	6.4
Q108K:K40L:**T53E**:R58L:Q38F:Q4F	387	480	0.04	< 0.02	nd
Q108K:K40L:T53A:R58L:Q38F:Q4F:**V62E**	386	474, 598	0.08	0.27	nd
Q108K:K40L:T53A:R58L:Q38F:Q4F:**L117E**	391	486, 653	-	0.74	nd
Q108K:K40L:T53A:R58L:Q38F:Q4F:**L119E**	384	479	0.05	0.03	nd
Q108K:K40L:T53A:R58L:Q38F:Q4F:**Q128E**	385	473	0.05	< 0.02	nd

[a]Half-lives were computed from the rate constant that was obtained from plotting data to pseudo-first order kinetics, to ensure the numbers are comparable (note, they should only be compared in relative terms). Second order rates are not given, as concentration is difficult to determine when both SB and PSB are formed in course of binding. Rates were measured via fluorescence monitoring SB emission.

Fluorescent properties obtained upon excitation of the SB were obtained (**Table IV-5**). Again, we defined Φ_{ESPT} as the fraction of photons absorbed at the absorption maximum that leads to photochemical protonation of the chromophore's imine. Φ_{ESPT} was derived by integrating the area of the emission spectra of both the deprotonated and protonated forms obtained upon excitation

of the SB; Φ_{ESPT} was then obtained as a ratio the red-shifted emission band area over total fluorescence area. High Φ_{ESPT} values were obtained for only K40E, I42E, and T51E mutations. Mutations V62E and L117E significantly reduced Φ_{ESPT}, while all other mutations tested led to typical SB emission as the major event.

Among those mutations exhibiting a majority of ESPT fluorescence upon excitation of the SB, high quantum yields were observed for K40E and I42E, while a pronounced reduction is observed for T51E. As seen in **Figure IV-16**, all three residues are in close proximity to the imine. However, crystal structures are not available for these hCRBPII mutant/**ThioFluor** complexes.

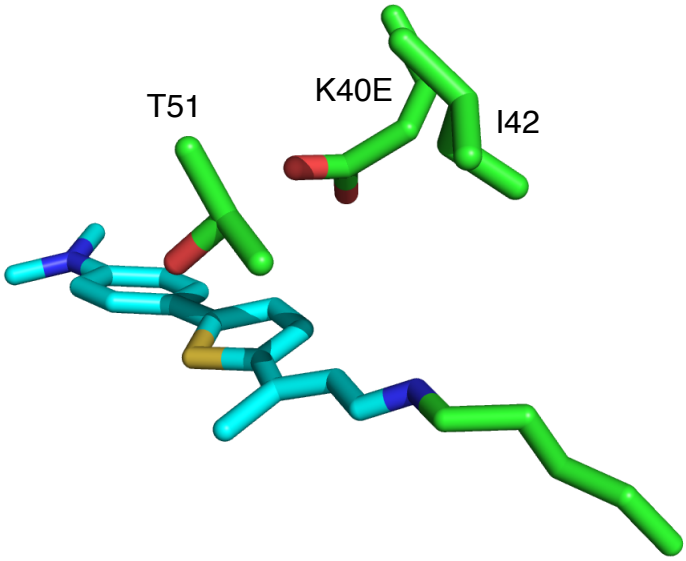


Figure IV-16. Crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor** (Chain A), showing the location of residues 40, 42, and 51.

The striking difference in quantum yields is quite interesting. It was first suspected that the hCRBPII/**ThioFluor** complexes have different excited state lifetimes. With the aid of Dr. Christopher Mancuso (Dantus lab), we were able to

obtain lifetimes for the red-shifted fluorescent species that were observed upon SB excitation at pH 7 (1060 nm laser corresponds to three-photon excitation). As seen in **Table IV-6**, there is no significant difference between the lifetimes associated with these hCRBPII mutants; it is concluded that the differences in quantum yield are not due to lifetime. Further crystallographic evidence may be useful to elucidate the decrease in quantum yield in the case of T51E, as compared to either K40E or I42E. It is possible that T51E adopts a rotamer, in which the side chain is not directly interacting with the imine.

Table IV-6. Fluorescence lifetimes of hCRBPII/**ThioFluor** complexes upon 1060 nm excitation (three photon excitation).

hCRBPII Mutant	Φ	Lifetime (ns)
Q108K:**K40E**:T53A:R58L:Q38F:Q4F	0.51	1.6
Q108K:K40L:T53A:R58L:Q38F:Q4F:**I42E**	0.41	1.9
Q108K:K40L:T53A:R58L:Q38F:Q4F:**T51E**	0.12	1.7

In preliminary studies, it was observed that when we added V62E to the hCRBPII mutant Q108K:K40D:T53A:R58L:Q38F:Q4F (Φ_{ESPT} = 0.80), ESPT was the major pathway for fluorescence. A major emission peak with maxima at 612 nm was observed (**Figure IV-6**); only a small blue-shifted shoulder corresponding to normal SB emission is apparent. Additionally, a quantum yield of 0.50 is obtained; this is a substantial increase than that observed in Q108K:K40D:T53A:R58L:Q38F:Q4F (0.34 for the monomer and 0.35 for the dimer). Based on these results, we sought to investigate whether the introduction of a second acidic residue to the optimized Q108K:K40E:T53A:R58L:Q38F:Q4F could increase quantum yield.

To this end, we attempted to introduced a second glutamic acid at the following sites: L115, F130, V62, T51, W8 and 42E. Unfortunately, no soluble protein was obtained during protein expression for L115E, F130E, W8E and I42E. hCRBPII was found to be in the pellet of lysed cells (as evidenced by protein gel electrophoresis). Two hCRBPII mutants, Q108K:K40E:T53A:R58L:Q38F:Q4F:V62E and Q108K:K40E:T53A:R58L:Q38F:Q4F:T51E provided soluble protein during expression. Unfortunately, only dimer was isolated from the expression of Q108K:K40E:T53A:R58L:Q38F:Q4F:T51E. On the other hand, Q108K:K40E:T53A:R58L:Q38F:Q4F:V62E led to both monomer and dimer (6.6 mg/L monomer and 13.8 mg/L dimer). Upon excitation of the SB with **ThioFluor**, for both monomer and dimer, single emission bands are observed at 605 nm and 608 nm, respectively. While ESPT properties did not change, quantum yield did decrease for the monomer (0.34) from the parent hCRBPII mutant Q108K:K40E:T53A:R58L:Q38F:Q4F (0.51). However, an increase in the quantum yield of the dimer upon SB was observed (0.67).

IV.4 Effect of point mutations in Q108K:K40E:T53A:R58L:Q38F:Q4F on ESPT fluorescent properties

Presumably, substitutions of residues in the putative proton wire, as well as point mutations in the immediate chromophore environment can change the ESPT pathway. Thus, we sought to investigate how the single point mutations in Q108K:K40E:T53A:R58L:Q38F:Q4F affect the fluorescent properties upon

excitation of the SB. To this end, we removed each mutation independently from the parent mutant. Unfortunate, it was observed that hCRBPII oligomerization is significantly affected by single point mutations (**Table IV-7**). Additionally, soluble protein is not obtained during expression in the absence of Q4F.

Table IV-7. Protein expression yields as a result of point mutations in Q108K:K40E:T53A:R58L:Q38F:Q4F.

hCRBPII Mutant	Monomer (mg/L)	Dimer (mg/L)	Mol% monomer
Q108K:K40E:T53A:R58L:Q38F:Q4F	22.6	-	> 95
Q108K:T53A:R58L:Q38F:Q4F	16.6	2.5	93
Q108K:K40E:R58L:Q38F:Q4F	66.7	4.3	97
Q108K:K40E:T53A:Q38F:Q4F	3.7	4.0	65
Q108K:K40E:T53A:R58L:Q4F	5.6	36.3	24

Interestingly, the single point mutations did not affect the degree of ESPT, as all showed a single emission band centered at approximately 600 nm (**Table IV-8**). This was the case for all monomers and dimers. Unfortunately, in the hCRBPII monomers, quantum yield of the SB was markedly decreased, almost in half for each mutant (**Table IV-8**). Rates for iminium formation, measured via fluorescence at 23 °C (plotted to pseudo-first order rate kinetics using 20 μM protein with 10 μM **ThioFluor**) did not substantially change between any of the mutants.

It was previously observed that the emission maxima observed upon excitation of the SB and PSB were drastically different in Q108K:K40E:T53A:R58L:Q38F:Q4F. We wanted to investigate whether this is a one time phenomena, or if it is true of mutants that contain K40E. To this end, we

explored the fluorescent properties of the PSB in the monomeric hCRBPII mutants just discussed. In comparing the SB emission maxima (**Table IV-8**) to the PSB emission maxima (**Table IV-9**), it is quite apparent that the emission obtained upon PSB excitation is redder than that obtained upon excitation of the SB, in all cases.

Table IV-8. Spectroscopic properties of the SB as a result of point mutations in Q108K:K40E:T53A:R58L:Q38F:Q4F.

hCRBPII Mutant	λ_{abs}	λ_{em}	Φ	$t_{1/2}$ (min)
Q108K:K40E:T53A:R58L:Q38F:Q4F	397	605	0.51	3.6
Q108K:K40E:R58L:Q38F:Q4F (monomer)	396	612	0.23	5.1
Q108K:K40E:R58L:Q38F:Q4F (dimer)	397	611	0.28	3.4
Q108K:K40E:T53A:Q38F:Q4F (monomer)	396	607	0.25	3.5
Q108K:K40E:T53A:Q38F:Q4F (dimer)	394	613	0.40	5.1
Q108K:K40E:T53A:R58L:Q4F (monomer)	397	619	0.23	2.3
Q108K:K40E:T53A:R58L:Q4F (dimer)	409	628	0.39	9.0

Table IV-9. Spectroscopic properties of the PSB as a result of point mutations in Q108K:K40E:T53A:R58L:Q38F:Q4F.

hCRBPII Mutant	λ_{abs}	λ_{em}
Q108K:K40E:T53A:R58L:Q38F:Q4F	575	671
Q108K:K40E:R58L:Q38F:Q4F (monomer)	554	651
Q108K:K40E:R58L:Q38F:Q4F (dimer)	557	660
Q108K:K40E:T53A:Q38F:Q4F (monomer)	568	664
Q108K:K40E:T53A:Q38F:Q4F (dimer)	566	674
Q108K:K40E:T53A:R58L:Q4F (monomer)	566	670
Q108K:K40E:T53A:R58L:Q4F (dimer)	583	670

We wanted to see if the same mutations also produced different amounts of monomer in the similar hCRBPII mutant Q108K:K40D:T53A:R58L:Q38F:Q4F.

Again, residues were changed individually back to wild type, and proteins were expressed at room temperature. As shown in **Table IV-10**, similar results were obtained; removal of R58L or Q38F results in substantially less monomer formation. In the case of R58L removal, no monomer was isolated during protein expression. Beneficially though, in this case, the removal of T53A led to a significant increase in the amount of monomer produced. Due to the large amount of dimer formation in these variants, they were not pursued further for development into a LSS fluorescent probe.

Table IV-10. Protein expression yields as a result of point mutations in Q108K:K40D:T53A:R58L:Q38F:Q4F.

hCRBPII Mutant	Monomer (mg/L)	Dimer (mg/L)	Mol% monomer
Q108K:K40D:T53A:R58L:Q38F:Q4F	8.3	30.8	35
Q108K:K40D:R58L:Q38F:Q4F	22.5	3.0	94
Q108K:K40D:T53A:Q38F:Q4F	-	22.5	< 5
Q108K:K40D:T53A:R58L:Q4F	3.2	45.7	12

IV.5 Energy transfer from tryptophan fluorescence to the bound ligand

As briefly discussed in Section **IV.2**, the emission spectra obtained upon excitation of tryptophan 280 nm suggests the existence of an energy transfer process from tryptophan to the bound **ThioFluor**-SB. **Figure IV-8** shows the spectral overlap between the protein emission and the **ThioFluor**-SB absorption spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor**. This spectral overlap indicates the possibility of energy transfer from the excited tryptophans to the bound ligand. This leads to over a 300 nm pseudo-Stokes shift (280 nm excitation with ~ 600 nm emission).

The energy transfer process between hCRBPII's tryptophans and bound **ThioFluor** has been preliminarily studied. Wild-type hCRBPII contains four tryptophan residues (**Figure IV-17**). In an attempt to understand which tryptophan is responsible for the energy transfer, each were independently mutated to phenylalanine in the parent template Q108K:K40E:R58L:Q38F:Q4F. This protein template was used because protein expression is higher without the T53A mutation (**Table IV-7**, 66.7 mg/L monomer), yet still shows complete ESPT upon excitation of the SB.

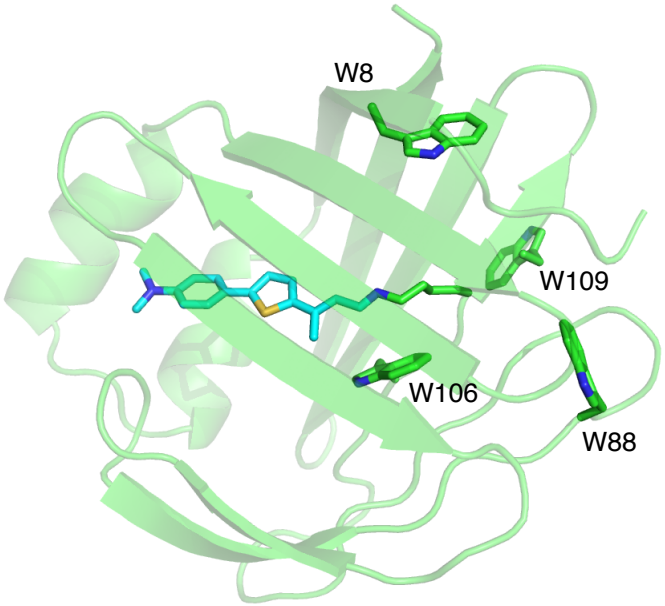


Figure IV-17. Four tryptophan residues are present in wild type hCRBPII. Based on the crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor**.

To this end, Q108K:K40E:R58L:Q38F:Q4F:W8F, Q108K:K40E:R58L:Q38F:Q4F:W88F, Q108K:K40E:R58L:Q38F:Q4F:W106F, and Q108K:K40E:R58L:Q38F:Q4F:W109F were expressed and bound with

ThioFluor. All complexes showed essentially complete ESPT upon excitation of the SB (**Table IV-11**).

Table IV-11. Spectroscopic properties of the SB with tryptophan to phenylalanine mutations in Q108K:K40E:R58L:Q38F:Q4F.

hCRBPII Mutant	Monomer Expression (mg/L)	λ_{abs}	λ_{em}	Φ
Q108K:K40E:R58L:Q38F:Q4F:W8F	27.0	398	616	0.38
Q108K:K40E:R58L:Q38F:Q4F:W88F	24.2	399	615	0.30
Q108K:K40E:R58L:Q38F:Q4F:W106F	8.9	394	615	0.33
Q108K:K40E:R58L:Q38F:Q4F:W109F	17.0	399	614	0.37

The spectroscopic properties of the parent mutant Q108K:K40E:R58L:Q38F:Q4F coupled with **ThioFluor** are shown in **Figure IV-18**. From the overlay of the emission spectrum (blue curve) upon excitation at 280 nm and the absorption spectrum of the hCRBPII/**ThioFluor** complex (green curve), it is obvious that that there is spectral overlap between tryptophan emission (λ_{max} = 355 nm) and SB absorbance (λ_{max} = 396 nm). This suggests that tryptophan should be able to transfer energy to the SB.

As shown in **Figure IV-18**, not only does excitation at 280 nm lead to tryptophan emission, but it also leads to a red-shifted emission band (λ_{max} = 612 nm; blue curve), similar to that observed upon irradiation of the SB at 396 nm (red curve). This indicates that excitation of tryptophan residues at 280 nm leads to energy transfer to the **ThioFluor**-SB, which subsequently undergoes a proton transfer in the excited state. The final result is a pseudo Stokes shift of 332 nm.

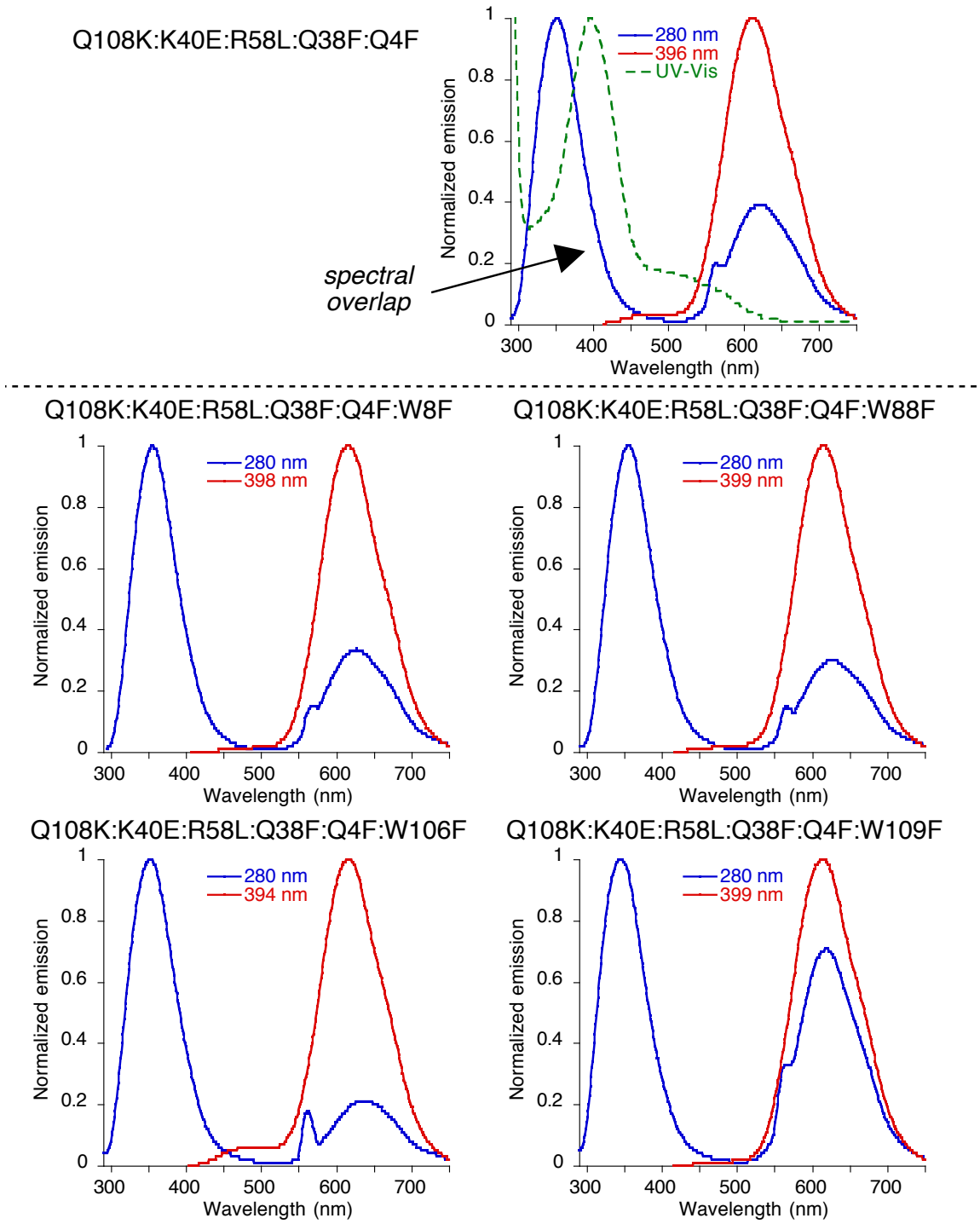


Figure IV-18. Emission spectra of Q108K:K40E:R58L:Q38F:Q4F and subsequent mutation of the four tryptophan residues in hCRBPII.

Similar results were obtained after each tryptophan was individually mutated to phenylalanine. Overlays of the emission spectra upon excitation at

280 nm (blue curve) and the SB at approximately 390 nm (red curve) of the hCRBPII/**ThioFluor** complexes are shown in **Figure IV-18**. In all cases the spectra are dominated by tryptophan emission at approximately 350 nm. However, the ratio of emissions from tryptophan (~350 nm) and the ESPT pathway (~615 nm) do vary. The most obvious difference is seen in Q108K:K40E:R58L:Q38F:Q4F:W109F. When the emission spectrum was obtained at 280 nm excitation, the quenched fluorescence emission of the tryptophan residue around 350 nm is more, as compared to the other mutants. It is interesting to note that W109 is located on the protein surface, directed toward the solvent exterior.

The emission spectra of the mutants containing W8F, W88F and W106F are comparable to that of the parent hCRBPII. Because the removal of W109 increases energy transfer, while the others do not change the fluorescent properties, it can be inferred that W109 impedes the energy transfer process to the SB. However, no other conclusions can be made at this time, except each tryptophan allows for at least partial energy transfer of their fluorescence to the bound SB. subsequent ESPT leads to the red-shifted emission band at approximately 600 nm.

When each tryptophan was removed individually, all of the mutants still showed partial energy transfer and complete ESPT upon 280 nm excitation. It can be concluded that more than one tryptophan residue is responsible for the energy transfer to the bound SB. We attempted to express and purify double

tryptophan to phenylalanine mutations, but no hCRBPII protein was isolated in any case.

IV.6 Live cell imaging of large Stokes shift hCRBPII/ThioFluor complexes

The mutant Q108K:K40E:T53A:R58L:Q38F:Q4F was chosen for live-cell imaging with **ThioFluor**. It displays ideal spectroscopic properties including fast iminium formation with a half-life of 1.7 min (measured via fluorescence with 20 μM protein with 10 μM **ThioFluor** and plotted to second order rate kinetics; concentration should be accurate as the pK_a is low enough that only SB is formed at pH 7.2), low pK_a of 5.3 (rendering almost complete SB formation at physiological pH), high quantum yield (0.51) and large Stokes shift (208 nm). Additionally, *in vitro* protein expression produces solely the monomer form of hCRBPII.

We cloned the hCRBPII hexamutant Q108K:K40E:T53A:R58L:Q38F:Q4F into plasmids containing EGFP, to serve as a control of transfection. The same backbone vector pFlag-CMV2 was used, as we have previously demonstrated that this plasmid efficiently expresses our protein in living HeLa, U2O2 and COS-7 cell lines. Cells were transfected with EGFP-Q108K:K40E:T53A:R58L:Q38F:Q4F and EGFP-Q108K:K40E:T53A:R58L:Q38F:Q4F-3NLS (nuclear localization) and labeled with 10 μM **ThioFluor** for one minute at 37 °C. Cells were subsequently washed twice and imaged immediately (**Figure IV-19**).

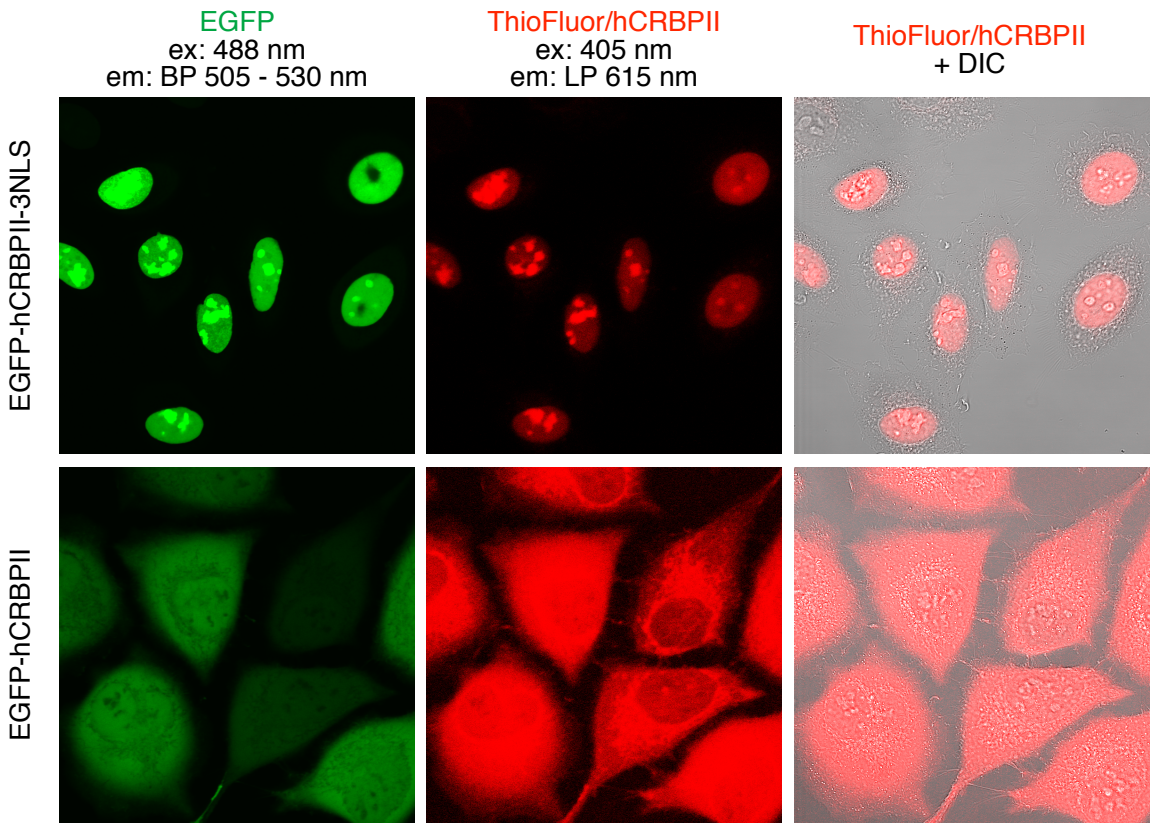


Figure IV-19. Live cell imaging of HeLa cells expressing EGFP-Q108K:K40E:T53A:R58L:Q38F:Q4F and EGFP-Q108K:K40E:T53A:R58L:Q38F:Q4F-3NLS, labeled with 10 μM **ThioFluor** for one minute at 37 °C. NLS = nuclear localization sequence.

As seen in **Figure IV-19**, **ThioFluor** was sufficient in labeling expressed hCRBPII. Excitation at 488 nm led to green fluorescence from EGFP, indicating that transfection was efficient. Fortuitously, excitation at 405 nm of the hCRBPII/**ThioFluor** complex resulted in red fluorescence, which was collected with a LP 615 nm emission filter.

While the system shows utility, the disadvantage of this system should be noted. For images shown in **Figure IV-19**, background fluorescence is not visible, presumably due to the short incubation time (one minute) required for **ThioFluor** binding to Q108K:K40E:T53A:R58L:Q38F:Q4F. Longer incubation

times and higher **ThioFluor** concentrations lead to background fluorescence under the same imaging conditions. It is hypothesized that the background fluorescence is due to unbound **ThioFluor**, due to its solvatochromic properties; **ThioFluor** absorbs at approximately 430 nm in solution and emits in the red wavelength regime (**Figure II-13**), which would disadvantageously show fluorescence under the imaging conditions for ESPT. Images of non-transfected cells, incubated with **ThioFluor**, are shown in **Figure IV-20**.

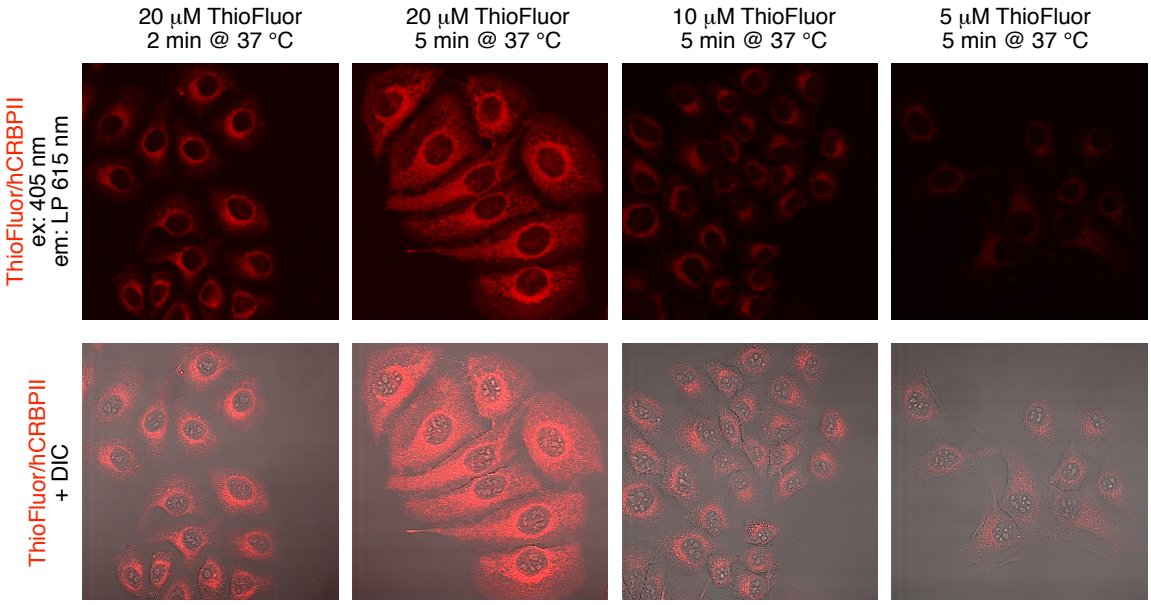


Figure IV-20. Non-transfected HeLa cells incubated with **ThioFluor** show background fluorescence when excited at 405 nm (LP 615 nm emission filter).

It is suggested that a fluorophore be used that does not display solvatochromic properties, because the far red emission observed from excitation at 405 nm is presumably due to the free chromophore. Spectroscopic properties of **ThioFluor** bound to *n*-butyl amine are shown in **Figure IV-21**. In ethanol, **ThioFluor** absorbs at 430 nm and emits at 654 nm. This large Stokes shift would allow for any unbound chromophore to be visible in the same LP 615 nm

emission channel, at which hCRBPII/**ThioFluor** fluorescence as a result of ESPT. When bound to *n*-butyl amine, the absorbance blue-shifts to 400 nm, and emits with a maximum of 540 nm. Only a small amount of ESPT occurs form the SB in ethanol. However, even this fluorescence over 615 nm may lead to background fluorescence upon excitation of non-specifically formed imine.

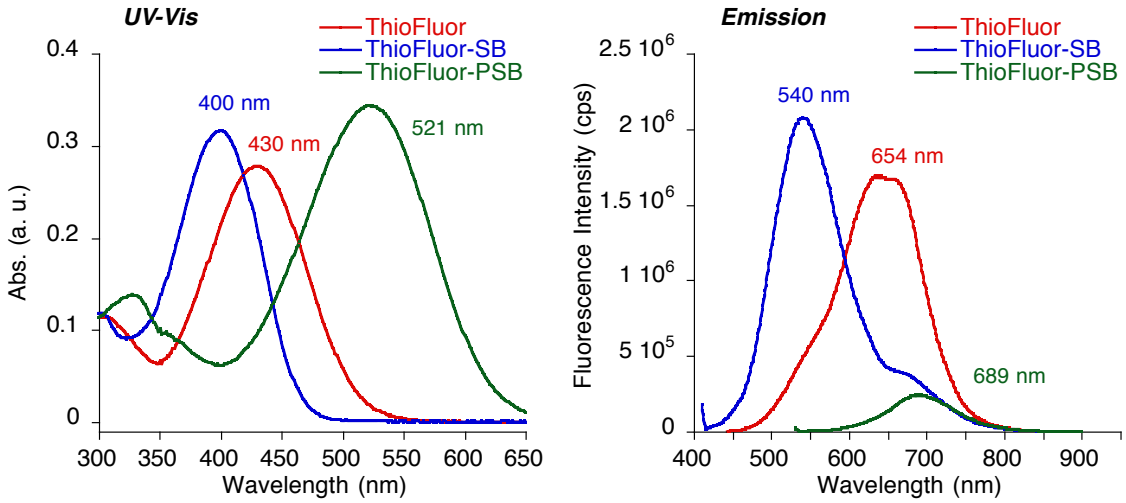


Figure IV-21. UV-Vis and emission spectra of **ThioFluor**, **ThioFluor**-SB and **ThioFluor**-PSB when bound to *n*-butyl amine.

To this end, it is suggested that a more hydrophilic fluorophore be synthesized and tested, as it is expected to stay in the cell's aqueous environment and remain nonfluorescent (recall **ThioFluor** is not fluorescent due to aggregation in PBS buffer). The addition of a sulfonic acid group to **ThioFluor** should fulfill this need, assuming other spectroscopic properties are not detrimentally altered.

IV.7 Excited state proton transfer is observed in ThioFluor analogues

Gratifyingly, the phenomenon of ESPT was also observed in **ThioFluor** analogues. **ThioFluor-3**, **ThioFluor-4**, **ThioFluor-7**, **ThioFluor-8**, **ThioFluor-9**,

and **ThioFluor-10** were analyzed with the hCRBPII mutant (Q108K:K40E:T53A:R58L:Q38F:Q4F) that was optimized for complete ESPT at physiological pH with **ThioFluor**.

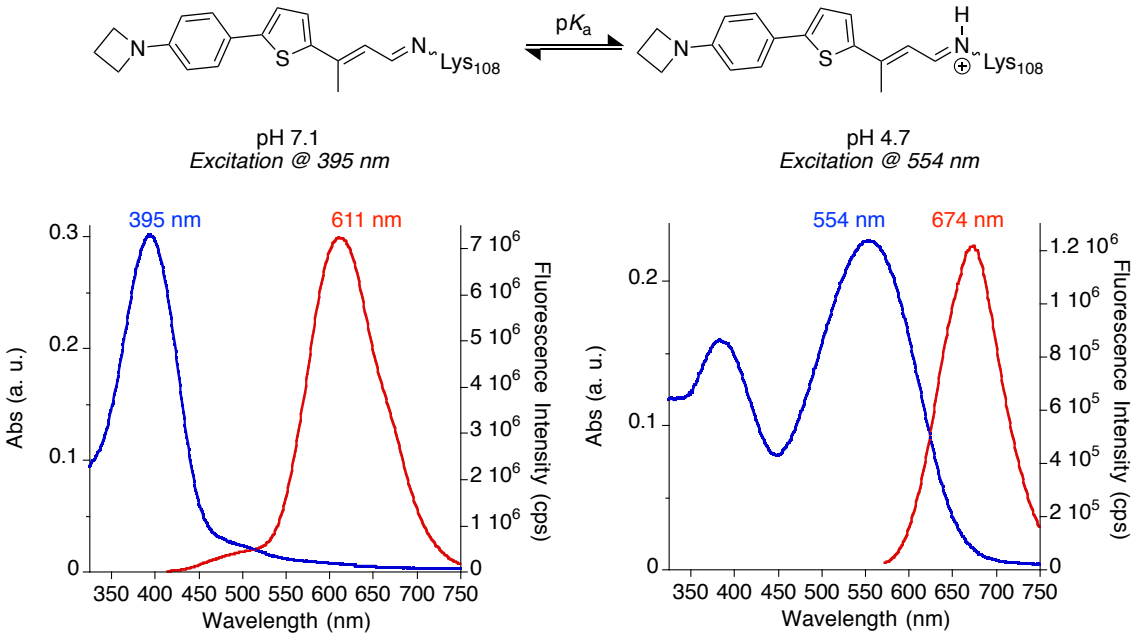


Figure IV-22. UV-Vis and emission spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-3** at pH 7.1 and pH 4.7.

Spectroscopic properties of Q108K:K40E:T53A:R58L:Q38F:Q4F bound with **ThioFluor-3** are shown in **Figure IV-22**. At neutral pH, SB is present, which upon excitation yields a red-shifted emission band at 611 nm. This is indicative of complete ESPT, as was observed for **ThioFluor**. The quantum yield is 0.58, a small increase over that observed for **ThioFluor** (0.51). Upon acidification and formation of PSB in the ground state, an absorption maximum of 554 nm is observed. Excitation of this peak results in an emission maximum of 674 nm.

Spectroscopic properties of Q108K:K40E:T53A:R58L:Q38F:Q4F bound with **ThioFluor-4** are shown in **Figure IV-23**. At neutral pH, only SB is present, which upon excitation yields emission with a maximum of band of 479 nm, with a

quantum yield of 0.02. Upon acidification and formation of PSB in the ground state, an absorption maximum of 470 nm is observed. Excitation of this peak results in an emission maximum of 550 nm. From this data it is obvious that substitution of the *N,N*-dimethyl donating group in **ThioFluor** for the less donating methoxy group in **ThioFluor-4** leads to no ESPT upon SB excitation.

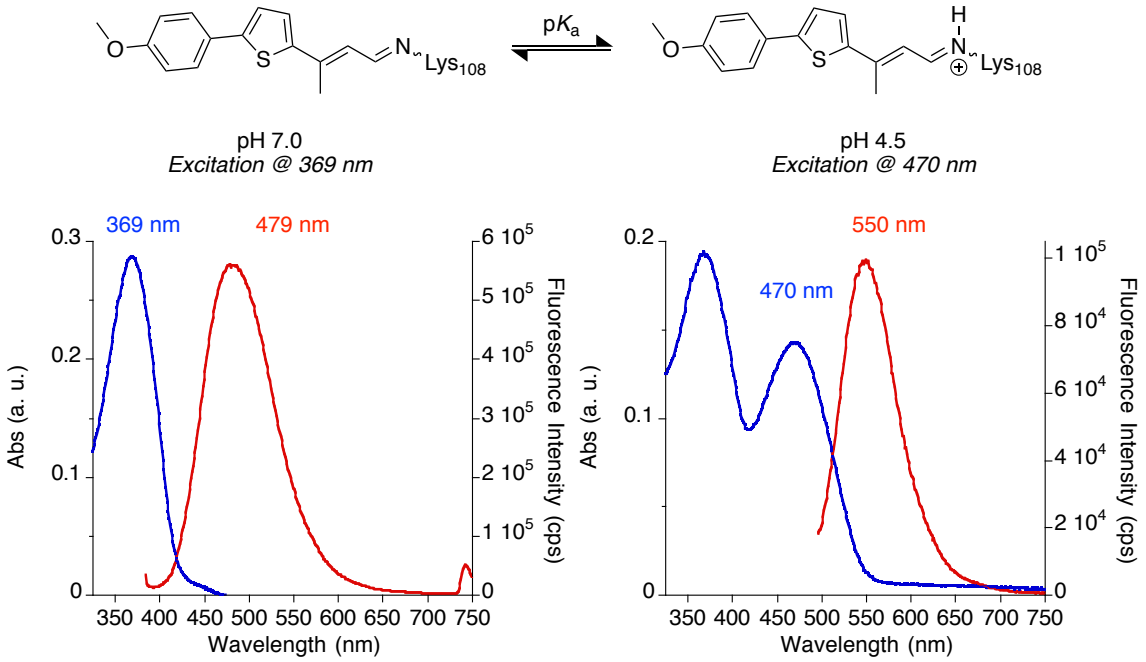


Figure IV-23. UV-Vis and emission spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-4** at pH 7.0 and pH 4.5.

Spectroscopic properties of Q108K:K40E:T53A:R58L:Q38F:Q4F bound with **ThioFluor-7** are shown in **Figure IV-24**. At neutral pH, only SB is present, which upon excitation yields emission at 516 nm, with a quantum yield of 0.11. Upon acidification no PSB was observed, even at pH 2.8. As the emission of the PSB cannot be determined, no definite conclusion can be drawn. However, an emission maximum of 516 nm is more indicative of SB emission (**ThioFluor-7**-SB

with *n*-butyl amine, **Table III-3**). It would stand to reason that the pK_a of the imine is too low and thus is incapable of abstracting a proton in the excited state.

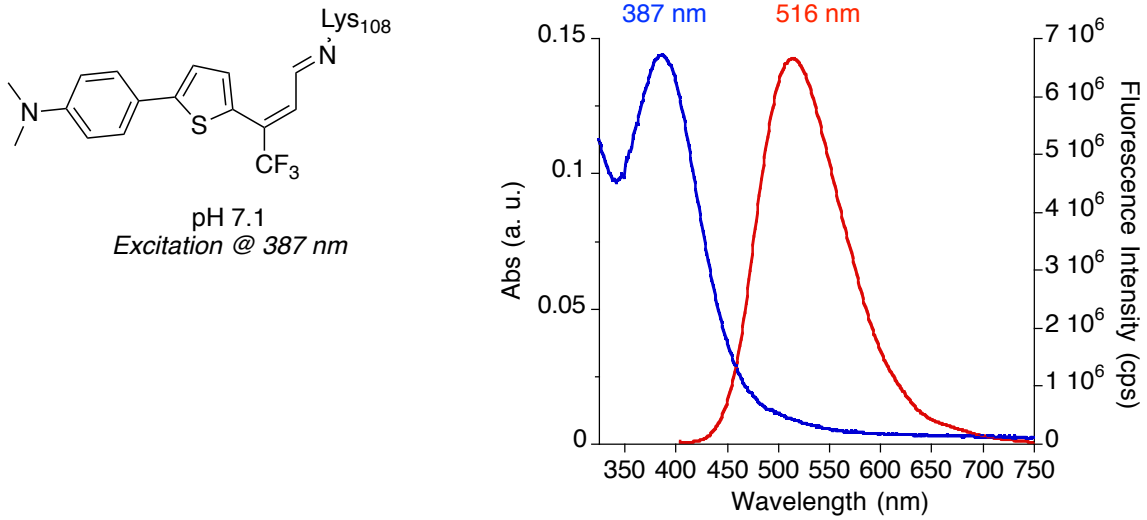


Figure IV-24. UV-Vis and emission spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-7** at pH 7.1.

Spectroscopic properties of Q108K:K40E:T53A:R58L:Q38F:Q4F bound with **ThioFluor-8** are shown in **Figure IV-25**. At neutral pH, SB is present, which upon excitation yields a red-shifted emission band at 642 nm. This is indicative of complete ESPT, as was observed for **ThioFluor**. The quantum yield is 0.45, a small decrease to that observed for **ThioFluor** (0.51), but about 30 nm more red-shifted in emission. Upon acidification and formation of the PSB in the ground state, an absorption maximum of 650 nm is observed. Excitation of this peak results in an emission maximum of 701 nm, which is 50 nm more red-shifted than the emission maximum obtained upon excitation of the SB. The iminium pK_a was determined to be 5.5, which is comparable to the pK_a determined for **ThioFluor** with the same hCRBPII mutant (5.3).

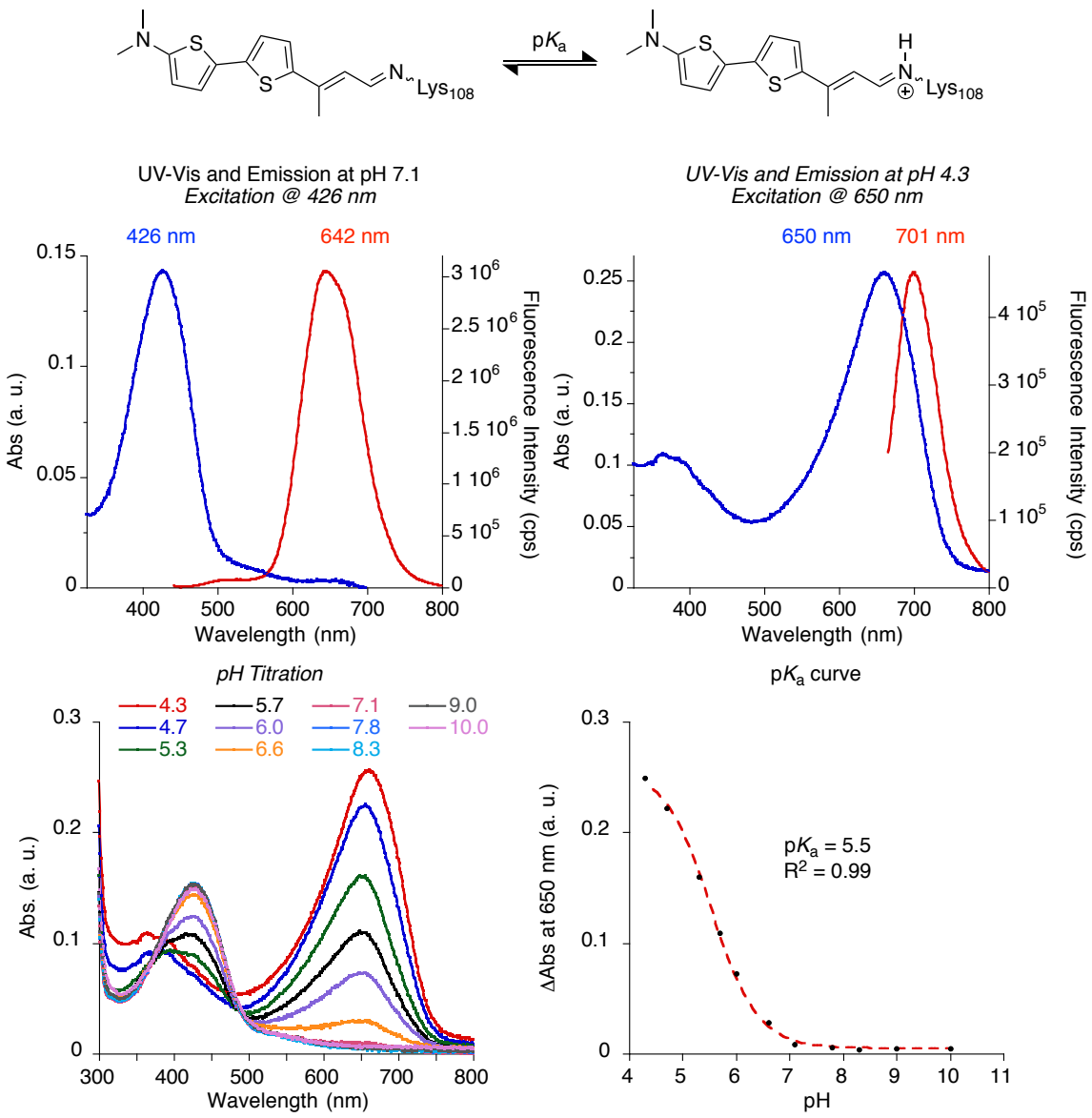


Figure IV-25. UV-Vis, emission spectra and pH titration of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-8**.

Spectroscopic properties of Q108K:K40E:T53A:R58L:Q38F:Q4F bound with **ThioFluor-9** are shown in **Figure IV-26**. At neutral pH, SB is present, which upon excitation yields dual fluorescence with emission maxima of 470 nm and 635 nm. This is indicative of only partial ESPT, as some of the typical SB emission occurs at 470 nm. The quantum yield is 0.10, a substantial decrease than that observed for **ThioFluor** (0.51). Upon acidification and formation of the

PSB in the ground state, an absorption maximum of 550 nm is observed. Excitation of this peak results in an emission maximum of 677 nm, which is approximately 40 nm more red-shifted than the red-shifted emission maximum obtained upon excitation of the SB. The iminium pK_a was determined to be 4.6, which is less than one unit different from the pK_a determined for **ThioFluor** (5.3) with the same hCRBPII mutant.

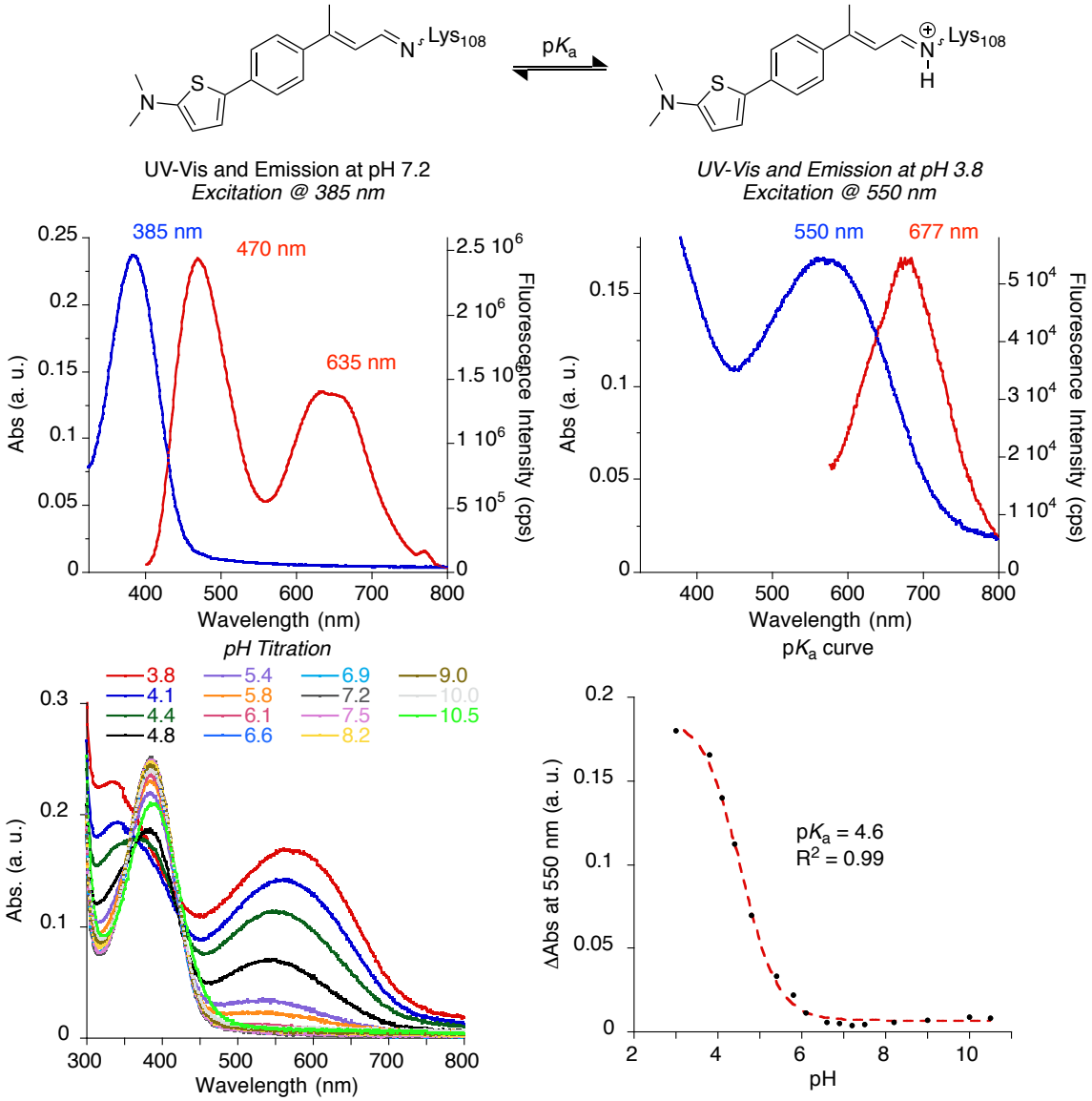


Figure IV-26. UV-Vis, emission spectra and pH titration of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-9**.

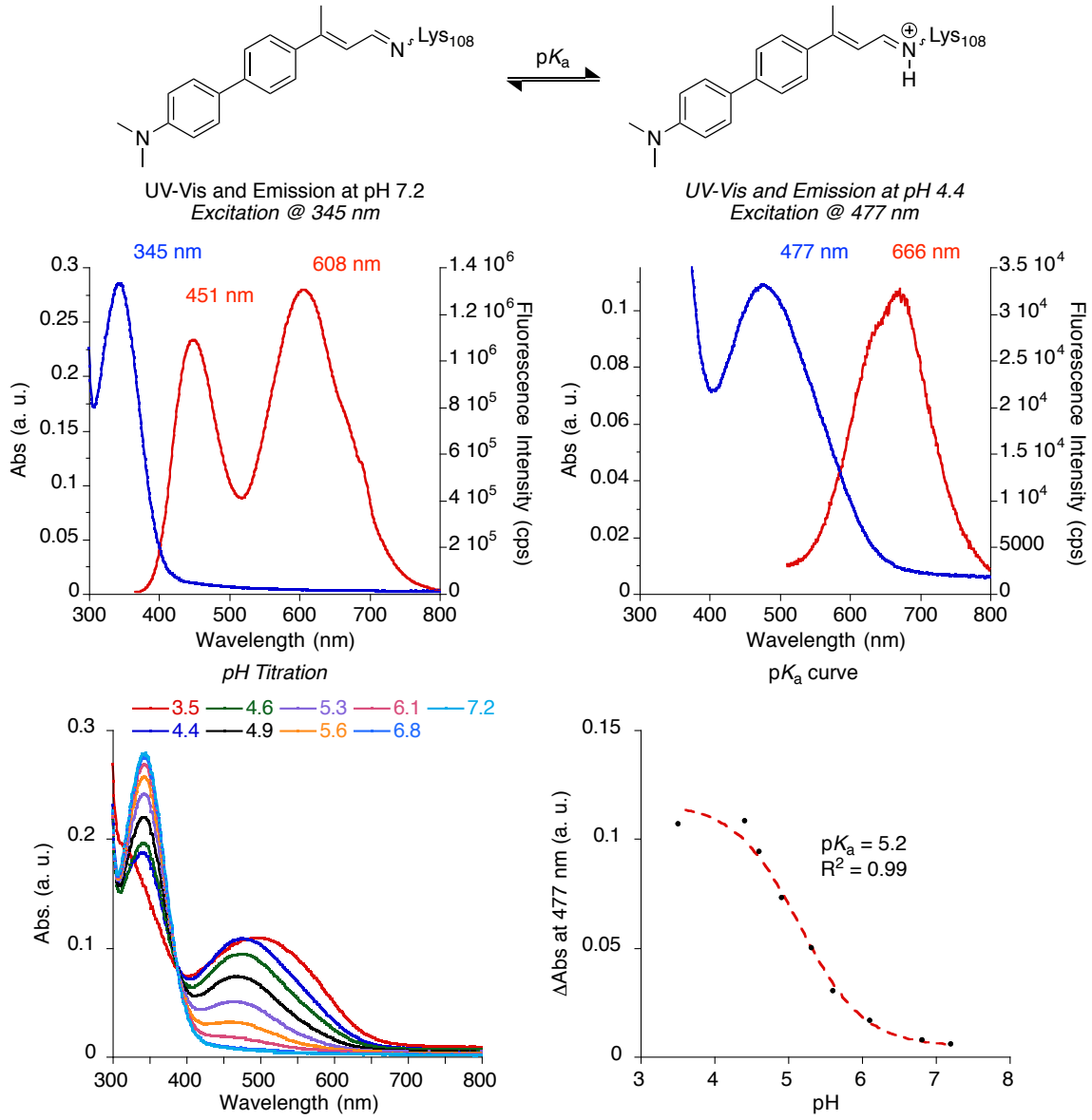


Figure IV-27. UV-Vis, emission spectra and pH titration of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor-10**.

Spectroscopic properties of Q108K:K40E:T53A:R58L:Q38F:Q4F bound with **ThioFluor-10** are shown in **Figure IV-27**. At neutral pH, SB is present, which upon excitation yields dual fluorescence with emission maxima of 451 nm and 608 nm. This is indicative of only partial ESPT, as some of the typical SB emission occurs at 451 nm. The quantum yield is 0.17, a substantial decrease than that observed for **ThioFluor** (0.51). Upon acidification and formation of the

PSB in the ground state, an absorption maximum of 477 nm is observed. Excitation of this peak results in an emission maximum of 666 nm, which is approximately 60 nm more red-shifted than the red-shifted emission maximum obtained upon excitation of the SB. The iminium pK_a was determined to be 5.2, which is almost the same as the pK_a determined for **ThioFluor** (5.3) with the same hCRBPII mutant.

Table IV-12. Spectroscopic properties of **ThioFluor** analogues coupled with Q108K:K40E:T53A:R58L:Q38F:Q4F.

Ligand	SB				PSB		pK_a
	λ_{abs}	λ_{em}	Φ	Φ_{ESPT}	λ_{abs}	λ_{em}	
ThioFluor	397	605	0.51	> 0.98	575	671	5.3
ThioFluor-3	395	611	0.58	> 0.98	554	674	nd
ThioFluor-4	369	479	0.02	< 0.02	470	550	nd
ThioFluor-8	426	642	0.45	> 0.98	650	701	5.5
ThioFluor-9	385	470, 635	0.10	0.45	550	677	4.6
ThioFluor-10	345	451, 608	0.17	0.66	477	666	5.2

The data for **ThioFluor** analogues bound with Q108K:K40E:T53A:R58L:Q38F:Q4F excited at SB and PSB is summarized in **Table IV-12**. In general, it seems that the highly conjugated systems (**ThioFluor**, **ThioFluor-3** and **ThioFluor-8**) result in only the red-shifted emission band upon SB excitation, indicating that the fluorescence occurs via a pronated excited state. These three complexes have the most red-shift SB and PSB, indicating the higher degree of conjugation. On the other hand, the SB of **ThioFluor-4** (with the less donating methoxy group) is almost 30 nm blue-shifted than that of **ThioFluor**; in addition, the PSB is 100 nm more blue-shifted. It can be inferred

that the less electron donating group leads to less overall conjugation to the iminium. As no ESPT emission is observed, it can be reasoned that the methoxy group cannot sufficiently stabilize a protonated excited state.

ThioFluor-9 and **ThioFluor-10** show partial ESPT when the SBs of the Q108K:K40E:T53A:R58L:Q38F:Q4F complexes are irradiated. At first, it was counterintuitive for **ThioFluor-10** to show more ESPT than **ThioFluor-9**. It was expected that it would be harder to break aromaticity in the biphenyl **ThioFluor-10**, leading to less conjugation than **ThioFluor-9**, which possess a phenyl ring and the less aromatic thiophene ring as the second aromatic linker.[44] The hypothesis that **ThioFluor-10** is less conjugated than **ThioFluor-9** is supported by the UV-Vis absorption of **ThioFluor-10**-SB (λ_{max} = 345 nm) and **ThioFluor-9**-SB (λ_{max} = 385 nm); **ThioFluor-10**-SB displays an absorption maxima 40 nm hypsochromically shifted of **ThioFluor-9**-SB.

From this data, it is apparent that another process is partially responsible for the dual fluorescence observed upon irradiation of **ThioFluor-10**-SB. Large Stokes shifts of hCRBPII/**ThioFluor-10** complexes were previously observed (**Chapter III**); we hypothesized that this was due to the formation of a twisted TICT state upon excitation.[10, 45, 46] In this case, it is hypothesized that excitation of **ThioFluor-10**-SB leads to a proton transfer in the excited state, assisted via TICT. These two processes can take place cooperatively, and lead to the emission profile observed. TICT facilitated ESPT was previously shown to be

energetically preferable (through computational studies) in hydrogen bonding donor-acceptor models.[47-49]

IV.8 Conclusions

Typically, large Stokes shift resulting from proton transfer do not lead to probes that are able to meet practical demands, as there are many parameters that are difficult to control, such as pH, solvent, etc. In contrast, our developed protein/probe complex exhibits excited state proton transfer only when an acidic residue is placed in the appropriate orientation to the protein-bound iminium, rendering the probe ideally suited for live-cell imaging and a turn on fluorescent response over a broad pH range. To our delight, the Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioFluor** probe can be used to image live cells with a Stokes shift of over 200 nm (and quantum yields up to 0.51).

The limitations of this system stem from the solvatochromic nature of the fluorophores, as the free aldehyde appears to emit background fluorescence above 600 nm. Efforts should be focused on fluorophore design applicable toward live-cell imaging without the need for a wash step, to allow real time measurements. Additionally, crystallographic studies and ultrafast spectroscopy may prove invaluable in shedding a detailed light on the mechanism of the excited state proton transfer process observed here.

REFERENCES

REFERENCES

1. Weissleder, R. A clearer vision for in vivo imaging. *Nature Biotechnology* **19**, 316-317 (2001).

2. Weissleder, R. & Ntziachristos, V. Shedding light onto live molecular targets. *Nature Medicine* **9**, 123-128 (2003).

3. He, B., Nie, H., Chen, L., Lou, X., Hu, R., Qin, A., Zhao, Z. & Tang, B.Z. High Fluorescence Efficiencies and Large Stokes Shifts of Folded Fluorophores Consisting of a Pair of Alkenyl-Tethered, pi-Stacked Oligo-p-phenylenes. *Organic Letters* **17**, 6174-6177 (2015).

4. Miura, T., Urano, Y., Tanaka, K., Nagano, T., Ohkubo, K. & Fukuzumi, S. Rational design principle for modulating fluorescence properties of fluorescein-based probes by photoinduced electron transfer. *Journal of the American Chemical Society* **125**, 8666-8671 (2003).

5. Trenor, S.R., Shultz, A.R., Love, B.J. & Long, T.E. Coumarins in polymers: From light harvesting to photo-cross-linkable tissue scaffolds. *Chemical Reviews* **104**, 3059-3077 (2004).

6. Loudet, A. & Burgess, K. BODIPY dyes and their derivatives: Syntheses and spectroscopic properties. *Chemical Reviews* **107**, 4891-4932 (2007).

7. Beija, M., Afonso, C.A.M. & Martinho, J.M.G. Synthesis and applications of Rhodamine derivatives as fluorescent probes. *Chemical Society Reviews* **38**, 2410-2433 (2009).

8. Lu, H., Mack, J., Yang, Y. & Shen, Z. Structural modification strategies for the rational design of red/NIR region BODIPYs. *Chemical Society Reviews* **43**, 4778-4823 (2014).

9. Grabowski, Z.R., Rotkiewicz, K. & Rettig, W. Structural changes accompanying intramolecular electron transfer: Focus on twisted intramolecular charge-transfer states and structures. *Chemical Reviews* **103**, 3899-4031 (2003).

10. Sasaki, S., Drummen, G.P.C. & Konishi, G. Recent advances in twisted intramolecular charge transfer (TICT) fluorescence and related phenomena in materials chemistry. *Journal of Materials Chemistry C* **4**, 2731-2743 (2016).

11. Dereka, B., Koch, M. & Vauthey, E. Looking at Photoinduced Charge Transfer Processes in the IR: Answers to Several Long-Standing Questions. *Accounts of Chemical Research* **50**, 426-434 (2017).

12. Klymchenko, A.S. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Accounts of Chemical Research* **50**, 366-375 (2017).

13. Loving, G.S., Sainlos, M. & Imperiali, B. Monitoring protein interactions and dynamics with solvatochromic fluorophores. *Trends in Biotechnology* **28**, 73-83 (2010).

14. Klymchenko, A.S. & Mely, Y. Fluorescent Environment-Sensitive Dyes as Reporters of Biomolecular Interactions. *Progress in Molecular Biology and Translational Science* **113**, 35-58 (2013).

15. Vollmer, F., Rettig, W. & Birckner, E. Photochemical mechanisms producing large fluorescence stokes shifts. *Journal of Fluorescence* **4**, 65-69 (1994).

16. Okamoto, A., Ichiba, T. & Saito, I. Pyrene-labeled oligodeoxynucleotide probe for detecting base insertion by excimer fluorescence emission. *Journal of the American Chemical Society* **126**, 8364-8365 (2004).

17. Honda, K., Fujishima, S.H., Ojida, A. & Hamachi, I. Pyrene excimer-based dual-emission detection of a oligoaspartate tag-fused protein by using a Zn(II)-DpaTyr probe. *ChemBiochem* **8**, 1370-1372 (2007).

18. Padalkar, V.S. & Seki, S. Excited-state intramolecular proton-transfer (ESIPT)-inspired solid state emitters. *Chemical Society Reviews* **45**, 169-202 (2016).

19. Serdiuk, I.E. & Roshal, A.D. Exploring double proton transfer: A review on photochemical features of compounds with two proton-transfer sites. *Dyes and Pigments* **138**, 223-244 (2017).

20. Song, P. & Ma, F.C. Intermolecular hydrogen-bonding effects on photophysics and photochemistry. *International Reviews in Physical Chemistry* **32**, 589-609 (2013).

21. Klymchenko, A.S., Pivovarenko, V.G., Ozturk, T. & Demchenko, A.P. Modulation of the solvent-dependent dual emission in 3-hydroxychromones by substituents. *New Journal of Chemistry* **27**, 1336-1343 (2003).

22. Zimmer, M. GFP: from jellyfish to the Nobel prize and beyond. *Chemical Society Reviews* **38**, 2823-2832 (2009).

23. Tsien, R.Y. The green fluorescent protein. *Annual Review of Biochemistry* **67**, 509-544 (1998).

24. Oltrogge, L.M., Wang, Q. & Boxer, S.G. Ground-state proton transfer kinetics in green fluorescent protein. *Biochemistry* **53**, 5947-5957 (2014).

25. Jung, G., Wiehler, J. & Zumbusch, A. The photophysics of green fluorescent protein: Influence of the key amino acids at positions 65, 203, and 222. *Biophysical Journal* **88**, 1932-1947 (2005).

26. Lossau, H., Kummer, A., Heinecke, R., PollingerDammer, F., Kompa, C., Bieser, G., Jonsson, T., Silva, C.M., Yang, M.M., Youvan, D.C. & MichelBeyerle, M.E. Time-resolved spectroscopy of wild-type and mutant Green Fluorescent Proteins reveals excited state deprotonation consistent with fluorophore-protein interactions. *Chemical Physics* **213**, 1-16 (1996).

27. 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. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8362-8367 (1996).

28. Brejc, K., Sixma, T.K., Kitts, P.A., Kain, S.R., Tsien, R.Y., Ormo, M. & Remington, S.J. Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 2306-2311 (1997).

29. Henderson, J.N., Osborn, M.F., Koon, N., Gepshtein, R., Huppert, D. & Remington, S.J. Excited State Proton Transfer in the Red Fluorescent Protein mKeima. *Journal of the American Chemical Society* **131**, 13212-13213 (2009).

30. Kogure, T., Karasawa, S., Araki, T., Saito, K., Kinjo, M. & Miyawaki, A. A fluorescent variant of a protein from the stony coral Montipora facilitates dual-color single-laser fluorescence cross-correlation spectroscopy. *Nature Biotechnology* **24**, 577-581 (2006).

31. Violot, S., Carpentier, P., Blanchoin, L. & Bourgeois, D. Reverse pH-Dependence of Chromophore Protonation Explains the Large Stokes Shift of the Red Fluorescent Protein mKeima. *Journal of the American Chemical Society* **131**, 10356-10357 (2009).

32. Bergeler, M., Mizuno, H., Fron, E. & Harvey, J.N. QM/MM-Based Calculations of Absorption and Emission Spectra of LSSmOrange Variants. *Journal of Physical Chemistry B* **120**, 12454-12465 (2016).

33. Chu, J., Oh, Y., Sens, A., Ataie, N., Dana, H., Macklin, J.J., Laviv, T., Welf, E.S., Dean, K.M., Zhang, F.J., Kim, B.B., Tang, C.T., Hu, M., Baird, M.A., Davidson, M.W., Kay, M.A., Fiolka, R., Yasuda, R., Kim, D.S., Ng, H.L. & Lin, M.Z. A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo. *Nature Biotechnology* **34**, 760-767 (2016).

34. Fron, E., De Keersmaecker, H., Rocha, S., Baeten, Y., Lu, G., Uji-I, H., Van der Auweraer, M., Hofkens, J. & Mizuno, H. Mechanism Behind the Apparent Large Stokes Shift in LSSmOrange Investigated by Time-Resolved Spectroscopy. *Journal of Physical Chemistry B* **119**, 14880-14891 (2015).

35. Pletnev, S., Shcherbakova, D.M., Subach, O.M., Pletneva, N.V., Malashkevich, V.N., Almo, S.C., Dauter, Z. & Verkhusha, V.V. Orange Fluorescent Proteins: Structural Studies of LSSmOrange, PSmOrange and PSmOrange2. *Plos One* **9** (2014).

36. Shcherbakova, D.M., Hink, M.A., Joosen, L., Gadella, T.W.J. & Verkhusha, V.V. An Orange Fluorescent Protein with a Large Stokes Shift for Single-Excitation Multicolor FCCS and FRET Imaging. *Journal of the American Chemical Society* **134**, 7913-7923 (2012).

37. Chen, F.S., Zeng, Q., Zhuang, W. & Liang, W.Z. Characterizing the Structures, Spectra, and Energy Landscapes Involved in the Excited-State Proton Transfer Process of Red Fluorescent Protein LSSmKate1. *Journal of Physical Chemistry B* **120**, 9833-9842 (2016).

38. Piatkevich, K.D., English, B.P., Malashkevich, V.N., Xiao, H., Almo, S.C., Singer, R.H. & Verkhusha, V.V. Photoswitchable Red Fluorescent Protein with a Large Stokes Shift. *Chemistry & Biology* **21**, 1402-1414 (2014).

39. Piatkevich, K.D., Hulit, J., Subach, O.M., Wu, B., Abdulla, A., Segall, J.E. & Verkhusha, V.V. Monomeric red fluorescent proteins with a large Stokes shift. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5369-5374 (2010).

40. Piatkevich, K.D., Malashkevich, V.N., Almo, S.C. & Verkhusha, V.V. Engineering ESPT Pathways Based on Structural Analysis of LSSmKate

Red Fluorescent Proteins with Large Stokes Shift. *Journal of the American Chemical Society* **132**, 10762-10770 (2010).

41. Randino, C., Moreno, M., Gelabert, R. & Lluch, J.M. Peek at the Potential Energy Surfaces of the LSSmKate1 and LSSmKate2 Proteins. *Journal of Physical Chemistry B* **116**, 14302-14310 (2012).

42. Shcherbo, D., Merzlyak, E.M., Chepurnykh, T.V., Fradkov, A.F., Ermakova, G.V., Solovieva, E.A., Lukyanov, K.A., Bogdanova, E.A., Zaraisky, A.G., Lukyanov, S. & Chudakov, D.M. Bright far-red fluorescent protein for whole-body imaging. *Nature Methods* **4**, 741-746 (2007).

43. Yang, J., Wang, L., Yang, F., Luo, H.M., Xu, L.L., Lu, J.L., Zeng, S.Q. & Zhang, Z.H. mBeRFP, an Improved Large Stokes Shift Red Fluorescent Protein. *Plos One* **8** (2013).

44. Chen, W., Li, H., Widawsky, J.R., Appayee, C., Venkataraman, L. & Breslow, R. Aromaticity decreases single-molecule junction conductance. *Journal of the American Chemical Society* **136**, 918-920 (2014).

45. Zhang, Z.Y., Zhang, G.Q., Wang, J.X., Sun, S.S. & Zhang, Z.Z. The mechanisms of Large Stokes Shift and Fluorescence Quantum Yields in anilino substituted Rhodamine analogue: TICT and PICT. *Computational and Theoretical Chemistry* **1095**, 44-53 (2016).

46. Haberhauer, G., Gleiter, R. & Burkhart, C. Planarized Intramolecular Charge Transfer: A Concept for Fluorophores with both Large Stokes Shifts and High Fluorescence Quantum Yields. *Chemistry-A European Journal* **22**, 971-978 (2016).

47. Kim, C.H., Park, J., Seo, J., Park, S.Y. & Joo, T. Excited state intramolecular proton transfer and charge transfer dynamics of a 2-(2'-hydroxyphenyl)benzoxazole derivative in solution. *The Journal of Physical Chemistry A* **114**, 5618-5629 (2010).

48. Wang, D.D., Lu, R., Yuan, M.H., Chen, J.S., Feng, L.Q., Fu, A.P., Tian, F.H., Varandas, A.J.C. & Chu, T.S. Photoinduced coupled twisted intramolecular charge transfer and excited-state proton transfer via intermolecular hydrogen bonding: A DFT/TD-DFT study. *Chemical Physics Letters* **610**, 179-185 (2014).

49. Wang, Y., Shi, Y., Cong, L. & Li, H. TDDFT study of twisted intramolecular charge transfer and intermolecular double proton transfer in the excited

state of 4 '-dimethylaminoflavonol in ethanol solvent. *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy* **137**, 913-918 (2015).

CHAPTER V: DEVELOPMENT OF A SINGLE PROTEIN FLUORESCENT RATIOMETRIC PH SENSOR

Our lab previously showed success in reengineering cellular retinoic acid binding protein II (CRABPII) to bind retinal as a protonated Schiff base (PSB), demonstrating wavelength tuning over a 166 nm range (474 nm to 640 nm) and pK_a modulation from 2.4 to 8.1.[1] This led to the development of a colorimetric pH sensor, in which a CRABPII-CRABPII protein fusion was used to measure pH ratiometrically. The two CRABPII mutants chosen were R111K:R132L:Y134F:T54V:L121Q:R59W (pK_a 4.4, λ_{max} 505 nm) and R111K:R132L:Y134F:T54V:R59W:A32W:M93L:E73A (pK_a 2.6, λ_{max} 607 nm). **Figure V-1a** depicts the absorption spectra of the individually expressed CRABPII mutants when bound to retinal.

The two proteins were cloned into the expression vector pET-17b with a three amino acid linker. The expressed protein fusion was incubated with 2 equiv of retinal in citrate buffer, and subsequently titrated (**Figure V-1b**). As expected, upon acidification, an iminium corresponding to the high pK_a mutant first appears; further acidification leads to the iminium of the low pK_a mutant. To build a standard curve, the ratio of absorbance intensity at 480 nm to 650 nm was plotted against pH (**Figure V-1c**); standard curves at concentration of 58 μM and 19 μM overlapped, purporting the concentration independent response of the system. As a proof of principle, the protein fusion was used to measure the pH of known buffer solutions with accuracy within 0.2 pH units.

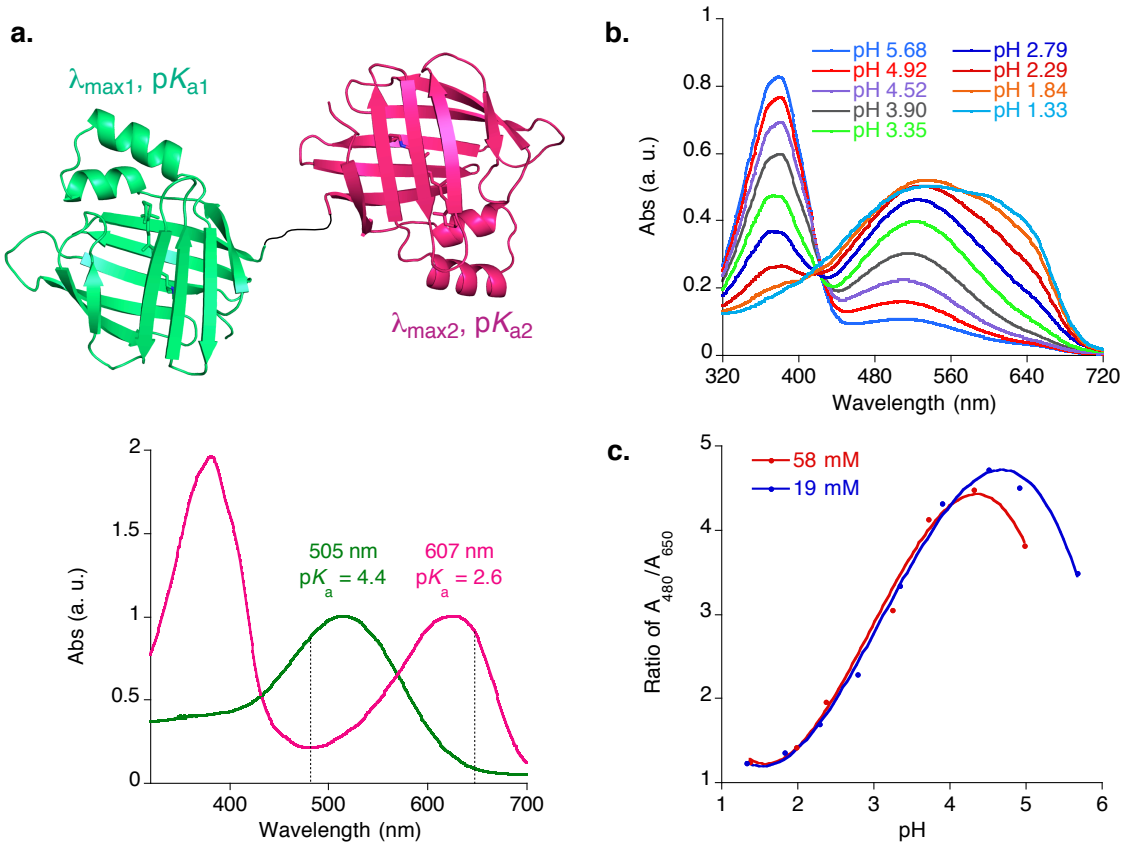


Figure V-1. Two-protein-based ratiometric probe design. **a**. The selected proteins have distinct wavelength (UV-Vis) and pK_a values. **b**. pH titration of the fused protein. **c**. Ratiometric analysis of the fused protein system at two different concentrations.

The system just discussed relied on the ratio of absorbances. However, fluorescence based pH probes are suitable for more applications. Fluorescence detection is greater than absorption spectrophotometric methods, which leads to greater limits of detection. Thus, Dr. Tetyana Berbasova pioneered efforts to develop a fluorescent ratiometric pH sensor. The fortuitous discovery that human Cellular Retinol Binding Protein (hCRBPII) mutants containing a L117E mutation showed a pH dependent wavelength shift of the protonated Schiff based (PSB) when bound to julolidine retinal analog (JRA) allowed for the generation of a single protein ratiometric pH sensor.[2] As seen in **Figure V-2a**, the pentamutant

Q108K:K40L:T51V:R58F:L117E showed two distinct absorption maxima at pH 5.0 and pH 8.0. The tetramutant lacking the L117E mutation shows the necessity of the titratable carboxylate side chain near the iminium (**Figure V-2b**). It was hypothesized that the carboxylate form of L117E localizes charge on the PSB, which leads to the observed blue-shift (**Figure V-2c**). Acidification results in protonation of the side chain, leading to better delocalization of charge and a red-shift in absorption wavelength.

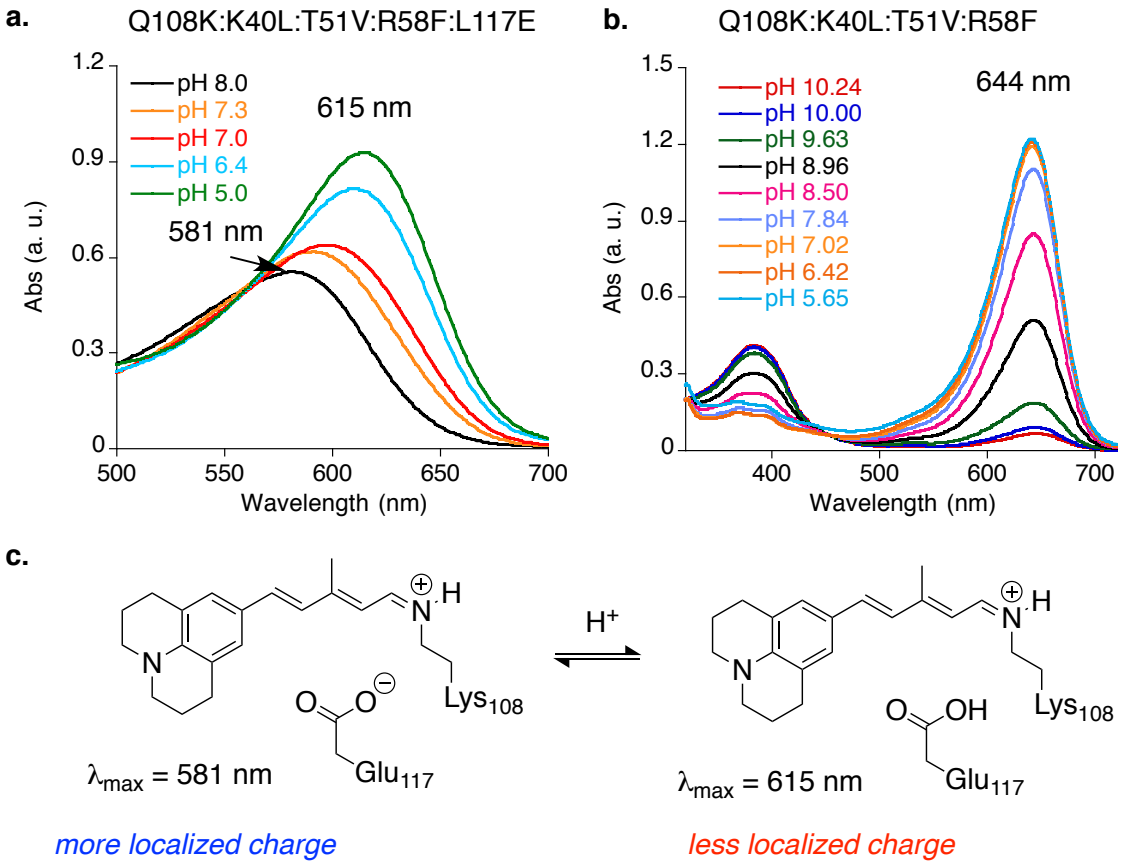


Figure V-2. Single protein ratiometric probe design. **a**. Titration of L117E leads to a shift in wavelength. **b**. Removal of L117E does not lead to the pH induced wavelength change. **c**. A blue-shift in wavelength results due to localizing charge on the iminium through interaction with the carboxylate, while protonation of the L117E side chain leads to less localized charge.

Through mutagenesis of active site residues, the largest shift between the two PSB maxima achieved was 95 nm (565 nm and 660 nm) in the nonamutant Q108K:K40L:T51V:T53C:R58W:T29L:A33W:Q4F:L117D, with a quantum yield of 0.18.[2] The large shift in absorption wavelength allows for excitation at two distinct wavelengths, from which the emission ratios can be plotted as a function of pH to develop a ratiometric standard curve. It was envisioned that the ratio of emission intensities at the two excitation wavelengths in live cells expressing hCRBPII-L117E could be obtained. From the standard curve, the *in vivo* pH could then be determined. Unfortunately live cell imaging was not fruitful; JRA leads to significant background fluorescence when incubated with both HeLa and U2OS cells expressing hCRBPII, while specific labeling was not achieved.

V.1 Initial observation of pH induced wavelength shifts in a ThioFluor/hCRBPII complex

In the course of optimizing binding of **ThioFluor** to hCRBPII mutants in wavelength regulation studies (**Chapter II**), it was observed that the inclusion of L117E produced the same pH induced wavelength shift as was observed for JRA. **ThioFluor** was coupled with the heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E (**Figure V-3**). At pH 7, an absorption maximum of 558 nm is apparent; acidification leads to a bathochromic shift in absorption with a maximum of 614 nm, a 56 nm change. When the absorption intensity at 614 nm was plotted as a function of pH, a double pK_a

curve is obtained. Fitting two pK_as to the curve yields values of 6.6 and 11.2, which presumably correspond to the pK_as of L117E and iminium, respectively.

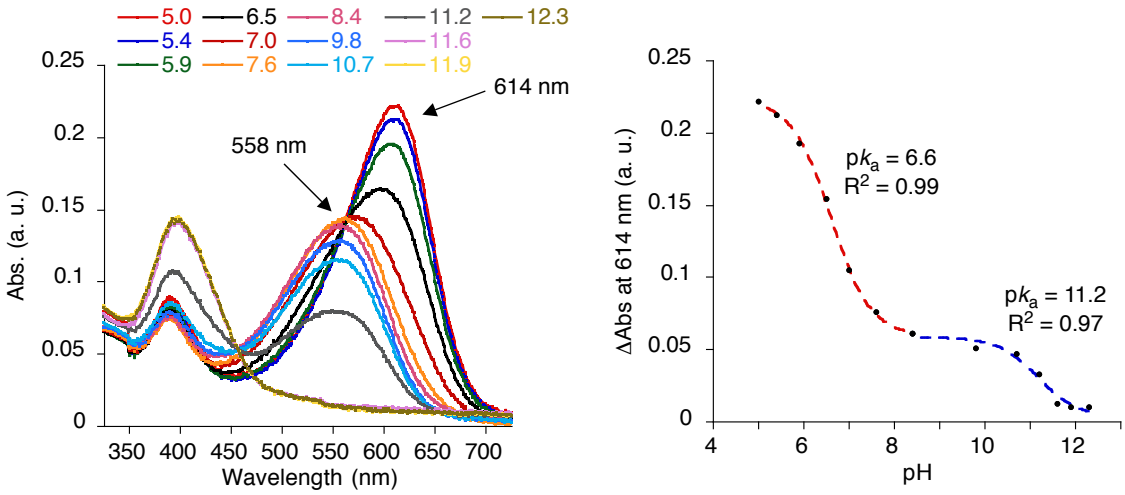


Figure V-3. Acid-base titration of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**.

To our delight, the crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E (crystallized at 1.2 Å resolution by Ms. Zahra Assar) shows a distance of 2.9 Å from the glutamic acid side chain to the iminium (**Figure V-4**). Unexpectedly, this high pK_a mutant exists as a *trans* iminium when crystallized at pH 4 – 4.5; in crystal structures of retinal/hCRBPII, it was previously shown that high pK_a mutants typically exhibited a *cis* iminium.[3, 4]

As previously stated, Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** has an absorption maximum of 558 nm; acidification leads to a bathochromic shift in absorption with a maximum of 614 nm, a 56 nm change. This bathochromic shift in wavelength with protonation of the L117E side chain is in agreement with the idea that an even distribution of electrostatic potential along the chromophore leads to red-

shifting. Moreover, the mutant without L117E (Q108K:K40L:T51V:T53S:R58W:Y19W) is more red-shifted, with an absorption maximum of 653 nm and extinction coefficient of approximately 40,000 $M^{-1}cm^{-1}$.

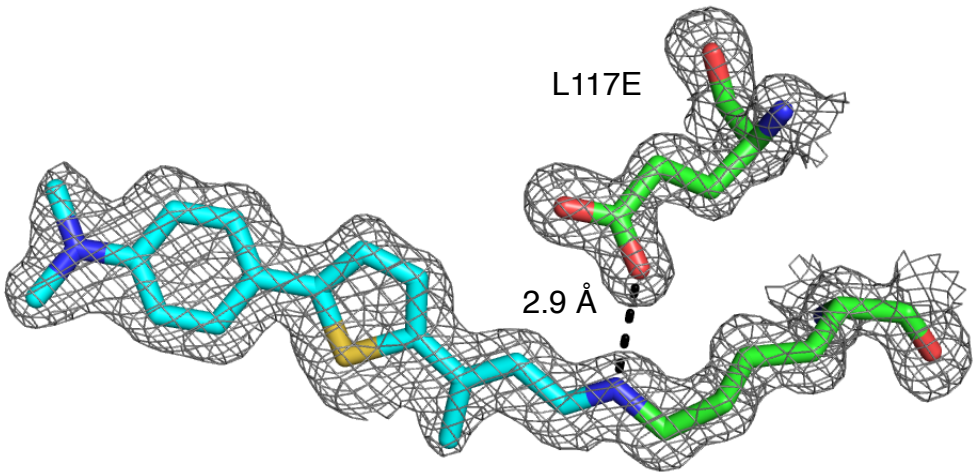


Figure V-4. Crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** highlighting the location of L117E. Density is shown at 1σ.

Not only does the addition of L117E to this mutant produce the pH induced wavelength change, but it also has the added benefit of increasing the rate of iminium formation by 4.5 fold ($t_{1/2}$ = 370 minutes for Q108K:K40L:T51V:T53S:R58W:Y19W and 82 minutes for Q108K:K40L:T51V:T53S:R58W:Y19W:L117E; iminium formation was measured via UV-Vis at 23°C with 20 μM protein and 0.5 equiv ligand, and plotted to second order rate kinetics).

A disadvantage of the latter system is that the extinction coefficient of hCRBPII/**ThioFluor** is significantly decreased with the addition of L117E (approximately two-fold). Ms. Zahra Assar was also successful in obtaining the crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor** (at 1.2 Å

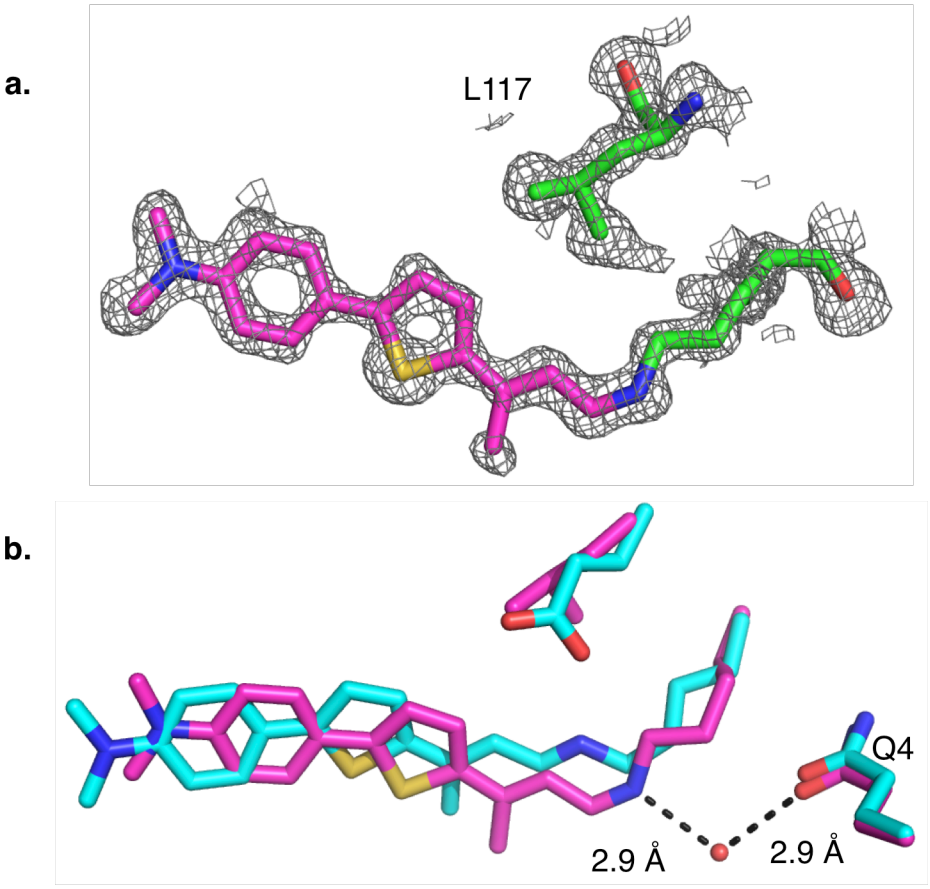


Figure V-5. **a**. Crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor**. Density is shown at 1σ. **b**. Overlay of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor** (magenta) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** (cyan).

resolution) which clearly shows a *cis* iminium stabilized by a water mediated hydrogen bond to Q4 (**Figure V-5a**). An overlay of Q108K:K40L:T51V:T53S:R58W:Y19W/**ThioFluor** and Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** shows that the chromophore is translocated toward the protein exterior when the *trans* iminium is present (**Figure V-5b**). However, there are no significant differences in any of the nearby residues. It is hypothesized that the decrease in extinction coefficient upon L117E introduction is due to the isomerization at the iminium, the

translocation of the chromophore, or because of simply introducing polarity near the iminium.

Fortuitously this Q108K:K40L:T51V:T53S:R58W:Y19W:L117E complex has shown success in labeling live HeLa cells with whole cell labeling, and both nuclear and cytosolic targeting (**Chapter II.11.2**). We first chose to explore how the excitation wavelengths affect the emission ratio. To this end, we collected emission spectra at four excitation wavelengths (at each pH) that correspond to lasers available on the confocal microscope: 514 nm, 543 nm, 594 nm and 633 nm. Each emission spectrum was then integrated from either 615 nm to 900 nm or 650 nm to 900 nm, in order to mimic the long pass (LP) 615 nm and 650 nm filters available on the confocal microscope. All data is shown in **Table V-1** for Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**.

Table V-1. Emission ratios (*in vitro*) as a function of pH for Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**.

pH	$\lambda_{em1}/\lambda_{em2}$ (LP 615 nm)			$\lambda_{em1}/\lambda_{em2}$ (LP 650 nm)					
	514/543	514/594	543/594	514/543	514/594	514/633	543/594	543/633	594/633
5.0	0.55	0.37	0.67	0.55	0.36	0.50	0.66	0.91	1.39
5.4	0.57	0.40	0.70	0.57	0.39	0.56	0.69	0.99	1.44
5.9	0.62	0.50	0.79	0.61	0.47	0.71	0.77	1.17	1.52
6.5	0.71	0.72	1.00	0.70	0.66	1.15	0.95	1.65	1.74
7.0	0.79	1.05	1.32	0.78	0.98	2.35	1.25	3.01	2.40
7.6	0.82	1.27	1.54	0.81	1.20	4.29	1.48	5.29	3.57
8.4	0.83	1.38	1.67	0.82	1.33	6.52	1.62	7.94	4.90
9.8	0.83	1.42	1.71	0.82	1.37	7.59	1.67	9.21	5.53
10.7	0.84	1.49	1.77	0.83	1.44	8.32	1.72	10.00	5.80

As seen in **Table V-1**, cutting the emission off at 615 nm or 650 nm makes essentially no difference in the ratio of emission intensities (i.e. 0.84 and 0.83 for excitation wavelengths of 514 nm and 543 nm, respectively). However, the emission ratio significantly varies depending on the two excitation wavelengths used. The smallest changes between pH 5.0 and pH 10.7 are observed when the excitation wavelengths are closer (594 and 633 nm). On the other hand, when the excitation wavelengths are further apart, and close to the maxima of the hCRBPII/**ThioFluor** complex, the emission ratio is higher.

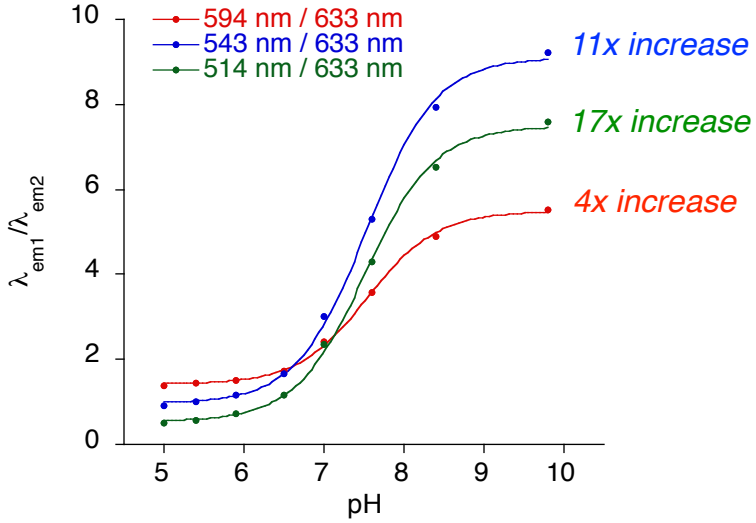


Figure V-6. Select standard curves obtained upon excitation of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** at 514/633 nm, 543/633 nm and 594/633 nm.

In the case of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**, the highest emission ratio is obtained upon using excitation wavelengths of 543 nm and 633 nm. Standard curves can be generated by plotting emission ratio versus pH. An overlay of the three best standard curves is shown in **Figure V-6**. The largest increase in fluorescence emission ratio is observed when 514 nm

and 633 nm excitation wavelengths are used, leading to a 17-fold dynamic range. Presumably this is because the two different iminiums can be selectively excited at the tail end of their absorbances. In doing so the emission ratio starts more toward zero, indicating that with an excitation wavelength of 514 nm, little fluorescence is observed from the iminium when L117E is protonated. Based on this curve, the dynamic range is 6.5 to 8.5, which should prove useful for measuring cells at physiological pH.

With this initial hit, we sought to optimize its use for live cell imaging. To this end, we made point mutations in the Q108K:K40L:T51V:T53S:R58W:Y19W:L117E heptamutant. The following requirements were kept in mind when engineering hCRBPII for applicability toward a single fluorescent protein ratiometric pH sensor. Firstly, we sought to increase the change between the two absorptive states to allow for excitation at two distinct wavelengths. It is also essential to have a high pK_a value of the iminium, such that the iminium remains protonated during the titration. This will ensure that a change to the absorption spectra of the iminium is registered as the result of titrating the side chain carboxylic acid. It is important to note that the titratable side chain should possess a pK_a that allows a pH working range that is useful at physiological pH. Lastly, as in all imaging methods, fast kinetics and high brightness (the product of extinction coefficient and quantum yield) are ideal in live cell imaging, so that measurements can be taken directly following fluorophore incubation.

V.2 Optimization of hCRBPII for a ratiometric pH probe

In screening hCRBPII mutants for **ThioFluor** binding, it was observed that T53D and R58H mutations increased the rate of iminium formation. To this end, they were introduced individually into the parent template Q108K:K40L:T51V:T53S:R58W:Y19W:L117E. In both cases, the absorption wavelengths of the iminium, upon titrating the L117E, are more separated. Advantageously, the quantum yield of the PSB was also increased. Unfortunately though, the introduction of T53D (**Table V-2**, entry 2) significantly decreased both the rate of iminium formation and the extinction coefficient of the bound complex.

Table V-2. Initial mutagenesis of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E for development toward a single fluorescent protein ratiometric sensor.

Entry	hCRBPII mutant[a]	λ_{abs}[b]	ΔAbs	pK_a[c]	Φ[d]	$t_{1/2}$ (min)[e]
1	KLVS:R58W:Y19W:L117E	558, 614	56	11.2, 6.6	0.31	82
2	KLVD:R58W:Y19W:L117E	530, 606	76	9.7, 6.9	0.41	217
3	KLVS:R58H:Y19W:L117E	530, 598	68	10.9, 6.6	0.37	17
4	KLVS:R58W:Y19W:L117E:A33W	539, 628	89	11.3, 8.0	0.28	241
5	KLVS:R58H:Y19W:L117E:A33W	547, 614	67	10.9, 7.0	0.29	30

[a]KLVX = Q108K:K40L:T51V:T53X. [b]λ_{max} of absorbance at basic pH and λ_{max} of absorbance at acidic pH. [c]pK_a corresponding to the iminium and pK_a corresponding to Glu117. [d]Quantum yield measured at pH ~7.2. [e]Kinetic measurements were performed at 23 °C with 20 μM protein and 0.5 equiv ligand. PSB formation was monitored by UV-Vis at λ_{max} for each complex over time at pH ~7.2. Data was then fit to second order kinetics.

To our delight the substitution of R58H to yield Q108K:K40L:T51V:T53S:R58H:Y19W:L117E (**Table V-2**, entry 3) led to fast iminium formation (**Figure V-7a**), and retained the high iminium pK_a (**Figure V-7b**

and **Figure V-7c**). Fortuitously, the highest emission ratio obtained when excited at 514 nm and 633 nm is also significantly increased to approximately 40 (**Figure V-7d**), which should allow for more sensitive detection. The disadvantage of this complex is that the working range (7.4 – 9.4) is not useful for most physiological applications, but it may find use in measuring the pH of mitochondria (pH ~ 8).[5]

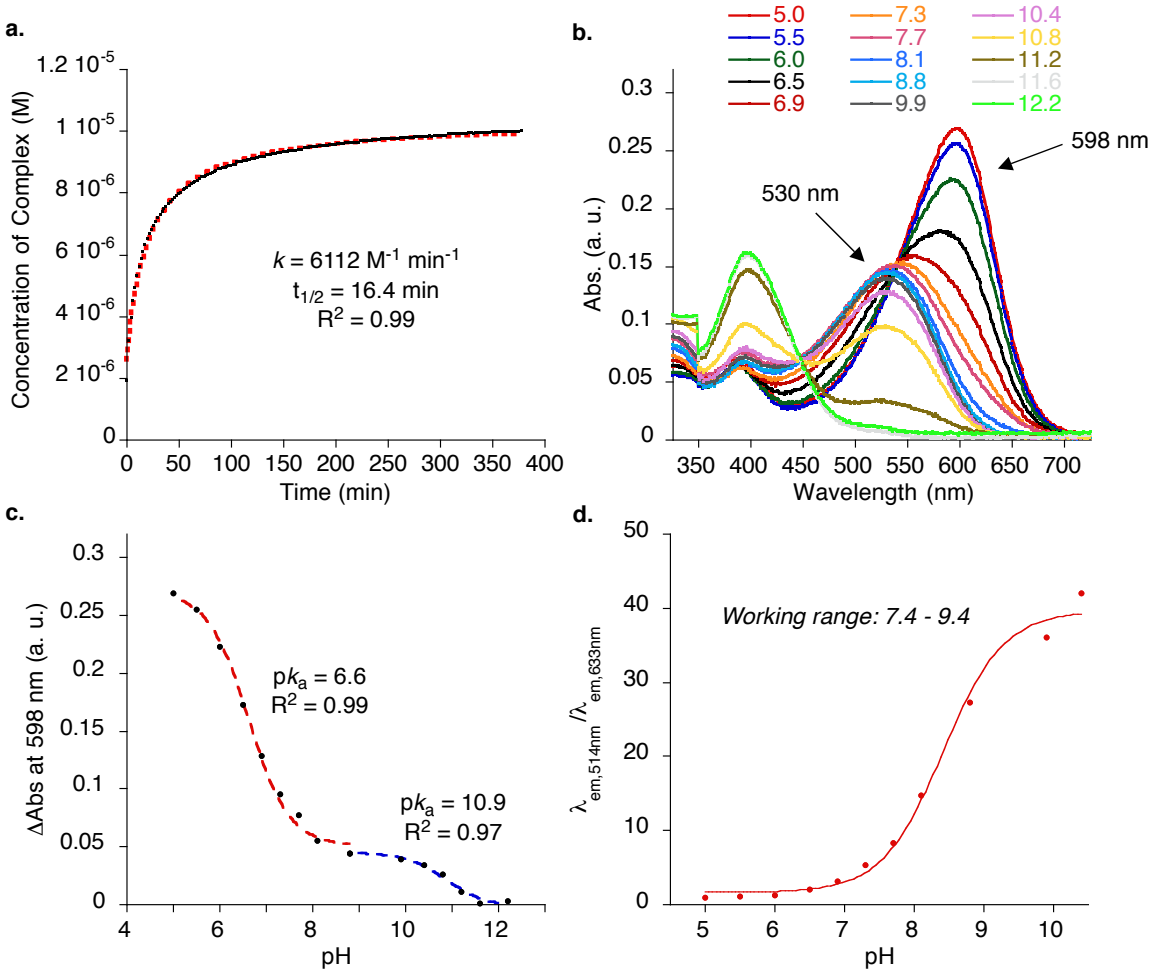


Figure V-7. *In vitro* characterization of Q108K:K40L:T51V:T53S:R58H:Y19W:L117E/**ThioFluor** as a single protein ratiometric pH sensor. **a**. Formation of PSB, monitored at 539 nm via UV-Vis, fit to second order rate kinetics. **b**. UV-Vis titration of the hCRBPII/**ThioFluor** complex. **c**. Double pK_a curve obtained by plotting the absorbance intensity at 598 nm as a function of pH. **d**. Standard curve generated by plotting the emission ratio of 514/633 nm versus pH.

Ms. Zahra Assar obtained the crystal structure of Q108K:K40L:T51V:T53S:R58H:Y19W:L117E with **ThioFluor** at 1.1 Å resolution. The crystal structures show similar trajectories for the chromophore and the active site lysine, location of L117E and all surrounding residues. Even histidine at residue 58 adopts the same rotamer as R58W. Assuming similar binding affinities of **ThioFluor** to the hCRBPII mutants, it is only the substitution at R58 that affects the rate of iminium formation. Thus, it is hypothesized that the less sterically encumbering histidine accelerates the reaction. It is also possible that the histidine hydrogen bonds with the *N,N*-dimethyl moiety of the chromophore leading to a partially positively charged nitrogen. Less electron donation from the *N,N*-dimethyl moiety would lead to a more electrophilic aldehyde.

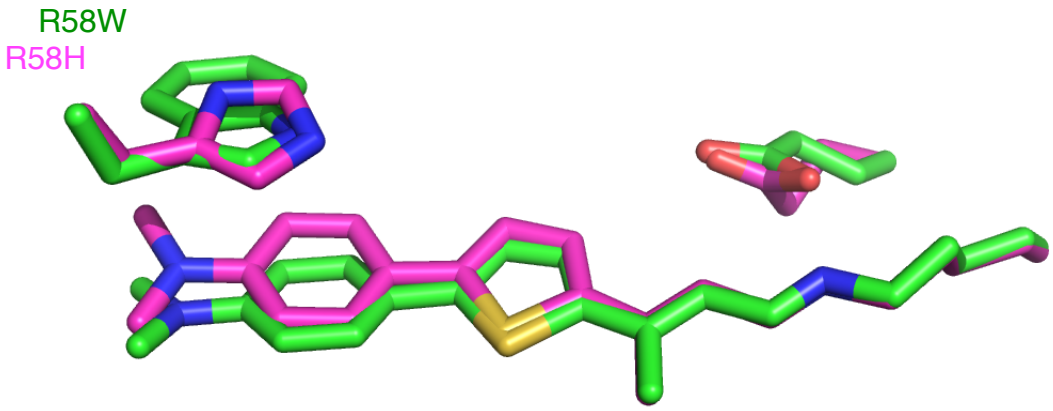


Figure V-8. Overlay of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** (green) and Q108K:K40L:T51V:T53:R58H:Y19W:L117E/**ThioFluor** (magenta), highlighting R58 and L117E.

Besides fast iminium formation, it would also be ideal to develop a bright hCRBPII/**ThioFluor** complex. The quantum yields obtained with L117E are sufficient, so attempts to increase brightness were undertaken by increasing the complex's extinction coefficient. In wavelength regulation studies, it was

observed that A33W led to an increase in extinction coefficient, presumably by encapsulation of the binding cavity. To this end, A33W was added to the parent hCRBPII heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117E (**Table V-2**, entry 4).

Interestingly the extinction coefficient at the carboxylate state of L117E is ~15,000 $M^{-1}cm^{-1}$, which is similar to that observed for Q108K:K40L:T51V:T53S:R58W:Y19W:L117E. However the extinction coefficient of the iminium with the protonated L117E residue increases by two-fold to ~40,000 $M^{-1}cm^{-1}$ (**Figure V-10a**). At the same time, the quantum yield is not significantly changed; this leads to a two-fold increase in brightness. The crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:A33W was crystallized with **ThioFluor** by Mr. Alireza Ghanbarpour with 1.3 Å resolution. It appears that the inclusion of A33W results in a *cis* iminium (which interacts with Q4 through a water molecule), with the concomitant translation of the chromophore deeper into the binding pocket (**Figure V-9**).

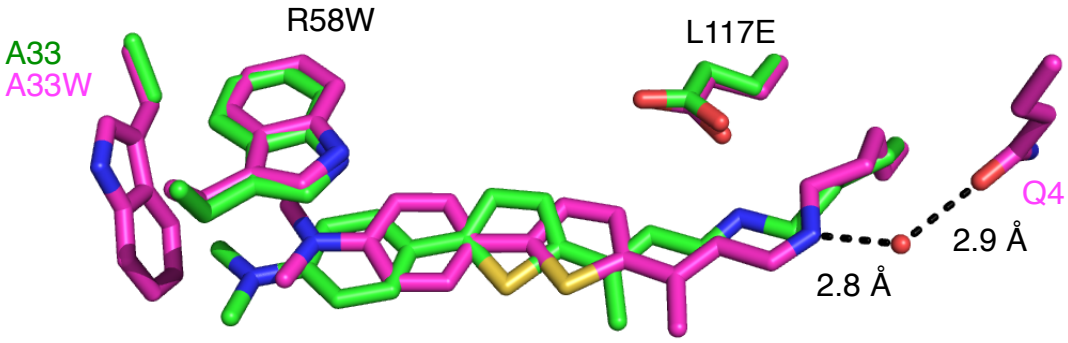


Figure V-9. Overlay of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** (green) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:A33W/**ThioFluor** (magenta).

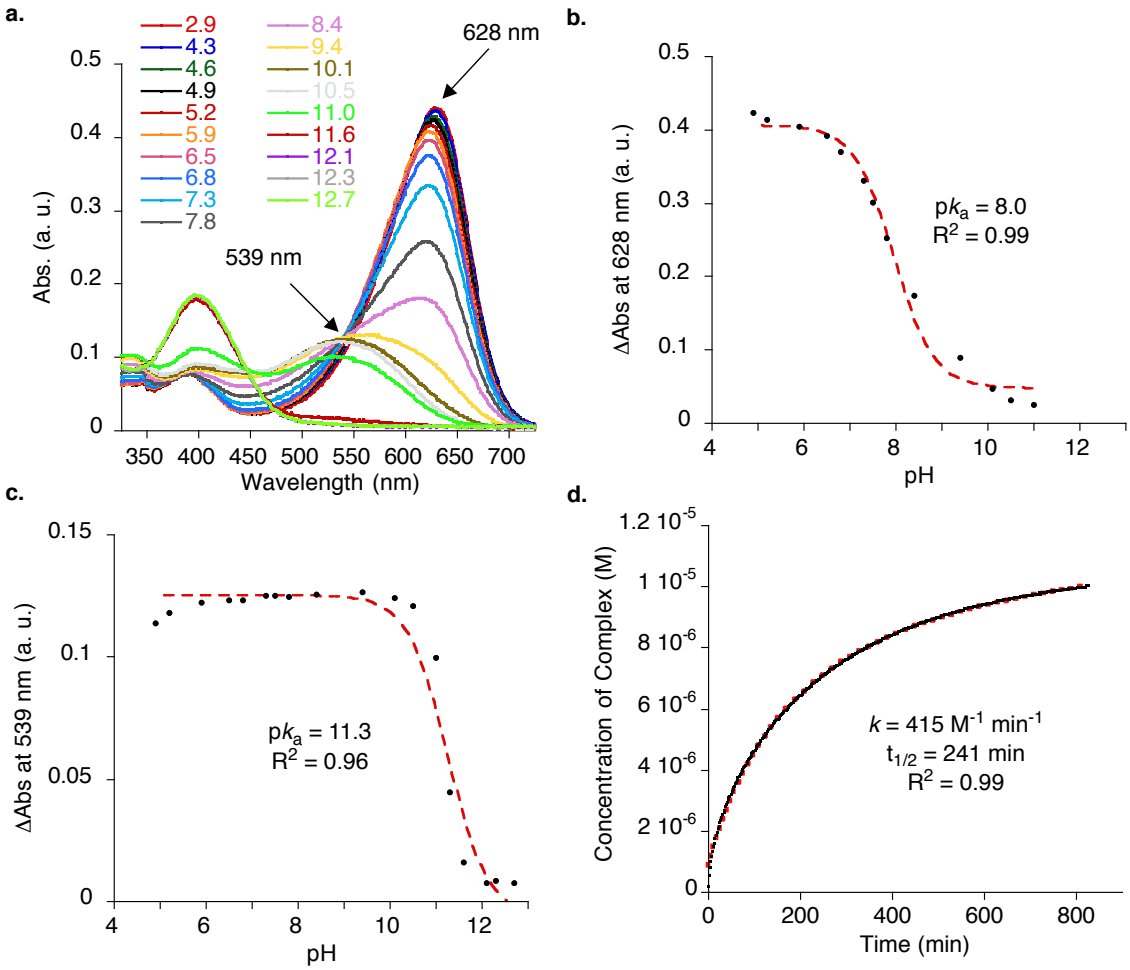


Figure V-10. *In vitro* characterization of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:A33W/**ThioFluor** as a single protein ratiometric pH sensor **a**. UV-Vis titration of the hCRBPII/**ThioFluor** complex. **b**. pK_a curve obtained by plotting the absorbance intensity at 628 nm as a function of pH. **c**. pK_a curve obtained by plotting the absorbance intensity at 539 nm as a function of pH. **d**. Formation of PSB, monitored at 628 nm via UV-Vis, fit to second order rate kinetics.

As shown in the overlay of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E (green) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:A33W (magenta), L117E does not change position. As discussed previously, L117E interacts with the *trans* imine nitrogen (2.9 Å) in Q108K:K40L:T51V:T53S:R58W:Y19W:L117E. However, in the case of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:A33W, it is quite apparent that L117E does not face the imine nitrogen atom when it is in

the *cis* orientation. It is hypothesized that this is why the pK_a of L117E is higher when A33W is present.

Fortuitously, a much larger change is also observed between the two iminium maxima. Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:A33W displays a 89 nm shift (**Figure V-10a**), as compared to 56 nm for Q108K:K40L:T51V:T53S:R58W:Y19W:L117E. Additionally, a high pK_a of 11.3 is achieved (**Figure V-10c**). A pK_a of 8.0 is observed for the glutamic acid side chain at position 117 (**Figure V-10b**); this is much higher than would be expected for a carboxylic acid moiety. While brightness is higher and the iminium maxima are more separated, this mutant does experience a much slower rate of iminium formation. For this reason, this hCRBPII mutant was not further explored for used as a single fluorescent protein ratiometric sensor.

We next made the mutant Q108K:K40L:T51V:T53S:R58H: Y19W:L117E:A33W (**Table V-2**, entry 5). The previously discussed Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:A33W mutant was ideal in terms of the large separation between the two iminium maxima. However, the rate of iminium formation was slow. The R58H mutation was added in order to increase the reaction rate. To our delight, the half-life of iminium formation was decreased from 241 minutes to 30 minutes (**Figure V-11a**), with the R58H substitution, comparable to that observed in Q108K:K40L:T51V:T53S:R58H:Y19W:L117E. However, the separation between the iminium maxima decreased to 67 nm (**Figure V-11b**). From these data, it is obvious that

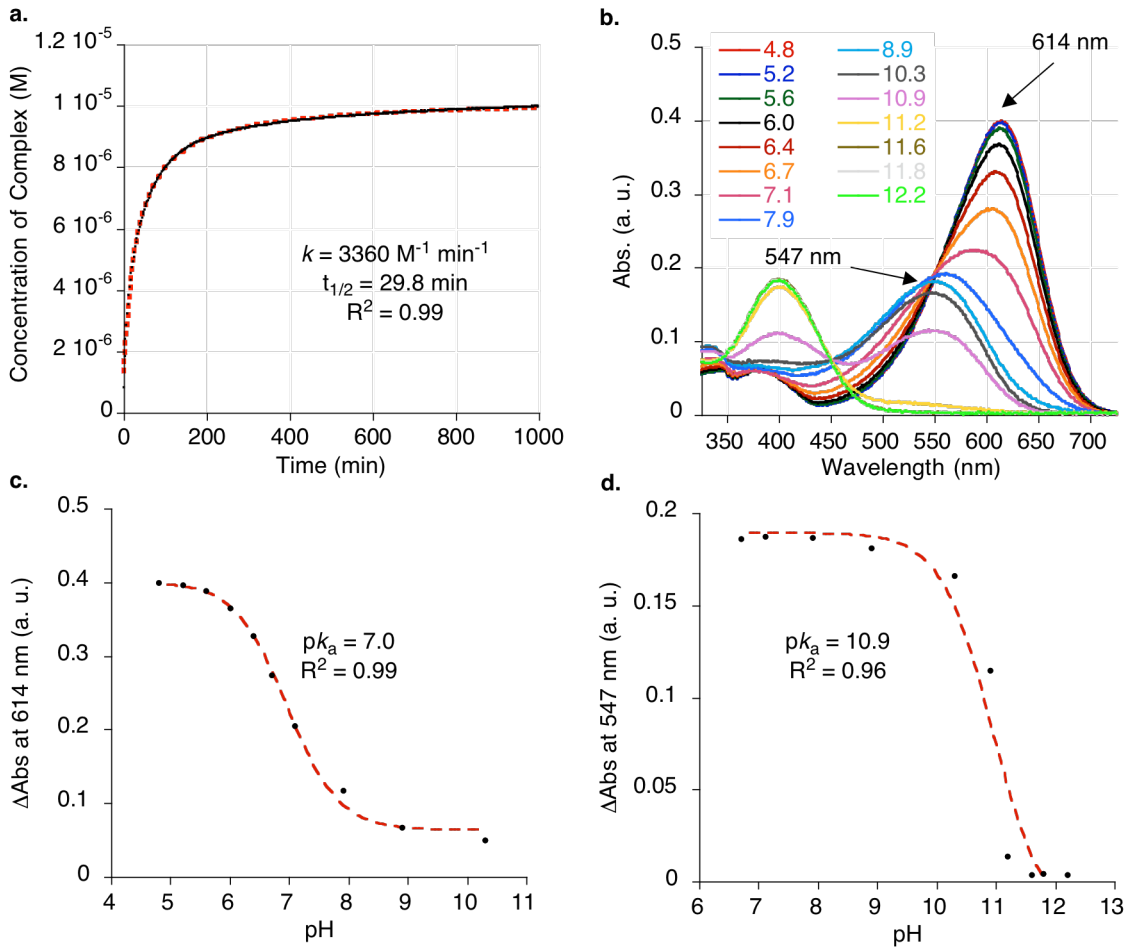


Figure V-11. *In vitro* characterization of Q108K:K40L:T51V:T53S:R58H:Y19W:L117E:A33W/**ThioFluor** as a single protein ratiometric pH sensor. **a**. Formation of PSB, monitored at 585 nm via UV-Vis, fit to second order rate kinetics. **b**. UV-Vis titration of the hCRBPII/**ThioFluor** complex. **c**. pK_a curve obtained by plotting the absorbance intensity at 614 nm as a function of pH. **d**. pK_a curve obtained by plotting the absorbance intensity at 547 nm as a function of pH.

Q108K:K40L:T51V:T53S:R58H:Y19W:L117E:A33W behaves similar to that of Q108K:K40L:T51V:T53S:R58H:Y19W:L117E when bound to **ThioFluor** and that A33W does not lead to a larger separation between the absorbance maxima in all cases. It should also be noted that the pK_a of the L117E side chain is similar to the mutant without A33W (**Figure V-11c** and **Figure V-11d**), but the extinction

coefficients at both iminium states are advantageously as high as Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:A33W (**Figure V-11b**).

V.3 L117D mutation leads to a much larger change in the iminium maxima

We next sought to compare the effect of L117D versus L117E as the titratable acid. The comparison was made in three different hCRBPII series. The rate of iminium formation was significantly increased with aspartate at residue 117, as compared to glutamate (**Table V-3**). Interestingly, L117D led to a more hypsochromic shift in absorption wavelength of the iminium at basic pH, as compared to L117E.

Table V-3. Effect of L117D versus L117E in the development toward a single fluorescent protein ratiometric sensor.

hCRBPII mutant[a]	λ_{absa}[b]	λ_{abs}[c]	ΔAbs	pK_a[d]	pK_a[e]	Φ[f]	$t_{1/2}$ (min)[g]	E.R.[h]
KLVSWW:L117E	558	614	56	11.2	6.6	0.31	82	7
KLVSWW:L117D	525	618	93	10.9	6.7	0.30	14	21
KLVSWW:L117E:A33W	539	628	89	11.3	8.0	0.28	241	-
KLVSWW:L117D:A33W	510	645	135	11.2	7.6	0.16	20	17
KLVSWW:L117E:Q4F	565	619	54	11.5	6.3	0.25	262	-
KLVSWW:L117D:Q4F	529	623	94	11.0	5.7	0.31	8	16

[a]KLVSWW = Q108K:K40L:T51V:T53S:R58W:Y19W. [b]λ_{max} of absorbance at basic pH. [c]λ_{max} of absorbance at acidic pH. [d]pK_a corresponding to the iminium. [e]pK_a corresponding to Glu/Asp117. [f]Quantum yield measured at pH ~7.2. [g]Kinetic measurements were performed at 23 °C with 20 μM protein and 0.5 equiv ligand. PSB formation was monitored by UV-Vis at λ_{max} for each complex over time at pH ~7.2. Data was then fit to second order kinetics. [h]ER = the highest emission ratio collected as a result of excitation at 514 nm and 633 nm.

Titration of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:A33W leads to the largest change in the two absorption maxima, 135 nm. However, as was the

case with L117E in this mutant, the working pH range of the complex is too high, rendering the probe not suitable for use at physiological pH.

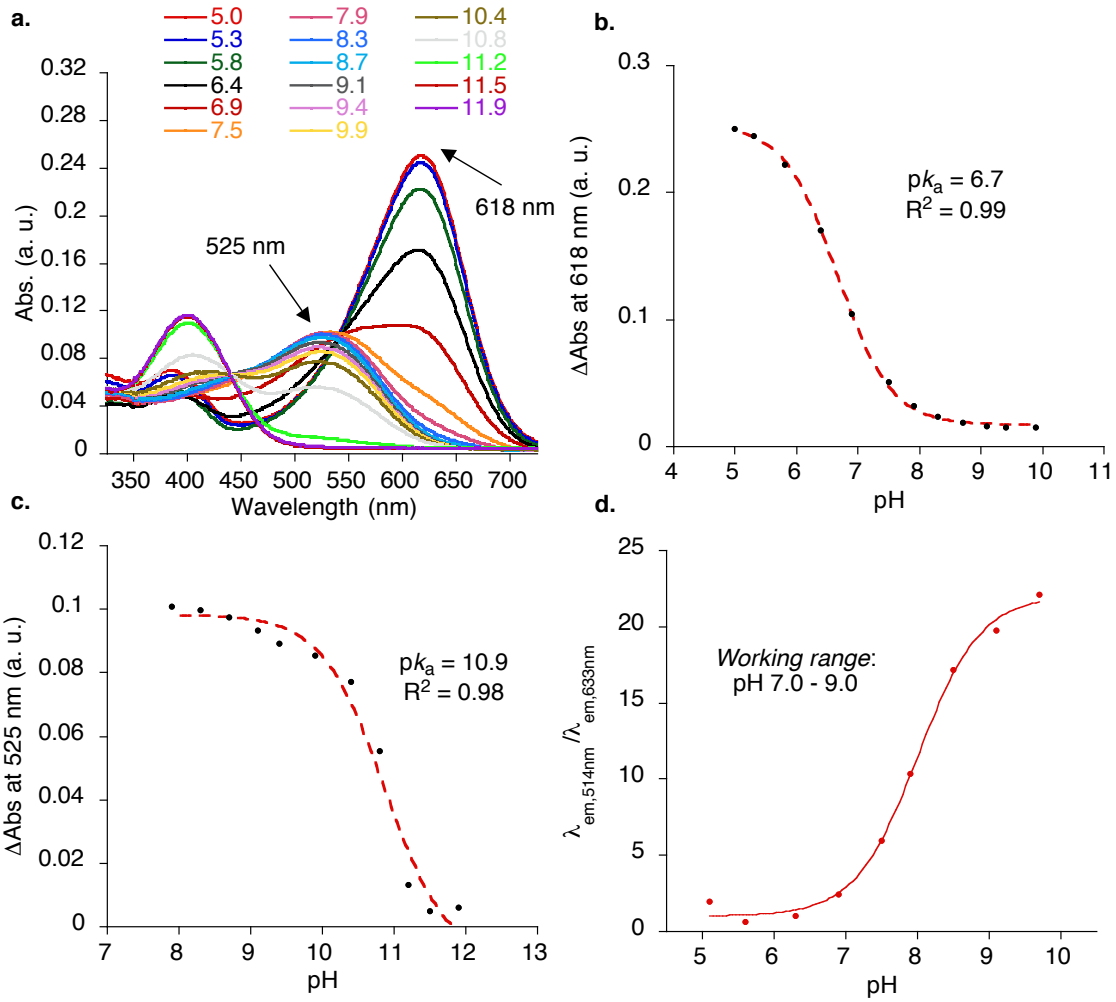


Figure V-12. *In vitro* characterization of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor** as a single protein ratiometric pH sensor. **a**. UV-Vis titration of the hCRBPII/**ThioFluor** complex. **b**. pK_a curve obtained by plotting the absorbance intensity at 618 nm as a function of pH. **c**. pK_a curve obtained by plotting the absorbance intensity at 525 nm as a function of pH. **d**. Standard curve generated by plotting the emission ratio of 514/633 nm versus pH.

The heptamutant Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor** iminium displayed a change of 93 nm upon titration of the aspartic acid side chain at residue 117. Characterization of the complex is shown in

Figure V-12. Similar to that previously observed for mutants containing L117E, the titration of L117D shows an effect on absorption wavelength of the iminium. The aspartic acid side chain exhibits a pKa of 6.7 (**Figure V-12b**), rendering an absorption maximum of 525 nm at pH 7.5. Acidification of the complex yields a bathochromic shift in absorption maximum to 618 nm (**Figure V-12a**). To our delight, generation of a standard curve based on excitation wavelengths of 514 nm and 633 nm yields a high emission ratio of approximately 20 (**Figure V-12d**), which is more than two-fold higher than for Q108K:K40L:T51V:T53S:R58W:Y19W:L117E (**Figure V-6**). Unfortunately, the working range (pH 7.0 – 9.0) is outside physiological pH. However, this probe would be useful to measure the pH of more basic organelles, such as the mitochondria (pH ~ 8).[5]

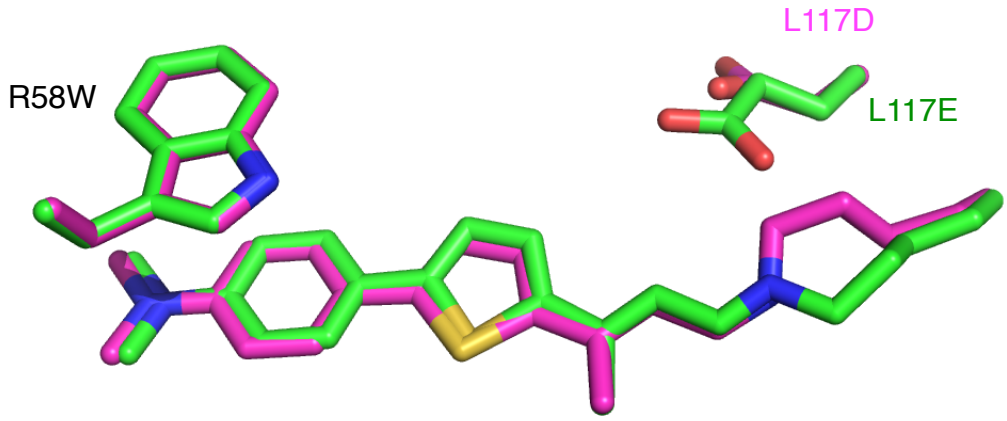


Figure V-13. Overlay of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** (green) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor** (magenta).

It is interesting to note that the absorption maxima of the iminium at low pH is similar for Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor** and Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**. However, a difference

of 33 nm is observed at high pH. According to the theory that blue-shifted absorption wavelengths occur from localization of charge, it has to be concluded that aspartate more efficiently localizes charge on the iminium as compared to glutamate. Fortuitously, Mr. Alireza Ghanbarpour provided the crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor** at 1.0 Å resolution. An overlay of the crystal structures Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor** is shown in **Figure V-13**. No significant differences are apparent when comparing the crystal structures of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor**, except for the isomeric state of the iminium. In the case of L117E, the *trans* iminium is stabilized by the glutamate side chain. Conversely, the shorter L117D mutation does not provide this interaction (it interacts with Q4 through an acetate molecule).

However, it is quite important to note that these structures were crystallized at low pH 4.0 – 4.5. From the UV-Vis, we can see that the absorption maxima of these two complexes are not significantly different at low pH. It is hypothesized that an acid induced isomerization occurs upon acidification when L117D is present. In the case of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/ **ThioFluor**, it is proposed that at low pH a *trans* iminium is present, such that the carboxylate of L117D can interact with the imine to localize charge. However, upon acidification, a *cis* iminium forms, in which L117D does not interact with the

iminium. This would lead to the observed bathochromic shift in absorption wavelength with acidification. This posited mechanism may also explain the difference in extinction coefficients of the iminium at low and high pH. It is possible that the *cis* or *trans* iminium leads to a higher extinction. However, evidence for these hypotheses has not been obtained.

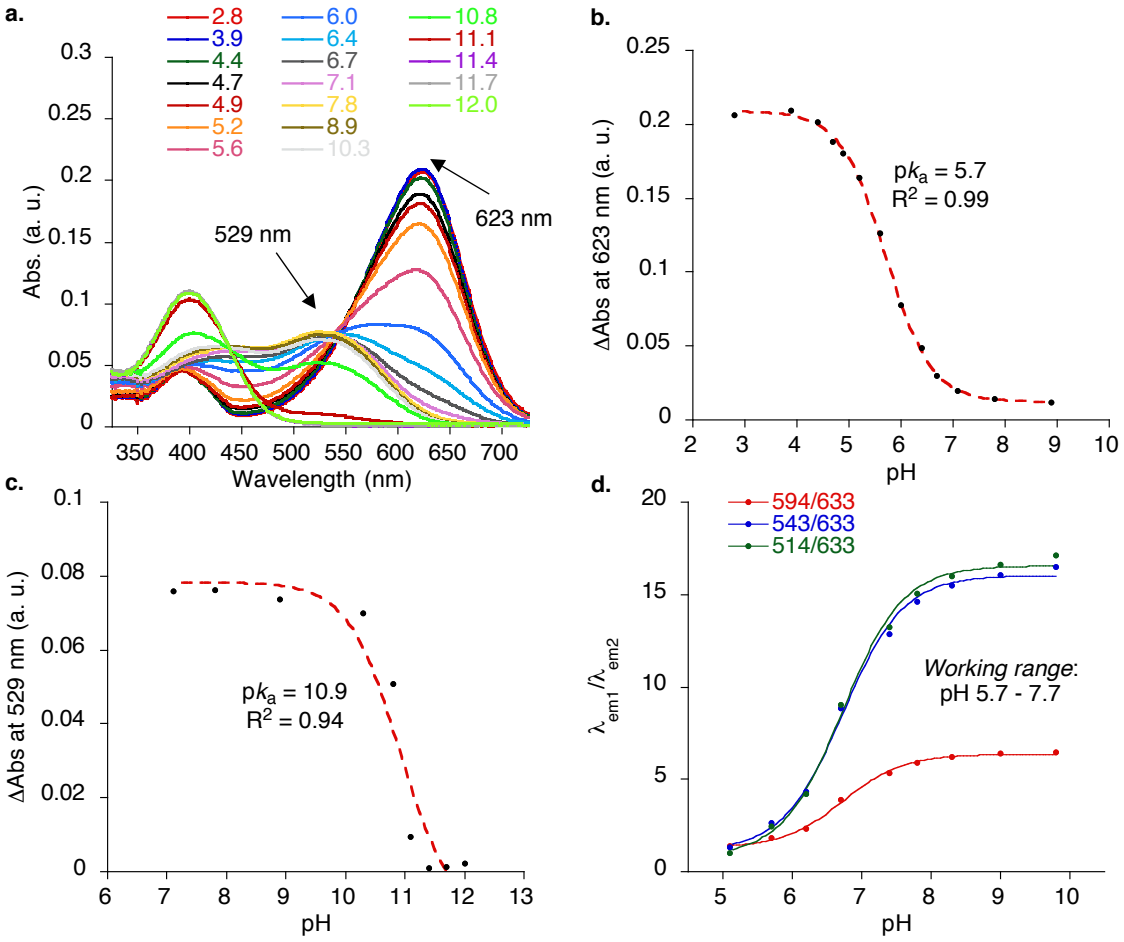


Figure V-14. *In vitro* characterization of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioFluor** as a single protein ratiometric pH sensor. **a**. UV-Vis titration of the hCRBPII/**ThioFluor** complex. **b**. pK_a curve obtained by plotting the absorbance intensity at 623 nm as a function of pH. **c**. pK_a curve obtained by plotting the absorbance intensity at 539 nm as a function of pH. **d**. Standard curves generated by plotting the emission ratio of 514/633 nm, 543/633 nm and 594/633 nm versus pH.

The working range of the Q108K:K40L:T51V:T53S:R58W:Y19W:L117D was too high, so we chose to introduce a Q4F mutation, in hopes that the pK_a of the aspartate side chain would decrease. To our delight, the addition of Q4F to Q108K:K40L:T51V:T53S:R58W:Y19W:L117D results in one unit decrease in the pK_a of the aspartic acid side chain (from 6.7 to 5.7), while having no effect on any of the other spectroscopic properties (**Figure V-14**). Standard curves plotted at different wavelengths are shown in **Figure V-14d**. Excitation wavelengths of 514/633 nm and 543/633 nm show high emission ratios. From the standard curve generated with excitation wavelengths of 514 nm and 633 nm, this hCRBPII/**ThioFluor** pair is shown to be useful at physiological pH with a working range of 5.7 to 7.7.

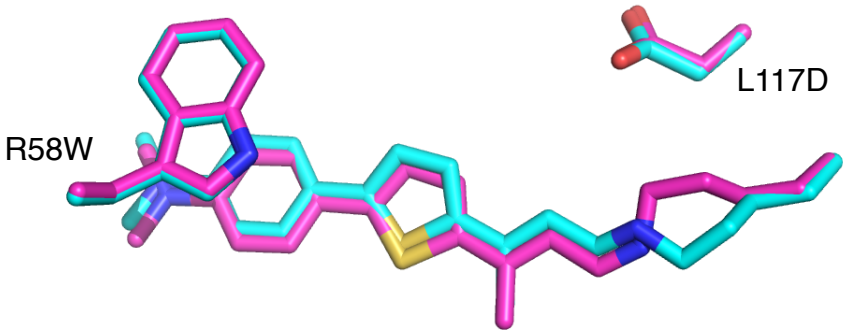


Figure V-15. Overlay of the crystal structures Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor** (magenta) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioFluor** (cyan).

Mr. Alireza Ghanbarpour was successful in obtaining the crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F with **ThioFluor** at 1.6 Å resolution. An overlay of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioFluor** and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D/**ThioFluor** indicates that with the

addition of Q4F a *trans* iminium is observed (**Figure V-15**). Neighboring residues, including L117D and R58W do not move. It is the closer distance between L117D and the iminium hydrogen atom in the *trans* isomer that presumably leads to the decreased pK_a.

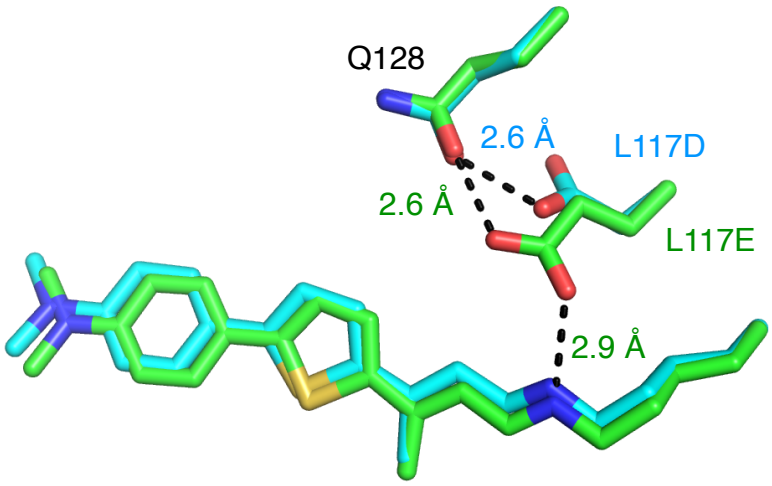


Figure V-16. Overlay of crystal structures of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** (green) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioFluor** (cyan).

It is most interesting to compare the structures of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioFluor** with Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor** (**Figure V-16**), as the chromophore and active site lysine adopt the same orientation, both displaying a *trans* iminium. The major difference between the structures is that L117E interacts with both the iminium and Q128, with distances of 2.9 Å and 2.6 Å, respectively. However, L117D, which is one methylene shorter, only forms a hydrogen bond with Q128 (2.6 Å). Upon titration of L117E in Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**, a shift of 56 nm is observed between the two iminium maxima. A larger shift of 94 nm is observed in

Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioFluor**. The respective absorbance maxima of the red-shifted iminiums are 614 nm and 623 nm. The insignificant change in absorption maxima is presumably due to the unchanging electrostatic potential along the chromophore when L117D and L117E are protonated; in other words, the polarity is essentially the same.

These crystals were obtained at pH 4.0 – 4.5, which yields no information about the state of the iminium with the carboxylate at position 117 (at more basic pH). Crystallization at a more basic pH, in which the imine is protonated and acidic side chain is deprotonated may provide insight into the reason why L117D mutants are more blue-shifted than with L117E, as well as the reason for the significant increase in extinction coefficient observed upon protonation of the acidic side chain.

Lastly, final efforts were made to optimize Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F. We chose to test the effect of an R58H substitution, as it was already shown to enhance the rate of iminium formation. It should be noted that we first tried the mutants Q108K:K40L:T51V:T53S:R58H:Y19W:L117D and Q108K:K40L:T51V:T53S:R58H:Y19W:L117D:A33W, but no soluble protein was obtained from the expressions. Fortunately, expression of Q108K:K40L:T51V:T53S:R58H:Y19W:L117D:Q4F was fruitful with an expression yield of 21 mg/mL. In this hCRBPII mutant, the titration of L117D in absorption is hard to observe, due to its extremely low absorbance intensity. Starting from pH

7.2, the Q108K:K40L:T51V:T53S:R58H:Y19W:L117D:Q4F/**ThioFluor** complex was first acidified (**Figure V-17a**). Upon acidification, the SB at ~400 nm decreases, while a peak at 617 nm forms. The pK_a for this species is 5.8. Presumably, this corresponds to the pK_a of the aspartic acid. Additionally, it appears that the small peak at ~500 nm decreases with acidification.

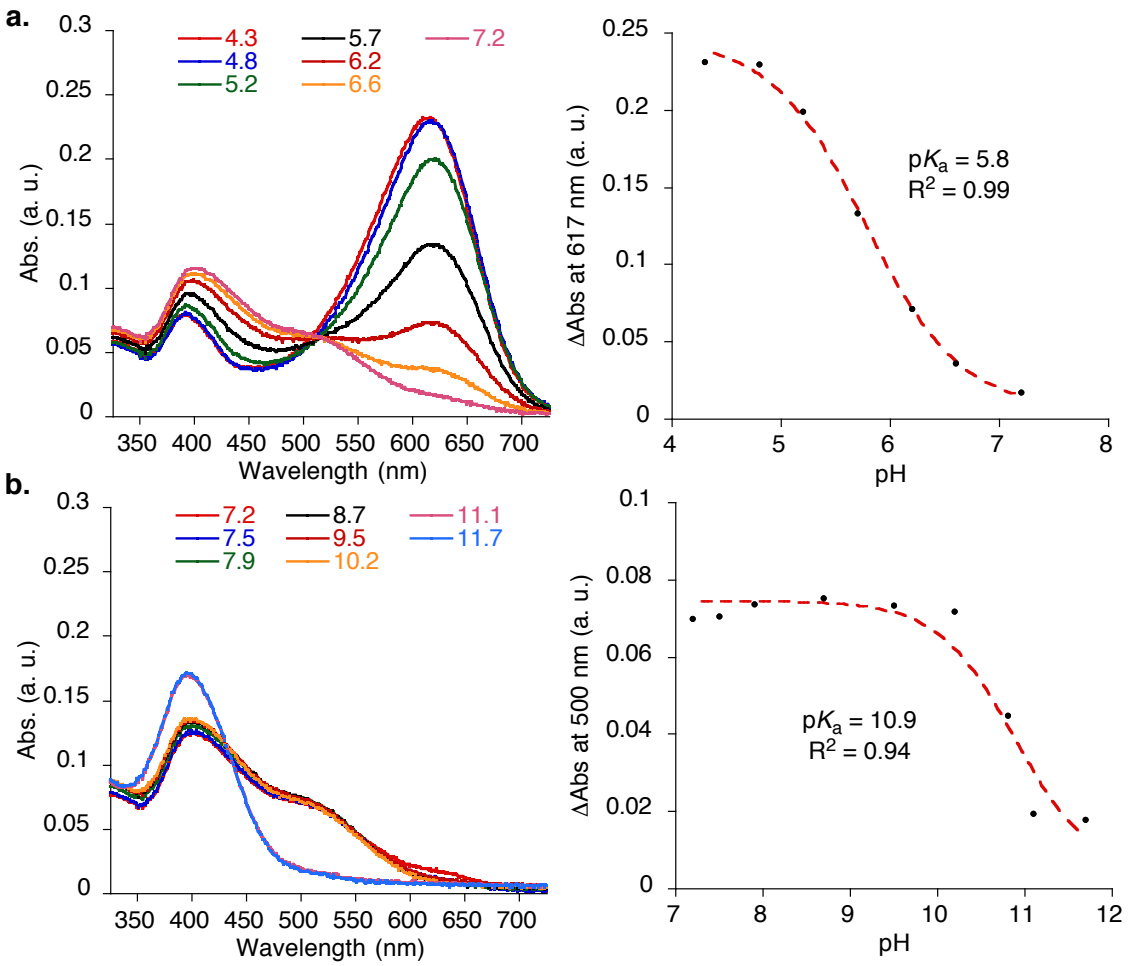


Figure V-17. **a**. Acidification of Q108K:K40L:T51V:T53S:R58H:Y19W:L117D:Q4F/**ThioFluor** complex. **b**. Basification of Q108K:K40L:T51V:T53S:R58H:Y19W:L117D:Q4F/**ThioFluor** complex.

On the other hand, basification of the Q108K:K40L:T51V:T53S:R58H:Y19W:L117D:Q4F/**ThioFluor** complex from pH

7.2 (**Figure V-17b**) leads to a decrease in the peak at ~500 nm and an increase in the SB at ~400 nm. It would stand to reason that this is the peak corresponding to the iminium when the aspartic acid is deprotonated (pK_a = 10.9). However, it is predicted that a mixture of imine isomers is present since the SB can be converted to both the 617 nm and ~500 nm peak. It should also be noted that the blue-shifted iminium is almost unobservable due to its low absorbance intensity, presumably due to its extremely low extinction coefficient, which is going to limit its use as a single protein ratiometric pH sensor.

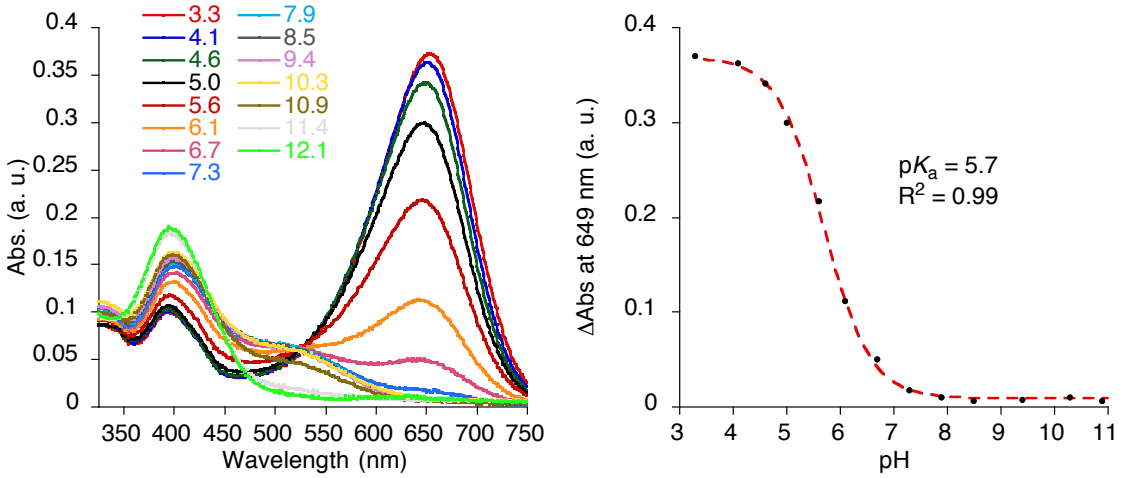


Figure V-18. Titration of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:A33W:Q4F/**ThioFluor** complex and pK_a obtained by plotting absorbance intensity versus pH at 649 nm.

The last hCRBPII mutant investigated was Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:A33W:Q4F because the mutant with A33W lead to the largest shift in absorption maxima upon titration of the aspartic acid side chain, but the pK_a led to a working range too high for use at physiological pH. Q4F was included to lower the pK_a of the side chain. This hCRBPII mutants behaves similar to the previous mutant; a blue-shifted iminium

at ~500 nm exists at pH 7.3, while a red-shifted peak with absorption maximum of 649 nm (pK_a = 5.7) is formed upon acidification (**Figure V-18**). The almost unobservable blue-shifted iminium renders this hCRBPII/**ThioFluor** complex also inadequate for application as a single protein ratiometric sensor.

V.4 Imaging of hCRBPII mutants displaying the pH induced wavelength shift with ThioFluor

We chose to test two hCRBPII mutants for their applicability as a single fluorescent protein ratiometric sensor. Q108K:K40L:T51V:T53S:R58W: Y19W:L117E was first chosen as **ThioFluor** has also shown success in labeling. It also shows an absorption wavelength shift with pH (56 nm) and an emission ratio of 7 when excitation wavelengths of 514 nm and 633 nm are used. HeLa cells expressing EGFP-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-3NLS were incubated with 10 μM **ThioFluor** for one hour at 37 °C. Cells were then washed twice with PBS buffer and subsequently imaged (**Figure V-19**).

Green fluorescence was observed when the cells excited at 488 nm (BP 505 – 530 nm), indicating that transfection was successful and the fusion protein is being expressed in the nucleus. Imaging wavelengths of 514 nm, 543 nm, 594 nm and 633 nm were tested for their ability to fluoresce the hCRBPII/**ThioFluor** complex. The gain and offset settings were optimized for the 594 nm excitation with a long pass (LP) 615 nm emission filter. As seen, there is minimal background fluorescence under these confocal settings. However, significant background fluorescence is observed at the lower excitation wavelengths of 514

nm and 543 nm, with both LP 615 nm and 650 nm emission filters. This is presumably due to nonspecific iminium formation or free aldehyde. Unfortunately, background fluorescence is also observed at the most red-shifted excitation wavelength of 633 nm. It only makes sense that this comes from two-photon excitation (~315 nm) of nonspecific imine formation.

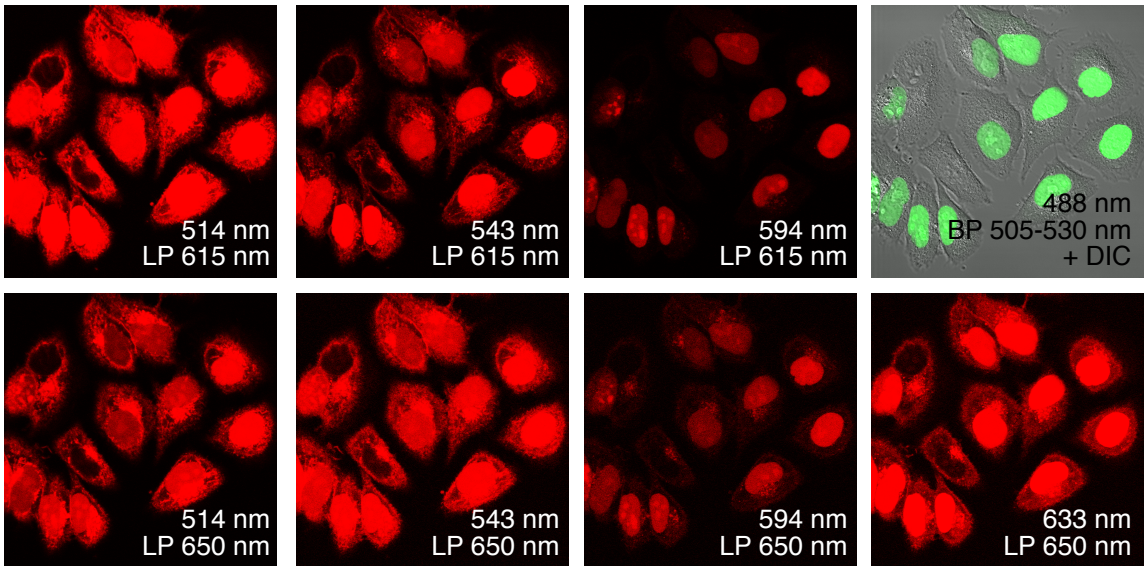


Figure V-19. Imaging of HeLa cells expressing EGFP-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-3NLS, labeled with **ThioFluor** for one hour at 37 °C.

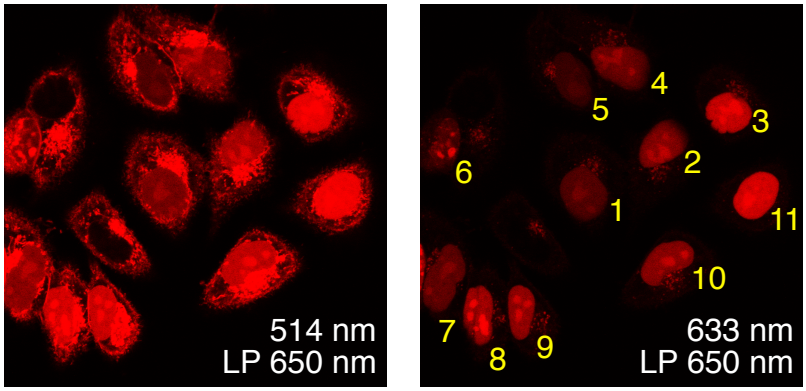


Figure V-20. Optimized imaging of HeLa cells expressing EGFP-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-3NLS upon excitation at 514 nm and 633 nm.

Nonetheless, it should be noted that background fluorescence observed upon excitation with 633 nm could be minimized upon optimization of the gain and offset settings in this channel (**Figure V-20**). Under the same settings, the background fluorescence with 514 nm is also drastically reduced.

In the optimized 633 nm excitation channel, the nucleus of each cell was selected (cells are numbered in **Figure V-20**). From the confocal software a histogram of fluorescence intensity was then output. The sum of the area under the curve was then computed for each individual cell. In order to ensure the exact same area is being analyzed when comparing the excitation of 514 nm and 633 nm, the same nucleus selections were superimposed on the imaged obtained from 514 nm. Again, the area under the curve of the generated histogram was determined. Lastly, the ratio of fluorescence intensity at 514 nm to 633 nm was determined for each cell. It should be noted that two cells did not provide reliable data, as their fluorescence intensities were outside of the threshold limits (cells 3 and 11 are too bright). This data is summarized in **Table V-4**.

Table V-4. Ratio of emissions obtained upon excitation at 514 nm and 633 nm for EGFP-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-3NLS/**ThioFluor**.

Cell	$\lambda_{em,514nm}/\lambda_{em,633nm}$	Cell	$\lambda_{em,514nm}/\lambda_{em,633nm}$
1	1.60	7	1.60
2	1.37	8	1.38
4	1.38	9	1.39
5	1.61	10	1.42
6	1.65		

Note: Cells 3 and 11 omitted were removed because their fluorescence intensities were outside of the threshold limits.

Based on the data collected from nine cells, the average emission ratio of 514 nm to 633 nm is 1.49 with a standard deviation of 0.12. The *in vitro* standard curve (**Figure V-6**) was previously obtained based on plotting the ratio of emissions at 514 nm and 633 nm versus pH. To make this plot useful, we fit a linear line to the linear portion of the standard curve (6.5 – 8.5). This plot is shown in **Figure V-21**. Substitution of the observed 1.49 ± 0.12 emission ratio average over the nine cells yields an *in vivo* pH of 6.65 ± 0.043 for the nucleus of these HeLa cells.

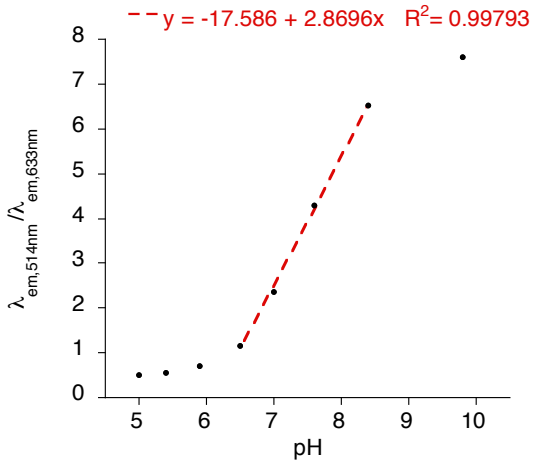


Figure V-21. Linear fit to the working range of the *in vitro* generated standard curve for Q108K:K40L:T51V:T53S:R58W:Y19W:L117E/**ThioFluor**.

Intracellular pH is a fundamental modulator of cell function. Abnormal pH values are associated with some common diseases such as cancer and Alzheimer's. Thus, the ability to measure intracellular pH is needed in order to study the physiological processes of organelles. To this end, research has focused on the development of intracellular pH sensors based on small fluorescent organic molecules, nanoparticles and fluorescent proteins.[6] More specifically, green fluorescent protein (GFP) variants were used to measure

cytosolic pH of HeLa cells (7.4).[7, 8] More recently, nano-pH probes were developed to measure the pH levels of individual cells. Pourmand and coworkers measured single-cell intracellular pH measurements using non-cancerous and cancerous cell lines, including human fibroblasts, HeLa, MDA-MB-231 and MCF-7. The average intracellular pH levels were 7.4, 6.8, 6.9 and 6.9 for human fibroblast, HeLa, MCF-7 and MDA-MB-231, respectively.[9] While most intracellular pH sensors do not differentiate between nuclear and cytoplasmic pH, small molecules have been targeted for this purpose. U2OS cells expressing HaloTag-3NLS showed a pH of 7.4, while cytosol targeting indicated a pH of 7.1.[10] From these data it is apparent that there is a gap in the field of intracellular pH measurement to which we can contribute.

As shown, *in vivo* pH can be measured in HeLa cells expressing EGFP-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-3NLS, labeled with **ThioFluor**. The only disadvantage with this mutant is that high background fluorescence is observed, presumably due to unbound chromophore or nonspecific imine formation. Additionally, the pH is on the edge of the working range for this mutant (6.5 – 8.5). To this end, we chose to test the utility of the fastest binding mutant developed in this study. Advantageously, it also has a lower working range of 5.7 – 7.7, which should allow for more accurate estimation of the *in vivo* pH.

Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F was cloned into the pFlag-CMV2 vector, retaining the nuclear localization sequence. It should be noted that EGFP was not retained, in order to eliminate any crosstalk that may

arise from undesired EGFP excitation at 514 nm or 543 nm (even though the emission should be filtered with both LP 615 nm and LP 650 nm emission filters).

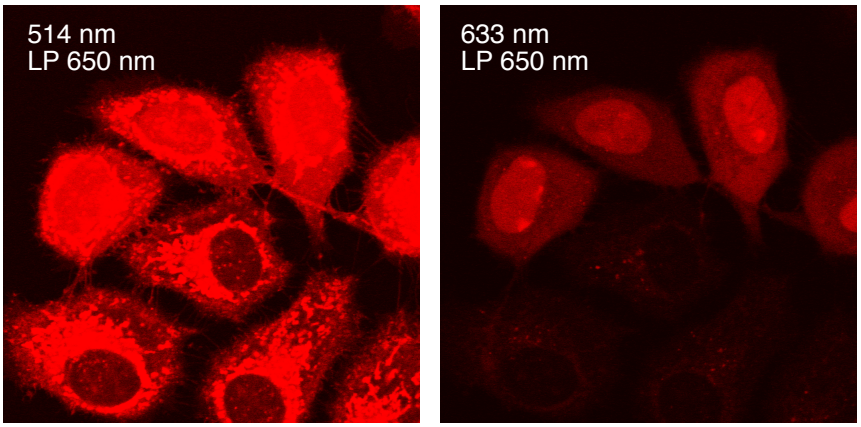


Figure V-22. Imaging of HeLa cells expressing Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F-3NLS, labeled with **ThioFluor** for ten minutes at 37 °C.

HeLa cells expressing Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F-3NLS were incubated with 10 μM **ThioFluor** for ten minutes at 37 °C. Images were collected with excitation wavelengths of 514 nm and 633 nm, with LP 650 nm emission filter. The gain and offset settings were optimized for the 633 nm excitation. Background fluorescence is negligible under excitation with 633 nm (**Figure V-22**), but unfortunately significant background fluorescence is again observed upon 514 nm excitation. The background fluorescence is even brighter than the fluorescence from hCRBPII/**ThioFluor** in all cells. As compared to HeLa cells expressing Q108K:K40L:T51V:T53S:R58W:Y19W:L117E, the background fluorescence is brighter at 514 nm; in vitro data shows that the extinction coefficient is decreased by about half (~8,000 $M^{-1}cm^{-1}$ for Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F and ~15,000 $M^{-1}cm^{-1}$ for

Q108K:K40L:T51V:T53S:R58W:Y19W:L117E) which leads to a reduction in brightness.

It should be noted that expression of this hCRBPII mutant leads to fluorescence from the hCRBPII/**ThioFluor** complex that seems to be 'leaky' as it is not fully localized to the nucleus. It is hypothesized that this is due to overexpression of the mutant *in vivo* (*in vitro* protein expression yield is 23 mg/mL). If this mutant it to be used for live cell imaging, it will be necessary to manipulate the *in vivo* protein expression, possibly by imaging earlier after transfection.

V.5 ThioFluor analogues also show a pH induced wavelength shift

While most of the **ThioFluor** analogues developed were not fluorescent (**Chapter III**), we sought to investigate whether they also showed a pH dependent wavelength shift of the iminium due to titration of the glutamic acid side chain in Q108K:K40L:T51V:T53S:R58W:Y19W:L117E. Results are summarized in **Table V-5**.

As expected, all of the chromophores show a bathochromic shift in wavelength upon putative protonation of the glutamic acid side chain, except **ThioFluor-2**, which was already shown not to be able to regulate wavelength. Small shifts are observed with weakly electron donating groups at the para position of the phenyl ring (**ThioFluor-6** = H, **ThioFluor-4** = methoxy). These chromophores also show a lower pK_a of the glutamic acid side chain. On the other hand, the more conjugated ThioFluor analogues (**ThioFluor**, **ThioFluor-3**,

ThioFluor-8) showed a larger difference between the iminium maxima and also an increased pK_a of the glutamic acid side chain at residue 117. Presumably the donation along the chromophore leads to the high pK_a in the other chromophores.

Table V-5. pH induced wavelength shift observed in Q108K:K40L:T51V:T53S:R58W:Y19W:L117E with **ThioFluor** analogues.

Ligand	λ_{absa}[b]	λ_{abs}[b]	ΔAbs	pK_a[d]	pK_a[e]	Φ[f]	$t_{1/2}$ (min)[g]
ThioFluor	558	614	56	11.2	6.6	0.31	82
ThioFluor-2	515	-	-	11.3	-	0.13	1124
ThioFluor-3	548	611	63	10.9	6.6	0.27	27
ThioFluor-4	466	493	27	10.9	5.7	-	17
ThioFluor-5	423	nd	nd	10.9	nd	-	9
ThioFluor-6	424	458	34	10.6	5.4	-	3
ThioFluor-8	641	687	46	11.2	6.8	0.09	352
ThioFluor-9	562	624	62	11.1	6.6	0.02	13
ThioFluor-10	489	545	56	11.3	6.1	0.06	92

[a]KLVSWW = Q108K:K40L:T51V:T53S:R58W:Y19W. [b]λ_{max} of absorbance at basic pH. [c]λ_{max} of absorbance at acidic pH. [d]pK_a corresponding to the iminium. [e]pK_a corresponding to Glu117/Asp117. [f]Quantum yield measured at pH ~7.2. [g]Kinetic measurements were performed at 23 °C with 20 μM protein and 0.5 equiv ligand. PSB formation was monitored by UV-Vis at λ_{max} for each complex over time at pH ~7.2. Data was fit to second order kinetics.

V.6 Conclusions

We have shown that the titration of an L117D or L117E mutation in hCRBPII mutants leads to a pH induced absorption wavelength shift of the iminium. Shifts up to 135 nm were observed, with quantum yields greater than 30%, and half-lives as fast as 8 minutes. Advantageously, the hCRBPII/**ThioFluor** system can successfully be used to image live cells. To our

delight, based on the ratio of emissions at two excitation wavelengths, the average over nine cells yields an *in vivo* pH of 6.65 ± 0.043 for the nucleus of HeLa cells.

The only pitfall is that background fluorescence is observed. To address this correctly, it is suggested that a fluorogenic analogue of **ThioFluor** be developed, such that it is only fluorescent when bound as the iminium to hCRBPII. It would also be beneficial to test the affect of **ThioFluor** concentration. A decrease in **ThioFluor** concentration would be expected to decrease the background fluorescence, if it is indeed due to unbound aldehyde.

REFERENCES

REFERENCES

1. Berbasova, T., Nosrati, M., Vasileiou, C., Wang, W.J., Sing, K., Lee, S., Yapici, I., Geiger, J.H. & Borhan, B. Rational Design of a Colorimetric pH Sensor from a Soluble Retinoic Acid Chaperone. *Journal of the American Chemical Society* **135**, 16111-16119 (2013).

2. Berbasova, T. in PhD Thesis, Michigan State University (2014).

3. Wang, W. in PhD Thesis, Michigan State University (2012).

4. 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).

5. Casey, J.R., Grinstein, S. & Orlowski, J. Sensors and regulators of intracellular pH. *Nature Reviews - Molecular Cell Biology* **11**, 50-61 (2010).

6. Han, J.Y. & Burgess, K. Fluorescent Indicators for Intracellular pH. *Chemical Reviews* **110**, 2709-2728 (2010).

7. Llopis, J., McCaffery, J.M., Miyawaki, A., Farquhar, M.G. & Tsien, R.Y. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6803-6808 (1998).

8. Kneen, M., Farinas, J., Li, Y.X. & Verkman, A.S. Green fluorescent protein as a noninvasive intracellular pH indicator. *Biophysical Journal* **74**, 1591-1599 (1998).

9. Ozel, R.E., Lohith, A., Mak, W.H. & Pourmand, N. Single-cell intracellular nano-pH probes. *RSC Advances* **5**, 52436-52443 (2015).

10. Benink, H.A., McDougall, M.G., Klaubert, D.H. & Los, G.V. Direct pH measurements by using subcellular targeting of 5(and 6-) carboxyseminaphthorhodafluor in mammalian cells. *Biotechniques* **47**, 769-774 (2009).

CHAPTER VI: PROMOTING MONOMERIZATION OF hCRBPII MUTANTS

Our lab has observed the formation of domain-swapped dimers during protein bacterial overexpression. Characterization of domain-swapped dimers from hCRBPII-Y60 mutants has been provided by both size-exclusion chromatography and crystallography. It was hypothesized that the combination of two open monomer intermediates leads to the formation of the domain-swapped dimer (**Figure VI-1**).[1]

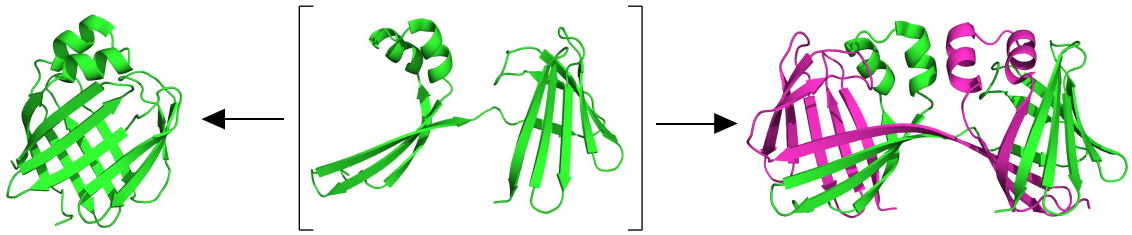


Figure VI-1. Formation of domain-swapped dimer in hCRBPII-Q108K:K40L:Y60W. Coordinates obtained from PDB 4ZR2 (dimer) and 4ZJ0 (monomer).

In the course of engineering hCRBPII to bind ThioFluor, it was observed that certain residues are more prone to foster hCRBPII dimerization. For example, T51 was shown to form a significant amount of dimer when mutated to alanine (**Table VI-1**), albeit in the presence of other mutations.

Table VI-1. Effect of T51 mutation in KL:**T51X**:T53S:R58W:Y19W:F16Y.

hCRBPII mutant	Major Fraction
Q10K:K40L:**T51V**:T53S:R58W:Y19W:F16Y	Monomer
Q10K:K40L:**T51C**:T53S:R58W:Y19W:F16Y	Monomer
Q10K:K40L:**T51A**:T53S:R58W:Y19W:F16Y	Dimer

Other mutations shown to affect the monomer/dimer ratio formed during bacterial expression were K40, T51, T53, V62, and R58, among others. A few

select examples are shown in **Table V1-2**. It should be noted that Q108K:K40L produces only monomer during expression. Single point mutations on wild type hCRBPII were also shown to promote dimerization, including I25E (27.5 mg/L monomer, 16.9 mg/L dimer; 77 mol% monomer), Y60F (40.6 mg/L monomer, 2.9 mg/L dimer; 97 mol% monomer), L77D (31.3 mg/L monomer, 19.3 mg/L dimer; 76 mol% monomer), L77E (45.7 mg/L monomer, 7.6 mg/L dimer; 92 mol% monomer), and Q128L (6.0 mg/L monomer, 5.6 mg/L dimer; 68 mol% monomer).

Table VI-2. hCRBPII dimerization promoted by T53S, V62N, T51A and T51N mutations.

hCRBPII mutant	mol % monomer	mg/L monomer	mg/L dimer
Q108K:K40E	31	3.3	14.6
Q108K:K40D	33	2.5	10.1
Q108K:K40D:Q4F	> 95	7.0	-
Q108K:K40L:T53A	63	5.5	6.6
Q108K:K40L:T53S	85	11.4	3.9
Q108K:K40L:T53V	84	10.6	4.0
Q108K:K40L:V62N	76	14.5	9.4
Q108K:K40L:T53S:V62N	69	8.0	7.1
Q108K:K40L:T53S:V62E	34	5.5	21.6
Q108K:K40L:T51A:T53S:V62N	< 5	-	5.8
Q108K:K40L:T51N:T53S:V62N	22	1.7	12.4

We previously showed that hCRBPII variants prone to oligomerization are not good candidates for fluorescent labeling in live cells (see **Chapter II.11.1** for data and discussion). For this reason, we wanted to determine if we could rationally introduce point mutations to a sequence of interest, in order to promote the formation of hCRBPII monomer over dimer. We chose to start by exploring

the affect of T51 on the formation of domain-swapped dimers, as this position was shown to drastically alter the ratio of monomer and dimer formed during bacterial protein overexpression (**Table VI-1**). Because we have also observed a difference between hCRBPII monomer and dimer binding to **ThioFluor** (**Chapter II.8.1**), we also sought to investigate the spectroscopic properties of these mutants when bound to retinal.

VI.1 Effect of T51 mutations on the monomer/dimer ratio during bacterial overexpression in hCRBPII mutants

In order to study the effect of T51 mutations (**Figure VI-2**) on the formation of dimer-swapped dimers in hCRBPII, we started from the parent template Q108K:K40L. This was chosen as the template for mutation for two reasons. Firstly, the inclusion of K40L was shown by Dr. Wenjing Wang to increase the iminium pK_a due to the removal of basic charge.[2] Additionally, we wanted to prevent the binding of retinal at the secondary K40 site. Dr. Meisam Nosrati previously showed occupancy of retinal at K40 in the dimer Q108K:T51D.[3]

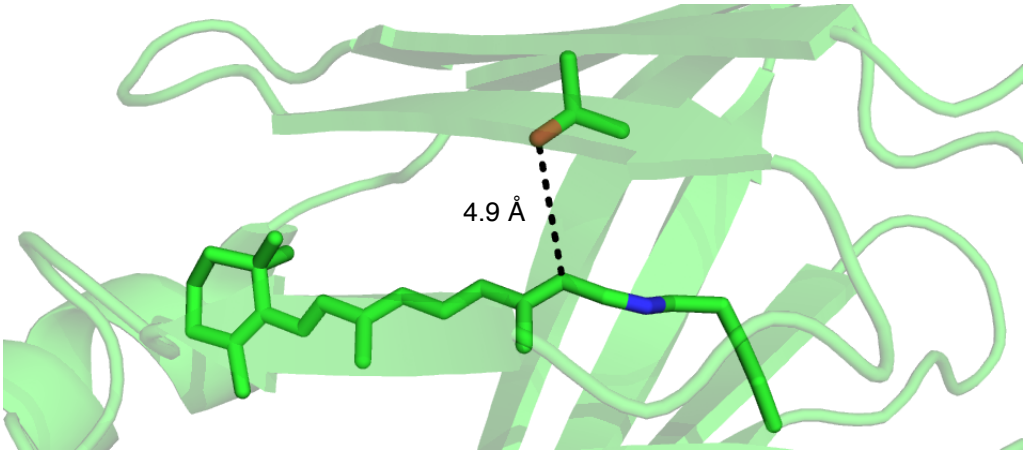


Figure VI-2. Location of T51 in Q108K:K40L/retinal. Coordinates were obtained from PDB 4EXZ.

To this end, all twenty amino acids were introduced at residue 51 in Q108K:K40L. Fortunately, all substitutions except T51D, T51G, T51P and T51R led to soluble protein during expression at room temperature in terrific broth (TB). Isolated protein yields are shown in **Table VI-3**, in increasing order of monomer. If only one oligomeric state of hCRBPII was isolated after size exclusion chromatography, mol percent monomer is expressed as 99% upon isolation of monomer, or 1% upon isolation of dimer.

Table VI-3. Effect of **T51X** mutations on monomer/dimer formation in Q108K:K40L.

hCRBPII mutant	mol % monomer	mg/L monomer	mg/L dimer
Q108K:K40L:T51K	1	-	6.1
Q108K:K40L:T51A	14	1.1	13.0
Q108K:K40L:T51N	17	2.7	25.6
Q108K:K40L:T51W	25	1.0	5.9
Q108K:K40L:T51S	28	1.7	8.7
Q108K:K40L:T51H	58	4.6	6.8
Q108K:K40L:T51Q	65	10.2	11.2
Q108K:K40L:T51Y	86	7.9	2.6
Q108K:K40L:T51F	90	10.2	2.3
Q108K:K40L:T51M	92	15.1	2.5
Q108K:K40L:T51E	96	44.7	3.8
Q108K:K40L:T51L	99	57.6	0.7
Q108K:K40L:T51C	99	9.5	-
Q108K:K40L:T51I	99	12.0	-
Q108K:K40L:T51V	99	12.5	-
Q108K:K40L	99	13.3	-

It was quickly realized that the mol percent monomer ratio cannot be correlated to any of the commonly used hydrophobicity scales, including Kyte-Doolittle,[4] Hoops-Woods,[5] Cornette,[6] Eisenberg,[7] Rose,[8] Janin[9] or Engelman.[10] Additionally, there also is no correlation between residue size and the monomer/dimer ratio obtained during expression.

Non polar side chains, including leucine, isoleucine and valine lead to monomer formation. However, alanine leads to majorly dimer formation. On the other hand, no general trend is observed for the polar side chains. Asparagine, glutamine and serine lead to substantial dimer formation, while threonine and cysteine lead to solely monomer formation. The aromatic residues of phenylalanine, tyrosine and tryptophan all produced both monomer and dimer, albeit at different monomer/dimer ratios. Charged side chains again presented contrasting results. Mutation to the basic lysine and histidine produced a large population of dimer. In fact T51K yields dimer only. Conversely, glutamic acid produces mostly monomer.

VI.2 Spectroscopic properties of Q108K:K40L:T51X mutants

Purified monomers and dimers were coupled with retinal, in order to see if any differences in spectroscopic properties, including absorption wavelength and half-life of binding, were different in the monomer and dimer of the same mutation (**Table VI-4**). All rates were measured via UV-Vis at 23 °C at approximately pH 7.2 (20 μM protein and 10 μM retinal, plotted according to pseudo-first order rate kinetics). It should be noted that half-lives less than two minutes should not be

seen as significantly different, as the initial rates may not be accurate due to the fast reaction rate. The most interesting mutants are highlighted, due to the different spectroscopic properties observed between the dimer and monomer.

Table VI-4. Spectroscopic properties of Q108K:K40L:**T51X** mutants bound to retinal.

hCRBPII mutant	λ_{abs} monomer	λ_{abs} dimer	$t_{1/2}$ (min) monomer	$t_{1/2}$ (min) dimer
Q108K:K40L:T51A	519	520	1.9	2.0
Q108K:K40L:T51C	489	-	5.3	-
Q108K:K40L:T51E	500	461	nd	nd
Q108K:K40L:T51F	529	538	108.3	6.3
Q108K:K40L:T51H	509	512	1.5	1.5
Q108K:K40L:T51I	544	-	12.9	-
Q108K:K40L:T51K	-	494	-	nd
Q108K:K40L:T51L	535	494/517	60.9	1.7
Q108K:K40L:T51M	536	nd	18.7	nd
Q108K:K40L:T51N	482/493	489/494	1.1	0.9
Q108K:K40L:T51Q	nd	nd	nd	nd
Q108K:K40L:T51S	503	503	1.2	ns
Q108K:K40L	505	-	3.3	-
Q108K:K40L:T51V	535	-	2.1	-
Q108K:K40L:T51W	503	503	78.8	86.6
Q108K:K40L:T51Y	nd	nd	nd	nd

ns = protein is not stable, and precipitates under the experimental conditions very quickly

In general, the rate of iminium formation is fast in the hCRBPII variants. The exceptions are when phenylalanine, tryptophan or leucine are substituted at position 51. These three residues lead to slow iminium formation, presumably due to their larger size. However, in the dimer, only tryptophan leads to slow

iminium formation. It is quite interesting to note that T51F and T51L actually lead to substantially faster iminium formation in the dimer. Attempts to understand the difference between Q108K:K40L:T51F monomer and dimer were undertaken as it produced more dimer during bacterial expression than Q108K:K40L:T51L (**Table VI-3**). This will be discussed in **Section VI.5**.

The other major noticeable difference between retinal bound monomer and dimer is seen in the absorption wavelengths of Q108K:K40L:T51E. The iminium in the monomer is almost 40 nm red-shifted of the dimer. Presumably, the glutamic acid adopts a different orientation in the two structures. However, crystal structures would be necessary to understand the difference. All other monomer and dimer pairs exhibit essentially the same absorption wavelengths and rates of iminium formation.

VI.3 Imine isomerization of Q108K:K40L:T51X mutants

As discussed above, the substitution at residue 51 can have a significant affect on the spectroscopic properties of bound retinal; additionally differences are apparent between the monomer and dimer. These include the absorption wavelength and rate of iminium formation. Because residue 51 was shown to drastically alter the rate of iminium formation, we sought to test whether it also affects the rate of imine isomerization. We again used Q108K:K40L as the parent mutant.

The spectroscopic properties of Q108K:K40L-monomer/retinal are shown in **Figure VI-4**. Iminium formation is fast, exhibiting a half-life of 3.3 minutes. The

highest intensity of iminium is observed at 24 minutes, which then decays into SB (**Figure VI-4a** and **Figure VI-4b**). This is indicative of *cis-trans* isomerization of the imine, which we have previously observed between retinal and both hCRBPII and CRABPII mutants. (**Figure VI-3**).[11, 12] briefly, the low pK_a *trans* imine can be converted to the high pK_a *cis* imine with UV light irradiation, which is subsequently protonated to yield the *cis* iminium. Thermal relaxation eventually produces the thermodynamic *trans* imine.

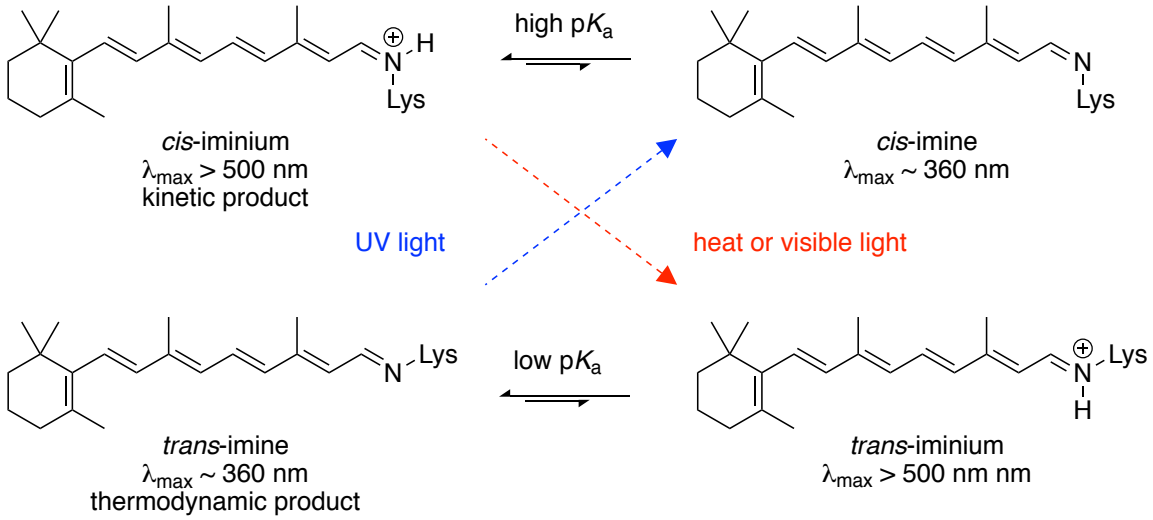


Figure VI-3. Cycle for imine isomerization in retinal bound hCRBPII and CRABPII variants.

The half-life of the thermal isomerization observed in Q108K:K40L-monomer/retinal is 110.0 minutes. Interestingly, upon UV irradiation of this thermodynamic state (365 nm TLC handlamp for one minute), only iminium is formed, evidenced by the lack of SB at approximately 360 nm (**Figure VI-4c**). The rate of isomerization after UV irradiation is almost identical to that observed thermally, with a half-life of 111.6 minutes (**Figure VI-4d**), indicating that the conversion is to the same species.

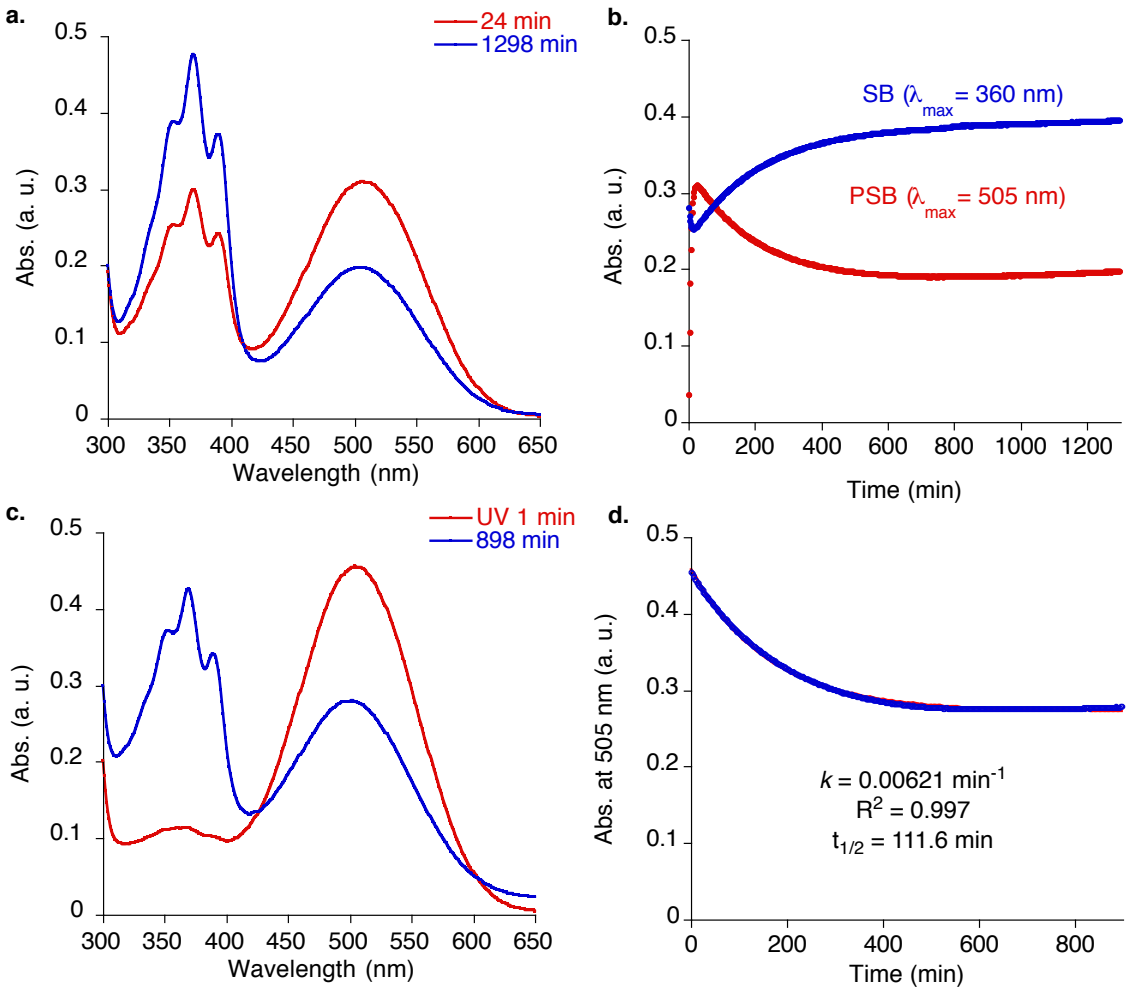


Figure VI-4. **a**. UV-Vis of Q108K:K40L-monomer/retinal. **b**. Absorbance intensity at SB and PSB as a function of time. **c**. UV irradiation and subsequent isomerization of the imine. **d**. Rate (pseudo-first order) of imine isomerization after UV irradiation.

Substitution of valine, alanine and isoleucine also allow imine isomerization to occur. However, the monomers display drastically different half-lives of thermal isomerization (**Table VI-5**). The isomerization of Q108K:K40L:T51I-monomer/retinal was extremely slow and was not complete within 800 minutes. On the other hand, it does not appear that imine isomerization occurs after iminium formation with substitutions of serine or cysteine (in the monomer or dimer). However, in the case of serine, the protein

complex does appear to precipitate after iminium formation. It can be concluded that residue 51 significantly affects the imine isomerization.

Table VI-5. Thermal isomerization half-lives with substitution of small aliphatic residues at T51.

hCRBPII mutant	$t_{1/2}$ (min)	
	Monomer	Dimer
Q108K:K40L	110.0	-
Q108K:K40L:T51A	36.5	36.5
Q108K:K40L:T51V	356.1	-

VI.4 Thermal imine isomerization that leads to a different absorption wavelength of the iminium

From the data we have collected, the absorption of the PSB remains consistent, whether formed immediately with retinal or after UV irradiation. However, here we show that with a T51N mutation, two absorption wavelengths of PSB are observed. Spectroscopic properties of Q108K:K40L:T51N-monomer/retinal are shown in **Figure VI-5**. Iminium formation is fast, exhibiting a half-life of 1.1 minutes and absorption wavelength of 482 nm. The highest intensity of iminium is observed at 10 minutes, which then decays into SB (**Figure VI-5a**). Again, this is indicative of imine isomerization. However, in this case, a bathochromic shift in the wavelength to 493 nm is also observed. This is the first case, in which a shift in the wavelength is observed, presumably of the *cis* and *trans* iminium or different conformations of active site residues. The half-life of this thermal isomerization is 43.5 minutes. From the plot of the intensity at 280 nm (corresponding to the protein absorbance), it is quite obvious that after

the thermodynamic product is formed, the complex is not stable and the protein unfolds over time (**Figure VI-5b**). It does not appear to precipitate, because the baseline at 750 nm does not change significantly.

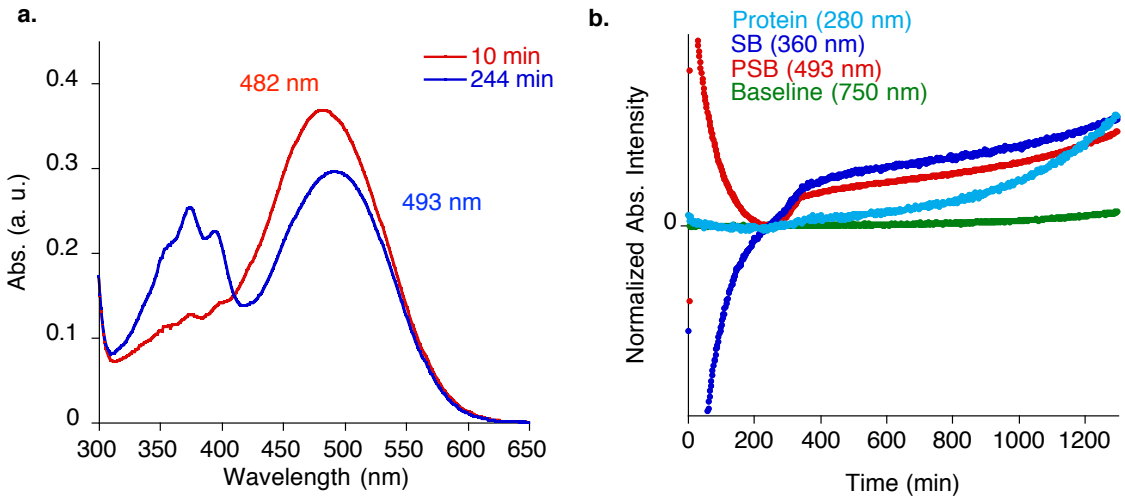


Figure VI-5. **a**. UV-Vis of Q108K:K40L:T51N-monomer/retinal. **b**. Normalized absorption intensity versus time indicates that protein unfolding occurs after iminium formation.

Q108K:K40L:T51N-dimer behaves the same as the monomer when bound to retinal. In this case, the absorption wavelength of the kinetically formed iminium is 489 nm. Subsequent thermal decay then yields an iminium with an absorption maximum of 494 nm. We were not able to obtain the crystal structure of Q108K:K40L:T51N-monomer/retinal, however, Mr. Alireza Ghanbarpour was successful in obtaining the crystal structure of Q108K:K40L:T51N-dimer/retinal. There were 12 molecules in the asymmetric unit, of which sufficient retinal density was observed in three chains (**Figure IV-6**). The sigma level (from Pymol) is given for each chain. An overlay of retinal in the three chains shows that the chromophore is quite variable. This conformational change may account for the shift in absorption wavelength observed over time.

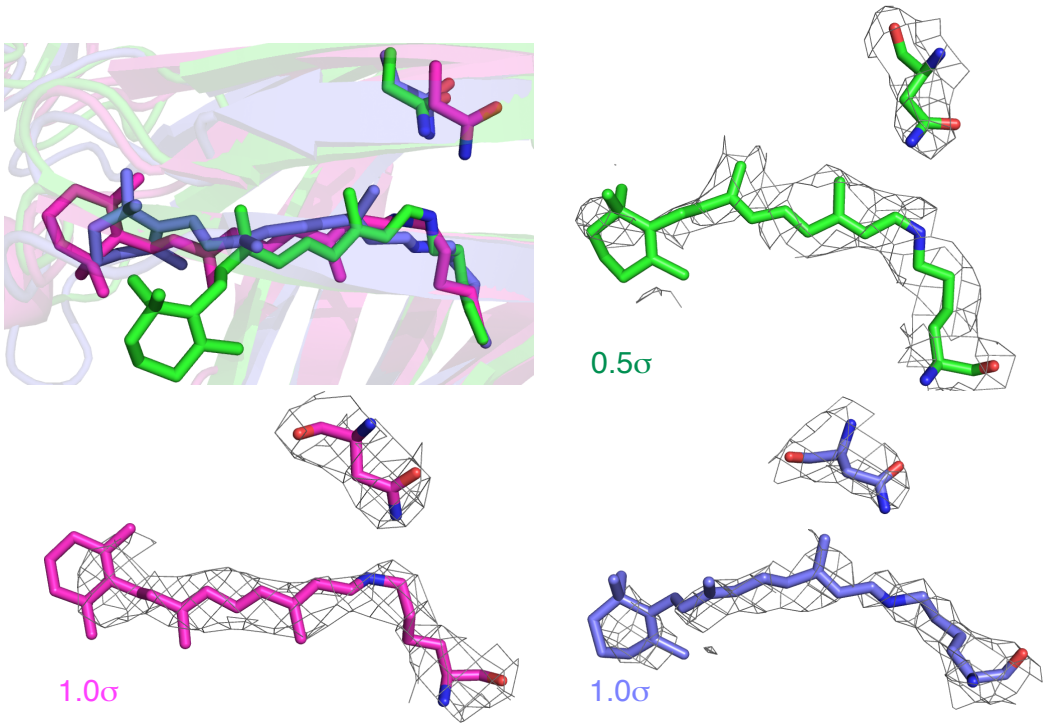


Figure VI-6. Crystal structure of Q108K:K40L:T51N-dimer/retinal.

VI.5 Different rates of iminium formation between monomer and dimer with retinal

As previously mentioned, the rate of iminium formation was drastically different in the monomer and dimer when phenylalanine and leucine were substituted at residue 51 (**Table VI-4**). The monomer of Q108K:K40L:T51L exhibited a half-life of 60.9 minutes, while the dimer was much faster with a half-life of 1.7 minutes. It should also be noted that the initial iminium of Q108K:K40L:T51L-dimer/retinal formed (presumably the kinetic product) had an absorption wavelength of 494 nm, while the thermodynamic iminium showed a maximum at 506 nm. This phenomenon was not observed in the monomer.

Similarly, the half-life of iminium formation of Q108K:K40L:T51F-monomer and dimer were 108.3 minutes and 6.3 minutes respectively. UV-Vis data is

shown in **Figure VI-7**. It is interesting to note that not only is the rate of iminium formation faster in the dimer, but so is the thermal isomerization. Additionally, the pK_as are drastically different between the monomer and dimer, as evidenced by the absorbance intensity of the iminium.

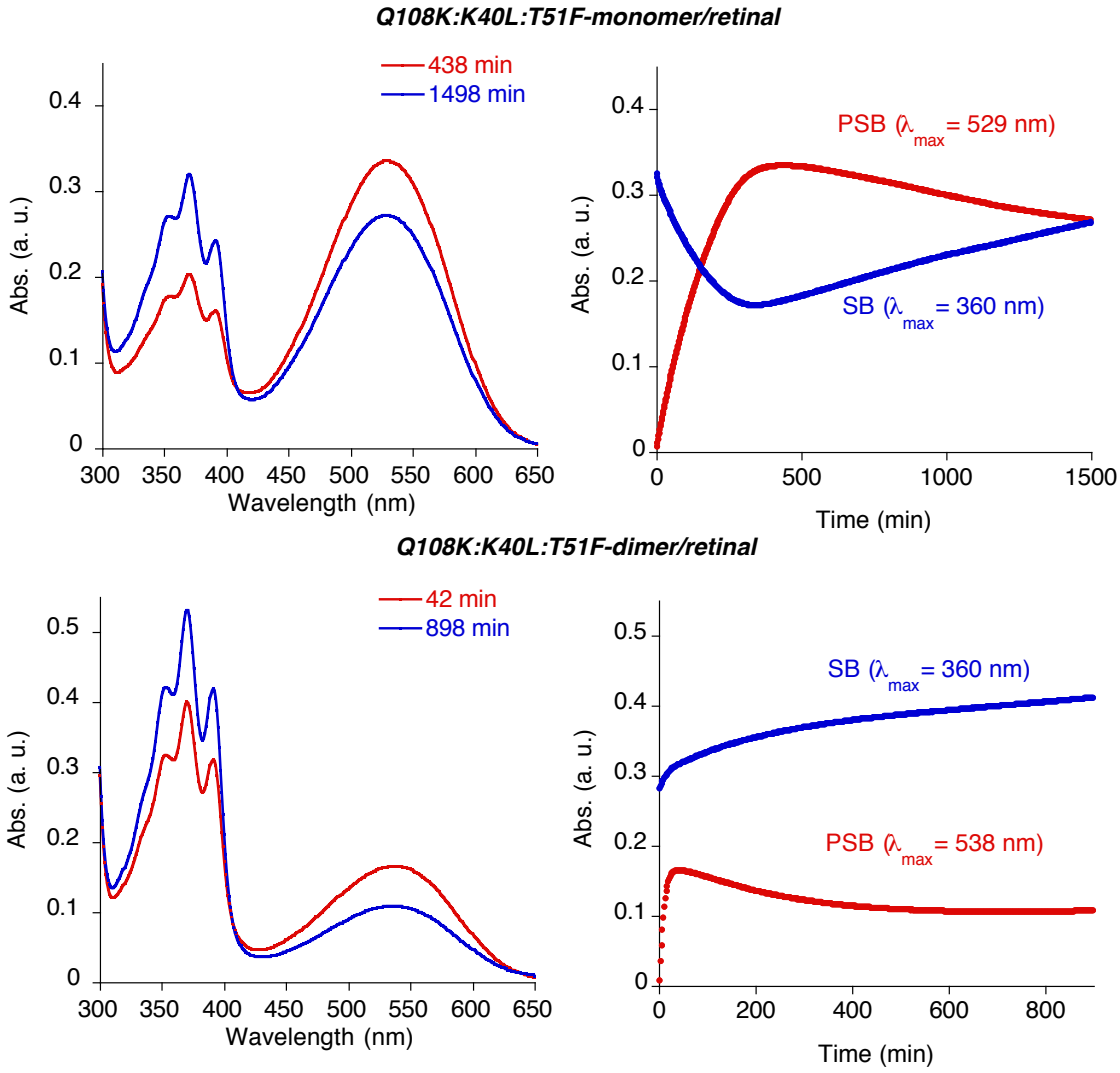


Figure VI-7. UV-Vis of Q108K:K40L:T51F-monomer/retinal and Q108K:K40L:T51F-dimer/retinal.

We wanted to see if structural data could provide insight into these very different rates of iminium formation (though structural data only provides the thermodynamically static picture), as it may yield hints on how to engineer faster

binders for fluorescence imaging. Mr. Alireza Ghanbarpour was successful in obtaining the crystal structures of apo-dimer (1.9 Å), holo-dimer (1.9 Å) and holo-monomer (1.6 Å) of Q108K:K40L:T51F. In the crystal structure of Q108K:K50L:T51F-dimer/retinal there were four molecules in the asymmetric unit, of which sufficient retinal density was observed in two chains (**Figure VI-8**); the sigma level shown is 1.0. An overlay of retinal in the two chains shows that both retinal and the residue T51F exist in two conformations. Both *cis* and *trans* isomers of the iminium are shown, which would be expected based on the existence of both *cis* iminium and *trans* imine at pH 7 (**Figure VI-7**).

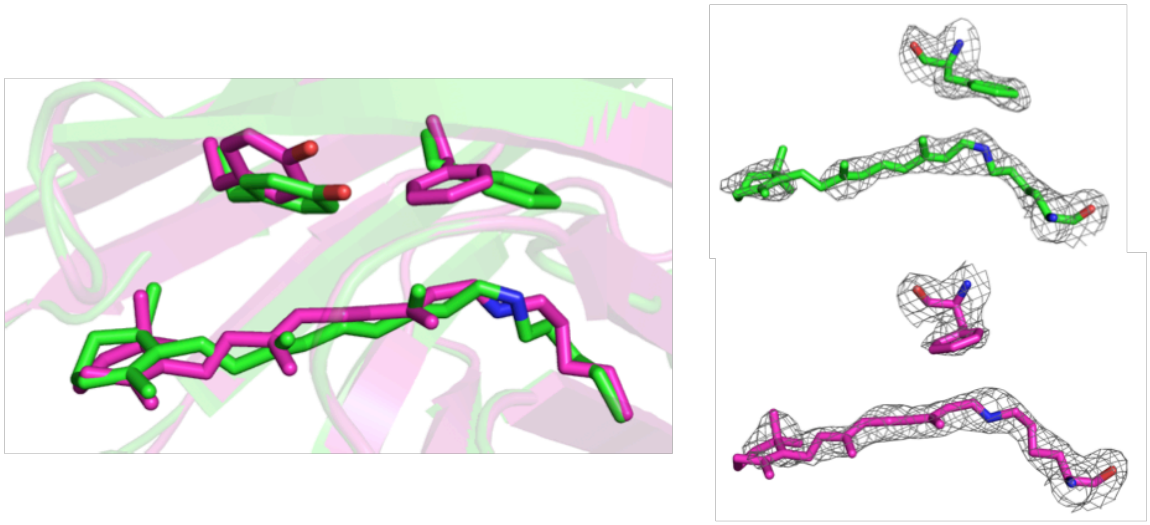


Figure VI-8. Overlay of two chains in the crystal structure of Q108K:K40L:T51F-dimer/retinal.

The crystal structure of Q108K:K40L:T51F-monomer/retinal showed four molecules in the asymmetric unit, with retinal occupancy in three molecules. In each, retinal adopts essentially the same conformation of the ligand. Again, both *cis* and *trans* isomers of the imine are observed (**Figure VI-9**). However, in this

case, T51F does not change conformation. Scrutiny of the crystal structure indicates that there are no significant differences of any active site residues.

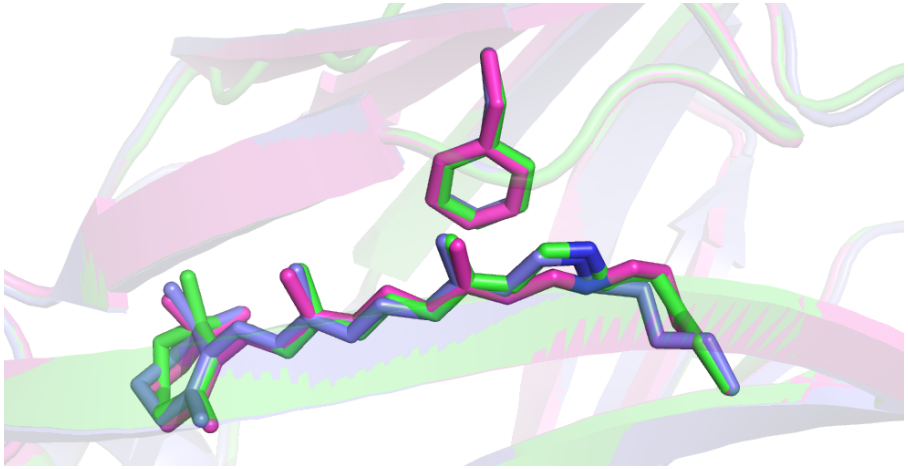


Figure VI-9. Overlay of three chains in the crystal structure of Q108K:K40L:T51F-monomer/retinal.

Both Q108K:K40L:T51F-monomer and dimer bind retinal as both *cis* and *trans* imine isomers. An overlay of the *cis* imine conformations of the monomer and dimer is shown in **Figure VI-10**, indicating that the trajectory of chromophore binding is quite different in the two cases. The major difference in the active-site residues is in the interaction of Y60 with E72.

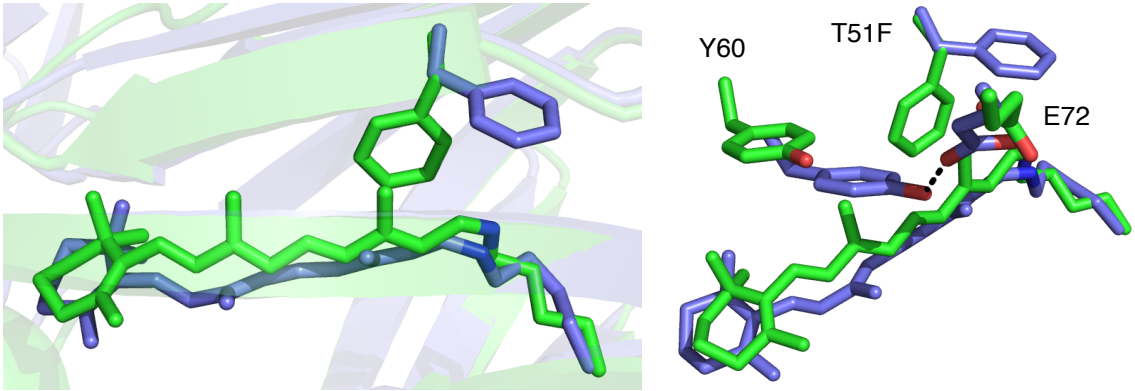


Figure VI-10. Overlay of the crystal structures of Q108K:K40L:T51F-monomer/retinal (green) and Q108K:K40L:T51F-dimer/retinal (purple).

In the domain-swapped dimer crystal structure, Y60 hydrogen bonds with E72 (2.2 Å). This hydrogen bond is not seen in the monomer structure because the residues are not close enough; this hydrogen bond presumably stabilizes the domain-swapped dimer when retinal is bound. Interestingly, in the apo Q108K:K40L:T51F-dimer structure, Y60 is flipped out of the binding cavity (**Figure V-11**), while N59 is flipped in. Careful inspection of the crystal structures does not reveal any other significant changes of the side chains in the active site. Binding of retinal requires N59 to flip out of the active site, or else it would sterically hinder the binding of retinal. As a result of this conformational change, Y60 flips in to hydrogen bond with E72 in the holo domain-swapped dimer. This hydrogen bond may prevent Y60 from impeding on imine formation in the case of the domain-swapped dimer. Conversely, the difference in rates of iminium formation may simply be due to the conformation of Phe51.

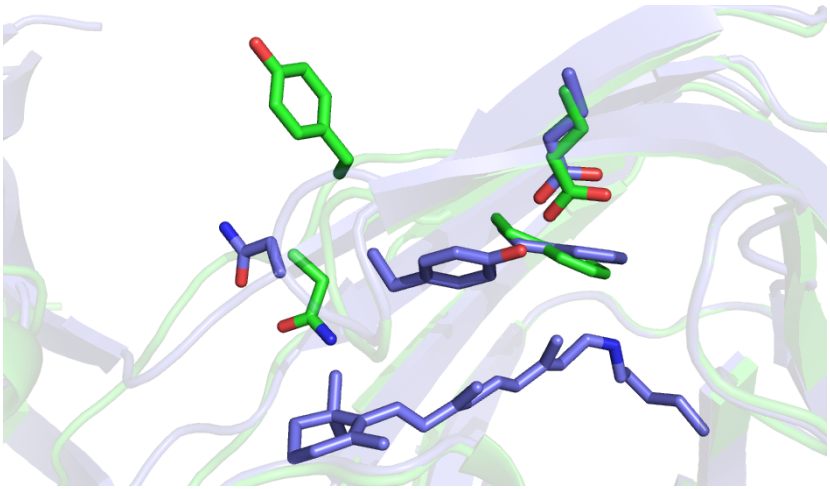


Figure VI-11. Overlay of the crystal structures of Q108K:K40L:T51F-dimer/retinal (purple) and apo Q108K:K40L:T51F-dimer (green).

Careful analysis of the UV-Vis binding data of the monomer and dimer with retinal indicate that it is binding of retinal that is the slow process in iminium

formation in the monomer. Free retinal absorbs at 380 nm, and upon binding to hCRBPII exhibits fine vibronic structure. This fine vibronic structure forms within two minutes in the case of Q108K:K40L:T51F-dimer/retinal (**Figure VI-12**). However, in the case of the monomer, vibronic structure at the SB is not apparent in such a time frame. Vibronic structure is only weakly apparent after 40 minutes (**Figure VI-12**). This may indicate that it is the step of retinal binding to the monomer that is slow. It would be interesting to measure the binding affinity of retinal to Q108K:K40L:T51F monomer and dimer to see if this is the reason, or if it is just a matter of the dimer binding pocket being more accessible.

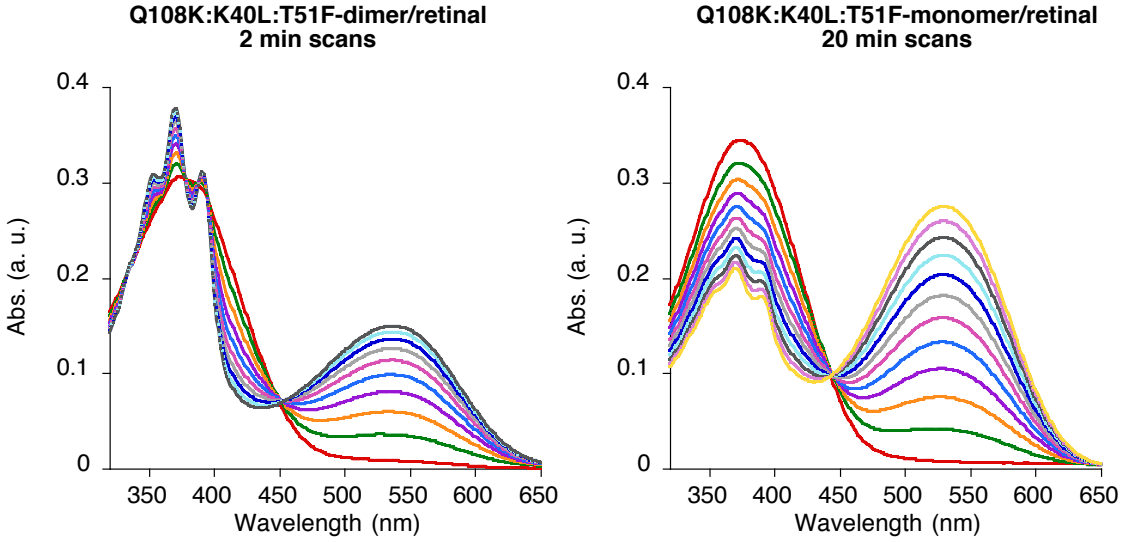


Figure VI-12. UV-Vis scanning of Q108K:K40L:T51F dimer and monomer over time with retinal.

VI.6 Mutation of Y60 in the triple mutant Q108K:K40L:T51F

As we have predicted that the Y60-E72 hydrogen bond is critical in minimizing the formation of the domain-swapped dimer and may play some role in retinal binding, we wanted to investigate the properties observed upon mutation of Y60. To this end, we mutated Y60 to alanine, phenylalanine, leucine,

histidine and tryptophan and E72 to leucine and alanine. Unfortunately, in the context of Q108K:K40L:T51F, E72A and E72L mutations did not provide soluble protein. Expression yields for the Y60 mutants are shown in **Table VI-6**.

Table VI-6. Effect of Y60 mutation in Q108K:K40L:T51F on protein expression yields and rate of iminium formation.

hCRBPII mutant	mol % monomer	mg/L monomer	mg/L dimer	$t_{1/2}$ (min) monomer	$t_{1/2}$ (min) dimer
Q108K:K40L:T51F	82	10.2	2.3	108.3	6.3
Q108K:K40L:T51F:Y60A	< 5	-	12.9	76.8	40.8
Q108K:K40L:T51F:Y60F	40	5.3	7.8	73.7	17.8
Q108K:K40L:T51F:Y60L	< 5	-	5.9	-	24.8
Q108K:K40L:T51F:Y60H	< 5	-	6.6	-	23.1
Q108K:K40L:T51F:Y60W	71	8.2	3.4	94.9	13.1

As previously shown, Y60 mutation led to the formation of more dimer.[1] Expression at room temperature led to only dimer formation in the case of Q108K:K40L:T51F:Y60A, Q108K:K40L:T51F:Y60L and Q108K:K40L:T51F:Y60H. However, expression at lower temperature can yield monomer (Mr. Alireza Ghanbarpour provided Q108K:K40L:T51F:Y60A monomer form an 19 °C expression).

Interestingly, only mutation to Y60A, in the context of Q108K:K40L:T51F led to significantly slower iminium formation in the dimer. The plot of absorbance intensity at the PSB and SB (**Figure VI-13**) is quite similar to that observed for Q108K:K40L:T51F-monomer/retinal (**Figure VI-7**), including the absorbance intensity at the PSB (indicating a similar pK_a to the monomer and higher than the dimer), and a slow rate of thermal isomerization. From this data, it is inferred that

the binding of retinal to Q108K:K40L:T51F:Y60A-dimer is similar to Q108K:K40L:T51F-monomer.

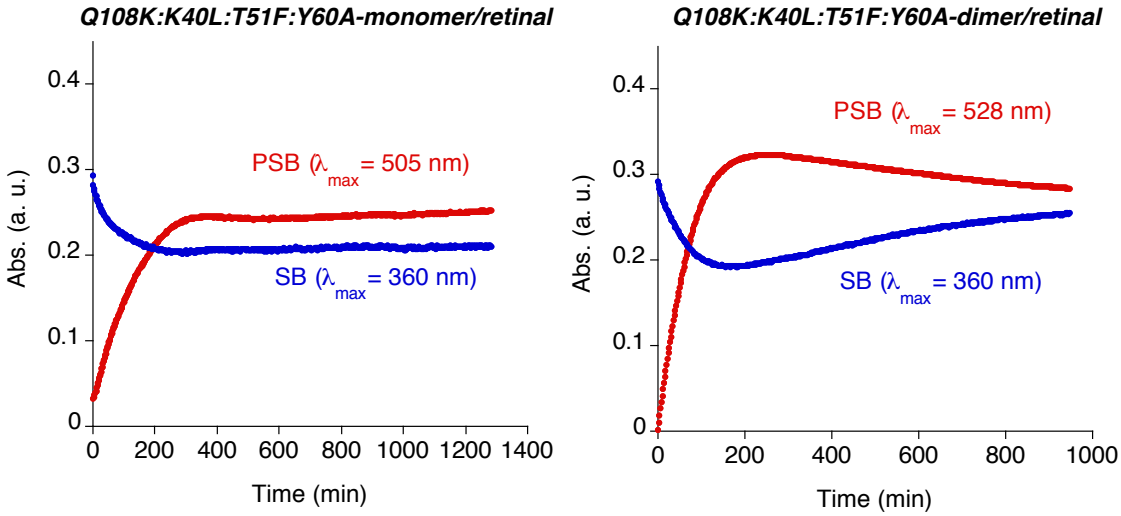


Figure VI-13. Absorbance intensity of PSB and SB as a function of time, followed by UV-Vis.

All of the holo domain-swapped dimer structures collected so far show an asymmetric orientation of the two domains, due to the flipped in orientation of N59 and flipping out of Y60. From the crystal structure of Q108K:K40L:T51F:Y60A-dimer/retinal, for the first time in a domain-swapped dimer structure, N59 is flipped into the binding pocket (**Figure VI-14**). This allows for a symmetric domain-swapped dimer as seen for most of the apo-dimer structures. Additionally, the ligand density is disordered; electron density is apparent for only the first four carbons of the bound retinylidene, suggesting that the ligand is not static inside of the protein cavity.

No concrete conclusions can be drawn as to the difference in the rates of iminium formation between the monomer and dimer. However, it is hypothesized that the slow step is in the rate of imine formation, presumably due to a sterically

hindered active site. Nonetheless, the most interesting observation made in this study is the mutation of Y60A, which leads to the first holo symmetric domain-swapped dimer, in which N59 adopts a flipped in conformation.

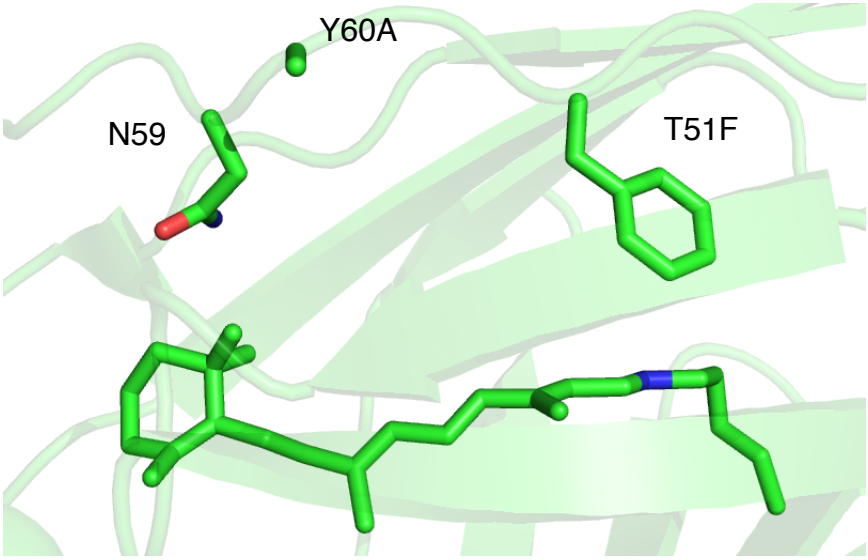


Figure VI-14. Crystal structure of Q108K:K40L:T51F:Y60A-dimer/retinal.

VI.7 Rational mutation to promote monomerization in hCRBPII mutants

As described in Chapter 4, we designed a hCRBPII mutant, Q108K:K40D:T53A:R58L:Q38F:Q4F:V62E, which displayed desirable spectroscopic properties when bound to ThioFluor as a large Stokes shift fluorescent protein (**Figure IV-6**). However, live cell imaging was not successful with this mutant. We hypothesized that it was due to the propensity of the mutant to oligomerize; this mutant produces only dimer under numerous protein expression conditions. To this end, we wanted to see if we could rationally monomerize the heptamutant to deem it suitable for *in vivo* experiments. As seen in **Table VI-3**, Q108K:K40L:T51V does not lead to any dimer expression at room

temperature. We hypothesized that T51V should be able to “monomerize” our dimer-forming mutant.

Therefore, we introduced T51V to the Q108K:K40D:T53A:R58L:Q38F:Q4F:V62E hCRBPII mutant in hopes of producing monomeric hCRBPII. To our delight, the addition of T51V does allow for a small amount of monomer formation (15 mol percent). Moreover, removal of T53A yields 72 mol percent monomer. Here we have shown that with the rational introduction of two mutations into Q108K:K40D:T53A:R58L:Q38F:Q4F:V62E (which produces only dimer during protein expression), we can obtain a greater ratio of monomer than dimer.

Table VI-7. Effect of T51 and T53 mutations on monomer/dimer ratio Q108K:K40D:T53A:R58L:Q38F:Q4F:V62E.

hCRBPII mutant	mol % monomer	mg/L monomer	mg/L dimer
Q108K:K40D:T53A:R58L:Q38F:Q4F:V62E	< 5	-	12.0
Q108K:K40D:T51V:T53A:R58L:Q38F:Q4F:V62E	15	1.5	17.4
Q108K:K40D:T51V:R58L:Q38F:Q4F:V62E	72	9.8	7.5

VI.8 Dimer formation in CRABPII, another iLBP

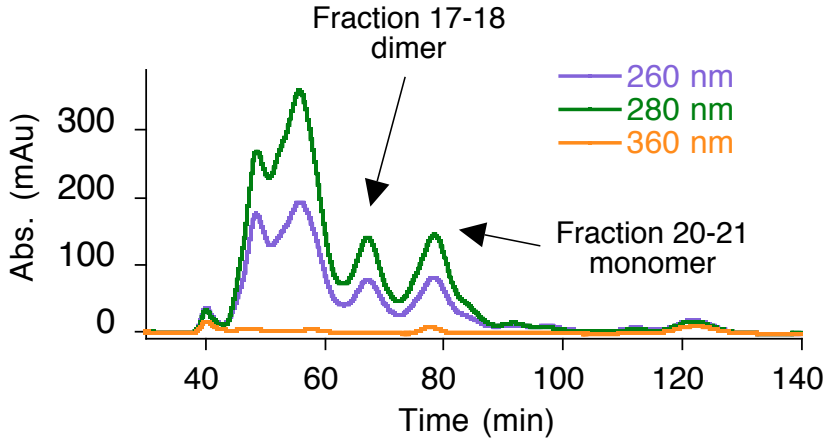
Because T51 mutations showed a propensity toward dimerization in hCRBPII, we wanted to see if the equivalent residue in another intracellular lipid binding protein (iLBP) displayed the same inclination. To this end, the equivalent residue I52 in cellular retinoic acid binding protein II (CRABPII) was mutated to alanine, asparagine, tryptophan and serine. In hindsight this was not a good position to test, as protein expression yields were low, presumably due to the fact that I52 is a conserved residue.

Nonetheless, only monomer was isolated under a number of different expression conditions for I52W, I52N and I52S. Fortuitously dimer was isolated from the expression of I52A under the following expression conditions: a single colony (from BL21(DE3) pLysS cells transformed with I52A DNA) was inoculated to 10 mL Luria broth (LB) with ampicillin (100 mg/L) and chloramphenicol (27 mg/L) and grown at 37 °C. After 16 hours, the solution was transferred to 1 L terrific broth (TB) with fresh ampicillin and chloramphenicol and grown for an additional 6 hours at 37 °C. Isopropyl-1-thio-D-galactopyranoside (IPTG) was added to the culture (at a final concentration 1 mM) to induce protein expression, and the culture was shaken at 18 °C for 29 hours.

The size exclusion chromatography (SEC) trace (run at 1 mL/min with 4 mL fractions) for this purification is shown in **Figure VI-15**. Based on the size of the fractions eluted, both monomer and dimer, as well as higher-order oligomers are isolated from this expression. To confirm that these proteins are CRABPII, rather than proteins from *E. coli*, a SDS-page denaturing gel was run (**Figure VI-15**). The order of fractions is in the order of elution from gel filtration.

From the denaturing gel, it is quite obvious that this expression does not cleanly produce only CRABPII. Fractions 12 to 17 from gel filtration are not CRABPII, but likely *E. coli* proteins. However, fractions 17 to 21 (though with some impurity) show the correct molecular weight for CRABPII. Based on this data, it is observed that CRABPII can indeed form the dimer species; whether it is domain swapped dimer or homo dimer, in which structural elements have not

been exchanged, is yet to be investigated. It is suggested that mutations that affect dimerization in hCRBPII be carried out on equivalent non-conserved residues in CRABPII.



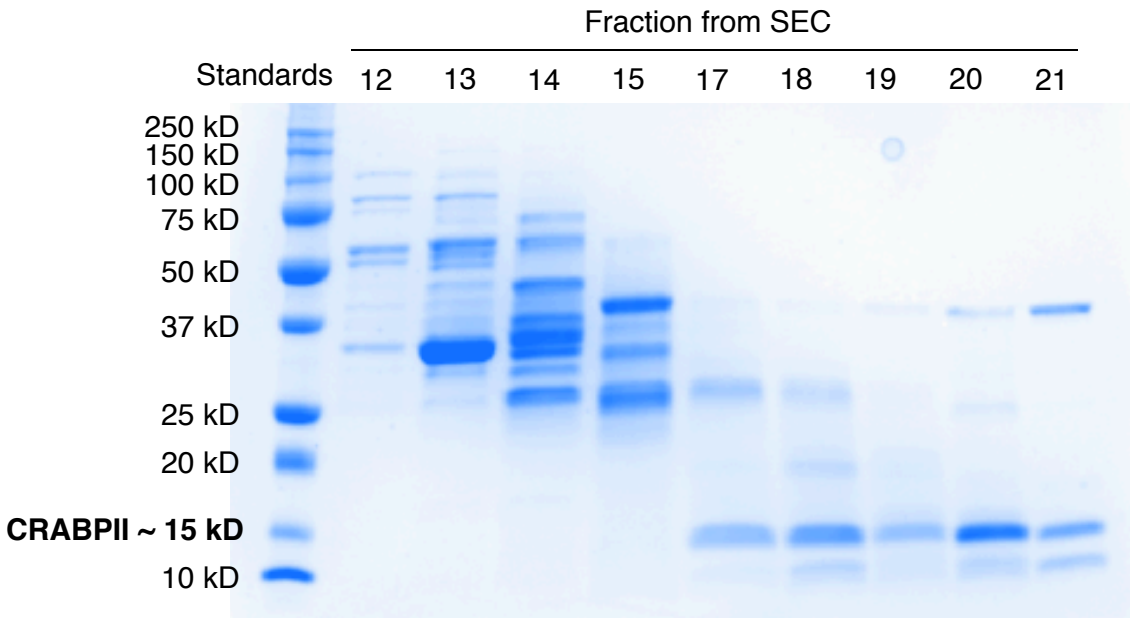


Figure VI-15. Size exclusion chromatography trace and denaturing protein gel for each fraction obtained from expression of I52A.

VI.9 Conclusions

The onset of this chapter described the seemingly random mutations that lead to dimer formation in hCRBPII. It was previously hypothesized that the interaction between Y60 and E72 was important for retaining the monomeric fold,

by controlling the conformation of the open monomer intermediate. It was further predicted that residues along the same strand (52 – 64) may also contribute to controlling the relative orientation of the two halves.[1]

The mutations shown to affect monomer/dimer ratio during bacterial protein overexpression are shown in the wild-type domain swapped dimer of hCRBPII (**Figure IV-16**). Presumably these mutations alter the hinge loop or the open interface, which can change the equilibrium between the monomer and the domain-swapped dimer. It is possible that new interactions (ionic, hydrophobic packing, etc.) at the open interface make the domain swapped dimer more favorable; this would explain why residues distal from the hinge loop (I25, K40, L77, Q128) could significantly alter the formation of domain swapped dimer in hCRBPII.

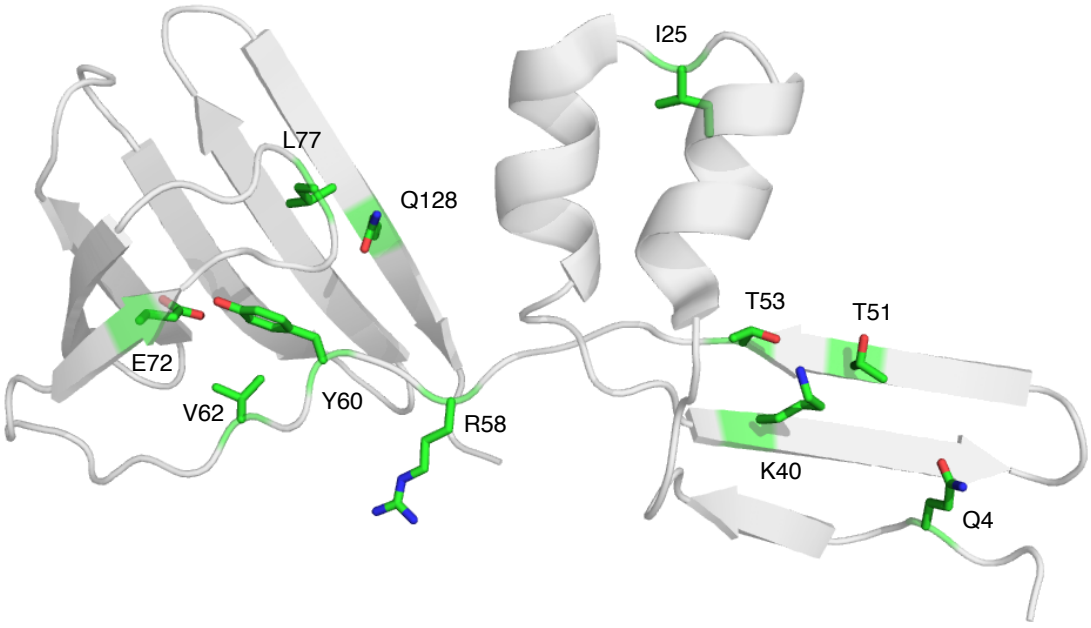


Figure VI-16. Residues shown to affect the formation of domain swapped dimer in hCRBPII. Coordinates obtained from 4ZH9.

In the absence of more crystal structures and protein folding assays, we cannot make any concrete conclusions as to why mutations at residue 51 render hCRBPII prone to dimerization. We hypothesize that the ratio of monomer/dimer formation is correlated to the stability of the monomer and dimer. If there exists an equilibrium between the closed monomer and open monomer intermediate, and the domain swapped dimer and the open monomer intermediate, it would stand to reason that the most stable product would prevail. In the same manner, destabilization of the monomer would lead to more domain swapped dimer. To this end, it would be interesting to measure the thermal stability of all the monomers and dimers developed in this study in order to ascertain whether the ratio of monomer/dimer corresponds with stability of the products.

REFERENCES

REFERENCES

1. Assar, Z., Nossoni, Z., Wang, W., Santos, E.M., Kramer, K., McCornack, C., Vasileiou, C., Borhan, B. & Geiger, J.H. Domain-Swapped Dimers of Intracellular Lipid-Binding Proteins: Evidence for Ordered Folding Intermediates. *Structure* **24**, 1590-1598 (2016).

2. Wang, W. in PhD Thesis, Michigan State University (2012).

3. Nosrati, M. in PhD Thesis, Michigan State University (2016).

4. Kyte, J. & Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105-132 (1982).

5. Hopp, T.P. & Woods, K.R. A computer program for predicting protein antigenic determinants. *Molecular Immunology* **20**, 483-489 (1983).

6. Cornette, J.L., Cease, K.B., Margalit, H., Spouge, J.L., Berzofsky, J.A. & DeLisi, C. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *Journal of Molecular Biology* **195**, 659-685 (1987).

7. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *Journal of Molecular Biology* **179**, 125-142 (1984).

8. Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H. & Zehfus, M.H. Hydrophobicity of amino acid residues in globular proteins. *Science* **229**, 834-838 (1985).

9. Janin, J. Surface and inside volumes in globular proteins. *Nature* **277**, 491-492 (1979).

10. Engelman, D.M., Steitz, T.A. & Goldman, A. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annual Review of Biophysics and Biophysical Chemistry* **15**, 321-353 (1986).

11. Berbasova, T., Santos, E.M., Nosrati, M., Vasileiou, C., Geiger, J.H. & Borhan, B. Light-Activated Reversible Imine Isomerization: Towards a Photochromic Protein Switch. *ChemBioChem* **17**, 407-414 (2016).

12. Nosrati, M., Berbasova, T., Vasileiou, C., Borhan, B. & Geiger, J.H. A Photoisomerizing Rhodopsin Mimic Observed at Atomic Resolution. *Journal of the American Chemical Society* **138**, 8802-8808 (2016).

CHAPTER VII: MATERIALS AND METHODS

VII.1 Site-directed mutagenesis of hCRBPII and CRABPII

The pET-17b plasmid, containing hCRBPII-Q108K:K40L cloned between NdeI and XhoI, was used as a template for mutagenesis of hCRBPII.[1, 2] Single point mutations were made on wild-type CRABPII in pET-17b (cloned between NdeI and EcoRI). Site-directed mutagenesis was conducted via polymerase chain reaction (PCR), with the specified cycling conditions shown in **Table VII-1**.

Table VII-1. PCR cycling conditions for site-directed mutagenesis.

	PCR Program	Time (min)
1x	94 °C	3:00
20x	94 °C	0:20
	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
DI 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. All primers correspond to hCRBPII, unless otherwise indicated.

Q4E-GACGAGGGACGAGAATGGAACC

Q4F-GACGAGGGACTTCAATGGAACC

Q4W-GACGAGGGACTGGAATGGAACC

W8E-AATGGAACCGAGGAGATGGAGAGT

W8F-CAGAATGGAACCTTCGAGATGGAGAGT

F16W:Y19W-AGTAATGAAAACTGGGAGGGCTGGATGAAGGCC

F16Y-GAGTAATGAAAACTATGAGGGCTACATG

F16Y:Y19W-GAGAGTAATGAAAACTATGAGGGCTGGATG

Y19F-CTTTGAGGGCTTCATGAAGGC

Y19W-CTTTGAGGGCTGGATGAAGGC

I25W-AAGGCCCTGGATTGGGATTTTGCCACC

T29W:A33W-GATTTTGCCTGGCGCAAGATTTGGGTACGTCTC

A33W-CGCAAGATTTGGGTACGTCTCAC

Q38A:K40L-GTACGTCTCACTGCGACGCTGGTTATTGATCAA

Q38E:K40L-GTACGTCTCACTGAGACGCTGGTTATTGATCAA

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

Q38M:K40L-CGTCTCACTATGACGCTGGTTATTG

Q38W:K40L-GTACGTCTCACTTGGACGCTGGTTATTGATCAA

K40D-CTCACTCAGACGGATGTTATTGATCAAGATGG

K40E-CTCACTCAGACGGAGGTTATTGATCAA

K40H-CTCACTCAGACGCACGTTATTGATCAA

K40L-CTCACTCAGACGCTGGTTATTGATCAAGATGG

K40W-GTCTCACTCAGACGTGGGTTATTGATCAAGA

F49E:T53A-GATGGTGATAACGAAAAGACAAAAGCAACTAGCACATTC

T51A-ATAACTTCAAGGCAAAAACCACTAGC

T51A:T53S-ATAACTTCAAGGCAAAATCCACTAGCACATT

T51C-GGTGATAACTTCAAGTGTAAAACCACTAGCAC

T51C:T53A-GATAACTTCAAGTGCAAAGCCACTAGCACATTC

T51C:T53S-GATAACTTCAAGTGTAAAAGCACTAGCACA

T51D-GGTGATAACTTCAAGGATAAAACCACTAGCAC

T51D:T53A-GATAACTTCAAGGACAAAGCCACTAGCACATTC

T51E-GATAACTTCAAGGAGAAAACCACTAGCAC

T51E:T53A-GATAACTTCAAGGAAAAAGCAACTAGCACATTC

T51F-GGTGATAACTTCAAGTTCAAAACCACTAGCAC

T51G-GGTGATAACTTCAAGGGCAAAACCACTAGCAC

T51H-GGTGATAACTTCAAG CAC AAAACCACTAGCAC

T51I-GGTGATAACTTCAAGATTAAAACCACTAGCAC

T51K-GATAACTTCAAGAAAAAAACCACTAGC

T51L-GGTGATAACTTCAAGCTGAAAACCACTAGCAC

T51M-GGTGATAACTTCAAGATGAAAACCACTAGCAC

T51N-GGTGATAACTTCAAGCTGAAAACCACTAGCAC

T51N:T53S-GTGATAACTTCAAGAATAAAAGCACTAGCAC

T51P-GGTGATAACTTCAAGCCAAAAACCACTAGCAC

T51Q-GGTGATAACTTCAAGCAAAAAACCACTAGCAC

T51R-GGTGATAACTTCAAGCGAAAAACCACTAGCAC

T51S-GGTGATAACTTCAAGTCGAAAACCACTAGCAC

T51V-GGTGATAACTTCAAGGTAAAAACCACTAGCAC

T51V:T53A-GATAACTTCAAGGTAAAAGCCACTAGCACATTC

T51V:T53D-CTTCAAGGTAAAAGACACTAGCACATTC

T51V:T53S-CAAGGTAAAAAGCACTAGCACATTC

T51W-GGTGATAACTTCAAGTGGAAAACCACTAGCAC

T51W:T53S-GATAACTTCAAGTGGAAAAGCACTAGCACATTC

T51Y-GGTGATAACTTCAAG TAC AAAACCACTAGCAC

T53A-TTCAAGACAAAAGCCACTAGCACATTC

T53E-TTCAAGACAAAAGAGACTAGCACATTC

T53S-CAAGACAAAAAGCACTAGCACATTCCG

T53V-CTTCAAGACAAAAGTCACTAGCACATTCCG

F57W:R58W-CACTAGCACATGGTGGAACTATGATG

R58E-CTAGCACATTCGAGAACTATGATGTG

R58F-CTAGCACATTCTTCAACTATGATGTG

R58G-CTAGCACATTCGGCAACTATGATGTG

R58H-CTAGCACATTCCACAACTATGATGTG

R58L-CTAGCACATTCCTGAACTATGATGTG

R58L:Y60F-ACTAGCACATTCCTCAACTTTGATGTGGATTTC

R58P-CTAGCACATTCCCCAACTATGATGTG

R58Q-CTAGCACATTCCAGAACTATGATGTG

R58W-CTAGCACATTCTGGAACTATGATGTG

R58W:Y60W-GCACATTCTGGAACTGGGATGTGGATTTC

R58Y-CTAGCACATTCTACAACTATGATGTG

Y60A-TTCCGCAACGCTGATGTGGAT

Y60F-CACATTCCGCAACTTTGATGTGGATTTCAC

Y60H-CACATTCCGCAACCATGATGTGGATTTCAC

Y60L-TTCCGCAACCTGGATGTGGAT

Y60W-CACATTCCGCAACTGGGATGTGGATTTCAC

V62E-AACTATGATGAGGATTTCACTGTTGGAGTA

F64E-AACTATGATGTGGATGAAACTGTTGGAGTAGAG

F64W-GATGTGGATTGGACTGTTGGAGTAG

V62N-AACTATGATAACGATTTCACTGTTGGAGTAG

E72A-GTAGAGTTTGACGCGTACACAAAGAGC

E72L-GTAGAGTTTGACCTGTACACAAAGAGC

S76A-GAGTACACAAAGGCCCTGGATAACCGG

S76W-GAGTACACAAAGTGGCTGGATAACCGG

L77W-GTACACAAAGAGCTGGGATAACCGGCATG

V86E-TTAAGGCACTGGAGACCTGGGAAGG

W88F-GCACTGGTCACCTTCGAAGGTGATGTC

L93E-TGGGAAGGTGATGTCGAAGTGTGTGTGCAAAAG

C95E-GATGTCCTTGTGGAGGTGCAAAAGGGG

W106E:Q108K-GAGAACCGCGGCGAGAAGAAGTGGATTGAGGGG

W106F:Q108K-GAGAACCGCGGCTTCAAGAAGTGGATT

Q108K:W109F-GGCTGGAAGAAGTTCATTGAGGGGGAC

L115E-GGGGACAAGGAGTACCTGGAGC

L117D-CAAGCTGTACGAGGAGCTGACC

L117E-CAAGCTGTACGACGAGCTGACC

L119E-GCTGTACCTGGAGGAGACCTGTGGTGAC

L119W-GTACCTGGAGTGGACCTGTGGTG

Q128E-CAGGTGTGCCGTGAGGTGTTCAAAAAG

Q128L-CAGGTGTGCCGTCTGGTGTTCAAAAAG

F130E-TGCCGTCAAGTGGAGAAAAAGAAGTGA

I52A (CRABPII)-GACACTTTCTACGCCAAAACCTCCACC

I52S (CRABPII)-GACACTTTCTACUCCAAAACCTCCACC

I52W (CRABPII)-GACACTTTCTACTGGAAAACCTCCACC

I52N (CRABPII)-GACACTTTCTACAACAAAACCTCCACC

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 min. Subsequently, the cells were heat shocked for 30 sec 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 h.

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 h. 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 sampled 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.

VII.2 Protein expression and purification of hCRBPII and CRABPII in pET-17b expression plasmid

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 min. Subsequently, the cells were heat shocked for 30 sec 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 h.

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 KH_2PO4 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 h. 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 h.

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 (100 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 VII-2**. 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 VII-2. 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 VII-3**. At all steps, the pH was set to 8.1.

Table VII-2. 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

VII.3 Protein extinction coefficient determination

The extinction coefficients of the proteins were measured at 280 nm, following the method described by Gill and von Hippel.[3] 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 and filtered before using.

The extinction coefficients of all purified hCRBPII and CRABPII monomers and dimers described in this thesis are given in **Table VII-3**. All proteins are hCRBPII, unless otherwise noted.

Table VII-3. Extinction coefficients of hCRBPII and CRABPII mutants.

Protein	$\varepsilon_{exp(280\ nm)}$	
	Monomer	Dimer
Q108K:K40C:T53A:R58L:Q38F:Q4F	27,948	-
Q108K:K40D	28,612	58,690
Q108K:K40D:Q4F	28,145	-
Q108K:K40D:R58L:Q38F:Q4F	29,070	58,142
Q108K:K40D:T51V:R58L:Q38F:Q4F:V62E	27,532	55,904
Q108K:K40D:T51V:T53A:R58L:Q38F:Q4F:V62E	26,357	53,118
Q108K:K40D:T53A:Q38F:Q4F	-	55,372
Q108K:K40D:T53A:R58L:Q38F:Q4F	28,682	56,298
Q108K:K40D:T53A:R58L:Q38F:Q4F:V62E	-	56,982
Q108K:K40D:T53A:R58L:Q4F	25,887	55,852
Q108K:K40D:T53A:R58Y:Q38F:Q4F	28,690	-
Q108K:K40D:T53A:R58Y:Q38F:Q4F:L117E	32,800	59,040
Q108K:K40E	31,611	56,834
Q108K:K40E:R58L:Q38F:Q4F	29,350	52,302
Q108K:K40E:R58L:Q38F:Q4F:W106F	21,632	-
Q108K:K40E:R58L:Q38F:Q4F:W109F	22,206	-
Q108K:K40E:R58L:Q38F:Q4F:W88F	20,811	-
Q108K:K40E:R58L:Q38F:Q4F:W8F	21,353	-
Q108K:K40E:T53A:Q38F:Q4F	28,083	53,656
Q108K:K40E:T53A:R58L:Q38F:Q4F	27,681	-
Q108K:K40E:T53A:R58L:Q38F:Q4F:V62E	25,267	53,774
Q108K:K40E:T53A:R58L:Q38F:Q4F:Y19W	32,892	-
Q108K:K40E:T53A:R58L:Q4F	28,344	54,796
Q108K:K40E:T53A:R58W:Q38F:Q4F	34,382	-
Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W	38,874	-
Q108K:K40H:T51V:T53S:R58W:Y19W:A33W	42,995	-
Q108K:K40H:T53A:R58L:Q38F:Q4F	27,742	-

Table VII-3 (cont'd)

Q108K:K40L	27,618	-
Q108K:K40L:L117E	28,540	-
Q108K:K40L:Q38F	29,409	-
Q108K:K40L:Q4F	30,761	-
Q108K:K40L:R58F	26,372	-
Q108K:K40L:R58W:Y19W	38,360	-
Q108K:K40L:R58Y	28,800	-
Q108K:K40L:T51A	29,926	59,182
Q108K:K40L:T51A:T53S:V62N	-	61,640
Q108K:K40L:T51C	31,624	-
Q108K:K40L:T51C:T53A:R58L:Q38F	28,527	-
Q108K:K40L:T51D:T53A:R58Y:Q38F:Q4F	28,763	-
Q108K:K40L:T51D:T53A:R58Y:Q38F:Q4F:L117E	30,317	-
Q108K:K40L:T51E	27,398	58,828
Q108K:K40L:T51F	30,065	63,540
Q108K:K40L:T51F:Y60A	-	49,397
Q108K:K40L:T51F:Y60F	29,077	56,631
Q108K:K40L:T51F:Y60H	-	56,329
Q108K:K40L:T51F:Y60L	-	59,024
Q108K:K40L:T51F:Y60W	30,651	62,700
Q108K:K40L:T51H	26,956	53,057
Q108K:K40L:T51I	28,736	-
Q108K:K40L:T51K	-	57,708
Q108K:K40L:T51L	27,987	64,330
Q108K:K40L:T51M	26,772	62,166
Q108K:K40L:T51N	30,281	60,452
Q108K:K40L:T51N:T53S:V62N	32,319	57,556
Q108K:K40L:T51Q	27,663	55,434

Table VII-3 (cont'd)

Q108K:K40L:T51S	33,186	56,882
Q108K:K40L:T51V	29,609	-
Q108K:K40L:T51V:R58W:Y19W	37,342	-
Q108K:K40L:T51V:T53D:R58W:Y19W:L117E	37,899	-
Q108K:K40L:T51V:T53S	30,888	-
Q108K:K40L:T51V:T53S:R58E	28,960	60,986
Q108K:K40L:T51V:T53S:R58F	29,833	-
Q108K:K40L:T51V:T53S:R58G	30,371	-
Q108K:K40L:T51V:T53S:R58H	28,504	-
Q108K:K40L:T51V:T53S:R58H:Y19W	34,571	-
Q108K:K40L:T51V:T53S:R58H:Y19W:L117D:Q4F	31,920	-
Q108K:K40L:T51V:T53S:R58H:Y19W:L117E	32,192	-
Q108K:K40L:T51V:T53S:R58H:Y19W:L117E:A33W	38,408	-
Q108K:K40L:T51V:T53S:R58L	29,079	60,938
Q108K:K40L:T51V:T53S:R58P	30,155	-
Q108K:K40L:T51V:T53S:R58Q	29,013	60,989
Q108K:K40L:T51V:T53S:R58W	33,984	-
Q108K:K40L:T51V:T53S:R58W:Y19W	36,880	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W	43,042	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:F16Y	44,754	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:F57W	47,166	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:F64W	49,830	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:I25W	54,006	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L117D	43,373	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L117E	42,130	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L119W	49,720	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:Q38W	48,544	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:Q4W	52,151	-

Table VII-3 (cont'd)

Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:Q4W:F16Y	57,192	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:S76W	51,910	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:T29W	52,454	-
Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:Y60W	49,545	118,327
Q108K:K40L:T51V:T53S:R58W:Y19W:F16Y	39,990	-
Q108K:K40L:T51V:T53S:R58W:Y19W:L117D	38,626	-
Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:A33W:Q4F	42,730	-
Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F	38,135	-
Q108K:K40L:T51V:T53S:R58W:Y19W:L117E	35,721	-
Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:Q4F	37,554	-
Q108K:K40L:T51V:T53S:R58W:Y19W:L77W	44,867	-
Q108K:K40L:T51V:T53S:R58W:Y19W:L77W:F16Y	45,310	-
Q108K:K40L:T51V:T53S:R58Y	28,768	-
Q108K:K40L:T51V:T53S:R58Y:Y19W	36,971	-
Q108K:K40L:T51V:T53S:Y19W	34,288	-
Q108K:K40L:T51V:T53W:R58W:Y19W:A33W	57,298	-
Q108K:K40L:T51W	38,181	75,400
Q108K:K40L:T51W:T53S:R58W:Y19W:A33W	57,571	-
Q108K:K40L:T51Y	28,044	54,498
Q108K:K40L:T53A	29,917	61,852
Q108K:K40L:T53A:Q38A	29,631	57,823
Q108K:K40L:T53A:Q38M	29,587	-
Q108K:K40L:T53A:R58F	28,499	-
Q108K:K40L:T53A:R58L	28,621	54,716
Q108K:K40L:T53A:R58L:A33W	35,160	-
Q108K:K40L:T53A:R58L:Q128L	33,769	65,392
Q108K:K40L:T53A:R58L:Q38A	29,882	58,515
Q108K:K40L:T53A:R58L:Q38F	29,023	-

Table VII-3 (cont'd)

Q108K:K40L:T53A:R58L:Q38F:F16Y	29,004	-
Q108K:K40L:T53A:R58L:Q38F:L77W	34,305	-
Q108K:K40L:T53A:R58L:Q38F:Q4F	31,014	-
Q108K:K40L:T53A:R58L:Q38F:Q4F:I42E	29,326	-
Q108K:K40L:T53A:R58L:Q38F:Q4F:F49E	32,681	-
Q108K:K40L:T53A:R58L:Q38F:Q4F:L117E	28,159	57,373
Q108K:K40L:T53A:R58L:Q38F:Q4F:L119E	29,595	-
Q108K:K40L:T53A:R58L:Q38F:Q4F:Q128E	30,636	-
Q108K:K40L:T53A:R58L:Q4F:Q38E	29,153	-
Q108K:K40L:T53A:R58L:Q38F:Q4F:T51E	29,050	63,812
Q108K:K40L:T53A:R58L:Q38F:Q4F:V62E	28,660	-
Q108K:K40L:T53A:R58L:Q38F:S76A	30,174	-
Q108K:K40L:T53A:R58L:Q38F:Y19W	36,749	-
Q108K:K40L:T53A:R58L:Q38F:Y60F	26,367	-
Q108K:K40L:T53A:R58L:Q38M	28,828	-
Q108K:K40L:T53A:R58L:Y19F	29,082	64,206
Q108K:K40L:T53A:R58W	37,876	71,046
Q108K:K40L:T53A:R58Y:Q38F:Q4F:L117E	28,820	-
Q108K:K40L:T53E:R58L:Q38F:Q4F	27,822	-
Q108K:K40L:T53S	28,195	64,629
Q108K:K40L:T53S:R58W:Y19W	38,340	-
Q108K:K40L:T53S:V62E	29,792	60,016
Q108K:K40L:T53S:V62N	32,430	59,200
Q108K:K40L:T53V	28,710	61,756
Q108K:K40L:T53V:R58L	27,460	-
Q108K:K40L:V62N	29,091	58,408
Q108K:K40Q:T53A:R58L:Q38F:Q4F	28,759	-
Q108K:K40R:T53A:R58L:Q38F:Q4F	27,776	-

Table VII-3 (cont'd)

Q108K:K40T:T53A:R58L:Q38F:Q4F	28,048	-
Q108K:K40W:T51V:T53S:R58W:Y19W:A33W	53,280	-
Q108K:K40Y:T53A:R58L:Q38F:Q4F	28,587	-
Q108K:T53A:R58L:Q38F:Q4F	29,370	-
I52A (CRABPII)	18,391	-
I52S (CRABPII)	16,914	-
I52N (CRABPII)	19,582	-
I52W (CRABPII)	26,101	-

VII.4 UV-Vis measurements of hCRBPII/chromophore complexes

UV-Vis spectra were recorded with a Cary 300 Bio WinUV, Varian spectrophotometer. For all experiments, unless otherwise indicated, 20 μM protein was incubated with ligand (0.5 equiv) in 2 x 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.

VII.4.1 pK_a measurements of hCRBPII/chromophore complexes

For pK_a determination, protein (20 μM in PBS) was incubated with ligand (0.5 equiv) and incubated 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 (with 1 M NaOAc, 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 are: 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 the deviation from zero absorbance intensity of the deprotonated PSB.

VII.4.2 Kinetic measurements of hCRBPII/chromophore PSB formation

For kinetic measurements, PSB formation was followed by UV-Vis in 2 x PBS buffer (~ pH 7.2) at 23 °C using a Cary temperature controller. pH was verified each time 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 absorbance intensity plotted as a function of time.

The data were fit to either a pseudo-first order rate equation or second order rate equation. Those samples that exhibited high PSB pK_a (pK_a > 9) were fit to a second order rate equation, as we can be confident in the chromophore concentration. However, when both SB and PSB are formed under the conditions for the kinetic measurements, we had to fit a pseudo-first order rate equation, as concentration could not be accurately determined.

The fit for pseudo-first order rate eis shown below; *A* is the absorbance value at each recorded time point, A_0 is the final absorbance value (at completion of the reaction), *k* is the rate constant and *t* is the time after ligand addition.

$$A = A_0 \times (1 - e^{-kt})$$

The data was fit in Kaleidagraph using the equation shown below, where m3 is a constant that accounts for the time delay after ligand addition before

recording the first time point. The half-life ($t_{1/2}$) of the reaction can be calculated by ln(2)/*k*, where m2 is the rate constant.

$$y = m1 \times (1 - e^{-m2 \times m0}) + m3$$

For those samples exhibiting high PSB pK_a, we chose to fit the second order rate equation (in Kaleidagraph) shown below. The second order rate equation for product growth was previously derived,[4, 5] where m3 is the concentration of the limiting reactant (in all cases this should be 10 μM, assuming greater than 50% of the protein is functional), m2 is the rate constant *k* and m1 is 1/m3. The half-life ($t_{1/2}$) of the reaction can be calculated by $1/(k^*[A]_0)$. In all cases 10 μM was assumed for $[A]_0$.

$$y = m3 - \frac{1}{(m2 \times m0) + m1}$$

First, the concentration of product (hCRBPII/ligand complex) versus time was plotted. The concentration of product at each time point was calculated by Beer's Law using the absorbance intensity at each time point and the extinction coefficient of the hCRBPII/ligand complex. For kinetic experiments conducted by fluorescence, product concentration was calculated by dividing the fluorescence intensity at each time point by the apparent extinction (maximum fluorescence intensity divided by the limiting reactant concentration).

VII.5 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.

VII.6 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. All absolute quantum yields were measured by Mr. Wei Sheng and taken as an average of five recordings, excited at lambda max.

VII.7 Live cell imaging in bacteria

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 min. Subsequently, the cells were heat shocked for 30 sec 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 h.

A single colony was used for inoculation to 10 mL LB. The culture was shaken at 37°C for 6 h followed by induction with IPTG (final concentration 1 mM). The culture was then shaken at 25°C for 16 h. For the control experiment, 5 µL of BL21 (DE3)pLysS cells was inoculated into 10 mL of LB. Culture growth and induction with IPTG proceeded the same as aforementioned.

The cells were harvested by centrifugation and resuspended in 1 mL of PBS buffer (pH 7.2). The cultures were heated to 37°C for 10 min. Subsequently, ThioFluor was added and incubated at 37 °C, at the indicated final concentration

and time. After the incubation period, the cells were harvested by centrifugation and the supernatant was discarded. The pellets were washed 3x with warm PBS buffer. After the last wash, the cells were resuspended in PBS buffer.

For imaging, the cells were plated on a microslide (VWR, 1 in x 3 in, 1.2 mm thick) and covered with a glass coverslip (*Corning, No.1, 22 mm sq.)*. In order to stop the bacteria from floating in solution, the slide was heated at 42 °C for ~1 min. The sample was placed with the coverslip facing the objective. For bacterial imaging, a 63 x oil objective was used.

VII.8 Cloning for mammalian expression vectors

VII.8.1 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 VII-4** using a Bio-Rad iCycler thermal cycler. Four reactions were set up for each cloning, to ensure that enough amplified product was obtained.

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/mL. The product was digested with the proper enzymes and ligated to a similarly prepared plasmid (50 ng/mL). 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 and grown on LB-agar plates supplemented with antibiotics (100

μg/mL ampicillin, 7.5 μg/mL tetracycline) at 37 ºC for 18 h. 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 h. DNA purification was performed using Promega Wizard® Plus SV Miniprep DNA purification kit (A1330) following the manufacturer's protocol.

Table VII-4. 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
DI water	50 μL – x – y – z – 17.5 μL

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'

CFP FOR: 5'-CGGTCGCCACCATGGTGAGCAAG-3'

VII.8.2 Preparation of plasmids for mammalian expression

Plasmid1 (HindIII-EGFP-NotI-hCRBPII$_{tetra}$-Stop-BamHI in pFlag-CMV2) and Plasmid2 (HindIII-EGFP-NotI-hCRBPII$_{tetra}^{Q108L}$-Stop-BamHI in pFlag-CMV2) were constructed by amplifying the hCRBPII gene of interest from pET-17b, Q108K:K40L:T51V:R58F and Q108L:K40L:T51V:R58F, respectively. Forward and reverse primers used were 5'-CGGCGGCCGCATGACGAGGGCC-3' and 5'-GCGGATCCTCACTTCTTTTTGAACAC-3' (see **Table VII-4** for protocol, extension time = 15 sec). The purified PCR product was then inserted between NotI and BamHI of Plasmid3. The products were sequenced with the primers CMVend_Seq and midGFP.

Plasmid3 (HindIII-EGFP-NotI-CRBPII$_{nona}$-Stop-BamHI in pFlag-CMV2) was prepared in multiple steps. The first step (completed by Dr. Tetyana Berbasova) served to obtain the construct HindIII-EGFP-hCRBPII$_{nona}$-Stop-BamHI in pFlag-CVM2.[6] The EGFP- hCRBPII$_{nona}$ insert was amplified in two rounds of PCR (the italicized portion of the primer indicates the overlap of the sequences): 1) EGFP was amplified from EGFP-RB (donated from Professor Bill Henry's Lab at MSU) with the forward primer 5'-CGAAGCTTATGGTGAGCAAGGGCGAG-3' and reverse 5'-CCGGA*CTTGTACAGC*-3' primer (see Table S3 for protocol, extension time = 30 sec). 2) hCRBPII$_{nona}$ was amplified from pET-17b with the

forward primer 5‘-*GCTGTACAAG*TCCGGAGCCGCTGCAGGAGGC AGCCAAATGACGAGGTTC-3’ and reverse primer 5’-GCGGATCCTCACTTCTTTTTGAACAC-3’ (see **Table VI-4** for protocol, extension time = 30 sec). The PCR products were combined and amplified with the abovementioned HindIII-EGFP forward primer and hCRBPII-Stop-BamHI reverse primer (see **Table VII-4** for protocol, extension time = 45 sec). The full-length insert was ligated between HindIII and BamHI of the digested pFLAG-CMV2-AML1B vector (Addgene #12504).

A NotI cutting site was then introduced by two sequential site-directed mutageneses. The DNA was purified and sequenced after each step. The first PCR was conducted with the forward primer 5’-CTGCAGGAGGCGGCCAAATGACGA-3’ and reverse primer 5’-CTCGTCATTTGGCCGCCTCCTGCAG-3’ to introduce the first half of the NotI cutting site (see Table S1 for protocol). The second PCR was conducted with the forward primer 5’-CAGGAGGCGGCCGCATGACGAGG-3’ and reverse primer 5’-CCTCGTCATGCGGCCGCCTCCTG-3’ to finish introducing the NotI cutting site (see **Table VII-1** for protocol). After the two PCRs, the product is HindIII-EGFP-NotI-hCRBPII$_{nona}$-Stop-BamHI in pFlag-CVM2. This was done to allow for easy exchange of the hCRBPII variant. The product was sequenced with the primers CMVend_Seq and midGFP.

Plasmid4 (HindIII-EGFP-NotI-hCRBPII$_{nona}$-EcoRI-3NLS-Stop in pFlag-CMV2) was prepared in multiple steps. The first step served to obtain the

construct HindIII-hCRBPII$_{nona}$-3NLS-Stop-EcoRI in pFlag-CVM2 (completed by Dr. Tetyana Berbasova).[6] Two sequential PCR amplifications were necessary to amplify the full-length insert. In the first step, hCRBPII$_{nona}$ was amplified from the pET-17b vector using a reverse primer containing a partial 3NLS sequence. The forward and reverse primers used are 5'-GCAAGCTTACGAGGGACTTCAATGGAACC-3' and 5'-CTTCTTGGGGTCCACCTTCCTCTTCTTCTTGGGGTCAGCCCTGCTCTTCTTTTTGAACAC-3' (see **Table VII-4** for protocol, extension time = 15 sec). The PCR product was purified by gel extraction as described above. A second PCR was performed to complete the full-length 3NLS localization signal, add a stop codon and an EcoRI site. The second PCR was done in two 50 µL reaction volume using 20 ng of the product from the first PCR. Forward and reverse primers used in the second reaction are 5'-GCAAGCTTACGAGGGACTTCAATGGAACC-3' and 5'-GAATTCACACCTTCCTCTTCTTCTTGGGGTCCACCTTCCTCTTCTTCTTGGGGTCCAC-3' (see Table **VII-4** for protocol, extension time = 15 sec). The purified PCR product was then inserted between HindIII and EcoRI of the digested pFLAG-CMV2-AML1B vector.

Subsequently, the existing EcoRI cutting site was removed by site-directed mutagenesis with the forward primer 5'-GGAAGGTGTGAAAACATCGATAGA-3' and reverse primer 5'-TCTATCGATGTTTTCACACCTTCC-3' (see **Table VII-1** for protocol) and a new EcoRI cutting site was introduced via site-directed mutagenesis before the 3NLS signal by PCR with the forward primer 5'-

AAAAAGAAGGAATTCGCTGACCCC-3' and reverse primer 5'-GGGGTCAGCGAATTCCTTCTTTT-3' (see **Table VII-1** for protocol). After these two PCRs, the product is HindIII-hCRBPII$_{nona}$-EcoRI-3NLS-Stop in pFlag-CVM2. This was done to allow for easy exchange of the hCRBPII variant.

EGFP-NotI-hCRBPII$_{nona}$ was then amplified from Plasmid3 with the forward primer 5'-CGAAGCTTATGGTGAGCAAGGGCGAG-3' and reverse primer 5'-CGGAATTCGCACTTCTTTTTGAACACTTG-3' (see **Table VII-4** for protocol, extension time = 30 sec). The purified PCR product was then inserted between HindIII and EcoRI of the HindIII-hCRBPII$_{nona}$-EcoRI-3NLS-Stop construct in pFlag-CVM2. The product was sequenced with the primers CMVend_Seq and midGFP.

To construct Plasmid5 (HindIII-EGFP-NotI-hCRBPII$_{nona}$-EcoRI-NES(ELAEKLAGLDIN)-Stop-BamHI in pFlag-CMV2), a gBlock gene fragment of hCRBPII$_{nona}$-EcoRI-NES was ordered from IDT. The fragment was amplified with the forward primer 5'-CGGCGGCCGCATGACGAGGTTC-3' and reverse primer 5'-CGGGATCCTCAATTTATGTCAAGCCCGGCAAG-3' (see **Table VII-4** for protocol). The purified PCR product was then inserted between NotI and BamHI of Plasmid3. The product was sequenced with the primers CMVend_Seq and midGFP.

Plasmid6 (XhoI-Q108K:K40L:T53S:V62N-BamHI-AgeI-ECFP-Stop-NotI in pECFP-N1) was constructed by amplifying Q108K:K40L:T53S:V62N from pET-17b. Forward and reverse primers used were 5'-

GTTTACTTTACTCGAGCTCGCCACCATGACGAGGGACC-3' and 5'-CGGTGGATCCTTCGACTTCTTTTTGAACACTTG-3' (see **Table VII-4** for protocol, extension time = 15 sec). The purified PCR product was then inserted between XhoI and BamHI of GalT-ECFP (Addgene #11937). The product was sequenced with the primers CMVend_Seq and CFP FOR.

Plasmid7 (HindIII-EGFP-NotI-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-BamHI-RVASL-Stop in pFlag-CMV2) was constructed by amplifying Q108K:K40L:T51V:T53S:R8W:Y19W:L117E from pET-17b. Forward and reverse primers used were 5'-CGGCGGCCGCATGACGAGGGAC-3' and 5'-GCCATGGATCCGCACTTCTTTTTGAACA-3' (see **Table VII-4** for protocol, extension time = 15 sec). The purified product was then inserted between NotI and BamHI of Plasmid3. The product was sequenced with the primers CMVend_Seq and midGFP.

Plasmid8 (HindIII-EGFP-NotI-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-EcoRI-3NLS-Stop in pFlag-CMV2) was made by amplifying Q108K:K40L:T51V:T53S:R8W:Y19W:L117E from pET-17b. Forward and reverse primers used were 5'-CGGCGGCCGCATGACGAGGGAC-3' and 5'-CGGAATTCGCACTTCTTTTTGAACACTTG-3' (see **Table VII-4** for protocol, extension time = 15 sec). The purified product was then inserted between NotI and EcoRI of Plasmid4. The product was sequenced with the primers CMVend_Seq and midGFP.

Plasmid9 (HindIII-EGFP-NotI-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-EcoRI-NES(ELAEKLAGLDIN)-Stop-BamHI in pFlag-CMV2) was constructed by amplifying Q108K:K40L:T51V:T53S:R8W:Y19W:L117E from pET-17b. The forward and reverse primers used were 5'-CGGCGGCCGCATGACGAGGGAC-3' and 5'-CGGAATTCGCACTTCTTTTTGAACACTTG-3' (see **Table VII-4** for protocol, extension time = 15 sec). The purified product was then inserted between NotI and EcoRI of Plasmid5. The product was sequenced with the primers CMVend_Seq and midGFP.

Plasmid10 (HindIII-EGFP-NotI-Q108K:K40E:T53A:R58L:Q38F:Q4F-EcoRI-3NLS-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 VII-4** for protocol, extension time = 15 sec). The purified product was then inserted between NotI and EcoRI of Plasmid8. The product was sequenced with the primers CMVend_Seq and midGFP.

Plasmid11 (HindIII-EGFP-NotI-Q108K:K40E:T53A:R58L:Q38F:Q4F-Stop-BamHI 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'-GCGGATCCTCACTTCTTTTTGAACAC-3' (see **Table VII-4** for protocol, extension time = 15 sec). The purified product was then inserted between NotI

and BamHI of Plasmid7. The product was sequenced with the primers CMVend_Seq and midGFP.

Plasmid12 (HindIII-Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F-EcoRI-3NLS-Stop in pFlag-CMV2) was constructed by amplifying Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F from pET-17b. The forward and reverse primers used were 5'-GATATAAAGCTTATGACGAGGGACCAG-3' and 5'-CGGAATTCGCACTTCTTTTTGAACACTTG-3'. The purified product was then inserted between HindIII and EcoRI of HindIII-hCRBPII$_{nona}$-EcoRI-3NLS-Stop in pFlag-CVM2 (an intermediate of Plasmid4). The product was sequenced with the primer CMVend_Seq.

VII.8.3 Sequences of plasmids described in this thesis

Plasmid1: HindIII-EGFP-NotI-Q108K:K40L:T51V:R58F-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACCAGAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATA
TTGATTTTGCCACCCGCAAGATTGCAGTACGTCTCACTCAGACGCTGGTTATTGATCAA
GATGGTGATAACTTCAAGGTAAAAACCACTAGCACATTCTTCAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACCTGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGAGGATCC

Plasmid2: HindIII-EGFP-NotI-Q108L:K40L:T51V:R58F-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACCAGAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATA
TTGATTTTGCCACCCGCAAGATTGCAGTACGTCTCACTCAGACGCTGGTTATTGATCAA
GATGGTGATAACTTCAAGGTAAAAACCACTAGCACATTCTTCAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGCTGTGGATTGAGGGGGACAAGCTGTACCTGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGAGGATCC

Plasmid3: HindIII-EGFP-NotI-Q108K:K40L:T51V:T53C:R58W:T29L:A33W:Q4F:L117E-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACTTCAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATA
TTGATTTTGCCCTGCGCAAGATTTGGGTACGTCTCACTCAGACGCTGGTTATTGATCAA
GATGGTGATAACTTCAAGGTAAAATGCACTAGCACATTCTGGAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC

CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACGAGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGAGGATCC

Plasmid4: HindIII-EGFP-NotI-Q108K:K40L:T51V:T53C:R58W:T29L:A33W:

Q4F:L117E-EcoRI-3NLS-Stop (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACTTCAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATA
TTGATTTTGCCCTGCGCAAGATTTGGGTACGTCTCACTCAGACGCTGGTTATTGATCAA
GATGGTGATAACTTCAAGGTAAAATGCACTAGCACATTCTGGAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACGAGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGAATTCGCTGACCCCAAGAAGAAGA
GGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGTGA

Plasmid5: HindIII-EGFP-NotI-Q108K:K40L:T51V:T53C:R58W:T29L:A33W:

Q4F:L117E-EcoRI-NES(ELAEKLAGLDIN)-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACTTCAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATA
TTGATTTTGCCCTGCGCAAGATTTGGGTACGTCTCACTCAGACGCTGGTTATTGATCAA

GATGGTGATAACTTCAAGGTAAAATGCACTAGCACATTCTGGAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACGAGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGAATTCGAGCTTGCCGAGAAACTTG
CCGGGCTTGACATAAATTGAGGATCC

Plasmid6: **XhoI-Q108K:K40L:T53S:V62N-BamHI-AgeI-ECFP-Stop-NotI**

(pECFP-N1)

CTCGAGCTCGCCACCATGACGAGGGACCAGAATGGAACCTGGGAGATGGAGAGTAATGA
AAACTTTGAGGGCTACATGAAGGCCCTGGATATTGATTTTGCCACCCGCAAGATTGCAG
TACGTCTCACTCAGACGCTGGTTATTGATCAAGATGGTGATAACTTCAAGACAAAAAGC
ACTAGCACATTCCGCAACTATGATAACGATTTCACTGTTGGAGTAGAGTTTGACGAGTA
CACAAAGAGCCTGGATAACCGGCATGTTAAGGCACTGGTCACCTGGGAAGGTGATGTCC
TTGTGTGTGTGCAAAAGGGGGAGAAGGAGAACCGCGGCTGGAAGAAGTGGATTGAGGGG
GACAAGCTGTACCTGGAGCTGACCTGTGGTGACCAGGTGTGCCGTCAAGTGTTCAAAAA
GAAGTCGAAGGATCCACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCG
GGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACGAGCTGGACGGC
GACGTAACGGCCACAGGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGC
AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCT
CGTGACCACCCTGACCTGGGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGC
AGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGTACCATCTTC
TTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCT
GGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGC
ACAAGCTGGAGTACAACTACATCAGCCACAACGTCTATATCACCGCCGACAAGCAGAAG
AACGGCATCAAGGCCCACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCT
CGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACA
ACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCAC
ATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTA
CAAGTAAAGCGGCCGC

Plasmid7: HindIII-EGFP-NotI-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-

BamHI-RVASL-Stop (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA

AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACCAGAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTGGATGAAGGCCCTGGATA
TTGATTTTGCCACCCGCAAGATTGCAGTACGTCTCACTCAGACGCTGGTTATTGATCAA
GATGGTGATAACTTCAAGGTAAAAAGCACTAGCACATTCTGGAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACGAGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGGATCCCGGGTGGCATCCCTGTGA

Plasmid8: **HindIII-EGFP-NotI-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-**

EcoRI-3NLS-Stop (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACCAGAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTGGATGAAGGCCCTGGATA
TTGATTTTGCCACCCGCAAGATTGCAGTACGTCTCACTCAGACGCTGGTTATTGATCAA
GATGGTGATAACTTCAAGGTAAAAAGCACTAGCACATTCTGGAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACGAGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGAATTCGCTGACCCCAAGAAGAAGA
GGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGTGA

Plasmid9: **HindIII-EGFP-NotI-Q108K:K40L:T51V:T53S:R58W:Y19W:L117E-**

EcoRI-NES(ELAEKLAGLDIN)-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG

ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCCCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACCAGAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTGGATGAAGGCCCTGGATA
TTGATTTTGCCACCCGCAAGATTGCAGTACGTCTCACTCAGACGCTGGTTATTGATCAA
GATGGTGATAACTTCAAGGTAAAAAGCACTAGCACATTCTGGAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACGAGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGAATTCGAGCTTGCCGAGAAACTTG
CCGGGCTTGACATAAATTGAGGATCC

Plasmid10: HindIII-EGFP-NotI-Q108K:K40E:T53A:R58L:Q38F:Q4F-EcoRI-3NLS-Stop (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACTTCAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATA
TTGATTTTGCCACCCGCAAGATTGCAGTACGTCTCACTTTTACGGAAGTTATTGATCAA
GATGGTGATAACTTCAAGACAAAAGCCACTAGCACATTCCTGAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACCTGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGAATTCGCTGACCCCAAGAAGAAGA
GGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGTGA

Plasmid11: HindIII-EGFP-NotI-Q108K:K40E:T53A:R58L:Q38F:Q4F-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA

CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACTTCAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATA
TTGATTTTGCCACCCGCAAGATTGCAGTACGTCTCACTTTTACGGAAGTTATTGATCAA
GATGGTGATAACTTCAAGACAAAAGCCACTAGCACATTCCTGAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACCTGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGAGGATCC

Plasmid12: HindIII-Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F-EcoRI-3NLS-Stop (pFlag-CMV2)

AAGCTTATGACGAGGGACTTCAATGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGA
GGGCTGGATGAAGGCCCTGGATATTGATTTTGCCACCCGCAAGATTGCAGTACGTCTCA
CTCAGACGCTGGTTATTGATCAAGATGGTGATAACTTCAAGGTAAAAAGCACTAGCACA
TTCTGGAACTATGATGTGGATTTCACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAG
CCTGGATAACCGGCATGTTAAGGCACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTG
TGCAAAAGGGGGAGAAGGAGAACCGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTG
TACGACGAGCTGACCTGTGGTGACCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGA
ATTCGCTGACCCCAAGAAGAAGAGGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGGACC
CCAAGAAGAAGAGGAAGGTGTGA

VII.9 Mammalian cell culture

All cell lines (HeLa, U2OS and COS-7) were cultured in Dulbecco's Modified Eagle medium (DMEM, supplemented with phenol red, 4.5 g/L D-glucose, L-glutamine and 110 mg/L Sodium Pyruvate; purchased from Sigma-Aldrich) supplemented with 10% (v/v) Fetal Bovine Serum (FBS, BioWest) and 1x Penicillin-Streptomycin-Glutamine (PSG; purchased from GIBCO) at 37 °C within a 5% CO_2 and 10% O_2 atmosphere.

For microscopic imaging, cells were seeded on an *ibidi 1 μ-Slide 8 well ibiTreat* plate. After approximately 12 h the cells were transiently transfected using Genjet Ver. II (purchased from SignaGen) according to the manufacturer's protocol. Following transfection (after 48 h), the media was removed and the cells were incubated with a media containing the fluorophore. To prepare the fluorophore solution, a stock solution in DMSO (~0.003 M) was warmed to room temperature. The exact concentration of the fluorophore stock solution was determined using UV-Vis (using the associated extinction coefficient). The stock solution was then diluted to the specified concentration with pre-heated (37 °C) DMEM.

After the indicated fluorophore incubation period, the cells were washed two times with Dulbecco's Phosphate Buffered Saline (DPBS, supplemented with calcium chloride and magnesium chloride; purchased from Sigma Aldrich) and incubated in RPMI-1640 medium (without phenol red) for imaging immediately.

VII.10 General confocal imaging methods

Before imaging, the slide lid was replaced with a DIC lid for *μ*-dishes (purchased from Ibidi). 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. hCRBPII/MCRA was imaged with 594 nm excitation, 594 nm primary dichroic, 545 nm secondary dichroic and LP 615 nm emission. EGFP was imaged using 488 nm excitation, 488 nm primary dichroic, 545 nm secondary

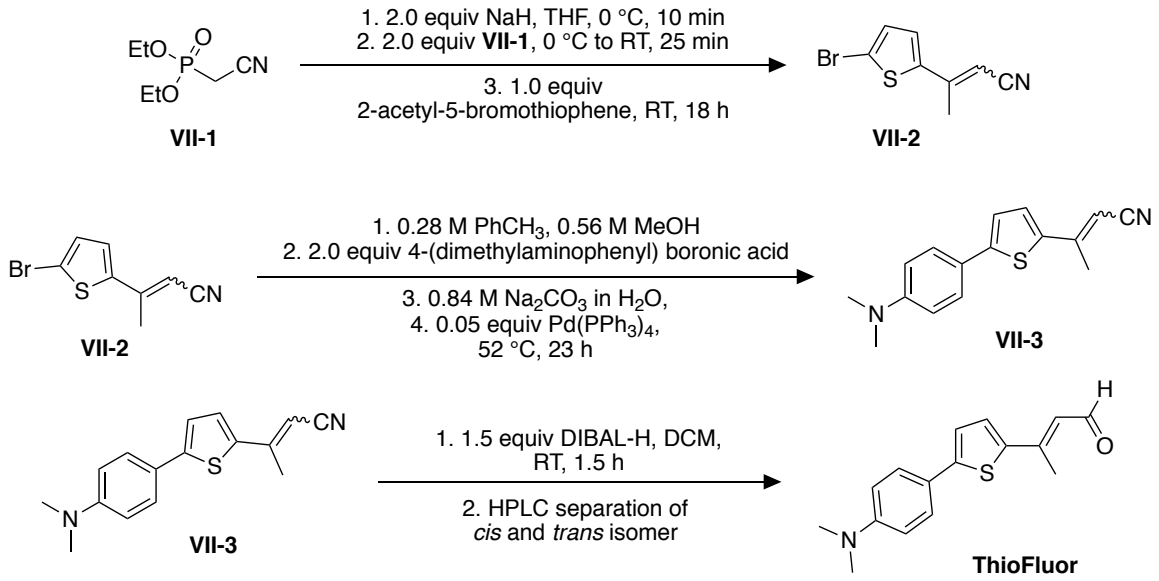
dichroic and BP 505-530 nm emission. ECFP was imaged using 458 nm excitation, 458 nm primary dichroic, 515 nm secondary dichroic and BP 475 – 525 nm emission.

hCRBPII/ThioFluor was imaged using multiple settings, as indicated in the main text: 1) 594 nm excitation, 594 nm primary dichroic, 545 nm secondary dichroic and LP 615 nm (or LP 650 nm) emission; 2) 633 nm excitation, 633 nm primary dichroic, 545 nm secondary dichroic and LP 650 nm emission; 3) 543 nm excitation, 543 nm primary dichroic, 545 nm secondary dichroic and LP 615 nm (or LP 650 nm) emission; 4) 514 nm excitation, 514 nm primary dichroic, 545 nm secondary dichroic and LP 615 nm (or LP 650 nm) emission; 5) 405 nm excitation, 405 nm primary dichroic, 545 nm secondary dichroic and LP 615 nm emission. DICII images were also collected. Kalman averaging 8 was applied in all confocal images. Fluorescence in each experiment was normalized to the same intensity adjusting the gain and amplifier offset. All images are pseudo-colored.

VII.11 Synthesis of chromophores

VII.11.1 Synthesis of ThioFluor

The synthesis of ThioFluor is shown in **Scheme VII-1**. The synthesis of ethyl phosphonate **VII-1** was described previously. 2-Acetyl-5-bromothiophene, (CAS number 5370-25-2) and 4-(dimethylaminophenyl) boronic acid (CAS number 28611-39-4) were purchased from Oakwood Chemical.



Scheme VII-1. Detailed synthesis of **ThioFluor**.

*Synthesis of compound **VII-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 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 min, gradually warming to room temperature. 2-Acetyl-5-bromothiophene (1.0 g, 4.9 mmol) in anhydrous THF (10 mL) was added dropwise. The reaction was kept under nitrogen and stirred for 18 h.

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 **VII-2** and the starting 2-acetyl-5-bromothiophene. This product was carried to the next step.

*Synthesis of compound **VII-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 **VII-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 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 h.

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 **ThioFluor**:* Compound **VII-3** (0.2 g) was added to a dry round bottom flask, flushed with nitrogen and dissolved in anhydrous dichloromethane (30 mL). Subsequently, 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 h.

Work up: The reaction was quenched by slowly adding five drops of water, followed by five drops of 10% HCl, followed by stirring for 15 min. 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 was 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 with 50% dichloromethane/hexane 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 **ThioFluor** was approximately 4% (red solid).

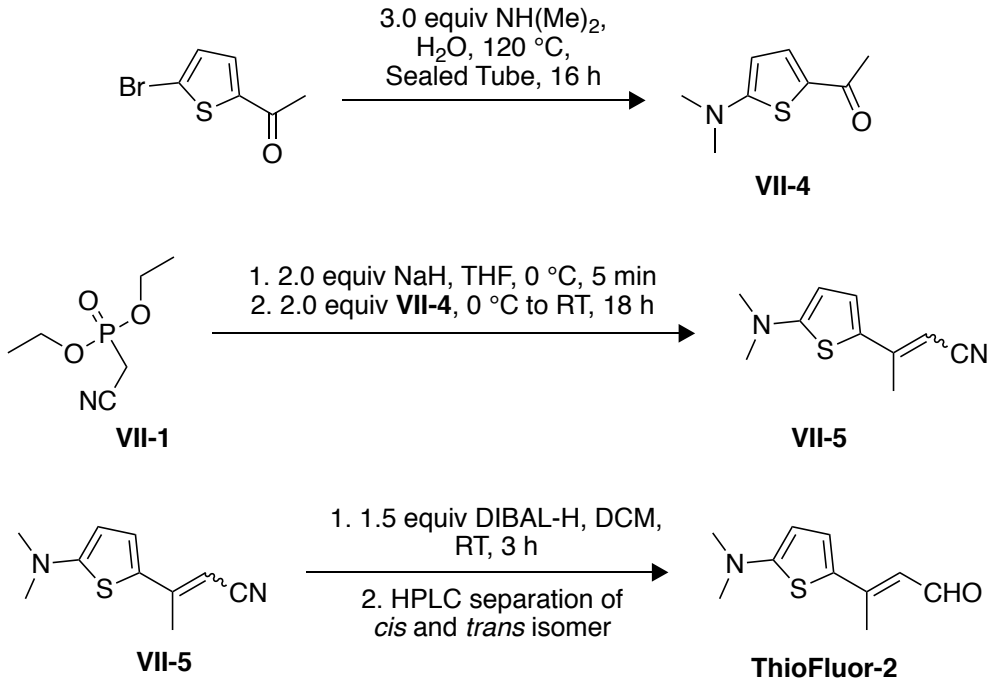
^1H NMR (500 MHz, $CDCl_3$): δ 10.10 (d, *J* = 8.0 Hz, 1H), 7.51 (d, *J* = 8.5 Hz, 2H), 7.38 (d, *J* = 4.0 Hz, 1H), 7.15 (d, *J* = 3.5 Hz, 1H), 6.71 (d, *J* = 9.0 Hz, 2H), 6.43 (d, *J* = 7.5 Hz, 1H), 3.01 (s, 6H), 2.56 (s, 3H)

^{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

HRMS (ES^+): m/z calcd for $C_{16}H_{17}NOS$: 272.1109, found: 272.1069 $[M+H]^+$.

VII.11.2 Synthesis of ThioFluor-2

The synthesis of **ThioFluor-2** is shown in **Scheme VII-2**. The synthesis of ethyl phosphonate **VII-1** was described previously. 2-Acetyl-5-bromothiophene, (CAS number 5370-25-2) was purchased from Oakwood Chemical.



Scheme VII-2. Detailed synthesis of **ThioFluor-2**.

*Synthesis of compound **VII-4**:* 2-Acetyl-5-bromothiophene (1.0 g, 5.0 mmol) was added to a sealed tube. Then *N,N*-dimethyl amine (40%, 2.0 mL, 15.0 mmol) and DI water (5 mL) were added. Subsequently, the reaction mixture was heated at reflux for 16 h.

Work up: Water was added to the reaction mixture and the resulting aqueous layer was extracted three times with dichloromethane (3 x 30 mL). The combined organic phase was dried over anhydrous Na_2SO_4. The solvent was then removed under reduced pressure and the crude product was carried to the next step.

*Synthesis of compound **VII-5**:* NaH (60% by weight, 0.40 g suspension containing 0.24 g NaH, 10.0 mmol) was placed in a dry round bottom flask and flushed with nitrogen. Anhydrous THF (25 mL) was added and then the solution was cooled to 0 °C via an ice-water bath and stirred for five min. Subsequently,

diethyl(cyanomethyl)phosphonate (1.8 g, 10.0 mmol) was added dropwise and stirred for 10 min, gradually warming to room temperature. Compound **VII-4** (0.85 g, 5.0 mmol) in 6 mL anhydrous THF was added dropwise. The reaction was kept under nitrogen and stirred for 18 h.

Work up: Water was added to the reaction mixture and the resulting aqueous layer was extracted three times with diethyl ether (3 x 40 mL). The combined organic phase was dried over anhydrous Na_2SO_4. The solvent was then removed under reduced pressure to yield a yellow-orange solid. The crude product was purified by silica gel column chromatography with 20% ethyl acetate/hexane mixture to yield **VII-5** and the starting 2-acetyl-5-bromothiophene. This product was carried to the next step.

*Synthesis of **ThioFluor-2**:* Compound **VII-5** (0.5 g) was added to a dry round bottom flask, flushed with nitrogen and dissolved in anhydrous dichloromethane (30 mL). Subsequently, a 1 M solution of DIBAL-H in THF (5.2 mL, 5.2 mmol) was added dropwise. The reaction was stirred at room temperature, under nitrogen, for 3 h. An additional 2 mL of 1 M DIBAL-H in THF was added and the reaction was stirred for 1.5 h.

Work up: The reaction was quenched by slowly adding 1 mL of water. The reaction was stirred vigorously for five min. Next, 3 mL of 10% HCl was added, stirred for five min, and an additional 20 mL of brine solution was added and the aqueous layer was extracted with dichloromethane (3 x 40 mL). The resulting organic layer was dried over anhydrous Na_2SO_4 and the solvent was

removed under reduced pressure. The crude product was initially purified by silica gel column chromatography using 10% ethyl acetate/hexane to 30% ethyl acetate/hexane (using a gradient of 10% ethyl acetate/hexane). The crude product after the second column was dissolved in 80% 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 30% ethyl acetate/hexane at 3 mL min^{-1}. The product was detected at 420 nm. The overall yield of the *trans* isomer of **ThioFluor-2** was 6% (brown solid).

^{1}H NMR (500 MHz, $CDCl_3$): δ 9.95 (d, J = 8.5 Hz, 1H), 7.25 (d, J = 4.5 Hz, 1H), 6.12 (d, J = 7.5 Hz, 1H), 5.85 (d, J = 4.5 Hz, 1H), 3.04 (s, 6H), 2.47 (s, 3H)

^{13}C NMR (125 MHz, $CDCl_3$): δ 189.72, 163.04, 150.98, 131.62, 126.50, 119.12, 103.3, 42.08, 14.82

HRMS (ES^+): m/z calcd for $C_{10}H_{13}NOS$: 196.0796, found: 196.0749 $[M+H]^+$.

REFERENCES

REFERENCES

1. Wang, W. in PhD Thesis, Michigan State University (2012).

2. 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).

3. Gill, S.C. & von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry* **182**, 319-326 (1989).

4. 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).

5. Yapici, I. in PhD Thesis, Michigan State University (2015).

6. Berbasova, T. in PhD Thesis, Michigan State University (2014).