DEVELOPMENT OF NOVEL FLUORESCENT PROTEIN TAGS FOR NO-WASH LIVE-CELL IMAGING WITH MINIMUM FLUORESCENT BACKGROUND

By

Rahele Esmatpour Salmani

A DISSERTATION

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

Chemistry—Doctor of Philosophy

2022

ABSTRACT

DEVELOPMENT OF NOVEL FLUORESCENT PROTEIN TAGS FOR NO-WASH LIVE-CELL IMAGING WITH MINIMUM FLUORESCENT BACKGROUND

By

Rahele Esmatpour Salmani

Recent fluorescence microscopy technologies have revolutionized many areas of biomedical research. Nonetheless, high brightness, far-red/near infra-red emission, deep tissue penetration, and selective fluorescent imaging with the minimum background are among the most desired novel fluorescent labeling. One of our primary goals is to develop flexible fluorescent protein tags capable of being tailored *ad infinitum.* We successfully demonstrated the ability to fine-tune the absorption and emission spectra of protein-bound chromophores over an unprecedented wide range (~200 nm). In contrast to intrinsically fluorescent proteins that are always "ON" in our systems, fluorescent is activated upon covalent binding of ligand and the target protein leading to temporal control of fluorescence. However, the fluorescence background from unbound free chromophore and non-specific binding has always been a deep concern in fluorescent labeling. This Ph.D. research aimed to develop novel protein-based fluorescent tags emitting in the far-red/NIR region of the spectrum for no-wash background-free live-cell imaging applications. This was accomplished by coupling novel synthetic fluorogenic chromophores with hCRBPII mutants. Unbound free aldehyde **ThioPhenol** and **CyThioPhenol** are non-emissive dyes that become highly fluorescent upon imine formation with an active site lysine residue engineered deep in the hCRBPII cavity. We created a hydrogen-bonding network around the **ThioPhenol** hydroxyl group through rational protein engineering that facilitates its deprotonation

upon photoexcitation. On the other hand, engineering the target protein to maintain a high iminium pK_a resulted in Protonated Schiff Base (PSB) formation. The resultant complex experiences a strong intramolecular charge transfer (ICT), leading to fluorescence and a large bathochromic shift in the emission (~700 nm). The designed protein-based photoacid provides an unprecedented spatiotemporal control for no-wash bright NIR imaging. Our most recent report demonstrated that hCRBPII/chromophore complexes could be developed as a photobase where the imine is converted to an iminium upon photoexcitation. In the course of optimizing hCRBPII to promote ESPT of the hydroxyl group, we discovered that **ThioPhenol** is capable of acting as both a photoacid and a photobase upon a single photoirradiation. When bound as a Schiff base (SB) to protein mutants that maintain a low iminium pK_a (~5), engineered to deprotonate the hydroxyl group, a dual ESPT process leads to protonation of the imino to iminium (the photobase) and deprotonation of the hydroxyl to alkoxide (the photoacid). This double ESPT feature is recapitulated in a protein-ligand micro-environment, yielding bright protein-dye complexes with unapparelled large pseudo-Stokes shifts (~250 nm). Additionally, the double ESPT **ThioPhenol**/hCRBPII complexes show fast binding rates (half-life of <3 min) that were successfully used to visualize whole-cell and the nucleus as a fluorogenic tag without any washing steps. Currently, further modifications are in progress to optimize the double ESPT systems with **CyThioPhenol** and further *in-vivo* applications.

Dedicated to my parents, my family and Mehdi for their endless love, sacrifice, and support.

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude and appreciation to my Ph.D. advisor, professor Babak Borhan for his endless patience, motivation, enthusiasm, and immense knowledge. Babak has been a tremendous mentor, being supportive and giving me the freedom to pursue various research projects. I want to thank him for encouraging my research and allowing me to grow as a scientist. I hope I'd be able to emulate his open-mindedness, creativity, and scientific rigor in my own career. He was more than just an advisor or professor to me, and numerous other students welcomed us at his home at Thanksgiving, Christmas, Persian, or Chinese New Year. These gatherings, along with some traditions unique to his lab, such as group camping in summer, fireworks on the 4th of July, and birthday cakes for everyone, helped me get through years far from my family. I cannot finish this part without mentioning how a great cook he is, being an expert in making the best Persian kabobs. Thanks for always being more than helpful; your advice on both research as well as on my career has been invaluable.

I want to express my appreciation and gratitude to Dr. Chrysoula Vasileiou, one of the kindest people I know. She was a great help to me. I will never forget the hours I spent in her office asking questions or just talking about many different things. She always welcomed me warmly and advised me like an older sister. I will miss her and our lab gathering at her office. I want to thank our former lab member Dr. Wei Sheng who guided me throughout my Ph.D. studies and always had new fascinating ideas about our research, and I always enjoyed talking to him about

chemistry. I am also very thankful to Dr. Elizabeth Santos, Dr. Hadi Gholami, and Dr. Setare Tahmasebi Nick, who did not hesitate to teach me their skills and share their valuable knowledge and research experience.

Our collaborators had a pivotal role in the success of many projects I would like to express my gratitude to Professor Geiger and his lab members, especially Dr. Alireza Ghanbarpour for his great work acquiring the crystal structures. Professor Gary Blanchard, and Dr. Wenjing Wang for their profitable share on the double ESPT project. I am very grateful to my committee members Professor Xuefei Huang, Professor Robert Maleczka and especially Professor James Geiger not only for their time and extreme patience, but for their intellectual contributions and support on my doctoral training and useful guidance and comments on my research projects. In addition, many multi-colored imaging microscopies would not have been possible without the guidance and expertise of Dr. Melinda Frame. A huge thank to my former and current lab mates and all my other friends, for their support, inspiration, and all fun we had during this journey.

Finally, I want to deeply thank my parents, and sibling for their unconditional trust, endless patience, and selfless encouragement that let me explore new directions in life and seek my own destiny.

I never can thank Mehdi enough; he is my best friend with an amazing unique personality, he made me happier than ever I imagined I could be. Without his positive sense of humor and sunny optimism it would be much more difficult. I am grateful to him because he has given up so much to make my career

successful. He has shared this entire journey with me and has seen me through the entire ups and down of the graduate life, so it only seems right that I dedicate this milestone to him.

TABLE OF CONTENTS

LIST OF TABLES xii

LIST OF FIGURES xvii

KEY TO ABBREVIATIONS xxvi

CHAPTER I: A BRIEF OVERVIEW OF METHODS DEVELOPED FOR FLUORESCENT LABELING AND NEAR-INFRARED IMAGING TAGS……**1**

I.1 Fluorescent proteins 1

I.1.1 Excited state proton transfer and the origin of intrinsic fluorescence in GFP……. 3

I.2 Far-red near-infrared synthetics fluorescent dyes 8

I.3 Extrinsically fluorescent proteins and site-specific labeling methods..... 10

REFERENCES..... 14

CHAPTER II: DEVELOPING FAR-RED/NEAR-INFRARED DYE-HCRBPII FLUORESCENT TAGS FOR NO-WASH BACKGROUND-FREE LIVE CELL IMAGING APPLICATIONS..... 29

II.1 Initial work towards developing hCRBPI-based fluorescent tags..... 30

II.2 Desired features of a practical fluorescence imaging tag 34

II.3 Attempts to minimize non-specific fluorescent background 35

II.4 No-wash background-free fluorogenic imaging tag design..... 38

II.5 Spectroscopic properties of **ThioPhenol** in solution..... 40

II.6 General protein host properties for this study 43

II.7 Primary observation of dual fluorescence in hCRBPII/**ThioPhenol** 44

II.8 The effect of rigidifying the chromophore through Y19W and A33W mutations on the absorption wavelength and Φ_{ESPT} 50

II.9 Exploring the effects of L117E and L117D mutations on the iminium pK_a and Φ_{ESPT}……. 52

II.9.1 Basic residues around the hydroxyl group..... 54

II.9.2 Acidic residues around the hydroxyl group 57

II.10 Investigating the effects of proton acceptor residues at positions A33 and F16 on the Φ_{ESPT}……..58

II.11 Attempts to enhance the red-shifted emission quantum efficiency 66

II.11.1 Through expansion of water-mediated hydrogen bonding network around the R58H residue 66

II.11.1 A. T53 mutations 69

II.11.1 B. S55 mutations 71

II.11.1 C. Q38 mutations 73

II.11.2 Enhancing the ESPT emission quantum efficiency through increasing the iminium pK_a 78

II.11.2 A. L117 mutation 81

II.11.2 B. Q4 and T51 mutation 88

II.11.2 C. K40 Mutation 89

II.11.3 Through expanding water-mediated network around the hydroxyl group……. 92

II.11.3 A. L77 mutation 93

II.11.3 B. T29 mutation 96

II.11.3 C. T53 mutation 97

II-12 **ThioPhenol**/hCRBPII binding kinetics 101

II.13 Visualization of hCRBPII/**ThioPhenol** in mammalian cells 102

II.13.1 **ThioPhenol**/**M3** successfully labels hCRBPII in HeLa cells 103

II.14 Conclusion and future research directions 105

REFERENCES 109

CHAPTER III: DOUBLE EXCITED-STATE PROTON TRANSFER PHENOMENON: DEVELOPING PROTEIN TAGS WITH APPLICATIONS IN NO-WASH FLUORESCENT IMAGING 122

III.1 Previous work toward developing large Stokes shift fluorescent protein-based photobases.. 124

III.2 Designing hCRBPII/dye complexes that undergo double ESPT upon a single excitation….. 130

III.2.1 General protein host properties required for a double ESPT system……..... 131

III.3 **ThioPhenol**-hCRBPII as a photobasic system 132

III.4 Developing **ThioPhenol**-hCRBPII complexes showing both photoacidic and photobasic characteristics 136

III.5 Attempts to increase the fluorescence quantum efficiency of DESPT derived emission……..... 140

III.5.1 Expanding the water-mediated hydrogen bonding network around the hydroxyl group 141

III.6. Application of the double ESPT systems as a fluorescent tag for live cells imaging……. 148

III.7 Enhancing the double ESPT process with brighter chromophore cores.... 151

III.7.1 **MR0**…..... 152

III.7.1.A. Spectroscopic properties of **MR0** in solution 152

III.7.1.B. Attempts to develop double ESPT process with **MR0**-hCRBPII complex……..... 159

III.7.2 **MR1**…..... 169

III.7.2.A. Spectroscopic properties of **MR1** in solution 170

III.7.2.B. Exploration about photoacidic and photobasic properties of **MR1** complexes… ..... .174

III.7.2.C. Developing protein-based double ESPT systems with **MR1** 186

III.8 **ThioPhenol** structure modification and its application in multicolor imaging……. 192

III.8.1 Spectroscopic properties of **CyThioPhenol** in solution 193

III.8.2 Developing **CyThioPhenol** photoacidic complexes 197

III.8.2. A. Kinetic measurements of **CyThioPhenol** binding hCRBPII Photoacid mutants 203

III.8.2. B. Visualization of hCRBPII/**CyThioPhenol** in mammalian cells……..... 204

III.8.3 Developing the double ESPT systems with **CyThioPhenol**-hCRBPII complexes…. 209

III.8.4 Application of **CyThioPhenol** in multicolor no-wash live-cell confocal imaging……. 220

III.8.4 A. Dual hCRBPII labeling expressed in HeLa cells..... 222

III.9 Conclusion and future research plans 224

REFERENCES..... 228

CHAPTER IV: MATERIALS AND METHODS 235

IV.1 Site-directed mutagenesis of hCRBPII and CRABPII..... 235

IV.2 hCRBPII and CRABPII expression and purification in pET-17b expression plasmids……. 245
IV.3 Protein characterization 248
IV.3.1 UV/Vis and fluorescence spectroscopy 248
IV.3.2 Extinction coefficient determination 249
IV.3.3 pK_a measurements of hCRBPII/chromophore complexes 259
IV.3.4 hCRBPII/chromophore binding Kinetic measurements 260
IV.3.4.A. Pseudo-first-order binding rate measurement 260
IV.3.4.B. Second-order binding rate measurement 261
IV.3.5. Absolute fluorescent quantum yield measurements 262
IV.4 Cloning to mammalian expression vectors 262
IV.4.1 General cloning protocol 262
IV.4.2 Sequences of plasmids described in this thesis 264
IV.5 Mammalian cell culture and transfection 271
IV.6 General methods for confocal imaging 272
IV.7 General synthetic procedures 274
IV.7.1 Synthesis of **Thiophenol** 275
IV.7.2 Synthesis of **ThioPhenol**-CF_3 278
IV.7.3 Synthesis of **ThioPhenol**-OMe_2 279
IV.7.4 Synthesis of **MR0** 280
IV.7.5 Synthesis of **MR1** 283
IV.7.6 Synthesis of **Me-TR1** 286
IV.7.7 Synthesis of **CyThioPhenol** 290
REFERENCES 296

LIST OF TABLES

Table II-1. Spectroscopic characterization of **ThioPhenol** in various solvents. . 40

Table II-2. Spectroscopic characterization of **ThioPhenol** and derivatives..... 43

Table II-3. Spectroscopic change as a result of mutating R58 in Q108K:K40L:T51V:T53S template. 46

Table II-4. Spectroscopic change upon the addition of Y19W and A33W mutations. 51

Table II-5. Spectroscopic change upon addition of L117E mutation. 53

Table II-6. Spectroscopic properties of the protein mutants designed to increase the photoacidity of the bound **ThioPhenol** by introducing basic residues. 56

Table II-7. Spectroscopic properties of the protein mutants designed to increase the photoacidity of the bound **ThioPhenol** by introducing acidic residues. 58

Table II-8. Spectroscopic change upon mutation of A33 to histidine, serine and glutamic acid, and tyrosine. 59

Table II-9. Spectroscopic change upon F16Y addition. 61

Table II-10. Spectroscopic changes as a result of mutating F16. 63

Table II-11. Spectroscopic changes as a result of mutating R58, F16, and A33 residues to tyrosine. 65

Table II-12. Spectroscopic changes as a result of mutating T53 to histidine, tyrosine, and glutamic acid. 69

Table II-13. Spectroscopic changes as a result of mutating T53. 70

Table II-14. Protein expression yields of hCRBPII mutants upon S55 mutation to histidine, tyrosine, and glutamic acid. 72

Table II-15. Spectroscopic properties of the PSB as a result of mutation of S55. 72

Table II-16. Spectroscopic properties of the PSB as a result of mutation of Q38. 74

Table II-17. Spectroscopic change as a result of mutation Q38 and T53 mutation to Glu..... 77

Table II-18. List of mutants that resulted in insoluble protein expression upon introduction of Glu resides in the proximity of iminium..... 81

Table II-19. Spectroscopic properties of protein mutants upon mutation of L117 in order to increase the iminium pK_a..... 82

Table II-20. Spectroscopic properties as the result of L117M mutation. 85

Table II-21. Spectroscopic properties of the SB as the result of Q4 mutation.... 88

Table II-22. Spectroscopic change as the result of the introduction of T51M mutation. 89

Table II-23. Spectroscopic properties as the result of K40 mutation. 91

Table II-24. Spectroscopic change as the result of L77 mutations. 93

Table II-25. Spectroscopic and pK_a change as the result of L77Y introduction.. 94

Table II-26. Spectroscopic change as the result introduction of T29Y mutation. 96

Table II-27. Spectroscopic properties as the result of introduction of cysteine in positions T51 and T53. 97

Table II-28. Spectroscopic properties of **ThioPhenol**/**M3** complex. 103

Table III-1. Spectroscopic properties in complexation with **M3**. 132

Table III-2. Spectroscopic properties in complexation with **M1**. 136

Table III-3. Spectroscopic properties of **ThioPhenol**/**M3** and **M4** complexes. complexation with **M1**. 139

Table III-4. Spectroscopic properties upon addition of F16Y, A33H and T29Y 141

Table III-5. Spectroscopic properties as the result of substituting 58H with 58Y. 142

Table III-6. Spectroscopic properties of R58H containing mutants. 145

Table III-7. Spectroscopic properties of **ThioPhenol**/**M4** complex.....148

Table III-8. Spectroscopic characterization of **MR0** in various solvents. 153

Table III-9. Spectroscopic features of **MR0**, **FR0**, and **ThioPhenol** in toluene and ethanol. 154

Table III-10. Spectroscopic features of free aldehyde **MR0** in water in acidic and basic pH. 156

Table III-11. Spectroscopic characterization of **MR0** and derivatives. 158

Table III-12. Spectroscopic properties of **MR0** Phenol-SB in PBS buffer. 159

Table III-13. Spectroscopic properties of **MR0**/**M3** and **MR0**/**M4**. 160

Table III-14. Spectroscopic properties of **MR0**/**M3** and **MR0**/**M4**, and **MR0**/**M5**. 161

Table III-15. Spectroscopic properties change as the result of the addition of proton acceptor residues. 166

Table III-16. Spectroscopic characterization of **MR1** in various solvents. 170

Table III-17. Comparison of spectroscopic features of free aldehyde **ThioPhenol** and **MR1** in ethanol. 172

Table III-18. Spectroscopic characterization of **MR1** and derivatives. 174

Table III-19. Comparison of the spectroscopic properties of **FR1/M3** and **MR1**/**M3** complexes. 175

Table III-20. Spectroscopic changes as the result of addition of R58H mutation. 176

Table III-21. Spectroscopic properties of **MR1** with high iminium and mutants 179

Table III-22. Spectroscopic properties of **MR1** with high iminium and mutants. 180

Table III-23. Spectroscopic change as the result of L77Y and Q38Y mutations. 182

Table III-24. Comparison of the spectroscopic properties of **FR1/M7** and **MR1**/**M7** complexes. 185

Table III-25. Comparison of spectroscopic properties of **MR1** in complexation with **M3** and **M4** mutants. 186

Table III-26. Spectroscopic change as the result of F77Y and Q38Y mutations. 188

Table III-27. Spectroscopic change as the result of the mutation of residues surrounding the **MR1** hydroxyl group. 190

Table III-28. Spectroscopic characterization of **CyThioPhenol** in various solvents. 194

Table III-29. Spectroscopic characterization of **CyThioPhenol** and derivatives. 195

Table III-30. Spectroscopic change as the result of the addition of tyrosine and histidine residues around the hydroxyl group. 198

Table III-31. Spectroscopic change as the result of the addition of L117C, T29Y, T51C and T53C mutations. 201

Table III-32. Spectroscopic properties of **CyThioPhenol**/**M6** complex compared with **ThioPhenol**/**M3** complex. 206

Table III-33. Spectroscopic properties of **ThioPhenol** and **CyThioPhenol** binding **M3** mutant. 209

Table III-34. Spectroscopic properties of mutants as the result of the introduction of V62E, I42E, and K40D mutations. 213

Table III-35. Spectroscopic properties of **ThioPhenol** and **CyThioPhenol** binding **M4** mutant. 214

Table III-36. Spectroscopic properties of mutants at positions 16, 20, 29, 33, 38, 58, and 77. 216

Table III-37. Spectroscopic properties upon mutating the K40 residue. 218

Table III-38. Spectroscopic properties of **CyThioPhenol**/**M6** complex compared with **CyThioPhenol**/**M3** complex. 220

Table IV-1. PCR cycling conditions for site-directed mutagenesis. 235

Table IV-2. FPLC Source 15Q method. 247

Table IV-3. FPLC SEC method. 247

Table IV-4. Extinction coefficients of hCRBPII and CRABPII mutants. 250

Table IV-5. PCR cycling conditions for cloning..... 263

LIST OF FIGURES

Figure I-1. A representative set of intrinsically fluorescent proteins and the corresponding chromophore structures responsible for fluorescence…… 2

Figure I-2. Proposed mechanism for the maturation of the GFP chromophore... 4

Figure I-3. **a**. Equilibrium between the neutral (phenol) species and the anionic phenolate species of wt-GFP. **b**. Normalized fluorescence excitation (dashed line) and emission (solid line) spectra for wild-type GFP. Excitation at either peak (~400 and 480 nm) leads to the characteristic green emission at 510 nm..... 5

Figure I-4. Proposed water hydrogen bonding network for the excited state proton transfer in GFP..... 7

Figure I-5. Schematic self-labeling reactions of SNAP-tag and Halo Tag. POI: protein of interest. The red star represents the conjugated fluorophore……... 11

Figure I-6. Design of the PYP-tag mutant PYP3R and its fluorogenic probe, with a focus on electrostatic interactions and the pK_a value of the leaving group.[110].... 12

Figure II-1. **a**. Crystal structure of Q108K:K40L-hCRBPII mutant complexed with all-*trans*-retinal. Key residues engineered to regulate the absorption wavelength of the bound retinylidene are highlighted (PDB 4EXZ). The scheme shows the iminium (PSB) formation between lysine 108 and retinal aldehyde. **b**. Protein solution of different hCRBPII mutants incubated with all-*trans* retinal…….. 31

Figure II-2. **a**. Structure of **MCRA** and its PSB formation with active Lysine 108 of hCRBPII. **b**. Spectroscopic properties of free aldehyde (blue) and PSB (red). .. 33

Figure II-3. Structures of Dapoxyl dye and **ThioFluor**……....37

Figure II-4. Non-specific fluorescent background originating from off-target iminium formation in labeled HeLa cells with 10 μM **ThioFluor** (incubated at 37 °C for 1 hour) expressing hCRBPII-EGFP-3NLS. NLS = nuclear localization sequence. 37

Figure II-5. Structures of free aldehyde **ThioPhenol**, **ThioPhenol**-imine, **ThioPhenol**-iminium, and **ThioPhenolate**-iminium..... 39

Figure II-6. Spectroscopic properties of **ThioPhenol** in different solvents. **a**. UV-Vis and **b**. Fluorescence spectra of **ThioPhenol**…… 41

Figure II-7. **a**. Schiff base and protonated Schiff base (PSB) of **ThioPheno**l with *n*-butyl amine in ethanol. **b**. Normalized spectra of **ThioPhenol** and derivatives in ethanol. Absorbance (left) and emission (right)……..... 42

Figure II-8. Flexible docking of **ThioPhenol** in the crystal structure of Q108K:K40L:T51V:T53S:R58W-hCRBPII/**ThioFluor**. The distance between the hydroxyl group and the indole nitrogen of 58W side chain is 3.9 Å…….. 45

Figure II-9. **a**. **ThioPhenolate**-PSB, the product of ESPT process, and normalized absorption/emission spectra of the iminium of Q108K:K40L:T51V:T53S:R58H/**ThioPhenol** complex. **b**. UV-Vis (left) and fluorescence (right) spectra of the same complex upon acidification of the solution. 47

Figure II-10. Formation of the **ThioPhenolate**-SB upon basification of the solution and the UV-Vis (left) and fluorescence (right) spectra of Q108K:K40L:T51V:T53S:R58H/**ThioPhenol** upon base titration…….. 49

Figure II-11. Absorption and emission spectra of KLVS:R58W:Y19W:L117E:A33W/**ThioPhenol** complex upon PSB excitation. . 54

Figure II-12. Closest residues to **ThioPhenol**'s hydroxyl group in the crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioPhenol** complex……... 55

Figure II-13. Hydrogen bonding network responsible for excited state proton transfer in LSS-mKate 1 and LSS-mKate 2…….. 57

Figure II-14. The crystal structure of Q108K:K40L:T51V:T53S:R58H:Y19W:A33Y/**ThioPhenol** complex with F16 mutated to tyrosine via mutagenesis in Pymol……. 60

Figure II-15. Schematic representation of the negative charge localization on the oxygen atom due to the electrostatic interaction with Tyr 16 side chain…….. 62

Figure II-16. **a**. Overlay of Q108K:K40L:T51V:T53S:R58W/**ThioPheno**l (magenta) and Q108K:K40L:T51V:T53S:R58H:Y19W:A33Y/**ThioPhenol** (green). **b**. The distance between the F16Y and the hydroxyl group. 64

Figure II-17. **a**. *In silico* modeling of R58H in the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioPheno**l and the distances between S55, Q38, and T53S residues to 58H. **b**. T53S distances to R58H residue and **ThioPhenol** hydroxyl group……. 67

Figure II-18. Crystal structures of WT-hCRBPII bound with all-*trans*-retinol showing the internal hydrogen bonding network among the T51 and T53 side chain residues and retinol……... 68

Figure II-19. Water mediated hydrogen bonding between Q38 and Q128 in the crystal structure of hCRBPII-Q108K:K40L/retinal PDB 4EXZ……73

Figure II-20. **a**. Overlay of Q108K:K40L:T51V:T53S:R58W/**ThioPhenol** (green) and Q108K:K40L/retinal (pink). **b**. Water network between T53S, Q38, and Q128 in Q108K:K40L:T51V:T53S:R58W/**ThioPhenol** crystal structure. 75

Figure II-21. **a**. UV-Vis (left) and fluorescence (right) spectra of Q108K:K40L:T51V:T53S:R58H:A33H:F16Y:Q38E/**ThioPheno**l upon acid titration. **b**. Excitation of the shoulder indicated by the black arrow results in the same emission as PSB excitation……... 76

Figure II-22. **a**. An overlay of crystal structures of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioPhenol** (purple) and **ThioFluor** (green) showing *cis* and *trans* iminium conformation, respectively. **b**. The distance between L117D residue and the iminium nitrogen atom for **ThioFluor** complex (left) and **ThioPhenol** (right)……79

Figure II- 23. Highlighted residues mutated to glutamic acid in order to interact with the iminium to increase its pK_a……80

Figure II-24. UV-Vis (left) and fluorescence (right) spectra of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y/**ThioPhenol** (red) and Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117T/**ThioPhenol** (blue).. 84

Figure II-25. **a**. Water mediated hydrogen bonding between Q4 and the iminium, and the distance between L117C and the iminium nitrogen atom in the crystal structure of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C/**ThioPhenol** complex. **b**. The different trajectory of chromophore shown upon the overlay of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C/**ThioPhenol** (green) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D: Q4F/**ThioPhenol** (purple) with F16 residue mutated to Try in the latter (shown in grey)…… 87

Figure II-26. The distance between the K40L residue and the iminium nitrogen atom in the crystal structure of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C/**ThioPhenol**. 90

Figure II-27. **a**. Cartoon representation of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C/**Thiophenol** complex

with T29 and L77 residues highlighted. **b**. Detailed hydrogen bonding network surrounding T29, A33H, R58H and the hydroxyl group in the same crystal…….95

Figure II-28. **a** and **b**. The absorption and emission spectra of Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51C:T53S:R58H:L117C/**ThioPhenol** complex. The shoulder corresponds to PSB-Phenol complex since excitation of both results to the same emission. **c**. The acid titration and **d**. the base titration of the same complex……. 99

Figure II-29. Comparison of absorption (left) and emission (right) of **M1** (T51VT:53S), **M2** (T51C:T53S), **M3** (T51V:T53C), and **M4** (T51M:T53S) protein/**ThioPhenol** complexes……. 100

Figure II-30. Rate of **ThioPhenol**/**M3** PSB formation, fitted to 2nd order kinetics with 20 μM protein and 10 μM **Thiophenol**. Plotted is the concentration of free chromophore vs. time. 101

Figure II-31. Spectroscopic properties of Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53C:R58H:L117C mutant with **ThioPhenol** including UV-Vis and fluorescence spectra (left) and pK_a titration (right)……. 102

Figure II-32. Maps of the EGFP-hCRBPII-SP fusion constructs. SP: signaling peptides. SP = 3×NLS (nuclear localization sequence), NES (nuclear export sequence), and CAAX (prenylation tag). 103

Figure II-33. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-*3NLS*. NLS = nuclear localization sequence. Cells were stained with 10 μM **ThioPhenol** and incubated at 37 °C for 1 h and 30 min. Cells were not washed before imaging……. 107

Figure II-34. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-*NES*. NES = nuclear export localization sequence. Cells were stained with 10 μM **ThioPhenol** and incubated at 37 °C for 1 h and 30 min. Cells were not washed before imaging……108

Figure III-1. Model for a large Stoke shift (LSS) vs. a standard fluorescence system. Large Stokes shift red-shifted emission results via conversion of the excited state species **A*** to a bathochromically distinct molecule **B***…….. 124

Figure III-2. **a**. **FR0** structure and formation of **FR0**-SB and **FR0**-PSB. **b**. UV-Vis (left) and fluorescence (right) spectra of **FR0**-SB (blue) and **FR0**-PSB (red) in ethanol…….. 125

Figure III-3. **a**. Formation of the imine and iminium of **ThioFluor** and **FR1** in ethanol. **b**. Spectroscopic properties of **FR0** (blue), **FR1** (red), and **ThioFluor** (green) in ethanol. UV-Vis (left) and fluorescence spectra upon SB excitation (right)..... 127

Figure III-4. UV-Vis and fluorescence spectra of **M3**/**ThioFluor** complex (394 nm excitation), exhibiting ESPT emission (left) vs. UV-Vis and fluorescence spectra of **M1**/**ThioFluor** complex (380 nm excitation) leads to SB emission at 474 nm (right). The spectra have been measured in PBS buffer at neutral pH (7.2)…….. 128

Figure III-5. **M3**/**FR1** imaging in live HeLa cells. NES = nuclear export sequence. Cells were stained with 500 nM **FR1** and incubated at 37 °C for 1 min. Cells were washed three times with DPBS before imaging. Scale bar, 10 μm……. 129

Figure III-6. **a**. Formation of **ThioPhenolate**-PSB complex through a double ESPT process. **b**. The internal charge transfer resulted from the **ThioPhenolate**-PSB complex, the product of the double ESPT process……..131

Figure III-7. UV-Vis, emission spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioPhenol** complex……133

Figure III-8. **a**. UV-Vis (left) and fluorescence (right) spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioPhenol** complex upon acidification of the solution. **b**. UV-Vis (left) and fluorescence (right) spectra of Q108K:K40D:T53A:R58L:Q38F:Q4F/**ThioFluor** complex upon acidification of the solution..... 134

Figure III-9. UV-Vis (left) and fluorescence (right) spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioPhenol** complex upon basification of the solution……... 135

Figure III-10. Normalized absorption and emission spectra of Q108K:K40E:T53A:R58H:Q38F:Q4F/**ThioPhenol** complex, when excited at 370 nm at neutral pH (~7.2)……... 137

Figure III-11. Schematic representations of single and double ESPT processes. The starting complex is different, while the product is the same. 138

Figure III-12. **a**. UV-Vis spectra of Q108K:K40E:T53A:**R58H**:Q38F:Q4F/**ThioPhenol** complex upon acid titration of the sample. **b**. Fluorescence spectra of the same complex upon SB excitation at neutral pH and PSB excitation in acidic pH. 139

Figure III-13. UV-Vis spectra of Q108K:K40E:T53A:**R58H**:Q38F:Q4F/**ThioPhenol** complex upon base titration (left). Fluorescence spectra of the same complex upon SB excitation in different pH..... 140

Figure III-14. The crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFlour** and the closest residues to the nitrogen atom of the dimethyl amino moiety of **ThioFluor**..... 144

Figure III-15. Products of single ESPT and double ESPT processes. Absorption and emission spectrum of Q108K:K40E:T53A:Q38F:Q4F:R58H:L77M/**ThioPhenol** upon SB excitation at 370 nm..... 146

Figure III-16. Absorption (left) and emission spectra (right) of Q108K:K40E:T53A:Q38F:Q4F:R58H:F16Y:M20Y/**ThioPhenol** complex. Upon excitation of SB (blue line) and excitation of the PSB (red line). DESPT= Double ESPT. SESPT= Single ESPT. 147

Figure III-17. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-*whole cell*. Cells were stained with 10 μM **ThioPhenol** and incubated at 37 °C for 5 min. Cells were not washed before imaging..... 150

Figure III-18. Structures of **ThioPhenol**, **MR0**, and **FR0**..... 152

Figure III-19. Spectroscopic properties of **MR0** in different solvents. **a**. UV-Vis and **b**. Fluorescence spectra. 154

Figure III-20. UV-Vis spectra (left) and fluorescence spectra (right) of free aldehyde **MR0** upon acidification and basification of the sample. Measurements were done in water..... 155

Figure III-21. **a**. Schiff base and protonated Schiff base (PSB) of **MR0** with *n*-butyl amine in ethanol. **b**. Absorbance (left) and emission (right) spectra of **MR0** and derivatives: Phenol-SB, Phenol-PSB, and Phenolate-PSB. 157

Figure III-22. Absorption (left) and emission (right) spectra of **M3** (**a**), **M4** (**b**), and **M5** (**c**) upon basification of their solutions. 162

Figure III-23. Formation of the hydrolyzed Phenolate-SB complex produced upon basification of the protein solution..... 163

Figure III-24. Absorption (blue line) and emission (red line) of Q108K:K40L:T51V:T53S:R58H:F16Y:Y19W:A33H:L117E/**MR0** complex at neutral (left) and acidic (right) pH..... 164

Figure III-25. The schematic representation of single and double ESPT processes and their photophysical properties……... 167

Figure III-26. Absorption and emission spectra of Q108K:K40E:T53A:Q38F:Q4F:R58H:F16Y/**MR0** complex. The blue-shifted shoulder corresponds to Phenol-PSB, or Phenolate-SB produced upon single ESPT, and the emission at 560 nm results from Phenolate-PSB complex produced upon the double ESPT process……. 168

Figure III-27. Structures of **FR1** and **MR1**…….. 170

Figure III-28. Spectroscopic properties of **MR1** in different solvents. **a**. UV-Vis and **b**. Fluorescence spectra. 171

Figure III-29. **a**. Schiff base and protonated Schiff base (PSB) of **MR1** with *n*-butylamine in ethanol. **b**. Absorbance (left) and emission (right) spectra of **MR1** and derivatives: Phenol-SB, Phenol-PSB, and Phenolate-PSB…….....173

Figure III-30. The normalized absorption (blue line) and emission (red line) spectra of **MR1**/**M3** complex……175

Figure III-31. The absorption and emission spectra of Q108K:K40L:T51V:T53S:R58H/**MR1** complex **a**. upon SB excitation in neutral pH 7.2 and **b**. PSB excitation in acidic pH 5.4……...177

Figure III-32. Proposed photoswitching cycle of Q108K:K40L:T51V:T53S:R58H/**MR1** complex……... 178

Figure III-33. Flexible docking of **MR1** in the crystal structure of Q108K:K40L:T51V:T53S:R58W-hCRBPII/**ThioFluor**. **MR1** is shown in purple, and **ThioFluor** is shown in cyan……... 181

Figure III-34. Absorption (blue line) and emission (red line) spectra of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y: L117C:L77Y/**MR1** complex upon excitation of **a**. SB at 364 nm and **b**. PSB at 470 nm in neutral pH……...183

Figure III-35. Absorption (left) and emission (right) spectra of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y: L117C:L77Y/**MR1** upon acid and base titrations……... 184

Figure III-36. The absorption and emission spectra of Q108K:K40E:T53A:Q38F:Q4F:**R58H/MR1** complex measured at neutral pH...187

Figure III-37. The absorption and emission spectra of Q108K:K40E:T53A:Q38F:Q4F**:**R58H:F16Y:Q38Y/**MR1** complex measured at neutral pH. 188

Figure III-38. Docked **MR1** in the crystal structure of Q108K:K40L:T51V:T53S:R58W-hCRBPII/**ThioFluor**, and the closest residues around the hydroxyl group…….. 189

Figure III-39. **a**. Structures of **ThioPhenol** and **CyThioPhenol**. **b**. Absorbance (left) and emission (right) spectra of **CyThioPhenol** in different solvents…..... 193

Figure III-40. Absorbance (left) and emission (right) spectra of **CyThioPhenol** and …derivatives: Phenol-SB, Phenol-PSB, and Phenolate-SB…….. 195

Figure III-41. Schematic comparison of ICT systems in Phenolate-PSB and Phenolate-SB complexes of **CyThioPhenol**…….. 197

Figure III-42. **a**. The absorption and **b**. the emission spectra of KLVSH is equal to Q108K:K40L:T51V:T53S:R58H:F16Y:L77Y/**CyThioPhenol** complex measured in neutral pH 7.3. **c**, **d**. The absorption and emission spectra of the same complex screened upon acidification pH 5.8 and basification pH 8.3 of the solution….. 199

Figure III-43. The absorption spectra of Q108K:K40L:T51V:T53C:R58H:Y19W:F16Y:T29Y:L77Y:L117C/**CyThioPhenol** complex recorded upon basifying the solution from pH 7.3 to pH 9.5. 202

Figure III-44. Rate of **CyThioPhenol**/**M6** PSB formation fitted to second-order kinetics with 20 μM protein and 10 μM **CyThioPhenol**. Plotted is the concentration of free chromophore vs. time……. 204

Figure III-45. Spectroscopic properties of Q108K:K40L:T51V:T53C:R58H:Y19W:A33H:F16Y:L77Y:L117C mutant with **CyThioPhenol** including UV-Vis and fluorescence spectra (left) and the pK_a titration (right)……... 205

Figure III-46. Schematic map of EGFP-hCRBPII-NES fusion construct. NES: nuclear export sequence……..... 207

Figure III-47. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-NES. NES: nuclear export sequence. Cells were stained with 5 μM **CyThioPhenol** and incubated at 37 °C overnight. Cells were not washed before imaging..... 208

Figure III-48. The absorption and emission spectra of **a**. **ThioPhenol/M3** (red lines) and **b**. the absorption and emission spectra of **CyThioPhenol/M3** complex (blue lines)..... 210

Figure III-49. The absorption and emission spectra of Q108K:**K40L**:T53A:Q38F:Q4F:R58L/**CyThioPheno**l complex measured in neutral pH..... 211

Figure III-50. The crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFlour** and highlighted V62, I42, and K40E residues..... 212

Figure III-51. Absorption and emission spectra of Q108K:K40E:T53A:Q38F:Q4F:R58H/**CyThioPhenol** complex measured in neutral pH. 214

Figure III-52. The absorption and emission spectra of **CyThioPhenol** upon binding **a**. Q108K:K40E:T53A:Q38F:Q4F:R58H:A33H and **b**. Q108K:K40E:Q4F:A33H:Q38F:T53A:R58H:L77Y mutants..... 217

Figure III-53. The absorption and emission spectra of Q108K:**K40D**:T53A:Q38F:Q4F:R58H/**CyThioPhenol** complex upon SB excitation..... 219

Figure III-54. **a**. **M3**/**CyThioPhenol** complex pK_a titration. **b**. The absorption (left) and emission (right) spectra of **M3** (blue) and **M6** (red) mutants upon binding **CyThioPhenol** measured at pH 7.2..... 221

Figure III-55. Schematic map of EGFP-**M6**-hCRBPII-NES and EGFP-**M3**-hCRBPII-3NLS fusion constructs. NLS: nuclear localization sequence and NES: nuclear export sequence..... 222

Figure III-56. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-*3NLS*. NLS = nuclear localization sequence. Cells were stained with 5 mM **CyThioPhenol** and incubated at 37 °C for 5 min. Cells were not washed before imaging..... 226

Figure III-57. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-*3NLS* and EGFP-hCRBPII-*NES*. NLS= nuclear localization sequence. NES= nuclear export sequence. Cells were stained with 5 μM **CyThioPhenol** and incubated at 37 °C for overnight and another 5 μM **CyThioPhenol** incubated at 37 °C for 5 minutes. Cells were not washed before imaging..... 227

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, 10^{-10} meter

Abs, λ_{abs} absorbance, absorption wavelength maximum

ACQ aggregation-caused quenching

AIE aggregation-induced emission

AIEgen aggregation-induced emission luminogen/fluorogen

aq. Aqueous

BFP blue fluorescent protein

BODIPY boron dipyrromethene

BP bandpass (filter)

°C degree Celsius

CAAX prenylation tag

CFP cyan fluorescent protein

cm centimeter

cm^{-1} wavenumber

CP chromoprotein

CRABPII cellular retinoic acid binding protein II

CuAAC CuI-catalyzed alkyne-azide cycloaddition

Da Dalton

DESPT double excited state proton transfer

DIBAL diisobutylaluminium hydride

DIC differential interference contrast

DMEM Dulbecco's Modified Eagle's Medium

DMSO dimethyl sulfoxide

DPBS Dulbecco's phosphate buffered saline

EDG electron donating group

eDHFR Escherichia coli dihydrofolate reductase

EGFP enhanced green fluorescent protein

Em, λ_{em} emission, emission wavelength maximum

equiv equivalent

ESCT excited state charge transfer

ESIPT (**ESiPT**) excited state intramolecular proton transfer

ESPT excited state proton transfer

EWG electron withdrawing group

Ex, λ_{ex} excitation wavelength

FABP fatty acid binding protein

FAP fluorogen-activating proteins

FBS fetal bovine serum

FL fluorescence

FlAsH fluorescein arsenical hairpin binder

FP fluorescent protein

FPLC fast protein liquid chromatography

FRET Förster resonance energy transfer

FRFP far-red fluorescent protein

g gram

GFP green fluorescent protein

h hour

hAGT human O^6-alkylguanine-DNA alkyl transferase

HBA H-bonding acceptor

HBD H-bonding donor

HBR 4-hydroxybenzylidene-rhodanine

hCRBPII human Cellular Retinol Binding Protein II

HOMO highest occupied molecular orbital

Hz hertz

IC internal conversion

ICG indocyanine green

ICT intramolecular charge transfer

IDT Integrated DNA Technologies

IEDDA Inverse-Electron-Demand Diels-Alder cycloaddition

iLBP intracellular lipid binding protein

IPTG isopropyl β-D-1-thiogalactopyranoside

IR infrared

ISC intersystem crossing

K kelvin

k kinetic rate constant

LB lysogeny broth

LP longpass (filter)

LplA lipoic acid ligase

LSS long/large Stokes shift

LUMO lowest unoccupied molecular orbital

M molar

mBeRFP monomer blue light-excited RFP

MBP maltose binding protein

MCRA merocyanine retinal aldehyde

mg milligram

min minute

mmol millimole

mol mole

MS molecular sieve

N.D. not determined

N.O. not observed

NBD nitrobenzoxadiazole

NES nuclear export sequence

NIR near-infrared

NLS nuclear localization sequence

nm nanometer

nM nanomolar

O/N overnight

OD optical density

OFP orange fluorescent protein

***p*-HBI** 4-(p-hydroxy-benzylidene)-5-imidazolinone

PA-FP photoactivatable fluorescent protein

PBS phosphate-buffered saline

PCR polymerase chain reaction

PDB protein data bank

PeT photoinduced electron transfer

PHY phytochrome

PLICT planarized intramolecular charge transfer

POI protein of interest

PPI protein-protein interaction

PSB protonated Schiff base

PSFP photoswitchable FP

PYP photoactive yellow protein

qABP quenched activity-based probe

QY quantum yield

ReAsH resofurin arsenical hairpin binder

RFP red fluorescent protein

ROI region of interest

SB Schiff base

SBR signal-to-background ratio

sec second

SEC size-exclusion chromatography

SNAr nucleophilic aromatic substitution

SNR signal-to-noise ratio

SPAAC strain-promoted alkyne-azide cycloaddition

SQ Source 15Q

SS Stokes shift

T temperature

t time

$\mathbf{t_{1/2}}$ maturation half-life

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

xs excess

Y-FAST Yellow Fluorescence-Activating and absorption-Shifting Tag

YFP yellow fluorescent protein

ε extinction coefficient

µg microgram

μm micrometer

μM micromolar

ν frequency

Φ quantum yield

CHAPTER I: A BRIEF OVERVIEW OF METHODS DEVELOPED FOR FLUORESCENT LABELING AND NEAR-INFRARED IMAGING TAGS

Revolutionary advances in science and technology have greatly enhanced our lives. The development of fluorescent proteins (FPs) and their subsequent use in fluorescent imaging is an excellent example of basic science leading to practical biotechnological and medical applications. Fluorescence-based assays are crucial tools, which enable studying detailed molecular mechanisms, including protein-protein interactions, enzymatic activity, conformational changes, and protein localization.[1-9]

I.1 Fluorescent proteins

Fluorescent proteins (FPs) have empowered researchers to visualize inside living cells, examine tissues and subcellular components, and track biological and cellular events with an unprecedented level of resolution and sensitivity.[10-12] Engineered fluorescent proteins from various species have expanded the color palette from blue to red and far-red spectral regions.[13] Notably, engineered fluorescent proteins from bacterial phytochromes emit around near-infrared (NIR) wavelengths (**Figure I-1**).[14-16]

The green fluorescent protein (GFP) and its color-shifted genetic variants have played an indispensable role in many live-cell imaging experiments. FPs are genetic labels and can be fused to the target proteins without the need for exogenous labeling agents or permeabilization procedures.[17,18] They are

expressed in a 1:1 ratio (if fused to one FP), which leads to extremely high labeling specificity.[19,20]

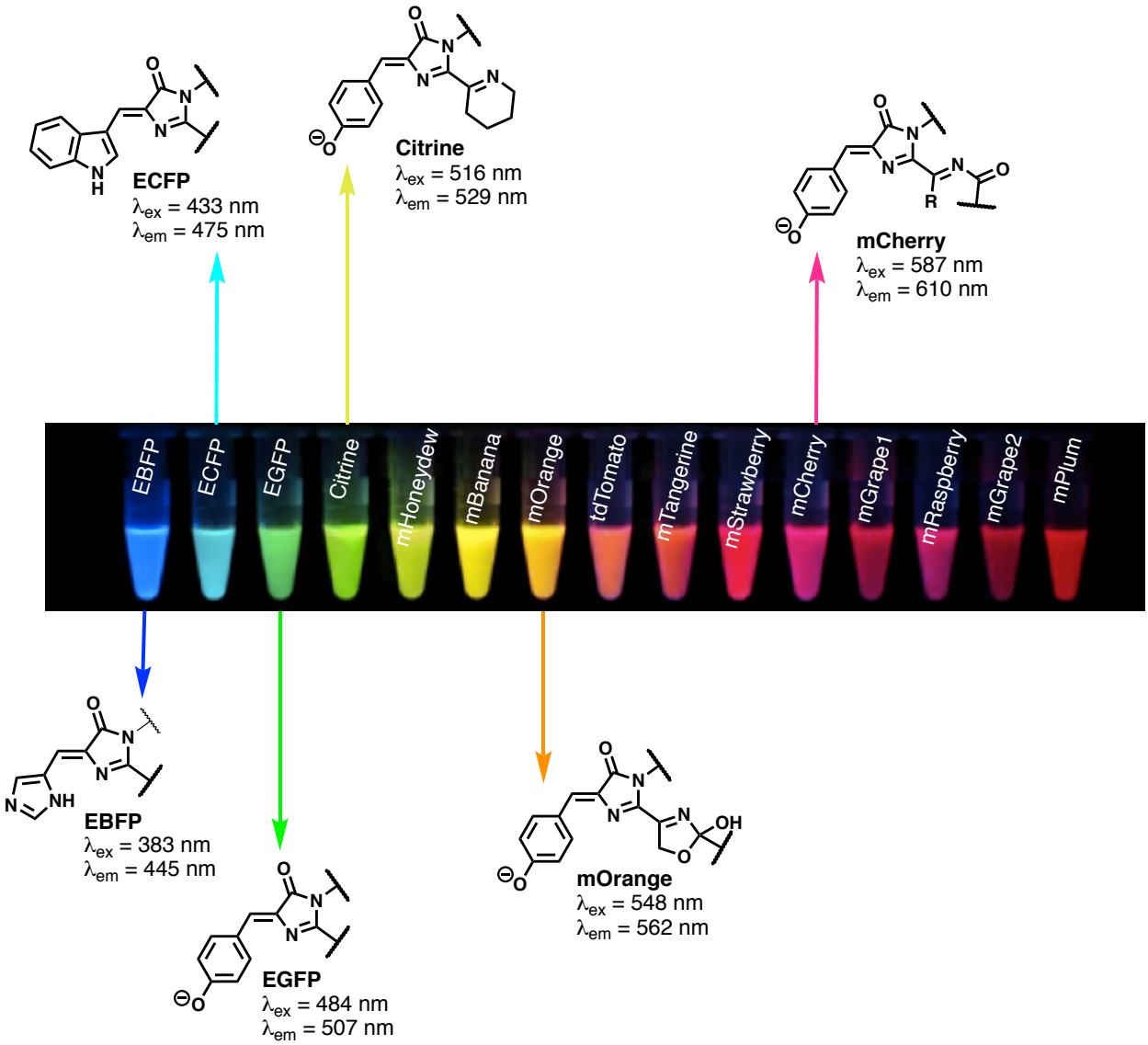


Figure I-1. A representative set of intrinsically fluorescent proteins and the corresponding chromophore structures responsible for fluorescence.

In recent years, red-shifted FPs have become more desired in bioimaging applications because cells are more transparent to far-red/NIR light while effectively absorb in shorter wavelengths.[21-24] In addition, cellular autofluorescence decreases as the excitation wavelengths increases.[24,25] Imaging

with fluorophores that have red-shifted excitation maxima can also benefit sample health. The light of a longer wavelength is less phototoxic to cells and will enable longer acquisition times.[26,27] However, known naturally occurring red fluorescent proteins (RFPs), such as DsRed,[28] and eqFP611,[29] tend to form obligate oligomers, and their usefulness as molecular fusion tags in *in-vivo* model systems are limited.[30,31]

Using directed evolution, significant effort has been put towards monomerizing the RFPs mentioned above. It is noteworthy that all the mFruits, including the most widely used mCherry,[32] tandem dimer tdTomato,[32] and TagRFP,[33] are derived from naturally tetrameric proteins. Random mutagenesis has successfully yielded several far-red (λ_{em} >630 nm) monomeric RFPs such as mPlum,[29] mKate2,[34] and mNeptune.[35]

Despite many advantageous features, FPs do not always exhibit large Stokes shifts and are oxygen-dependent. In addition, some have residual tendencies to oligomerize and aggregate, which can cause localization defects that are detrimental to certain applications.[36-38]

I.1.1 Excited state proton transfer and the origin of intrinsic fluorescence in GFP

As discussed above, FPs do not require an accessory cofactor, external enzymes, or an exogenous substrate to fluoresce. Particularly in green fluorescent protein (GFP), the light-emitting molecular unit, the chromophore, is p-hydroxybenzylidene-2,3-dimethylimidazolinone (HBDI) in its anionic form (**Figure**

I-2). The maturation of the embedded chromophore is accomplished through an autocatalytic process. The most accepted mechanism for the maturation of *p*-HBI, which was acquired based on several crystallographic studies, is depicted in **Figure 1-2**.[39]

As shown below, formation of the chromophore happens in three subsequent steps: 1) internal nucleophilic attack of Gly67 amide nitrogen to the carbonyl carbon of Ser65 that results in the formation of a five-membered imidazolone ring, 2) dehydration of the hemi-aminal to form an imidazolin-5-one intermediate, (3) oxidation of Tyr66 by molecular oxygen that completes the conjugation of the ring system (**Figure 1-2**).[40,41]

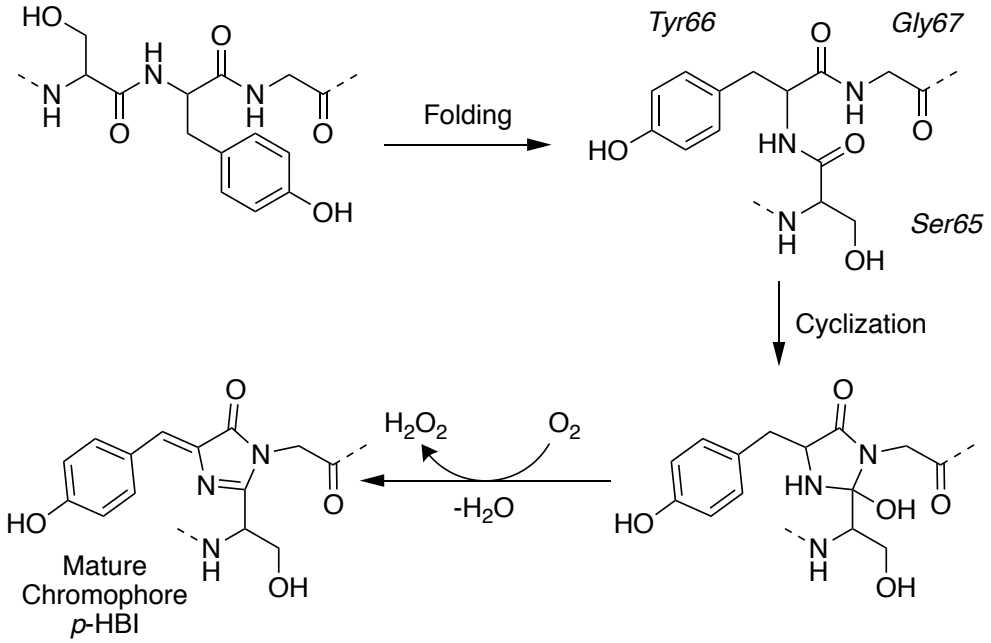


Figure I-2. Proposed mechanism for the maturation of the GFP chromophore.

However, neither unfolded GFP, nor the naked chromophore (*p*-HBI), are fluorescent proving that the stable protein structure is a requirement for efficient fluorescence. Notably, GFP is extremely resistant to protease enzymes and remains stable under very harsh conditions making it exceptionally suitable as a fusion tag in both prokaryotic and eukaryotic organisms.[42-44]

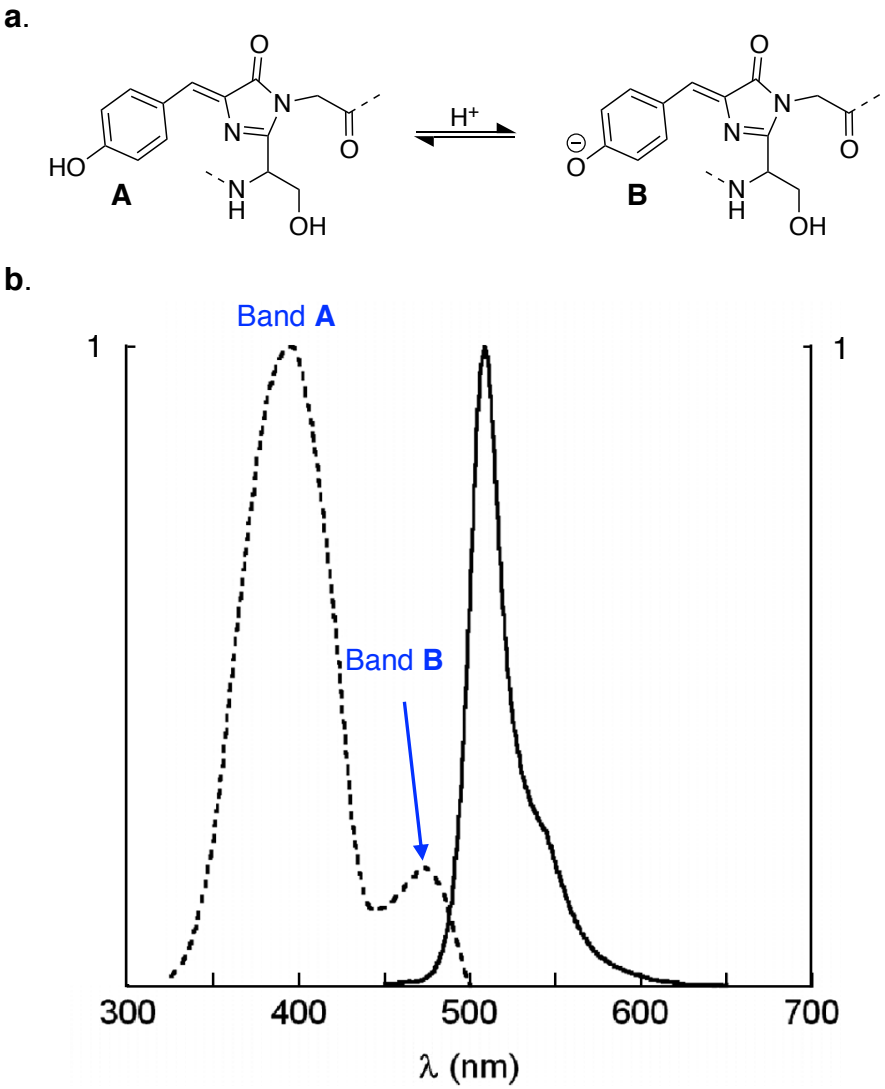


Figure I-3. **a**. Equilibrium between the neutral (phenol) species and the anionic phenolate species of wt-GFP. **b**. Normalized fluorescence excitation (dashed line) and emission (solid line) spectra for wild-type GFP. Excitation at either peak (~400 and 480 nm) leads to the characteristic green emission at 510 nm.

Although several GFP homologs have been isolated from marine organisms, the original GFP isolated from *Aequorea victoria* shows unique and complex excitation/emission spectra. The absorption spectrum of the wild-type green fluorescent protein (wt-GFP) contains two peaks: at 395 nm, referred to as band A, and at 475 nm, referred to as band B.[45] Excitation of either of the two absorption bands leads to emission at 510 nm. This observation suggests an equilibrium exists between two distinct chromophore states within the protein with similar emission wavelengths (**Figure 1-3**).

Boxer and co-workers discovered that the apparent large shift of the emission to longer wavelengths after excitation at band A is due to the deprotonation of the neutral chromophore (phenolate) induced by the excited-state proton transfer (ESPT) process occurring via a hydrogen-bonded network.[46-48] They showed that the light-driven conversion between the neutral (A) and anionic (B) form of the chromophore passes through the intermediate state I (**Figure I-4**).

Moreover, X-ray crystallographic data provided by Remington, Tsien, and co-workers confirmed a hydrogen-bonding network that links the chromophore hydroxyl group and E222. It is proposed that photoexcited A-form (A*) transfers a proton to E222 via a water molecule and S205 (**Figure I-4**).[45] Quantum-classical calculations support these findings and reveal that the phenolic oxygen acidity increases and its pK_a drops by several units upon photoexcitation, which in correlation with the heterocyclic ring, provides an emissive push-pull system.[45,46,49]

GFP is an exemplar of a naturally occurring biological system in which the ESPT phenomenon strongly influences its fluorescence properties.[39,48,50] Notably, the dual excitation behavior of GFP have found applications in designing active biosensors of various cellular phenomena.[51]

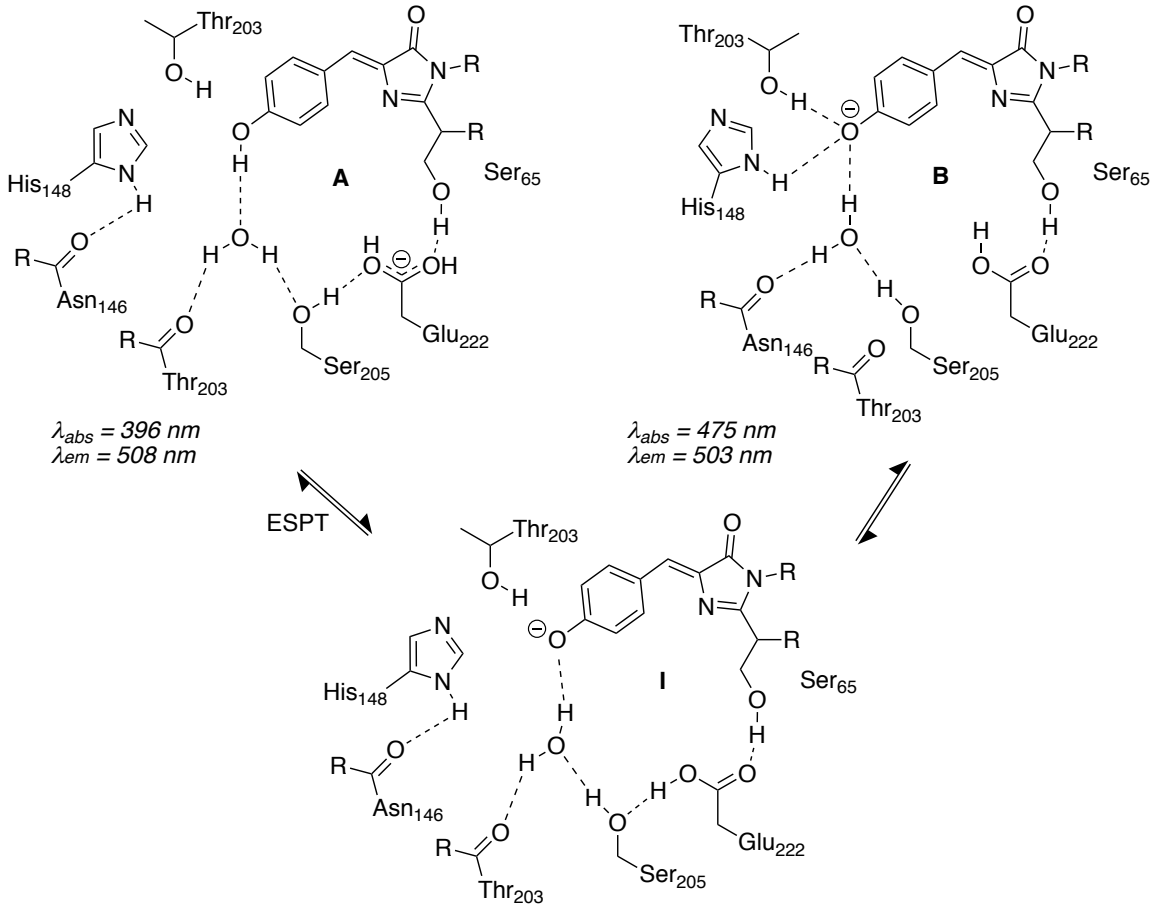


Figure I-4. Proposed water hydrogen bonding network for the excited state proton transfer in GFP.

The ESPT phenomenon has been further adapted to engineer FPs in order to extend the Stokes Shift, such as mKeima, LSSmKates, mBeRFP, CyOFP1, Sandercyanin, and TagRFP675, to name a few.[37,52-57] The chromophore environment of this class of proteins shares noticeable structure similarities; as in

GFP, the hydroxyl group of the chromophore forms a hydrogen-bonding network with active site neighboring residues. The web of hydrogen bonding both stabilizes the neutral form of the chromophore in the ground state and facilitates deprotonation of the tyrosine hydroxyl group through ESPT process in the excited state.

I.2 Far-red near-infrared synthetics fluorescent dyes

Recently, noninvasive optical imaging in the far-red/near infra-red (NIR) region has attracted significant attention due to deeper penetration (approximately 5-20 mm) of biological tissues, reduced autofluorescence background, and improved signal-to-noise ratio.[22,58-62]

Lately, organic fluorescent dyes have become increasingly attractive from a practical perspective because their photophysical properties are well-tunable by various structural modifications.[63-65] Additionally, their small size can be of great benefit, which ensures low interference to the function of the labeled protein. Some organic dyes also offer higher photostability and brightness.[66,67]

Red- and NIR- emitting organic fluorophores belong to various dye families, such as cyanines,[68,69] Si-rhodamines,[70,71] and BODIPY[72,73] derivatives. Most red-emitting fluorophores have extensive π-electron conjugation, usually containing large polycyclic aromatic hydrocarbons. Some other dyes have a push-pull type electronic structure that renders red/far-red fluorescence when placed in a polar milieu and favors positive solvatochromism.

However, oftentimes structural modifications are needed to improve their permeability and solubility. For example, because of intermolecular stacking or attractive dipole-dipole interactions, many organic dyes are prone to form aggregates in aqueous environments, resulting in drastic self-quenching, aggregation-caused quenching (ACQ),[38,74-76] and can significantly impede their applications in many biological systems.[77,78]

Nonetheless, the AIEgen materials, known to have aggregation-induced emission (AIE), are weakly emissive or non-emissive in dilute solutions but become highly emissive in aggregated or solid form and can have potential use in some systems.[79,80] Well-known cyanine dyes also have small Stokes shifts (typically less than 35 nm), which leads to severe self-quenching[81] and cause poor Signal-to-Noise Ratio (SNR) in imaging.[82]

Organic fluorescent dyes do not need maturation half-time that can vary from minutes to hours for FPs; however, the unbound fluorophore and/or nonspecifically bound fluorophores can cause severe fluorescent background. consequently, the use of non-fluorogenic fluorophores that are always “ON” (here defined as always being fluorescent) requires extensive washing steps to remove excess dye prior to imaging.[83]

An alternative approach to overcome this problem is to use fluorogenic dyes. These fluorochromes fluoresce upon binding to the target and are particularly useful for bioimaging.[84,85] Solvatochromicity,[86] pH sensitivity,[87] photoinduced electron transfer (PeT),[88] and other dark state quenching

mechanisms[89,90] are among the current methods employed to produce fluorogenic dyes in order to avoid the need for washing steps.

For example, quenched activity-based probes (qABPs)- they are basically small molecule reporters of enzymatic activity- are commonly designed to compose a ligand-tethered fluorophore and a quencher via a certain linkage that can be cleaved by key catalytic residues of the enzyme with high selectivity. This enzymatic activity-dependent cleavage leads to the release of the quencher moiety and subsequent restoration of reporter fluorescence.[90-92]

Tetrazine-derived fluorogens are another example that can put the fluorophores into quiescence before target binding. However, it requires the expression of non-canonical amino acids chemically modified with olefinic side chains to facilitate the fluorescence activation via sequential inverse electron demand Diels-Alder (IEDDA) reactions and dinitrogen repulsion.[93,94] Despite laborious design efforts to customize quenching mechanisms for a given fluorophore, this fluorogenic approach sometimes suffers the drawbacks such as modest fluorescent enhancement or incomplete fluorescence quenching.

I.3 Extrinsically fluorescent proteins and site-specific labeling methods

The conjugation of small dyes with genetically encodable protein/peptide tags has led to the development of flexible labeling systems that can be tailored *ad infinitum*. These site-selective chemical labeling systems include self-labeling enzymes SNAP/CLIP-tag and HaloTag (**Figure 1-5**),[95,96] short peptide tags that bind biarsenical FlAsH and ReAsH[97,98], PYP-based fluorogenic protein tags

(**Figure 1-6**),[99,100] protein affinity ligands (Y-FAST, TMP, and FAP),[9,101-103] and bioorthogonal tags namely copper-catalyzed alkyne-azide cyclization (CuAAC), strain-promoted alkyne-azide cycloaddition (SPAAC), Staudinger ligation, and other copper-free click reactions.[91,104-106]

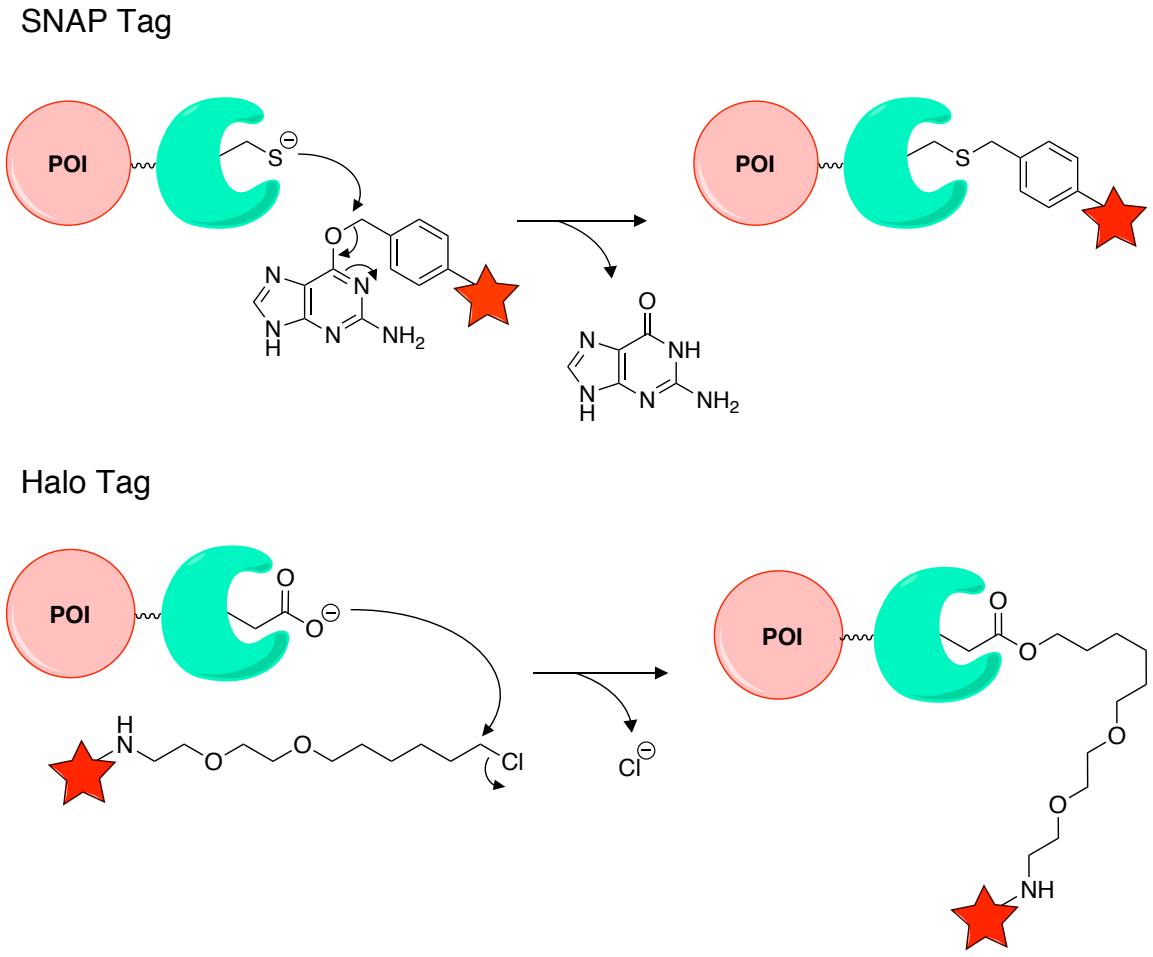


Figure I-5. Schematic self-labeling reactions of SNAP-tag and Halo Tag. POI: protein of interest. The red star represents the conjugated fluorophore.

Combination of the genetical addressability of protein tags and the tunability of small molecules renders a variety of customizable methods that have greatly complemented FPs in bioimaging applications. However, the need for the addition

of exogenous chromophores raises the concern of cell/organelle permeability and cytotoxicity. Nonetheless, the fluorescence background remains challenging unless a fluorogenic tag is used or the unbound chromophore emits in a different spectral region from that of the fluorescent complex.[85,107-109]

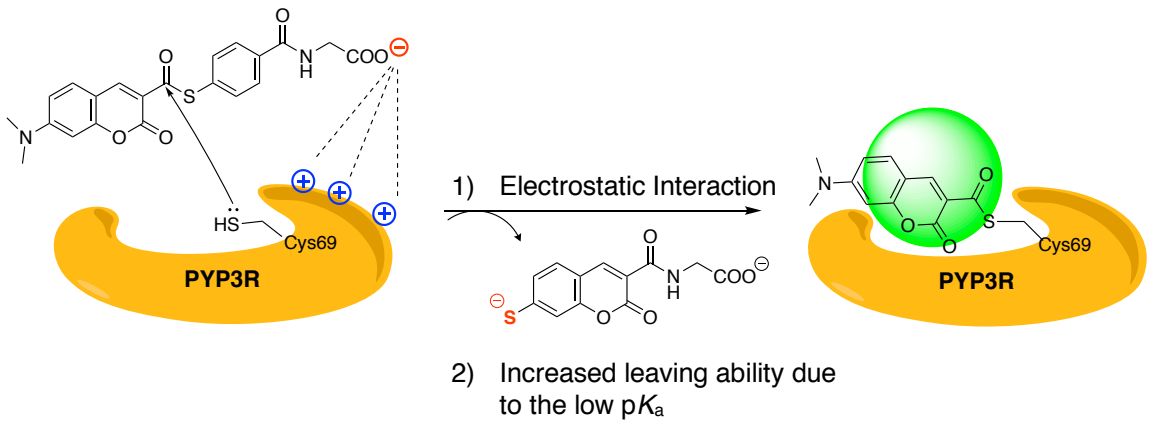


Figure I-6. Design of the PYP-tag mutant PYP3R and its fluorogenic probe, with a focus on electrostatic interactions and the pK_a value of the leaving group.[110]

As discussed earlier, a major attraction to the bioimaging community is the ability to turn "ON" the fluorescence signal on-demand to prevent extensive washing prior to imaging.[111-115] Among all possible triggers, photochemical activation offers precise control over others. Light of specific wavelength can provide rapid, remote, and noninvasive control on photoactive delivery platforms in a spatiotemporal manner.[116] These chemical changes affect the photophysical properties of the chromophore, preferably leading to a bathochromic shift in emission or even a dual-emission,[117] large Stokes shift,[117] and spectral sensitivity of the emitters to the surrounding medium.[118]

Notably, such systems lead to reduced or eliminated background since the fluorescence of corresponding complexes is separated from nonspecific fluorescence.[119-122]

REFERENCES

REFERENCES

1 Theissen, B., Karow, A. R., Kohler, J., Gubaev, A. & Klostermeier, D. Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 548-553, doi:10.1073/pnas.0705488105 (2008).

2 Altman, D., Goswami, D., Hasson, T., Spudich, J. A. & Mayor, S. Precise positioning of myosin VI on Endocytic vesicles in vivo. *Plos Biology* **5**, 1712-1722, doi:10.1371/journal.pbio.0050210 (2007).

3 Bianco, P. R. *et al.* Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature* **409**, 374-378, doi:10.1038/35053131 (2001).

4 Fili, N. *et al.* Visualizing helicases unwinding DNA at the single molecule level. *Nucleic Acids Research* **38**, 4448-4457, doi:10.1093/nar/gkq173 (2010).

5 Miyawaki, A. *et al.* Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. *Nature* **388**, 882-887, doi:10.1038/42264 (1997).

6 Sakamoto, T., Webb, M. R., Forgacs, E., White, H. D. & Sellers, J. R. Direct observation of the mechanochemical coupling in myosin Va during processive movement. *Nature* **455**, 128-U199, doi:10.1038/nature07188 (2008).

7 Li, H. Y. *et al.* Protein-protein interaction of FHL3 with FHL2 and visualization of their interaction by green fluorescent proteins (GFP) two-fusion fluorescence resonance energy transfer (FRET). *Journal of Cellular Biochemistry* **80**, 293-303, doi:10.1002/1097-4644(20010301)80:3<293::Aid-jcb10>3.3.Co;2-l (2001).

8 Tong, H. J. *et al.* Enzymatic Cleavage and Subsequent Facile Intramolecular Transcyclization for in Situ Fluorescence Detection of

gamma-Glutamyltranspetidase Activities. *Analytical Chemistry* **88**, 10816-10820, doi:10.1021/acs.analchem.6b03448 (2016).

9 Plamont, M. A. *et al.* 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, doi:10.1073/pnas.1513094113 (2016).

10 Toomre, D. & Bewersdorf, J. in *Annual Review of Cell and Developmental Biology, Vol 26* Vol. 26 *Annual Review of Cell and Developmental Biology* (eds R. Schekman, L. Goldstein, & R. Lehmann) 285-314 (2010).

11 Aper, S. J. A. *et al.* Colorful Protein-Based Fluorescent Probes for Collagen Imaging. *Plos One* **9**, doi:10.1371/journal.pone.0114983 (2014).

12 Shcherbakova, D. M., Sengupta, P., Lippincott-Schwartz, J. & Verkhusha, V. V. in *Annual Review of Biophysics, Vol 43* Vol. 43 *Annual Review of Biophysics* (ed K. A. Dill) 303-329 (2014).

13 Kremers, G. J., Gilbert, S. G., Cranfill, P. J., Davidson, M. W. & Piston, D. W. Fluorescent proteins at a glance (vol 124, pg 157, 2011). *J. Cell Sci.* **124**, 2676-2676, doi:10.1242/jcs.095059 (2011).

14 Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R. & Piston, D. W. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophysical Journal* **73**, 2782-2790, doi:10.1016/s0006-3495(97)78307-3 (1997).

15 Piatkevich, K. D. *et al.* Near-Infrared Fluorescent Proteins Engineered from Bacterial Phytochromes in Neuroimaging. *Biophysical Journal* **113**, 2299-2309, doi:10.1016/j.bpj.2017.09.007 (2017).

16 Shcherbakova, D. M., Baloban, M. & Verkhusha, V. V. Near-infrared fluorescent proteins engineered from bacterial phytochromes. *Current Opinion in Chemical Biology* **27**, 52-63, doi:10.1016/j.cbpa.2015.06.005 (2015).

17 Thorn, K. Genetically encoded fluorescent tags. *Mol. Biol. Cell* **28**, 848-857, doi:10.1091/mbc.E16-07-0504 (2017).

18 Cardarelli, F. Back to the Future: Genetically Encoded Fluorescent Proteins as Inert Tracers of the Intracellular Environment. *International Journal of Molecular Sciences* **21**, doi:10.3390/ijms21114164 (2020).

19 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, doi:10.1002/anie.201200408 (2012).

20 Smith, E. M. & Mueller, J. D. The statistics of protein expression ratios for cellular fluorescence studies. *European Biophysics Journal with Biophysics Letters* **41**, 341-352, doi:10.1007/s00249-012-0792-x (2012).

21 Bashkatov, A. N., Genina, E. A., Kochubey, V. I. & Tuchin, V. V. Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm. *Journal of Physics D-Applied Physics* **38**, 2543-2555, doi:10.1088/0022-3727/38/15/004 (2005).

22 Frangioni, J. V. In vivo near-infrared fluorescence imaging. *Current Opinion in Chemical Biology* **7**, 626-634, doi:10.1016/j.cbpa.2003.08.007 (2003).

23 Rodriguez, E. A. *et al.* A far-red fluorescent protein evolved from a cyanobacterial phycobiliprotein. *Nature Methods* **13**, 763-769, doi:10.1038/nmeth.3935 (2016).

24 Shcherbo, D. *et al.* Bright far-red fluorescent protein for whole-body imaging. *Nature Methods* **4**, 741-746, doi:10.1038/nmeth1083 (2007).

25 Diao, S. *et al.* Biological imaging without autofluorescence in the second near-infrared region. *Nano Research* **8**, 3027-3034, doi:10.1007/s12274-015-0808-9 (2015).

26 Hong, G. S., Antaris, A. L. & Dai, H. J. Near-infrared fluorophores for biomedical imaging. *Nat. Biomed. Eng* **1**, 22, doi:10.1038/s41551-016-0010 (2017).

27 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, doi:10.1073/pnas.97.22.11984 (2000).

28 Wiedenmann, J. *et al.* A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor (Anthozoa, Actinaria). *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11646-11651, doi:10.1073/pnas.182157199 (2002).

29 Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp red fluorescent protein. *Nature Biotechnology* **22**, 1567-1572, doi:10.1038/nbt1037 (2004).

30 Costantini, L. M., Fossati, M., Francolini, M. & Snapp, E. L. Assessing the Tendency of Fluorescent Proteins to Oligomerize Under Physiologic Conditions. *Traffic* **13**, 643-649, doi:10.1111/j.1600-0854.2012.01336.x (2012).

31 Piatkevich, K. D., Efremenko, E. N., Verkhusha, V. V. & Varfolomeev, D. D. Red fluorescent proteins and their properties. *Russian Chemical Reviews* **79**, 243-258, doi:10.1070/RC2010v079n03ABEH004095 (2010).

32 Merzlyak, E. M. *et al.* Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature Methods* **4**, 555-557, doi:10.1038/nmeth1062 (2007).

33 Shcherbo, D. *et al.* Far-red fluorescent tags for protein imaging in living tissues. *Biochemical Journal* **418**, 567-574, doi:10.1042/bj20081949 (2009).

34 Lin, M. Z. *et al.* Autofluorescent Proteins with Excitation in the Optical Window for Intravital Imaging in Mammals. *Chemistry & Biology* **16**, 1169-1179, doi:10.1016/j.chembiol.2009.10.009 (2009).

35 Costantini, L. M. & Snapp, E. L. Fluorescent proteins in cellular organelles: serious pitfalls and some solutions. *DNA Cell Biol* **32**, 622-627, doi:10.1089/dna.2013.2172 (2013).

36 Snapp, E. Design and use of fluorescent fusion proteins in cell biology. *Curr Protoc Cell Biol* **Chapter 21**, 21 24 21-21 24 13, doi:10.1002/0471143030.cb2104s27 (2005).

37 Yanushevich, Y. G. *et al.* A strategy for the generation of non-aggregating mutants of Anthozoa fluorescent proteins. *Febs Letters* **511**, 11-14, doi:10.1016/s0014-5793(01)03263-x (2002).

38 Kwok, R. T. K., Leung, C. W. T., Lam, J. W. Y. & Tang, B. Z. Biosensing by luminogens with aggregation-induced emission characteristics. *Chemical Society Reviews* **44**, 4228-4238, doi:10.1039/c4cs00325j (2015).

39 Zimmer, M. Green fluorescent protein (GFP): Applications, structure, and related photophysical behavior. *Chemical Reviews* **102**, 759-781, doi:10.1021/cr010142r (2002).

40 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, doi:10.1073/pnas.91.26.12501 (1994).

41 Zhang, L. P., Patel, H. N., Lappe, J. W. & Wachter, R. M. Reaction progress of chromophore biogenesis in green fluorescent protein. *Journal of the American Chemical Society* **128**, 4766-4772, doi:10.1021/ja0580439 (2006).

42 Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509-544, doi:10.1146/annurev.biochem.67.1.509 (1998).

43 Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Circular permutation and receptor insertion within green fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 11241-11246, doi:10.1073/pnas.96.20.11241 (1999).

44 Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A. & Getzoff, E. D. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12111-12116, doi:10.1073/pnas.2133463100 (2003).

45 Brejc, K. *et al.* 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, doi:10.1073/pnas.94.6.2306 (1997).

46 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, doi:10.1073/pnas.93.16.8362 (1996).

47 Stoner-Ma, D. *et al.* An alternate proton acceptor for excited-state proton transfer in green fluorescent protein: Rewiring GFP. *Journal of the American Chemical Society* **130**, 1227-1235, doi:10.1021/ja0754507 (2008).

48 Shu, X. *et al.* An alternative excited-state proton transfer pathway in green fluorescent protein variant S205V. *Protein Science* **16**, 2703-2710, doi:10.1110/ps.073112007 (2007).

49 Palm, G. J. *et al.* The structural basis for spectral variations in green fluorescent protein. *Nat. Struct. Biol.* **4**, 361-365, doi:10.1038/nsb0597-361 (1997).

50 Seward, H. E. & Bagshaw, C. R. The photochemistry of fluorescent proteins: implications for their biological applications. *Chemical Society Reviews* **38**, 2842-2851, doi:10.1039/b901355p (2009).

51 Giepmans, B. N. G., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. Review - The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217-224, doi:10.1126/science.1124618 (2006).

52 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, doi:10.1021/ja101974k (2010).

53 Piatkevich, K. D. *et al.* 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, doi:10.1073/pnas.0914365107 (2010).

54 Yang, J. *et al.* mBeRFP, an Improved Large Stokes Shift Red Fluorescent Protein. *Plos One* **8**, doi:10.1371/journal.pone.0064849 (2013).

55 Chu, J. *et al.* A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo. *Nature Biotechnology* **34**, 760-+, doi:10.1038/nbt.3550 (2016).

56 Ghosh, S. *et al.* Blue protein with red fluorescence. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 11513-11518, doi:10.1073/pnas.1525622113 (2016).

57 Piatkevich, K. D. *et al.* Extended Stokes Shift in Fluorescent Proteins: Chromophore-Protein Interactions in a Near-Infrared TagRFP675 Variant. *Scientific Reports* **3**, doi:10.1038/srep01847 (2013).

58 Lavis, L. D. & Raines, R. T. Bright Building Blocks for Chemical Biology. *Acs Chemical Biology* **9**, 855-866, doi:10.1021/cb500078u (2014).

59 Shuhendler, A. J., Pu, K. Y., Cui, L., Uetrecht, J. P. & Rao, J. H. Real-time imaging of oxidative and nitrosative stress in the liver of live animals for drug-toxicity testing. *Nature Biotechnology* **32**, 373-U240, doi:10.1038/nbt.2838 (2014).

60 Pu, K. Y. *et al.* Semiconducting polymer nanoparticles as photoacoustic molecular imaging probes in living mice. *Nature Nanotechnology* **9**, 233-239, doi:10.1038/nnano.2013.302 (2014).

61 Han, M. Y., Gao, X. H., Su, J. Z. & Nie, S. Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nature Biotechnology* **19**, 631-635, doi:10.1038/90228 (2001).

62 Lei, Z. *et al.* Bright, Stable, and Biocompatible Organic Fluorophores Absorbing/Emitting in the Deep Near-Infrared Spectral Region. *Angewandte Chemie-International Edition* **56**, 2979-2983, doi:10.1002/anie.201612301 (2017).

63 Jun, J. V., Petersson, E. J. & Chenoweth, D. M. Rational Design and Facile Synthesis of a Highly Tunable Quinoline-Based Fluorescent Small-

Molecule Scaffold for Live Cell Imaging. *Journal of the American Chemical Society* **140**, 9486-9493, doi:10.1021/jacs.8b03738 (2018).

64 Lavis, L. D. & Raines, R. T. Bright ideas for chemical biology. *Acs Chemical Biology* **3**, 142-155, doi:10.1021/cb700248m (2008).

65 Urano, Y. *et al.* Evolution of fluorescein as a platform for finely tunable fluorescence probes. *Journal of the American Chemical Society* **127**, 4888-4894, doi:10.1021/ja043919h (2005).

66 Toseland, C. P. Fluorescent labeling and modification of proteins. *Journal of Chemical Biology* **6**, 85-95, doi:10.1007/s12154-013-0094-5 (2013).

67 Grimm, J. B. *et al.* A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nature Methods* **12**, 244-+, doi:10.1038/nmeth.3256 (2015).

68 Sun, W., Guo, S. G., Hu, C., Fan, J. L. & Peng, X. J. Recent Development of Chemosensors Based on Cyanine Platforms. *Chemical Reviews* **116**, 7768-7817, doi:10.1021/acs.chemrev.6b00001 (2016).

69 Zhu, S. J., Tian, R., Antaris, A. L., Chen, X. Y. & Dai, H. J. Near-Infrared-II Molecular Dyes for Cancer Imaging and Surgery. *Advanced Materials* **31**, 25, doi:10.1002/adma.201900321 (2019).

70 Koide, Y., Urano, Y., Hanaoka, K., Terai, T. & Nagano, T. Development of an Si-Rhodamine-Based Far-Red to Near-Infrared Fluorescence Probe Selective for Hypochlorous Acid and Its Applications for Biological Imaging. *Journal of the American Chemical Society* **133**, 5680-5682, doi:10.1021/ja111470n (2011).

71 Wang, L. L. *et al.* Hybrid Rhodamine Fluorophores in the Visible/NIR Region for Biological Imaging. *Angewandte Chemie-International Edition* **58**, 14026-14043, doi:10.1002/anie.201901061 (2019).

72 Zhang, J. *et al.* BODIPY-Based Fluorescent Probes for Biothiols. *Chem.-Eur. J.* **26**, 4172-4192, doi:10.1002/chem.201904470 (2020).

73 Kowada, T., Maeda, H. & Kikuchi, K. BODIPY-based probes for the fluorescence imaging of biomolecules in living cells. *Chemical Society Reviews* **44**, 4953-4972, doi:10.1039/c5cs00030k (2015).

74 Huang, Y. J. *et al.* Reducing aggregation caused quenching effect through co-assembly of PAH chromophores and molecular barriers. *Nature Communications* **10**, doi:10.1038/s41467-018-08092-y (2019).

75 Ma, X. F. *et al.* Fluorescence Aggregation-Caused Quenching versus Aggregation-Induced Emission: A Visual Teaching Technology for Undergraduate Chemistry Students. *Journal of Chemical Education* **93**, 345-350, doi:10.1021/acs.jchemed.5b00483 (2016).

76 Qi, J. P. *et al.* Towards more accurate bioimaging of drug nanocarriers: turning aggregation-caused quenching into a useful tool. *Advanced Drug Delivery Reviews* **143**, 206-225, doi:10.1016/j.addr.2019.05.009 (2019).

77 Forster, T. & Kasper, K. EIN KONZENTRATIONSUMSCHLAG DER FLUORESZENZ DES PYRENS. *Zeitschrift Fur Elektrochemie* **59**, 976-980 (1955).

78 Ren, F. *et al.* Real time bioimaging for mitochondria by taking the aggregation process of aggregation-induced emission near-infrared dyes with wash-free staining. *Materials Chemistry Frontiers* **3**, 57-63, doi:10.1039/c8qm00425k (2019).

79 Wang, Y. F., Zhang, Y. X., Wang, J. J. & Liang, X. J. Aggregation-induced emission (AIE) fluorophores as imaging tools to trace the biological fate of nano-based drug delivery systems. *Advanced Drug Delivery Reviews* **143**, 161-176, doi:10.1016/j.addr.2018.12.004 (2019).

80 Zhao, N. *et al.* Aggregation-Induced Emission Luminogens with the Capability of Wide Color Tuning, Mitochondrial and Bacterial Imaging, and Photodynamic Anticancer and Antibacterial Therapy. *Acs Applied Materials & Interfaces* **11**, 11227-11237, doi:10.1021/acsami.9b01655 (2019).

81 Fan, J. L., Hu, M. M., Zhan, P. & Peng, X. J. Energy transfer cassettes based on organic fluorophores: construction and applications in ratiometric

sensing. *Chemical Society Reviews* **42**, 29-43, doi:10.1039/c2cs35273g (2013).

82 Liu, C. *et al.* A unique rectilinearly pi-extended rhodamine dye with large Stokes shift and near-infrared fluorescence for bioimaging. *Chemical Communications* **53**, 10727-10730, doi:10.1039/c7cc06220f (2017).

83 Grimm, J. B. *et al.* A general method to fine-tune fluorophores for live-cell and in vivo imaging. *Nature Methods* **14**, 987-+, doi:10.1038/nmeth.4403 (2017).

84 Li, C., Tebo, A. G. & Gautier, A. Fluorogenic Labeling Strategies for Biological Imaging. *International Journal of Molecular Sciences* **18**, doi:10.3390/ijms18071473 (2017).

85 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, doi:10.1016/j.cbpa.2013.05.015 (2013).

86 Prifti, E. *et al.* A Fluorogenic Probe for SNAP-Tagged Plasma Membrane Proteins Based on the Solvatochromic Molecule Nile Red. *Acs Chemical Biology* **9**, 606-612, doi:10.1021/cb400819c (2014).

87 Han, J. Y. & Burgess, K. Fluorescent Indicators for Intracellular pH. *Chemical Reviews* **110**, 2709-2728, doi:10.1021/cr900249z (2010).

88 Shieh, P. *et al.* CalFluors: A Universal Motif for Fluorogenic Azide Probes across the Visible Spectrum. *Journal of the American Chemical Society* **137**, 7145-7151, doi:10.1021/jacs.5b02383 (2015).

89 Sun, X. L. *et al.* Development of SNAP-Tag Fluorogenic Probes for Wash-Free Fluorescence Imaging. *Chembiochem* **12**, 2217-2226, doi:10.1002/cbic.201100173 (2011).

90 Blum, G. *et al.* Dynamic imaging of protease activity with fluorescently quenched activity-based probes. *Nature Chemical Biology* **1**, 203-209, doi:10.1038/nchembio728 (2005).

91 Lang, K. & Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chemical Reviews* **114**, 4764-4806, doi:10.1021/cr400355w (2014).

92 Bender, K. O. *et al.* Design of a Highly Selective Quenched Activity-Based Probe and Its Application in Dual Color Imaging Studies of Cathepsin S Activity Localization. *Journal of the American Chemical Society* **137**, 4771-4777, doi:10.1021/jacs.5b00315 (2015).

93 Devaraj, N. K. & Weissleder, R. Biomedical Applications of Tetrazine Cycloadditions. *Accounts of Chemical Research* **44**, 816-827, doi:10.1021/ar200037t (2011).

94 Kormos, A. *et al.* Bistetrazine-Cyanines as Double-Clicking Fluorogenic Two-Point Binder or Crosslinker Probes. *Chem.-Eur. J.* **24**, 8841-8847, doi:10.1002/chem.201800910 (2018).

95 Dean, K. M. & Palmer, A. E. Advances in fluorescence labeling strategies for dynamic cellular imaging. *Nature Chemical Biology* **10**, 512-523, doi:10.1038/nchembio.1556 (2014).

96 Los, G. V. *et al.* HatoTag: A novel protein labeling technology for cell imaging and protein analysis. *Acs Chemical Biology* **3**, 373-382, doi:10.1021/cb800025k (2008).

97 Griffin, B. A., Adams, S. R. & Tsien, R. Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269-272, doi:10.1126/science.281.5374.269 (1998).

98 Adams, S. R. *et al.* 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, doi:10.1021/ja017687n (2002).

99 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, doi:10.1111/j.1751-1097.2008.00335.x (2008).

100 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-+, doi:10.1021/ja904800k (2009).

101 Szent-Gyorgyi, C. *et al.* Fluorogen-activating single-chain antibodies for imaging cell surface proteins. *Nature Biotechnology* **26**, 235-240, doi:10.1038/nbt1368 (2008).

102 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, doi:10.1038/nmeth749 (2005).

103 Ozhalici-Unal, H. *et al.* 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-+, doi:10.1021/ja805042p (2008).

104 Ziegler, M. S., Lakshmi, K. V. & Tilley, T. D. Dicopper Cu(I)Cu(I) and Cu(I)Cu(II) Complexes in Copper-Catalyzed Azide-Alkyne Cycloaddition. *Journal of the American Chemical Society* **139**, 5378-5386, doi:10.1021/jacs.6b13261 (2017).

105 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 (vol 126, pg 15046, 2004). *Journal of the American Chemical Society* **127**, 11196-11196, doi:10.1021/ja059912x (2005).

106 Zimmerman, J. R. *et al.* The Synthesis of a New Class of Highly Fluorescent Chromones via an Inverse-Demand Hetero-Diels-Alder Reaction. *Organic Letters* **17**, 3256-3259, doi:10.1021/acs.orglett.5b01417 (2015).

107 Chen, Y. C. *et al.* Coumarin-Based Fluorogenic Probes for No-Wash Protein Labeling. *Angewandte Chemie-International Edition* **53**, 13785-13788, doi:10.1002/anie.201408015 (2014).

108 Liu, W. *et al.* A Rapid and Fluorogenic IMP-AcBODIPY Probe for Covalent Labeling of Proteins in Live Cells. *Journal of the American Chemical Society* **136**, 4468-4471, doi:10.1021/ja500170h (2014).

109 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, doi:10.1139/cjc-2014-0405 (2015).

110 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, doi:10.1002/anie.201506935 (2015).

111 Barth, C. W. & Gibbs, S. L. Direct Administration of Nerve-Specific Contrast to Improve Nerve Sparing Radical Prostatectomy. *Theranostics* **7**, 573-593, doi:10.7150/thno.17433 (2017).

112 Altman, R. B. *et al.* Cyanine fluorophore derivatives with enhanced photostability. *Nature Methods* **9**, 68-U178, doi:10.1038/nmeth.1774 (2012).

113 Tai, C. H., Lu, C. P., Wu, S. H. & Lo, L. C. Synthesis and evaluation of turn-on fluorescent probes for imaging steroid sulfatase activities in cells. *Chemical Communications* **50**, 6116-6119, doi:10.1039/c4cc01282h (2014).

114 Collot, M. *et al.* Ultrabright and Fluorogenic Probes for Multicolor Imaging and Tracking of Lipid Droplets in Cells and Tissues. *Journal of the American Chemical Society* **140**, 5401-5411, doi:10.1021/jacs.7b12817 (2018).

115 Yan, R. Q. *et al.* Activatable NIR Fluorescence/MRI Bimodal Probes for in Vivo Imaging by Enzyme-Mediated Fluorogenic Reaction and Self-Assembly. *Journal of the American Chemical Society* **141**, 10331-10341, doi:10.1021/jacs.9b03649 (2019).

116 Ning, C., Xiao, C. Y., Lu, R. F. & Zhou, P. W. Whether the excited state intramolecular proton transfer of 1-hydroxy-2-acetonaphthone will happen? *Journal of Luminescence* **217**, doi:10.1016/j.jlumin.2019.116825 (2020).

117 Kwon, J. E. & Park, S. Y. Advanced Organic Optoelectronic Materials: Harnessing Excited-State Intramolecular Proton Transfer (ESIPT) Process. *Advanced Materials* **23**, 3615-3642, doi:10.1002/adma.201102046 (2011).

118 Cheng, J. L. *et al.* A Single 2-(2 '-Hydroxyphenyl)benzothiazole Derivative Can Achieve Pure White-Light Emission. *Chemistry-an Asian Journal* **9**, 3215-3220, doi:10.1002/asia.201402779 (2014).

119 Karpenko, I. A. *et al.* Fluorogenic Squaraine Dimers with Polarity-Sensitive Folding As Bright Far-Red Probes for Background-Free Bioimaging. *Journal of the American Chemical Society* **137**, 405-412, doi:10.1021/ja5111267 (2015).

120 Klymchenko, A. S. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Accounts of Chemical Research* **50**, 366-375, doi:10.1021/acs.accounts.6b00517 (2017).

121 Li, X. H., Gao, X. H., Shi, W. & Ma, H. M. Design Strategies for Water-Soluble Small Molecular Chromogenic and Fluorogenic Probes. *Chemical Reviews* **114**, 590-659, doi:10.1021/cr300508p (2014).

122 Lukinavicius, G. *et al.* Fluorogenic probes for live-cell imaging of the cytoskeleton. *Nat. Methods* **11**, 731-U168, doi:10.1038/nmeth.2972 (2014).

CHAPTER II: DEVELOPING FAR-RED/NEAR-INFRARED DYE-HCRBPII FLUORESCENT TAGS FOR NO-WASH BACKGROUND-FREE LIVE CELL IMAGING APPLICATIONS

Over the past 25 years, the ready availability of genetically encoded fluorescent tags has revolutionized cell biology and live-cell imaging.[1-6] New tags continue to be developed that are brighter, more red-shifted, expanding the available spectral range, and better tolerated by fusions. As described in Chapter I, fluorescent proteins (FPs) provide toolkits that enable various technologies and applications in a variety of biomedical research fields.[7-15] FPs are bright enough and provide emission spectra spanning the visible spectrum from the blue to the near-infrared (NIR).[16-21] However, the evolving field of fluorescence microscopy demands more sophisticated FPs to address early-generation candidates' deficiencies. Optimization of characteristics such as photostability, red-shifted emission, high fluorescence quantum yield (QY), high extinction coefficient (ε), and large Stokes shift (>100 nm) is always pursued. Nonetheless, certain pitfalls of FPs may hamper their application; for example, their Stokes shifts are usually small, leading to the significant overlap in the absorption and emission spectra, self-quenching, and severe limitation in the imaging depth and overall emission brightness.[22-26]

As discussed in **Section I-2**, far-red/NIR emission suffers less from interference with cellular milieu and is more advantageous for deep tissue imaging applications.[27-33] Nevertheless, despite extensive efforts to develop monomeric FPs, known naturally occurring red fluorescent proteins (RFPs), such as DsRed

and eqFP611, tend to form obligate oligomers, which due to their large sizes, has raised serious concerns about their usefulness as molecular fusion tags in *in vivo* model systems.[34-39] Besides, being oxygen dependent for the fluorophore maturation also limits their application in some anaerobic experiment settings.[40,41]

Conjugation of small synthetic dyes with genetically encodable protein/peptide tags has led to the development of flexible labeling systems that can be tailored ad infinitum and open a wide window for the development of novel imaging tags.

II.1 Initial work towards developing hCRBPI-based fluorescent tags

One of the main goals of our research group is to develop fluorescent tags that emit in the far-red/NIR region of the spectrum where conventional fluorescent proteins can be inadequate. Attempts towards this initiated in 2012 when our lab demonstrated the ability to regulate the absorption wavelength of all-*trans*-retinal in type II human cellular retinoid binding protein (hCRBPII) complexes from 425 nm to 644 nm.[42] In this system, the protein was engineered to encapsulate fully and covalently bind retinal as a Schiff base (**Figure II-1a**). Through rational point mutagenesis, the absorption profile of the corresponding complexes shifted over a range of more than 200 nanometers as the result of altering the electrostatic environment of bound retinal within the binding pocket of the host protein (**Figure II-1b**).

hCRBPII was selected as the target of protein engineering since it has a large binding cavity and a high tolerance for mutations that allows flexibility in the

protein redesign and enables binding of a wide range of different ligands.[43-45] We envisioned replacing all-*trans*-retinal with synthetic chromophores to turn hCRBPII into a robust platform to design novel fluorescent tags.

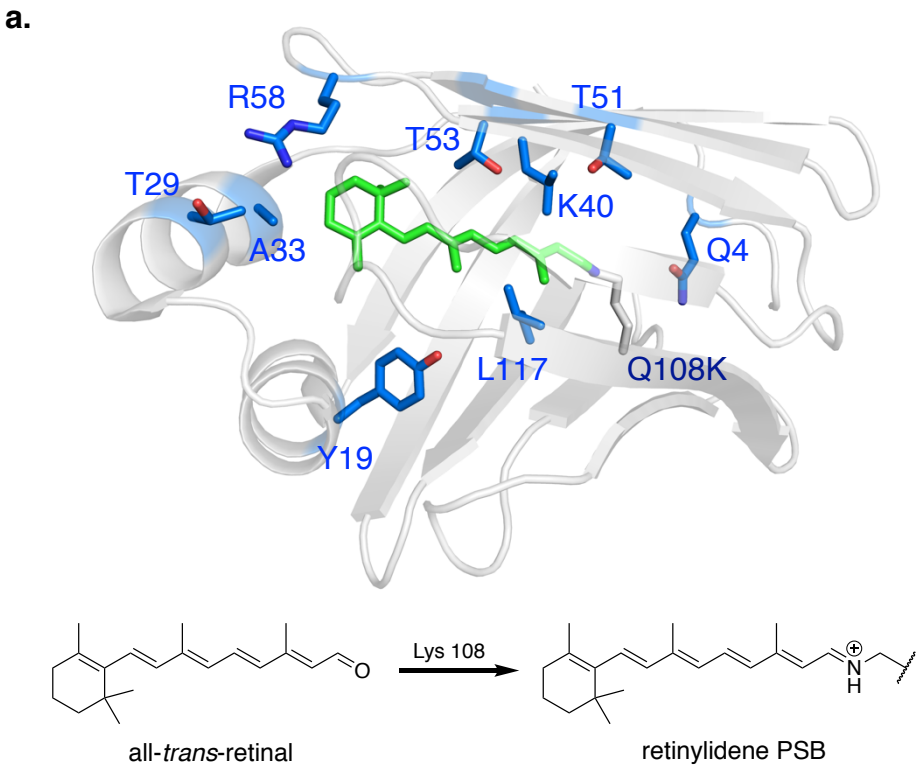


Figure II-1. **a**. Crystal structure of Q108K:K40L-hCRBPII mutant complexed with all-*trans*-retinal. Key residues engineered to regulate the absorption wavelength of the bound retinylidene are highlighted (PDB 4EXZ). The scheme shows the iminium (PSB) formation between lysine 108 and retinal aldehyde. **b**. Protein solution of different hCRBPII mutants incubated with all-*trans* retinal.

Ideally, in this approach, the unbound chromophore is not fluorescent and becomes activated only upon binding to the target protein and irradiation of the corresponding complexes. In contrast to the intrinsically fluorescent proteins that are constantly on, this method can provide spatiotemporal control of the signal. Besides, hCRBPII is a small protein (~15 KDa) that does not require oxygen to fold or bind the chromophore [46,47] and, therefore, can find potential applications in obligate anaerobes.

The following briefly describes the pairing of different engineered hCRBPII mutants with various fluorophores to generate tailor-made fluorogenic protein fusion tags. As a proof of principle, in 2015, our lab successfully showed the application of merocyanine retinal analog (**MCRA**) as a "turn-on" fluorescent tag in conjugation with type II cellular retinoic acid binding protein (CRABPII).[48] CRABPII is also a small cytosolic lipid-binding protein and is a structural homolog of hCRBPII.[49,50] The spectroscopic characterizations of the complexes were examined with a variety of mutants, and their use as fluorescent tags to image *E. coli* was demonstrated. Nonetheless, efforts to use MCRA/CRABPII probes as a fluorescent tag to image mammalian cells were not fruitful, most probably due to protein misfolding

On the contrary, hCRBPII variants showed no deleterious problem in expression or folding in mammalian cells environment. Our research group produced highly fluorescent red pigments through covalent conjugation of the merocyanine aldehyde to the active site lysine residue of hCRBPII (**Figure II-2a**).[51]

Upon extensive protein engineering, **MCRA** can bind some hCRBPII mutants almost instantaneously (less than 1 minute) with rapid cellular penetration. It binds hCRBPII as Protonated Schiff Base (PSB) under physiological pH resulting in a substantial bathochromic shift in its absorption and emission, which is entirely separated from unbound **MCRA** absorption and should mitigate background fluorescence from the free reagent (**Figure II-2b**).

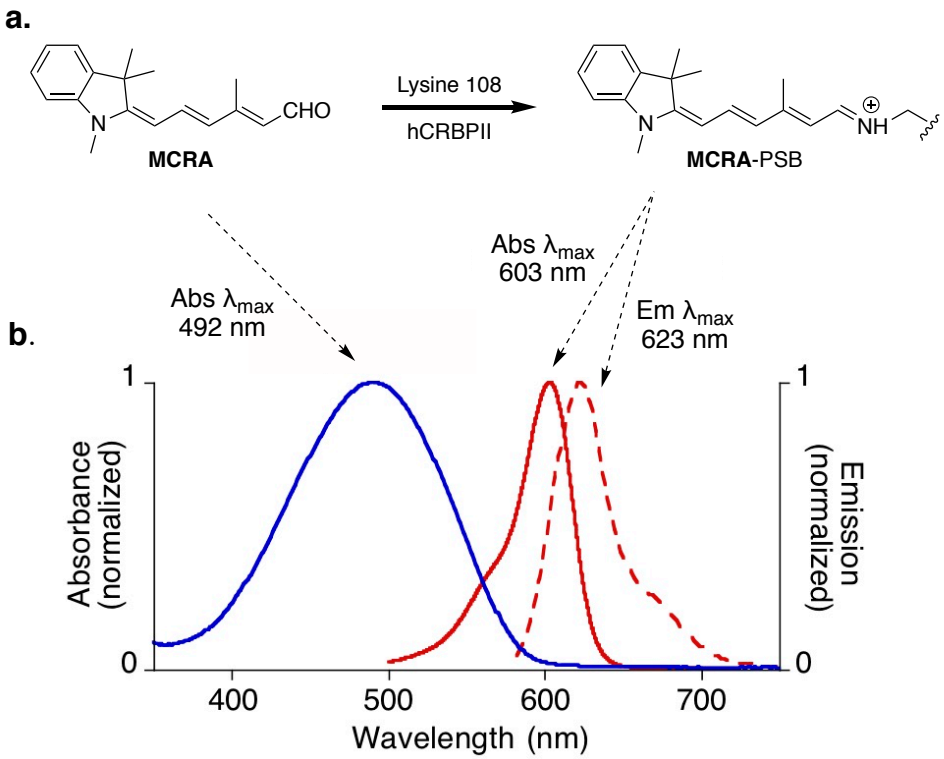


Figure II-2. **a**. Structure of **MCRA** and its PSB formation with active Lysine 108 of hCRBPII. **b**. Spectroscopic properties of free aldehyde (blue) and PSB (red).

The required condition for cell imaging is 250 nM **MCRA** as the staining concentration for HeLa, U2OS, and COS-7 cells and 1 min incubation at 37 °C, followed by washing two times with DPBS to remove excess **MCRA**. However, non-specific iminium formation with other proteins in the cells leads to significant

fluorescent background that extra washing steps cannot diminish.[51] Furthermore, the absorption maxima of the **MCRA** complexes with different hCRBPII variants is essentially constant (centered around 600 nm) due to the high degree of conjugation between the iminium and terminal nitrogen, and they often show Stokes shifts smaller than 30 nm.

II.2 Desired features of a practical fluorescence imaging tag

To this end, we showed that hCRBPII could serve as a robust platform to develop protein-based fluorescent tags; however, improving to practical tags requires particular modification of the selected fluorophore's structure and its photophysical properties. One of the most critical issues to overcome when using tags that require supplementing exogenous synthetic ligands is the fluorescent background that can originate eighter free chromophore or off-target bindings.

One way to avoid this and achieve a high signal-to-noise ratio is to use fluorescent probes that display no fluorescence until labeling occurs. Such probes are often called fluorogenic probes to highlight their ability to show an increase in fluorescence upon binding their targets.[52] Fluorogenic probes can provide high sensitivity and the ability to monitor diverse events selectively and are essential components in the toolkit of chemical biology.[53-55] Besides, using fluorogenic probes eliminates washing steps required for the removal of free chromophores and enables real-time measurements.[56-61]

Other main requirements of successful imaging tags can be summarized as followed. The probe must be bright enough to be detected well, which requires it

to show both high extinction coefficient and high fluorescence quantum yield as the product of these two determines the absolute brightness. High photostability that enables noninvasive long-term imaging and large Stokes shifts (typically over 80 nm) are highly preferred. As previously discussed, probes with large Stokes shifts can minimize cross-talk between the excitation source and the fluorescent emission, which generally results in poor signal-to-noise ratio and self-quenching due to back-scattering from biological samples.[62] [63-68]

Another photophysical requirement is the far-red/NIR wavelengths. As visible light is strongly scattered in deep tissues, and thus its penetration is limited. In recent decades, the development of NIR fluorophores has enabled the visualization of deep tissues of living organisms with detailed spatiotemporal information.[69-74] The tissue is more transparent to NIR (~650–900 nm) optical window light due to less absorbance by body pigments and is characterized by low autofluorescence and minimal phototoxicity for living cells.[75,76] Lastly, but important, the synthetic chromophore should not be cytotoxic and possess high cell permeability.

II.3 Attempts to minimize non-specific fluorescent background

One of our approaches to prevent fluorescent background from non-specific iminium formation has been to design chromophores with strong solvatochromic properties. We hypothesized that using environmentally sensitive probes will enable us to tailor the specific hCRBPII/chromophore emission by changing the electrostatic environment around the chromophore inside the protein cavity. As

such, the fluorescence from the desired reaction is red-shifted compared to that of non-specific iminium formation.

Fluorescent solvatochromic dyes are characterized as push-pull dyes containing electron-donating and electron-withdrawing groups conjugated through a π-electron spacer. These dyes indicate a feature that favors an Intramolecular Charge Transfer (ICT) process upon light excitation making these chromophores have a larger dipole moment in the excited state.[77-80] Consequently, polar solvents will stabilize the excited state better than the ground state resulting in a reduction in the energy gap between the ground state and excited state energy levels, leading to bathochromic shifts in emission.[81-83] Consequently, solvatochromic compounds usually show the same absorption maxima in various solvents but can have the most bathochromic shift in emission wavelength in a polar solvent environment.

After an extensive search of literature for different chromophores that exhibit substantial sensitivity to solvent polarity, **ThioFlour**, a derivative of Dapoxyl dyes, were synthesized to pair with hCRBPII as a solvatochromic probe (**Figure II-3**); Dapoxyl dye family show distinctive variations in quantum yields, fluorescence wavelength maxima (i.e., larger Stokes shifts), and extinction coefficients upon changes in polarity.[84-86] Conjugating the solvatochromic fluorophore **ThioFluor** to engineered hCRBPII mutants yielded various **ThioFluor**-PSB complexes with emission maxima varying from 613 nm to 744 nm, thus exhibit about 130 nm

variation in emission wavelength, covering both the red and far-red fluorescence wavelength regimes.

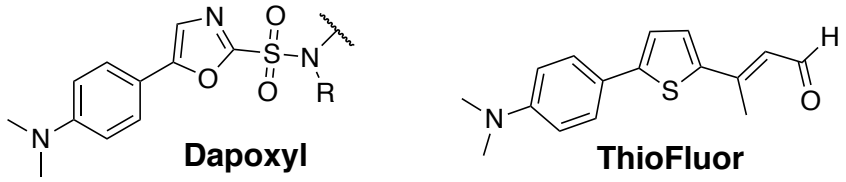


Figure II-3. Structures of Dapoxyl dye and **ThioFluor**.

Furthermore, to prove its utility in live-cell imaging, whole-cell and subcellular compartments such as nuclei and cytosol were imaged by targeting the hCRBPII mutants with signaling peptides like nucleus-localizing sequence (NES) and nucleus-exclusion sequence (NES). However, since unbound **ThioFluor** is strongly emissive in hydrophobic environments, washing steps were required before fluorescence imaging experiments to remove the excess unbound dyes after incubation.

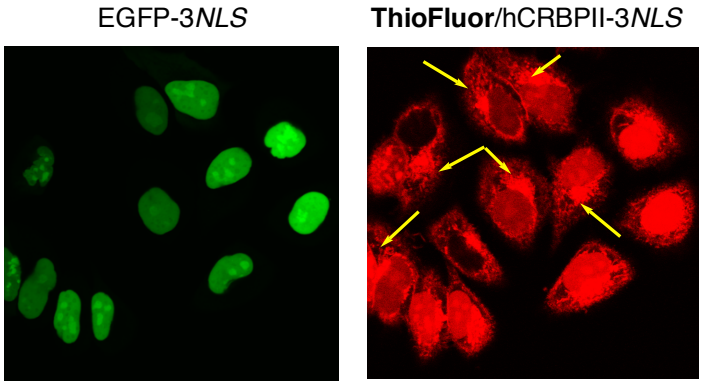


Figure II-4. Non-specific fluorescent background originating from off-target iminium formation in labeled HeLa cells with 10 μM **ThioFluor** (incubated at 37 °C for 1 hour) expressing hCRBPII-EGFP-3NLS. NLS = nuclear localization sequence.

Nevertheless, following optimized imaging condition, using 10 μM **ThioFluor** incubated at 37 °C for 1 hour provides enough time for non-specific iminium formations that emits in the same window as **ThioFluor-**PSB emission is collected to images cells leading in the fluorescent background even after washing steps (**Figure II-4**).[87] However, Dr. Elizabeth Santos later successfully developed engineered hCRBPII/**ThioFlour** complexes as Large Stokes Shift (LSS) fluorescent tags with fluorescent properties and applications in "washed" live-cell imaging (see Chapter III).

II.4 No-wash background-free fluorogenic imaging tag design

An important photochemical feature of an advanced imaging tag is efficient "on/off" switching properties of the fluorescence extended from the fluorogenicity. If the fluorescence can be turned on and off iteratively by fast photochemical transformations, then the spatiotemporal control of the fluorescence signals can be realized.[88] It also helps to eliminate laborious washing steps prior to imaging experiments. We have pursued developing protein-based no-wash background-free imaging via modifying the chromophore structure and reengineering the protein cavity accordingly. In this study, we have designed a chromophore that becomes fluorescent only upon binding hCRBPII mutants.

It was envisioned that replacing the *N,N*-dimethyl amino group in **ThioFluor** with a phenol moiety, **ThioPhenol**, would make it fluorogenic (**Figure II-5**). Since the hydroxyl group is a weak electron donor, it fails to produce a strong ICT system, leading to a weakly emissive molecule. Furthermore, the phenol moiety's high pK_a

prevents its deprotonation in the ground state at neutral pH. Therefore, the free aldehydic **ThioPhenol** does not fluoresce either in aqueous or organic solvents (<1% QY in solution). However, we envisaged that we could create a hydrogen-bonding network around the hydroxyl group through rational protein engineering and thus facilitate its deprotonation through the Excited State Proton Transfer (ESPT) process. Deprotonation of phenol moiety either through adjacent residues or water molecules upon photoexcitation liberates the negatively charged oxygen atom. The corresponding alkoxide has the same Hammett value as *N,N*-dimethyl amino group, i.e., both have equal electron-donating strenghths.[89]

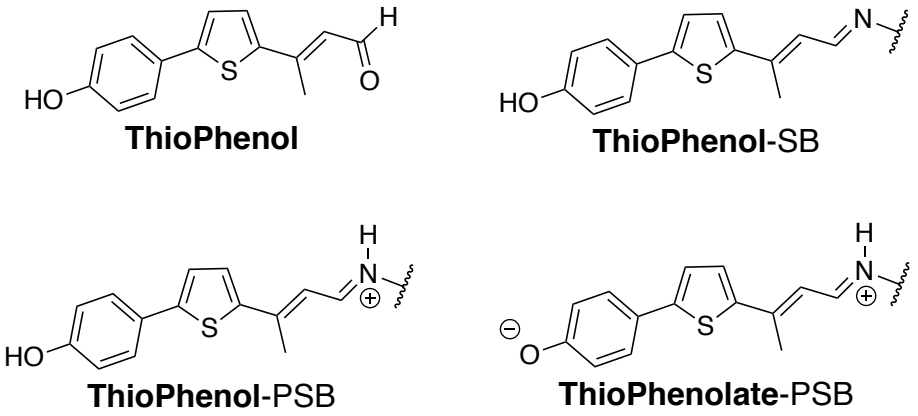


Figure II-5. Structures of free aldehyde **ThioPhenol**, **ThioPhenol**-imine, **ThioPhenol**-iminium, and **ThioPhenolate**-iminium.

On the other hand, engineering the target protein to maintain a high iminium pK_a will result in Protonated Schiff Base (PSB) formation. Therefore, the resulting photoexcited product, **ThioPhenolate**-PSB, complex possesses a potent electron donor and electron acceptor character on both ends of the complex, forming the D–π–A structure with increased ICT emitting in far-red to NIR region of the

spectrum depending on the protein mutant (**Figure II-5**). The designed system will put the target protein in control of fluorescence selectivity and effectively diminish fluorescence from non-specific iminium formation. The same as the free aldehyde, the **ThioPhenol**-iminium complex (**Figure II-5**) is also unable to make a robust ICT and is extremely weakly fluorescent (less than 2% QY depending on the protein mutation). Thus, the designed probe is a light-triggered OFF-to-ON system that enables spatiotemporal control of the **ThioPhenol**/protein complex's NIR fluorescence emission as an imaging probe.

II.5 Spectroscopic properties of ThioPhenol in solution

First, **ThioPhenol**'s absorption and emission were measured in various organic solvents with different polarities to investigate its spectroscopic behavior. Results are summarized in **Table II-1.**

Table II-1. Spectroscopic characterization of **ThioPhenol** in various solvents.

Solvent	λ_{abs} (nm)	λ_{em} (nm)	SS (nm)	ε ($M^{-1}cm^{-1}$)	Φ [a]
Toluene	386	461	75	29,715	0.00
Tetrahydrofuran	382	463	81	31,299	0.00
Ethyl acetate	378	465	87	34,319	0.00
Dimethyl sulfoxide	392	495	103	28,560	0.015
Ethanol	391	515	124	29,752	0.01
PBS buffer	389	-	-	4,193	0.00

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

ThioPhenol shows almost the same absorbance in different solvents, with absorption maximum centered around 390 nm (**Figure II-6a**). It is not soluble in

aqueous solutions and forms aggregates leading to its small extinction coefficient and quenched emission in PBS buffer. However, its extinction coefficient is as high as **ThioFlour**'s in organic solvents, about 30,000 $M^{-1}cm^{-1}$.

ThioPhenol's emission also barely changes as a function of the solvent's E_T (30) value.[90] The emission maximums span a small range of 54 nm (from 461 nm in Toluene to 515 nm in EtOH) (**Figure II-6b**), and it hardly shows solvatochromic properties. Such behavior is expected as the hydroxyl group is a weak electron-donating group and fails to produce a strong Internal Charge Transfer (ICT) system upon excitation. Subsequently, **ThioPhenol** is less prone to solvent relaxations, and its emission slightly red-shifts in polar solvents.

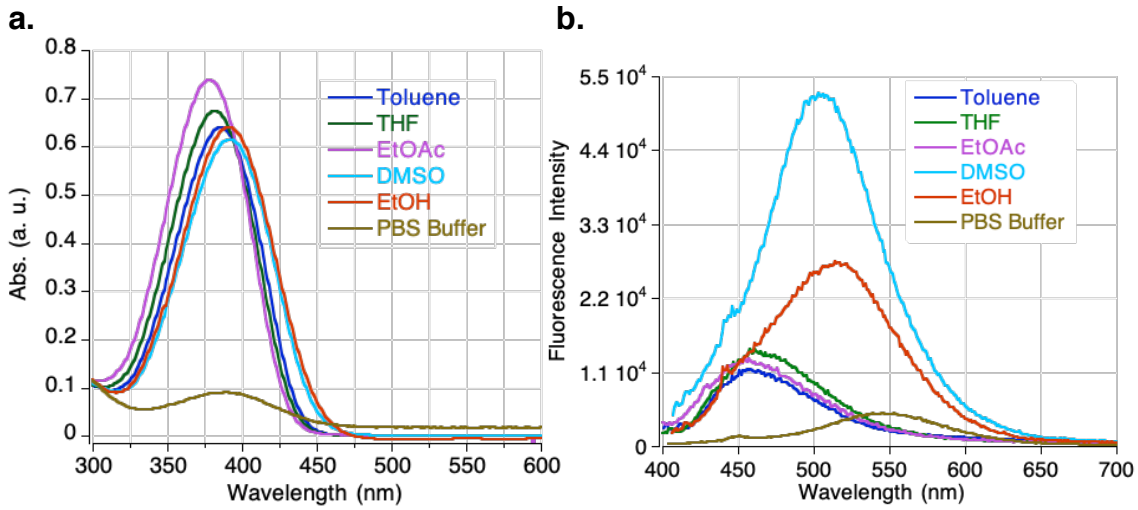


Figure II-6. Spectroscopic properties of **ThioPhenol** in different solvents. **a**. UV-Vis and **b**. Fluorescence spectra of **ThioPhenol**.

In order to mimic the product of lysine 108 and free chromophore condensation in the protein pocket, the aldehyde was coupled with *n*-butyl amine in ethanol (**Figure II-7a**). The absorption and emission maxima of the

corresponding *n*-butyl imine Schiff base blue-shifts by about 25 and 50 nm, respectively. Acidification of **ThioPhenol**-SB in ethanol with concentrated hydrochloric acid (aq.) solution gives the protonated iminium (**ThioPhenol**-PSB), resulting in a large red-shift in UV-Vis with the absorption maximum at 450 nm. Furthermore, basification of the imine solution with 1M NaOH (pH>9.5-10) yields the corresponding alkoxide (**ThioPhenolate**-SB), showing bathochromic shifts in both absorption and emission spectra (**Figure II-7b** and **Table II-2**).

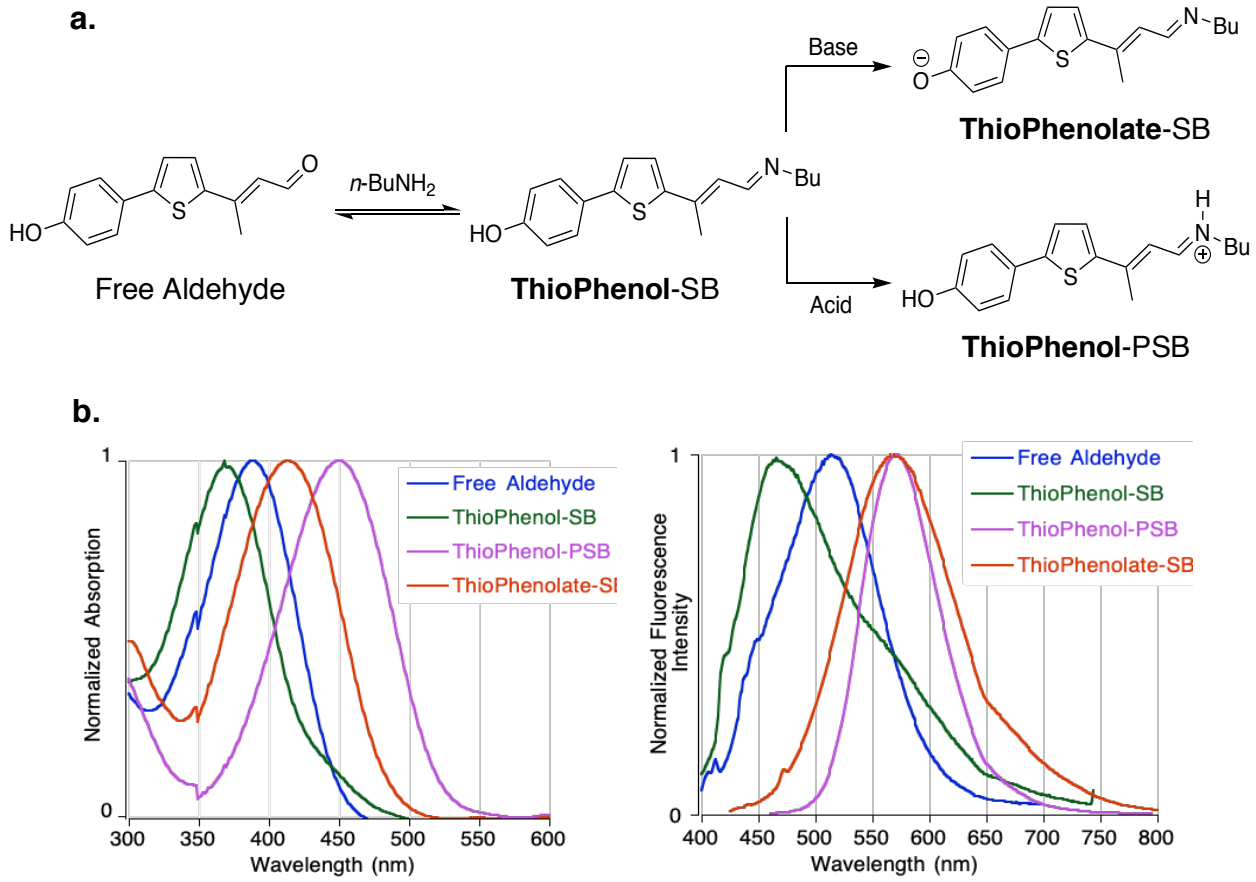


Figure II-7. **a**. Schiff base and protonated Schiff base (PSB) of **ThioPhenol** with *n*-butyl amine in ethanol. **b**. Normalized spectra of **ThioPhenol** and derivatives in ethanol. Absorbance (left) and emission (right).

Notably, the maximum emission wavelength is almost the same (~570 nm) for both **ThioPhenol**-PSB and **ThioPhenolate**-SB; however, the latter has a broad emission peak that reaches the far-red/NIR with a small shoulder around 670 nm. Nevertheless, it is not impossible to produce the **ThioPhenolate**-PSB complex in solution, due to lower iminium pK_a (pK_a~7), it will deprotonate before the phenol moiety (pK_a~10) in aqueous media. As the result of failure in producing a strong ICT system, all species formed in solution and the free aldehyde are very low emissive molecules showing less than 3% quantum yield in different solvents **Table II-2**. This leads to essentially non-fluorescent unbound chromophore and insignificant fluorescence from off-target imine or minimum formation. Therefore, engineering hCRBPII mutants to fluoresce upon binding **ThioPhenol** will result in fluorogenic probes with applications in no-wash live-cell imaging.

Table II-2. Spectroscopic characterization of **ThioPhenol** and derivatives.

Compound	λ_{abs} (nm)	λ_{em} (nm)	SS (nm)	ε ($M^{-1}cm^{-1}$)	Φ^a
Free Aldehyde	391	515	124	29,752	0.01
ThioPhenol-SB	368	468	100	33,623	0.0
ThioPhenol-PSB	450	571	121	45,752	0.02
ThioPhenolate-SB	413	568	155	36,341	0.02

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

II.6 General protein host properties for this study

The formation of the chromophore-protein complex provides a platform to control the embedded chromophore's photophysical properties exquisitely. The

protein host should fulfill certain requirements to be chosen as a good candidate for its fluorogenic partner, **ThioPhenol**, in this system. Fortunately, hCRBPII has a large binding cavity and a high tolerance to mutations without affecting its structure making it a robust platform for the flexible design of fluorescent tags.[42,51,91]

In this study, the pK_a of the chromophore Schiff Base (SB) should be optimized so that an iminium is formed in neutral pH. Second, the protein cavity surrounding the hydroxyl group should provide a path for deprotonation through the ESPT process. It is essential that forming these species in the ground state will make a constantly "ON" system without any temporal control. Therefore, protein engineering should be adjusted to not deprotonate the hydroxyl group in the ground state. What follows is our approach to engineer (hCRBPII) as a host having such properties.

II.7 Primary observation of dual fluorescence in hCRBPII/ThioPhenol

To deprotonate the hydroxyl group in the excited state, basic or proton acceptor amino acids should be placed at its close proximity. Picking the suitable amino acids as well as their distances to the hydroxyl group is critical for success. The crystal structures of several hCRBPII mutants with **ThioFluor** helped to identify amino acids putatively close to the hydroxyl group.

Arg 58 is located in the hCRBPII binding pocket entrance and partially covers the mouth of the cavity. Previously we reported on how its mutation to larger, hydrophobic residues could cap and sequester the binding pocket from the

bulk medium.[42] Docking simulation of **ThioPhenol** in the crystal structure of Q108K:K40L:T51V:T53S:R58W-hCRBPII/**ThioFluor** showed that Arg 58 is at a decent distance from the hydroxyl group and can be a proper candidate to tackle (**Figure II-8**).

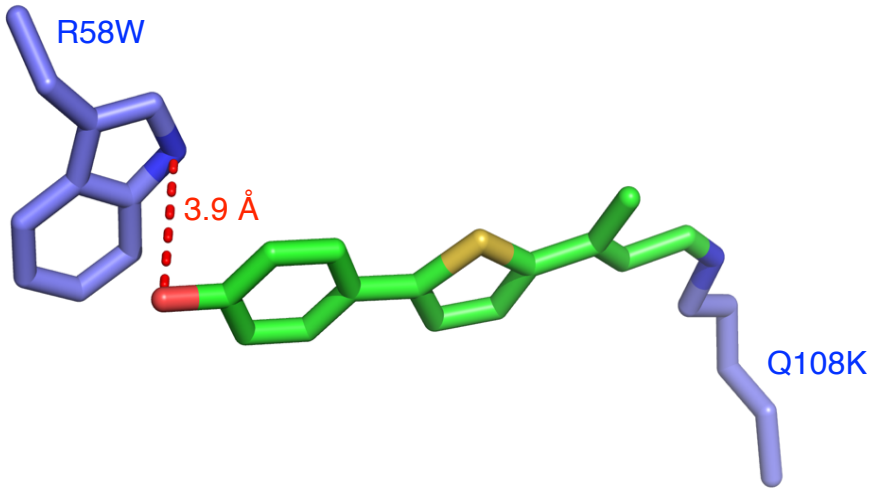


Figure II-8. Flexible docking of **ThioPhenol** in the crystal structure of Q108K:K40L:T51V:T53S:R58W-hCRBPII/**ThioFluor**. The distance between the hydroxyl group and the indole nitrogen of 58W side chain is 3.9 Å.

It was hypothesized that substitution of Arg 58 for amino acids involved in ESPT of FP's chromophore, such as serine and glutamic acid,[92] along with amino acids capable of forming a hydrogen bond network with the hydroxyl group including histidine, tryptophan, and tyrosine, could result in deprotonation of the hydroxyl group. A list of all mutations on Arg 58 is listed in **Table II-3**, along with the corresponding complexes' photophysical properties. For all in-protein experiments in this study, 20 μM protein and 0.5 equivalent **ThioPhenol** are incubated at room temperature (23 °C) at pH ~7.2.

As shown in **Table II-3**, all mutants show absorption around 525-530 nm and emission around 580-590 nm (the blue emission). However, mutation of Arg 58 to histidine and tryptophan yielded another red-shifted emission peak around 690 nm (the red emission) (**Table II-3** entries 2 and 3).

Table II-3. Spectroscopic change as a result of mutating R58 in Q108K:K40L:T51V:T53S template.

Entry	hCRBPII Mutant	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	SS	Φ_{ESPT}	pK_a
1	Q108K:K40L:T51V:T53S	531	585	–	54	0.11	6.1
2	Q108K:K40L:T51V:T53S:**R58W**	530	586	693	56/163	0.24	6.0
3	Q108K:K40L:T51V:T53S:**R58H**	530	590	691	60/161	0.49	6.1
4	Q108K:K40L:T51V:T53S:**R58E**	519	586	–	67	0.16	5.9
5	Q108K:K40L:T51V:T53S:**R58L**	517	590	–	73	0.11	5.9
6	Q108K:K40L:T51V:T53S:**R58Q**	522	591	–	69	0.13	5.9
7	Q108K:K40L:T51V:T53S:**R58K**	531	583	–	52	0.16	6.8
8	Q108K:K40L:T51V:T53S:**R58S**	532	585	–	53	0.09	6.3
9	Q108K:K40L:T51V:T53S:**R58C**	522	589	–	67	0.19	6.2
10	Q108K:K40L:T51V:T53S:**R58Y**	522	587	–	65	0.11	7.1

We postulated that the blue emission is attributed to **ThioPhenol**-PSB and the red emission to **ThioPhenolate**-PSB (**Figure II-9a**), since for non-polar residues such as Leu only the blue emission is observed (**Table II-3**, entry 5). As explained earlier, it is not possible to form the **ThioPhenolate**-PSB complex in solution. Therefore we tried to prove its existence by acidifying the Q108K:K40L:T51V:T53S:R58H/**ThioPhenol** sample. Iminium absorption and emission peak's intensity (at 530 nm and 590 nm, respectively) increase upon

direct acid titration of the sample to pH=4.7 (**Figure II-9b**). However, the intensity of the emission peak at 690 nm decreases with the reduction of pH.

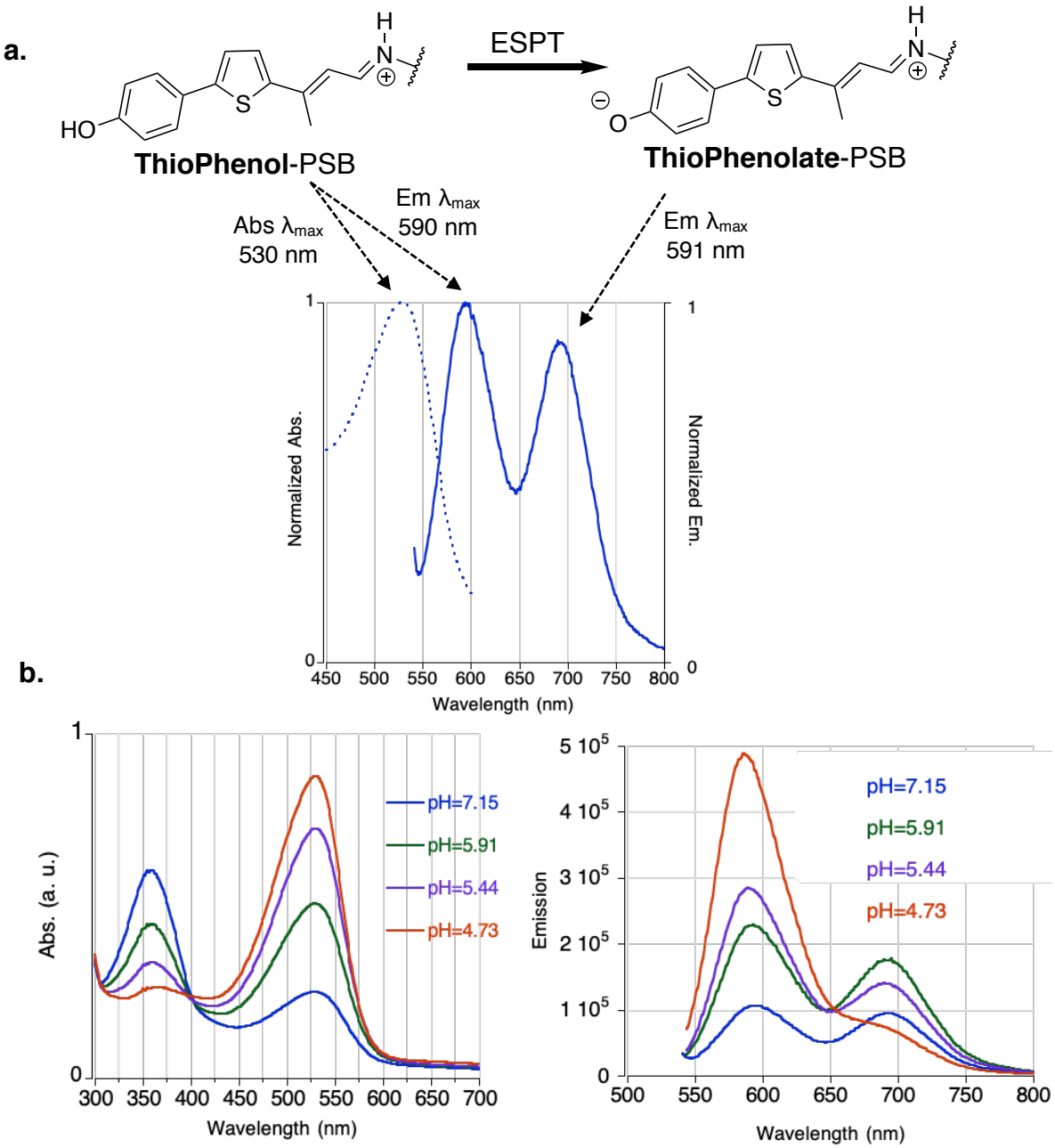


Figure II-9. **a**. **ThioPhenolate**-PSB, the product of ESPT process, and normalized absorption/emission spectra of the iminium of Q108K:K40L:T51V:T53S:R58H/**ThioPhenol** complex. **b**. UV-Vis (left) and fluorescence (right) spectra of the same complex upon acidification of the solution.

We presumed that the phenol moiety becomes deprotonated upon photoexcitation and forms the **ThioPhenolate**-PSB complex due to hydrogen bonding interaction between the sp^2 hybridized nitrogen of histidine's side chain and the hydroxy group of the chromophore. This nitrogen atom is neutral in physiological pH (pK_a~6) and capable of building a hydrogen bond bridge with the hydroxyl group's oxygen and possibly deprotonate it in the excited state. However, the histidine side chain is protonated in acidic pHs and unable to make such interaction, and thus **ThioPhenolate**-PSB emission decreases at pH=4.7. Q108K:K40L:T51V:T53S:R58W/**ThioPhenol** mutant follows the same trend over pH changes but with lower ESPT quantum yield due to the high pK_a of the nitrogen atom in the indole ring.

To compare different mutants' ability to develop a hydrogen-bonding network around the hydroxyl group and generate the ESPT product, we defined Φ_{ESPT} as the fraction of total fluorescence from the excited state **ThioPhenolate**-PSB. Briefly, the fluorescence spectrum was deconvolved to give separate traces representing the emission from the neutral and negatively charged alkoxide, respectively. The area under each trace was then integrated to give the relative percentage of total fluorescence quantum yield. The number derived from the **ThioPhenolate**-PSB was determined as Φ_{ESPT}. The highest Φ_{ESPT} values (majorly yielded red-shifted emission) were obtained for hCRBPII/**ThioPhenol** complexes that contained R58W and R58H mutations (**Table II-3**, entries 2 and 3).

As detailed in our previous studies, bound chromophore could exist as SB or PSB; however, the current designed structure can form four different complexes depending on whether the hydroxyl group is protonated or deprotonated, **ThioPhenol**-SB and **ThioPhenolate**-SB when the chromophore is bound as imine and **ThioPhenol**-PSB and **ThioPhenolate**-PSB as iminium. Except for, **ThioPhenolate**-PSB, the other three complexes' photophysical properties were measured in solution (**section II.5**). Nonetheless, to characterize this system in the protein cavity's complex environment, the sample was basified as well (**Figure II-10**).

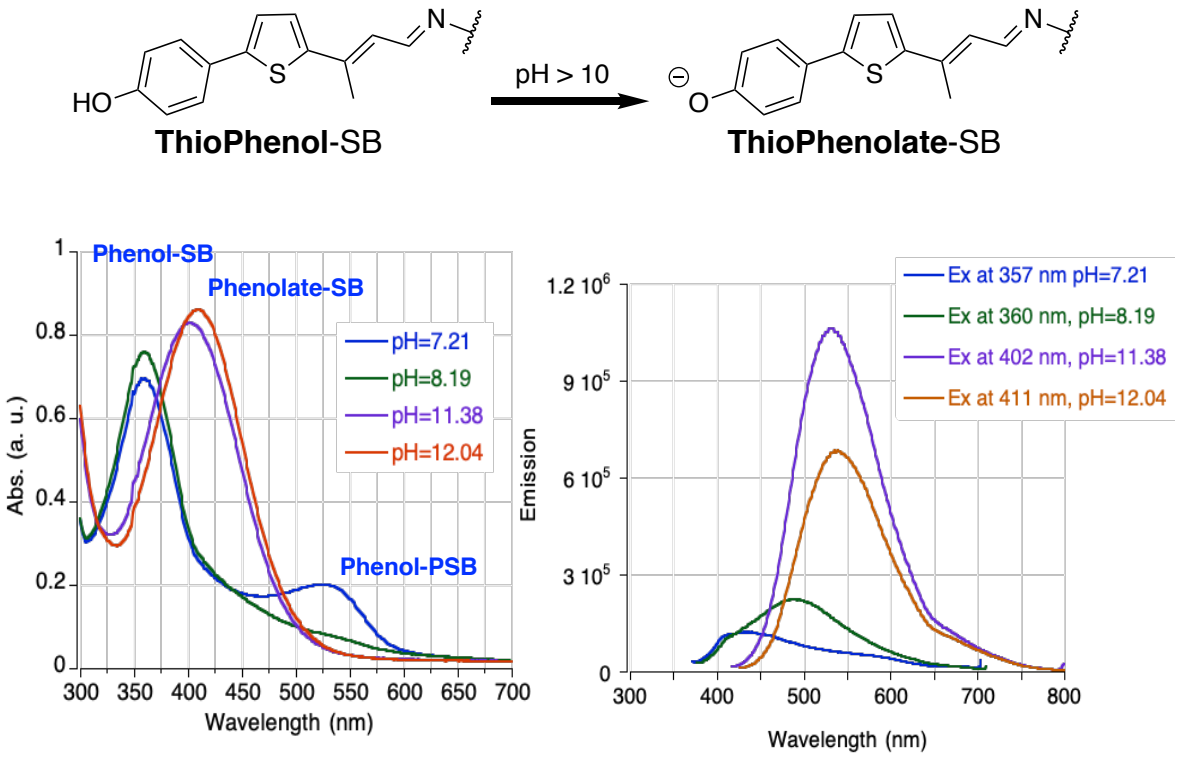


Figure II-10. Formation of the **ThioPhenolate**-SB upon basification of the solution and the UV-Vis (left) and fluorescence (right) spectra of Q108K:K40L:T51V:T53S:R58H/**ThioPhenol** upon base titration.

As shown, basification of the Q108K:K40L:T51V:T53S:R58H/**ThioPhenol** sample from pH=7.2 to pH=8.2 results in the disappearance of PSB absorption peak (~530 nm) and an increase in the SB absorption peak intensity at 360 nm. Further basification of the solution red-shifts the SB absorption by about 50 nm, due to the hydroxyl group deprotonation and **ThioPhenolate**-SB complex formation absorbing at 409 nm, very close to its absorption in EtOH solution (**Table II-2**). An interesting observation is that the emission intensity for the neutral complex, **ThioPhenol**-SB, is very low but enhances about six-fold upon deprotonation of the hydroxyl group. Such results are significantly important for imaging applications as non-specific imine formation has always been the primary source of fluorescence background in our studies. However, with the newly designed structure, it is expected to have the most negligible background fluorescence from off-target imine formation as these species are hardly emissive, and second, the emission is well-separated from the ESPT derived emission.

II.8 The effect of rigidifying the chromophore through Y19W and A33W mutations on the absorption wavelength and Φ_{ESPT}

In our previous studies with Professor Geiger's lab, we demonstrated that aromatic residues such as A33W and Y19W provide tight packing of the chromophore leading to chromophore rigidification and restriction. Besides, the introduction of these bulky residues results in further delocalization of the charge and a red-shift in the bound chromophore wavelength by dispelling some of the ordered water molecules out of the binding cavity as previously reported for

retinylidene variants.[42] We chose to add these mutations, Y19W and A33W, to Q108K:K40L:T51V:T53S:R58H and Q108K:K40L:T51V:T53S:R58W templates to explore their effect on Φ_{ESPT} and other photophysical properties of the corresponding mutants (**Table II-4**).

Table II-4. Spectroscopic change upon the addition of Y19W and A33W mutations.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	SS	Φ_{ESPT}	pK_a
1	KLVS:R58H:Y19W	533	–	698	165	96.8	5.3
2	KLVS:R58H:Y19W:A33W	530	588	694	58/164	31.3	6.0
3	KLVS:R58W:Y19W	529	591	–	62	< 2	5.7
4	KLVS:R58W:Y19W:A33W	534	594	676	60/142	< 2	5.1

[a] KLVS equals to Q108K:K40L:T51V:T53S.

No significant bathochromic shift was observed in the absorption wavelength of resultant mutants. However, interestingly, the addition of Y19W had two opposite effects on the Φ_{ESPT} of the parent mutants. It increased the Φ_{ESPT} to 95% in Q108K:K40L:T51V:T53S:R58H:Y19W mutant, the highest Φ_{ESPT} acquired up to this point, but it ceased the ESPT of the hydroxyl group and decreased the Φ_{ESPT} to less than 2% in Q108K:K40L:T51V:T53S:R58W:Y19W (entries 1 and 3). The addition of both Y19W and A33W mutations reduced the Φ_{ESPT} (entries 2 and 4).

Unfortunately, the fluorescence quantum efficiency of the red-shifted emission (~690 nm) generated upon ESPT of the hydroxyl group is less than 5%, and the brightness needed to be increased to be applicable for imaging purposes.

We pursued two approaches to address this issue by expanding the hydrogen bonding network around the hydroxyl group and increasing the iminium pK_a.

II.9 Exploring the effects of L117E and L117D mutations on the iminium pK_a and Φ_{ESPT}

As described in **Section II-6**, the protein host should be reengineered to maintain a high iminium pK_a (>8.5) since the fluorescent complex, **ThioPhenolate**-PSB, is produced upon photoexcitation of the iminium. Clearly, more iminium concentration leads to more ESPT product; however, as shown in **Table II-3** and **Table II-4**, mutants that yielded the highest Φ_{ESPT} exhibit relatively low iminium pK_a (<6), resulting in the formation of imine more than twice of iminium at neutral pH according to the according to the Henderson-Hasselbalch equation (see **Figure II-10** to compare SB vs PSB ratio).

Thus, we sought to investigate the effect of high iminium pK_a on the Φ_{ESPT}. Previous studies in CRABPII showed that L121E was successful at enhancing the rate of iminium formation.[48] Mutation of the analogous residue, L117, in hCRBPII to aspartic acid and glutamic acid residues increased the rate of iminium formation with **ThioFluor** as well and retained the high iminium pK_a. To this end, L117E was added to the hCRBPII mutants showing the highest Φ_{ESPT}, Q108K:K40L:T51V:T53S:R58H, and Q108K:K40L:T51V:T53S:R58H:Y19W (**Table II-5**, entries 1 and 2). In addition, several other mutants contained R58H and R58W and showed high iminium pK_a with **ThioFluor**, were tested with the new structure (**Table II-5**, entries 3-7).

The PSB absorption wavelength blue-shifted by about 70 nm for entries 1-3 and 55 nm for entries 4 and 5 upon the addition of L117E mutation. An average blue-shift of 35 nm was observed for the PSB emission.

Table II-5. Spectroscopic change upon addition of L117E mutation.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	SS	Φ_{ESPT}	pK_a
1	KLVS:R58H:L117E	460	552	92	< 2	6.7
2	KLVS:R58H:Y19W:L117E	456	557	101	5.9	10.3
3	KLVS:R58W:Y19W:L117E	468	554	86	< 2	10.0
4	KLVS:R58W:Y19W:L117E:A33W	477	563	86	< 2	9.4
5	KLVS:R58H:Y19W:L117E:A33W	478	559	81	3.8	7.0
6	KLVS:R58W:Y19W:L117D:Q4F	462	549	87	< 2	10.0
7	KLVS:R58H:Y19W:L117D:Q4F	378	548	178	27.3	5.4

[a] KLVS equals to Q108K:K40L:T51V:T53S.

We hypothesized it is due to aspartate acid and glutamate's ability to localize the positive charge on iminium, as we previously showed that blue-shifted absorption wavelengths occur from localization of charge. Regardless, none of the mutants showed the red-shifted emission corresponding to the **ThioPhenolate**-PSB complex resulting in very low Φ_{ESPT} (<5%). As shown for KLVS:R58W:Y19W:L117E:A33W/**ThioPhenol** complex (entry 4), excitation of the PSB maximum yields the blue-shifted emission at 563 nm (**Figure II-11**).

Interestingly, KLVS:R58H:Y19W:L117D:Q4F mutant (entry 7) shows a very low iminium pK_a, and **ThioPhenol**-SB is the main species at neutral pH. However, excitation of the imine at 378 nm yields the PSB emission with a large Stokes shift (178 nm) as a result of the ESPT of the hydroxyl group. For unknown reasons, it appears that increased iminium pK_a stopped the ESPT of the hydroxyl group.

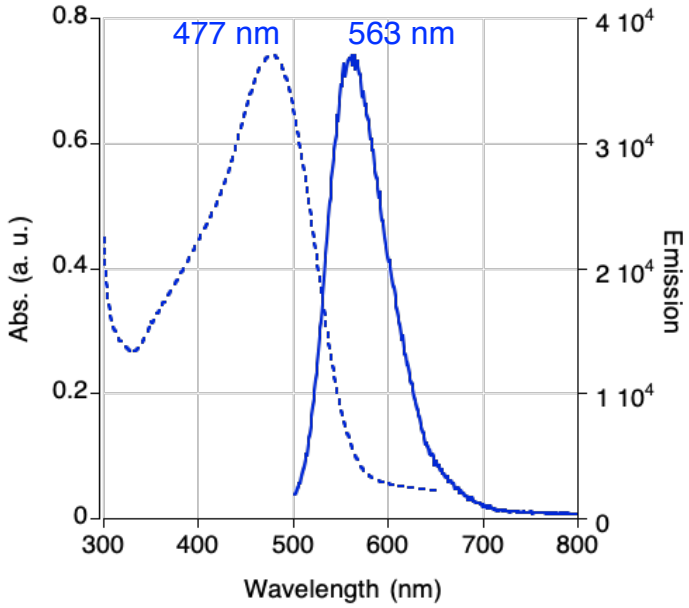


Figure II-11. Absorption and emission spectra of KLVS:R58W:Y19W:L117E:A33W/**ThioPhenol** complex upon PSB excitation.

II.9.1 Basic residues around the hydroxyl group

In the next step, we attempted to expand the hydrogen bonding network around the hydroxyl group to see if the nearby basic residues could activate the bound chromophore's photoacidic properties.

Fortuitously, Dr. Alireza Ghanbarpour provided the crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioPhenol** that was used to

investigate positions near the hydroxy group inside the hCRBPII cavity for the introduction of basic residues (**Figure II-12**).

As shown in **Figure II-12**, M20, T29, A33, S76, and L77 are within 8 Å distance to the hydroxyl group, which were individually mutated to basic amino acids such as lysine, arginine, and histidine. The mutations were added to the Q108K:K40L:T51V:T53S:R58W:Y19W:L117E:A33W and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F templates that showed the highest pK_a to ensure most of the bound chromophore is in the protonated state.

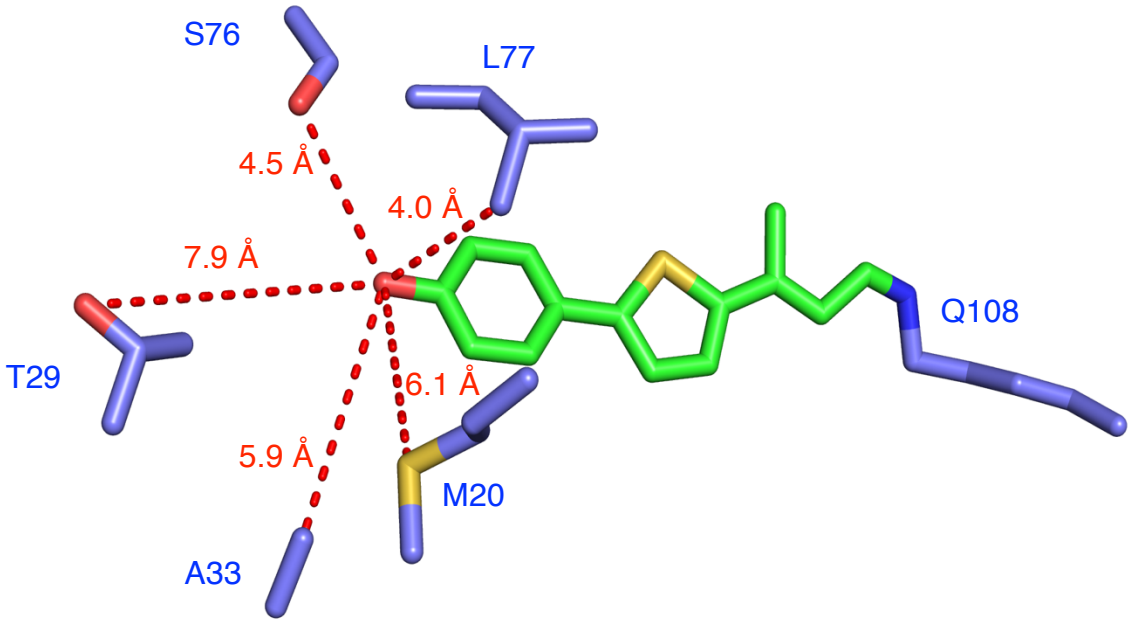


Figure II-12. Closest residues to **ThioPhenol**'s hydroxyl group in the crystal structure of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioPhenol** complex.

Unfortunately, a number of the mutants led to insoluble protein expression; expressed proteins' spectroscopic properties are shown in **Table II-6**. Nonetheless, none of the mutants found success in activating **ThioPhenol**-PSB

as a photoacid, and all of them showed the PSB emission with less than 100 nm Stokes shift and Φ_{ESPT} less than 2 percent.

Table II-6. Spectroscopic properties of the protein mutants designed to increase the photoacidity of the bound **ThioPhenol** by introducing basic residues.

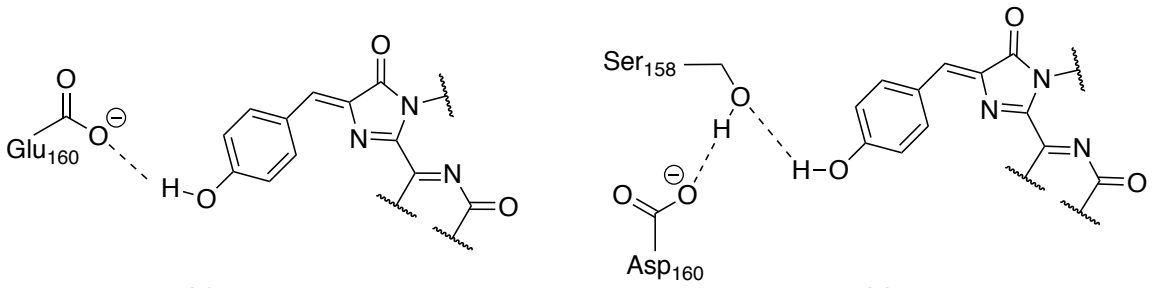
Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	SS	Φ_{ESPT}	pK_a
1	KLVS:R58W:Y19W:L117E:A33W:**L77K**	471	562	91	< 2	9.2
2	KLVS:R58W:Y19W:L117E:A33W:**S76K**	473	563	90	< 2	4.8
3	KLVS:R58W:Y19W:L117E:A33W:**S76R**	472	564	92	3.2	9.1
4	KLVS:R58W:Y19W:L117E:A33W:**S76H**	472	565	93	< 2	9.4
5	KLVS:R58W:Y19W:L117E:A33W:**T29H**	470	558	88	< 2	10.1
6	KLVS:R58W:Y19W:L117E:A33W:**T29R**	478	569	91	< 2	9.9
7	KLVS:R58W:Y19W:L117E:A33H	467	549	82	3.6	9.2
8	KLVS:R58W:Y19W:L117D:Q4F:**L77H**	468	557	89	< 2	9.8
9	KLVS:R58W:Y19W:L117D:Q4F:**S76R**	464	553	89	< 2	8.1
10	KLVS:R58W:Y19W:L117D:Q4F:**S76H**	463	559	96	3.1	7.9
11	KLVS:R58W:Y19W:L117D:Q4F:**A33H**	465	554	89	< 2	9.5
12	KLVS:R58W:Y19W:L117D:Q4F:**T29K**	463	552	89	5.3	8.0
13	KLVS:R58W:Y19W:L117D:Q4F:**M20K**	483	562	79	< 2	10.5
14	KLVS:R58W:Y19W:L117D:Q4F:**M20H**	464	555	91	4.8	9.3

[a] KLVS equals to Q108K:K40L:T51V:T53S.

Presumably, due to the high pK_a of lysine and arginine side chains' (about 10.5 and 12, respectively), they are protonated in neutral pH and unable to act as a proton acceptor of the hydroxyl group upon excitation. Furthermore, although R58H mutation was the most successful in deprotonating the hydroxyl group, insertion of histidine at these selected positions was not fruitful.

II.9.2 Acidic residues around the hydroxyl group

As described in **Section I-1**, Large Stokes shift fluorescent proteins such as mKeima and LSS-mKates exhibit high similarity in the interactions between the chromophores and neighboring residues. In this class of fluorescent proteins, the active site residues have been optimized such that the chromophore is stabilized through hydrogen bonding with the acidic residues (**Figure II-13**). Carboxylate moiety in acidic residues acts as a proton acceptor as the chromophore's pK_a is decreased several units upon excitation.[93,94] Thus, we next sought to investigate whether the selected residues' mutation to aspartic and glutamic acid can enhance the ESPT process and fluorescent properties of **ThioPhenol**-PSB.



LSS-mKate 1 **LSS-mKate 2**

Figure II-13. Hydrogen bonding network responsible for excited state proton transfer in LSS-mKate 1 and LSS-mKate 2.

Several of the designed mutants led to insoluble protein expressions, as evidenced by the presence of hCRBPII in the pellet of lysed cells via gel electrophoresis. Nevertheless, no enhancement was observed in the chromophore's ESPT activity bound to the expressed ones (**Table II-7**).

Upon examining all protein mutants and their photophysical properties till now, we realized none of the variants with high iminium pK_a could yield the ESPT derived red-shifted emission. Presumably, it is the presence of L117E and L117D mutations that prevent the ESPT, although the reasons are unknown at this time. Hence, we chose to first expand a well-developed water-mediated hydrogen bonding network around the hydroxyl group, enhance the Φ_{ESPT} as much as possible, and then increase the iminium pK_a through other approaches except the L117E/D mutation.

Table II-7. Spectroscopic properties of the protein mutants designed to increase the photoacidity of the bound **ThioPhenol** by introducing acidic residues.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	SS	Φ_{ESPT}	pK_a
1	KLVS:R58W:Y19W:L117D:Q4F:**L77D**	466	554	88	< 2	8.4
2	KLVS:R58W:Y19W:L117D:Q4F:**S76E**	460	557	97	< 2	8.5
3	KLVS:R58W:Y19W:L117D:Q4F:**A33E**	463	552	89	< 2	8.1
4	KLVS:R58W:Y19W:L117D:Q4F:**T29E**	467	560	93	< 2	9.6
5	KLVS:R58W:Y19W:L117D:Q4F:**T29D**	465	556	91	< 2	9.5

[a] KLVS equals to Q108K:K40L:T51V:T53S.

II.10 Investigating the effects of proton acceptor residues at positions A33 and F16 on the Φ_{ESPT}

Among the five selected closest residues to the hydroxyl group (M20, T29, A33, S76, and L77), mutation of A33 resulted in higher protein yield and showed more stability toward the pH changes of the solution required for pK_a titrations. Hence, we sought to investigate whether its mutation to proton acceptor residues such as histidine, serine, glutamic acid, and tyrosine can improve the Φ_{ESPT}. As

discussed previously, these residues proved to facilitate deprotonation of the phenol moiety in various FPs through direct or water-mediated hydrogen bonding in the excited state. Thus, A33 was substituted with those residues in Q108K:K40L:T51V:T53S:R58H and Q108K:K40L:T51V:T53S:R58H:Y19W templates. The highest Φ_{ESPT} was obtained with Q108K:K40L:T51V:T53S:R58H:Y19W:**A33H** mutant (**Table II-8,** entry 5).

Table II-8. Spectroscopic change upon mutation of A33 to histidine, serine and glutamic acid, and tyrosine.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	SS	Φ_{ESPT}	pK_a
1	KLVS:R58H:**A33H**	530	586	693	56/163	54.2	6.3
2	KLVS:R58H:**A33S**	529	595	691	66/162	55.1	6.5
3	KLVS:R58H:**A33E**	523	584	694	61/171	51.4	6.1
4	KLVS:R58H:**A33Y**	530	593	-	63	< 2	7.1
5	KLVS:R58H:Y19W:**A33H**	533	588	694	55/161	96.8	6.5
6	KLVS:R58H:Y19W:**A33S**	534	590	693	56/159	89.1	6.4
7	KLVS:R58H:Y19W:**A33E**	531	593	695	62/164	95.7	6.4
8	KLVS:R58H:Y19W:**A33Y**	533	596	693	63/160	53.6	6.8

[a] KLVS equals to Q108K:K40L:T51V:T53S.

As was expected, removing L117E/L117D mutations recovered the red emission, although the iminium pK_a reduced by about two to three units. Alternative protein designs to increase the pK_a are discussed in **Section II.11.2**.

Most of the protein variants shown in **Table II-8** provide improved Φ_{ESPT}. However, the ESPT efficiency is higher when Y19W is present. Therefore,

Q108K:K40L:T51V:T53S:R58H:Y19W sequence is retained in most mutants tested. Noteworthy, mutation of A33 to tyrosine enhanced the minimum pK_a.

To our delight, the crystal structure of Q108K:K40L:T51V:T53S:R58H:Y19W:A33Y/**ThioPhenol** was obtained and solved by Dr. Nona Ehyaei. Meticulous examination of the crystal structure revealed that F16 residue is very close to the hydroxyl group. Its mutation to tyrosine via mutagenesis in Pymol suggested a 3.4 Å distance, which is the shortest distance compared to previously selected positions (**Figure II-14**).

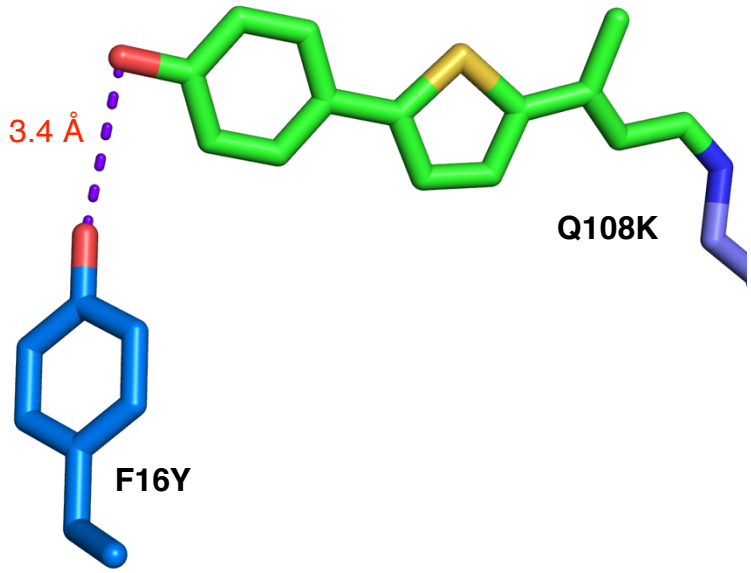


Figure II-14. The crystal structure of Q108K:K40L:T51V:T53S:R58H:Y19W:A33Y/**ThioPhenol** complex with F16 mutated to tyrosine via mutagenesis in Pymol.

Therefore, we introduced F16Y into the parent template mutant: Q108K:K40L:T51V:T53S:R58H (**Table II-9**, entry 1). Additionally, in order to expand the hydrogen bonding network around the hydroxyl group, A33 was mutated to amino acid residues having proton acceptor side chains in combination with F16Y.

Table II-9. Spectroscopic change upon F16Y addition.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	SS	Φ_{ESPT}	pK_a
1	KLVS:R58H:F16Y	530	596	677	66/147	62.1	7.0
2	KLVS:R58H:A33H:F16Y	533	589	677	56/144	49.8	7.2
3	KLVS:R58H:Y19W:A33Y:F16Y	531	594	678	63/147	65.2	6.9
4	KLVS:R58H:Y19W:A33W:F16Y	533	590	677	57/144	47.3	7.0
5	KLVS:R58H:Y19W:A33H:F16Y	530	-	680	150	97.6	6.9

[a] KLVS equals to Q108K:K40L:T51V:T53S.

Remarkably, with the newly designed mutants, Φ_{ESPT} increased to more than 95% (**Table II-9**, entry 5), meaning almost all bound chromophores as **ThioPhenol**-PSB converts to **ThioPhenolate**-PSB upon excitation. Comparing the Φ_{ESPT} of entries 3,4, and 5 where A33 is mutated to tyrosine, tryptophan, and histidine, respectively, proves that A33H works best in conjugation with F16Y and R58H mutations. The triple mutations (F16Y, A33H, and R58H) must have aligned well toward the hydroxyl group and form a water-mediated network capable of activating the phenol moiety as a photoacid.

Notably, although absorption and the blue emission wavelength corresponding to the **ThioPhenol**-PSB complex remained the same, the red

emission wavelength hypsochromically shifted about 15 nm from the parent mutants upon F16Y addition. In our previous studies with hCRBPII/retinal, Dr. Wenjing Wang demonstrated that an even distribution of electrostatic potential across the polyene led to a bathochromic shift in absorption while localizing the cation on the iminium nitrogen resulted in blue-shifted wavelengths.[42 95]

The emission wavelength is also correlated to the degree of intramolecular charge transfer (ICT) from the electron donor to the electron acceptor.[96,97] In this system, deprotonation of the hydroxyl moiety upon excitation leads to an ICT from the correspondent alkoxide to the iminium, which results in red-shifted emission and larger Stokes shifts. A 15 nm hypsochromic shift in the emission wavelength shows that the charge transfer has decreased, and the negative charge is more localized on oxygen rather than the whole chromophore toward the iminium. We hypothesized that this is due to the electrostatic interaction between the positively charged protonated 16Y and the alkoxide group, the product of the ESPT process (**Figure II-15**).

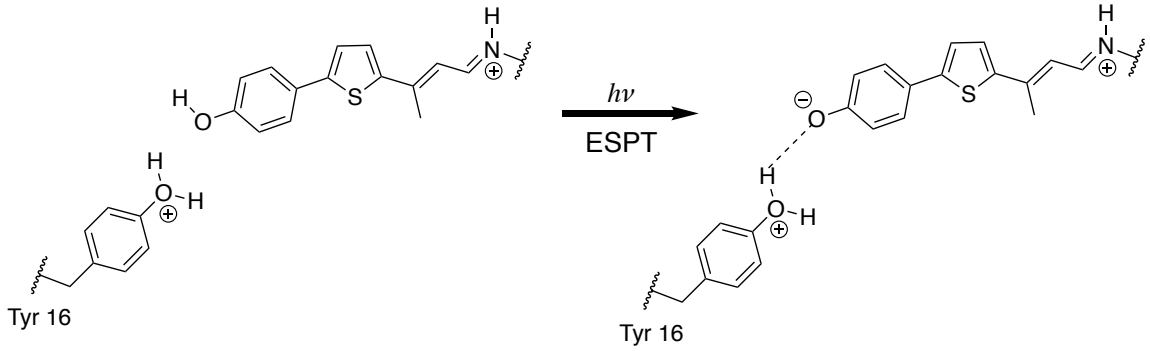


Figure II-15. Schematic representation of the negative charge localization on the oxygen atom due to the electrostatic interaction with Tyr 16 side chain.

We sought to explore whether tyrosine is the best choice for position 16; hence, other proton acceptor residues such as histidine, glutamic acid, tryptophan, and cysteine were introduced to the template protein mutants, and their fluorescence properties were measured (**Table II-10**).

Table II-10. Spectroscopic changes as a result of mutating F16.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	SS	Φ_{ESPT}
1	KLVS:R58H:**F16E**	530	610	-	80	< 2
2	KLVS:R58H:A33H:**F16E**	531	611	-	80	< 2
3	KLVS:R58H:A33H:**F16W**	531	583	675	51/144	40.2
4	KLVS:R58H:Y19W:A33W:**F16H**	531	591	678	60/147	46.2
5	KLVS:R58H:Y19W:A33W:**F16E**	530	595	-	65	35.3
6	KLVS:R58H:Y19W:A33Y:**F16E**	530		-	72	23.2

[a] KLVS equals to Q108K:K40L:T51V:T53S.

Unfortunately, most of the designed mutants led to insoluble protein expressions. For example, none of the protein mutants containing F16C mutation were expressed, and the Φ_{ESPT} for the successfully expressed ones was less than 50%. Presumably, since F16 is located at the interior pocket, it can be envisioned that mutation of F16 to tryptophan or histidine would lead to the steric clash of F16W/F16H with R58H and may be able to flip R58H outward and thus less ESPT of the hydroxyl group.

Nevertheless, one interesting new observation was made concerning the addition of F16E. The emission of the **ThioPhenol**-PSB complexes red-shifts about 10 nm when F16 is mutated to glutamic acid. It seems the interaction

between the carboxylate and the hydroxyl group leads to a partial proton transfer. As a result, the emission wavelength (~610 nm) is between the **ThioPhenol**-PSB and **ThioPhenolate**-PSB complexes' emission.

A most interesting observation was made when overlaying the crystal structures of Q108K:K40L:T51V:T53S:R58H:Y19W:A33Y/**ThioPhenol** (magenta) with Q108K:K40L:T51V:T53S:R58W/**ThioPhenol** (green), obtained and solved by Dr. Alireza Ghanbarpour, indicates that the trajectories of 58H and 58W in these two structures are vastly different (**Figure II-16**). The magenta histidine adopts the flipped conformation outward to avoid the steric interaction with 33Y residue. Yet, it shows a slightly higher Q_{ESPT} compared to Q108K:K40L:T51V:T53S:R58H mutant (53% VS 50%).

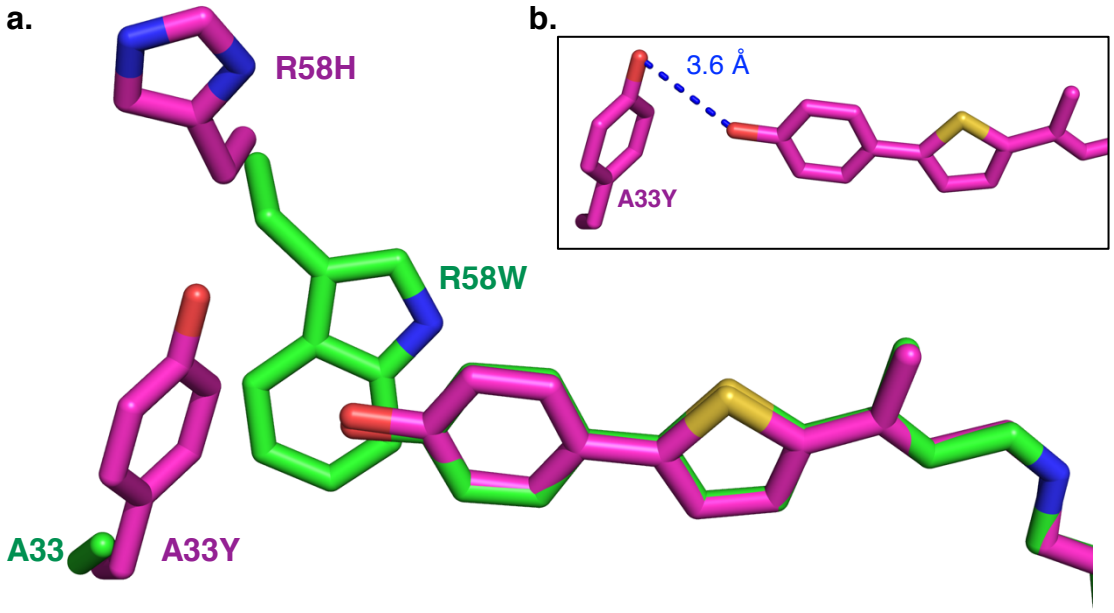


Figure II-16. **a**. Overlay of Q108K:K40L:T51V:T53S:R58W/**ThioPhenol** (magenta) and Q108K:K40L:T51V:T53S:R58H:Y19W:A33Y/**ThioPhenol** (green). **b**. The distance between the F16Y and the hydroxyl group.

As R58H and A33Y showed almost the same efficiency in deprotonating the hydroxyl group, we decided to probe the effect of more tyrosine residues around the hydroxyl group on the Φ_{ESPT}. To this end, R58Y, A33Y, and F16Y mutations were added to the Q108K:K40L:T51V:T53S:R58H template (**Table II-11**).

Table II-11. Spectroscopic changes as a result of mutating R58, F16, and A33 residues to tyrosine.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	Φ_{ESPT}
1	KLVS:R58H	530	590	691	50.2
2	KLVS:R58H:**F16Y**	530	596	677	62.1
3	KLVS:R58H:Y19W:**A33Y**	533	596	693	53.3
4	KLVS:R58H:**A33Y**:**F16Y**	536	598	-	28.8
5	KLVS:**R58Y**:Y19W:**A33Y**	531	591	-	< 2
6	KLVS:**R58Y**:**F16Y**	535	590	-	24.6
7	KLVS:**R58Y**:**A33Y**:**F16Y**	545	593	-	9.3

[a] KLVS equals to Q108K:K40L:T51V:T53S.

Unfortunately, double or triple mutations to tyrosine did not lead to an expanded hydrogen bonding network, and in fact, the resultant mutants yielded very poor Φ_{ESPT} (**Table II-11**). Presumably, tyrosine residues turn away from the cavity and further from the hydroxyl group to relieve the steric hindrance. Mutants with R58Y residue resulted in the lowest Φ_{ESPT}, but the insertion of a single tyrosine at positions 16 and 33 enhances the ESPT process (entries 2 and 3). Although, as shown previously, F16Y introduces a blue-shift of 20 nm to the ESPT derived emission.

II.11 Attempts to enhance the red-shifted emission quantum efficiency

Thus far, we have discovered that R58H is the essential mutation in the ESPT process. As shown earlier, Φ_{ESPT} would decrease dramatically or was eliminated in the absence of this mutation. We also realized that R58H, A33H, and F16Y are the three key mutations required to get the maximum ESPT quantum yield. The Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y/**ThioPhenol** complex containing these residues yielded the highest Φ_{ESPT} (>90%); however, the ESPT derived far-red emission (680-700 nm) fluorescence quantum yield is less than 5% for this mutant. Our attempts were then focused on expanding the water-mediated hydrogen bonding network around the hydroxyl group via point mutagenesis. As we presumed, suppressing the blue-shifted non-ESPT emission and increasing the Φ_{ESPT} will lead to higher fluorescence quantum efficiency.

II.11.1 Through expansion of water-mediated hydrogen bonding network around the R58H residue

We sought to investigate whether providing a proton transfer network around R58H would improve the Φ_{ESPT}. It was hypothesized that hydrogen bonds from ionizable amino acids to 58H directly or through water molecules would stabilize protonated histidine in the excited state and increase its basicity. *In silico* modeling of R58H in the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioPhenol** complex helped to identify such residues: T53S, S55, and Q38, all are located within 5 Å from the 58H, with presumably close interactions (**Figure II-17a**). We chose to introduce amino acids

having proton acceptor side chains at these positions such as histidine, tyrosine, and glutamic acid, to facilitate deprotonation of the hydroxyl group upon excitation.

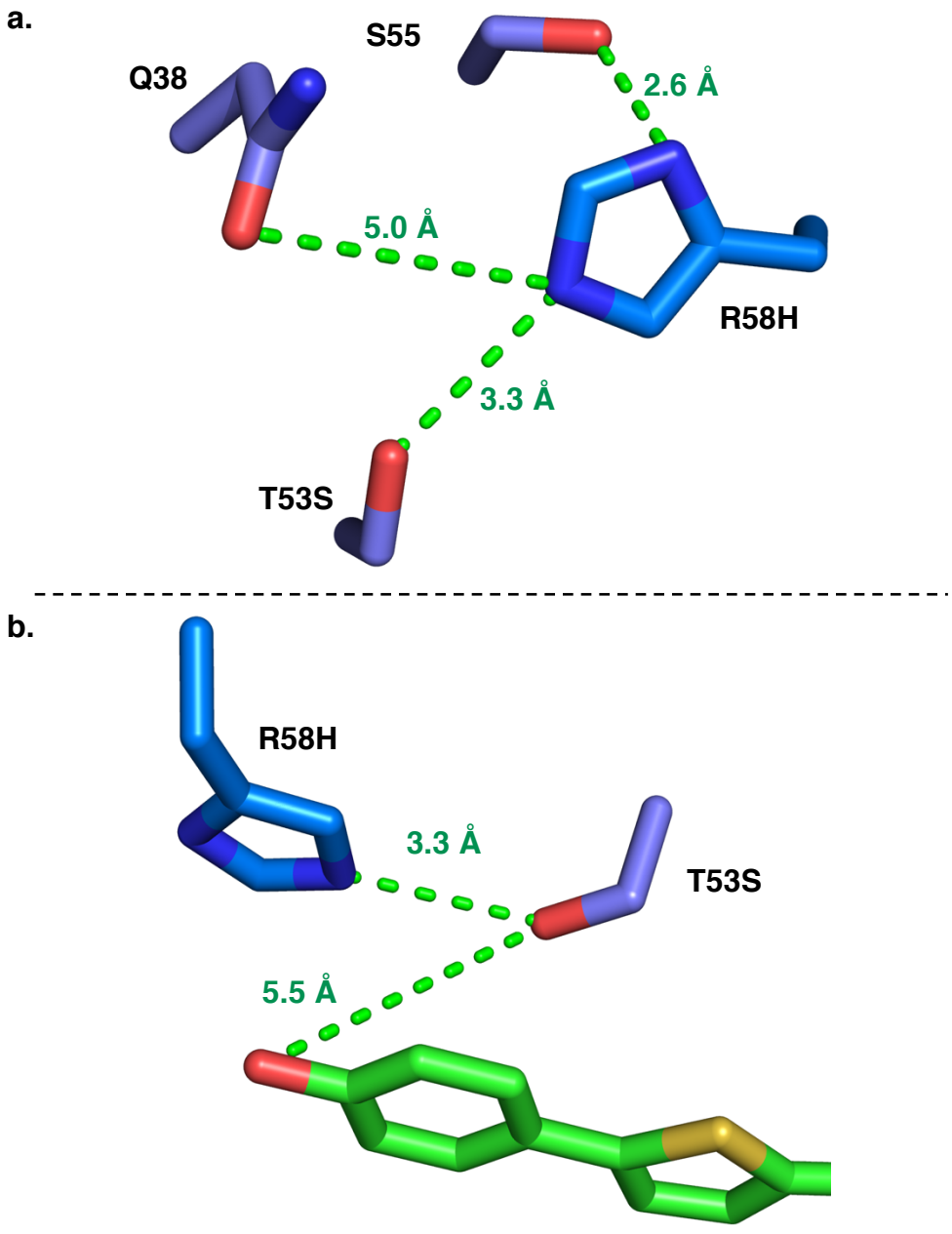


Figure II-17. **a**. *In silico* modeling of R58H in the crystal structure of Q108K:K40L:T51V:T53S:R58W/**ThioPhenol** and the distances between S55, Q38, and T53S residues to 58H. **b**. T53S distances to R58H residue and **ThioPhenol** hydroxyl group.

To this point, the template Q108K:K40L:T51V:T53S was retained in all protein engineering efforts in this study, as recently Dr. Elizabeth Santos showed T51V:T53S double mutation leads to a red-shift in the absorption and emission wavelength of bound **ThioFlour**. In addition, T51V proved to be capable of monomerizing hCRBPII variants.[87]

Earlier, Dr. Wenjing Wang demonstrated that T51V mutation results in bathochromic shifts in bound retinal absorption wavelength. As seen in the crystal structure of Q108K:K40L/retinal (**Figure II-18**), T53 makes a water-mediated hydrogen bond with T51, which is abolished when T51 is mutated to valine, leading to the red-shifted wavelength.

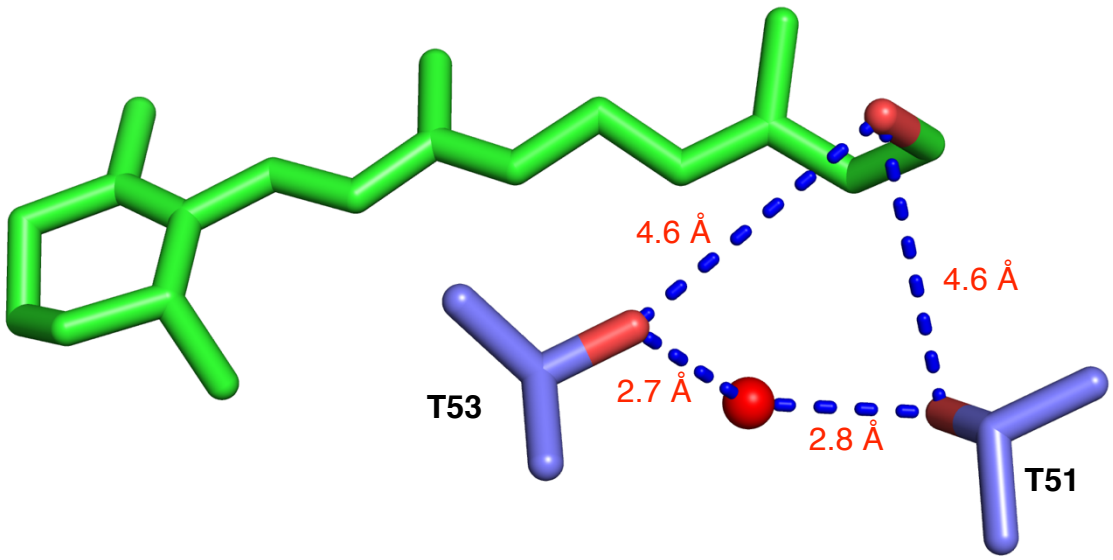


Figure II-18. Crystal structures of WT-hCRBPII bound with all-*trans*-retinol showing the internal hydrogen bonding network among the T51 and T53 side chain residues and retinol.

II.11.1 A. T53 mutations

T53S is situated near the phenol moiety and shows a close distance to both the hydroxyl group and 58H, 5.5 and 3.3 Å, respectively (**Figure II-17b**). We decided to mutate T53 to Tyr, Glu, and His residues in three different templates: Q108K:K40L:T51V:T53S:R58H, Q108K:K40L:T51V:T53S:R58H:Y19W:A33H, and Q108K:K40L:T51V:T53S:R58H:Y19W:A33W in order to study the effect of T53 mutations on emission wavelength and the Φ_{ESPT} (**Table II-12**).

Table II-12. Spectroscopic changes as a result of mutating T53 to histidine, tyrosine, and glutamic acid.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	SS	Φ_{ESPT}	Φ
1	KLV:**T53S**:R58H	530	590	691	60/161	50.2	0.01
2	KLV:**T53Y**:R58H	525	594	681	69/156	48.7	< 0.01
3	KLV:**T53E**:R58H	526	589	703	63/177	52.3	< 0.01
4	KLV:**T53S**:R58H:Y19W:A33H	533	588	693	55/161	96.8	< 0.01
5	KLV:**T53H**:R58H:Y19W:A33H	520	586	691	66/171	48.9	< 0.01
6	KLV:**T53Y**:R58H:Y19W:A33H	527	590	693	63/166	91.6	< 0.01
7	KLV:**T53E**:R58H:Y19W:A33H	531	-	704	60/173	97.8	0.01
8	KLV:**T53S**:R58H:Y19W:A33W	530	588	694	58/164	31.3	0.01
9	KLV:**T53H**:R58H:Y19W:A33W	521	592	-	71	18.7	< 0.01
10	KLV:**T53Y**:R58H:Y19W:A33W	528	599	690	71/162	68.3	< 0.01
11	KLV:**T53E**:R58H:Y19W:A33W	521	587	702	66/180	84.8	0.01

[a] KLV equals to Q108K:K40L:T51V.

As shown, T53H mutation reduced the Φ_{ESPT} dramatically (entries 5 and 9); it also decreased the protein expression yield. On the other hand, T53E mutation enhanced both ESPT and fluorescence quantum efficiency (entries 3, 7, and 11). T53Y mutation does not provide such interaction, and proteins with 53Y show almost the same Φ_{ESPT} as their parent templates. Another interesting observation is that T53E mutation bathochromicaly shifts the ESPT emission by about 7 nm and yields the most red-shifted emission wavelength (entries 3, 7, and 11). The glutamate side chain must be oriented toward the hydroxyl group facilitating its deprotonation leading to a stronger ICT.

Table II-13. Spectroscopic changes as a result of mutating T53.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	SS	Φ_{ESPT}	Φ
1	KLVH:**T53S**:Y19W:A33H:F16Y	530	-	680	150	97.6	0.01
2	KLVH:**T53Y**:Y19W:A33H:F16Y	525	581	682	56/157	98.2	0.01
3	KLVH:**T53E**:Y19W:A33H:F16Y	524	583	681	59/157	97.5	0.01
4	KLVH:**T53S**:Y19W:A33W:F16Y	533	590	677	57/144	47.3	0.01
5	KLVH:**T53E**:Y19W:A33W:F16Y	521	592	679	71/158	46.9	< 0.01
6	KLVH:**T53S**:Y19W:A33Y	533	596	693	63/160	53.3	0.02
7	KLVH:**T53E**:Y19W:A33Y	526	590	702	64/176	78.2	0.01
8	KLVH:**T53S**:A33H:F16Y	533	589	677	56/144	49.8	0.02
9	KLVH:**T53E**:A33H:F16Y	525	588	682	63/157	57.1	< 0.01
10	KLVH:**T53S**:A33H	530	586	693	56/163	54.2	0.01
11	KLVH:**T53E**:A33H	525	588	702	63/177	66.9	0.01

[a] KLVH equals to Q108K:K40L:T51V:R58H.

To further explore the effect of T53E and T53Y mutations on the formation of **ThioPhenolate**-PSB and the fluorescence quantum yield, other protein

templates were tested, and their spectral properties were compared to the parent mutants (**Table II-13**).

As was observed before, a blue-shift of 15-20 nm happens upon the addition of F16Y mutation; for example, see entries 5 and 6, **Table II-13**. Nevertheless, unfortunately, neither quantum efficiency nor Φ_{ESPT} changed markedly in these newly designed mutants.

Q108K:K40L:T51V:**T53E**:R58H:Y19W:A33H/**ThioPhenol** complex yielded the highest Φ_{ESPT} and fluorescence quantum efficiency when comparing to the other protein mutants in **Tables II-12** and **II-13**. However, the addition of F16Y to this mutant did not enhance its fluorescence properties (**Table II-13**, entry 3). In general, it seems that in this study, the effect of each individual residue does not add up, and it is the combination of residues that improve the ESPT process.

II.11.1 B. S55 mutations

As was indicated in **Figure II-17a**, another residue in the close vicinity of R58H is S55, which shows a distance of 2.6 Å to the histidine side chain. However, unfortunately, the expansion of the hydrogen bonding network through this residue was unsuccessful. S55 mutation leads to majorly dimer formations (**Table II-14**), which are not suitable for imaging purposes. Nonetheless, we decided to measure the fluoresce properties of the PSB in the expressed dimers to see how the Φ_{ESPT} varies compared to the monomeric forms of the protein (**Table II-15**).

As shown, dimeric proteins showed low Φ_{ESPT}; another interesting new observation is that even for the same mutant, the dimer shows less Φ_{ESPT} than the monomeric form (**Table II-15**, entries 3 and 4). While the dimeric species of hCRBPII are not desirable, these results seem to indicate that Φ_{ESPT} is critically dependent on the environment around the bound chromophore, changing one residue even if it is far from the hydroxyl group can affect its deprotonation.

Table II-14. Protein expression yields of hCRBPII mutants upon S55 mutation to histidine, tyrosine, and glutamic acid.

hCRBPII Mutant [a]	Monomer (mg/L)	Dimer (mg/L)	Mol% dimer
KLVS:**S55E**:R58H:A33H	-	15 .4	> 98
KLVS:**S55Y**:R58H	-	19.2	> 98
KLVS:**S55E**:R58H	3.1	16.2	> 95
KLVS:**S55H**:R58H	-	17.3	> 98

[a] KLVS equals to Q108K:K40L:T51V:T53S.

Table II-15. Spectroscopic properties of the PSB as a result of mutation of S55.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	SS	Φ_{ESPT}
1	KLVS:**S55E**:R58H:A33H (dimer)	530	590	-	60	27.4
2	KLVS:**S55Y**:R58H (dimer)	530	591	-	61	18.5
3	KLVS:**S55E**:R58H (dimer)	529	589	-	60	31.2
4	KLVS:**S55E**:R58H (monomer)	523	590	674	67/151	50.2
5	KLVS:**S55H**:R58H (dimer)	532	593	-	61	22.7

[a] KLVS equals to Q108K:K40L:T51V:T53S.

II.11.1 C. Q38 mutations

A water-mediated hydrogen bonding network between Q38 and Q128 is observed in most hCRBPII mutants if the two residues are maintained (**Figure II-19**).

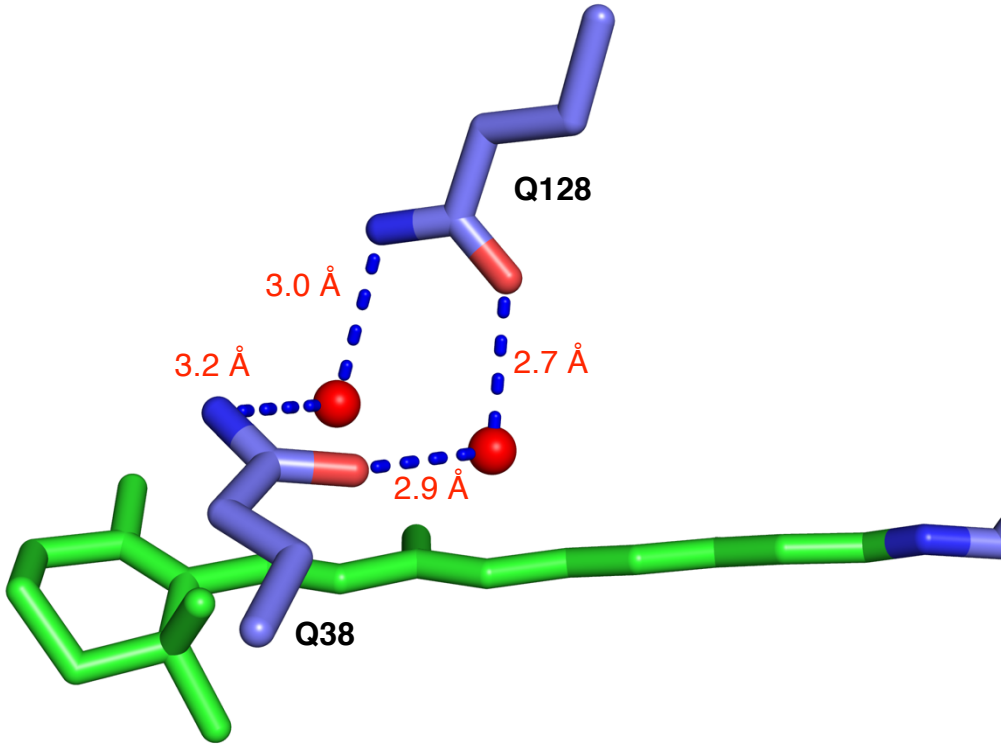


Figure II-19. Water mediated hydrogen bonding between Q38 and Q128 in the crystal structure of hCRBPII-Q108K:K40L/retinal PDB 4EXZ.

The addition of T53S mutation disturbs part of the conserved water-mediated network through the formation of another tight hydrogen bond between T53S and Q38. An overlay of the crystal structure of Q108K:K40L/retinal with Q108K:K40L:T51V:T53S:R58W/**ThioPhenol** indicates that while Q128 and Q38 maintain almost the same conformation in both structures, the 53S side chain has rotated toward Q38 to form the hydrogen bond with Q38 (**Figure II-20**).

As shown earlier in **Figure II-17**, the hydrogen bonding interaction between Q38 and R58H leads to the imidazoline ring of the histidine to adopt a parallel trajectory toward the phenol moiety. We sought to examine whether Q38 mutation and elimination of this interaction would result in the imidazoline ring rotation toward the hydroxyl group and consequently shorter distance.

Table II-16. Spectroscopic properties of the PSB as a result of mutation of Q38.

Entry	hCRBPII Mutant [a,b]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	SS	Φ_{ESPT}	Φ
1	KLVSH:**Q38H**	515	588	-	73	9.8	< 0.01
2	KLVSH:A33H:**Q38Y**	525	591	685	66/160	48.7	< 0.01
3	KLVSH:**Q38E**	528	586	689	58/161	49.8	< 0.01
4	KLVSH:F16Y:**Q38E**	533	595	674	62/141	48.3	0.02
5	KLVSH:A33H:F16Y:**Q38E**	534	596	673	62/139	47.2	0.01
6	KLVSHW:A33H:F16Y:**Q38E**	533	-	677	144	66.4	0.01
7	KLVSHW:A33Y:F16Y:**Q38E**	529	588	678	59/149	51.5	0.01

[a] KLVSH equals to Q108K:K40L:T51V:T53S:R58H. [b] KLVSHW equals to Q108K:K40L:T51V:T53S:R58H:Y19W/

Mutation of Q38 to histidine in Q108K:K40L:T51V:T53S:R58H template significantly reduced Φ_{ESPT} from 50% to less than 10% (**Table II-16**, entry 1). Moreover, Q108K:K40L:T51V:T53S:R58H:A33H:**Q38Y** showed diminished Φ_{ESPT} and fluorescence Φ (**Table II-16**, entry 2). In contrast, **Q38E/ThioPhenol** showed elevated Φ_{ESPT} and, more importantly, the fluorescence Φ (**Table II-16**, entry 3). However, the addition of Q38E to other templates did not improve the fluorescent properties compared to the parent mutants.

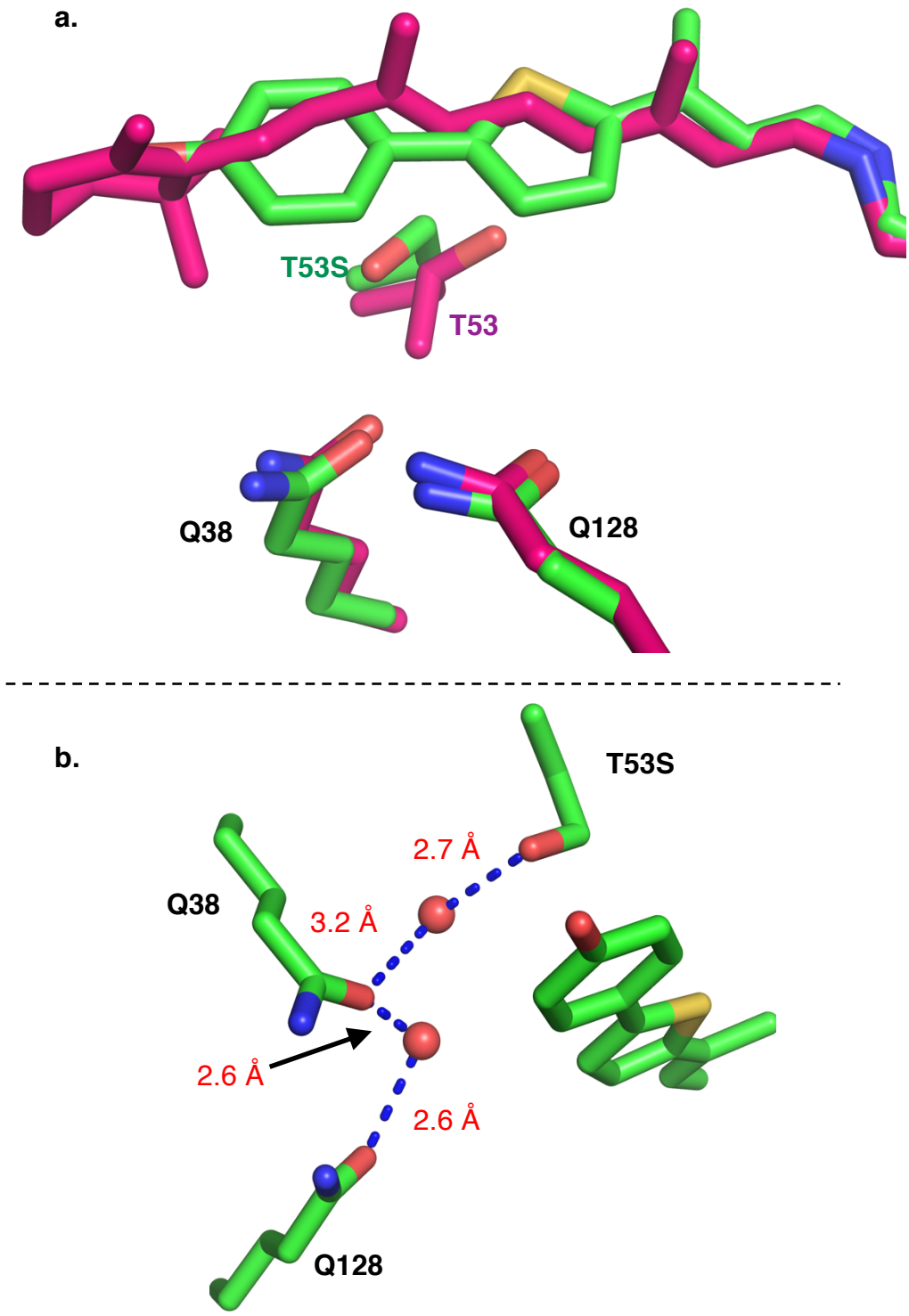


Figure II-20. **a**. Overlay of Q108K:K40L:T51V:T53S:R58W/**ThioPhenol** (green) and Q108K:K40L/retinal (pink). **b**. Water network between T53S, Q38, and Q128 in Q108K:K40L:T51V:T53S:R58W/**ThioPhenol** crystal structure.

Nonetheless, an interesting trend was observed in the absorption and emission of mutants containing Q38E upon acidification. During acidifying these complexes to pH less than 6, a small shoulder appears to the right side of the PSB absorption peak. Notably, the excitation of this shoulder around 610 nm leads to emission maximized at the same wavelength when PSB is excited. However, with lower intensity (**Figure II-21b**), thus, we surmised the shoulder corresponds to **ThioPhenolate**-PSB produced in the ground state (see **Section II.11.3 C** for more data and discussion).

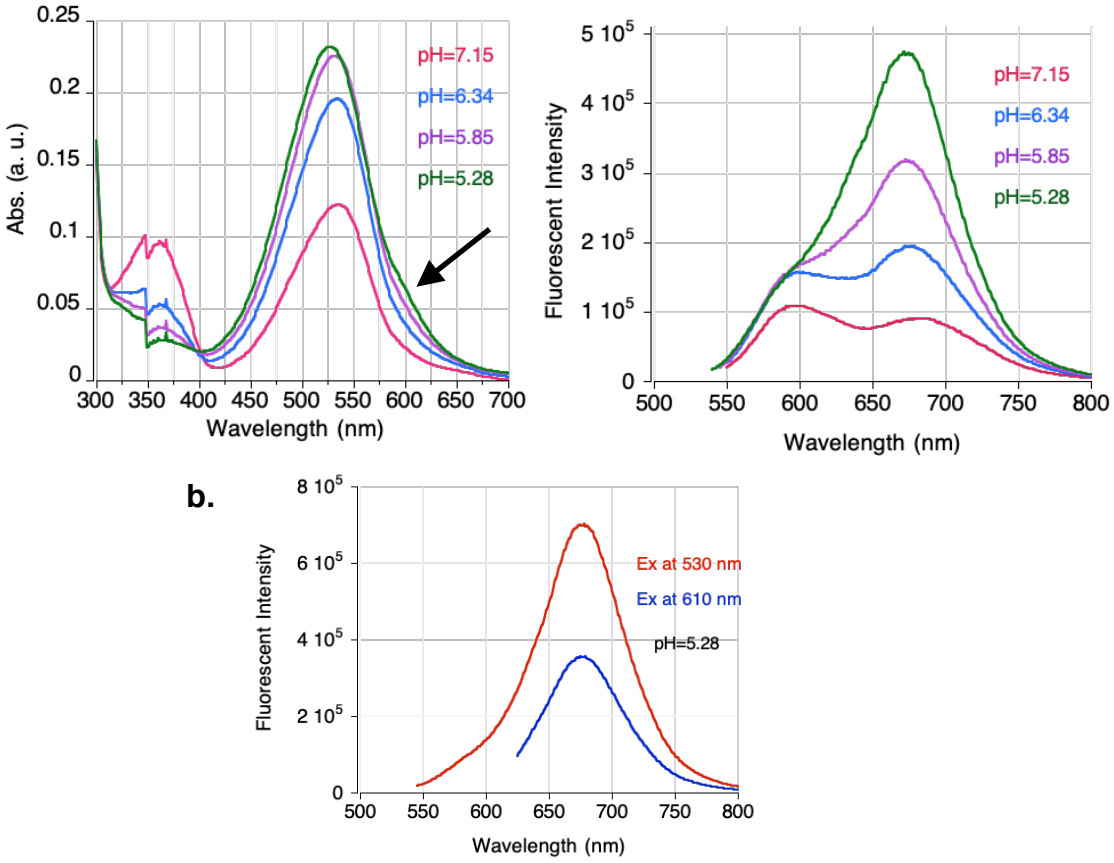


Figure II-21. **a**. UV-Vis (left) and fluorescence (right) spectra of Q108K:K40L:T51V:T53S:R58H:A33H:F16Y:Q38E/**ThioPheno**l upon acid titration. **b**. Excitation of the shoulder indicated by the black arrow results in the same emission as PSB excitation.

Furthermore, as opposed to previous observations, protein variants containing Q38E showed a remarkable increase in the red-shifted emission upon acidification. For example, in Q108K:K40L:T51V:T53S:R58H:A33H:F16Y:**Q38E** mutant, the Φ_{ESPT} increases from 45% in pH 7.3 to more than 90% in pH 5.3 (**Figure II-21a**). As discussed earlier in this section, we hypothesize that the lower pK_a of the Glu side chain next to 58H produces a hydrogen bond network and facilitates the proton transfer from the hydroxyl group in acidic pH. Unfortunately, there are no crystal structures containing R58H and Q38E mutations with **ThioFluor** or **ThioPhenol** to prove this claim.

Among all protein mutants designed to extend the water-mediated hydrogen bonding network around R58H and the hydroxyl group, only the ones that contained Q38E and T53E mutations showed enhanced properties. Hence, we decided to measure the fluorescence characteristics of variants that contained both mutations together (**Table II-17**).

Table II-17. Spectroscopic change as a result of mutation Q38 and T53 mutation to Glu.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	Φ_{ESPT}	Φ
1	KLV:**T53E**:R58H:**Q38E**	527	588	709	49.1	< 0.01
2	KLV:**T53E**:R58H:Y19W:A33H:**Q38E**	526	-	707	97.8	0.01

[a] KLV equals to Q108K:K40L:T51V.

The combination of the two mutations resulted in the most red-shifted emission wavelengths (>705 nm); however, only with the presence of A33H fluorescence properties improved compared to the parent mutant. Therefore, it

cannot be concluded that the addition of Q38E and T53E effects are additive and would always result in brighter complexes.

II.11.2 Enhancing the ESPT emission quantum efficiency through increasing the iminium pK_a

To this end, we could design several **ThioPhenol**/hCRBPII complexes showing more than 99% Φ_{ESPT} through rational point mutagenesis. Yet, it was necessary to increase the fluorescence quantum yield. As discussed previously, increasing the iminium pK_a is another approach for enhancing the QY. Unfortunately, for the protein mutants that yielded the highest Φ_{ESPT}, the iminium pK_a was determined to be in the range of 5.0 to 6.5, which leads to the SB concentration more than twice of the PSB's according to the Henderson-Hasselbalch equation.

An interesting observation was made when comparing the crystal structures of **ThioPhenol** and **ThioFluor** bound to Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F mutant. Both chromophores and the 117D residue adopt the same conformation, but the iminium's configuration is different. **ThioPhenol** forms a *cis* iminium, while **ThioFluor** adopts the *trans* isomer (**Figure II-22a**). As shown, there is a closer distance between L117D and the iminium nitrogen atom in the *trans* isomer, 4.5 Å compared to 5.3 Å, which presumably should lead to the decreased pK_a. Our lab also has formerly observed the trend that for both CRABPII and hCRBPII bound to retinal, *cis* iminiums show a higher pK_a than the *trans* iminiums; [98,99] a similar trend is

observed with **ThioFluor** as well. Nonetheless, the pK_a of the *cis* isomer formed upon binding to **ThioPhenol** is lower by 1 unit. Generally, it was observed that **ThioPhenol** tends to show lower iminium pK_a compared to **ThioFluor** when bound to the same mutant.

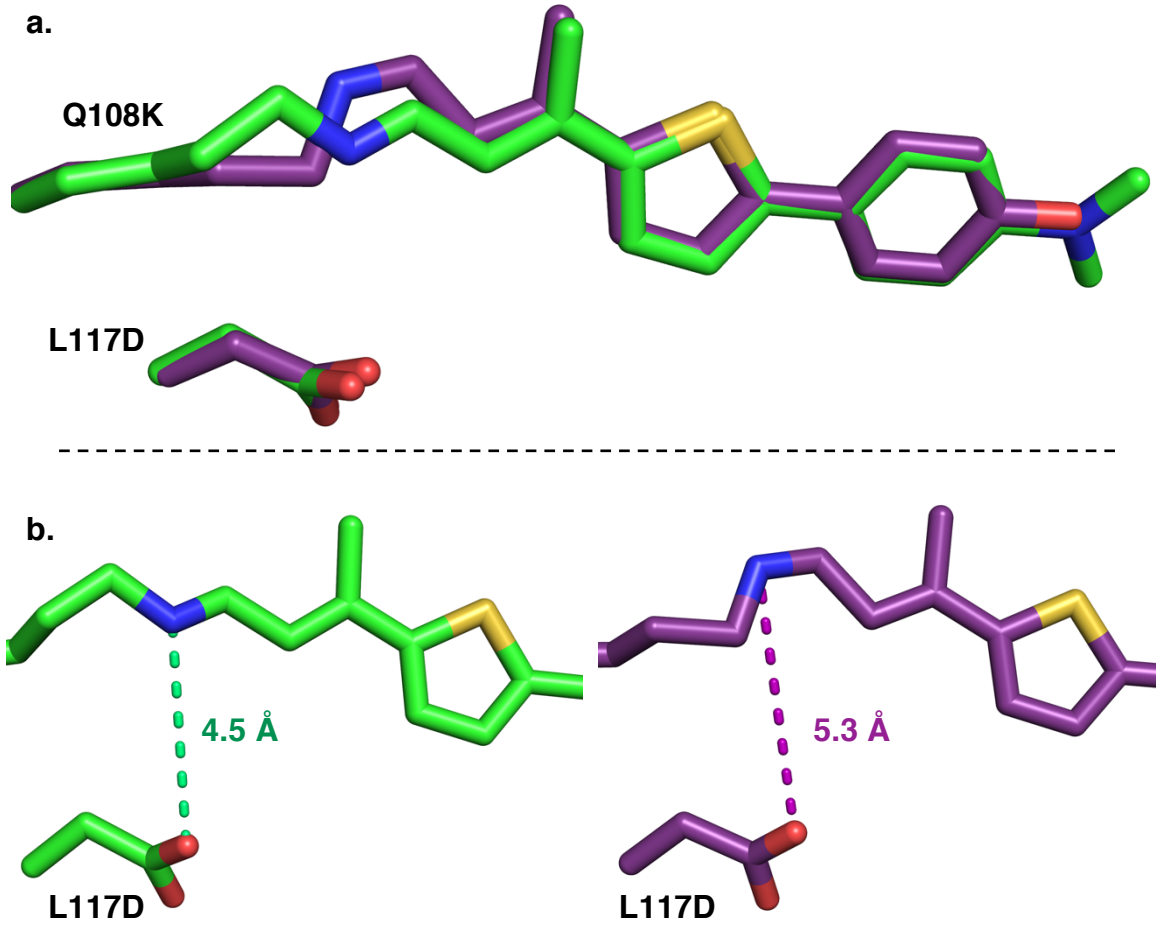


Figure II-22. **a**. An overlay of crystal structures of Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioPhenol** (purple) and **ThioFluor** (green) showing *cis* and *trans* iminium conformation, respectively. **b**. The distance between L117D residue and the iminium nitrogen atom for **ThioFluor** complex (left) and **ThioPhenol** (right).

As shown in **Section II-9**, the introduction of L117E or L117D mutations accelerate the iminium formation process and provides high iminium pK_a values.

However, in this system, the addition of L1117E/D mutations prevent from deprotonation of the hydroxyl group upon excitation (see **Tables II-5** and **II-6**). Hence, we sought to investigate whether mutation of other residues in the vicinity of the iminium to acidic amino acids can provide a high pK_a without impeding the Q_{ESPT}.

As seen in the crystal structure of Q108K:K40L:T51V:T53S:R58H:Y19W:A33Y/**Thiophenol**, Q4, I42, 51V, F64, L93, L115, and L117 are all within 6 Å from the iminium (**Figure II-23**).

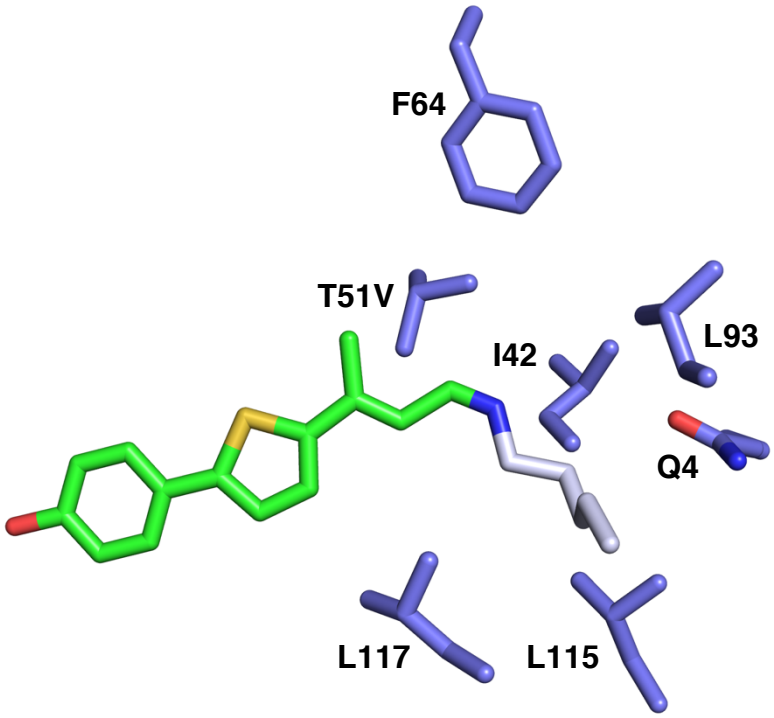


Figure II- 23. Highlighted residues mutated to glutamic acid in order to interact with the iminium to increase its pK_a.

All these residues are located in the hydrophobic protein binding pocket, and it is widely accepted that the hydrophobic interactions can influence protein folding and stability. [100-102] It was more of a challenge to express soluble proteins

when substituting non-charged hydrophobic residues with polar acidic residues. Unfortunately, except for L117E mutation, most of the designed mutants led to insoluble protein expression (**Table II-18**), misfolded and aggregated in the form of insoluble inclusion bodies during protein expression, and so exploration at these positions could not be followed.

Table II-18. List of mutants that resulted in insoluble protein expression upon introduction of Glu resides in the proximity of

Entry	hCRBPII Mutant [a]
1	KLVS:R58H:Y19W:A33W:F16Y:**Q4E**
2	KLVS:R58H:**I42E**
3	KL:**T51E**:T53S:R58H
4	KLVS:R58H:A33W:Y19W:F16Y:**F64E**
5	KLVS:R58H:**L93E**
6	KLVS:R58H:A33W:Y19W:F16Y:**L115E**

[a] KL equals to Q108K:K40L and KLVS equals to Q108K:K40L:T51V:T53S.

II.11.2 A. L117 mutation

We then sought to explore if other residues at Leu 117 can stabilize the PSB. As is shown in **Figure II-22b**, the distance between L117D residue and the iminium nitrogen is not that close to interacts directly, 5.3 Å, and thus it is presumably a water-mediated stabilization. It was envisioned that the introduction of other ionizable amino acids such as Ser, Thr, Gln, Try, His, and Cys residues at position 117 might form such water network and increase the iminium pK_a without

the prevention of the ESPT process. The spectroscopic properties of all expressed mutants are listed in **Table II-19**.

Table II-19. Spectroscopic properties of protein mutants upon mutation of L117 in order to increase the iminium pK_a.

Entry	hCRBPII Mutant [a,b]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	Φ_{ESPT}	Φ	pK_a
1	KLVSH:Q38E	528	586	689	49.8	< 0.01	5.9
2	KLVSH:Q38E:**L117T**	487	564	-	16.4	0.09	7.6
3	KLVSH:Q38E:**L117H**	483	576	-	28.2	0.11	5.6
4	KLVSH:Q38E:**L117C**	513	580	-	31.5	0.07	8.1
5	KLVSH:F16Y	530	596	677	62.1	0.04	7.1
6	KLVSH:F16Y:**L117Y**	532	598	-	8.4	0.02	8.2
7	KLVSH:F16Y:**L117C**	480	581	679	34.2	0.07	8.6
8	KLVSH:A33H:F16Y	533	589	677	49.8	0.02	7.3
9	KLVSH:A33H:F16Y:**L117T**	505	573	-	25.3	0.04	7.8
10	KLVSH:A33H:F16Y:**L117Y**	536	595	-	4.3	0.09	8.6
11	KLVSH:A33H:F16Y:**L117S**	515	562	-	< 0.02	0.05	6.9
12	KLVSH:A33H:F16Y:**L117C**	523	579	678	54.6	0.05	8.8
13	KLVSH:Y19W:A33H:F16Y	530	-	680	97.6	0.01	7.0
14	KLVSH:Y19W:A33H:F16Y:**L117T**	498	568	-	19.8	0.08	7.7
15	KLVSH:Y19W:A33H:F16Y:**L117S**	492	571	-	8.6	0.07	8.1
16	KLVSH:Y19W:A33H:F16Y:**L117E**	511	564	-	5.2	0.06	7.5
17	KLVSH:Y19W:A33H:F16Y:**L117C**	520	577	680	95.5	0.08	8.6
18	KLV:T53E:R58H:Y19W:A33H	531	-	704	97.8	0.01	6.0
19	KLV:T53E:R58H:Y19W:A33H:**L117Q**	470	560	670	37.6	0.02	5.8
20	KLV:T53E:R58H:Y19W:A33H:**L117H**	361	473	-	-	0.03	5.3
21	KLV:T53E:R58H:Y19W:A33H:**L117C**	505	573	698	48.3	0.02	5.5

[a] KLV equals to Q108K:K40L:T51V. KLVSH equals to Q108K:K40L:T51V:T53S:R58H.

We chose to make L117 mutations with five of the protein templates that showed higher Φ_{ESPT}. Substitution of hydrophobic leucine at 117 with ionizable residues such as Thr, His, and Cys led to a blue-shift in the absorption wavelength of 41 nm, 45 nm, and 15 nm from the parent mutant Q108K:K40L:T51V:T53S:R58H:**Q38E**, respectively (**Table II-19**, entries 2, 3 and 4). Presumably, the introduction of polarizable amino acids and the carboxylate from Q38E in the proximity of the iminium localizes the positive charge, leading to the observed blue shift. Proposedly, for the same reasons, L117 mutations to Gln and Cys on the Q108K:K40L:T51V:**T53E**:R58H:Y19W:A33H:L117C template containing T53E resulted in hypsochromic shifts in the absorption wavelength, 61 nm, and 26 nm respectively (entries 19 and 21). Although Q38E and T53E mutations initially increased the Φ_{ESPT}, it was observed that the polarity changes in the protein cavity, as a result of the presence of their carboxylate side chain plus polar residues at position 117 exerts a significant impact on the absorption and emission profile of **ThioPhenol**-PSB and its Φ_{ESPT}.

The introduction of L117T did not change the iminium pK_a, but it reduced the Φ_{ESPT} dramatically compared to the parent mutants (entries 2, 9, and 14). Interestingly, the intensity of the blue emission corresponding to the **ThioPhenol**-PSB complex increased drastically upon L117T addition, while the red-shifted ESPT emission intensity remained the same, leading to decreased Φ_{ESPT} (**Figure II-24**). Unfortunately, the addition of L117H and L117Q severely hampered the chromophore's binding; monitoring the SB/PSB formation via the UV-Vis

absorption peak showed that even after 12 hours of incubation, part of the chromophore remained unbound. Nonetheless, the iminium pK_a of the produced complexes was determined to be lower than the initial mutants; for instance, binding to Q108K:K40L:T51V:T53E:R58H:Y19W:A33H:**L117H** mutant resulted in mere SB formation (entry 20).

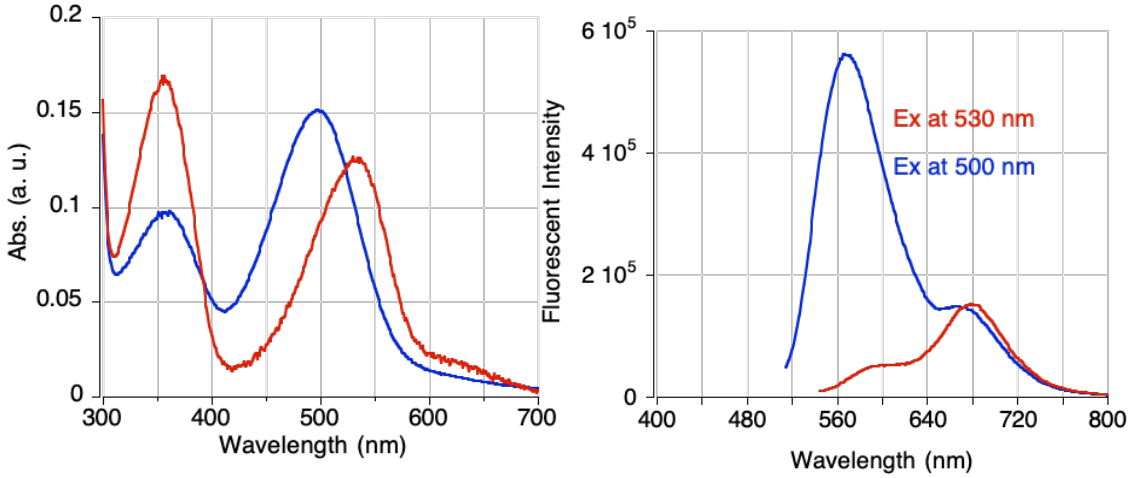


Figure II-24. UV-Vis (left) and fluorescence (right) spectra of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y/**ThioPhenol** (red) and Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117T/**ThioPhenol** (blue).

Inclusion of L117Y and L117S mutations led to relatively high iminium pK_a values; for example, a pK_a of 8.5 was observed with Q108K:K40L:T51V:T53S:R58H:F16Y:**L117Y**, which is higher by about two units as compared to its parent mutant (entry 10). Regardless, both Tyr and Ser at position 117 suppressed the ESPT emission and led to $\Phi_{ESPT} < 0.05$ (entries 6, 10, 11, and 15).

Fortuitously, however, mutation of Leu 117 to cysteine yields the best results; the iminium pK_a increased to more than 8, leading to the PSB as the major component. As a result, the emission intensity enhanced significantly; it led to

almost 2, 4, and 5-fold emission enhancement for entries 7, 12, and 17, respectively. To our delight, while the introduction of L117C provides mutants with higher fluorescence quantum yield, it does not affect the Φ_{ESPT} efficiency. Excellent data was acquired with Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:**L117C** mutant, the iminium pK_a increased to 8.6 with the Φ_{ESPT} as efficient as 95%, and the fluorescence quantum yield elevated to 8% (entry 17). Interestingly, the substitution of glutamic acid with cysteine in the same mutant shows a high pK_a of 7.5; however, as previously reported, the ESPT emission was absent (entry 16). As only the addition of L117C to protein variants contained the Q108K:K40L:T51V:T53S:R58H pentamutant, resulted in high iminium pK_a and enhanced fluorescence properties, this template was retained in most of the further protein engineering.

Great results with L117C piqued our interest to explore the effects of Leu 117 mutation to methionine on the iminium pK_a and the Φ_{ESPT}. The resultant mutant Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:**L117M** showed slightly higher iminium pK_a but lowered fluorescence quantum yield **Table II-20**.

Table II-20. Spectroscopic properties as the result of L117M mutation.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	Φ_{ESPT}	Φ	pK_a
1	KLVSHWH:F16Y:**L117C**	520	577	680	95.5	0.08	8.6
2	KLVSHWH:F16Y:**L117M**	525	583	672	84.3	0.05	8.9

[a]KLVSHWH equals to Q108K:K40L:T51V:T53S:R58H:Y19W:A33H.

Fortunately, Dr. Nona Ehyaei was able to obtain the crystal structure of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C/**ThioPhenol**, in which the 117C shows a distance of 5.4 Å to the iminium nitrogen atom (**Figure II-25a**). In addition, as shown, the *cis* iminium is stabilized through a water-mediated hydrogen bond with the Q4 residue. Previously, Dr. Wenjing Wang and Dr. Elizabeth Santos noticed the same interaction is stabilizing the retinal/**ThioFluor**'s PSB bound to hCRBPII.

Elimination of the water-mediated interaction through mutation of Q4 to Phe results in the chromophore movement toward the mouth of the protein cavity, which results in more exposure to the bulk solvent and non-radiative relaxation pathways. In addition, An overlay of the crystal structures of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C/**ThioPhenol** with Q108K:K40L:T51V:T53S:R58W:Y19W:L117D:Q4F/**ThioPhenol** shows that the distance between 117C/117D and the corresponding iminium nitrogen is almost the same. However, mutation of F16 to Tyr in Pymol for the latter complex shows a further distance to the hydroxyl group, 3.9 Å vs. 4.5 Å, respectively (**Figure II-25b**). These observations help to understand that the improved hydrogen bonding network around the hydroxyl group and stabilizing the iminium positive charge led to the increased Φ_{ESPT} and the fluorescence efficiency with the former complex.

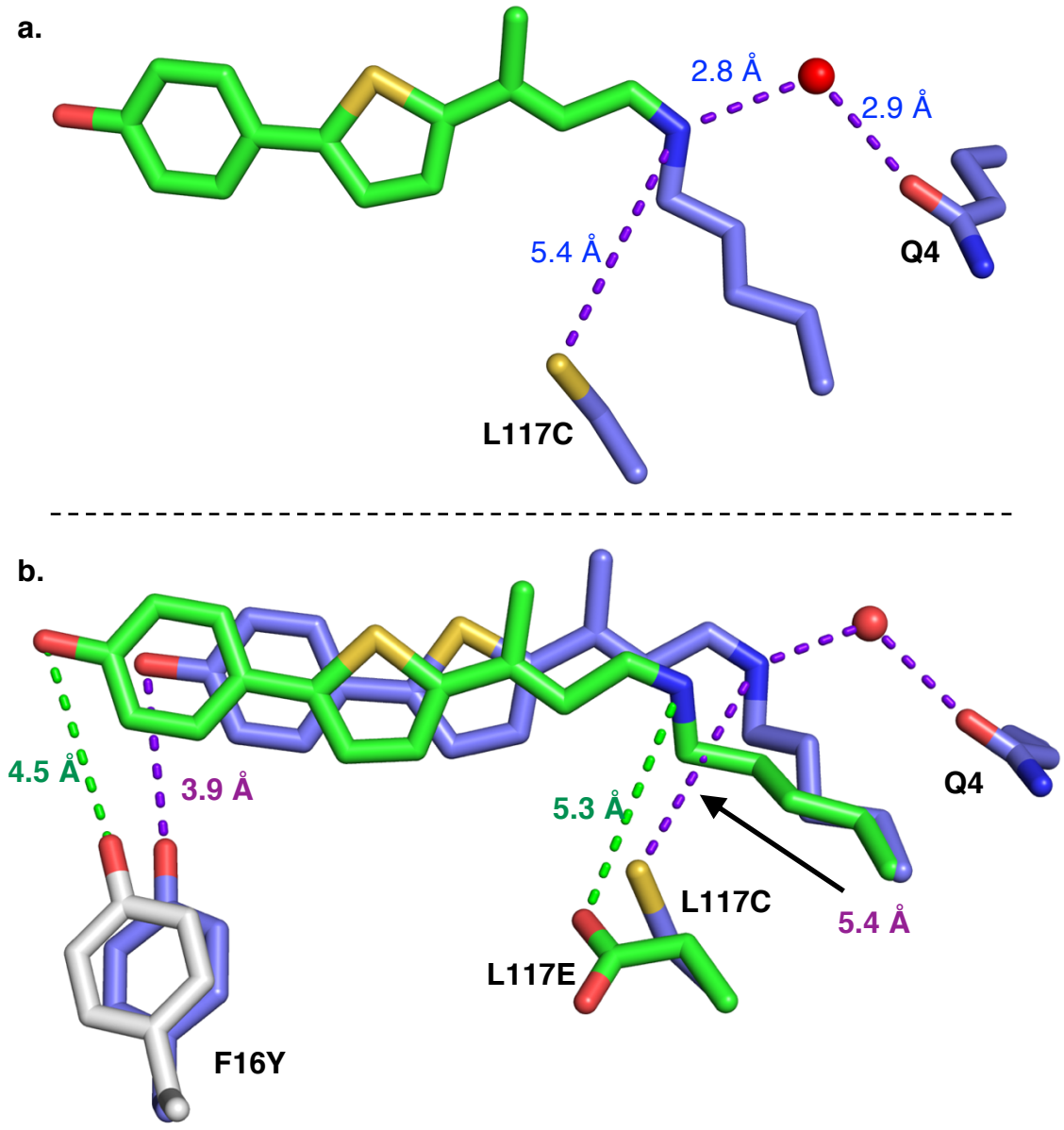


Figure II-25. **a**. Water mediated hydrogen bonding between Q4 and the iminium, and the distance between L117C and the iminium nitrogen atom in the crystal structure of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C/**ThioPhenol** complex. **b**. The different trajectory of chromophore shown upon the overlay of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C/**ThioPhenol** (green) and Q108K:K40L:T51V:T53S:R58W:Y19W:L117D: Q4F/**ThioPhenol** (purple) with F16 residue mutated to Try in the latter (shown in grey).

Notably, although **ThioFluor** forms a *trans* iminium in the presence of L117D/E mutations (**Figure II-22a**); analysis of all crystal the structures acquired from hCRBPII/**ThioPhenol** complexes shows a *cis* minimum with all mutants.

II.11.2 B. Q4 and T51 mutation

Next, we sought to explore whether mutation of Gln 4 to ionizable amino acids can strengthen the water-mediated interaction with PSB and increase the iminium pK_a. Most of the designed mutants led to insoluble protein expression or proteins that precipitated during binding with **ThioPhenol**. However, all expressed mutants showed low iminium pK_a (~5), with no PSB apparent at neutral pH. The photophysical properties of the SBs were measured and are listed in **Table II-21**.

Table II-21. Spectroscopic properties of the SB as the result of Q4 mutation.

Entry	hCRBPII Mutant[a]	λ_{abs}	λ_{em}	pK_a
1	KLVSH:Y19W:A33H:F16Y:**Q4T**	360	477	4.9
2	KLVSH:A33H:F16Y:**Q4S**	365	467/528	5.3
4	KLV:T53E:R58H:Y19W:A33H:**Q4Y**	361	458	-

[a]KLV equals to Q108K:K40L:T51V, and KLVSH equals to Q108K:K40L:T51V:T53S:R58H.

In **Section II-11**, we discussed the effects of T51V mutation on the absorption and emission wavelength of the corresponding complexes and its ability in monomerizing hCRBPII variants. As shown in **Figure II-23**, T51V is one of the closest residues to the iminium nitrogen atom, 4.7 Å, and prone to impact the pK_a.

Previously Dr. Elizabeth Santos illustrated that acidic residues at position T51 close to the putative aldehyde binding site facilitate the PSB formation

considerably, presumably through activation of the retinal aldehyde group. Additionally, the introduction of polar residues in close proximity to the iminium localizes the charge, leading to blue shift in wavelength.

Nonetheless, we decided to explore the effect of introducing the acidic and polarizable amino acids such as cysteine and methionine on the iminium pK_a in this system. The mutation was done on Q108K:K40L:T51V:T53S:R58H:F16Y:Y19W:A33H:L117C template that showed the highest Φ_{ESPT} (**Table II-22**). Both the iminium pK_a and Φ_{ESPT} were enhanced upon introduction of the T51M mutation (entry 2). Mutation of T51 to methionine via mutagenesis in Pymol suggests almost the same distance as T51V (4.4Å); however, the polarizability of the sulfur atom in the methionine side chain can stabilize the iminium and explain the higher iminium pK_a.

Table II-22. Spectroscopic change as the result of the introduction of T51M mutation.

Entry	hCRBPII Mutant	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	Φ_{ESPT}	pK_a
1	T51V	520	577	680	95.5	8.6
2	T51M	523	-	674	>99	9.0

II.11.2 C. K40 Mutation

As our last attempt to increase the pK_a, we chose to manipulate the environment around the iminium through K40 mutation. K40L has been retained in all protein mutants studied in this chapter. In studies with hCRBPII/retinal, Dr. Wenjing Wang previously demonstrated that K40L increases the pK_a of retinal-PSB, 8.5 for Q108K:K40L compared to <6.5 for Q108K. Additionally, it was shown

that Lys 40 was disturbing the stability of the apo protein upon introduction of Lys108, due to charge repulsion of the two residues in close proximity.

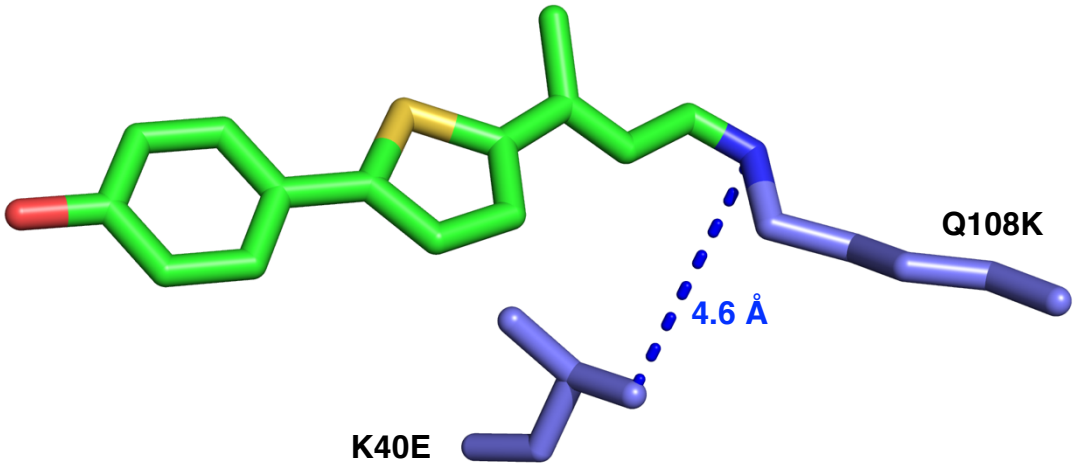


Figure II-26. The distance between the K40L residue and the iminium nitrogen atom in the crystal structure of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C/**ThioPhenol**.

K40 residue shows a close distance of 4.6 Å with the iminium nitrogen (**Figure II-26**), and as discussed above, mutation of K40 to leucine increased the retinal-PSB pK_a remarkably. Therefore, we sought to investigate whether ionizable residues at position 40 are capable of stabilizing the iminium without reducing the Φ_{ESPT}. For the sake of comparison, K40 mutations were done on the protein template yielded the best results, Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C.

Several neutral polar residues were tested since the positive charge on the basic residues' side chain can destabilize the PSB, as discussed above. As the protein yield was low for some of the mutations, the Q4F mutation was included since it has been shown to lead to higher protein expression yields without

negatively affecting the complex's quantum yield (denoted by a star in **Table II-23**). The photophysical and iminium pK_a of the expressed mutants are listed in (**Table II-23**).

Table II-23. Spectroscopic properties as the result of K40 mutation.

Entry	Residue at 40[a]	λ_{abs}	λ_{em}	Φ_{ESPT}	pK_a	Dipole moment (D)
1	K40L	520	577	95.5	8.6	3.78
2	K40T	495	572	22.6	6.7	9.30
3	K40S*	493	574	28.6	5.9	9.84
4	K40Y	505	575	10.5	7.9	10.41
5	K40C*	500	570	9.3	6.4	10.74
6	K40N*	370	465	< 0.02	5.0	18.89
7	K40D	460	550	4.3	5.3	29.49
8	K40Q	480	564	< 0.02	>5	39.89
9	K40H*	490	576	19.4	5.7	20.44

[a] The template for starred mutants is Q108K:K40L:T51V:T53S:R58H:F16Y:Y19W:A33H:L117C:Q4F, and for others is Q108K:K40L:T51V:T53S:R58H:F16Y:Y19W:A33H:L117C.

All mutations showed hypsochromic shifts in the absorption wavelength relative to the K40L template due to the localization of the positive charge on the iminium nitrogen atom. For instance, K40D mutation resulted in the most blue-shifted PSB absorptions wavelength; negative charge of the aspartic acid side chain exerts an electrostatic interaction on the PSB that leads to iminium stabilization and the observed blueshift (entry 7). Protein variant with K40Q mutation had the lowest iminium pK_a resulting in mere SB formation (entry 8). Interestingly it seems there is a correlation between the residue's dipole moment

value and the complex absorption wavelength; in general, less polar residues at position 40 show more red-shifted absorption wavelengths due to the same results discussed above. Nonetheless, the Φ_{ESPT} of the designed mutants was diminished drastically (<0.30) compared to the parent mutant.

There have been many attempts to analyze and optimize **ThioPhenol**'s iminium pK_a following several different approaches. Our studies show that the conventional mutations proved to increase the pK_a for retinal and **ThioFluor** such as L117D/E have a detrimental effect on the Φ_{ESPT} of the current system. The best results were obtained upon the combination of Q4, K40L and L117C mutations, which give rise to the maximum pK_a of 9.6.

II.11.3 Through expanding water-mediated network around the hydroxyl group

In the early stages of this study, several spots (M20, T29, A33, S76, and L77) were recognized as the candidates for point mutagenesis to facilitate the deprotonation of the hydroxyl group upon excitation (**Figure II-12**). However, later we realized most of those mutations were done on protein templates not optimized to yield high Φ_{ESPT} due to the presence of L117D or L117E. Unfortunately, mutation of M20, S76 primarily resulted in very low protein expressions. Hence, we chose to mutate T29 and L77 to proton acceptor residues in the newly designed templates.

II.11.3 A. L77 mutation

In **Section II-9-1**, it was shown that the L77 residue is close to the hydroxyl group (4Å, see **Figure II-12**); however, attempts to enhance the Φ_{ESPT} through its mutation to basic residues were not fruitful, as later we realized L117E or L117D mutations inhibit the process of proton transfer from the hydroxyl group to its neighboring residues upon excitation. Now that we could successfully design protein mutants showing more than 99% Φ_{ESPT} through rational point mutagenesis, we south to investigate whether mutation of L77 to proton acceptor residue will improve the fluoresce quantum efficiency of the corresponding complexes. None of the protein variants contained L77C, and L77M led to soluble protein expressions; L77H mutation also resulted in low protein yield. Photophysical properties of the expressed mutants are listed in (**Table II-24**).

Table II-24. Spectroscopic change as the result of L77 mutations.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	Φ_{ESPT}	Φ
1	KLVS:R58H:A33H	530	586	693	54.2	0.01
2	KLVS:R58H:A33H:**L77H**	536	584	693	53.2	0.02
3	KLVS:R58H:A33H:**L77Y**	520	584	692	64.8	0.05
4	KLVS:R58H:F16Y:**L77Y**	512	587	673	88.7	0.02

[a] KLVS equals to Q108K:K40L:T51V:T53S,

The addition of L77H mutation resulted in almost the same properties (compare entry 1 with entry 2). On the other hand, mutation to tyrosine enhanced both the Φ_{ESPT} and fluorescence quantum efficiency. L77Y, in combination with F16Y, improved the Φ_{ESPT} to more than 85%, although the inclusion of F16Y led to

the 20 nm blue-shift in the ESPT emission wavelength (entry 4). The highest Φ was observed for Q108K:K40L:T51V:T53S:R58H:A33H:L77Y mutant (entry 3).

Comparing the properties of entry 3 and 4 shows that inclusion of A33H improves the fluorescence efficiency, while F16Y enhances the Φ_{ESPT}. Therefore, we sought to investigate the effect of L77Y when both mutations are present on the template showed the highest pK_a Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C (**Table II-25**). The resultant mutant was less stable, evident from the absorption spectrum of the complex, and interestingly the iminium pK_a dropped to 6.0 with lower Φ_{ESPT} (entry 2).

Table II-25. Spectroscopic and pK_a change as the result of L77Y introduction.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	Φ_{ESPT}	pK_a
1	KLVSHWH:F16Y:L117C	520	577	680	95.5	8.6
2	KLVSHWH:F16Y:L117C:**L77Y**	502	571	670	41.5	6.0

[a] KLVSHWH equals to Q108K:K40L:T51V:T53S:R58H:Y19W:A33H.

L77 is located on the loop at the entrance of the protein's binding cavity (**Figure 11-27a**), and as previously Dr. Wenjing Wang showed for retinal/hCRBPII complexes, switching from a hydrophobic residue to more polar residues such as histidine, tyrosine, and cysteine could move the conformation of the loop so that it can flip out of the binding pocket to get more solvated in the aqueous solution. Such changes can affect the protein expression yield or its stability. For instance, in this study, none of the L77 mutations to serine or cysteine led to soluble proteins,

and for the expressed mutants, the formation of the aggregates was quite apparent from the UV-Vis spectrum of the corresponding complexes. Hence mutation of L77 was not continued in further studies.

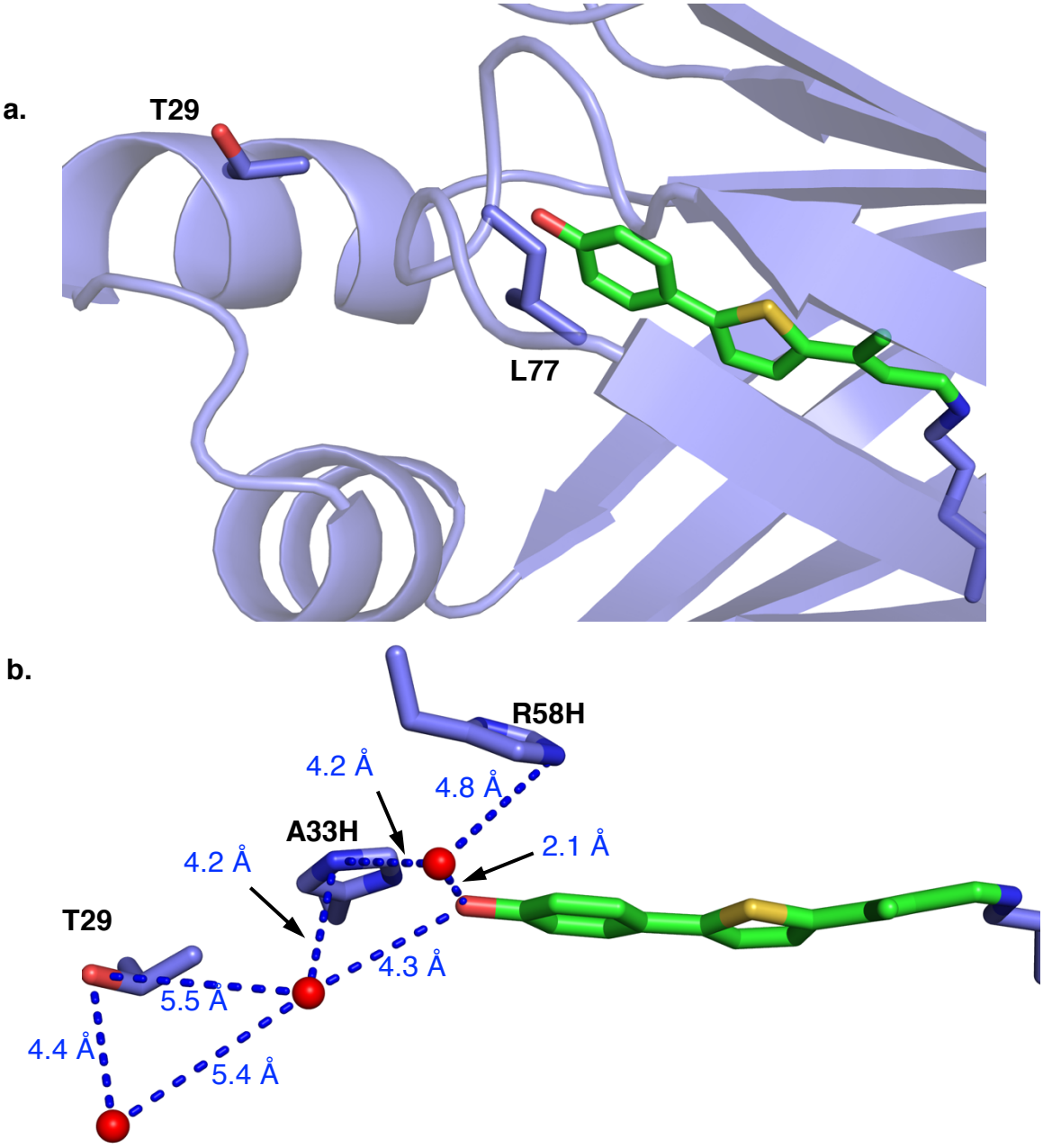


Figure II-27. **a**. Cartoon representation of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C/**Thiophenol** complex with T29 and L77 residues highlighted. **b**. Detailed hydrogen bonding network surrounding T29, A33H, R58H and the hydroxyl group in the same crystal.

II.11.3 B. T29 mutation

T29 resides on the α-helices segment with its side chain pointing outside of the binding pocket to get better solvation of the polar hydroxyl group leading to a long distance (9.1Å) between threonine and the chromophore's hydroxyl groups (**Figure II-27**). Hence, it is not expected for residues at position 29 to significantly affect the chromophore's photoacidity. However, the crystal structure of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:L117C/**ThioPhenol** complex reveals that there is a conserved water-mediated hydrogen bonding network including T29, A33H, the chromophore's hydroxy group, and two organized water molecules (**Figure II-27b**). The well-developed network around the hydroxyl group plus the close-distanced R58H explains the high ESPT quantum efficiency with this mutant. We sought to examine whether mutation of T29 to tyrosine would influence the water network and/or the Φ_{ESPT} (**Table II-26**).

Table II-26. Spectroscopic change as the result introduction of T29Y mutation.

Entry	hCRBPII Mutant [a,b]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	Φ_{ESPT}	Φ	pK_a
1	KLVSHWHY:L117C	520	577	680	95.5	0.08	8.6
2	KLVSHWHY:L117C:**T29Y**	527	-	681	>99	0.13	8.9

[a] KLVSHWHY equals to Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y.

Interestingly, the introduction of T29Y led to an increase in both the ESPT and the red-shifted fluorescence efficiency as the **ThioPhenol**-PSB emission is almost completely suppressed for the corresponding complex (entry 2). Unfortunately, there is no crystal structure for this mutant to demonstrate the

orientation of T29Y, but presumably strong water-mediated hydrogen bonding with the chromophore is the reason for the enhanced fluorescence properties. T29Y mutation was then retained in the protein template for further engineering.

II.11.3 C. T53 mutation

Formerly, Dr. Wenjing Wang reported that T53C mutation could lead to a significant bathochromic shift in the absorption wavelength of retinal/hCRBPII complex due to the polarizability of the cysteine residue. Earlier in **Section II.11.1**, we described the effect of T53E on increasing the Φ_{ESPT} and bathochromic shifts in the emission of **ThioPhenolate**-PSB; however, this mutation was eliminated for further studies as the addition of L117C could not increase the iminium pK_a and therefore mutants that contained T53E showed lower fluorescence quantum yield. We surmised that the polarizability of the T53C mutation could stabilize the excited state better than serine.

Table II-27. Spectroscopic properties as the result of introduction of cysteine in positions T51 and T53.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Blue)	λ_{em} (Red)	Φ_{ESPT}	Φ	pK_a
1	T51V:T53S (**M1**)	527	-	681	>99	0.13	8.9
2	**T51C**:T53S (**M2**)	524	-	672	>99	0.11	9.4
3	T51V:**T53C** (**M3**)	517	-	680	>99	0.15	9.8
4	**T51M**:T53S (**M4**)	523	-	674	>99	0.12	8.8

[a] The mutations are introduced to Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53S:R58H:L117C mutant.

In addition, in **Section II.11.2,** we showed that T51M mutation resulted in the PSB stabilization and increased pK_a due to the same reasons. Thus, cysteine was introduced at positions 51 and 53 of the template (Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53S:R58H:L117C), results in in the highest Φ_{ESPT} observed for this series (**Table II-27**).

To our delight, as in the parent protein (entry 1), the blue emission of **ThioPhenol**-PSB is completely suppressed in all of the resultant mutants. The highest Φ_{ESPT} (>99.5%) was acquired upon the introduction of cysteine either at position 51 or 53. Although with T51C, the emission is slightly blue shifted.

An interesting observation was made with T51C, as shown in **Figure II-28a**; there is a shoulder to the right side of the PSB absorption corresponding to the deprotonated hydroxy group of the chromophore, **ThioPhenolat**e-PSB complex. As excitation of this small shoulder at neutral pH results in emission maximum at 673 nm, same as the ESPT product emission (**Figure II-28b**).

The complex absorption was monitored via UV-Vis spectra upon both acid and base titration to ensure this claim is valid. Acidification of the solution leads to the disappearance of the shoulder (**Figure II-28c**). Upon basification of the sample, the intensity of the PSB absorption peak decreases while the shoulder grows, redshifting to 615 nm corresponding to **ThioPhenolate**-PSB. Upon basification to pH above 10, almost all of the PSB is deprotonated, and the absorption maximum around 400 nm corresponding to **ThioPhenolate**-SB appears (**Figure II-28d**).

These observations suggest that the shoulder is the deprotonated hydroxyl group in the ground state.

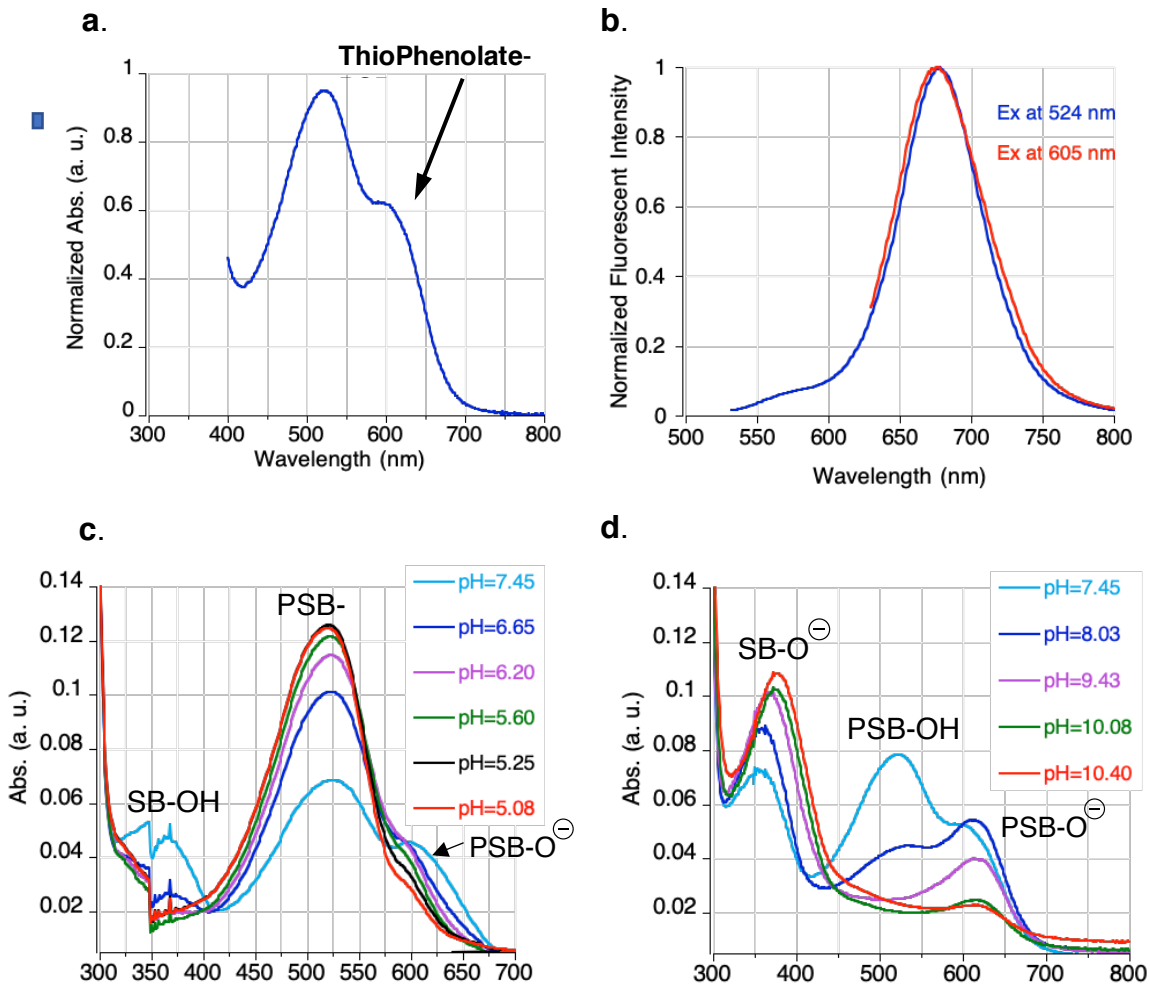


Figure II-28. **a** and **b**. The absorption and emission spectra of Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51C:T53S:R58H:L117C/**ThioPhenol** complex. The shoulder corresponds to PSB-Phenol complex since excitation of both results to the same emission. **c**. The acid titration and **d**. the base titration of the same complex.

Although the hydroxyl group's pK_a value is quite high, in the range of 9.2 to 10.5 depending on the protein mutant, it can partially get deprotonated at neutral pH in the ground state upon binding to protein mutants that have a well-developed water-mediated hydrogen bonding network. Nonetheless, the formation of **ThioPhenolate**-PSB in the ground state is not desired in this study as we are focused on developing far-red/NIR-emitting tags through the ESPT of a photoacid.

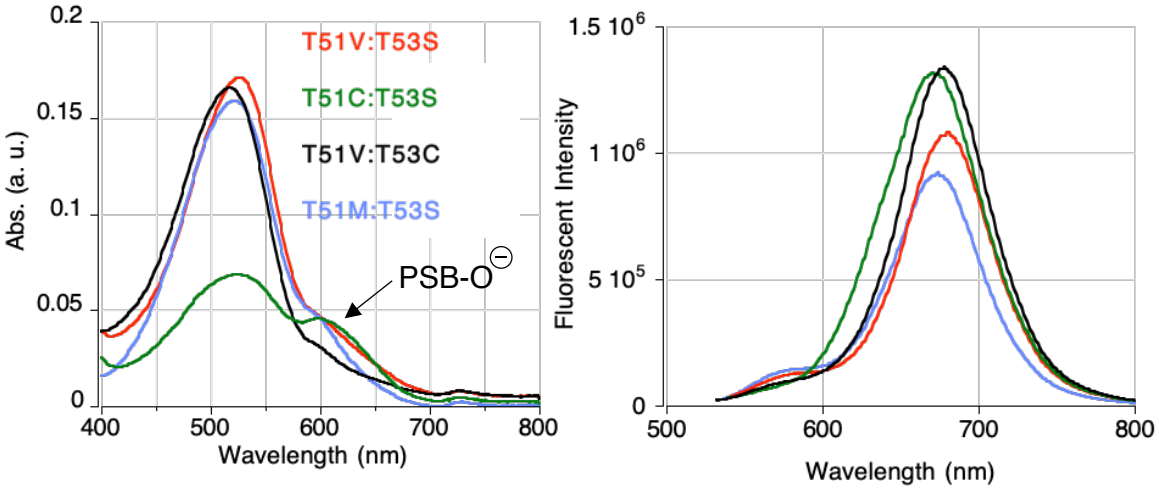


Figure II-29. Comparison of absorption (left) and emission (right) of **M1** (T51VT:53S), **M2** (T51C:T53S), **M3** (T51V:T53C), and **M4** (T51M:T53S) protein/**ThioPhenol** complexes.

As listed in **Table II-27**, all four protein mutants exhibit the Φ_{ESPT} (>99%) and high iminium pK_a values (8.8-9.8). However, the highest fluorescence quantum efficiency was acquired with **M3/ThioPhenol** complex (entry 3). In addition, comparing the absorption and emission spectra of these mutants shows that the shoulder corresponding to the unwanted **ThioPhenolate**-PSB complex is

the smallest with **M3** (**Figure II-29**). Thus, this mutant was chosen for live-cell imaging experiments.

II-12 ThioPhenol/hCRBPII binding kinetics

To investigate the proper staining duration, binding kinetics were measured prior to the confocal imaging experiments. 20 μM of the protein in PBS buffer at neutral pH (7.2-7.4) were incubated with 0.5 equivalent of the ligand at 23 °C and increase in absorbance of the corresponding PSB at its λ_{max} was recorded. Collected data points were fit with a second-order rate equation considering multiple reagents protein and the ligand with non-equal concentrations (**Figure II-30**, see section IV.3.4 for detailed fitting description).

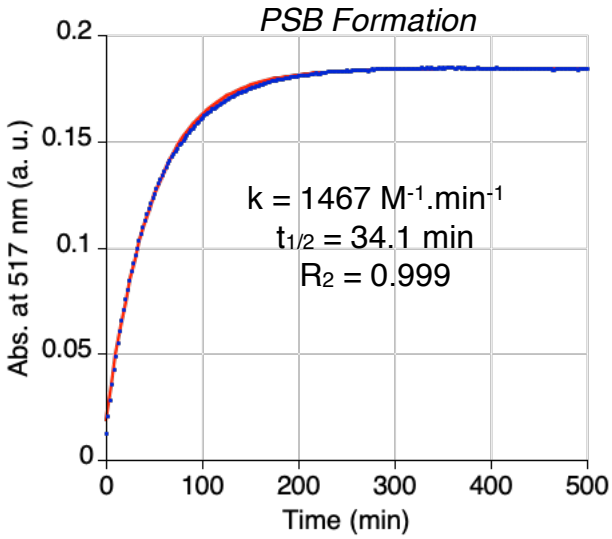


Figure II-30. Rate of **ThioPhenol**/**M3** PSB formation, fitted to 2nd order kinetics with 20 μM protein and 10 μM **Thiophenol**. Plotted is the concentration of free chromophore vs. time.

The next section of this chapter will demonstrate this probe's usefulness in live-cell imaging; most importantly, **ThioPhenol** cell permeability and its efficient

target binding are shown. As was predicted, acquired confocal images display no fluorescent background even with overnight incubation with the ligand and without any washing steps prior to imaging.

II.13 Visualization of hCRBPII/ThioPhenol in mammalian cells

Next, the performance of engineered hCRBPII/**ThioPhenol** complexes as a no-wash live-cell imaging system was investigated. The mutant Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53C:R58H:L117C (**M3**) was chosen for these studies. The iminium pK_a of this mutant is high (9.85), resulting in predominant PSB formation, the Φ_{ESPT} (>99%) and fluorescence quantum yield (0.15) is the highest among all mutants with, and moreover, the concentration of **ThioPhenolate**-PSB in the ground state was negligible with **M3** (**Figure II-31**).

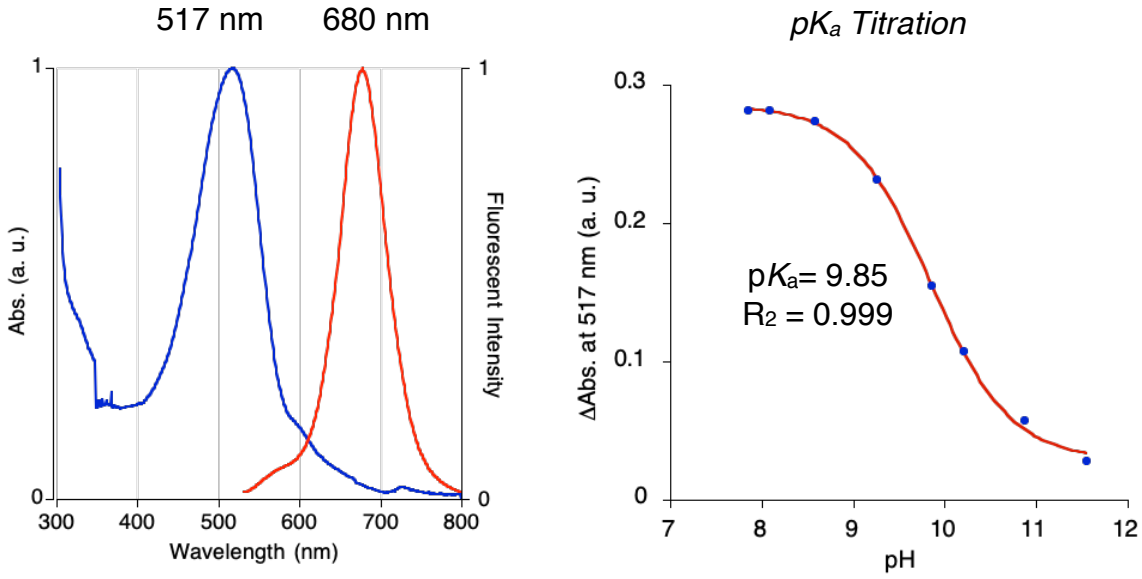


Figure II-31. Spectroscopic properties of Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53C:R58H:L117C mutant with **ThioPhenol** including UV-Vis and fluorescence spectra (left) and pK_a titration (right).

Additionally, binding was relatively fast; the second-order half-life for iminium formation is 34 minutes at 23 °C, and it is solely expressed in the monomeric form (**Figure II-30**). All spectroscopic properties of **ThioPhenol/M3** are summarized in (**Table II-28**).

Table II-28. Spectroscopic properties of **ThioPhenol/M3** complex.

Complex	λ_{abs}	λ_{em}	ε ($M^{-1}.cm^{-1}$)	Φ_{ESPT}	Φ [a]	pK_a	k ($M^{-1}.min^{-1}$)	$t_{1/2}$ [b] (min)
ThioPhenol/M3	517	679	33,743	>99	0.15	9.8	1467	34

[a] Absolute quantum yield was measured on a Quantaurus-QY. [b] Half-life based on the rate constant obtained from second order rate fitting; measured at 23 °C with 20 μM protein and 0.5 equiv **ThioPhenol** at pH 7.2.

II.13.1 ThioPhenol/M3 successfully labels hCRBPII in HeLa cells

To test the performance of **ThioPhenol** as a no-wash tag for *in vivo* imaging experiments, **M3**-hCRBPII was cloned into the pFlag-CMV2 vector containing EGFP fused on the N-terminus of hCRBPII to label the whole cell.

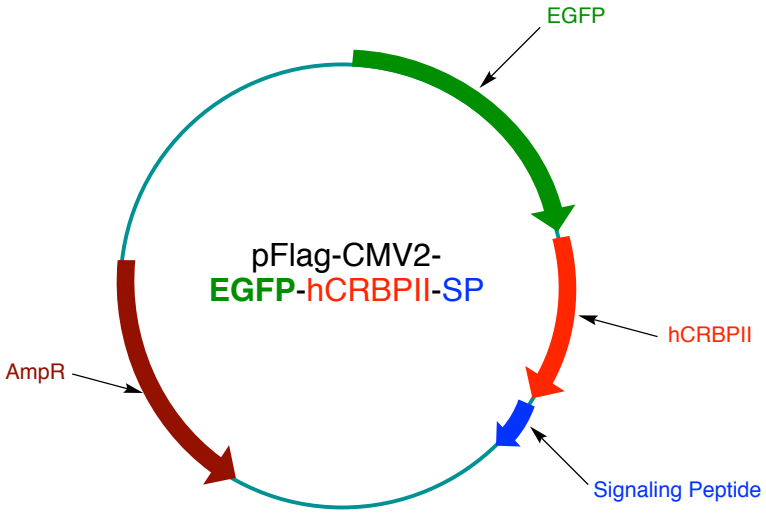


Figure II-32. Maps of the EGFP-hCRBPII-SP fusion constructs. SP: signaling peptides. SP = 3×NLS (nuclear localization sequence), NES (nuclear export sequence), and CAAX (prenylation tag).

Additionally, to target **M3** at cell nuclei, cytosol, or plasma membranes, the signaling peptides NLS (nuclear localization sequence), NES (nuclear export sequence), and CAAX (prenylation tag) were fused to the C-terminus of **M3**, respectively as illustrated in **Figure II-32**.

The fused constructs were transfected and expressed in HeLa cancer cell lines to test the applicability of **ThioPhenol**. Imaging was performed by incubating HeLa cells with 10 μM **ThioPhenol** for one and half hours at 37 °C. The cells were then directly subjected to the confocal imaging without any washing steps prior to imaging.

In all cases, the green fluorescence from EGFP was observed when excited at 488 nm, indicating that transfection was successful, and the fusion protein has been expressed (**Figures II-33** and **II-34**, the green channel). To collect the NIR emission, stained cells were excited at 514 nm. In every triple-fused construct, the NIR emission 620 nm-720 nm window) demonstrated the same pixel specificity as that of EGFP, confirming that the NIR fluorescence is purely emitted from the activated **ThioPhenol/M3** complex without any signal contamination from non-specific labeling that results in identical images in the green and red channel (**Figures II-33** and **II-34**, the red channel). Apparently, no fluorescence background is observed in the red channel, indicating that **Thiophenol** does not label off-target lysines or non-specific bindings does not lead to fluorescence signal. These results prove the utility of the no-wash labeling of

Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53C:R58H:L117C with **ThioPhenol** in mammalian cells.

Furthermore, as discussed above, size exclusion chromatography results confirmed that **M3** is expressed solely in monomeric form (up to 19 mg/L). This is advantageous, assuming the monomeric state dominates in cellular environments as well.

To ensure that labeling with **ThioPhenol** does not develop fluorescent background over long incubation times, HeLa cell lines were incubated with 10 μM **ThioPhenol** at 37 °C for about 12-15 h and were imaged without washing steps. Notably, no non-specific fluorescent background was observed even after overnight incubation time, proving the selectivity of fluorescent signal of **ThioPhenol/M3**-hCRBPII tag.

II.14 Conclusion and future research directions

As stated at the beginning of this chapter, developing fluorogenic background-free tags is one of the most important goals of our research. This chapter describes how this goal is achieved by coupling a non-fluorescent chromophore with various hCRBPII mutants. In this designed system, the protein is engineered to maintain a high iminium pK_a binding to the non-emissive chromophore as a PSB. Further modification of the protein host (hCRBPII) via point mutagenesis makes it possible to activate the complex to function as a photoacid and generate a NIR fluorescence signal. The strong ICT system formed between the alkoxide and the iminium in the **ThioPhenolate**-PSB complex, upon

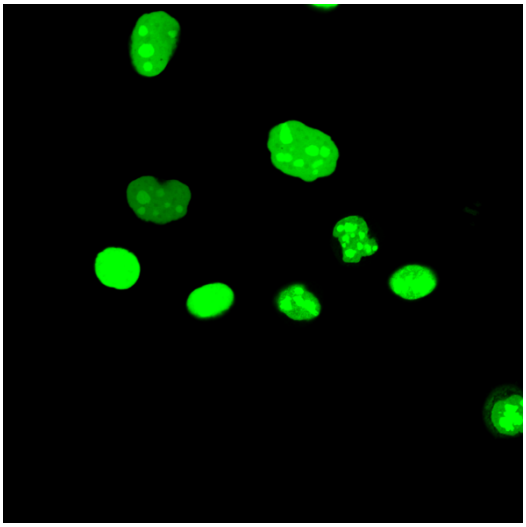
photoirradiation results in a bathochromically-shifted emission. No fluorescent background signal was observed as a result of non-specific imine or iminium formation in the cellular milieu. This is due to the high pK_a of the phenol moiety (9.5-10), which cannot be deprotonated in non-target hosts. Also, the NIR emission of the **ThioPhenolate**-PSB is well separated from unbound chromophore or non-specific bindings emission. The fluorogenic characteristics of this system make it useful in no-wash background-free NIR imaging applications, as demonstrated by the **ThioPhenol/M3** complex, which successfully labeled HeLa cell lines without any fluorescent background.

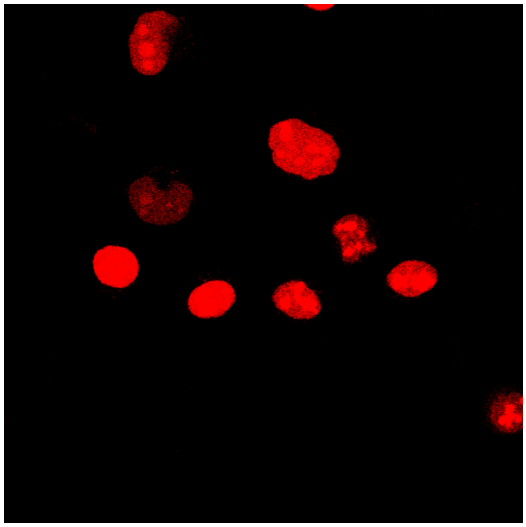
Our next goal is to modify the **ThioPhenol** structure to increase the quantum yield and, ultimately, the brightness of the tags. **ThioPhenol** structure optimization and rigidifying its structure to reduce the non-radiative relaxations of the excited state are discussed in Chapter III. In addition, extending the application of this system in two-color or multi-color imaging is followed and described in the next chapter.

Nucleus localization

EGFP-*3NLS*
Ex: 488 nm
Em: BP 500 - 550 nm

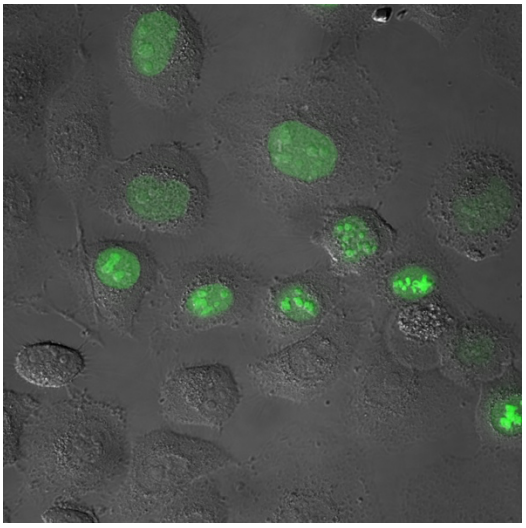
ThioPhenol/hCRBPII-*3NLS*
Ex: 515 nm
Em: BP 620 - 720 nm





EGFP-*3NLS* + DIC

ThioPhenol/hCRBPII-*3NLS* + DIC



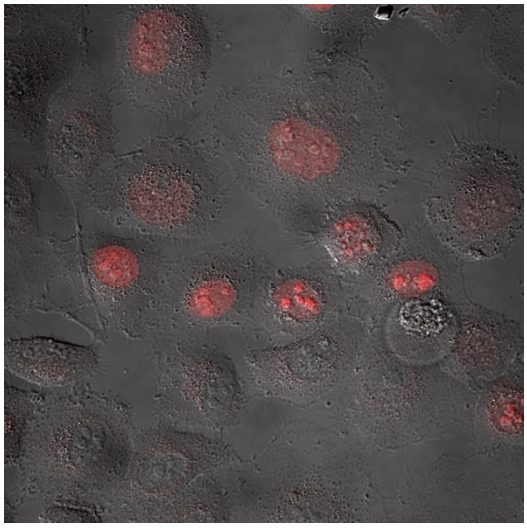
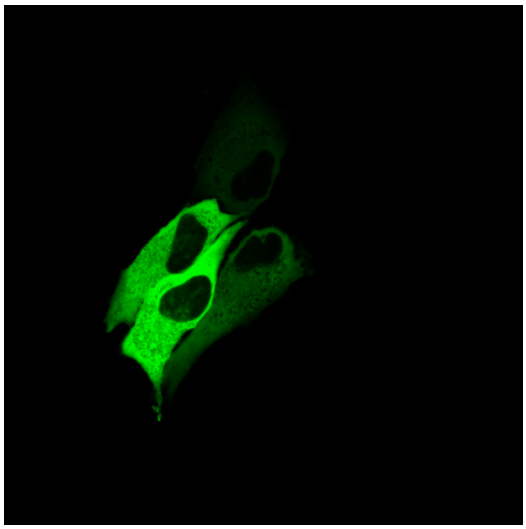


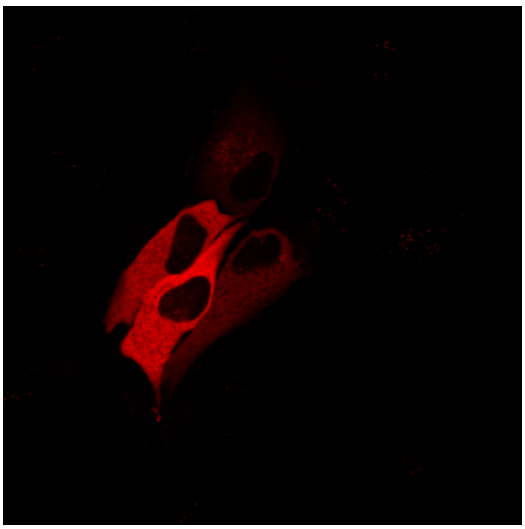
Figure II-33. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-*3NLS*. NLS = nuclear localization sequence. Cells were stained with 10 μM **ThioPhenol** and incubated at 37 °C for 1 h and 30 min. Cells were not washed before imaging.

Nucleolus export localization

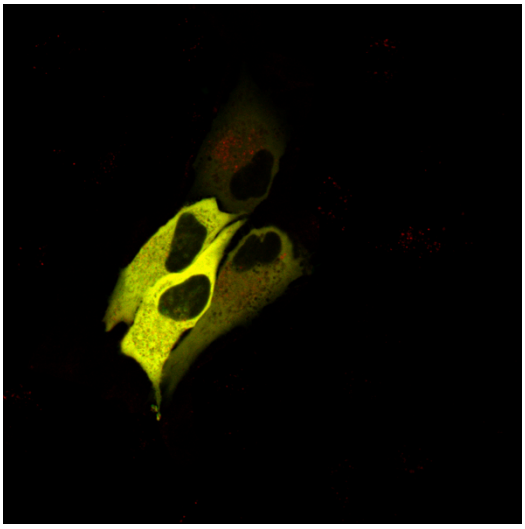
EGFP-*NES*
Ex: 488 nm
Em: BP 500 - 550 nm



ThioPhenol/hCRBPII-*3NES*
Ex: 515 nm
Em: BP 620 - 720 nm



Green + Red Channel



DIC

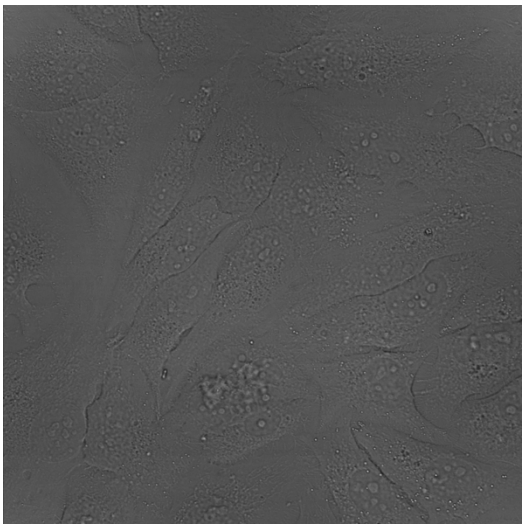


Figure II-34. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-*NES*. NES = nuclear export localization sequence. Cells were stained with 10 μM **ThioPhenol** and incubated at 37 °C for 1 h and 30 min. Cells were not washed before imaging.

REFERENCES

REFERENCES

1 Toomre, D. & Bewersdorf, J. in *Annual Review of Cell and Developmental Biology, Vol 26* Vol. 26 *Annual Review of Cell and Developmental Biology* (eds R. Schekman, L. Goldstein, & R. Lehmann) 285-314 (2010).

2 Thorn, K. Genetically encoded fluorescent tags. *Mol. Biol. Cell* **28**, 848-857, doi:10.1091/mbc.E16-07-0504 (2017).

3 Zhao, N. *et al.* A genetically encoded probe for imaging nascent and mature HA-tagged proteins in vivo. *Nature Communications* **10**, doi:10.1038/s41467-019-10846-1 (2019).

4 Ni, Q., Mehta, S. & Zhang, J. Live-cell imaging of cell signaling using genetically encoded fluorescent reporters. *Febs Journal* **285**, 203-219, doi:10.1111/febs.14134 (2018).

5 Tantama, M., Hung, Y. P. & Yellen, G. Imaging Intracellular pH in Live Cells with a Genetically Encoded Red Fluorescent Protein Sensor. *Journal of the American Chemical Society* **133**, 10034-10037, doi:10.1021/ja202902d (2011).

6 Shaner, N. C., Patterson, G. H. & Davidson, M. W. Advances in fluorescent protein technology (vol 120, pg 4247, 2007). *J. Cell Sci.* **124**, 2321-2321, doi:10.1242/jcs.094722 (2011).

7 Chudakov, D. M., Matz, M. V., Lukyanov, S. & Lukyanov, K. A. Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiological Reviews* **90**, 1103-1163, doi:10.1152/physrev.00038.2009 (2010).

8 Nienhaus, K. & Nienhaus, G. U. Fluorescent proteins for live-cell imaging with super-resolution. *Chemical Society Reviews* **43**, 1088-1106, doi:10.1039/c3cs60171d (2014).

9 Day, R. N. & Davidson, M. W. The fluorescent protein palette: tools for cellular imaging (vol 38, pg 2887, 2009). *Chemical Society Reviews* **40**, 5923-5923 (2011).

10 Remington, S. J. Green fluorescent protein: A perspective. *Protein Sci.* **20**, 1509-1519, doi:10.1002/pro.684 (2011).

11 Sample, V., Newman, R. H. & Zhang, J. The structure and function of fluorescent proteins. *Chemical Society Reviews* **38**, 2852-2864, doi:10.1039/b913033k (2009).

12 Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat. Methods* **2**, 905-909, doi:10.1038/nmeth819 (2005).

13 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, doi:10.1016/j.ceb.2010.12.004 (2011).

14 Kremers, G. J., Gilbert, S. G., Cranfill, P. J., Davidson, M. W. & Piston, D. W. Fluorescent proteins at a glance (vol 124, pg 157, 2011). *J. Cell Sci.* **124**, 2676-2676, doi:10.1242/jcs.095059 (2011).

15 Chudakov, D. M., Lukyanov, S. & Lukyanov, K. A. Fluorescent proteins as a toolkit for in vivo imaging. *Trends in Biotechnology* **23**, 605-613, doi:10.1016/j.tibtech.2005.10.005 (2005).

16 Mena, M. A., Treynor, T. P., Mayo, S. L. & Daugherty, P. S. Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. *Nature Biotechnology* **24**, 1569-1571, doi:10.1038/nbt1264 (2006).

17 Stepanenko, O. V., Verkhusha, V. V., Kuznetsova, I. M., Uversky, V. N. & Turoverov, K. K. Fluorescent proteins as biomarkers and biosensors: Throwing color lights on molecular and cellular processes. *Current Protein & Peptide Science* **9**, 338-369, doi:10.2174/138920308785132668 (2008).

18 Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp red fluorescent protein. *Nature Biotechnology* **22**, 1567-1572, doi:10.1038/nbt1037 (2004).

19 Lukyanov, K. A., Chudakov, D. M., Lukyanov, S. & Verkhusha, V. V. Photoactivatable fluorescent proteins. *Nat. Rev. Mol. Cell Biol.* **6**, 885-891, doi:10.1038/nrm1741 (2005).

20 Shaner, N. C. *et al.* Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat. Methods* **5**, 545-551, doi:10.1038/nmeth.1209 (2008).

21 Costantini, L. M. *et al.* A palette of fluorescent proteins optimized for diverse cellular environments. *Nature Communications* **6**, doi:10.1038/ncomms8670 (2015).

22 McKinney, S. A., Murphy, C. S., Hazelwood, K. L., Davidson, M. W. & Looger, L. L. A bright and photostable photoconvertible fluorescent protein (vol 6, pg 133, 2009). *Nat. Methods* **8**, 606-606 (2011).

23 Durisic, N., Cuervo, L. L. & Lakadamyali, M. Quantitative super-resolution microscopy: pitfalls and strategies for image analysis. *Current Opinion in Chemical Biology* **20**, 22-28, doi:10.1016/j.cbpa.2014.04.005 (2014).

24 Costantini, L. M. & Snapp, E. L. Fluorescent Proteins in Cellular Organelles: Serious Pitfalls and Some Solutions. *DNA and Cell Biology* **32**, 622-627, doi:10.1089/dna.2013.2172 (2013).

25 Zhang, M. S. *et al.* Rational design of true monomeric and bright photoactivatable fluorescent proteins. *Nat. Methods* **9**, 727-U297, doi:10.1038/nmeth.2021 (2012).

26 Landgraf, D., Okumus, B., Chien, P., Baker, T. A. & Paulsson, J. Segregation of molecules at cell division reveals native protein localization. *Nat. Methods* **9**, 480-U498, doi:10.1038/nmeth.1955 (2012).

27 Jun, Y. W., Kim, H. R., Reo, Y. J., Dai, M. & Ahn, K. H. Addressing the autofluorescence issue in deep tissue imaging by two-photon microscopy:

the significance of far-red emitting dyes. *Chemical Science* **8**, 7696-7704, doi:10.1039/c7sc03362a (2017).

28 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, doi:10.1039/c2cs35313j (2013).

29 Hong, G. S., Antaris, A. L. & Dai, H. J. Near-infrared fluorophores for biomedical imaging. *Nat. Biomed. Eng* **1**, 22, doi:10.1038/s41551-016-0010 (2017).

30 Liu, P. F., Mu, X. Y., Zhang, X. D. & Ming, D. The Near-Infrared-II Fluorophores and Advanced Microscopy Technologies Development and Application in Bioimaging. *Bioconjugate Chemistry* **31**, 260-275, doi:10.1021/acs.bioconjchem.9b00610 (2020).

31 Chen, C., Tian, R., Zeng, Y., Chu, C. C. & Liu, G. Activatable Fluorescence Probes for "Turn-On" and Ratiometric Biosensing and Bioimaging: From NIR-I to NIR-II. *Bioconjugate Chemistry* **31**, 276-292, doi:10.1021/acs.bioconjchem.9b00734 (2020).

32 Chen, G. C. *et al.* Advanced Near-Infrared Light for Monitoring and Modulating the Spatiotemporal Dynamics of Cell Functions in Living Systems. *Advanced Science* **7**, doi:10.1002/advs.201903783 (2020).

33 Wu, S. & Butt, H. J. Near-infrared photochemistry at interfaces based on upconverting nanoparticles. *Physical Chemistry Chemical Physics* **19**, 23585-23596, doi:10.1039/c7cp01838j (2017).

34 Costantini, L. M., Fossati, M., Francolini, M. & Snapp, E. L. Assessing the Tendency of Fluorescent Proteins to Oligomerize Under Physiologic Conditions. *Traffic* **13**, 643-649, doi:10.1111/j.1600-0854.2012.01336.x (2012).

35 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, doi:10.1073/pnas.97.22.11984 (2000).

36 Wannier, T. M. *et al.* Monomerization of far-red fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E11294-E11301, doi:10.1073/pnas.1807449115 (2018).

37 Wiedenmann, J. *et al.* A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor (Anthozoa, Actinaria). *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11646-11651, doi:10.1073/pnas.182157199 (2002).

38 Fradkov, A. F. *et al.* Far-red fluorescent tag for protein labelling. *Biochemical Journal* **368**, 17-21, doi:10.1042/bj20021191 (2002).

39 Yanushevich, Y. G. *et al.* A strategy for the generation of non-aggregating mutants of Anthozoa fluorescent proteins. *Febs Letters* **511**, 11-14, doi:10.1016/s0014-5793(01)03263-x (2002).

40 Coralli, C., Cemazar, M., Kanthou, C., Tozer, G. M. & Dachs, G. U. Limitations of the reporter green fluorescent protein under simulated tumor conditions. *Cancer Research* **61**, 4784-4790 (2001).

41 Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509-544, doi:10.1146/annurev.biochem.67.1.509 (1998).

42 Wang, W. J. *et al.* Tuning the Electronic Absorption of Protein-Embedded All-trans-Retinal. *Science* **338**, 1340-1343, doi:10.1126/science.1226135 (2012).

43 Gunasekaran, K., Hagler, A. T. & Gierasch, L. M. Sequence and structural analysis of cellular retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions. *Proteins-Structure Function and Bioinformatics* **54**, 179-194, doi:10.1002/prot.10520 (2004).

44 Zhang, Y. R., Zhao, Y. Q. & Huang, J. F. Retinoid-Binding Proteins: Similar Protein Architectures Bind Similar Ligands via Completely Different Ways. *Plos One* **7**, doi:10.1371/journal.pone.0036772 (2012).

45 Folli, C., Calderone, V., Ramazzina, I., Zanotti, G. & Berni, R. Ligand binding and structural analysis of a human putative cellular retinol-binding

protein. *Journal of Biological Chemistry* **277**, 41970-41977, doi:10.1074/jbc.M207124200 (2002).

46 Ghyselinck, N. B. *et al.* Cellular retinol-binding protein I is essential for vitamin A homeostasis. *Embo Journal* **18**, 4903-4914, doi:10.1093/emboj/18.18.4903 (1999).

47 Nossoni, Z. *et al.* Structures of holo wild-type human cellular retinol-binding protein II (hCRBPII) bound to retinol and retinal. *Acta Crystallographica Section D-Structural Biology* **70**, 3226-3232, doi:10.1107/s1399004714023839 (2014).

48 Yapici, I. *et al.* "Turn-On" Protein Fluorescence: In Situ Formation of Cyanine Dyes. *Journal of the American Chemical Society* **137**, 1073-1080, doi:10.1021/ja506376j (2015).

49 Fiorella, P. D. & Napoli, J. L. EXPRESSION OF CELLULAR RETINOIC ACID BINDING-PROTEIN (CRABP) IN ESCHERICHIA-COLI - CHARACTERIZATION AND EVIDENCE THAT HOLO-CRABP IS A SUBSTRATE IN RETINOIC ACID METABOLISM. *Journal of Biological Chemistry* **266**, 16572-16579 (1991).

50 Donovan, M., Olofsson, B., Gustafson, A. L., Dencker, L. & Eriksson, U. THE CELLULAR RETINOIC ACID-BINDING PROTEINS. *J. Steroid. Biochem. Mol. Biol.* **53**, 459-465, doi:10.1016/0960-0760(95)00092-e (1995).

51 Santos, E. M. *et al.* Engineering of a Red Fluorogenic Protein/Merocyanine Complex for Live-Cell Imaging. *Chembiochem* **21**, 723-729, doi:10.1002/cbic.201900428 (2020).

52 Wang, L. *et al.* A general strategy to develop cell permeable and fluorogenic probes for multicolour nanoscopy. *Nature Chemistry* **12**, 165-+, doi:10.1038/s41557-019-0371-1 (2020).

53 Chan, J., Dodani, S. C. & Chang, C. J. Reaction-based small-molecule fluorescent probes for chemoselective bioimaging. *Nature Chemistry* **4**, 973-984, doi:10.1038/nchem.1500 (2012).

54 Grimm, J. B., Heckman, L. M. & Lavis, L. D. in *Fluorescence-Based Biosensors: From Concepts to Applications* Vol. 113 *Progress in Molecular Biology and Translational Science* (ed M. C. Morris) 1-34 (Elsevier Academic Press Inc, 2013).

55 Nadler, A. & Schultz, C. The Power of Fluorogenic Probes. *Angewandte Chemie-International Edition* **52**, 2408-2410, doi:10.1002/anie.201209733 (2013).

56 Chyan, W. & Raines, R. T. Enzyme-Activated Fluorogenic Probes for Live-Cell and in Vivo Imaging. *Acs Chemical Biology* **13**, 1810-1823, doi:10.1021/acschembio.8b00371 (2018).

57 Mizukami, S., Watanabe, S., Akimoto, Y. & Kikuchi, K. No-Wash Protein Labeling with Designed Fluorogenic Probes and Application to Real-Time Pulse-Chase Analysis. *Journal of the American Chemical Society* **134**, 1623-1629, doi:10.1021/ja208290f (2012).

58 Lukinavicius, G. *et al.* Fluorogenic probes for live-cell imaging of the cytoskeleton. *Nat. Methods* **11**, 731-U168, doi:10.1038/nmeth.2972 (2014).

59 Lukinavicius, G. *et al.* Fluorogenic Probes for Multicolor Imaging in Living Cells. *Journal of the American Chemical Society* **138**, 9365-9368, doi:10.1021/jacs.6b04782 (2016).

60 Tachibana, R. *et al.* Molecular design strategy of fluorogenic probes based on quantum chemical prediction of intramolecular spirocyclization. *Communications Chemistry* **3**, doi:10.1038/s42004-020-0326-x (2020).

61 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, doi:10.1016/j.cbpa.2013.05.015 (2013).

62 Ren, T. B. *et al.* A General Method To Increase Stokes Shift by Introducing Alternating Vibronic Structures. *Journal of the American Chemical Society* **140**, 7716-7722, doi:10.1021/jacs.8b04404 (2018).

63 Gao, Z., Hao, Y. C., Zheng, M. L. & Chen, Y. A fluorescent dye with large Stokes shift and high stability: synthesis and application to live cell imaging. *Rsc Advances* **7**, 7604-7609, doi:10.1039/c6ra27547h (2017).

64 Yang, C. C. *et al.* Impact of Stokes Shift on the Performance of Near-Infrared Harvesting Transparent Luminescent Solar Concentrators. *Scientific Reports* **8**, doi:10.1038/s41598-018-34442-3 (2018).

65 Horvath, P., Sebej, P., Solomek, T. & Klan, P. Small-Molecule Fluorophores with Large Stokes Shifts: 9-Iminopyronin Analogues as Clickable Tags. *Journal of Organic Chemistry* **80**, 1299-1311, doi:10.1021/jo502213t (2015).

66 Piatkevich, K. D. *et al.* Extended Stokes Shift in Fluorescent Proteins: Chromophore-Protein Interactions in a Near-Infrared TagRFP675 Variant. *Scientific Reports* **3**, doi:10.1038/srep01847 (2013).

67 Yang, Q. Z. *et al.* A mitochondria-targeting nitroreductase fluorescent probe with large Stokes shift and long-wavelength emission for imaging hypoxic status in tumor cells. *Analytica Chimica Acta* **1103**, 202-211, doi:10.1016/j.aca.2019.12.063 (2020).

68 Bogh, S. A. *et al.* Unusually large Stokes shift for a near-infrared emitting DNA-stabilized silver nanocluster. *Methods and Applications in Fluorescence* **6**, doi:10.1088/2050-6120/aaa8bc (2018).

69 Wang, S. F., Li, B. H. & Zhang, F. Molecular Fluorophores for Deep-Tissue Bioimaging. *Acs Central Science* **6**, 1302-1316, doi:10.1021/acscentsci.0c00544 (2020).

70 Kim, K. H. *et al.* Far-red/near-infrared emitting, two-photon absorbing, and bio-stable amino-Si-pyronin dyes. *Chemical Science* **10**, 9028-9037, doi:10.1039/c9sc02287b (2019).

71 Piatkevich, K. D., Subach, F. V. & Verkhusha, V. V. Far-red light photoactivatable near-infrared fluorescent proteins engineered from a bacterial phytochrome. *Nature Communications* **4**, doi:10.1038/ncomms3153 (2013).

72 Sansalone, L. *et al.* A Photoactivatable Far-Red/Near-Infrared BODIPY To Monitor Cellular Dynamics in Vivo. *Acs Sensors* **3**, 1347-1353, doi:10.1021/acssensors.8b00262 (2018).

73 Diao, S. *et al.* Biological imaging without autofluorescence in the second near-infrared region. *Nano Research* **8**, 3027-3034, doi:10.1007/s12274-015-0808-9 (2015).

74 Hilderbrand, S. A. & Weissleder, R. Near-infrared fluorescence: application to in vivo molecular imaging. *Current Opinion in Chemical Biology* **14**, 71-79, doi:10.1016/j.cbpa.2009.09.029 (2010).

75 Weissleder, R. & Ntziachristos, V. Shedding light onto live molecular targets. *Nature Medicine* **9**, 123-128, doi:10.1038/nm0103-123 (2003).

76 Smith, A. M., Mancini, M. C. & Nie, S. M. BIOIMAGING Second window for in vivo imaging. *Nature Nanotechnology* **4**, 710-711, doi:10.1038/nnano.2009.326 (2009).

77 Klymchenko, A. S. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Accounts of Chemical Research* **50**, 366-375, doi:10.1021/acs.accounts.6b00517 (2017).

78 Niko, Y., Didier, P., Mely, Y., Konishi, G. & Klymchenko, A. S. Bright and photostable push-pull pyrene dye visualizes lipid order variation between plasma and intracellular membranes. *Scientific Reports* **6**, doi:10.1038/srep18870 (2016).

79 Carlotti, B. *et al.* Evaluation of Hyperpolarizability from the Solvatochromic Method: Thiophene Containing Push-Pull Cationic Dyes as a Case Study. *Journal of Physical Chemistry C* **122**, 2285-2296, doi:10.1021/acs.jpcc.7b10647 (2018).

80 Dziuba, D. *et al.* Solvatochromic fluorene-linked nucleoside and DNA as color-changing fluorescent probes for sensing interactions. *Chemical Science* **7**, 5775-5785, doi:10.1039/c6sc02548j (2016).

81 Nigam, S. & Rutan, S. Principles and applications of solvatochromism. *Applied Spectroscopy* **55**, 362A-370A, doi:10.1366/0003702011953702 (2001).

82 Rauf, M. A. & Hisaindee, S. Studies on solvatochromic behavior of dyes using spectral techniques. *Journal of Molecular Structure* **1042**, 45-56, doi:10.1016/j.molstruc.2013.03.050 (2013).

83 Testoni, F. M., Ribeiro, E. A., Giusti, L. A. & Machado, V. G. Merocyanine solvatochromic dyes in the study of synergistic effects in mixtures of chloroform with hydrogen-bond accepting solvents. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **71**, 1704-1711, doi:10.1016/j.saa.2008.06.023 (2009).

84 Diwu, Z., Lu, Y. X., Zhang, C. L., Klaubert, D. H. & Haugland, R. P. Fluorescent molecular probes .2. The synthesis, spectral properties and use of fluorescent solvatochromic Dapoxyl(TM) dyes. *Photochemistry and Photobiology* **66**, 424-431, doi:10.1111/j.1751-1097.1997.tb03168.x (1997).

85 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, doi:10.1021/cc0700546 (2007).

86 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, doi:10.1016/j.dyepig.2009.04.008 (2009).

87 Santos, E. M. in PhD thesis. *Michigan State University* (2017).

88 Choquet, D., Sainlos, M. & Sibarita, J.-B. Advanced imaging and labelling methods to decipher brain cell organization and function. *Nature reviews. Neuroscience*, doi:10.1038/s41583-021-00441-z (2021).

89 Hansch, C., Leo, A. & Taft, R. W. A SURVEY OF HAMMETT SUBSTITUENT CONSTANTS AND RESONANCE AND FIELD

PARAMETERS. *Chemical Reviews* **91**, 165-195, doi:10.1021/cr00002a004 (1991).

90 Reichardt, C. SOLVATOCHROMIC DYES AS SOLVENT POLARITY INDICATORS. *Chemical Reviews* **94**, 2319-2358, doi:10.1021/cr00032a005 (1994).

91 Sheng, W. *et al.* A Near-Infrared Photoswitchable Protein-Fluorophore Tag for No-Wash Live Cell Imaging. *Angewandte Chemie-International Edition* **57**, 16083-16087, doi:10.1002/anie.201810065 (2018).

92 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. *J. Am. Chem. Soc.* **132**, 10762-10770, doi:10.1021/ja101974k (2010).

93 Piatkevich, K. D. *et al.* 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, doi:10.1073/pnas.0914365107 (2010).

94 Piatkevich, K. D. *et al.* Photoswitchable Red Fluorescent Protein with a Large Stokes Shift. *Chem. Biol.* **21**, 1402-1414, doi:10.1016/j.chembiol.2014.08.010 (2014).

95 Wang, W. in PhD thesis. *Michigan State University* (2012).

96 Jana, D. & Jana, S. Donor-Pyrene-Acceptor Distance-Dependent Intramolecular Charge-Transfer Process: A State-Specific Solvation Preferred to the Linear-Response Approach. *Acs Omega* **5**, 9944-9956, doi:10.1021/acsomega.0c00265 (2020).

97 Long, Y. *et al.* Cooperation of ESIPT and ICT Processes in the Designed 2-(2 '-Hydroxyphenyl)benzothiazole Derivative: A Near-Infrared Two-Photon Fluorescent Probe with a Large Stokes Shift for the Detection of Cysteine and Its Application in Biological Environments. *Analytical Chemistry* **92**, 14236-14243, doi:10.1021/acs.analchem.0c03490 (2020).

98 Berbasova, T. *et al.* Light-Activated Reversible Imine Isomerization: Towards a Photochromic Protein Switch. *Chembiochem* **17**, 407-414, doi:10.1002/cbic.201500613 (2016).

99 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, doi:10.1021/jacs.6b03681 (2016).

100 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, doi:10.1073/pnas.84.13.4441 (1987).

101 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, doi:10.1098/rstb.2004.1500 (2004).

102 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, doi:10.1074/jbc.R800080200 (2009).

CHAPTER III: DOUBLE EXCITED-STATE PROTON TRANSFER PHENOMENON: DEVELOPING PROTEIN TAGS WITH APPLICATIONS IN NO-WASH FLUORESCENT IMAGING

Proton-transfer reactions remain one of the most fundamental and vital processes in chemistry and biology.[1-3] Excited-state proton transfer (ESPT) plays a crucial role in many biological systems, such as photosystem II,[1,4,5] DNA,[6,7] bacteriorhodopsin,[8,9] and green fluorescent protein.[10,11] Given the importance of ESPT processes and its role as a functional tool in different areas such as fluorescent probes,[12-14] white light-emitting materials,[15,16] biological imaging,[17,18] pH jumps,[19,20] and triggers for protein folding,[21] a large number of experimental and theoretical studies have been focused on these systems to investigate the underlying dynamics and principles.

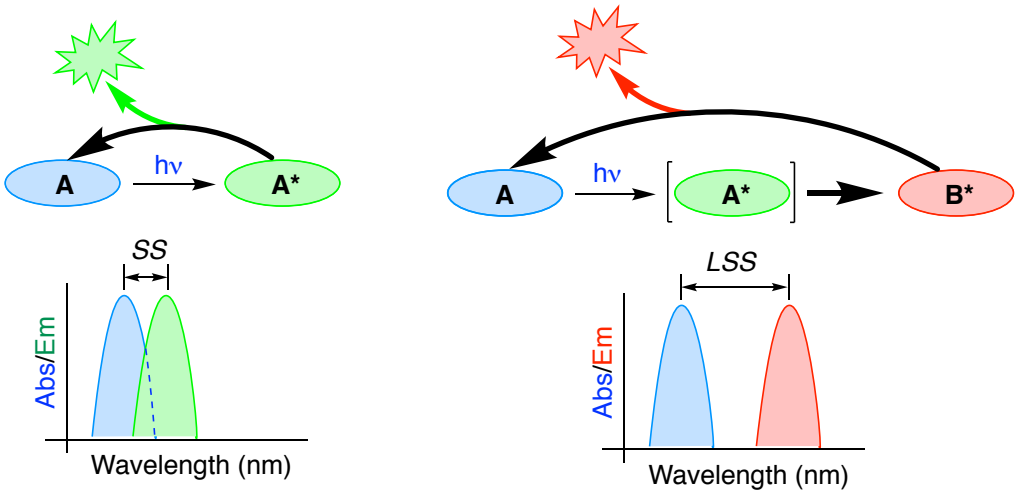
Photoacids and photobases are a class of molecular systems that exhibit a profound increase in acidity or basicity, respectively, upon photoexcitation into the first excited electronic state. As described in Chapter II, photoacidity is a reversible process as relaxation to the ground state changes the molecule back to the low-acidity state. Photoacids have been studied since the 1970s,[22-25] and they have found profound application in organic optoelectronic materials,[26,27] dye-sensitized ion exchange membranes,[28,29] large Stokes shift fluorescent proteins,[30,31] and to control molecular and supramolecular events. Generally, the ground state pK_a values range from 5 to 10, and the ΔpK_a (the difference between the ground and excited-state pK_a values) ranges from 3 to 13.[32]

As mentioned earlier, during the ESPT process, the chromophore moves along a proton to the protein through a proton wire inside the protein cavity. This phenomenon has been found in the Large Stokes Shift (LSS) red fluorescent proteins such as LSSmKate1 and LSSmKate2.[33-35] Structural analysis of other examples of ESPT-capable GFP variants, including mKeima, LSSmOrange, and mBeRFP, shows that the active site of these proteins can also be optimized to provide analogous proton relays to red-shift their emission. In particular, based on these predictions and rational engineering of the immediate chromophore environment, Piatkevich and coworkers produced LSS variants of several conventional orange and red FPs, including mCherry, mNeptune, mStrawberry, mOrange, and mKO through generating proton wires around the chromophore.[33]

Unlike photoacids, there are only a few reports of photobases in the literature, which are limited to heterocyclic amines such as acridines,[36,37] aminoanthraquinones,[38] 3-styrylpyridines,[39] Schiff bases,[40] and quinolines.[41] These photobases are distinct from photobase generators (PBGs) generated upon light irradiation of their salts, which have their pitfalls of being irreversible with slow proton transfer rates.[42] The paucity of photobases capable of generating more basic species upon photoirradiation has hindered their exploitation. Nevertheless, the ability to control the basicity of a compound upon photoexcitation is as important as those exhibited by photoacids and open to discovery.

III.1 Previous work toward developing large Stokes shift fluorescent protein-based photobases

As discussed previously, most ESPT-capable FPs are photoacids, wherein fluorescence is activated through proton transfer to the nearby residues upon excitation. A complementary approach would be a photobase in which photoirradiation leads to a more basic species capable of proton abstraction that leads to the generation of a cationic system. It was envisaged that a protein/chromophore complex with photobasic properties is capable of intramolecular charge transfer (ICT). If the photogenerated cation resides in conjugation with a polyene that is terminated with an electron-donating substituent, the photoactivation will lead to an electronic 'push-pull' system typical of large Stokes shift bathochromic pigments with ICT characteristics (**Figure III-1**).



Standard Fluorescence *Fluorescence from an Altered Excited State*

Figure III-1. Model for a large Stoke shift (LSS) vs. a standard fluorescence system. Large Stokes shift red-shifted emission results via conversion of the excited state species **A*** to a bathochromically distinct molecule **B***.

The inspiration to design such a system was our recent discovery. Dr. Wei Sheng reported the photobasic behavior and proton transfer dynamics of a fluorene-based imine, **FR0**-SB, a conjugated Schiff Base (SB) obtained upon the imine formation between the aldehydic form of **FR0** dye with *n*-butylamine.[43] The huge change in this system's basicity, 14 unit increase in pK_a, upon photoirradiation enables proton abstraction from protic solvents in its excited state, which generates the subsequent iminium that emit with over 200 nm apparent Stokes shift (**Figure III-2**).[43]

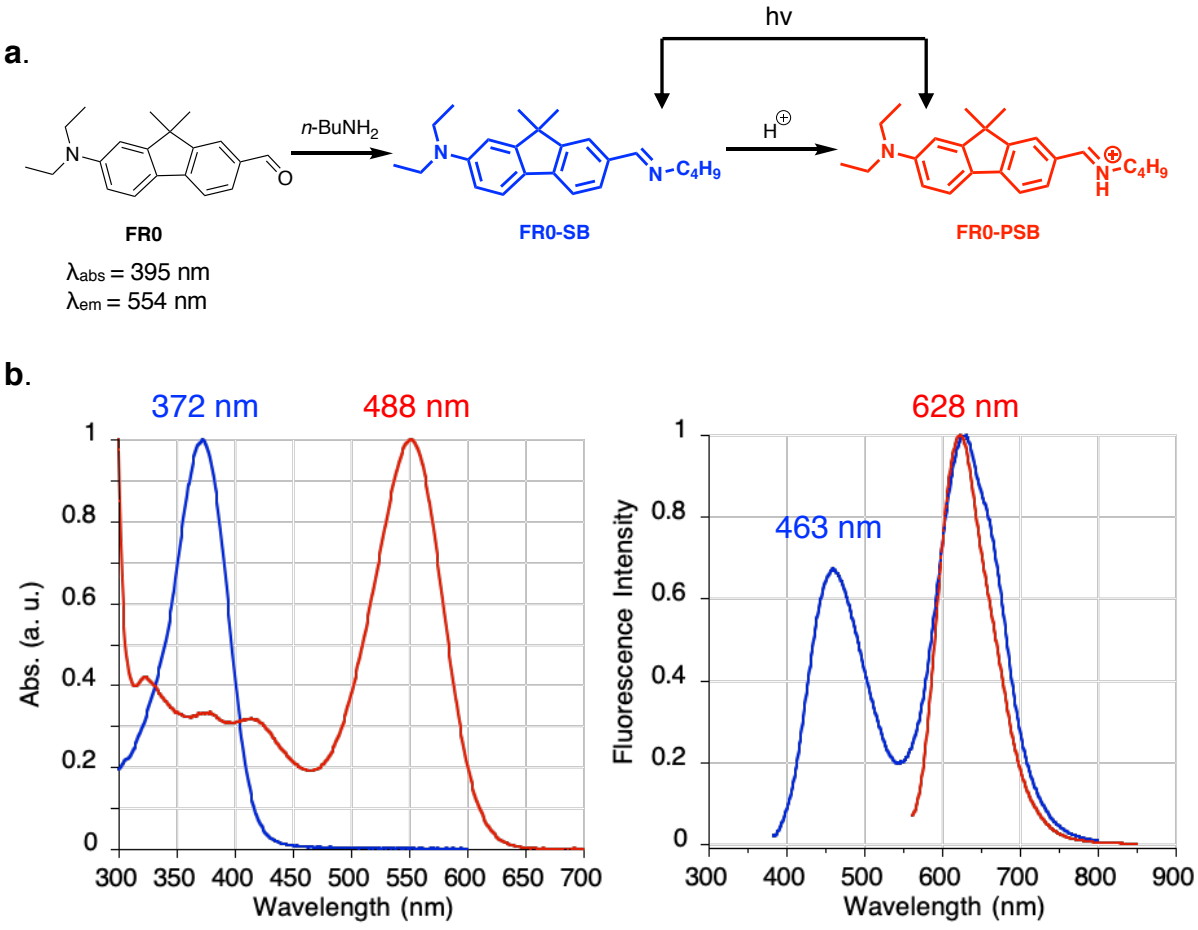


Figure III-2. **a**. **FR0** structure and formation of **FR0**-SB and **FR0**-PSB. **b**. UV-Vis (left) and fluorescence (right) spectra of **FR0**-SB (blue) and **FR0**-PSB (red) in ethanol.

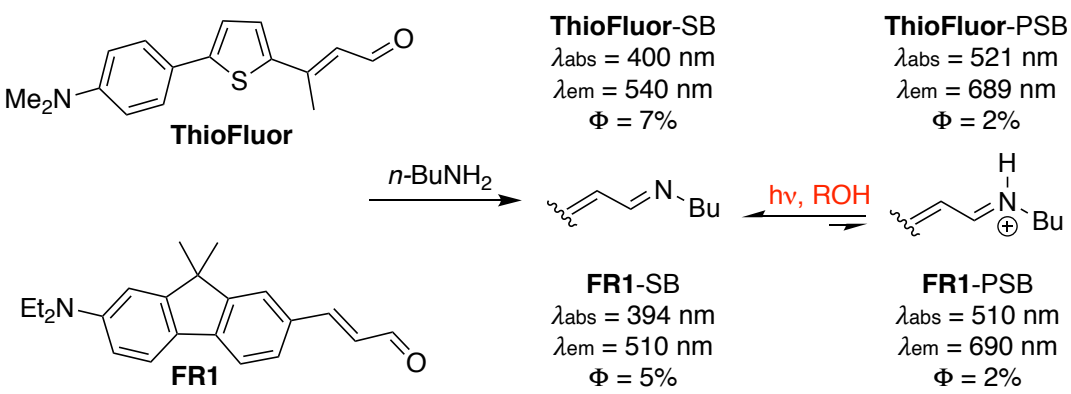
Excitation of **FR0**-SB at its maximum absorption wavelength (372 nm), when dissolved in protic solvents, results in dual emission bands with maxima at 463 nm and 628 nm corresponding to **FR0**-SB and **FR0**-PSB complexes, respectively. This assignment was confirmed as the excitation of the **FR0**-PSB, which can be produced upon acidification of the SB sample, gives rise to the same, red-shifted emission at about 628 nm (**Figure III-2**).

This discovery encouraged us to design fluorescent tags produced upon incorporating a fluorophoric photobase into a rationally engineered protein carrier to create photoactivatable LSS-FPs. Recently, Dr. Elizabeth Santos and Dr. Wei Sheng successfully designed protein-based photobases capable of generating highly conjugated polar ICT systems in the excited state that are well red-shifted in emission and show high fluorescence quantum efficiency. In their designed system, the fluorophore covalently binds hCRBPII through the reaction of the aldehydic ligands with an active site lysine residue. Additionally, the protein needs to maintain a low iminium pK_a to bind the fluorophore as an imine. The appropriate positioning of acidic amino acid sidechains is also critical in this study as they facilitate the proton transfer to the imine upon excitation and subsequent iminium generation.

Although **FR0** was the first ligand that showed photobasic properties, its strong activity as an ESPT-capable fluorophore is detrimental in achieving selectivity, and therefore, its application as a low background imaging tag is significantly hampered. On the contrary, the imine of **ThioFluor** and **FR1**

(derivative of **FR0**) exhibited weak ESPT in protic solvents (**Figure III-3b**), which is essential for cell imaging since then the iminium is only generated in the binding cavity of the engineered target protein.

a.



b.

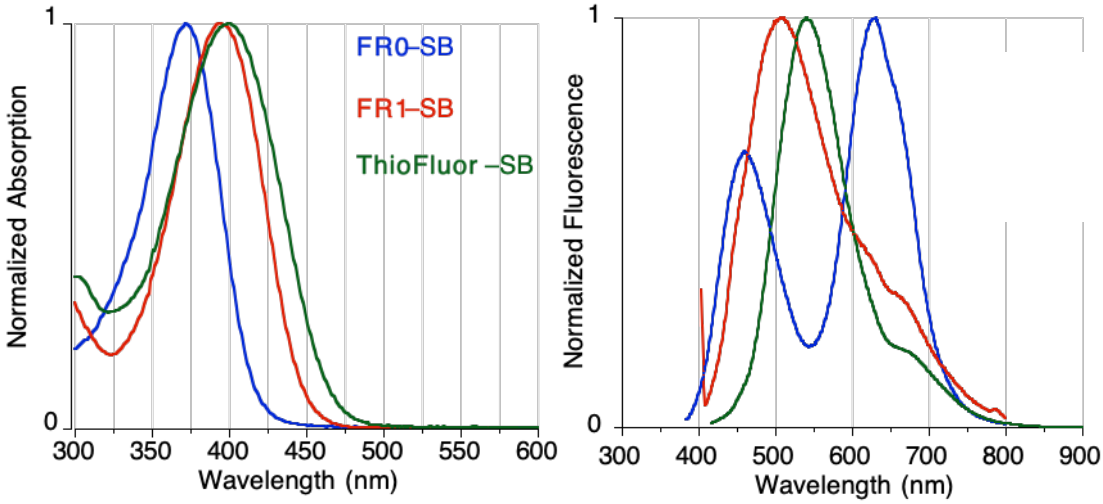


Figure III-3. **a**. Formation of the imine and iminium of **ThioFluor** and **FR1** in ethanol. **b**. Spectroscopic properties of **FR0** (blue), **FR1** (red), and **ThioFluor** (green) in ethanol. UV-Vis (left) and fluorescence spectra upon SB excitation (right).

Same as **FR0**, the selected ligands, **ThioFluor** and **FR1** (a derivative of **FR0**), provide the structural requirements for generating an ICT system; an

alkylamino group as the electron-donor, a π-spacer, and the imine unit as the electron-withdrawing group. The formation of the corresponding iminium upon the ESPT process results in a strong push-pull system, leading to an apparent large Stokes shift over 200 nm (**Figure III-3a**).

After extensive experiments, K40 was recognized as the best position to introduce acidic residues as the proton source for the ESPT process. We arrived at the Q108K:**K40E**:T53A:R58L:Q38F:Q4F-hCRBPII (**M3**) mutant exhibiting the highest values for Φ_{ESPT}. As described in Chapter II, we defined Φ_{ESPT} as the fraction of total fluorescence originating from the excited state iminium.

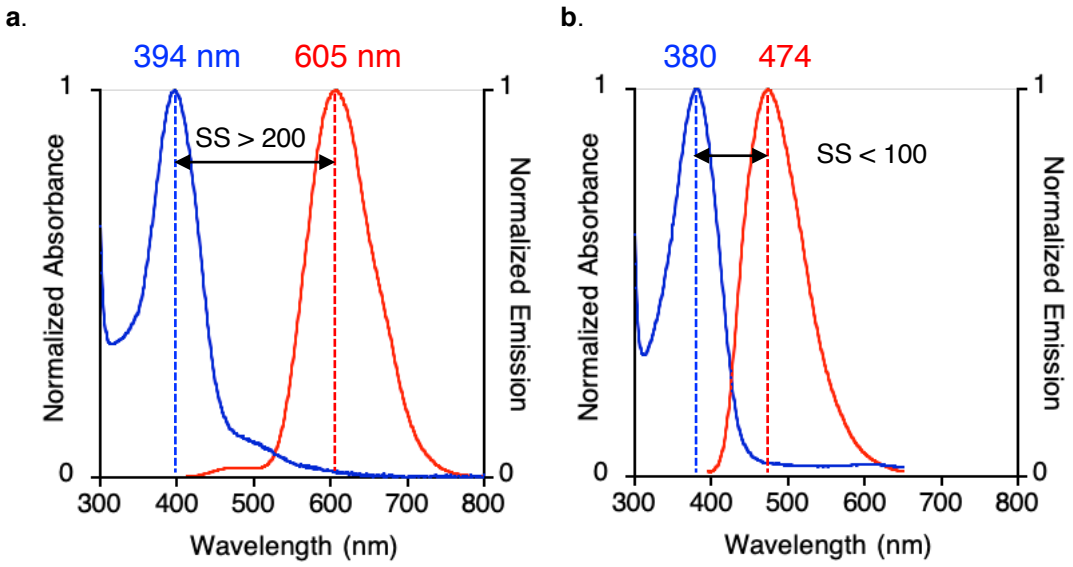


Figure III-4. UV-Vis and fluorescence spectra of **M3/ThioFluor** complex (394 nm excitation), exhibiting ESPT emission (left) vs. UV-Vis and fluorescence spectra of **M1/ThioFluor** complex (380 nm excitation) leads to SB emission at 474 nm (right). The spectra have been measured in PBS buffer at neutral pH (7.2).

Comparing the absorption and emission spectra of this mutant with that of Q108K:**K40L**:T53A:R58L:Q38F:Q4F-hCRBPII (**M1**) /**ThioFluor** complex in which

the acidic residue at position 40 (**40E**) is substituted with a neutral amino acid (**40L**) confirms that the ESPT originates from K40E mutation to the imine. As the emission wavelength is blue-shifted by about 130 nm and the Stokes shift is much smaller with the K40L mutant, which is an indication of the emission from the imine and not the iminium (**Figure III-4**).

M3 was selected for live-cell fluorescent imaging experiments. It shows low iminium pK_a values (~5.2); the Q4F mutation typically increases soluble protein expression, but more importantly, it helps to suppress the iminium pK_a leading to more imine formation in the ground state. Moreover, the three hydrophobic mutations, T53A, R58L, and Q38F, were introduced to increase binding affinity and enhance the rate of chromophore binding. The binding kinetics are also fast with this mutant (the binding is complete in less than 5 minutes) compared to non-specific imine formations, which is significantly important in increasing the selectivity for imaging the desired target.

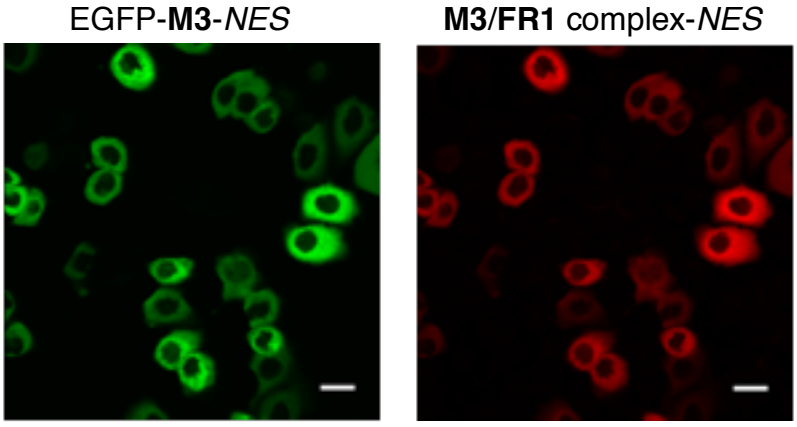


Figure III-5. **M3**/**FR1** imaging in live HeLa cells. NES = nuclear export sequence. Cells were stained with 500 nM **FR1** and incubated at 37 °C for 1 min. Cells were washed three times with DPBS before imaging. Scale bar, 10 μm.

Notably, using such a strategy to produce fluorescent tags with minimum fluorescence background is unique and unprecedented. Next, we chose to show the application of the photobasic complexes in live-cell imaging (**Figure III-5**). The required condition for cell imaging with **FR1** is one-minute incubation (500 nM) HeLa cells at 37 °C, followed by three washing steps with DPBS to remove unbound **FR1**. However, imaging experiments with **ThioFluor** complexes were not as successful since residual free ligand led to fluorescence background.

III.2 Designing hCRBPII/dye complexes that undergo double ESPT upon a single excitation

Chapter II described the photoacidic properties and excited-state proton transfer dynamics of **ThioPhenol** in complexation with hCRBPII. **ThioPhenol** as a photoacid emits in the far-red/NIR region of the spectrum (~700 nm) with a large Stokes shift of about 150 nm. We were able to develop a no-wash live-cell imaging using this system as the free **ThioPhenol** is unable to form a push-pull system and is not emissive. Furthermore, since the phenol moiety of the chromophore has a high pK_a, its deprotonation is only feasible in an engineered hCRBPII mutant leading to a substantial decrease in fluorescence background originating from non-specific imine or iminium formation.

Developing protein-based fluorescent tags with minimum background has been one of our lab's overarching goals. Successful no-wash background-free imaging experiments with **ThioPhenol** inspired us to explore its properties as a photobase. In addition, we sought to investigate whether **ThioPhenol** complexes

can be developed as a system that can show both photobasic and photoacidic characteristics upon a single excitation.

III.2.1 General protein host properties required for a double ESPT system

In order to develop the photobase part of the complex, the protein should be able to maintain a low iminium pK_a, so the imine is the only species formed in the ground state. At the same time, the mutations that helped deprotonate the phenol moiety should be retained to enhance photoacidic properties.

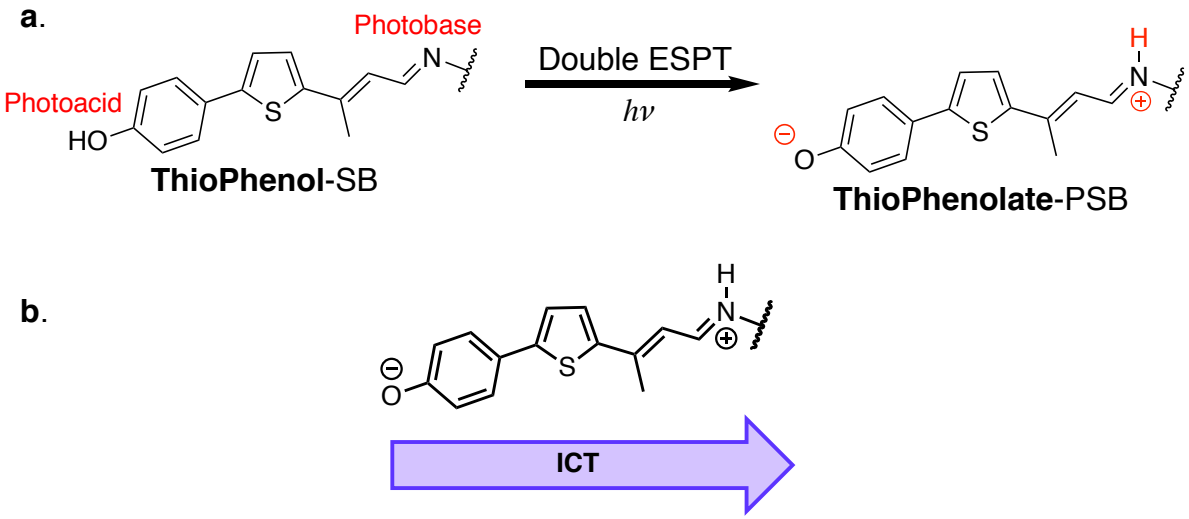


Figure III-6. **a**. Formation of **ThioPhenolate**-PSB complex through a double ESPT process. **b**. The internal charge transfer resulted from the **ThioPhenolate**-PSB complex, the product of the double ESPT process.

ThioPhenol binds the engineered protein as the **ThioPhenol**-SB and converts to **ThioPhenolate**-PSB through a double ESPT process in our proposed system (**Figure III-6a**). During this process, the SB grabs a proton from its vicinity and forms a Protonated Schiff Base (PSB), and simultaneously the phenol moiety loses a proton and generates the negatively charged oxygen (phenolate). The

excited state complex, **ThioPhenolate**-PSB, possesses a strong D–π–A and thus the resultant ICT leads to fluorescence (**Figure III-6b**).

III.3 ThioPhenol-hCRBPII as a photobasic system

As discussed in **Section III-1**, we recently discovered that glutamic acid at position 40 is the key residue in photobasic hCRBPII/fluorophore complexes that provides the proton for the ESPT process. The highest Φ_{ESPT} value was obtained with Q108K:K40E:T53A:R58L:Q38F:Q4F(**M3**)-hCRBPII mutant (>99%). We sought to investigate whether **ThioPheno**l/**M3** complex could provide the same results in the first step. For all measurements, 20 μM of protein is incubated with 0.5 equivalent of the ligands in PBS buffer at pH 7.2 (**Table III-1**).

Table III-1. Spectroscopic properties in complexation with **M3**.

Entry	Ligand	λ_{abs}	λ_{em}	SS	Φ_{ESPT}	Φ [a]	pK_a
1	**FR1**	392	595	203	> 0.99	0.72	5.2
2	**ThioFluor**	397	605	208	> 0.99	0.51	5.1
5	**ThioPhenol**	373	534	161	> 0.99	0.05	< 5

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

Both absorption and emission wavelengths are blue-shifted with **ThioPhenol** compared to the other ligands, about 21 nm and 60 nm, respectively. Due to the weak electron-donating properties of the hydroxyl group, **ThioPhenol** is unable to form a strong ICT system, which is effective in producing larger Stokes

shifts and red-shifted emission. Another outcome of the weak ICT characteristic of the **ThioPhenol/M3** complex is its decreased fluorescence quantum efficiency.

As shown in **Table III-1**, the emission wavelength with **ThioPhenol** is not as red-shifted as with the other ligands; however, emitting at 534 nm verifies the PSB formation upon excitation. Nonetheless, pH titration of the complex and monitoring its UV-Vis and emission spectra illustrates that the emission originates from the PSB.

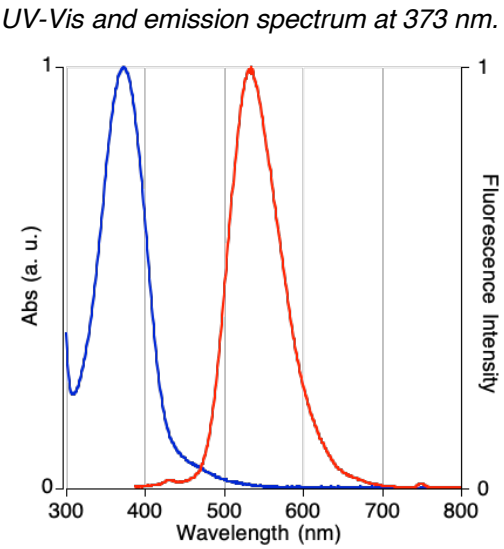


Figure III-7. UV-Vis, emission spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioPhenol** complex.

The absorption and emission spectra of the **ThioPhenol/M3** complex measured at neutral pH 7.4 are shown in **Figure III-7**. The lower iminium pK_a of this complex eliminates the residual PSB formation observed with **ThioFluor** and **FR1**. The acidification of the complex solution leads to the PSB formation absorbing at 488 nm; the excitation of PSB gives rise to the emission at 571 nm, which is 37 nm more red-shifted than the PSB produced upon excitation of the SB

(**Figure III-8a**). Such a trend in emission wavelength of PSB was observed with **ThioFlour** as well. A red-shift of 25 nm is observed for the emission wavelength of the PSB produced upon acidification of the sample compared to the PSB as the ESPT product (**Figure III-8b**).

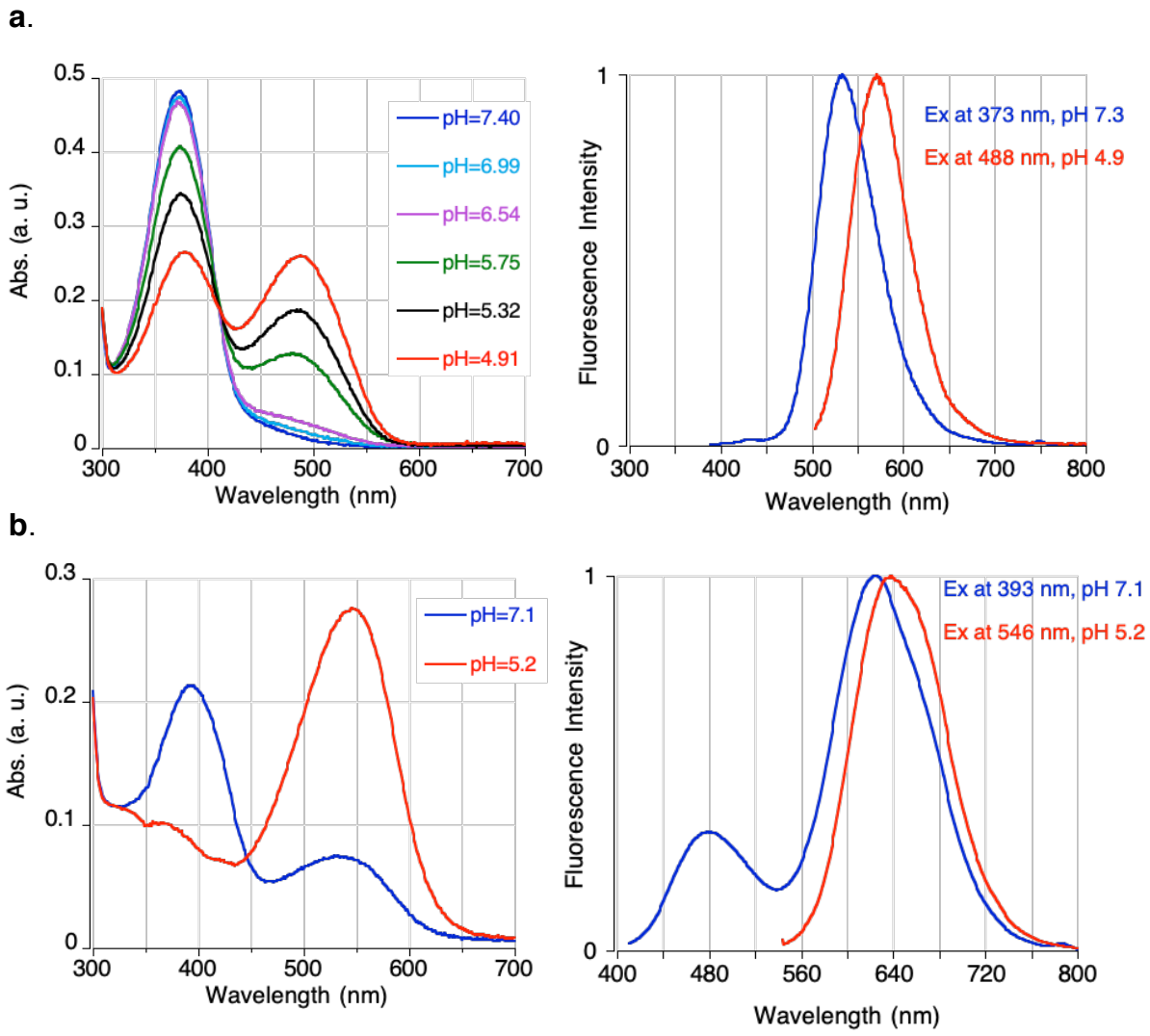


Figure III-8. **a**. UV-Vis (left) and fluorescence (right) spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioPhenol** complex upon acidification of the solution. **b**. UV-Vis (left) and fluorescence (right) spectra of Q108K:K40D:T53A:R58L:Q38F:Q4F/**ThioFluor** complex upon acidification of the solution.

The **ThioPhenol/M3** complex was stable enough upon basification of the sample to a pH value of more than 10. The absorption wavelength slightly red-shifts (about 10 nm) as the hydroxyl group becomes deprotonated and the **ThioPhenolat**e-SB complex forms. An interesting observation was made in this study as the excitation of SB even at pH 9.2 results in the same emission collected at neutral pH but with lower intensity, which is surprising since in such a basic pH, glutamic acid is not expected to act as a proton source for the ESPT process (**Figure III-9**).

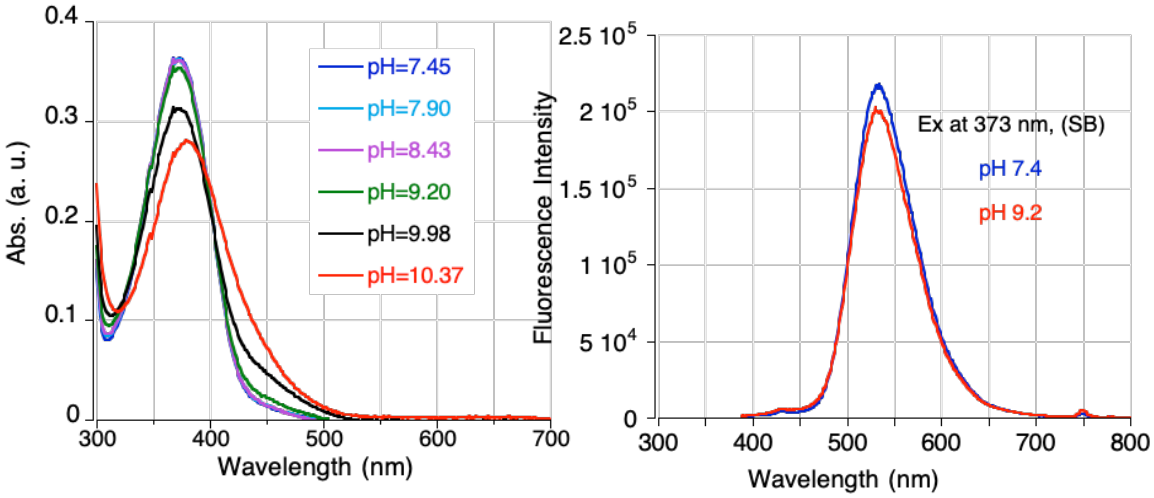


Figure III-9. UV-Vis (left) and fluorescence (right) spectra of Q108K:K40E:T53A:R58L:Q38F:Q4F/**ThioPhenol** complex upon basification of the solution.

To further study **ThioPhenol** complexes as a photobase, we chose to measure the photophysical properties in conjugation with Q108K:**K40L**:T53A:R58L:Q38F:Q4F (**M1**) mutant wherein the glutamic acid at position 40, the proton source, is substituted with leucine. The resultant complex, **ThioPhenol/M1**, shows a low iminium pK_a (5.1), leading to SB formation only.

However, surprisingly, in contrast to **ThioFluor**/**M1**, excitation of the SB at 370 nm led to the PSB emission at 531 nm (**Table III-2**), the same as what was collected with **M3** (see **Figure III-4b** and **Figure III-2**). Unfortunately, we were unable to obtain a crystal structure to investigate the possibility of other proton sources.

Table III-2. Spectroscopic properties in complexation with **M1**.

Entry	Ligand	λ_{abs}	$\lambda_{em\ (SB)}$	$\lambda_{em\ (PSB)}$	SS	Φ_{ESPT}	pK_a
1	**ThioFluor**	379	474	–	95	< 0.02	5.1
2	**ThioPhenol**	370	–	532	162	0.97	< 5

III.4 Developing ThioPhenol-hCRBPII complexes showing both photoacidic and photobasic characteristics

In Chapter II gradual evolution of **ThioPhenol**-hCRBPII complexes as a photoacidic system was elaborated; photoirradiation of the complex led to deprotonation of the chromophore's hydroxyl group and subsequent NIR emission with a large Stokes shift. Additionally, the successful application of **ThioPhenol**-SB as a protein-based photobase was discussed in the previous section. Thus, we sought to investigate engineering hCRBPII into a new class of fluorescent proteins that exhibits both photoacidic and photobasic properties upon a single excitation.

As shown in **Section II-7**, R58H is an essential mutation in deprotonating the hydroxyl group; therefore, leucine in **M3** was substituted with histidine, and the photophysical properties of the resultant complex (Q108K:**K40E**:T53A:**R58H**:Q38F:Q4F-hCRBPII (**M4**)/**ThioPhenol**) were obtained. **ThioPhenol** binds **M4** solely as imine with a maximum absorption wavelength of

370 nm. Gratifyingly, excitation of the SB results in an unprecedently large Stokes shift ~ 242 nm indicative of the double ESPT processes. To our delight, the Φ_{DESPT} (the quantum yield of double ESPT process) is more than 99% resulting in a well separated absorption and emission spectra and minimum self-absorption (**Figure III-10**). Φ_{DESPT} is defined as the fraction of total fluorescence from the excited state **ThioPhenolate**-PSB to the total fluorescence (see **Section II-7** for detailed description).

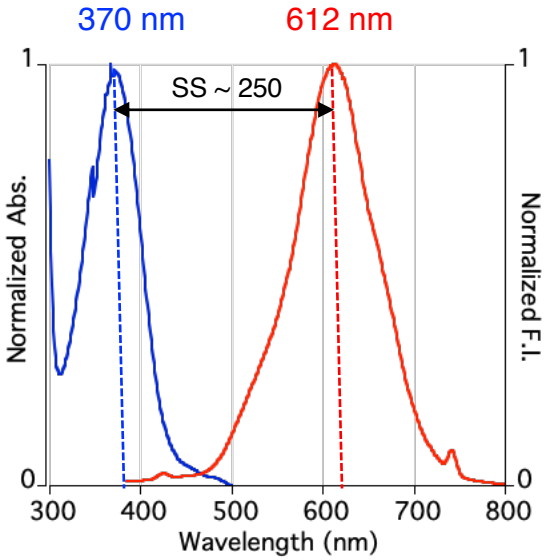


Figure III-10. Normalized absorption and emission spectra of Q108K:K40E:T53A:R58H:Q38F:Q4F/**ThioPhenol** complex, when excited at 370 nm at neutral pH (~7.2).

As proposed in **Section III-2**, two ESPT processes happen upon photoirradiation of this system. Presumably, the hydroxyl group acts as a photoacid losing a proton to R58H. At the same time, the SB segment acts as a photobase and grabs a proton resulting in PSB formation (see **Figure III-6**). In fact, the products of single and double ESPT processes are the same: **ThioPhenolate-**

PSB complex and the hydroxyl group becomes deprotonated in both designs. The only difference is that for the single ESPT processes, the iminium pK_a value is engineered to be high enough (~10), leading to a ground state PSB. While for dual proton transfer systems, the chromophore binds as a SB due to the low iminium pK_a (<5) and converts to the PSB upon photoexcitation (**Figure III-11**).

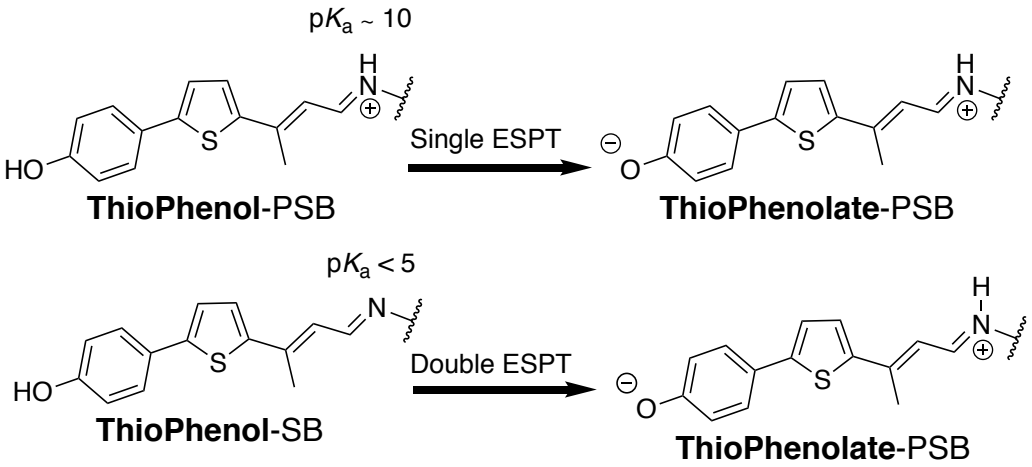


Figure III-11. Schematic representations of single and double ESPT processes. The starting complex is different, while the product is the same.

As described earlier, **ThioPhenol** photobasic complexes' quantum efficiency is low due to the weak electron-donating effect of the hydroxyl group. Therefore, enhanced brightness was expected for the double ESPT complexes as the Hammett value for the alkoxide group is much higher and almost the same as the *N,N* dimethyl amino group. The photophysical properties of **ThioPhenol**/**M3** and **M4** are compared in **Table III-3**.

Both mutants show the same imine absorption wavelength; however, excitation of **ThioPhenol**/**M4** imine at 370 nm results in 80 nm bathochromic shifts

in the emission wavelength compared to **ThioPhenol**/**M3** complex. Notably, the fluorescence quantum efficiency with **M4** is increased to 16% due to the strong ICT system formation upon excitation.

Table III-3. Spectroscopic properties of **ThioPhenol**/**M3** and **M4** complexes. complexation with **M1**.

Entry	hCRBPII mutant	λ_{abs}	λ_{em}	SS	Φ_{ESPT}	Φ [a]
M3	Q108K:K40E:T53A:**R58L**:Q38F:Q4F	370	532	162	0.97	0.05
M4	Q108K:K40E:T53A:**R58H**:Q38F:Q4F	370	612	242	0.98	0.16

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

To ensure that the red-shifted emission of the **ThioPhenol**/**M4** complex is the result of a double ESPT process, the UV-Vis and fluorescence spectra of this complex were monitored upon acid and base titration of the sample.

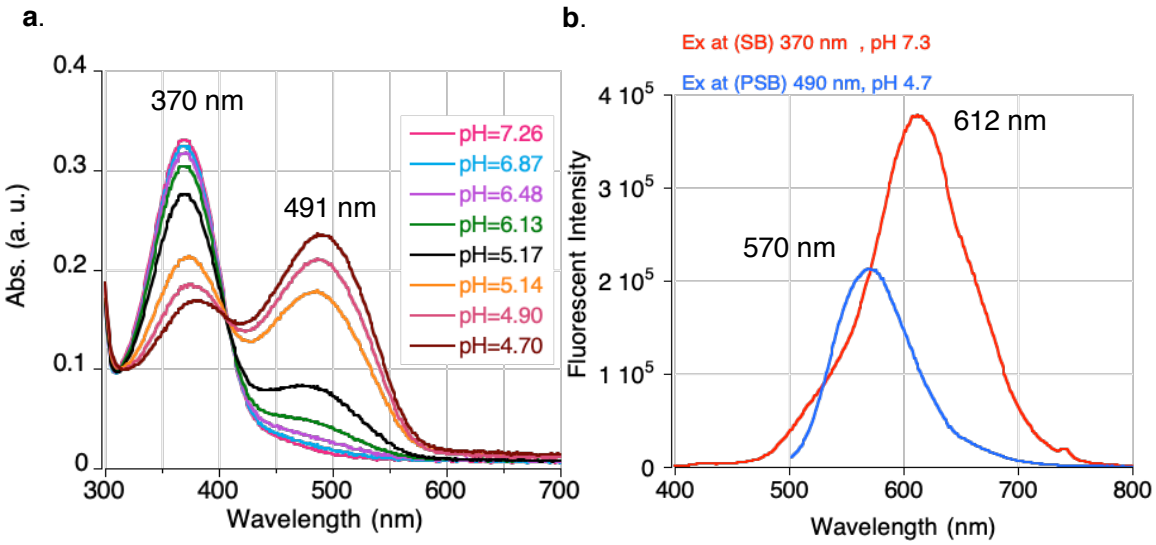


Figure III-12. **a**. UV-Vis spectra of Q108K:K40E:T53A:**R58H**:Q38F:Q4F/**ThioPhenol** complex upon acid titration of the sample. **b**. Fluorescence spectra of the same complex upon SB excitation at neutral pH and PSB excitation in acidic pH.

The PSB absorption peak at 491 nm starts to appear upon acidification of the complex solution to pH < 6. Putatively, the SB absorption decreases; however, it never disappears due to the low pK_a (**Figure III-12a**). PSB excitation at pH 4.7 leads to an emission wavelength at 570 nm, more than 40 nm blue-shifted when excited at SB, proving that the emission peak at 612 is due to the **ThioPhenolate**-PSB complex formation (**Figure III-12b**). Interestingly, the double ESPT process happens even at pH values up to 9.1. Excitation of the SB at pH more than 10 leads to a dual emission at 507 and 612 nm, which originates from the **ThioPhenolate**-SB and the **ThioPhenolate** PSB, respectively (**Figure III-13**).

III.5 Attempts to increase the fluorescence quantum efficiency of DESPT derived emission

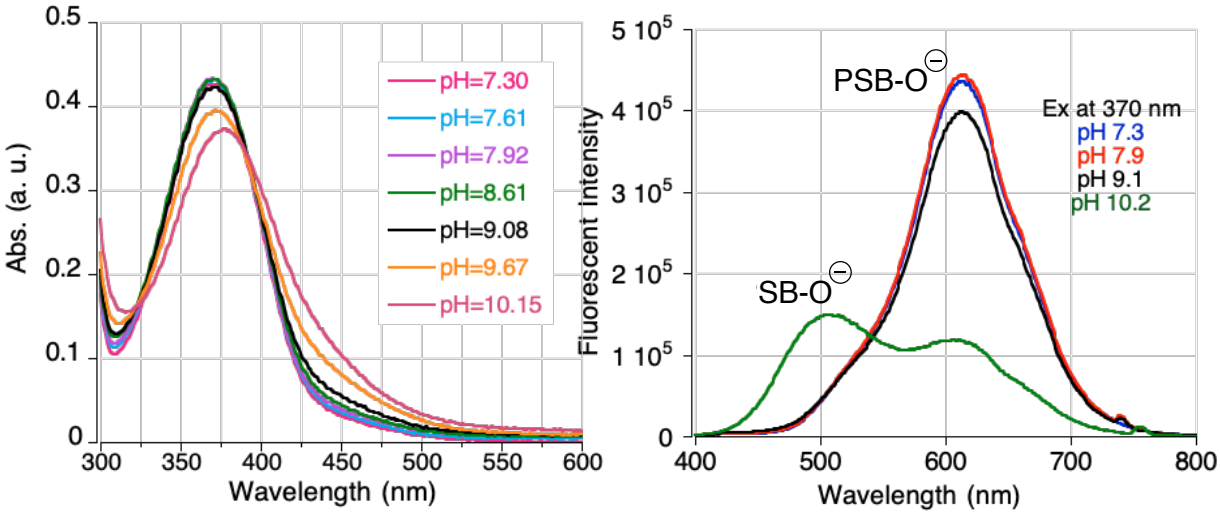


Figure III-13. UV-Vis spectra of Q108K:K40E:T53A:**R58H**:Q38F:Q4F/**ThioPhenol** complex upon base titration (left). Fluorescence spectra of the same complex upon SB excitation in different pH.

In our first attempt to generate a double ESPT system that shows both photoacidic and photobasic properties, R58H mutation, recognized as the key residue in the hydroxyl group deprotonation, was added to the photobasic protein mutant. The excitation of the SB in the corresponding complex, **ThioPhenol/M4**, results in an emission spectrum, in which its maximum wavelength is even more red-shifted than the PSB emission. A large Stokes shift (~250 nm) and high quantum yield (16%) of the red-shifted emission were acquired with this single mutation.

III.5.1 Expanding the water-mediated hydrogen bonding network around the hydroxyl group

Table III-4. Spectroscopic properties upon addition of F16Y, A33H and T29Y.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Red)	SS	Φ_{DESPT}	Φ
1	KEAFF:R58H:**F16Y**	370	594	224	> 0.99	0.16
2	KEAFF:R58H:**A33H**	375	613	238	0.98	0.15
3	KEAFF:R58H:**F16Y**:**A33H**	373	596	223	> 0.99	0.15
4	KEAFF:R58H:**Y19W**:**A33H**	370	614	244	> 0.99	0.14
5	KEAFF:R58H:**T29Y**	371	614	243	> 0.99	0.16
6	KEAFF:R58H:**F16Y**:**T29Y**	371	594	223	> 0.99	0.14

[a] KEAFF equals to Q108K:K40E:T53A:Q38F:Q4F.

As described in Chapter II, a few residues have played a critical role in expanding the hydrogen bonding network and consequently enhancing the Φ_{ESPT}. We sought to investigate whether the addition of those residues such as F16Y, A33H, T29Y, and their combinations could increase the fluorescence quantum

yield (**Table III-4**). The SB was the only species formed in neutral pH for all mutations due to the low iminium pK_a (in the range of <5 to 5.2). Interestingly, all mutants show excellent DESPT characteristics with none showing single ESPT emission around 530 nm (imine to iminium). The Φ_{DESPT} was more than 97% for all mutants. Noteworthy, mutants with F16Y mutation showed about 20 nm blue-shift in their emission wavelength (entries 1, 3, and 6), which is presumably due to the interaction of the phenol moiety and the tyrosine residue at position 16, leading to localization of the resultant phenoxide negative charge in the excited state (see **Section II-10**).

Table III-5. Spectroscopic properties as the result of substituting 58H with 58Y.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em} (Red)	SS	Φ_{DESPT}	Φ
1	KEAFF:**R58H**:F16Y	370	594	224	> 0.99	0.16
2	KEAFF:**R58Y**:F16Y	363	535	137	0.13	0.07

[a] KEAFF equals to Q108K:K40E:T53A:Q38F:Q4F.

To ensure that R58H is the essential residue for the dual proton transfer process, histidine at position 58 was substituted with tyrosine, another proton acceptor residue in the Q108K:K40E:T53A:Q38F:Q4F:R58H:F16Y template. The photophysical properties of the resultant mutant are shown in **Table III-5**. Interestingly, changing the histidine to tyrosine results in a broad peak maximized at 534 nm, the same as the emission wavelength of photobasic complexes such as **ThioPhenol/M3**. As a result, the Φ_{DESPT} decreased to less than 10%. These

results show that R58H mutation is vital in the double ESPT process; thus, this mutation is retained in all further protein engineering for this study.

In the course of engineering hCRBPII as a protein host capable of activating the photoacidity of **ThioPhenol**, we learned that methionine at position 51 helps to increase the Φ_{ESPT} (see **Section II-11.2 B**). Hence, we sought to explore the effect of this mutation on the fluorescence quantum efficiency of the double ESPT derived emission (**Table III-6**, entries 1 and 2).

Unfortunately, there is no crystal structure available for double ESPT proteins. However, overlaying different **ThioPhenol** crystal structures with **ThioFluor** shows that in the same protein mutant, the orientation of these two structures is similar. Even with a different iminium configuration, the residues surrounding the end of the chromophore are almost in the same orientation (see **Figure II-22**) Hence, with this assumption in mind, we decided to examine close residues to the nitrogen atom of the dimethylamino moiety in the crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFlour** complex (**Figure III-14**).

As shown in **Figure III-14**, F16, M20, A33, S76, and L77 are the closest residues to the nitrogen atom. However, our previous studies had demonstrated that the mutation of S76 is not well tolerated, and thus, this residue remained unchanged. Due to their close distance to the hydroxyl group, F16Y and A33H are the two key mutations that enhanced the Φ_{ESPT} and photoacid single ESPT emission quantum yield; however, the addition of these mutations to **M4** did not

result in improved photophysical properties (**Table III-4**). Hence, we chose to explore the effect of other proton acceptor residues at these positions on the double ESPT fluorescence quantum yield (**Table III-6**, 5, 6, and 7).

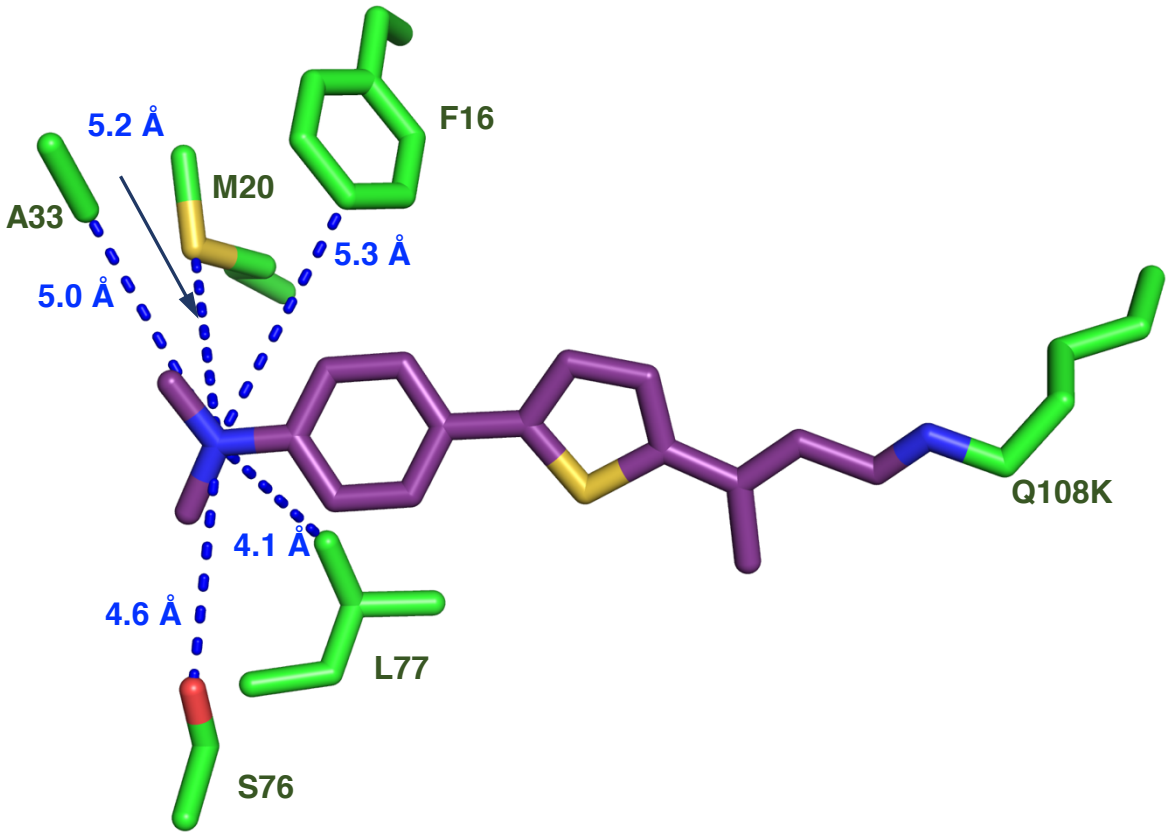


Figure III-14. The crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFlour** and the closest residues to the nitrogen atom of the dimethyl amino moiety of **ThioFluor**.

Additionally, M20 and L77 are in relatively close vicinity of the hydroxyl group, and thus several proton acceptors such as methionine, cysteine, and tyrosine were inserted at these positions. Unfortunately, most of these mutants did not yield soluble protein expressions. The photophysical properties of those that did express are listed in **Table III-6**, entries 2, 3, 8, and 9.

Table III-6. Spectroscopic properties of R58H containing mutants.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em}	SS	Φ_{DESPT}	Φ
1	KEAFF:R58H:**T51M**	375	611	236	0.97	0.07
2	KEAFF:R58H:**F16Y**:**T51M**	370	583	213	> 0.99	0.16
3	KEAFFQ:R58H:**L77M**	370	542/609	172/239	0.51	0.10
4	KEAFF:R58H:**F16Y**:L**77M**	370	596	226	> 0.99	0.12
5	KEAFF:R58H:**A33M**	375	539/608	164/233	0.47	0.05
6	KEAFF:R58H:**F16Y**:**A33M**	375	592	217	0.97	0.09
7	KEAFF:R58H:**F16Y**:**A33C**	375	595	220	> 0.99	0.14
8	KEAFF:R58H:**M20Y**	370	613	243	0.96	0.10
9	KEAFF:R58H:**F16Y**:**M20Y**[b]	380	545	165	0.36	0.04

[a] KEAFF equals to Q108K:K40E:T53A:Q38F:Q4F. [b] This mutant shows a higher iminium pK_a.

The addition of T51M mutation reduced the fluorescence quantum yield to 7% (entry 1), however in combination with F16Y mutation, it recovered to 16% same as for the **M4** complex (entry 2). Nonetheless. the Φ_{DESPT} is more than 0.95 for both mutants. Interestingly, the single addition of L77M mutation results in a doubly maximized fluorescence spectrum around 542 nm and 609 nm, corresponding to photobase single ESPT and doubles ESPT emission (**Figure III-**

15). Again, better results are acquired upon the addition of both F16Y and L77M mutation (entries 3 and 4).

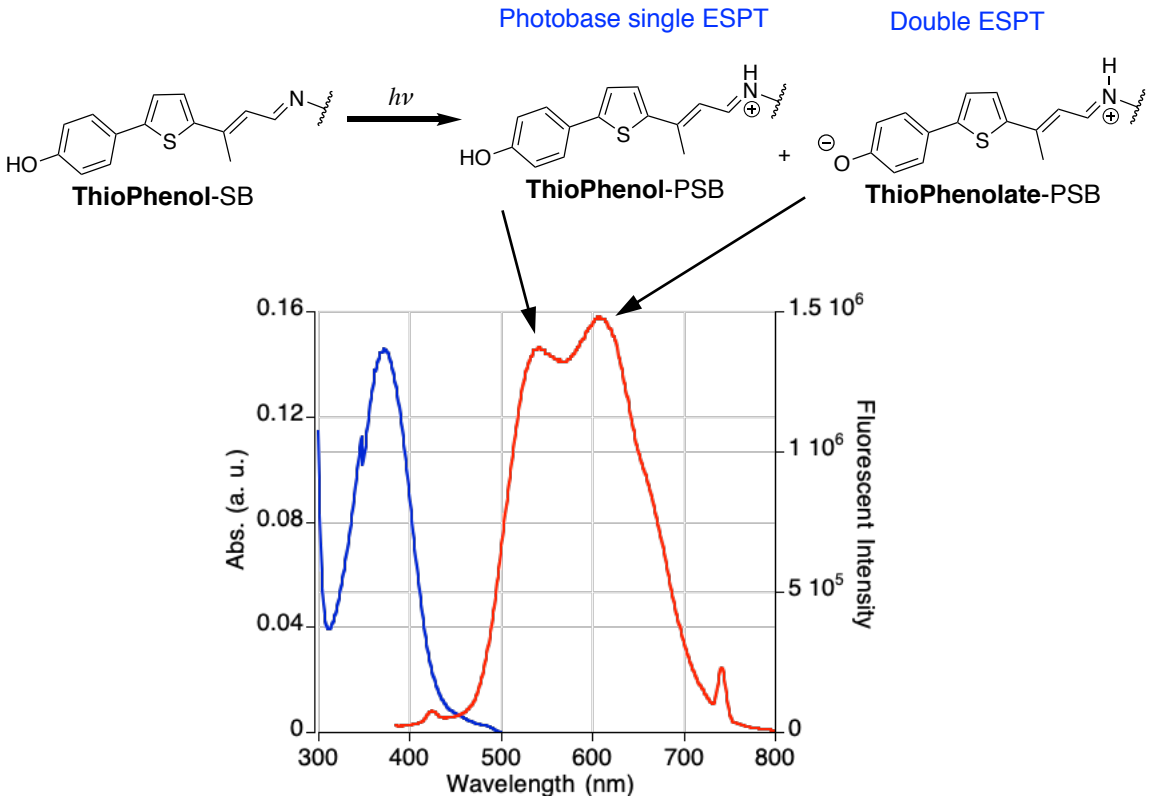


Figure III-15. Products of single ESPT and double ESPT processes. Absorption and emission spectrum of Q108K:K40E:T53A:Q38F:Q4F:R58H:L77M/**ThioPhenol** upon SB excitation at 370 nm.

The Φ_{DESPT} dramatically decreased upon the addition A33M mutation (entry 7). The excitation of the SB at 375 nm led to an emission spectrum maximized at 539 nm with a shoulder at 608 nm that can be characterized as shown above. The Φ_{DESPT} recovered with addition of F16Y mutation; however, the double ESPT emission quantum yield was slightly higher with Q108K:K40E:T53A:Q38F:Q4F:R58H:**F16Y**:**A33C** compared to the

Q108K:K40E:T53A:Q38F:Q4F:R58H:**F16Y**:**A33M** mutant (entries 6 and 7). The smallest Φ_{DESPT} was obtained with Q108K:K40E:T53A:Q38F:Q4F:R58H:**F16Y**:**M20Y** (entry 9). The iminium pK_a of this mutant is 5.9 resulting in residual PSB formation at neutral pH in the ground state. In addition, an interesting observation was made while measuring photophysical properties of this mutant: excitation of the PSB peak that appears as a shoulder at 520 nm leads to a low-intensity emission peak maximized at 674 nm, which is the wavelength for photoacidic single ESPT complexes.

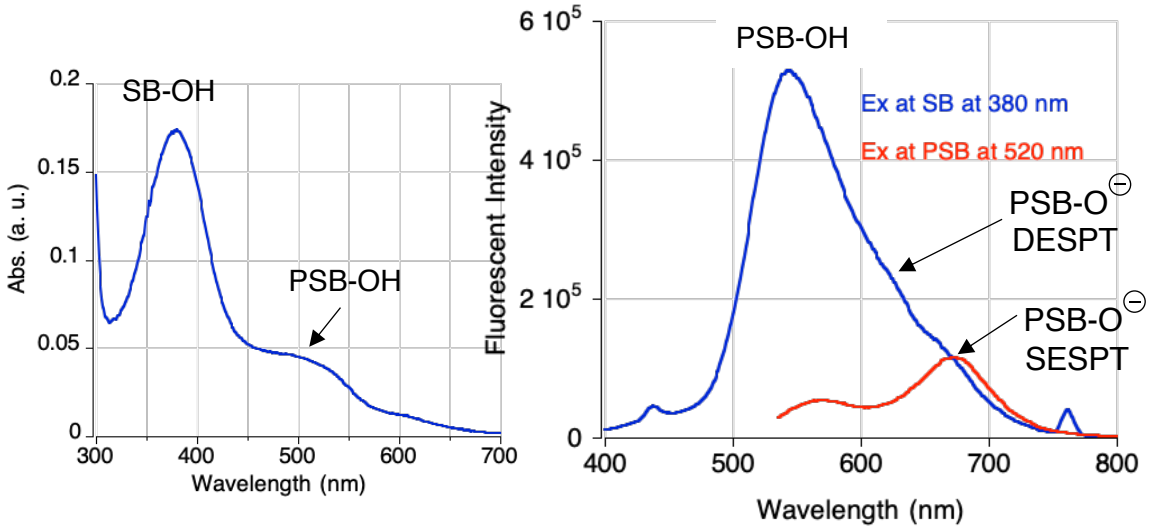


Figure III-16. Absorption (left) and emission spectra (right) of Q108K:K40E:T53A:Q38F:Q4F:R58H:F16Y:M20Y/**ThioPhenol** complex. Upon excitation of SB (blue line) and excitation of the PSB (red line). DESPT= Double ESPT. SESPT= Single ESPT.

Overlaying the emission spectra when exited at SB (380 nm) and PSB (520 nm) proves that the small shoulder in the right side of the double ESPT emission

of **M4/ThioPhenol** complex correlates to the photoacid single ESPT as discussed in **Section III.5.1** (**Figure III-16**).

III.6. Application of the double ESPT systems as a fluorescent tag for live cells imaging

We chose to explore the performance of hCRBPII/**ThioPhenol** double ESPT complexes as a no-wash live-cell imaging tag. The mutant Q108K:K40E:T53A:Q38F:Q4F:R58H/**ThioPhenol** (**M4**) was chosen for these studies. This mutant shows a low iminium pK_a leading to pure SB formation in the ground state. This complex yielded the highest Φ_{ESPT} and fluorescence quantum yield. Moreover, size exclusion chromatography data indicates that **M3** is expressed solely in the monomeric form, which is advantageous for imaging purposes. All spectroscopic properties of the **ThioPhenol/M4** complex are summarized below. Notably, the fast binding enables a quick labeling protocol with 5 min staining at 37 °C.

Table III-7. Spectroscopic properties of **ThioPhenol/M4** complex.

Complex	λ_{abs}	λ_{em}	ε ($M^{-1}.cm^{-1}$)	Φ_{ESPT}	Φ [a]	pK_a	$t_{1/2}$ [b] (min)
ThioPhenol/M4	370	612	35,081	0.98	0.16	< 5	< 3

[a] Absolute quantum yield was measured on a Quantaurus-QY. [b] Half-life based on the rate constant obtained from Second order rate fitting; measured at 23 °C with 20 µM protein and 0.5 equiv **ThioPhenol** at pH 7.2.

The **M3**-hCRBPII was cloned into the pFlag-CMV2 vector containing EGFP fused on the N-terminus of hCRBPII to label the whole cell. Additionally, to target

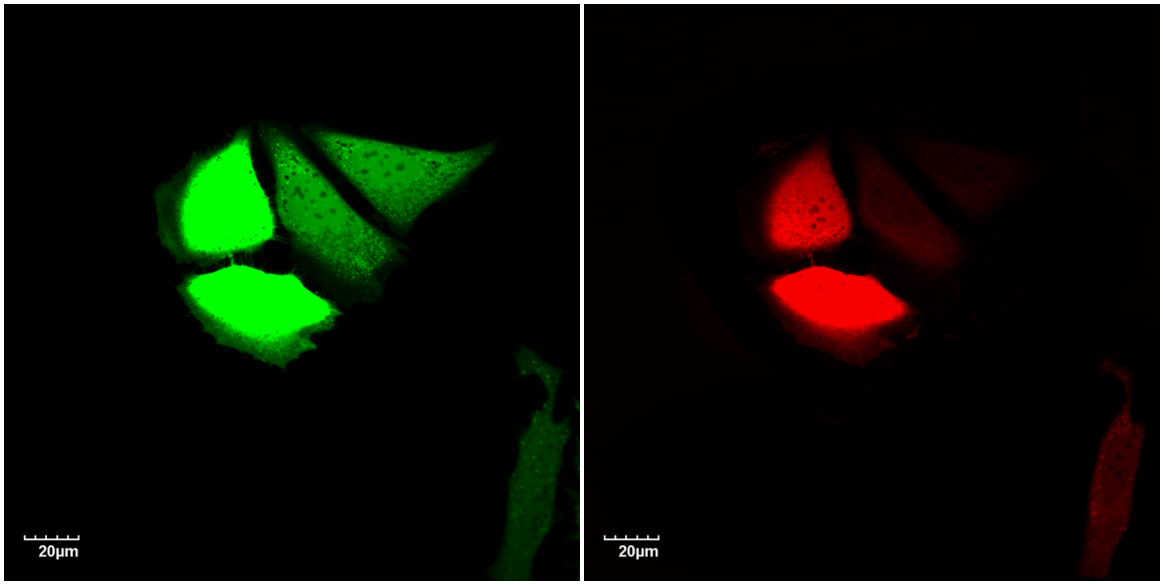
M4 at cell nuclei, the signaling peptide NLS (nuclear localization sequence) was fused to the C-terminus of **M4** (see **Figure II-32**).

Cells were transfected, and the fused constructs were expressed in HeLa cancer cell lines. Imaging was performed by incubating the HeLa cells with 10 μM **ThioPhenol** for five minutes at 37 °C. The cells were then directly subjected to confocal imaging without any washing steps prior to imaging.

In all cases, the green fluorescence was observed upon excitation at 488 nm, indicating that transfection was successful, and the fusion protein was expressed (**Figure III-17**, the green channel). Stained cells were then excited via a 405 nm laser to collect the far-red emission in the 550 nm-650 nm window (**Figure III-17**, the red channel). Apparently, no fluorescence background is observed in the red channel, indicating that **Thiophenol** does not label off-target lysines or non-specific bindings does not lead to fluorescence signal. These results prove the utility of no-wash labeling of Q108K:K40E:T53A:Q38F:Q4F:R58H mutant with **ThioPhenol** in mammalian cells with no fluorescent background.

EGFP-*whole cell*
Ex: 488 nm
Em: BP 500 nm- 550 nm

ThioPhenol/hCRBPII-*whole cell*
Ex: 405 nm
Em: BP 550 nm- 650 nm



Green + Red channel

DIC
Ex: 488 nm
BP 500 nm- 550 nm

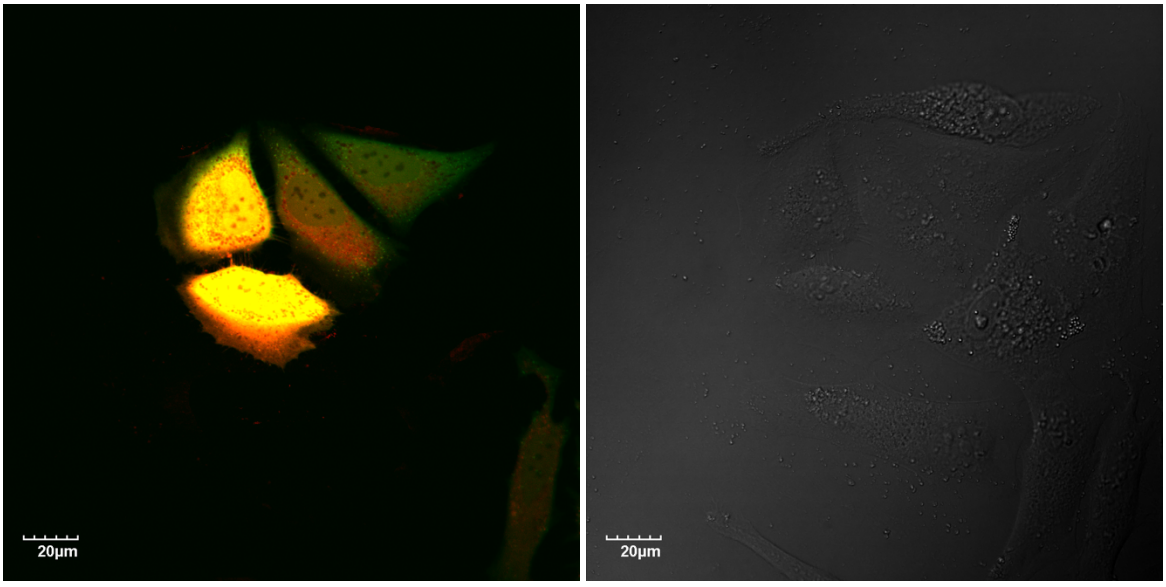


Figure III-17. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-*whole cell*. Cells were stained with 10 µM **ThioPhenol** and incubated at 37 °C for 5 min. Cells were not washed before imaging.

III.7 Enhancing the double ESPT process with brighter chromophore cores

Attempts to enhance the quantum efficiency of the dual ESPT emission through the insertion of basic residues in the vicinity of the **ThioPhenol**'s hydroxyl group were not fruitful. The highest Φ_{DESPT} and fluoresce quantum efficiency was acquired with the Q108K:K40E:T53A:Q38F:Q4F:R58H- (**M4**)/**ThioPhenol** complex, 98% and 16%, respectively, the very first complex we tested to examine the double ESPT process. Association of R58H with other proton acceptor residues leads to either the same photophysical characteristics as **M4** or reduced Φ_{DESPT} as the result of contribution from photobase single ESPT emission (see **Tables III-4 and III-6**).

As discussed in Chapter II, free **ThioPhenol** is not emissive due to its weak ICT in the ground state. Deprotonation of the hydroxyl group, which occurs upon excitation produces the corresponding alkoxide, a strong electron-donating group, leading to fluorescence activation. This system is perfect for no-wash background-free imaging applications with unprecedented fluorescent signal specificity, as the deprotonation happens only upon binding to the engineered target protein.

However, in comparison to other commercial fluorophores such as FRs, rhodamine, or cyanine dyes, the brightness of our system needs to be improved. Therefore, the next part of this chapter describes our attempts to increase the brightness of the double ESPT system while the specificity is kept intact. Hence, in the next step, we sought to investigate if the double ESPT process is possible with brighter chromophore cores such as **FR0** and **FR1**.

III.7.1 MR0

To examine the possibility of the double ESPT process with other chromophores, we decided to optimize the structure of **FR0.** This scaffold was selected for its extremely high quantum yield and relatively high extinction coefficient. Mr. Mehdi Moemeni successfully synthesized and purified **MR0**, wherein the *N,N* diethyl group is substituted with the hydroxyl group (**Figure III-18,** see **Section IV-7** for detailed synthesis).

Figure III-18. Structures of **ThioPhenol**, **MR0**, and **FR0**.

III.7.1.A. Spectroscopic properties of MR0 in solution

First, we measured **MR0** absorption, emission, and quantum yield in various organic solvents with different polarities to investigate its spectroscopic behavior. Results are summarized in **Tables III-7**. **MR0** shows almost the same absorbance wavelength in different solvents, with the maximum centered around 339 nm. On the other hand, the maximum emission wavelength is more sensitive to the environment's polarity; spanning from 382 nm in toluene to 568 nm in aqueous solvents (see **Figure III-19** and **Table III-7**).

The most interesting result of this study is, however, the significant increase of the fluorescence quantum yield upon increasing solvent polarity. **MR0** is not emissive in less polar solvents; the quantum yield is less than 0.01 in toluene and ethyl acetate, but it reaches to 0.33 in ethanol. Fortunately, these characteristics help to eliminate the fluorescent background from unbound **MR0** in the hydrophobic cell environments.

Table III-8. Spectroscopic characterization of **MR0** in various solvents.

Entry	Solvent	λ_{abs} (nm)	λ_{em} (nm)	SS (nm)	ε ($M^{-1}cm^{-1}$)	Φ [a]
1	Toluene	341	382	41	30,025	0.00
2	Ethyl acetate	338	394	56	29,548	0.00
3	Dimethyl sulfoxide	334	434	90	30,501	0.13
4	Ethanol	341	459	118	23,511	0.33
5	PBS buffer	339	492/568	153/229	25,974	0.15

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

Next, we compared the spectral properties of the three chromophores, **FR0**, **MR0**, and **ThioPhenol**, in non-polar (toluene), and polar (ethanol) environments (**Table III-8**). **FR0** shows the highest quantum yields, while **ThioPhenol** is not emissive as a free aldehyde regardless of the solvent polarity. However, **MR0** exhibits zero quantum yield in non-polar solvents and becomes bright in ethanol.

Although there is the same number of conjugated double bonds in **MR0** and **FR0 s**tructures, the absorption wavelength of the former is blue-shifted by 55 nm due to the less electron-donating ability of the hydroxyl group. **MR0** absorption wavelength is blue-shifted by about 50 nm compared to **ThioPhenol** as well. Nonetheless, the brightness ($\varepsilon * \Phi$) of **MR0** is 26 times higher than **ThioPhenol**.

Table III-9. Spectroscopic features of **MR0**, **FR0**, and **ThioPhenol** in toluene and ethanol.

	MR0			**FR0**			**ThioPhenol**		
	λ_{abs}	λ_{em}	Φ^a	λ_{abs}	λ_{em}	Φ^a	λ_{abs}	λ_{em}	Φ^s
Toluene	341	381	0.00	396	434	0.70	386	461	0.0
Ethanol	341	459	0.33	396	556	0.66	391	515	0.01

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

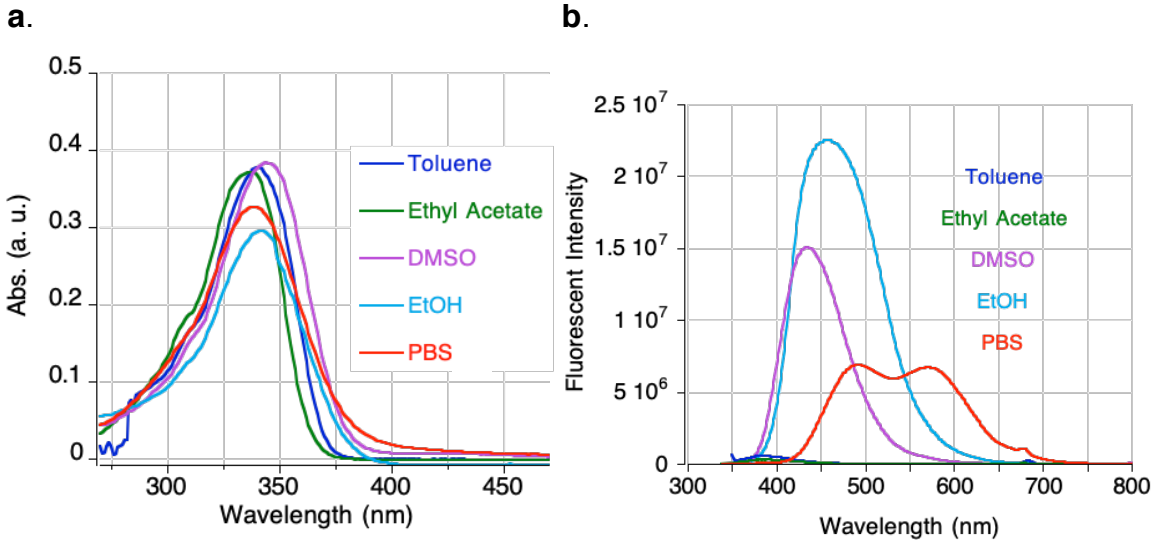


Figure III-19. Spectroscopic properties of **MR0** in different solvents. **a**. UV-Vis and **b**. Fluorescence spectra.

An interesting observation made in this study is that free **MR0** emission measurement in water or other aqueous solutions such as BPS buffer results in two maxima at 492 nm and 568 nm, respectively (see **Figure III-19** and **Table III-7**). Hence, we speculated that the red-shifted second peak at 568 nm might be due to the deprotonation of the **MR0** hydroxyl group. To confirm if this assumption is correct, the absorption and emission of **MR0** were measured upon acid and base titrations (**Figure III-20**).

a. **b**.

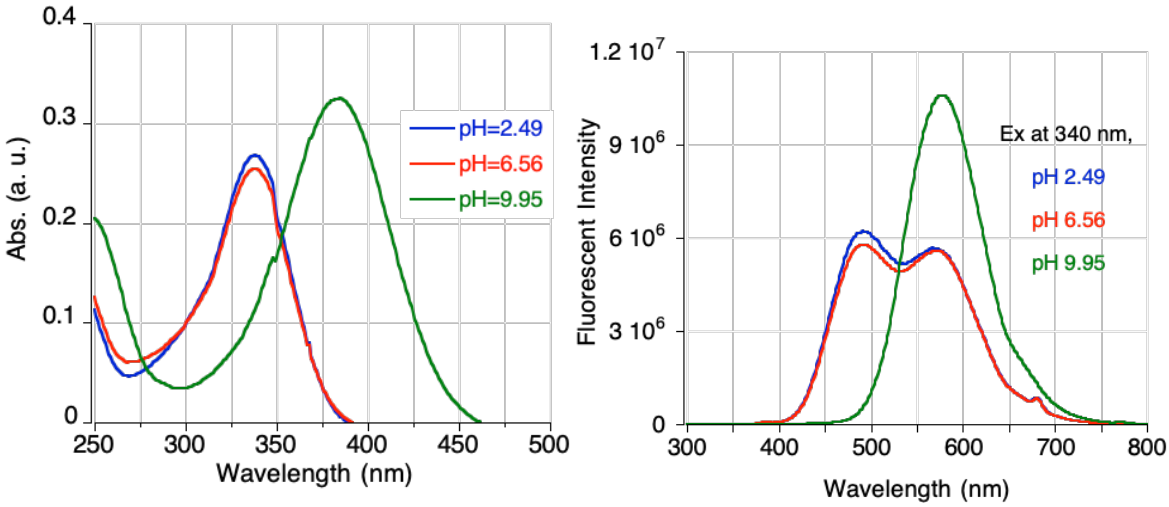


Figure III-20. UV-Vis spectra (left) and fluorescence spectra (right) of free aldehyde **MR0** upon acidification and basification of the sample. Measurements were done in water.

The absorption and emission spectra of **MR0** did not change upon acidification of the sample to pH 2.5. However, increasing the pH of the sample to 10 red shifts the absorption wavelength from 338 nm to 384 nm (46 nm). Apparently, deprotonation of the hydroxyl group of **MR0** results in the red-shifted absorption.

Interestingly, excitation of the **MR0**-phenolate leads to the emission wavelength maximized at 577 nm, close to the red-shifted emission peak of **MR0**. This observation verifies that in contrast to **ThioPhenol**, about half of the **MR0** becomes deprotonated in aqueous solutions at neutral pH (results are summarized in **Table III-9**).

Table III-10. Spectroscopic features of free aldehyde **MR0** in water in acidic and basic pH.

Free Aldehyde **MR0**	λ_{abs}	λ_{em}	SS (nm)
Neutral/Acidic pH	338	492/570	154/232
Basic pH	384	578	194

Next, in order to mimic the product of the protein bound complex, the aldehyde was coupled with *n*-butyl amine in ethanol (**Figure III-21**). As was previously discussed for ICT systems, it is expected that the reduced electron-withdrawing ability of the acceptor results in a blue-shift of the absorption and emission wavelength. Accordingly, in this system, imine formation leads to 9 and 54 nm blue-shifts in the absorption and emission wavelength, respectively, due to the replacement of the oxygen atom with the less electronegative nitrogen atom. Furthermore, to better characterize this system before moving into the complex environment of the protein cavity, the imine sample was acidified and basified, and the photophysical properties of PSB and phenolate-SB-complex were measured (the results are summarized in **Table III-9**).

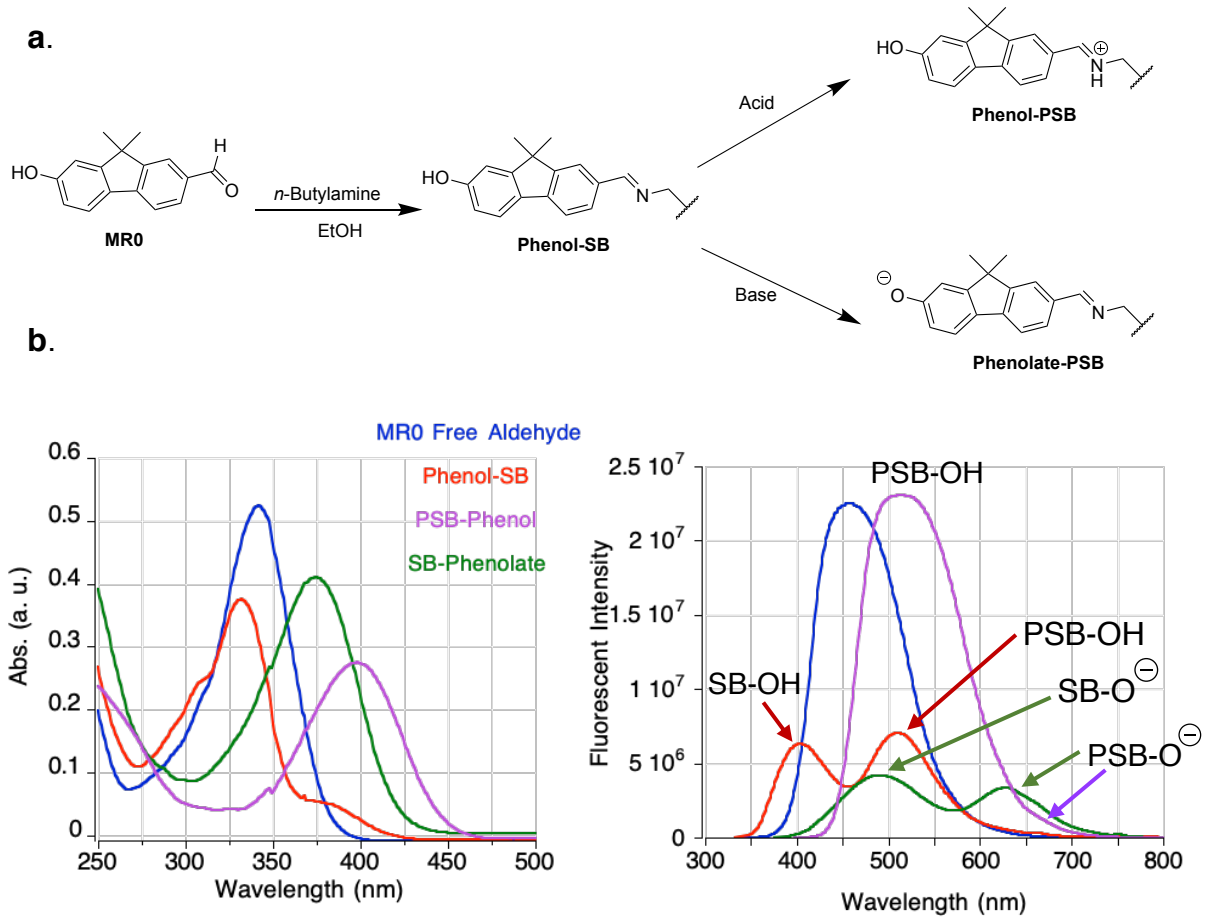


Figure III-21. **a**. Schiff base and protonated Schiff base (PSB) of **MR0** with *n*-butyl amine in ethanol. **b**. Absorbance (left) and emission (right) spectra of **MR0** and derivatives: Phenol-SB, Phenol-PSB, and Phenolate-PSB.

Most of the **MR0** forms Phenol-SB upon condensation with *n*-butylamine; however, a small portion of the product is protonated, which appears as a small shoulder to the right side of the SB absorption peak at about 383 nm. Excitation of Phenol-SB complex at its maximum absorption wavelength (332 nm) leads to an emission spectrum with maxima at 405 and 509 nm. Apparently, the phenol-SB complex acts as a photobase and can abstract a proton upon photoirradiation,

resulting in a Stokes shift of more than 175 nm. The maximum emission wavelength of the Phenol-PSB complex at 510 nm verifies the accuracy of this statement.

Table III-11. Spectroscopic characterization of **MR0** and derivatives.

Compound	λ_{abs}	λ_{em}	SS (nm)	ε ($M^{-1}.cm^{-1}$)	Φ [a]	Φ_{ESPT}
Free Aldehyde **MR0**	341	459	118	23,511	0.33	-
Phenol-SB	332	405/509	73/177	16,928	0.11	51.2
Phenol-PSB	398	510	112	12,360	0.40	-
Phenolate-SB	375	489/628	114/**253**	18,406	0.07	48.6

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

Interestingly, excitation of the Phenolate-SB produces a two-maxima emission peak as well at 489 and 628 nm, corresponding to Phenolate-SB and Phenolate-PSB complexes, respectively. This highlights the strong photobasic characteristics of the **MR0**-SB complex; even in such a basic environment where even though the hydroxyl group is deprotonated, photoexcitation of SB leads to PSB formation. Such photobasic properties were not observed for **ThioPhenol** in solution (see **Section II. 5**). Nonetheless, **MR0** photoacidic properties are not as strong, since photoirradiation of the Phenol-PSB complex shows a small peak at 650 nm, which might be due to the deprotonation of the hydroxyl group (**Figure III-21b**).

To investigate further about photobasic features of **MR0**, we chose to measure the absorption and emission of the phenol-SB in PBS, which might better resemble the protein cavity environment (**Table III-10**). Both the absorption and emission wavelength of Phenol-SB show bathochromic shifts, and the Φ_{ESPT} (the percentage of conversion to PSB upon excitation) is significantly increased (>90%) in BPS buffer as compared to when measured in ethanol.

Table III-12. Spectroscopic properties of **MR0** Phenol-SB in PBS buffer.

Compound	λ_{abs}	λ_{em}	SS (nm)	ε ($M^{-1}.cm^{-1}$)	Φ_{ESPT}
Phenol-SB	348	470/598	122/250	18,439	92.6

Same as **FR0**, the strong photobasic activity of **MR0** could be detrimental for developing selective low-background fluorescent tags. However, this section aims to investigate whether the double ESPT process is possible with brighter chromophore structures, which then facilitates studying the kinetics of this unique phenomenon. Nonetheless, the emission of the double ESPT process is most red-shifted and is well-separated from the Phenol-PSB emission; thus, the chances of background form cross-talk emissions will be diminished.

III.7.1.B. Attempts to develop double ESPT process with MR0-hCRBPII complex

As discussed previously, **FR1**, **ThioFluor**, and **ThioPhenol** exhibit photobasic properties in complexation with **M3**. Additionally, **M4** is the optimized

mutant for the double ESPT process. Hence, we decided to measure the photophysical properties of **MR0** with Q108K:K40E:T53A:**R58L**:Q38F:Q4F (**M3**) and Q108K:K40E:T53A:**R58H**:Q38F:Q4F (**M4**) hCRBPII mutants. Evident from the absorption, **MR0** is bound as a SB to both mutants. According to in-solution studies with **MR0**, the photogenerated Phenol-PSB emission wavelength is about 510 nm, which correlates with the **M3** complex emission wavelength (**Table III-11**, entry 1). [a] Absolute quantum yield was measured on a Quantaurus-QY.

Interestingly, the emission wavelength with **M4** (the double ESPT mutant)

Table III-13. Spectroscopic properties of **MR0**/**M3** and **MR0**/**M4**.

Entry	Complex	λ_{abs}	λ_{em}	SS	Φ [a]
1	**MR0**/**M3**	348	505	157	0.56
2	**MR0**/**M4**	347	525	178	0.31

is only 20 nm more red-shitted than **M3**. We expected that the double ESPT derived emission to be more red-shifted as the photogenerated Phenolate-PSB complex maximum emission wavelength is 628 nm in ethanol. Furthermore, the fluorescence quantum efficiency of **MR0**/**M3** is higher, which might be due to the absence of the double ESPT process with the **MR0**/**M4** complex. As discussed in **Section III.5.1**, in contrast to Phenolate-PSB, Phenol-PSB or Phenolate-SB complexes are unable to produce a strong ICT system and are less emissive (see **Table III-3**).

We presumed that the photoexcitation of the **MR0**/**M4** complex might lead to deprotonation of the hydroxyl group and Phenolate-SB complex formation. To test this hypothesis, we measured the photophysical properties of the **MR0**/**M4** and **MR0**/**M3** complexes in basic pH (**Table III-12**).

In addition, we chose to explore and compare the photophysical properties of **MR0** with Q108K:K40L:T51V:T53S:**R58H**:**F16Y**:**A33H**:L117C (**M5**) mutant that shows photoacidic characteristics with **ThioPhenol** (**Table III-12**). The absorption and emission spectra of **MR0** with **M3**, **M4**, and **M5** mutants were monitored upon base titration of their solutions in PBS buffer (**Figure III-22**).

Table III-14. Spectroscopic properties of **MR0**/**M3** and **MR0**/**M4**, and **MR0**/**M5**.

Entry	hCRBPII Mutant [a]	Neutral pH 7.3		Basic pH 10	
		λ_{abs}	λ_{em}	λ_{abs}	λ_{em}
1	**M3**	348	505	336/371	430/576
2	**M4**	347	525	332/364	430/575
3	**M5**	340	493/572	382	576

[a] **M3**: Q108K:K40E:T53A:R58L:Q38F:Q4F (photobase mutant),
M4: Q108K:K40E:T53A:R58H:Q38F:Q4F (double ESPT mutant),
M5: Q108K:K40L:T51V:T53S:R58H:F16Y:A33H:L117C (photoacid mutant).

Interestingly, all mutants show a two-maximized emission peak upon basification of the sample. The maximum emission wavelengths are the same for all mutants; however, the ratio of the red-shifted peak at 575 nm to the blue-shifted one at 430 nm is different (see **Figure III-22**, emission spectrum at pH ~10, the green line). Nonetheless, monitoring the UV-Vis spectra of these mutants upon

basification reveals that the absorption spectrum measured at pH ~10 (the green line) shows an irregular shape, and, in fact, it consists of two peaks.

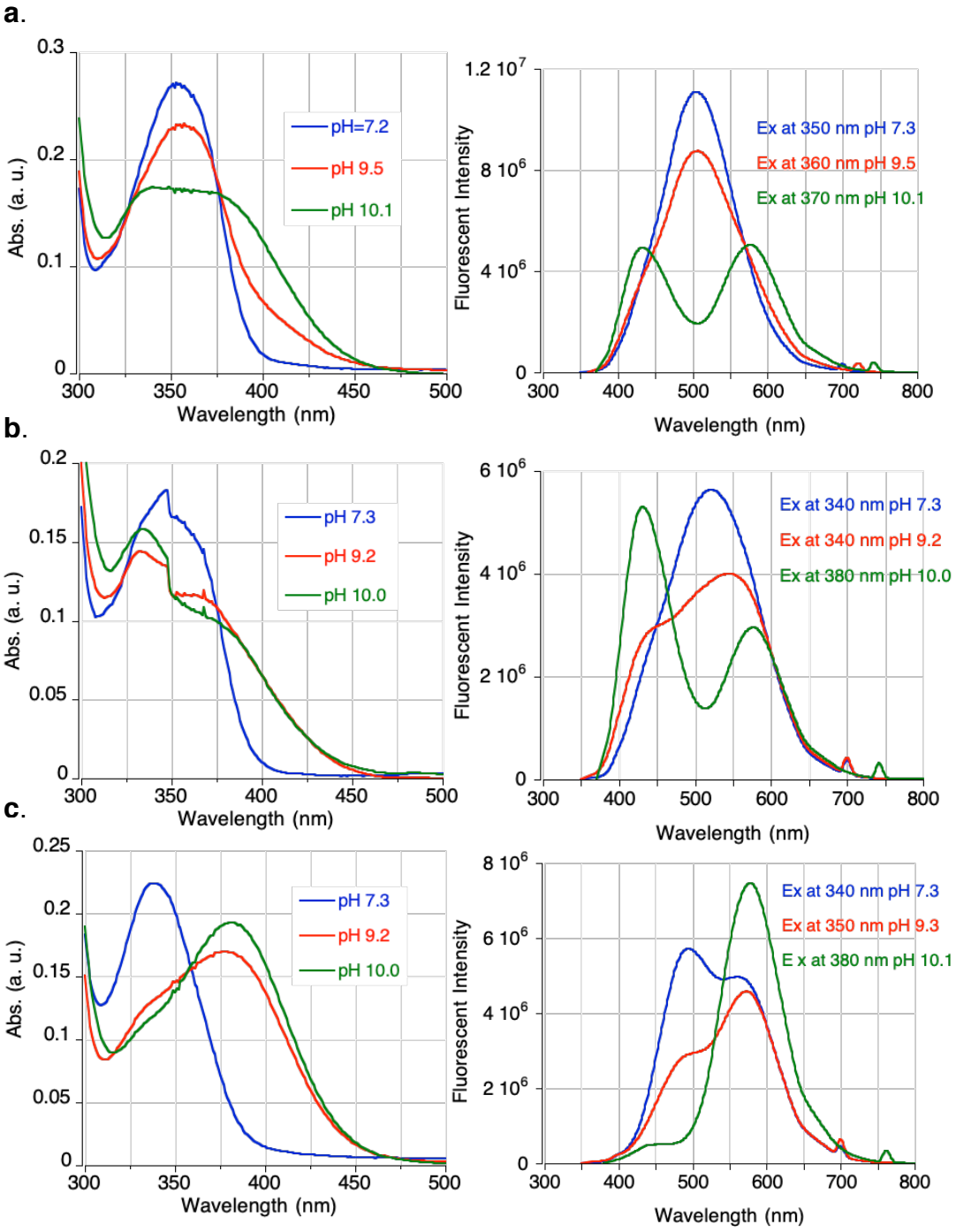
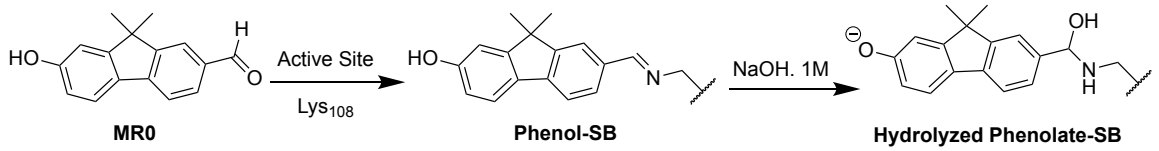


Figure III-22. Absorption (left) and emission (right) spectra of **M3** (**a**), **M4** (**b**), and **M5** (**c**) upon basification of their solutions.

Upon comparing each mutant's absorption and emission spectra at basic pH, we realized that the blue and red-shifted peaks in absorption and emission are in relation to each other. For example, for the **M3** complex, the absorption peaks at 336 and 371 nm have the same intensity, resulting in the broad absorption peak. Accordingly, the emission spectrum of **M3** shows two equal emission peaks at 430 and 576 nm (**Figure III-22a**). On the other hand, the blue-shifted absorption has a higher intensity for the **M4** complex, resulting in higher intensity for emission peak at 430 nm (**Figure III-22b**).



MR0 **Phenol-SB** **Hydrolyzed Phenolate-SB**

λ_{abs} 339nm
λ_{em} 492 nm

λ_{abs} 334nm
λ_{em} 430 nms

Figure III-23. Formation of the hydrolyzed Phenolate-SB complex produced upon basification of the protein solution.

Prior solution studies showed that none of the **MR0** species absorb as blue-shifted as 336 nm. Additionally, emission at 430 and 575 nm at a high pH could not match any bound **MR0** derivatives. We speculated that the basification of the sample with aqueous sodium hydroxide (1M. NaOH) solution leads to the hydrolyzation of the imine bond. At the same time, in the basic environment at pH ~10, the hydroxyl group gets deprotonated. Thus, the blue-shifted absorption and emission peaks with maximum wavelength at about 330 and 430 nm correspond to the phenolate hydrolyzed Phenolate-SB complex (see **Figure III-23**). Precise

characterization of the products is required to verify this statement. Presumably, the red-shifted absorption and emission peak at about 370 and 576 nm corresponds to the deprotonated hydroxyl group of non-hydrolyzed imines, Phenolate-SB complex (**Figure III-23**).

As discussed above, the reason we were interested in the photophysical properties of the **MR0/M5** complex is that this protein mutant, Q108K:K40L:T51V:T53S:R58H:F16Y:A33H:L117C, is engineered to show a high iminium pK_a with **ThioPhenol** (8.8). However, **MR0** binds this mutant as pure SB with low iminium pK_a (<5). We sought to investigate whether **MR0** can bind hCRBPII as an iminium; hence, it was incubated with Q108K:K40L:T51V:T53S:R58H:F16Y:Y19W:A33H:L117E mutant, which exhibit high pK_a with both **ThioPhenol** and **ThioFluor** due to the L117E mutation.

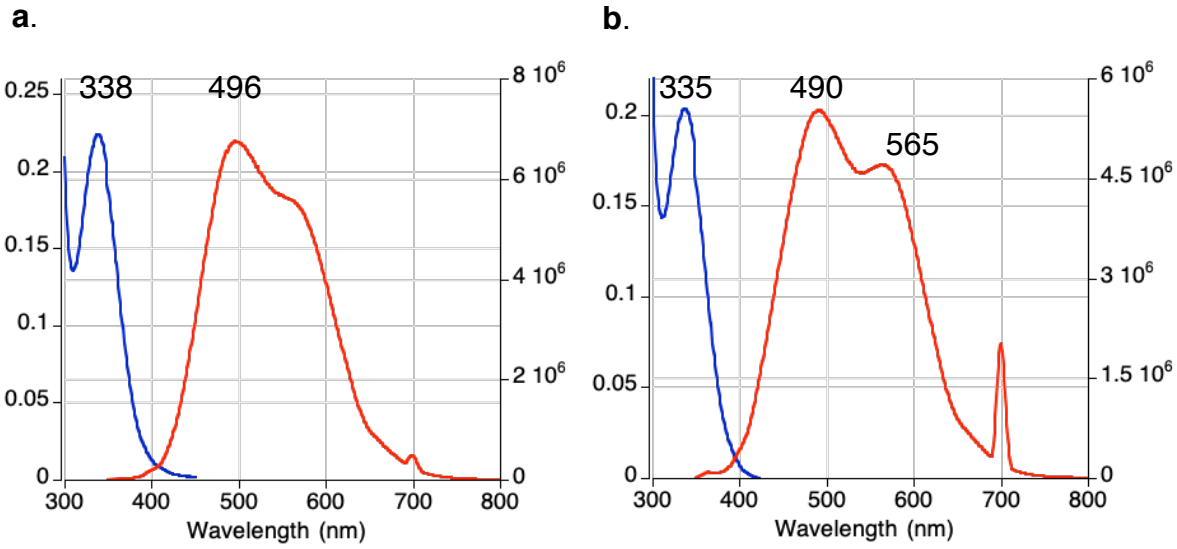


Figure III-24. Absorption (blue line) and emission (red line) of Q108K:K40L:T51V:T53S:R58H:F16Y:Y19W:A33H:L117E/**MR0** complex at neutral (left) and acidic (right) pH.

We measured the absorption and emission spectra of Q108K:K40L:T51V:T53S:R58H:F16Y:Y19W:A33H:L117E/**MR0** complex in both neutral and acidic pH (**Figure III-24**). Surprisingly, as shown, even acidification of the solution to pH less than five did not lead to **MR0** iminium formation. However, the red-shifted shoulder seems to be the PSB emission produced upon photoirradiation of SB (**Figure III-24b**). In general, it seems **MR0** cannot bind hCRBPII as an iminium, but it can generate PSB upon photoexcitation.

As described earlier, basification of the protein solutions showed that the **MR0** SB-Phenolate complex emits around 575 nm, but it does not help much to characterize the **MR0**/**M3** and **M4** complexes emissions at 505 nm and 525 nm, respectively. Presumably, the collected emission spectrum is not from a single species, and the broad shape of the peak is an indication of other species formed upon excited-state proton transfer to the imine or from the hydroxyl group (see **Figure III-22**). Nonetheless, it is clear that the emission of **MR0**/**M4** complex is not due to the double ESPT process, as solution studies suggested that the **MR0** Phenolate-PSB complex would emit beyond 600 nm.

In the next step, we sought to explore whether the introduction of proton-acceptor residues around the **MR0** hydroxyl group would facilitate the double ESPT process. Unfortunately, there are no crystal structures from **MR0** bound to hCRBPII. Thus, the most effective mutations in increasing the Φ_{ESPT} with **ThioPhenol** such as F16Y, T29Y, A33H, A33M, and L77M were introduced to **M4**,

the mutants yielded the highest Φ_{ESPT}, in hopes of improving the double ESPT process (**Table III-14**).

Table III-15. Spectroscopic properties change as the result of the addition of proton acceptor residues.

Entry	hCRBPII Mutant [a]	λ_{abs}	λ_{em}	SS (nm)	Φ_{DESPT}
1	KEAFF:R58H	347	525	178	0.02 >
2	KEAFF:R58H:**T29Y**	348	516	168	0.08
3	KEAFF:R58H:**L77M**	341	525	184	0.11
4	KEAFF:R58H:**F16Y**	329	560	231	0.53
5	KEAFF:R58H:**A33H**	334	512	178	0.07
6	KEAFF:R58H:**F16Y**:**A33H**	326	567	239	0.58
7	KEAFF:R58H:**Y19W**:**A33H**	345	498	159	0.43
8	KEAFF:R58H:**F16Y**:**T29Y**	329	562	233	0.57
9	KEAFF:R58H:**F16Y**:**L77M**	330	493/562	163/232	0.54
10	KEAFF:R58H:**F16Y**:**A33M**	329	435/565	106/236	0.61

[a] KEAFF equals to Q108K:K40E:T53A:Q38F:Q4F.

As listed above, the largest Stokes shifts (>230 nm) were acquired with mutants containing F16Y, entries 4 and 6, leading to the maximum emission wavelength around 565 nm. The addition of other mutations such as A33H, T29Y, and L77M either did not change or blue-shifted the emission wavelength compared to the **M4** complex (entry 1).

The Phenolate-PSB complex produced upon single ESPT of imine to iminium in ethanol emits around 630 nm with Stokes shift more than 250 nm (**Table III-10**). However, none of the tested protein mutants could result in such red-shifted emission. As was observed with **ThioPhenol**, the emission of Phenolate-PSB is

always blue-shifted for the double ESPT process as compared to its formation upon a single ESPT process (**Figure III-25**).

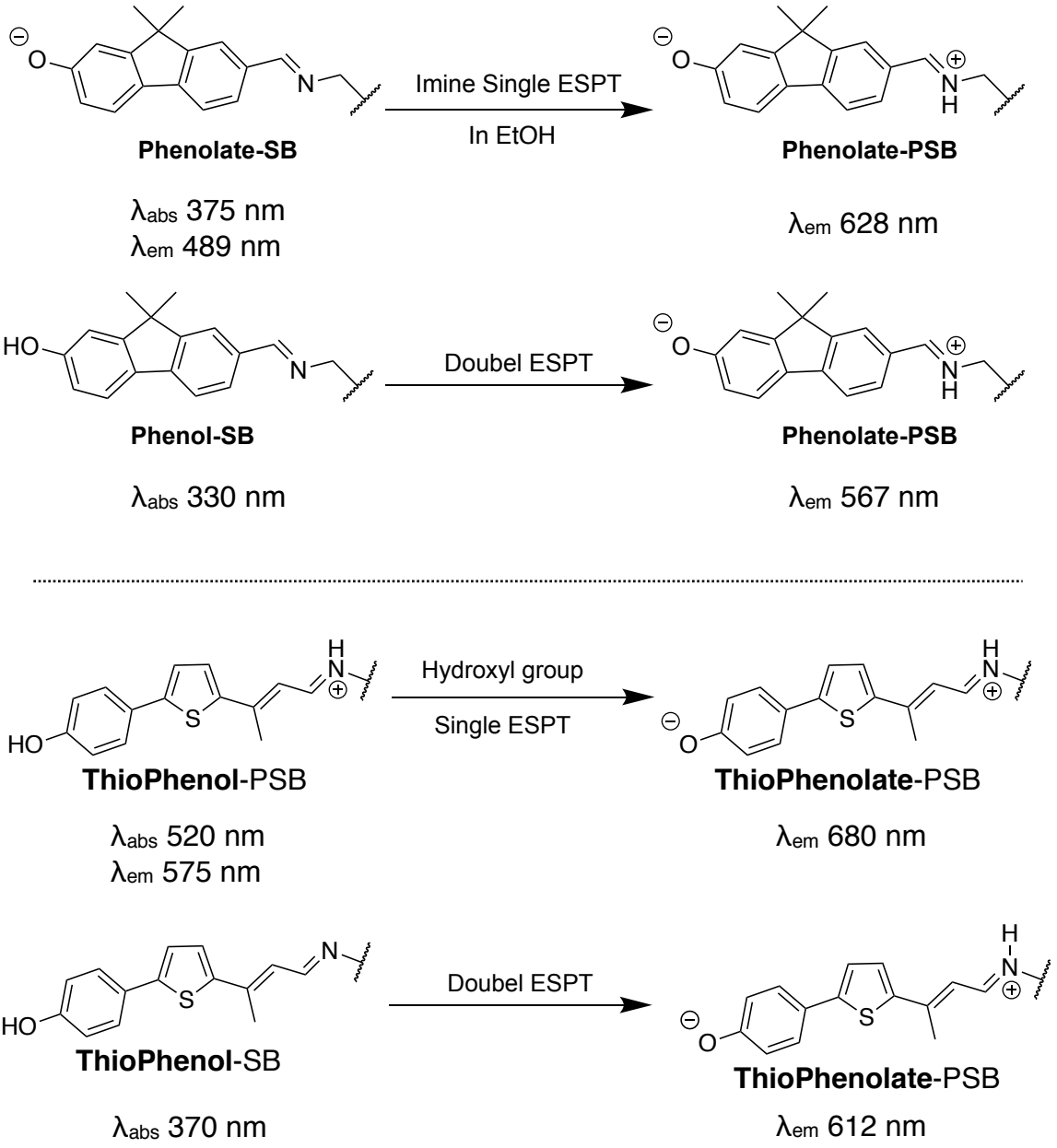


Figure III-25. The schematic representation of single and double ESPT processes and their photophysical properties.

The K40E mutation is retained in all double ESPT mutants as the proton source, and thus could putatively lead to the localization of the iminium positive charge and its blue-shifts in emission. Therefore, we surmised that the emission around 570 nm for entries 4 and 6 could be from the double ESPT process.

In contrast to double ESPT systems with **ThioPhenol** that can provide quite a narrow emission, all mutants tested with **MR0** produced a broad emission spectrum (**Figure III-26**). We attempted to narrow the emission spectrum by adding other proton acceptor residues besides F16Y; however, this was not fruitful (**Table III-6**, entries 8-10).

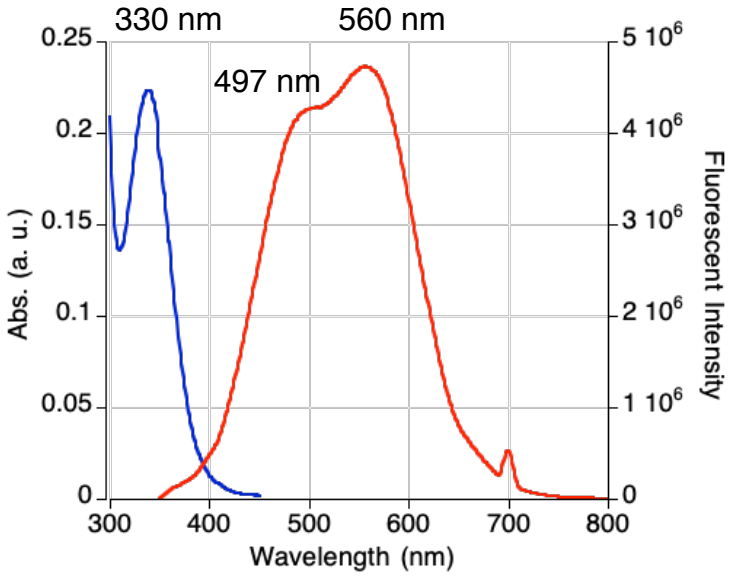


Figure III-26. Absorption and emission spectra of Q108K:K40E:T53A:Q38F:Q4F:R58H:F16Y/**MR0** complex. The blue-shifted shoulder corresponds to Phenol-PSB, or Phenolate-SB produced upon single ESPT, and the emission at 560 nm results from Phenolate-PSB complex produced upon the double ESPT process.

As discussed at the beginning of this section, one of our goals is to develop double ESPT systems with higher fluorescent quantum yield. **MR0** is a bright chromophore that binds designed hCRBPII mutants as a SB within 1 or 2 minutes. However, **MR0** complexes are not selective; photoirradiation of the SB results in other ESPT products besides Phenolate-PSB (the product of double ESPT), leading to a broad emission band. For instance, as shown in **Figure III-26**, excitation of Phenol-SB at 330 nm leads to a broad emission spectrum. Presumably, the left shoulder maximized at 497 nm results from imine to iminium or phenol to phenoxide single ESPT, and the peak at 560 nm is correlated to the Phenolate-PSB complex. More experiments are required to be able to characterize the emission precisely.

On the other hand, the live-cell applications are limited because, as discussed in **Section III.7.1.B** the **MR0** complexes are unstable and tend to hydrolyze easily in basic or acidic environments. Additionally, the absorption and emission wavelengths with **MR0** are blue-shifted, and the double ESPT Stokes shift is smaller as compared to **ThioPhenol**. As shown in **Table III-6**, all **MR0** SB complexes absorb in the UV region of the spectrum (less than 350 nm), which again can hamper its utility.

III.7.2 MR1

Next, we continued to pursue our goal through optimization of the **FR1** structure. Mr. Mehdi Moemeni successfully synthesized **MR1** wherein *N,N* diethyl nitrogen is replaced with a hydroxyl group (**Figure III-27**). **FR1**, a derivative of **FR0**,

was first designed and synthesized by Dr. Wei Sheng and exhibited the optimal extinction coefficient, fluorescence quantum yield, and proper absorption and emission wavelength.

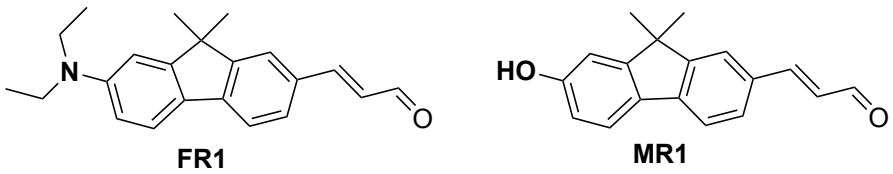


Figure III-27. Structures of **FR1** and **MR1**.

Dr. Wei Sheng successfully reported the first photoswitchable NIR tags that employ a synthetic dye, **FR1**, and a fusion protein. **FR1**/hCRBPII complexes showed great ability to photoactivably turn "ON" and turn "OFF" the fluorescence signal using an ICT-capable chromophore.[44]

III.7.2.A. Spectroscopic properties of MR1 in solution

To know about the spectroscopic characteristics of the newly designed chromophore, we measured its photophysical properties in various organic solvents with different polarities (**Table III-15**).

Table III-16. Spectroscopic characterization of **MR1** in various solvents.

Entry	Solvent	λ_{abs} (nm)	λ_{em} (nm)	SS (nm)	ε ($M^{-1}cm^{-1}$)	Φ [a]
1	Toluene	364	412	48	83,890	0.03
2	Ethyl acetate	357	398	41	86,611	0.04
3	Dimethyl sulfoxide	369	497	128	71,162	0.30
4	Ethanol	368	525	157	62,437	0.23
5	PBS buffer	362	562	200	51,230	0.03

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

As shown above, in **Table III-15**, the absorption wavelengths are close in different solvents centered around 364 nm. However, the emission wavelength shifts bathochromically in more polar solvents; it shifts more than 160 nm, from 398 nm in toluene to 562 nm in PBS buffer. **FR1** shows a strong solvatochromism in its fluorescence in various solvents (see **Figure III-28** and **Table III-15**). Such emission solvatochromism was expected since, like **FR1**, the polar excited state of **MR1** is more prone to solvent relaxations.

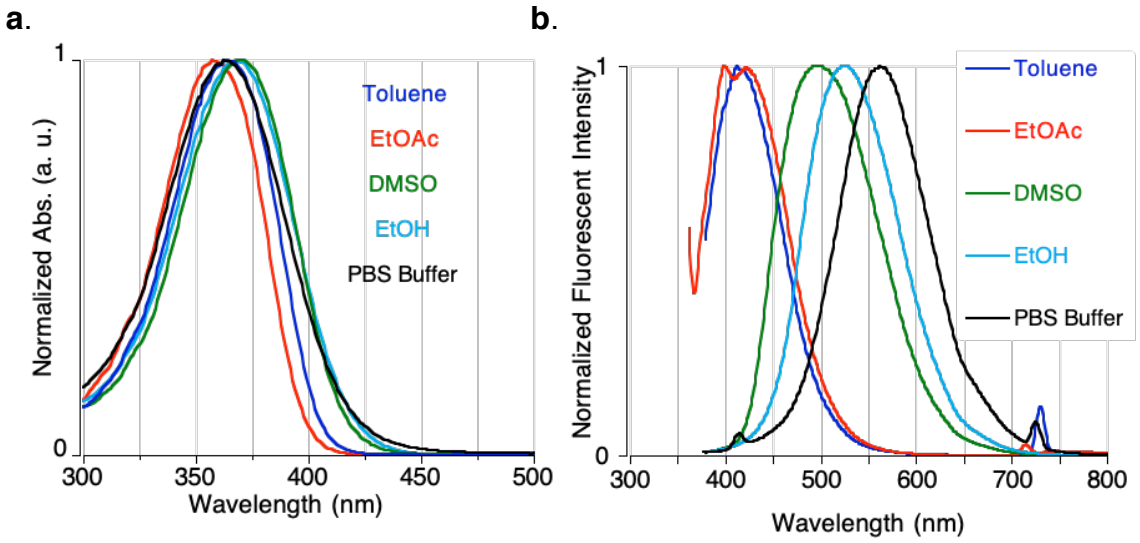


Figure III-28. Spectroscopic properties of **MR1** in different solvents. **a**. UV-Vis and **b**. Fluorescence spectra.

Interestingly, the fluorescence quantum efficiency of **MR1** in less polar solvents such as toluene and ethyl acetate is less than 5% (**Table III-15**, entries 1 and 2). Moreover, **MR1** is not quite soluble in aqueous solutions such as PBS buffer and shows a quantum yield of 3% (entry 5). These properties are advantageous for imaging applications as the free unbound **MR1** is not emissive in hydrophobic or aqueous parts of the cell milieu and thus, eliminates the need to

wash off free chromophore prior to imaging. Furthermore, in contrast to **FR1**, no deprotonation of the hydroxyl group was observed upon excitation in BPS buffer, leading to a less fluorescent background for fluorescent confocal imaging; since the corresponding alkoxide emission is red-shifted and might leak to the emission window of the designed fluorescent complex.

Comparing the photophysical properties of **ThioPhenol** with **MR1** in the same solvents indicates that for the latter, the average absorption and emission wavelengths have red-shifted 23 nm and 65 nm, respectively. In addition, the extinction coefficient and fluorescence quantum yield of **MR1** are significantly higher and lead to a 48-fold increase in the total brightness of the chromophore (**Table III-16**).

Table III-17. Comparison of spectroscopic features of free aldehyde **ThioPhenol** and **MR1** in ethanol.

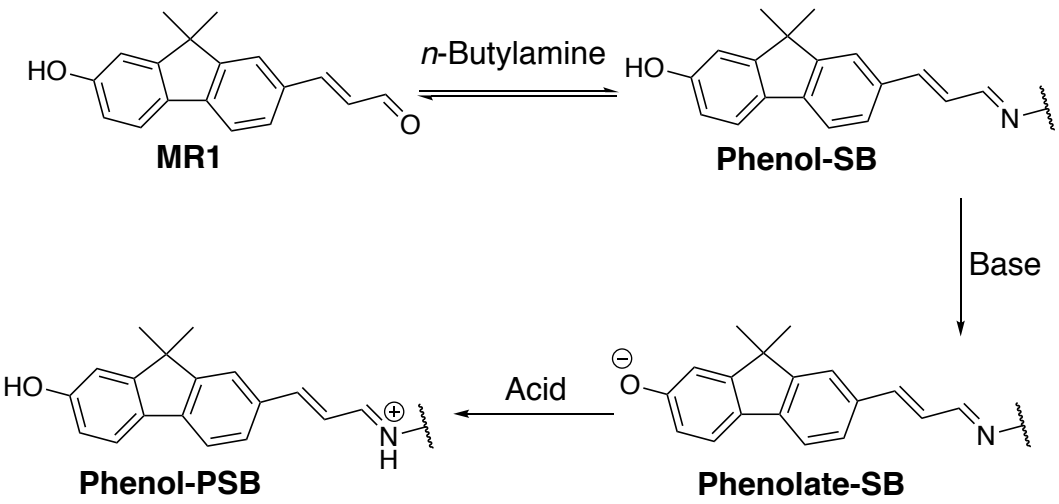
Chromophore	λ_{abs}	λ_{em}	SS (nm)	ε ($M^{-1}.cm^{-1}$)	Φ	Brightness (ε*Φ)
ThioPhenol	342	462	120	29,752	0.01	298
MR1	368	525	157	62,437	0.23	14,361

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

As for previous chromophores, it is essential to characterize the newly designed dye in solution before moving to the complex environment of the protein's

cavity. Hence, to mimic the product of **MR1** condensation in the protein pocket, the aldehyde was coupled with the *n*-butyl amine in ethanol (**Figure III-29**).

a.



b.

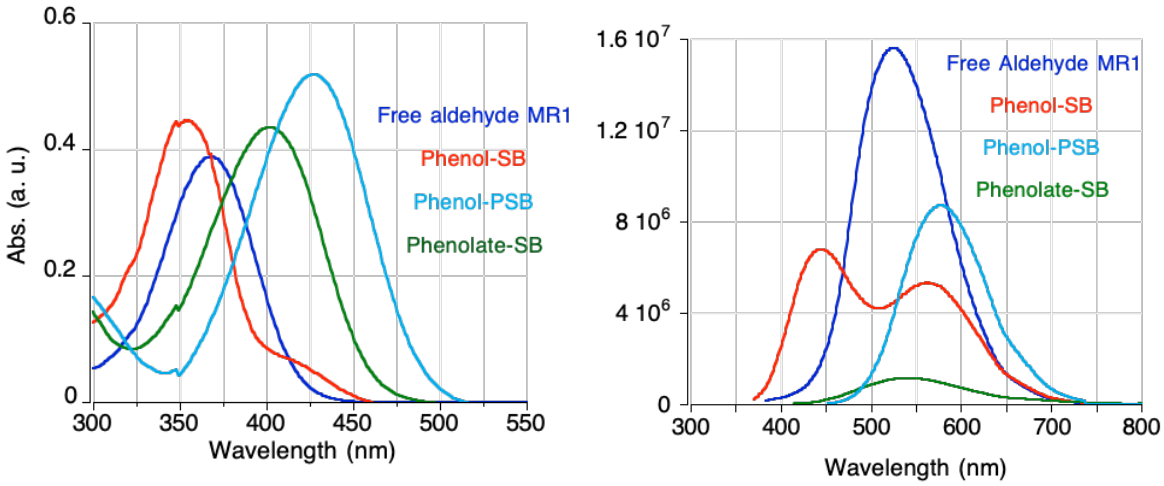


Figure III-29. **a**. Schiff base and protonated Schiff base (PSB) of **MR1** with *n*-butylamine in ethanol. **b**. Absorbance (left) and emission (right) spectra of **MR1** and derivatives: Phenol-SB, Phenol-PSB, and Phenolate-PSB.

As described at the beginning of this chapter (**Section III-1**), **FR0** possesses superior photobase activity. Its derivative **MR0** is a strong photobase as well that

even in a basic solution where the hydroxyl group is deprotonated, the corresponding imine can abstract a proton from ethanol the (see **Figure III-21** and **Table III-10**). It is clear from solution studies that **FR1** and **MR1** are weaker photoacids/bases. Excitation of the Phenol-SB activates the imine as a photobase and results in a two-maxima emission spectrum at 445 nm, and 564 nm, which corresponds to Phenol-SB and PSB, respectively (**Table III-17**, second entry). The emission of Phenol-PSB produce in the ground state upon acidification of the solution at 575 nm can verify this claim. However, in contrast to **MR0**, no imine ESPT was observed for Phenolate-SB (third and fourth entry).

Table III-18. Spectroscopic characterization of **MR1** and derivatives.

Compound	λ_{abs}	λ_{em}	SS (nm)	ε ($M^{-1}.cm^{-1}$)	Φ [a]	Φ_{ESPT}
Free Aldehyde **MR1**	368	525	157	62,437	0.33	-
Phenol-SB	355	445/564	90/209	71,727	0.16	43.9
Phenol-PSB	427	575	151	83,544	0.13	-
Phenolate-SB	401	541	140	70,117	0.04	< 2

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

III.7.2.B. Exploration about photoacidic and photobasic properties of MR1 complexes

In the first step, we sought to investigate whether **MR1** can act as a photobase in the protein environment. Thus, we measured its photophysical properties upon binding with the **M3** mutant, which is optimized as the best photobase in complexation with **ThioFluor** and **FR1** chromophores (**Table III-18**).

Monitoring the UV-Vis spectra indicates that the binding is fast, as the **MR1** Phenol-SB complex forms within 1-2 minutes. In addition, the imine pK_a is less than 5 (slightly lower than the previous chromophore), resulting in pure SB formation in neutral pH and a narrow Phenol-PSB emission (**Figure III-30**). Nonetheless, the fluorescence quantum yield with **MR1** has decreased to 26% due to the weak electron-donating effects of the hydroxyl group and failure to form an ICT system upon photoexcitation.

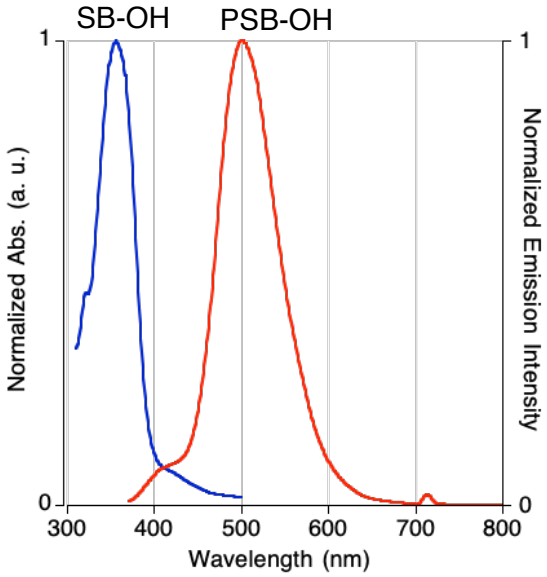


Figure III-30. The normalized absorption (blue line) and emission (red line) spectra of **MR1**/**M3** complex.

Table III-19. Comparison of the spectroscopic properties of **FR1/M3** and **MR1**/**M3** complexes.

Entry	Complex	λ_{abs}	λ_{em}	SS (nm)	Φ [a]
1	**FR1**/**M3** [b]	392	595	203	0.72
2	**MR1**/**M3**	357	502	145	0.26

[a] Absolute quantum yield was measured on a Quantaurus-QY. [b] **M3**: Q108K:K40E:T53A:Q38F:Q4F:R58L.

As shown in **Section III.7.1.B**, **MR0** could not form PSB with hCRBPII mutants; even with mutants containing L117E, the iminium pK_a was less than 5. Therefore, in order to study the **MR1** iminium and explore if it can show photoacidic properties the same as **ThioPhenol**, we chose to measure the photophysical characteristic of **MR1** and Q108K:K40L:T51V:T53S:**R58H**, the mutant that showed Φ_{ESPT} of 51% with **ThioPhenol**. Additionally, as a control experiment, **MR1** was incubated with Q108K:K40L:T51V:T53S mutant, which lacks the essential mutation (R58H) for deprotonation of the hydroxyl group (**Table III-19**).

The imine pK_a is low with both mutants resulting in a small amount of minimum formation in the ground state. The Q108K:K40L:T51V:T53S:R58H/**MR1** complex resulted in a rather broad emission spectrum with a maximum at 461 nm, which most probably is due to the Phenol-SB emission. On the other hand, unexpectedly, the Q108K:K40L:T51V:T53S/**MR1** complex behaved as a photobase since its emission wavelength is the same as **M3**/**MR1** complex with Stokes shift of more than 200 nm (**Table III-19**, entry 1).

Table III-20. Spectroscopic changes as the result of addition of R58H mutation.

Entry	hCRBPII mutant	λ_{abs}	λ_{em}	SS (nm)
1	Q108K:K40L:T51V:T53S	348	501	203
2	Q108K:K40L:T51V:T53S:**R58H**	345	461	145

This phenomenon was observed with **ThioPhenol** previously as well. As described in detail in **Section III-3** and **Table III-2**, **ThioPhenol** can appear as a

photobase even when K40E mutation is substituted with K40L in **M3** mutant. Nonetheless, the Q108K:K40L:T51V:T53S:**R58H/MR** complex was acidified to investigate if the PSB can act as a photoacid.

Interestingly, excitation at the PSB absorption wavelength (500 nm) results in a two-maxima emission spectrum at 602 nm and 692 nm. The latter red-shifted emission corresponds to the Phenolate-PSB complex product of the photoacid process formed upon ESPT of the hydroxyl group (**Figure III-31**).

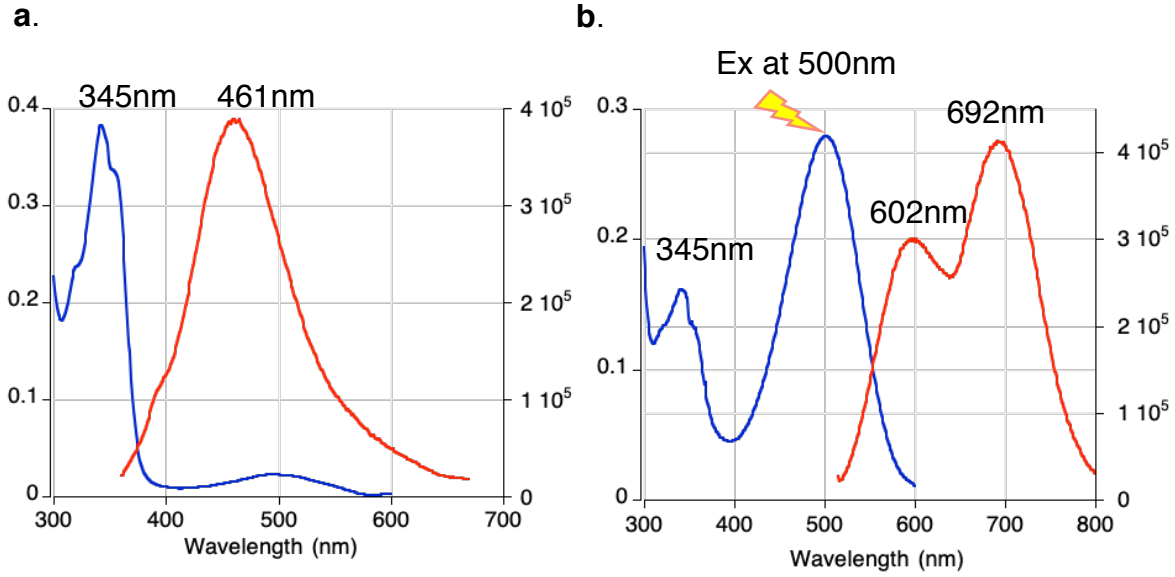


Figure III-31. The absorption and emission spectra of Q108K:K40L:T51V:T53S:R58H/**MR1** complex **a**. upon SB excitation in neutral pH 7.2 and **b**. PSB excitation in acidic pH 5.4.

Notably, same as **FR1** complexes, Q108K:K40L:T51V:T53S:R58H/**MR1** represents photoswitchable properties. Upon binding, the complex shows a maximum absorption wavelength at 345 nm with a minor peak at 600 nm. A 30-second UV irradiation (using a ~365 nm handheld UV lamp or a Xenon lamp

equipped with BP 300-400 band filter) can switch the thermal "OFF" state to a kinetic "ON" state that has a predominant absorption at 500 nm (**Figure III-32**). Excitation of the "OFF" state at 345 nm results in an intense 461-nm blue emission, with a tail around 600 to 680 nm. "ON" state excitation at 500 nm results in the same emission obtained upon acidification of the solution (see **Figure III-31**).

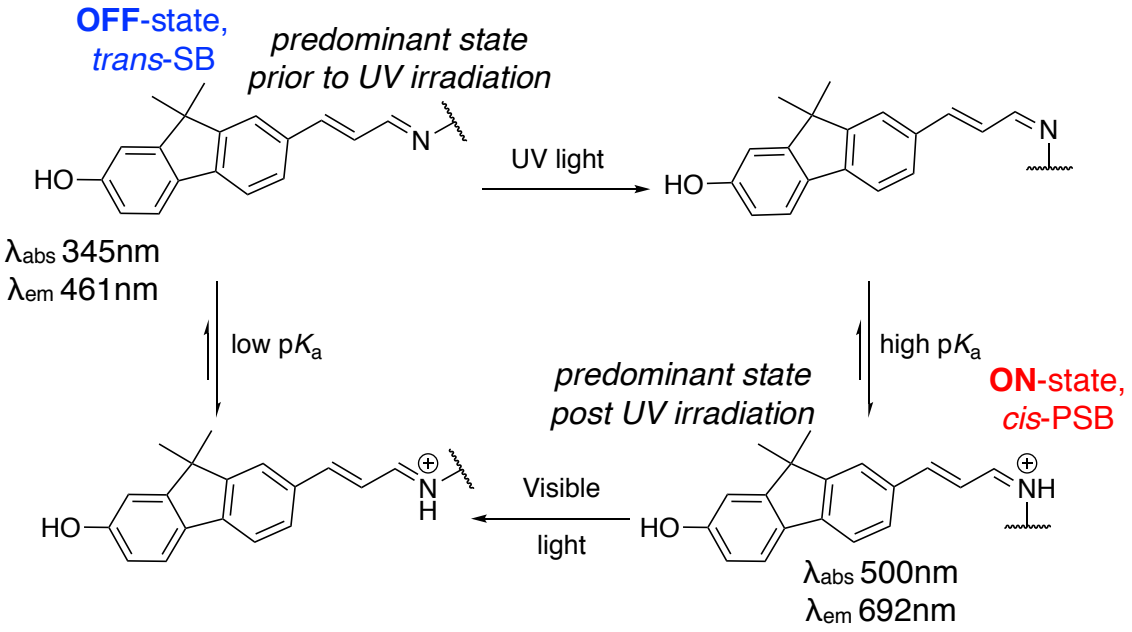


Figure III-32. Proposed photoswitching cycle of Q108K:K40L:T51V:T53S:R58H/**MR1** complex.

Studies of the photoswitchable properties of **MR1** were not pursued any further as this section aims to investigate whether it is possible to develop the double ESPT process with **MR1** complexes. Besides, Dr. Wei Sheng has studied and described this phenomenon with **FR1** chromophore previously. [44]

Experiments with Q108K:K40L:T51V:T53S:R58H/**MR1** complex indicates that **MR1** Phenol-PSB can show photoacidic properties; however, only in an acidic

solution, where a decent amount of PSB is present. We attempted to explore if **MR1** can appear as a photoacid in neutral pH or form stable PSB in the ground state. To this end, two protein mutants that show relatively high Φ_{ESPT} and iminium pK_a with **ThioPhenol** were chosen to test with **MR1** (**Table III-20**).

Table III-21. Spectroscopic properties of **MR1** with high iminium and mutants

Entry	hCRBPII mutant [a]	λ_{abs}	λ_{em}	SS (nm)	pK_a
1	KLVS:R58H:A33H:F16Y:L117C	363	519	156	5.2
2	KLVS:R58H:Y19W:A33H:F16Y:T29Y:L117C	368	550	182	< 5

[a] KLVS is equal to Q108K:K40L:T51V:T53S.

As listed above, binding to both mutants gives low iminium pK_a, and as a result, Phenol-SB is the dominant formed species in neutral pH. There is about a 10 nm red-shift in the SB absorption wavelength compared to the **MR1**/**M3** complex. Nonetheless, excitation of the SB in Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:T29Y:L117C/**MR1** complex gives a broad emission maximized at 550 nm, which is too red-shifted to be SB emission and thus, presumably is the emission of SB-Phenolate complex. However, the emission spectrum is broad enough for both mutants suggesting other species besides the Phenolate-SB. To confirm if the hydroxyl group becomes deprotonated upon excitation of the Phenol-SB, we decided to monitor the absorption and emission spectra of the same complex upon acid and base titration of its solution (results are summarized in **Table III-21**).

The absorption and emission wavelengths of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y:T29Y:L117C/**MR1** in acidic and basic pH are listed below. Due to the low iminium pK_a, acidic pH (5.1) a tiny fraction of the PSB peak was observed at 465 nm, which its excitation leads to ESPT of the hydroxyl group and Phenolate-PSB formation. Basifying the protein solution to pH more than 10 deprotonates the hydroxyl group in the ground state and gives the Phenolate-SB complex. Interestingly, the emission of this complex at 548 nm is close to the emission wavelength in neutral pH, which verifies that **MR1** complexes can appear as a strong photoacid however the iminium pK_a should increase.

Table III-22. Spectroscopic properties of **MR1** with high iminium and mutants.

Acidic pH 5.1			Neutral pH 7.3			Basic pH 10.7		
λ_{abs}	λ_{em}	SS	λ_{abs}	λ_{em}	SS	λ_{abs}	λ_{em}	SS
465	571/677	106/212	368	550	182	410	548	138

In the next step, we sought to investigate whether it is possible to enhance the photoacidic properties of **MR1** complexes, whether through increasing the iminium pK_a or inserting more proton acceptor residues around the hydroxyl group. **MR1** is a much brighter dye than **ThioPhenol**, which emits in the far-red NIR region of the spectrum and can find deep tissue fluorescent imaging applications.

To identify the closest residues to the hydroxyl group of **MR1**, we docked the energy-optimized **MR1** structure into the crystal structure of

Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFlour** complex (**Figure III-33**). The docking simulation was done using the pair fitting function in Pymol, and it indicates that Q38F and L77 (if the **MR1** structure is flipped) are the closest to the **MR1** hydroxyl group (**Figure III-33**). The protein mutants obtained upon mutation of these residues and their properties are listed in **Table III-22**.

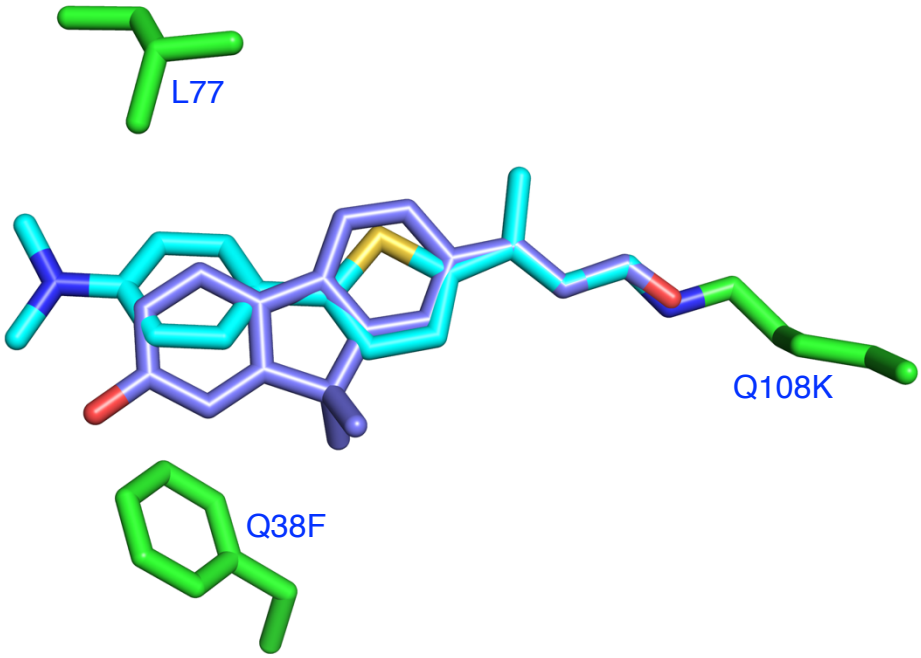


Figure III-33. Flexible docking of **MR1** in the crystal structure of Q108K:K40L:T51V:T53S:R58W-hCRBPII/**ThioFluor**. **MR1** is shown in purple, and **ThioFluor** is shown in cyan.

As described in Chapter II in Sections **II.11.1C** and **II.11.3A**, tyrosine works best at positions 77 and 38 for protein expression yield and deprotonation of the hydroxyl group of bound **ThioPhenol**. Hence, we sought to investigate the effect of these mutations on the photoacidic properties of **MR1**. Comparing the emission wavelengths of entries 1 and 2 indicates that Q38Y mutation can effectively activate the photoacidity of the bound **MR1**. With this mutation, the emission

wavelength is at 500 nm, which corresponds to the Phenolate-SB complex; however, the emission wavelength with L77Y at 471 nm represents the Phenol-SB complex. Nonetheless, both mutants show a low iminium pK_a, and no PSB formation is observed in neutral pH.

Table III-23. Spectroscopic change as the result of L77Y and Q38Y mutations.

Entry	hCRBPII mutant [a]	λ_{abs}	λ_{em}	SS (nm)
1	KLVS:R58H:A33H:**Q38Y**	350	500	150
2	KLVS:R58H:F16Y:**L77Y**	345	471	126
3	KLVS:R58H:Y19W:A33H:F16Y: L117C:**L77Y**	364	493/595	129/231

[a] KLVS is equal to Q108K:K40L:T51V:T53S.

In studies with **ThioPhenol**, we realized that L117C mutation usually increases the iminium pK_a, but notably, for all the same protein mutants, **MR1** complexes show lower iminium pK_a than **ThioPhenol**. The iminium pK_a in Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y: L117C:L77Y/**MR1** is 5.4 and gives a slightly higher amount of PSB in neutral pH. In order to better characterize this system, the emission spectrum of this complex was collected upon excitation of the SB and PSB (**Figure III-34**). An interesting observation was made during this study; excitation of Phenol-SB results in a broad two-maxima emission spectrum at 493 nm and 595 nm. As discussed above with other protein mutants, the blue-shifted emission results from the hydroxyl group deprotonation in the excited state. The more intense hypsochromic shifted emission is presumably due to the double ESPT process that leads to more than 230 nm Stokes shift, the

largest Stokes shift of **MR1**/hCRBPII complexes up to this point. Surprisingly, this mutant is not equipped for proton transfer to imine (K40E); however, such a large Stokes shift verifies the occurrence of the double ESPT process. Furthermore, we surmised that the collected emission acquired upon excitation of the PSB-Phenol complex at 470 corresponds to the Phenolate-PSB complex (**Figure III-34a**). However, as discussed earlier, the emission of the Phenolate-PSB complex is more red-shifted when it is the product of the single ESPT process than the product of double ESPT, 661 nm, and 595 nm, respectively, for the current system.

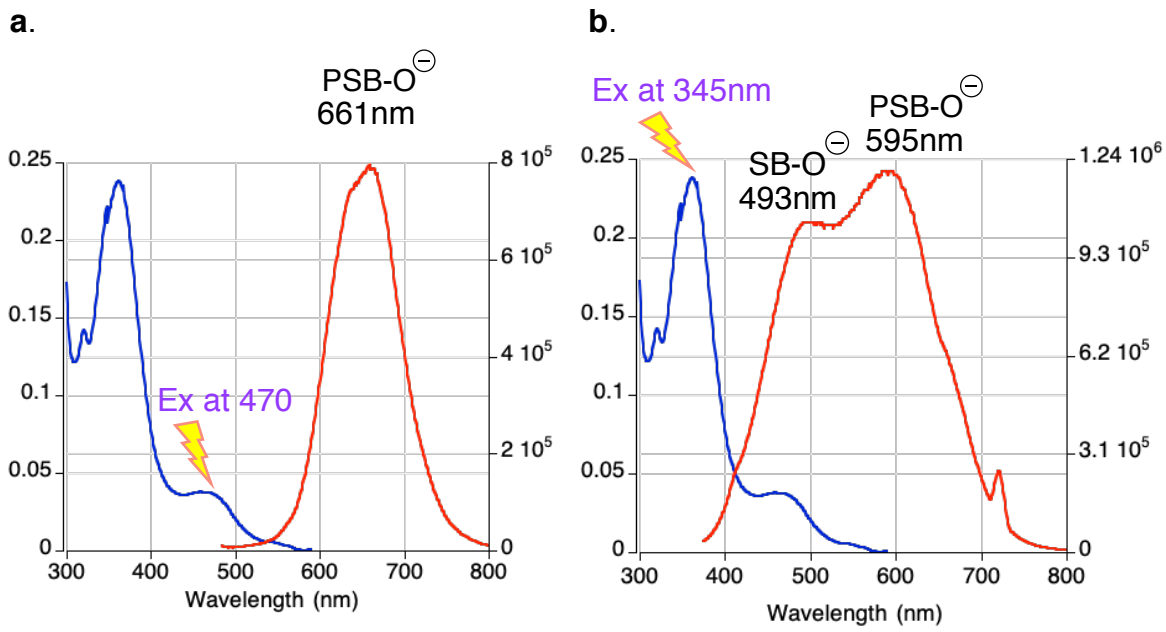


Figure III-34. Absorption (blue line) and emission (red line) spectra of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y: L117C:L77Y/**MR1** complex upon excitation of **a**. SB at 364 nm and **b**. PSB at 470 nm in neutral pH.

We monitored the absorption and emission spectra of this complex over acid and base titrations as well. When measured in a basic solution (the dark blue line) with the neutral solution, the comparison of the Phenol-SB emission confirms that

the emission peak maximized at 493 nm is correlated to the deprotonated hydroxyl. The emission spectrum measured in the acidic pH (4.8 the purple line) is quite broad and could indicate the exitance of more than one species. Presumably, all Phenol-PSB, Phenolate-SB, and Phenolate-PSB complexes contribute to this emission (**Figure III-35**).

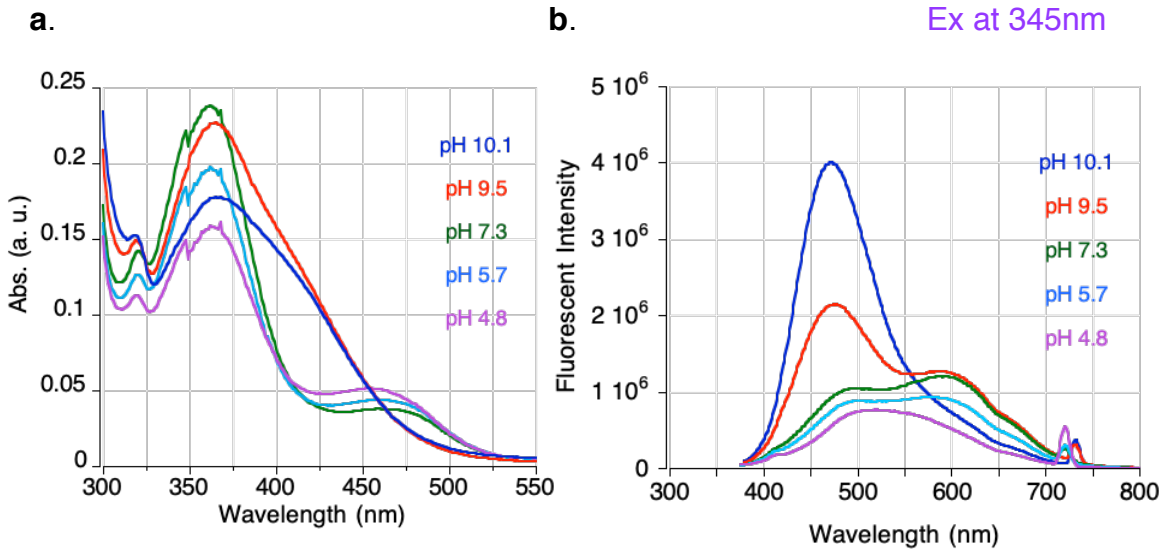


Figure III-35. Absorption (left) and emission (right) spectra of Q108K:K40L:T51V:T53S:R58H:Y19W:A33H:F16Y: L117C:L77Y/**MR1** upon acid and base titrations.

Up to this point, all mutants which showed high iminium pK_a with **ThioPhenol** led to a residual PSB formation with **MR1** in neutral pH. However, previously, Dr. Wei Sheng demonstrated that **FR1** binds Q108K:**K40H**:T53A:R58L:Q38F:Q4F (**M7**) mutant mainly as an iminium.[45] Hence, we sought to investigate the photophysical properties of the **MR1**/**M7** complex and explore if the excitation of the PSB can result in the deprotonation of the hydroxyl group and photoacidity. The iminium pK_a of this complex is 6.9, which is the highest

pK_a measured for **MR1** complexes, leading to almost equal SB and PSB formation. The comparison of **MR1**/**M7** absorption and emission wavelengths with **FR1** is shown in **Table III-23**.

Table III-24. Comparison of the spectroscopic properties of **FR1/M7** and **MR1**/**M7** complexes.

Entry	Complex	SB λ_{abs}	SB λ_{em}	SS (nm)	PSB λ_{abs}	PSB λ_{em}	SS (nm)	pK_a
1	**FR1**/**M7**[a]	398	450/630	52/232	516	630	114	> 7
2	**MR1**/**M7**	355	410	55	446	539	93	6.9

[a] **M7**:Q108K:**K40H**:T53A:R58L:Q38F:Q4F.

As shown above, in contrast to **FR1**, binding to **M7** does not lead to imine to iminium ESPT. Additionally, excitation of the PSB results in a single emission of the PSB, and no hydroxyl group photoacidity was observed.

We presume that **MR1** can function as a strong photoacid; however, its iminium pK_a with the same mutants is generally lower than **ThioPhenol**, which results in primary SB and residual PSB formation. Clearly, the emission due to photoacid activity of Phenol-SB complexes is blue-shifted in comparison with Phenol-PSB photoacids, simply because the former absorbs at about 350 nm vs. PSB absorption in the range of 470-500 nm. Therefore, future studies could be focused on enhancing the iminium pK_a of **MR1**-hCRBPII complexes to benefit from its potential as a bright and far-red emitting fluorescent tag for deep tissue imaging applications.

III.7.2.C. Developing protein-based double ESPT systems with MR1

Thus far in this section, we have discussed the photobasic and photoacidic properties of **MR1** complexes. The following describes our attempts to develop the double ESPT system with **MR1** complexes.

In the first step, we measured the physical characteristics of the **MR1**/**M4** complex. **M4** (Q108K:K40E:T53A:Q38F:Q4F:R58H) mutant has yielded the highest Φ_{DESPT} with **ThioPhenol**, and we sought to investigate whether the designed mutant can activate the ESPT process on both ends of bound **MR1** (imine and the hydroxyl group) as well (**Table III-24**).

Table III-25. Comparison of spectroscopic properties of **MR1** in complexation with **M3** and **M4** mutants.

Entry	hCRBPII mutant	λ_{abs}	λ_{em}	SS (nm)	Φ [a]
1	Q108K:K40E:T53A:Q38F:Q4F:**R58L** (**M3**)	357	502	145	0.26
2	Q108K:K40E:T53A:Q38F:Q4F:**R58H** (**M4**)	357	503	146	014

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

As shown above, the addition of R58H mutation does not change the maximum emission wavelength; however, in contrast to the narrow emission spectrum with **M3** mutant (entry 1, see **Figure III-30**), **MR1**/**M4** complex yields a wide emission spanning 370 nm to 780 nm (**Figure III-36**). Such a broad emission spectrum indicates that **MR1**/**M4** complex mainly acts as a photobase, with minimum photoacid activity of the hydroxyl group upon excitation.

Apparently, contrary to **ThioPhenol**, the addition of R58H alone cannot deprotonate the hydroxyl group in the excited state. Thus, we attempted to mutate amino acids surrounding the **MR1** hydroxyl group to proton acceptor residues.

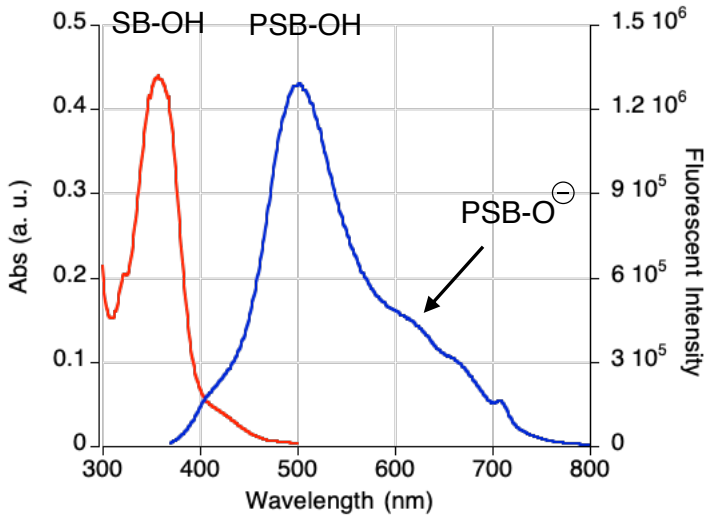


Figure III-36. The absorption and emission spectra of Q108K:K40E:T53A:Q38F:Q4F:**R58H/MR1** complex measured at neutral pH.

As is illustrated in **Figure III-38**, F16 is one the closest residues to the hydroxyl group, 3.7 Å, and studies with **MR0** showed that its mutation to tyrosine could substantially increase the Φ_{DESPT} (see **Table III-4**). Besides, Q38Y mutation enhanced the photoacidic properties of **MR1** complexes (see **Table III-22**). Therefore, we sought to explore the effect of F16Y and Q38Y mutations on the Φ_{DESPT} (**Table III-25**).

As shown below, the addition of F16Y and its combination with Q38Y mutation improved the Φ_{DESPT} considerably; the small bump in the emission spectrum of **MR1**/**M4** has grown to a great extent (**Figure III-37**). For both mutants, the emission spectrum shows two maxima. The blue-shifted emission peak at 490 nm corresponds to the Phenolate-SB complex; the emission of the Phenol-PSB complex is usually more red-shifted, and as in **MR1**/**M3** complex, is slightly beyond 500 nm.

Table III-26. Spectroscopic change as the result of F77Y and Q38Y mutations.

Entry	hCRBPII mutant [a]	λ_{abs}	λ_{em}	SS (nm)	Φ_{DESPT}
1	KEAFF:R58H:**F16Y**	357	490/603	133/246	0.45
2	KEAFF:R58H:**F16Y**:**Q38Y**	355	489/618	134/**263**	0.54

[a] KEAFF equals to Q108K:K40E:T53A:Q38F:Q4F.

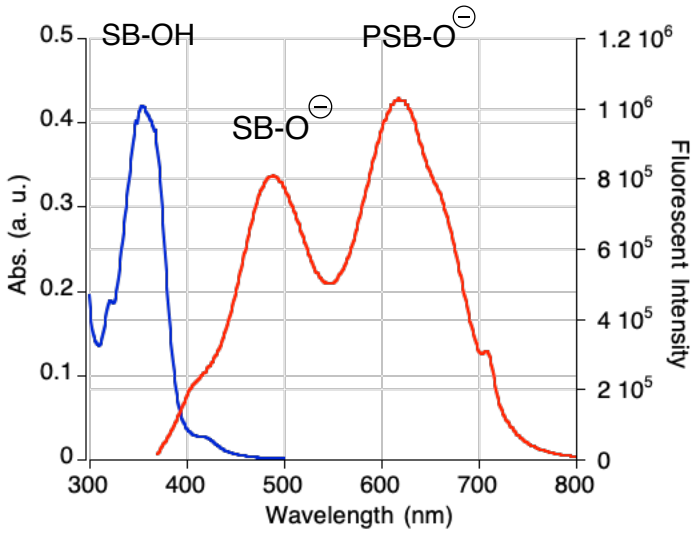


Figure III-37. The absorption and emission spectra of Q108K:K40E:T53A:Q38F:Q4F**:**R58H:F16Y:Q38Y/**MR1** complex measured at neutral pH.

Furthermore, the red-shifted emission peak is correlated with the double ESPT product, the **ThioPhenolate**-PSB complex. Noteworthy, the highest Φ_{DESPT} and the largest Stokes shift, 263 nm, was obtained upon the addition of both F16Y and Q38Y mutations to the **M4** mutant (entry 2).

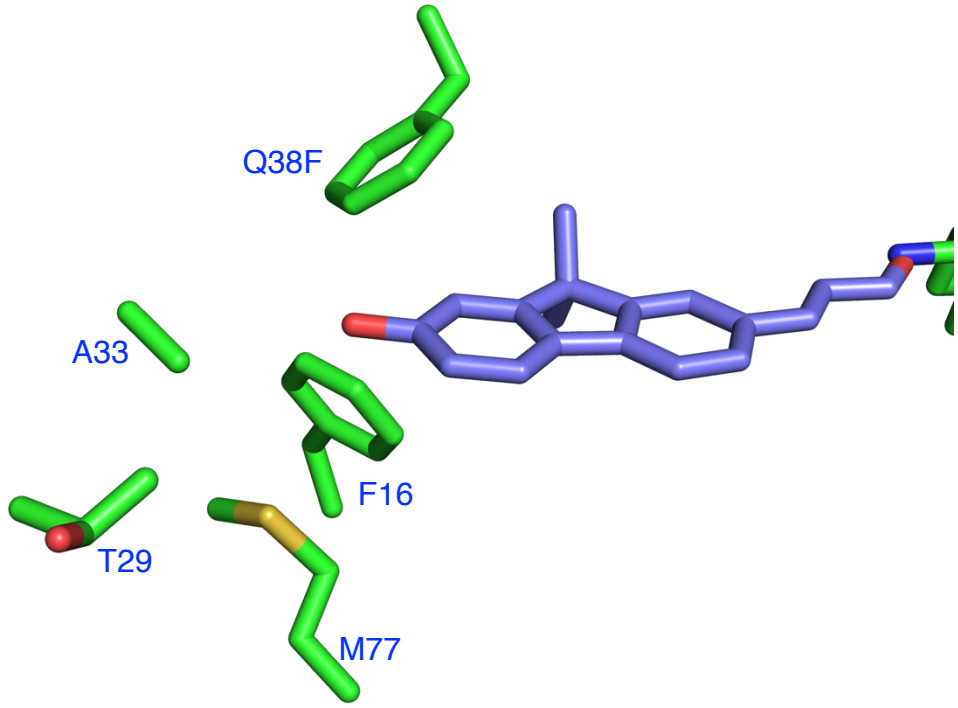


Figure III-38. Docked **MR1** in the crystal structure of Q108K:K40L:T51V:T53S:R58W-hCRBPII/**ThioFluor**, and the closest residues around the hydroxyl group.

Next, we sought to investigate whether it is feasible to suppress the Phenolate-SB or Phenol-PSB emission peaks through mutation of the surrounding residues. We envisioned developing a complex that emits only as the double ESPT process (**Table III-26**). Unfortunately, there is no crystal structure available from **MR1** complexes to learn about the exact distances between the hydroxyl and surrounding residues. However, studies with **ThioPhenol** showed a water-

mediated hydrogen bonding network between the T29 and the chromophore's hydroxyl group and that the introduction of T29Y mutation improved the Φ_{ESPT} for the resulted photoacid emission (see **Figure II-27** and **Table II-25**).

Table III-27. Spectroscopic change as the result of the mutation of residues surrounding the **MR1** hydroxyl group.

Entry	hCRBPII mutant [a]	λ_{abs}	λ_{em}	SS (nm)	Φ_{DESPT}
1	KEAFF:R58H:**T29Y**	357	501/616	144/259	0.41
2	KEAFF:R58H:**F16Y**:**T29Y**	358	493/608	135/250	0.53
3	KEAFF:R58H:**M20Y**	358	500	142	0.25
4	KEAFF:R58H:**F16Y**:**M20Y**	353	489	136	0.18
5	KEAFF:R58H:**F16Y**:**A33M**	355	492	137	0.21
6	KEAFF:R58H:**F16Y**:**T51M**	357	485/591	128/234	0.46
7	KEAFF:R58H:**F16Y**:**T29Y**:**Q38Y**	356	485	129	0.36

[a] KEAFF equals to Q108K:K40E:T53A:Q38F:Q4F.

The addition of M20Y and A33M in combination with F16Y reduced the Φ_{DESPT} dramatically and mainly resulted in Phenol-PSB or Phenolate-SB complexes emission in the range of 485 nm to 505 nm (entries 3, 4, and 5). As described in **Section III.5.1**, the introduction of T51M facilitates the deprotonation of the hydroxyl group and increases the Φ_{DESPT} with **ThioPhenol**. The same effect was observed with **MR1** complexes, as the addition of T51M mutation to Q108K:K40E:T53A:Q38F:Q4F:R58H:F16Y template enhanced the double ESPT quantum yield to more than 45% (entry 6). However, the highest Φ_{DESPT} in this table was obtained with the addition of both F16Y and T29Y mutations (entry 2).

Nonetheless, none of the newly designed mutants could exceed the Φ_{DESPT} acquired with F16Y:Q38Y. Thus, we chose to investigate if the addition of F16Y, Q38Y, and T29Y mutations altogether would further facilitate the deprotonation of the hydroxyl group and enhance the Φ_{DESPT} accordingly (entry 7). The resultant mutant, KEAFF:R58H:F16Y:T29Y:Q38Y, showed lower Φ_{DESPT}. Presumably, the presence of three close tyrosine residues leads to a steric clash, and tyrosine residues turn away from the cavity and further from the hydroxyl group to relieve the steric hindrance.

To summarize our research about developing the double ESPT system with **MR1** complexes, we noted that the introduction of proton acceptor residues such as tyrosine at positions F16, Q38, or T29 could enhance the Φ_{DESPT} to more than 50%. However, as described earlier in this section, the double ESPT emission with **MR1** complexes is associated with the emission of single ESPT of imine or the hydroxyl group products, Phenol-PSB and Phenolate-SB, respectively. Our attempts to suppress the blue-shifted emission of unwanted excited-state proton transfer were not fruitful up to this point with the handful of protein mutants we tried for this study.

As solution studies showed, the free aldehyde and Phenol-SB of **MR1** have absorption around 360 nm, close to the Phenol-SB complex with protein, and their emission tails reaching past 600 nm (see **Section III.7.2.A**). Therefore, controlling the ESPT process selectivity through engineering the hCRBPII mutants would result in a narrow, red-shifted double ESPT emission essential to prevent the

emission from free **MR1** or unspecific imine or iminium bindings. The above results suggest that **MR1** has a great potential for fluorescent imaging as fusion protein tag.

III.8 ThioPhenol structure modification and its application in multicolor imaging

Our studies indicate that **ThioPhenol** is the most selective chromophore for the ESPT processes; its complexes can act as a photobase, photoacid, or both upon single excitation, depending on the protein mutant. However, its brightness needs to be improved compared to commercially available dyes such as BODIPY, cyanines, and rhodamines, to name a few.

Basically, aside from fluorescence and phosphorescence, non-radiative processes such as internal conversion and intersystem crossing are the primary mechanisms responsible for the excited state relaxation. Most current theories suggest that a restriction in the rotational freedom of a fluorophore will lead to an increase in quantum yield and, consequently, fluorescence brightness.

Hence, we sought to investigate whether rigidifying the **ThioPhenol** structure by preventing the rotation around the carbon-carbon single bond would enhance its brightness. Mr. Mehdi Moemeni successfully designed a synthetic path to cyclize and rigidify the **ThioPhenol** structure (**Figure III-39a**). The synthetic procedure of newly designed chromophore Cyclized **ThioPhenol** (**CyThioPhenol**) and all other chromophore structures discussed in this thesis are described in Section **IV.7**.

III.8.1 Spectroscopic properties of CyThioPhenol in solution

CyThioPhenol is an unknown chromophore structure, and same as previous chromophore structures discussed, it is essential to characterize **CyThioPhenol** in solution before moving into the protein environment (**Figure III-39b).**

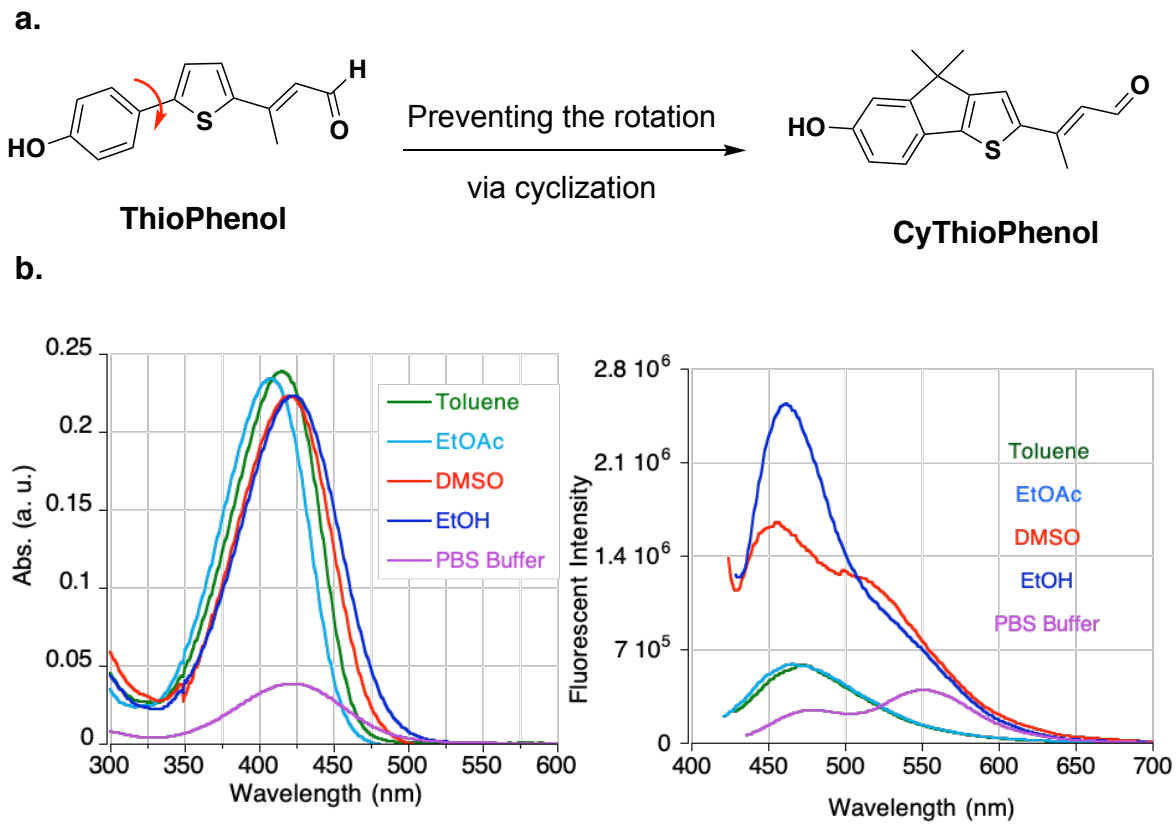


Figure III-39. **a**. Structures of **ThioPhenol** and **CyThioPhenol**. **b**. Absorbance (left) and emission (right) spectra of **CyThioPhenol** in different solvents.

The absorption wavelength of the cyclized structure shows ~30 nm bathochromic shift as compared to **ThioPhenol**, changing in a small range of 16 nm (from 408 nm to 424 nm). In addition, the molar extinction coefficient of

CyThioPhenol is larger in all solvents (see **Table II-1** and **Table III-27**). On the other hand, the average emission wavelength is blue-shifted by about 13 nm upon cyclization of the **ThioPhenol** structure, leading to smaller Stokes shifts. Furthermore, the fluorescence quantum efficiency has increased from 1 to 5% in ethanol and DMSO. Noteworthy, this structure shows low solubility in aqueous solutions resulting in aggravation and less intense absorption and emission peak. However, in contrast to **ThioPhenol**, the emission spectrum in PBS buffer shows two maxima presumably the red-shifted one at 550 nm corresponding to the deprotonated hydroxyl group in the excited state (**Table III-27**, entry 5).

Table III-28. Spectroscopic characterization of **CyThioPhenol** in various solvents.

Entry	Solvent	λ_{abs} (nm)	λ_{em} (nm)	SS (nm)	ε ($M^{-1}.cm^{-1}$)	Φ [a]
1	Toluene	416	473	57	42,482	0.01
2	Ethyl acetate	408	466	58	41,594	0.01
3	Dimethyl sulfoxide	421	455	34	39,620	0.05
4	Ethanol	424	462	38	39,619	0.05
5	PBS buffer	423	479/550	56/127	6,935	0.015

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

Evidently, **CyThioPhenol** is not a solvatochromic chromophore; there is no correlation between the emission wavelength and polarity of the solvents, and the emission wavelength spans a narrow range of 24 nm (from 455 nm to 479 nm if the hydroxyl group ESPT is not considered).

To mimic the binding in the protein, **CyThioPhenol** was coupled with *n*-butyl amine in ethanol. The sample was acidified and basified to characterize the corresponding PSB and deprotonated hydroxyl group in the ground state (**Table III-28**, and **Figure III-40**).

Table III-29. Spectroscopic characterization of **CyThioPhenol** and derivatives.

Entry	Compound	λ_{abs} (nm)	λ_{em} (nm)	SS (nm)	ε ($M^{-1}.cm^{-1}$)	Φ [a]
1	Free Aldehyde	424	462	38	39,619	0.05
2	Phenol-SB	397	460/550	63/153	45,860	0.11
3	Phenol-PSB	499	580	81	62,732	0.04
4	Phenolate-SB	448	555	107	49,538	0.12

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

a. **b.**

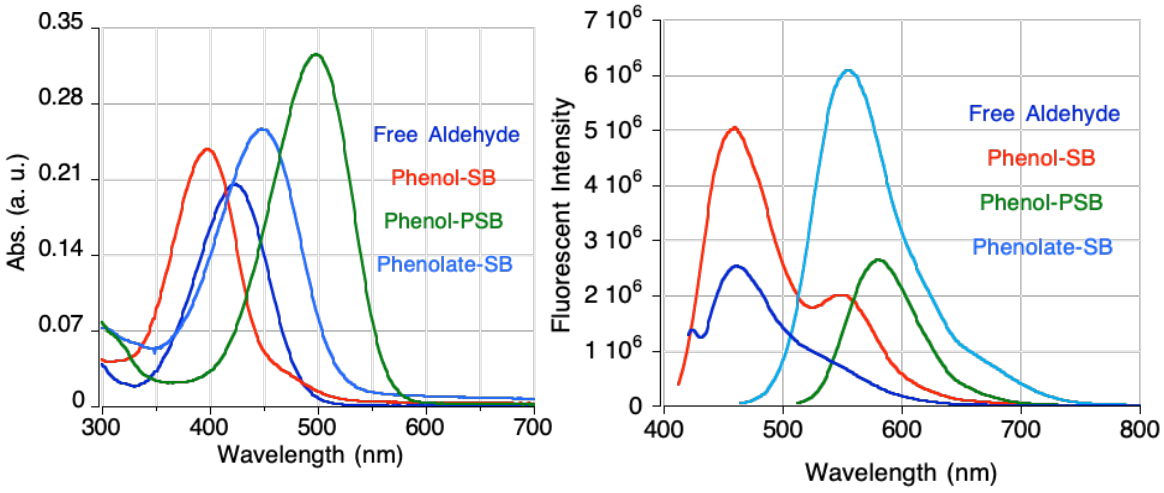


Figure III-40. Absorbance (left) and emission (right) spectra of **CyThioPhenol** and derivatives: Phenol-SB, Phenol-PSB, and Phenolate-SB.

As shown above, due to the less electron withdrawing nature of the imine as compared to the aldehyde, the absorption wavelength of the Phenol-SB is blue-shifted by 27 nm. Excitation of the Phenol-SB leads to a two maxima emission spectrum at 460 nm and 550 nm. The red-shifted emission peak results from the deprotonation of the hydroxyl group upon excitation and leads to a Stokes shift of more than 150 nm. The maximum emission wavelength of Phenolate-SB obtained upon basifying the solution with 1M NaOH at 555 nm verifies this observation. The most red-shifted absorption and emission wavelength, 499 nm and 580 nm, respectively, results from Phenol-PSB acquired via acidification of the sample with concentrated hydrochloric acid (aq.) solution.

As described in Chapter II, we cannot characterize the Phenolate-PSB since it is impossible to produce it in solution. However, we anticipate a substantial enhancement in quantum yield upon forming this complex in the engineered protein pocket due to the formation of a strong ICT system between the alkoxide and the iminium. Furthermore, solution studies show that the emission quantum yield enhances more than twice upon deprotonation of the hydroxyl group (compare phenolate-SB to free aldehyde), which forms a rather weak push-pull system as the result of the less electron-withdrawing effect of SB as compared to the PSB (**Figure III-41**).

Theoretically, it is not possible to deprotonate the hydroxyl group in neutral pH due to the high pK_a of the phenol moiety (>9). Additionally, Phenol-SB and Phenol-PSB are barely emissive beyond 650 nm. Therefore, we sought to pursue

the development of **CyThioPhenol**-hCRBPII complexes as a no-wash background-free fluorescent tag.

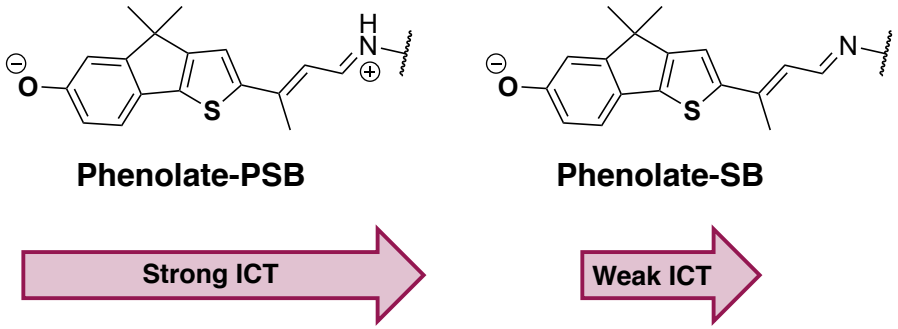


Figure III-41. Schematic comparison of ICT systems in Phenolate-PSB and Phenolate-SB complexes of **CyThioPhenol**.

III.8.2 Developing CyThioPhenol photoacidic complexes

There is no crystal structure of any **CyThioPhenol**-hCRBPII complexes up to this point. However, studies with **ThioPhenol** demonstrated that mutating a few residues around the hydroxyl group to proton acceptor residues, such as histidine, tyrosine, and cysteine, would enhance the deprotonation of the hydroxyl group and Φ_{ESPT}. Hence, we chose to measure the photophysical properties upon adding those mutations (**Table III-29**). Q108K:K40L:T51V:T53S:**R58H** template is retained in all mutants tested as screening different amino acids at position 58, including retention of R58, indicated that histidine yields the highest Φ_{ESPT} (see **Table II-3**). Φ_{ESPT} is defined as the fraction of total fluorescence from the excited state **ThioPhenolate**-PSB, which is obtained as explained in **Section II-7**.

In the first step, we sought to investigate the effect of key mutations, F16Y, A33H, A33Y, and L77Y on the Φ_{ESPT}. Upon testing with the first three mutants, **CyThioPhenol** showed higher iminium pK_a and Φ_{ESPT} as compared with the same protein variants binding **ThioPhenol**. In addition, it appears that the concentration of the deprotonated hydroxyl group or Phenolate-PSB complex in the ground state is higher with the cyclized structure.

Table III-30. Spectroscopic change as the result of the addition of tyrosine and histidine residues around the hydroxyl group.

Entry	hCRBPII mutant [a]	λ_{abs}	λ_{em}	pK_a	Φ_{ESPT}
1	KLVSH:Y19W:**A33Y**:**F16Y**	562	598/665	7.3	0.24
2	KLVSH:**A33H**:**L77Y**	563	660	6.9	0.73
3	KLVSH:**F16Y**:**L77Y**	560/628	650	7.1	> 0.99

[a] KLVSH is equal to Q108K:K40L:T51V:T53S:R58H.

As shown above, the highest Φ_{ESPT} was obtained upon the addition of F16Y:L77Y mutations (entry 3). Interestingly, the absorption spectrum of this mutant measured in neutral pH shows three maxima at 382 nm, 560 nm, and 628 nm, which are correlated with Phenol-SB, Phenol-PSB, and Phenolate-PSB, respectively (**Figure III-42a**). Therefore, we measured the emission spectrum upon excitation at both 560 nm and 628 nm to verify the accuracy of this statement, and as expected, both excitations result in the same emission. Interestingly, even excitation of the SB leads to residual Phenolate-PSB formation showing a substantial Stokes shift of 268 nm (**Figure III-42b**). Besides, the complex's

absorption and emission spectra were monitored upon both acid and base titrations to characterize this system (**Figure III-42 c**, **d**).

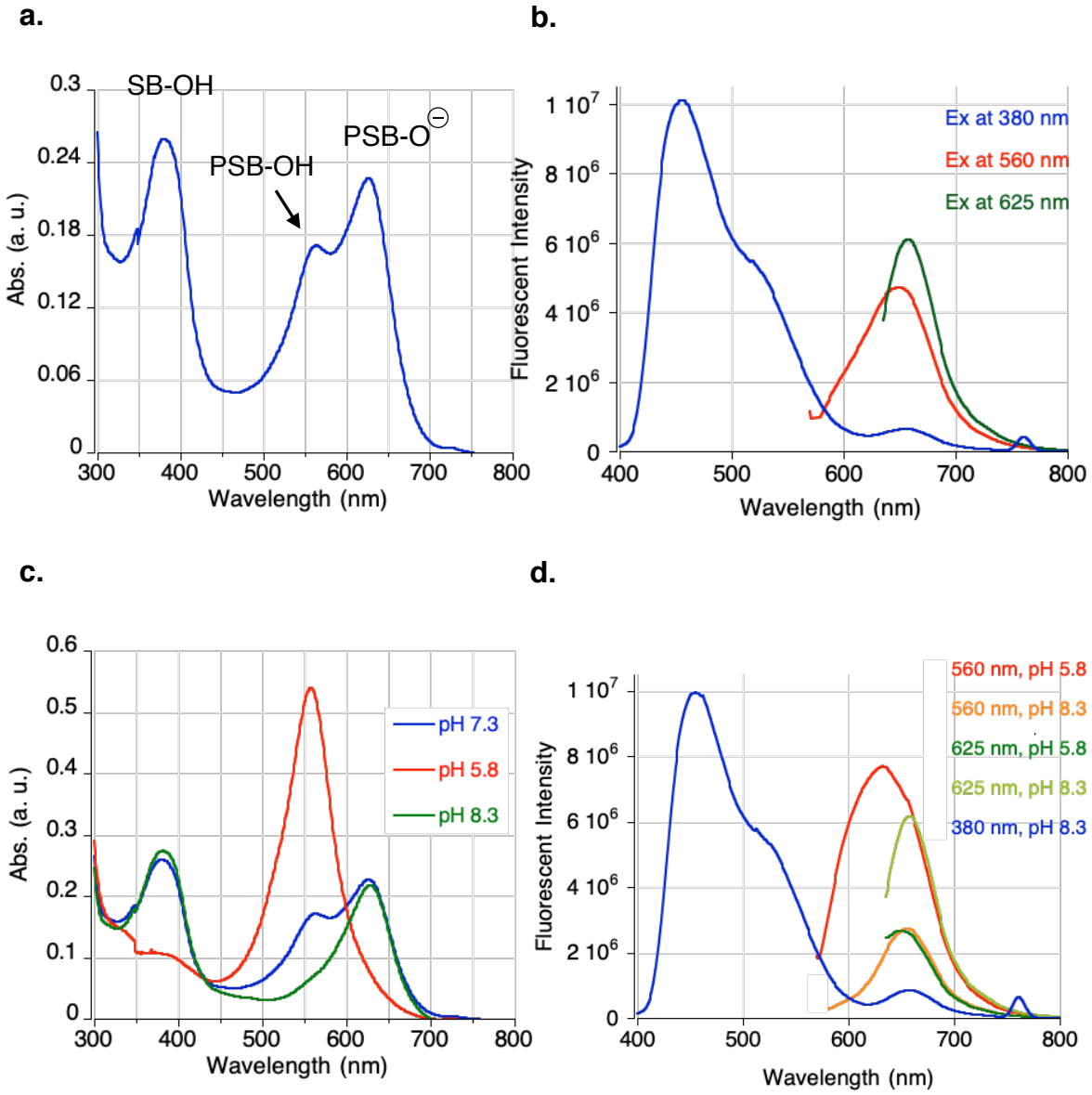


Figure III-42. **a**. The absorption and **b**. the emission spectra of KLVSH is equal to Q108K:K40L:T51V:T53S:R58H:F16Y:L77Y/**CyThioPhenol** complex measured in neutral pH 7.3. **c**, **d**. The absorption and emission spectra of the same complex screened upon acidification pH 5.8 and basification pH 8.3 of the solution.

As shown in **Figure III-42**, acidification of the solution leads to a significant increase in Phenol-PSB absorption and emission. Furthermore, clearly due to the high pK_a of the hydroxyl group, the majority of Phenolate-PSB is protonated in pH 5.8, resulting in the disappearance of its absorption peak (compare the blue line and the red line). On the contrary, the concentration of Phenolate-PSB increases upon basifying the solution. However, unfortunately, this protein complex was not stable in more basic environments, but we presume that in higher pH values, the concentration of Phenolate-PSB would decrease, and eventually, the Phenolate-SB will be the single species present in solution.

Nonetheless, for photoacids, the iminium pK_a should be high enough to produce maximum PSB in order to form a strong ICT system upon deprotonation of the hydroxyl group in the excited state. Our previous studies show that L117E/D mutations effectively enhance the iminium pK_a through electrostatic interactions between the iminium and the aspartic/glutamic acid side chains. Unfortunately, as described in Chapter II, the addition of these mutations prevents the ESPT of the hydroxyl group. Upon extensive experiments, we realized that L117C gives the best results in terms of increasing the pK_a and maintaining the photoacidic properties of the complex. Thus, we sought to explore whether the addition of L117C mutation can increase the pK_a with the cyclized structure. Additionally, we measured the photophysical properties of **CyThioPhenol** upon the addition of T29Y, T51C, T53C mutations, the same residues that yielded the highest Φ_{ESPT} with **ThioPhenol** (**Table III-30**).

To our delight, the addition of L117C mutation increased the iminium pK_a without preventing or reducing the Φ_{ESPT}. Notably, the average iminium pK_a with **CyThioPhenol** is slightly higher than **ThioPhenol** leading to minimal SB concertation. Interestingly, the least amount of Φ_{ESPT} was obtained upon binding to Q108K:K40L:T51C:T53S:R58H:Y19W:A33H:F16Y:T29Y:L117C and Q108K:K40L:T51V:T53C:R58H:Y19W:A33H:F16Y:T29Y:L117C mutants (entries 3 and 5), which yielded the highest Φ_{ESPT} with **ThioPhenol**. These results prove that L77Y mutation is vital for ESPT of the hydroxyl group, as all other mutants containing this mutation result in higher Φ_{ESPT}.

Table III-31. Spectroscopic change as the result of the addition of L117C, T29Y, T51C and T53C mutations.

Entry	hCRBPII mutant	λ_{abs}	λ_{em}	pK_a	Φ_{ESPT}	Φ
1	KLVSH:Y19W:A33H:F16Y:L77Y:L117C [a]	552/626	653	9.1	> 0.99	0.31
2	KLVSH:A33H:L77Y:L117C	558	658	9.2	0.54	0.33
3	KLCSH:Y19W:A33H:F16Y:T29Y:L117C [b]	548/631	585	8.9	0.18	0.30
4	KLVCH:Y19W:A33H:L77Y:L117C [c]	551	666	10	0.84	0.30
5	KLVCH:Y19W:A33H:F16Y:T29Y:L117C	550	582	9.2	0.20	0.36
6	KLVCH:Y19W:A33H:F16Y:L77Y:L117C	549	652	9.3	> 0.99	0.35
7	KLVCH:Y19W:F16Y:T29Y:L77Y:L117C	549	647	9.8	> 0.99	0.32
8	KLVCH:Y19W:A33H:F16Y:T29Y:L77Y:L117C	548	650	9.9	> 0.99	0.35

[a] KLVSH is equal to Q108K:K40L:T51V:T53S:R58H.
[b] KLCSH is equal to Q108K:K40L:T51C:T53S:R58H.
[c] KLVCH is equal to Q108K:K40L:T51V:T53C:R58H.

Moreover, comparing the protein mutants and the Φ_{ESPT} in **Table III-30** indicates that F16Y is also essential in facilitating the hydroxyl group's deprotonation. entries 2 and 4 contain the L77Y mutation but lack F16Y, which leads to the second-lowest Φ_{ESPT} values. Nonetheless, substitution of serine with cysteine in entry 4, Q108K:K40L:T51V:T53C:R58H:Y19W:A33H:L77Y:L117C, leads to higher Φ_{ESPT}, 84% vs. 54%, and thus Q108K:K40L:T51V:T53C:R58H was retained for further protein engineering.

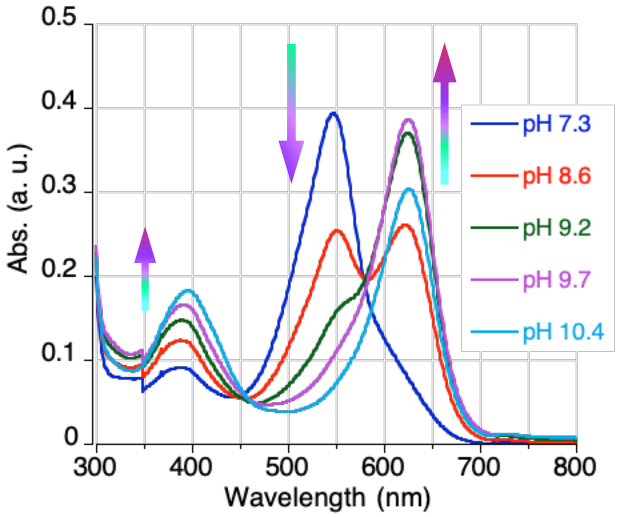


Figure III-43. The absorption spectra of Q108K:K40L:T51V:T53C:R58H:Y19W:F16Y:T29Y:L77Y:L117C/**CyThioPheno**l complex recorded upon basifying the solution from pH 7.3 to pH 9.5.

All other three protein mutants containing F16Y and L77Y along with T53C resulted in a pretty narrow emission spectrum corresponding to Phenolate-PSB complex and Φ_{ESPT} more than 99% (entries 6-8). However, the highest soluble protein expression yield was obtained with Q108K:K40L:T51V:T53C:R58H:Y19W:A33H:F16Y:L77Y:L117C (**M6**) mutant,

entry 6. Additionally, the red-shifted shoulder corresponding to the **ThioPhenolate**-PSB complex is the smallest with this mutant. Hence, this mutant was chosen for live-cell imaging experiments.

Noteworthy, **CyThioPhenol** complex with Q108K:K40L:T51V:T53C:R58H:Y19W:F16Y:T29Y:L77Y:L117C mutant (entry 7) was more stable in basic pH, and we were able to monitor the absorption spectra change upon base titration. As expected, in higher pH values, all Phenol-PSB is converted to Phenolate-PSB, but with increasing pH values, the Phenolate-PSB concertation decreases upon conversion to the Phenolate-SB complex (**Figure III-43**).

III.8.2. A. Kinetic measurements of CyThioPhenol binding hCRBPII Photoacid mutants

Binding kinetics were measured prior to confocal imaging experiments to explore the proper incubation time; 20 μM protein in PBS buffer at neutral pH 7.3 was incubated with 0.5 equivalent of the ligand at 23 °C, and the increase in absorbance of the corresponding PSB at its λ_{max} was recorded over time. Collected data points were fitted with a second-order rate equation considering multiple reagents protein and the ligand with non-equal concentrations (**Figure III-44**). The equation and detailed fitting process are described in **Section IV.3.5**.

Compared to **ThioPhenol** binding photoacid mutants, the half-time of the cyclized structure binding to **M6** is longer. However, we envisioned that even longer incubation times would not lead to the fluorescent background as the

fluoresce quantum yield of **CyThioPhenol** significantly increases upon binding the target mutant as compared to the free aldehyde or non-specific bindings (see **Table III-30** and **Table III-28**). The next section of this chapter aims to demonstrate this probe's utility in live-cell imaging; most importantly, **CyThioPhenol** cell permeability and its efficient target binding are shown.

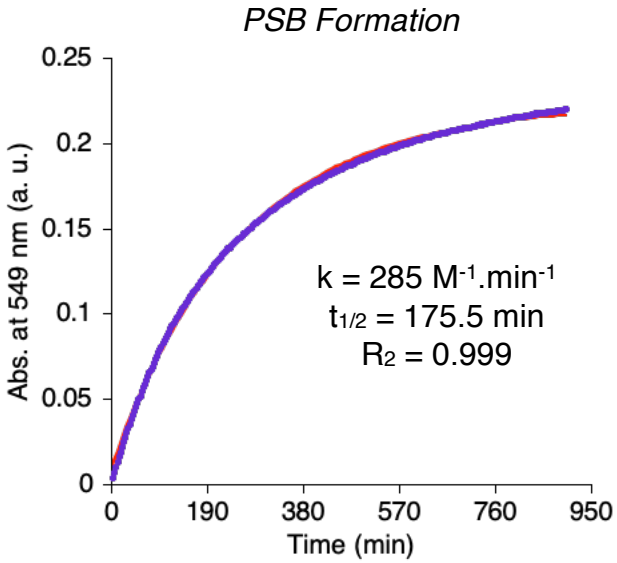


Figure III-44. Rate of **CyThioPhenol/M6** PSB formation fitted to second-order kinetics with 20 µM protein and 10 µM **CyThioPhenol**. Plotted is the concentration of free chromophore vs. time.

III.8.2. B. Visualization of hCRBPII/CyThioPhenol in mammalian cells

Next, we sought to probe the performance of the engineered hCRBPII/**CyThioPhenol** photoacid complexes as a no-wash live-cell imaging system. As described earlier in this section, the mutant Q108K:K40L:T51V:T53C:R58H:Y19W:A33H:F16Y:L77Y:L117C (**M6**) was chosen for these studies. As with this mutant, the iminium pK_a is high enough (9.3, **Table**

III-30, entry 6) to result in mainly PSB formation, the Φ_{ESPT} (>99%), and fluorescence quantum yield (0.35) is the highest with this mutant. In addition, the concentration of **ThioPhenolate**-PSB in the ground state was negligible, with **M6** leading to a single peak for PSB absorption. Furthermore, the protein expression yield with the **M6** mutant is relatively high, and it only expresses in the monomeric form. The absorption and emission wavelength of **CyThioPhenol/M6** plus its pK_a titrations are depicted in **Figure III-45**. Additionally, all spectroscopic properties of **CyThioPhenol/M6** are summarized in **Table III-31**.

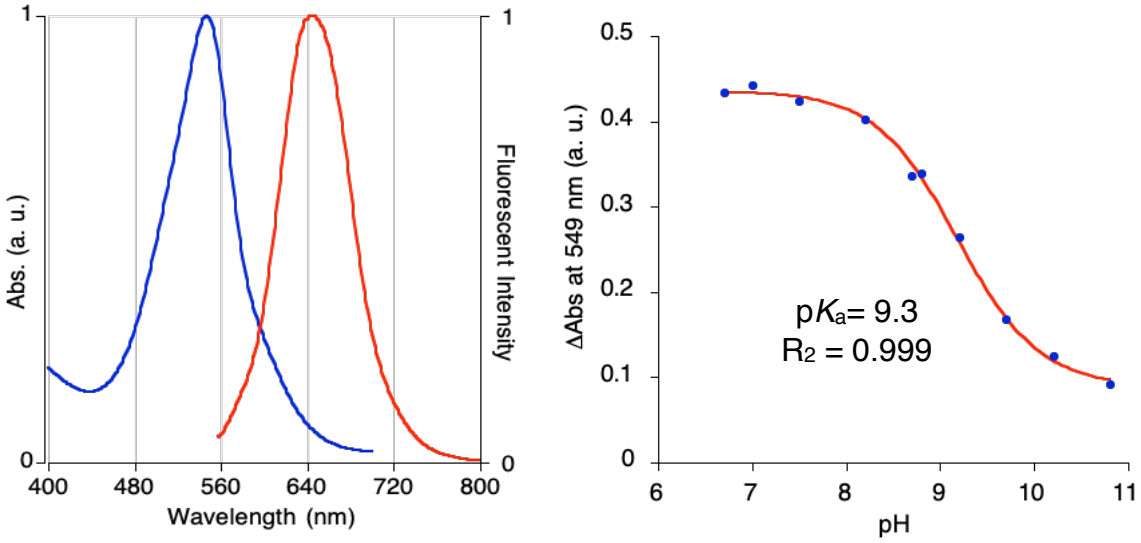


Figure III-45. Spectroscopic properties of Q108K:K40L:T51V:T53C:R58H:Y19W:A33H:F16Y:L77Y:L117C mutant with **CyThioPhenol** including UV-Vis and fluorescence spectra (left) and the pK_a titration (right).

The spectroscopic properties of **CyThioPhenol** and **ThioPhenol** as protein-based photoacidic fluorescent tags are compared in **Table III-31**. The total brightness has enhanced more than 2.5 times with the cyclized structure, although the binding rate has decreased.

Table III-32. Spectroscopic properties of **CyThioPhenol/M6** complex compared with **ThioPhenol/M3** complex.

Complex	λ_{abs}	λ_{em}	ε ($M^{-1}.cm^{-1}$)	Φ_{ESPT}	Φ [a]	pK_a	k ($M^{-1}.min^{-1}$)	$t_{1/2}$ [b] (min)
CyThioPhenol/M6	549	652	38,024	> 0.99	0.35	9.3	285	175.5
ThioPhenol/M3	517	679	33,743	> 0.99	0.15	9.8	1467	34

[a] Absolute quantum yield was measured on a Quantaurus-QY. [b] Half-life based on the rate constant obtained from second-order rate fitting; measured at 23 °C with 20 μM protein and 0.5 equiv **ThioPhenol** at pH 7.2.
M6: Q108K:K40L:T51V:T53C:R58H:Y19W:A33H:F16Y:L77Y:L117C.
M3: Q108K:K40L:T51V:T53C:R58H:Y19W:A33H:F16Y:T29Y:L117C.

In the next step, we sought to explore the performance of **CyThioPhenol** as a no-wash tag for *in vivo* confocal imaging experiments. Thus, the **M6**-hCRBPII mutant was cloned into the same vector described in Chapter II, pFlag-CMV2 vector containing EGFP as an internal standard to assess the specificity of the fluorescent signal. This vector was used for whole-cell labeling; additionally, to target **M6** in the cytosol, the signaling peptide NES (nuclear export sequence) was fused to the C-terminus of **M6**, as illustrated in **Figure III-46**.

HeLa cancer cells were then transiently transfected with the fused constructs to express the corresponding protein. Transfected cells were incubated with 5 μM **CyThioPhenol** overnight at 37 °C. The cells were then directly subjected to confocal imaging without any washing steps prior to confocal imaging experiments.

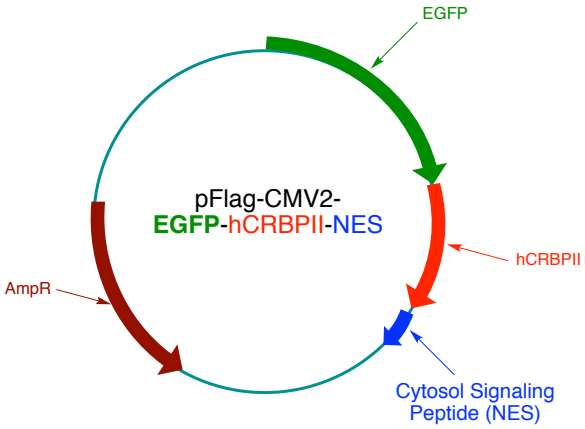


Figure III-46. Schematic map of EGFP-hCRBPII-NES fusion construct. NES: nuclear export sequence.

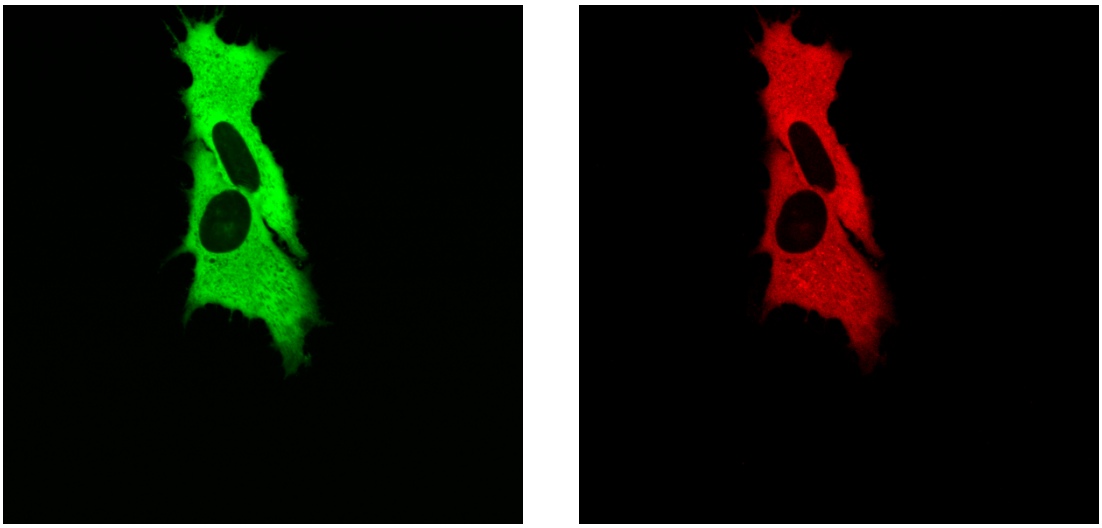
The green fluorescence observed upon excitation at 488 nm verifies the successful cell transfection and the fusion protein expression (**Figure III-47**, the green channel). The stained cells were excited with a 559 nm laser, and the far-red emission was collected with a 600 nm to 700 nm bandpass for the red channel.

Interestingly, even after overnight incubation of **CyThioPhenol** with cells (>12 hours), no unspecific red fluorescent signal was observed, confirming the selectivity of this probe (**Figure III-47**, the red channel). As discussed in **Section III.8.1**, free **CyThioPhenol** and off-target imine or iminium formation do not lead to the fluorescent background. First, these species' (Phenol-SB and Phenol-PSB) total fluorescence quantum yields are much lower than the imaging complex.

Second, only a minor part of their emission is beyond 600 nm making no-wash imaging feasible.

EGFP-*NES*
Ex: 488 nm
Em: BP 500 nm- 550 nm

CyThioPhenol/hCRBPII-*NES*
Ex: 559 nm
Em: BP 600 nm- 700 nm



DIC
Ex: 488 nm
BP 500 nm- 550 nm

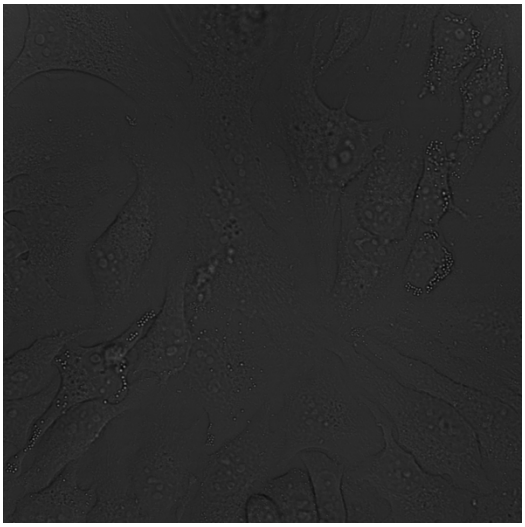


Figure III-47. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-NES. NES: nuclear export sequence. Cells were stained with 5 μM **CyThioPhenol** and incubated at 37 °C overnight. Cells were not washed before imaging.

III.8.3 Developing the double ESPT systems with CyThioPhenol-hCRBPII complexes

In the previous section, we showed the application of **CyThioPhenol** as a fluorescent photoacid. The successful application of this novel structure in no-wash background-free imaging inspired us to examine imaging with a double ESPT systems. In the first step, we measured the photophysical properties with **M3**, the photobase mutant that yields the highest Φ_{ESPT} for the ESPT of imine to iminium. The obtained results are compared with **ThioPhenol** binding the same mutant (**Table III.32**).

Table III-33. Spectroscopic properties of **ThioPhenol** and **CyThioPhenol** binding **M3** mutant.

Entry	Complex	λ_{abs}	λ_{em}	SS (nm)	Φ_{ESPT}	Φ [a]
1	**ThioPhenol/M3** [b]	370	532	162	0.98	0.05
2	**CyThioPhenol/M3**	398	541	143	> 0.99	0.21

[a] Absolute quantum yield was measured on a Quantaurus-QY. [b] **M3** mutation: Q108K:K40E:T53A:Q38F:Q4F:R58L.

As shown above, the bathochromic shift in the absorption wavelength (28 nm) is less than for the emission wavelengths (9 nm), leading to a smaller Stokes shift for the cyclized structure. Nonetheless, the blue-shifted absorption wavelength of **ThioPhenol** bound to photobase or double ESPT mutants limits its application for live-cell imaging. The most blue-shifted laser confocal microscopes are equipped with a 400 nm laser beam that can excite less than 50% of the corresponding complex and reduces the tag's total brightness (**Figure III-48**).

Interestingly, **CyThioPhenol** has a 15-fold enhancement in emission, and the fluorescence quantum efficiency is significantly increased (from 0.05 to 0.21). Note the emission intensities of **CyThioPhenol/M3** (left) and **ThioPhenol/M3** (right).

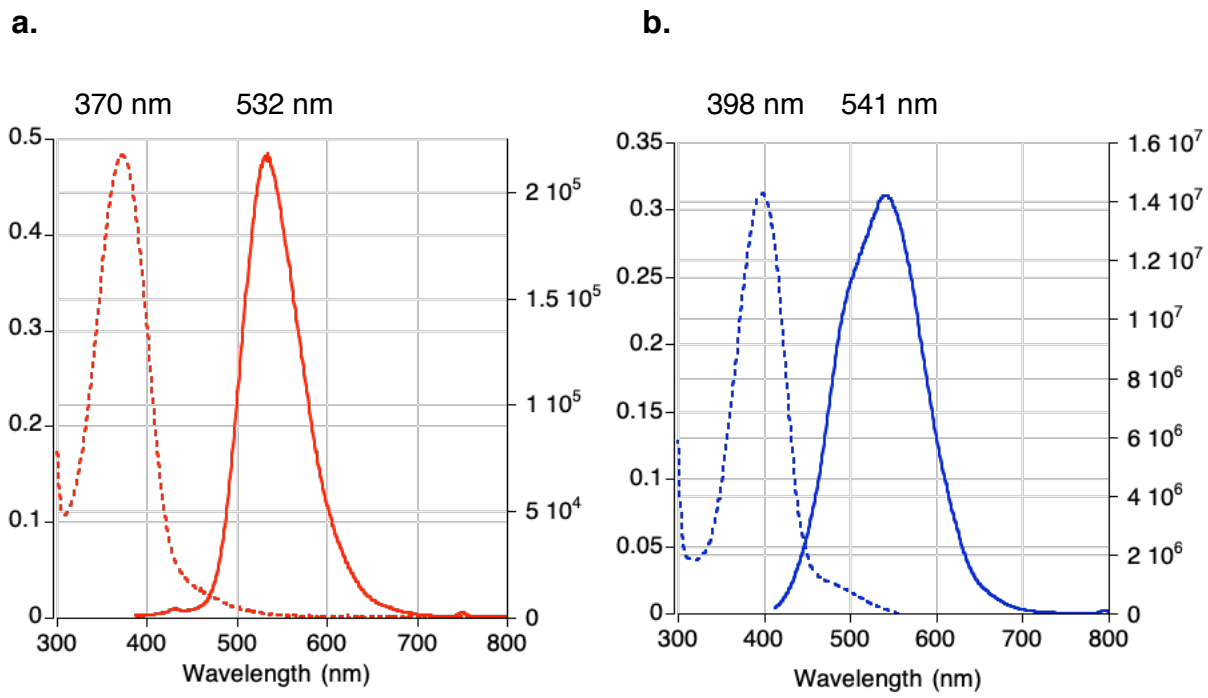


Figure III-48. The absorption and emission spectra of **a**. **ThioPhenol/M3** (red lines) and **b**. the absorption and emission spectra of **CyThioPhenol/M3** complex (blue lines).

As a control experiment, the photophysical properties were measured upon binding to Q108K:**K40L**:T53A:Q38F:Q4F:R58L, wherein the proton source, the glutamic acid at position 40, is substituted with leucine. As expected, the maximum emission wavelength blue-shifted to 471 nm corresponding to the Phenol-SB complex verifying the R58H as the proton donor source. However, the broad

emission spectrum indicates that the ESPT process is not entirely stopped with the K40L mutation (**Figure III-49**).

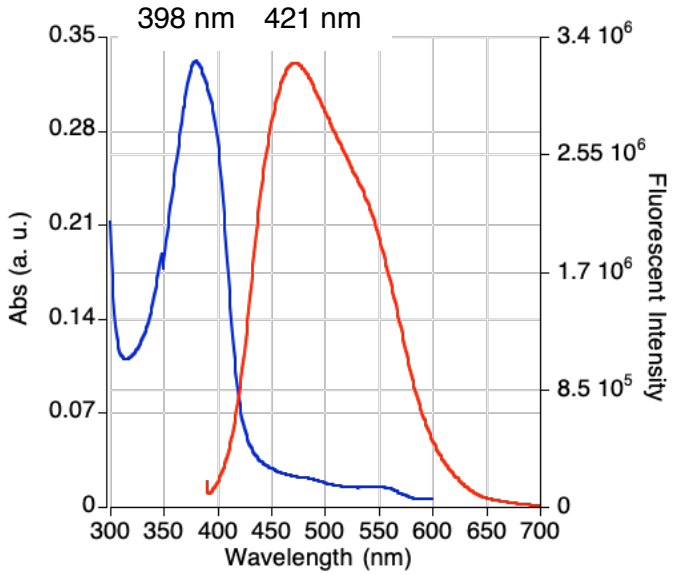


Figure III-49. The absorption and emission spectra of Q108K:**K40L**:T53A:Q38F:Q4F:R58L/**CyThioPheno**l complex measured in neutral pH.

Next, we sought to investigate whether adding more acidic residues around the imine bond would facilitate the ESPT process and enhance the emission quantum yield. Notably, residues V62 and I42 are located at a proper distance from imine, 6.7 Å and 5.4 Å, respectively; not too close to protonate the imine in the ground state, and yet not too far to prevent the proton transfer in the excited state (**Figure III-50**). We also measured the photophysical properties upon mutation to aspartic acid, K40D. Unfortunately, many of the mutants led to insoluble protein expression; expressed proteins' spectroscopic properties are shown in **Table III-33**.

Substitution of K40E with K40D mutation in **M3** complex, Q108K:**K40D**:T53A:Q38F:Q4F:R58L, decreases the Φ_{ESPT} to less than 50% (entry 1). Additionally, in contrast to **M3**, this mutant expresses as a mixture of monomer and dimer, which can hamper its further applications in cell imaging.

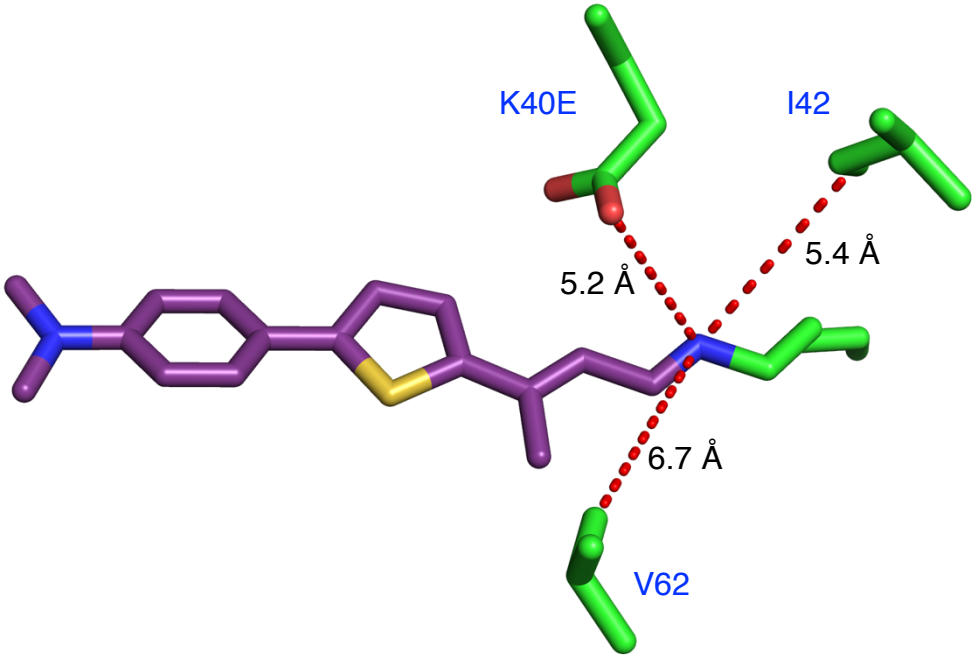


Figure III-50. The crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFlour** and highlighted V62, I42, and K40E residues.

As listed in **Table III-33**, the addition of V62E and I42E mutations with aspartic acid, glutamic acid, and leucine residues at position 40 (entries 2, 3, and 4, respectively) results in reduced Φ_{ESPT}, higher iminium pK_a values, and more dimer expressions. Previously Dr. Elizabeth Santos demonstrated that T51V mutation is significantly effective in monomerizing hCRBPII variants.[46] Thus, we sought to explore if more monomer formation would have any influence on the

Φ_{ESPT}. Interestingly, the only mutant, Q108K:**K40D**:**T51V**:T53A:R58L:Q38F:Q4F:**V62E** showing a lower dimer percentage, gave the highest Φ_{ESPT} (entry 5).

Table III-34. Spectroscopic properties of mutants as the result of the introduction of V62E, I42E, and K40D mutations.

Entry	hCRBPII mutant	λ_{abs}	λ_{em}	Dimer%	pK_a	Φ_{ESPT}
1	Q108K:**K40D**ALFF [a]	396	496/558	53	6.4	0.49
2	Q108K:**K40D**ALFF:**V62E**	397	550	78	5.6	0.73
3	Q108K:**K40E**ALFF:**V62E**	401	549	69	5.3	0.84
4	Q108K:**K40L**ALFF:**I42E**	395	446	9	5.2	0.27
5	Q108K:**K40D**:**T51V**ALFF:**V62E**	398	539	12	< 5	> 0.99

[a] ALFF is equal to T53A:R58L:Q38F:Q4F

These measurements indicate that **CyThioPhenol** can appear as a photobase, and thus, we sought to develop double ESPT systems with this chromophore. The following aims to describe our attempts to engineer hCRBPII so that **CyThioPhenol** can act as both photobase and photoacid upon a single photoirradiation.

First, we measured **CyThioPhenol** photophysical properties upon binding **M4**, Q108K:K40E:T53A:R58H:Q38F:Q4F, the mutant that yielded the highest quantum yield of double ESPT with **ThioPhenol** (**Table III-34**). As described earlier in this chapter, **M4** was designed simply with the following objectives: K40E

is the proton source for imine to iminium ESPT, and R58H acts as the proton acceptor residue for the hydroxyl to alkoxide ESPT process.

Table III-35. Spectroscopic properties of **ThioPhenol** and **CyThioPhenol** binding **M4** mutant.

Entry	Complex	λ_{abs}	λ_{em}	SS (nm)	Φ_{ESPT}	Φ [a]
1	**ThioPhenol/M4** [b]	370	612	242	0.98	0.16
2	**CyThioPhenol/M4**	399	553	154	< 0.50	0.20

[a] Absolute quantum yield was measured on a Quantaurus-QY. [b]. **M4** mutation: Q108K:K40E:T53A:Q38F:Q4F:R58H.

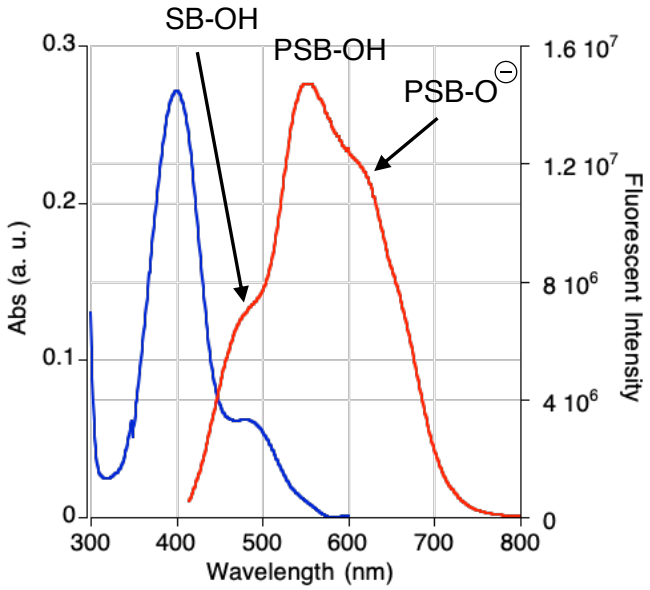


Figure III-51. Absorption and emission spectra of Q108K:K40E:T53A:Q38F:Q4F:R58H/**CyThioPhenol** complex measured in neutral pH.

Evidently, the maximum emission wavelength of **CyThioPhenol/M4** complex at 553 nm is due to the imine ESPT process (see **Table III-32**). However, its emission spectrum is extremely broad, spanning from 414 nm to 738 nm,

indicating that Phenol-PSB is not the only species formed upon excitation (**Figure III-51**). Therefore, we sought to investigate whether the insertion of proton acceptor residues around the hydroxyl group would enhance the double ESPT emission and consequently give a narrow emission spectrum. Although there is no available crystal structure from any of the **CyThioPhenol**-hCRBPII complexes, upon studying the **ThioPhenol** structures, we could identify 7 residues adjacent to the hydroxyl group. We surmised that a nearby basic or proton acceptor residue could abstract the proton from the hydroxyl group, a weak photoacid, and thus, these residues were individually mutated to tyrosine, cysteine, methionine, and histidine (F16, M20, T29, A33, Q38, R58, L77).

The photophysical properties of the mutants that led to soluble protein expressions are listed in **Table III-35**. All mutants show low iminium pK_a values (5.0-5.6), leading to SB formation in the ground state. However, upon excitation of the SB at its maximum absorption wavelength (~400 nm), a broad emission spectrum was obtained that is maximized in the range of 550 nm-570 nm, corresponding to the Phenol-PSB complex. Additionally, we observed a small shoulder to the left side of the PSB emission around 485 nm correlating to the Phenol-SB complex for most of the mutants. The absorption and emission spectra of Q108K:K40E:T53A:Q38F:Q4F:R58H:F16Y/**CyThioPhenol** complex (entry 1) is shown in **Figure III-52a**.

Table III-36. Spectroscopic properties of mutants at positions 16, 20, 29, 33, 38, 58, and 77.

Entry	hCRBPII mutant [a]	λ_{abs}	λ_{em}	SS (nm)
1	KEAFF:R58H:**F16Y**	399	485/583	86/184
2	KEAFF:R58H:**A33H**	402	552/618	150/216
3	KEAFF:R58H:**A33H**:**F16Y**	400	483/566	83/166
4	KEAFF:R58H:**A33M**	396	476/553	80/157
5	KEAFF:R58H:**A33M**:**F16Y**	400	480/557	80/157
6	KEAFF:R58H:**A33C**:**F16Y**	400	483/566	83/166
7	KEAFF:R58H:**F16Y**:**Q38Y**	398	486/549	88/151
8	KEAFF:R58H:**F16Y**:**Q38H**	401	480/569	79/168
9	KEAFF:R58H:**F16Y**:**Q38Y**:**T29Y**	398	483/553	85/155
10	KEAFF:R58H:**M20Y**	400	553	53
11	KEAFF:R58H:**A33H**:**L77Y**	408	503/555/619	95/147/211
12	KEAFF:R58H:**F16Y**:**Q38Y:L117E**	425	475/581	50/156
13	KEAFF:**R58Y**:**F16Y**	400	555	155

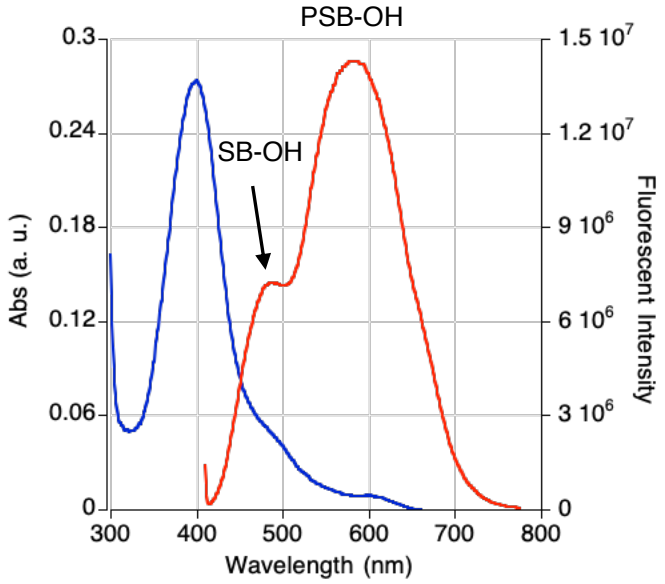
[a] KEAFF is equal to Q108K:K40E:T53A:Q38F:Q4F.

Unfortunately, none of the designed mutants gave a narrow emission derived from the double ESPT process, which expectedly should appear post 600 nm. However, we noticed a small, red-shifted shoulder next to the Phenol-PSB emission that supposedly corresponds to the Phenolate-PSB complex, the double ESPT process product for most mutants.

Particularly for Q108K:K40E:T53A:Q38F:Q4F:R58H:**A33H** and Q108K:K40E:Q4F:A33H:Q38F:T53A:R58H:**L77Y** mutants, entries 2 and 11,

respectively, the shoulder is more evident and appears around 620 nm. The absorption and emission spectra of the former are shown in **Figure III-52b**.

a.



b.

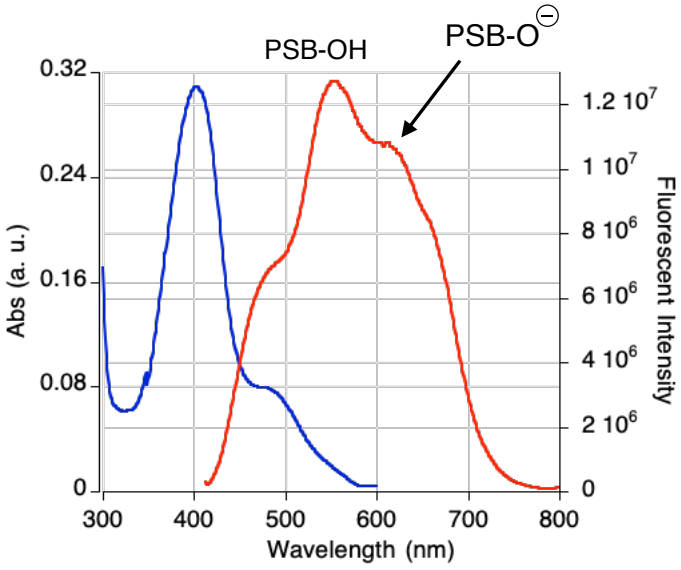


Figure III-52. The absorption and emission spectra of **CyThioPhenol** upon binding **a**. Q108K:K40E:T53A:Q38F:Q4F:R58H:A33H and **b**. Q108K:K40E:Q4F:A33H:Q38F:T53A:R58H:L77Y mutants.

Noteworthy, the substitution of histidine with tyrosine in Q108K:K40E:T53A:Q38F:Q4F:R58H:F16Y mutant reduces the red-shifted shoulder correlated with Phenolate-PSB complex and results in a relatively narrow emission spectrum that corresponds to the Phenol-PSB (entry 13). This observation confirms our previous studies that K40H mutation works best to deprotonate the hydroxyl group, but evidently, it is not optimal for this structure.

We next sought to explore how changing the residue at position 40 would affect the double ESPT of **CyThioPheno**l-hCRBPII complexes. Thus, K40 residue was mutated to several different amino acids in the Q108K:K40E:T53A:Q38F:Q4F:R58H template mutant (**Table III-36**).

Table III-37. Spectroscopic properties upon mutating the K40 residue.

Entry	hCRBPII mutant [a]	λ_{abs} [b]	λ_{em}	SS (nm)	pKa
1	K40	392	457/565	65/173	8.2
2	K40E	399	553	154	5.2
3	K40D	395	476/560/623	81/165/228	6.9
4	K40H	396	453/550	57/154	8.4
5	K40R	392	449/562	57/170	5.1
6	K40N	393	473/520	80/127	5.3

[a] The K40 mutations of **M4**: Q108K:K40E:T53A:Q38F:Q4F:R58H. [b] Only the SB maximum absorption wavelength is listed in this table.

Same as the template mutant, the designed mutants mainly gave rise to Phenol-SB and Phenol-PSB emission. However, interestingly, excitation of the SB of Q108K:**K40D**:T53A:Q38F:Q4F:R58H mutant (entry 3) resulted in a triple-

maxima emission spectrum. Moreover, this mutant leads to the highest intensity for the Phenolate-PSB and largest stokes shift, 228 nm (**Figure III-53**). Nonetheless, as shown below, the double ESPT derived emission maximized at 623 nm is associated with other species emission leading to an excessively broad spectrum.

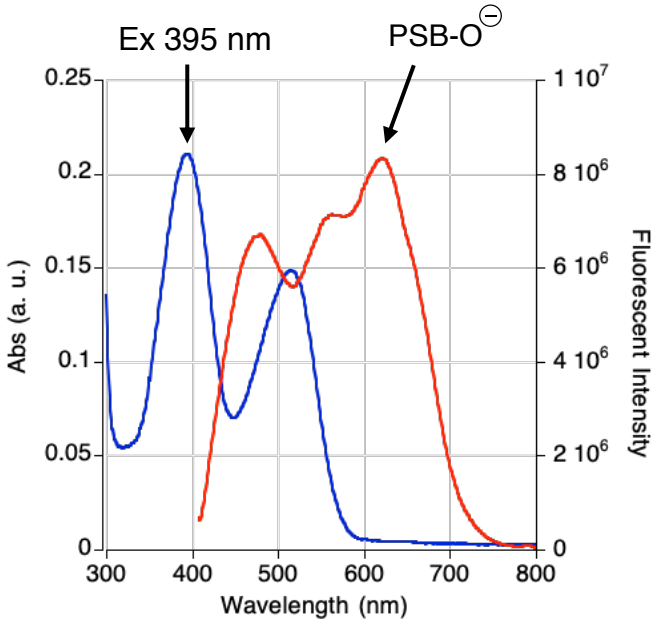


Figure III-53. The absorption and emission spectra of Q108K:**K40D**:T53A:Q38F:Q4F:R58H/**CyThioPhenol** complex upon SB excitation.

This section showed the successful development of **CyThioPhenol** as a fluorescent photobase and photoacid upon binding hCRBPII mutants. However, further protein engineering is required to optimize the protein cavity for a double ESPT system. In the light of measuring **CyThioPhenol** properties upon binding more than 35 different mutants, we realized the position of the proton acceptor

residues, such as R58, F16, A33, which were optimized for **ThioPhenol**, is not well optimized for the cyclized structure. Undoubtedly **CyThioPhenol** crystal structures will help identify the closet residues that could facilitate its deprotonation.

III.8.4 Application of CyThioPhenol in multicolor no-wash live-cell confocal imaging

The distinctive absorption and emission features of **CyThioPhenol** when it is bound to a photoacid vs. photobase mutant are well suited for two-color imaging applications. The photophysical properties of the photobase and photoacid complexes are compared in **Table III-37**.

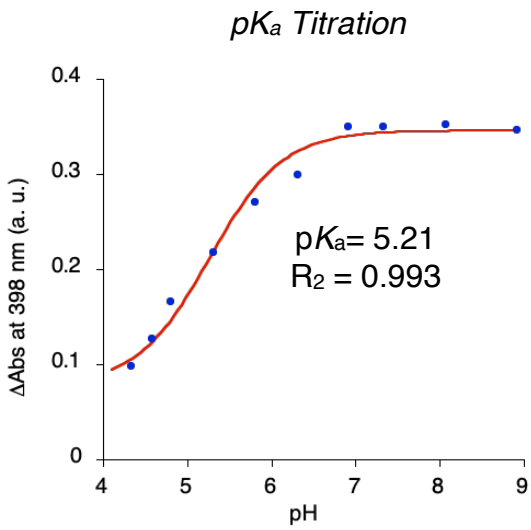
Table III-38. Spectroscopic properties of **CyThioPhenol/M6** complex compared with **CyThioPhenol/M3** complex.

Complex	λ_{abs}	λ_{em}	ε ($M^{-1}.cm^{-1}$)	Φ_{ESPT}	Φ [a]	pK_a	k ($M^{-1}.min^{-1}$)	$t_{1/2}$ [b] (min)
CyThioPhenol/M6	549	652	38,024	> 0.99	0.35	9.3	285	175.5
CyThioPhenol/M3	398	541	36,543	> 0.99	0.21	5.2	N.D	1.6

[a] Absolute quantum yield was measured on a Quantaurus-QY. [b] Half-life based on the rate constant obtained from second-order rate fitting; measured at 23 °C with 20 μM protein and 0.5 equiv **CyThioPhenol** at pH 7.2.
M6:Q108K:K40L:T51V:T53C:R58H:Y19W:A33H:F16Y:L77Y:L117C (Photoacid).
M3: Q108K:K40E:T53A:Q38F:Q4F:R58L (Photobase).

As shown in **Figure III-54b**, these two complexes' absorption and emission spectra are well separated, 151 nm for the absorption and 111 nm for the maximum emission wavelengths, making it possible to excite and collect the corresponding emission without any fluorescence leaking into another channel.

a.



b.

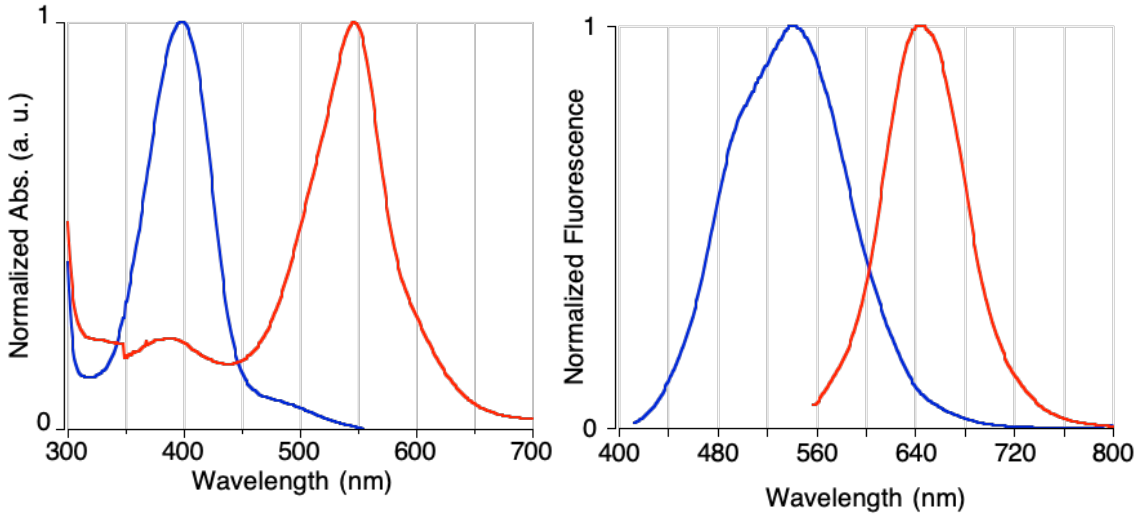


Figure III-54. **a**. **M3/CyThioPhenol** complex pK_a titration. **b**. The absorption (left) and emission (right) spectra of **M3** (blue) and **M6** (red) mutants upon binding **CyThioPhenol** measured at pH 7.2.

This is primarily due to the pK_a values of the photoacid and photobase mutants, **M6** and **M3**, leading to pure PSB and SB formation in the neutral pH (**Figure III-52a**, see **Figure III-45**). Noteworthy, the reaction between Q108K:K40E:T53A:Q38F:Q4F:R58L and **CyThioPhenol** is complete in less than 5 minutes, with a half-life of 1.6 minutes at 23 °C (measured via SB, the only species in neutral pH, absorption spectra monitoring with 20 μM protein with 5 μM **CyThioPhenol** and plotted to second-order rate kinetics).

III.8.4 A. Dual hCRBPII labeling expressed in HeLa cells

We sought to examine the performance of dual fluorescent labeling of hCRBPII in live HeLa cells. Thus, **M3** and **M6** mutants were cloned into the pFlag-CMV2 vector containing EGFP fused to the N-terminus of the proteins.

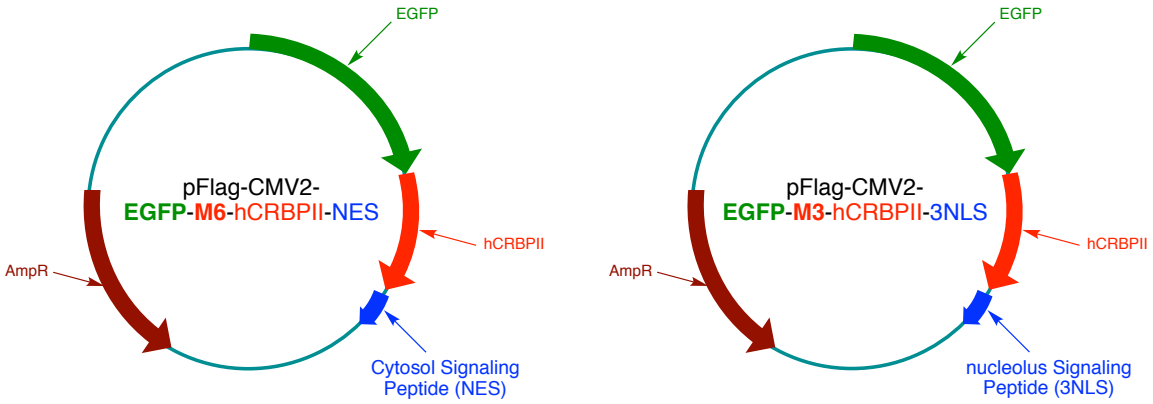


Figure III-55. Schematic map of EGFP-**M6**-hCRBPII-NES and EGFP-**M3**-hCRBPII-3NLS fusion constructs. NLS: nuclear localization sequence and NES: nuclear export sequence.

To image different cell organelles with two distinct colors, we aimed to express **M6** in cytosol and **M3** in the cell nucleus by fusing 3NLS and NES

localization peptides to the hCRBPII C-terminus, respectively (**Figure III-55**). The cells were co-transfected with the two vectors and then stained twice with 5 μM of **CyThioPhenol** each, first for an overnight incubation and second 10 minutes prior to imaging. The cells were then imaged without any washing steps.

First, we examine the performance of **CyThioPhenol/M3** as a fluorescent photobasic tag (**Figure III-56**). The cells were transfected solely with the EGFP-**M3**-hCRBPII-3NLS fusion construct and were imaged only after 5 minutes incubation of 5 μM **CyThioPhenol** in 37 °C. Cells were excited with a 405 nm laser, and the emission was collected 480 nm-580 nm bandpass (**Figure III-56**, the red channel).

Next, we were able to collect the two colors emission sequentially using the Olympus microscope virtual channel option. First, the **CyThioPhenol/M6** complex was excited via a 559 nm laser beam to collect the far-red emission in the 600 nm-700 nm bandpass, the red channel in **Figure III-57**. Then we collected the emission of **CyThioPhenol/M3** complex in the 480-580 nm window upon its excitation by the 405 nm laser (cyan channel, **Figure III-57**). Additionally, for both the red and cyan channels, the green fluorescence was collected to explore the **CyThioPhenol** cell permeabilization and its distribution in each organelle. Fortunately, as shown, no fluorescent background was detected in non-transfected cells. More importantly, none of the complexes showed fluorescence leakage to another organelle. As a proof of concept, this imaging assay successfully demonstrated the feasibility of labeling two sub-cellular compartments with

CyThioPhenol and two different hCRBPII tags that act as photoacid and photobase.

III.9 Conclusion and future research plans

This section described the photoacid and photobasic properties of a novel synthetic dye **ThioPhenol** and its cyclized derivative, **CyThioPhenol**, upon binding hCRBPII mutants. Additionally, we designed a protein/**ThioPhenol** complex for the first time that can act as both photoacid and photobase, able to transfer two protons upon a single photoirradiation called double ESPT. This system provides an unprecedently large Stokes shift (>240 nm), leading to well-separated absorption and emission and successful application in fluorescent labeling of the target protein. Notably, the cyclized derivative shows a higher fluorescence quantum yield and total brightness, and we illustrated its application as a practical fluorescent tag as a photobase.

Both chromophores developed as no-wash fluorescent tags with the minimal background because the structure is designed as the ICT, and fluorescence is activated upon binding the target protein. Besides, none of the chromophores display solvatochromic properties, which helps to eliminate washing steps in order to remove unbound chromophores.

Nonetheless, our efforts to develop a double ESPT system with **CyThioPhenol** were not fruitful. Future studies should be focused on developing the double ESPT with brighter cores such as the cyclized derivative or other novel designed structures potential that might find application in medicinal chemistry.

Additionally, crystallographic studies and ultrafast spectroscopy should be pursued to understand the mechanism of the double ESPT in the current system and evolve it with other structures.

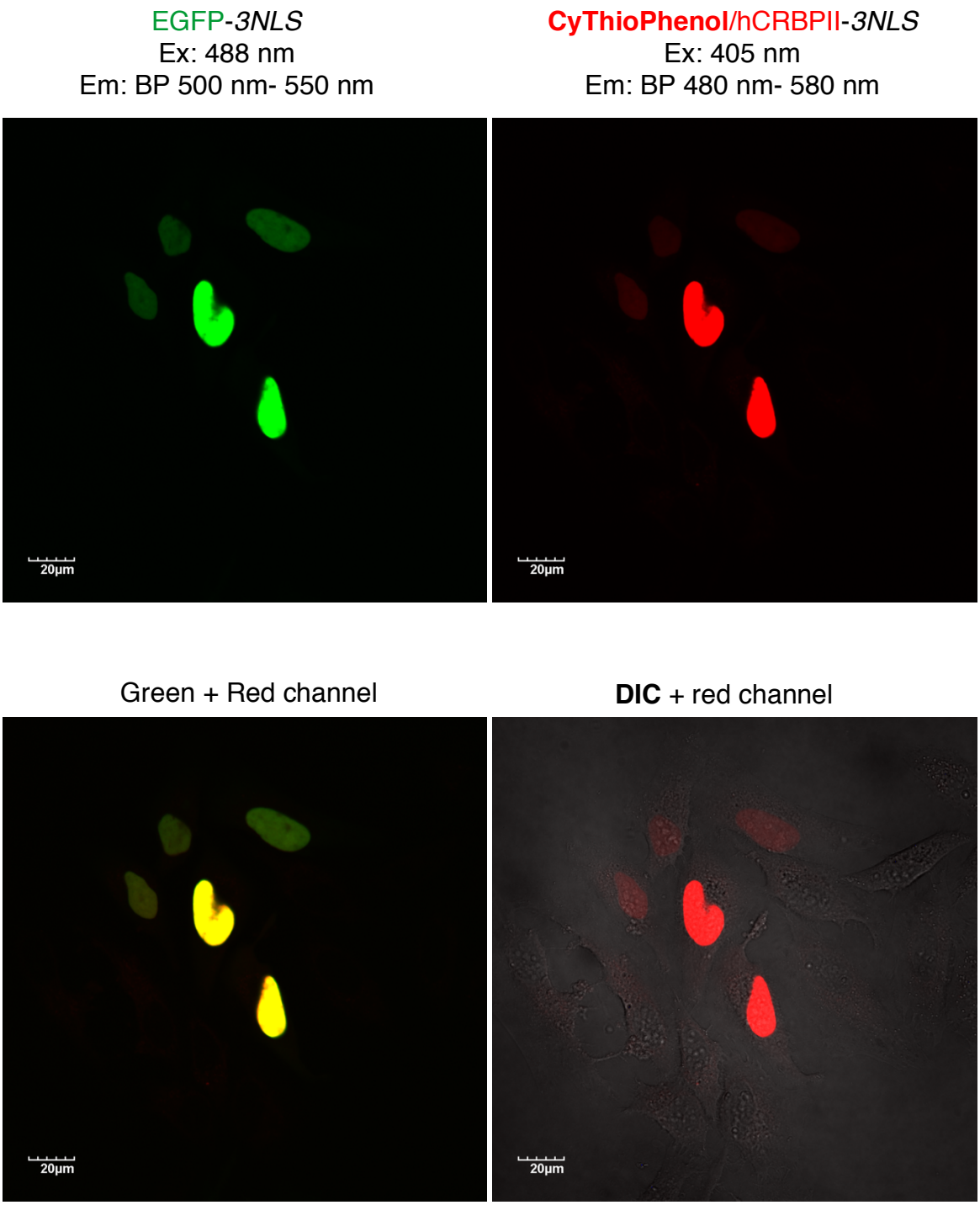
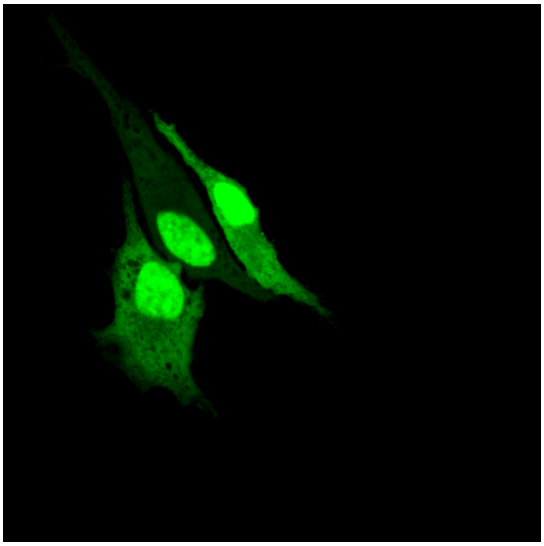
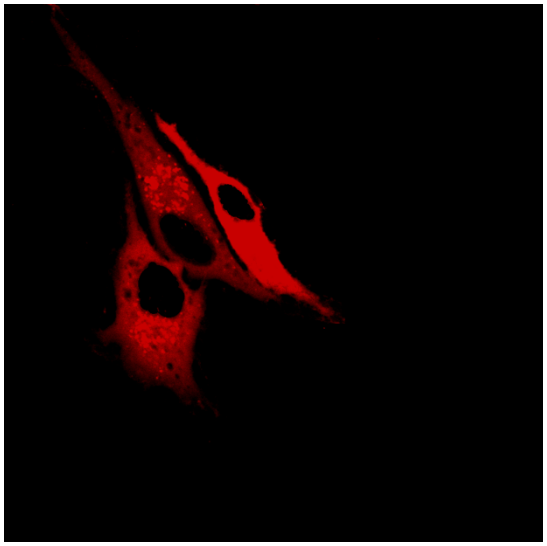


Figure III-56. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-*3NLS*. NLS = nuclear localization sequence. Cells were stained with 5 mM **CyThioPhenol** and incubated at 37 °C for 5 min. Cells were not washed before imaging.

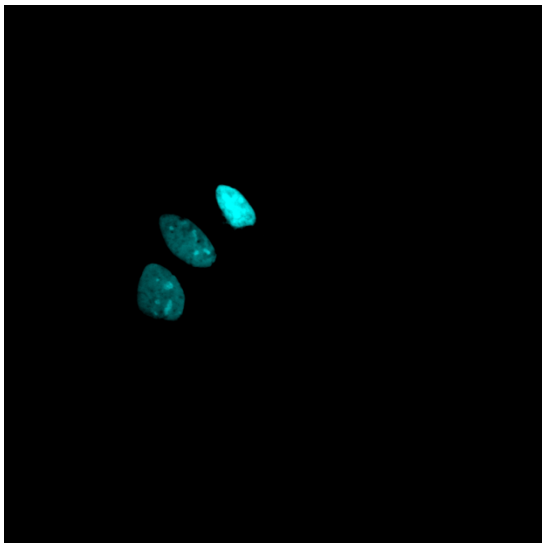
EGFP
Ex: 488 nm
Em: BP 500 nm- 550 nm



CyThioPhenol/hCRBPII-*NES*
Ex: 559 nm
Em: BP 600 nm- 700 nm



CyThioPhenol/hCRBPII-3*NLS*
Ex: 405 nm
Em: BP 480 nm- 580 nm



Cyan + Red channel

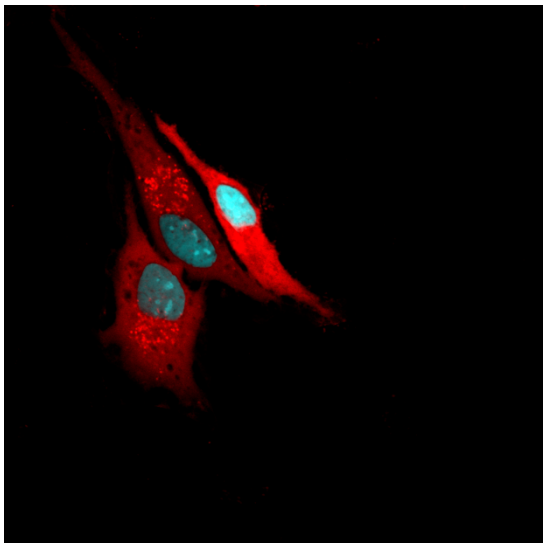


Figure III-57. Confocal imaging of labeled HeLa cells expressing EGFP-hCRBPII-*3NLS* and EGFP-hCRBPII-*NES*. NLS= nuclear localization sequence. NES= nuclear export sequence. Cells were stained with 5 μM **CyThioPhenol** and incubated at 37 °C for overnight and another 5 μM **CyThioPhenol** incubated at 37 °C for 5 minutes. Cells were not washed before imaging.

REFERENCES

REFERENCES

1 Huynh, M. H. V. & Meyer, T. J. Proton-coupled electron transfer. *Chem. Rev.* **107**, 5004-5064, doi:10.1021/cr0500030 (2007).

2 Lee, J., Kim, C. H. & Joo, T. Active Role of Proton in Excited State Intramolecular Proton Transfer Reaction. *J. Phys. Chem. A* **117**, 1400-1405, doi:10.1021/jp311884b (2013).

3 Joshi, H. C. & Antonov, L. Excited-State Intramolecular Proton Transfer: A Short Introductory Review. *Molecules* **26**, doi:10.3390/molecules26051475 (2021).

4 Amdursky, N., Simkovitch, R. & Huppert, D. Excited-State Proton Transfer of Photoacids Adsorbed on Biomaterials. *J. Phys. Chem. B* **118**, 13859-13869, doi:10.1021/jp509153r (2014).

5 Chowdhury, R. *et al.* Excited State Proton Transfer in the Lysosome of Live Lung Cells: Normal and Cancer Cells. *J. Phys. Chem. B* **119**, 2149-2156, doi:10.1021/jp503804y (2015).

6 Jacquemin, D., Zuniga, J., Requena, A. & Ceron-Carrasco, J. P. Assessing the Importance of Proton Transfer Reactions in DNA. *Accounts of Chemical Research* **47**, 2467-2474, doi:10.1021/ar500148c (2014).

7 Perun, S., Sobolewski, A. L. & Domcke, W. Role of electron-driven proton-transfer processes in the excited-state deactivation adenine-thymine base pair. *J. Phys. Chem. A* **110**, 9031-9038, doi:10.1021/jp061945r (2006).

8 Hayashi, S., Tajkhorshid, E. & Schulten, K. Structural changes during the formation of early intermediates in the bacteriorhodopsin photocycle. *Biophysical Journal* **83**, 1281-1297, doi:10.1016/s0006-3495(02)73900-3 (2002).

9 Kouyama, T., Kinosita, K. & Ikegami, A. EXCITED-STATE DYNAMICS OF BACTERIORHODOPSIN. *Biophysical Journal* **47**, 43-54, doi:10.1016/s0006-3495(85)83875-3 (1985).

10 Zhou, P. W. & Han, K. Unraveling the Detailed Mechanism of Excited-State Proton Transfer. *Accounts of Chemical Research* **51**, 1681-1690, doi:10.1021/acs.accounts.8b00172 (2018).

11 Mandal, D., Tahara, T. & Meech, S. R. Excited-state dynamics in the green fluorescent protein chromophore. *J. Phys. Chem. B* **108**, 1102-1108, doi:10.1021/jp035816b (2004).

12 Sedgwick, A. C. *et al.* Excited-state intramolecular proton-transfer (ESIPT) based fluorescence sensors and imaging agents. *Chem. Soc. Rev.* **47**, 8842-8880, doi:10.1039/c8cs00185e (2018).

13 Zhao, J. F., Dong, H. & Zheng, Y. J. Theoretical Insights Into the Excited State Double Proton Transfer Mechanism of Deep Red Pigment Alkannin. *J. Phys. Chem. A* **122**, 1200-1208, doi:10.1021/acs.jpca.7b10492 (2018).

14 Zhao, J. Z., Ji, S. M., Chen, Y. H., Guo, H. M. & Yang, P. Excited state intramolecular proton transfer (ESIPT): from principal photophysics to the development of new chromophores and applications in fluorescent molecular probes and luminescent materials. *Physical Chemistry Chemical Physics* **14**, 8803-8817, doi:10.1039/c2cp23144a (2012).

15 Kwon, J. E. & Park, S. Y. Advanced Organic Optoelectronic Materials: Harnessing Excited-State Intramolecular Proton Transfer (ESIPT) Process. *Adv. Mater.* **23**, 3615-3642, doi:10.1002/adma.201102046 (2011).

16 Tseng, H. W. *et al.* Harnessing Excited-State Intramolecular Proton-Transfer Reaction via a Series of Amino-Type Hydrogen-Bonding Molecules. *Journal of Physical Chemistry Letters* **6**, 1477-1486, doi:10.1021/acs.jpclett.5b00423 (2015).

17 Loken, M. R., Brand, L., Hayes, J. W. & Gohlke, J. R. EXCITED-STATE PROTON-TRANSFER AS A BIOLOGICAL PROBE - DETERMINATION OF RATE CONSTANTS BY MEANS OF NANOSECOND FLUOROMETRY. *Biochemistry* **11**, 4779-&, doi:10.1021/bi00775a022 (1972).

18 Bera, M. K. *et al.* Fluorene-based chemodosimeter for "turn-on" sensing of cyanide by hampering ESIPT and live cell imaging. *Journal of Materials Chemistry B* **2**, 4733-4739, doi:10.1039/c4tb00388h (2014).

19 Pines, E. & Huppert, D. PH JUMP - A RELAXATIONAL APPROACH. *Journal of Physical Chemistry* **87**, 4471-4478, doi:10.1021/j100245a029 (1983).

20 Gutman, M., Huppert, D. & Pines, E. THE PH JUMP - A RAPID MODULATION OF PH OF AQUEOUS-SOLUTIONS BY A LASER-PULSE. *Journal of the American Chemical Society* **103**, 3709-3713, doi:10.1021/ja00403a016 (1981).

21 Abbruzzetti, S. *et al.* Fast events in protein folding: Structural volume changes accompanying the early events in the N -> I transition of apomyoglobin induced by ultrafast pH jump. *Biophysical Journal* **78**, 405-415, doi:10.1016/s0006-3495(00)76603-3 (2000).

22 Vos, J. G. EXCITED-STATE ACID-BASE PROPERTIES OF INORGANIC-COMPOUNDS. *Polyhedron* **11**, 2285-2299, doi:10.1016/s0277-5387(00)83515-9 (1992).

23 Green, O., Gajst, O., Simkovitch, R., Shabat, D. & Huppert, D. New Phenol Benzoate Cyanine Picolinium Salt Photoacid Excited-State Proton Transfer. *J. Phys. Chem. A* **121**, 3079-3087, doi:10.1021/acs.jpca.7b02490 (2017).

24 Shizuka, H. EXCITED-STATE PROTON-TRANSFER REACTIONS AND PROTON-INDUCED QUENCHING OF AROMATIC-COMPOUNDS. *Accounts of Chemical Research* **18**, 141-147, doi:10.1021/ar00113a003 (1985).

25 Arnaut, L. G. & Formosinho, S. J. EXCITED-STATE PROTON-TRANSFER REACTIONS .1. FUNDAMENTALS AND INTERMOLECULAR REACTIONS. *J. Photochem. Photobiol. A-Chem.* **75**, 1-20, doi:10.1016/1010-6030(93)80157-5 (1993).

26 Wan, P. & Shukla, D. UTILITY OF ACID-BASE BEHAVIOR OF EXCITED-STATES OF ORGANIC-MOLECULES. *Chem. Rev.* **93**, 571-584, doi:10.1021/cr00017a024 (1993).

27 Charas, A. *et al.* Photoacid cross-linkable polyfluorenes for optoelectronics applications. *Synth. Met.* **158**, 643-653, doi:10.1016/j.synthmet.2008.02.016 (2008).

28 Tolbert, L. M. & Solntsev, K. M. Excited-state proton transfer: From constrained systems to "super" photoacids to superfast proton transfer. *Accounts of Chemical Research* **35**, 19-27, doi:10.1021/ar990109f (2002).

29 White, W., Sanborn, C. D., Fabian, D. M. & Ardo, S. Conversion of Visible Light into Ionic Power Using Photoacid-Dye-Sensitized Bipolar Ion-Exchange Membranes. *Joule* **2**, 94-109, doi:10.1016/j.joule.2017.10.015 (2018).

30 Piatkevich, K. D. *et al.* Extended Stokes Shift in Fluorescent Proteins: Chromophore-Protein Interactions in a Near-Infrared TagRFP675 Variant. *Scientific Reports* **3**, doi:10.1038/srep01847 (2013).

31 Fron, E. *et al.* Mechanism Behind the Apparent Large Stokes Shift in LSSmOrange Investigated by Time-Resolved Spectroscopy. *J. Phys. Chem. B* **119**, 14880-14891, doi:10.1021/acs.jpcb.5b09189 (2015).

32 Piatkevich, K. D. *et al.* 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, doi:10.1073/pnas.0914365107 (2010).

33 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, doi:10.1021/ja101974k (2010).

34 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-+, doi:10.1021/ja903695n (2009).

35 Henderson, J. N. *et al.* Excited State Proton Transfer in the Red Fluorescent Protein mKeima. *Journal of the American Chemical Society* **131**, 13212-13213, doi:10.1021/ja904665x (2009).

36 Ryan, E. T., Xiang, T., Johnston, K. P. & Fox, M. A. Absorption and fluorescence studies of acridine in subcritical and supercritical water. *Journal of Physical Chemistry A* **101**, 1827-1835, doi:10.1021/jp962676f (1997).

37 Pines, E., Huppert, D., Gutman, M., Nachliel, N. & Fishman, M. THE POH JUMP - DETERMINATION OF DEPROTONATION RATES OF WATER BY 6-METHOXYQUINOLINE AND ACRIDINE. *Journal of Physical Chemistry* **90**, 6366-6370, doi:10.1021/j100281a061 (1986).

38 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, doi:10.1021/jp970581n (1997).

39 Favaro, G., Mazzucat.U & Masetti, F. EXCITED-STATE REACTIVITY OF AZA AROMATICS .1. BASICITY OF 3-STYRYLPYRIDINES IN FIRST EXCITED SINGLET-STATE. *Journal of Physical Chemistry* **77**, 601-604, doi:10.1021/j100624a007 (1973).

40 Jimenez-Sanchez, A. & Santillan, R. A photochromic-acidochromic HCl fluorescent probe. An unexpected chloride-directed recognition. *Analyst* **141**, 4108-4120, doi:10.1039/c6an00509h (2016).

41 Munitz, N., Avital, Y., Pines, D., Nibbering, E. T. J. & Pines, E. Cation-Enhanced Deprotonation of Water by a Strong Photobase. *Israel Journal of Chemistry* **49**, 261-272, doi:10.1560/ijc.49.2.261 (2009).

42 Suyama, K. & Shirai, M. Photobase generators: Recent progress and application trend in polymer systems. *Progress in Polymer Science* **34**, 194-209, doi:10.1016/j.progpolymsci.2008.08.005 (2009).

43 Sheng, W. *et al.* Ultrafast Dynamics of a "Super" Photobase. *Angewandte Chemie-International Edition* **57**, 14742-14746, doi:10.1002/anie.201806787 (2018).

44 Sheng, W. *et al.* A Near-Infrared Photoswitchable Protein-Fluorophore Tag for No-Wash Live Cell Imaging. *Angewandte Chemie-International Edition* **57**, 16083-16087, doi:10.1002/anie.201810065 (2018).

45 Sheng, W. In PhD Thesis. *Michigan State University* (2019).

46 Santos, E. M. in PhD thesis. *Michigan State University* (2017).

CHAPTER IV: MATERIALS AND METHODS

IV.1 Site-directed mutagenesis of hCRBPII and CRABPII

Table IV-1. PCR cycling conditions for site-directed mutagenesis.

	PCR Program	Time (min)
1x	94 °C	3:00
30x	94 °C	00:20
	3-5 °C below T_m	00:55
	72 °C	03:30
1x	72 °C	10:00
1x	4 °C	5:00

Reactant	volume
DNA (Template Plasmid)	70 ng (x µl)
Primer forward	20 pmol (y µl)
Primer reverse	20 pmol (z µl)
10 mM dNTP	1 µL
DMSO	5 µL
50 mM $MgCl_2$	5 µL
10 x Cloned Pfu Reaction Buffer	5 µL
Pfu Turbo DNA Polymerase (2.5 U/µl)	1 µL
DI Water	50 – x – y – z – 17 µL

All DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT), with melting temperatures (Tm) from approximately 55 °C to 65 °C (depending on primer's sequence). All *E. coli* constructs were in the pET-17b vector (Addgene). The pET-17b plasmid, containing hCRBPII- Q108K:K40L

cloned between NdeI and XhoI, was used as a template for mutagenesis of hCRBPII.[1,2]

The template was used to do single point mutations on CRABPII was the pET-17b plasmid containing wild-type CRABPII cloned between NdeI and EcoRI. Site-directed mutagenesis was conducted via polymerase chain reaction (PCR), using PfuTurbo DNA Polymerase (Agilent) following the specified cycling conditions shown in **Table IV-1**.

The sequences of the forward primers are listed below. It is worth noting that, in all cases, the reverse primer is the reverse complement of the forward primer. All primers correspond to hCRBPII unless otherwise indicated.

Q4C: 5'-CG AGG GAC **TGC** AAT GGA ACC TGG-3'

Q4E: 5'-ACG AGG GAC **GAA** AAT GGA ACC TGG-3'

Q4F: 5'-G ACG AGG GAC **TTC** AAT GGA ACC-3'

Q4K: 5'-G ACG AGG GAC **AAG** AAT GGA ACC TGG G-3'

Q4S: 5'-CG AGG GAC **AGC** AAT GGA ACC TGG-3'

Q4T: 5'-G ACG AGG GAC **ACA** AAT GGA ACC-3'

Q4Y: 5'-CG AGG GAC **TAC** AAT GGA ACC TGG GAG-3'

F16C: 5'-GAG AGT AAT GAA AAC **TGC** GAG GGC TAC ATG-3'

F16C:Y19W: 5'-G AGT AAT GAA AAC **TGC** GAG GGC TGG ATG-3'

F16E: 5'-GAG AGT AAT GAA AAC **GAG** GAG GGC TAC ATG-3'

F16E:Y19W: 5'-G AGT AAT GAA AAC **GAG** GAG GGC TGG ATG-3'

F16F:Y19W: 5'-GAG AGT AAT GAA AAC **TTT** GAG GGC TGG ATG-3'

F16H: 5’-GAG AGT AAT GAA AAC **CAC** GAG GGC TAC ATG-3’

F16H:Y19W: 5’-G AGT AAT GAA AAC **CAC** GAG GGC TGG ATG-3’

F16W: 5’-G AGT AAT GAA AAC **TGG** GAG GGC TAC ATG-3’

F16W:Y19W: 5’-AGT AAT GAA AAC **TGG** GAG GGC TGG ATG AAG GCC-3’

F16Y: 5’-G AGT AAT GAA AAC **TAT** GAG GGC TAC ATG-3’

F16Y:Y19W: 5’-GA GAGT AAT GAA AAC **TAT** GAG GGC TGG ATG-3’

Y19L: 5’-TTT GAG GGC **CTG** ATG AAG GCC-3’

Y19W: 5’-C TTT GAG GGC **TGG** ATG AAG GCC CTG-3’

F16Y:**Y19W**: 5’-C TAT GAG GGC **TGG** ATG AAG GCC CTG-3’

Y19W:**M20D**: 5'-C TTT GAG GGC TGG **GAT** AAG GCC CTG-3'

Y19W:**M20E**: 5'-C TTT GAG GGC TGG **GAG** AAG GCC CTG-3'

Y19W:**M20H**: 5’-GAG GGC TGG **CAT** AAG GCC CTG-3’

Y19W:**M20K**: 5’-GAG GGC TGG **AAA** AAG GCC CTG-3’

M20L: 5’-GAG GGC TAC **CTG** AAG GCC CTG-3’

Y19W:**M20R**: 5’-GAG GGC TGG **AGA** AAG GCC CTG-3’

M20Y: 5’-GAG GGC TAC **TAC** AAG GCC CTG G-3’

T29D: 5'-GAT TTT GCC **GAT** CGC AAG ATT GC-3'

T29E: 5'-GAT TTT GCC **GAG** CGC AAG ATT GC-3'

T29F:A33H: 5'-G GAT ATT GAT TTT GCC **TTC** CGC AAG ATT CAC-3'

T29H:A33W: 5’-GAT TTT GCC **CAC** CGC AAG ATT TGG-3’

T29K:A33W: 5’-ATT GAT TTT GCC **AAG** CGC AAG ATT TGG-3’

T29R:A33W: 5’-GAT TTT GCC **CGG** CGC AAG ATT TGG-3’

T29W:A33H: 5’-GAT TTT GCC **TGG** CGC AAG ATT CAC-3’

T29Y: 5’-G GAT ATT GAT TTT GCC **TAC** CGC AAG ATT GC-3’

A33C: 5’-C CGC AAG ATT **TGC** GTA CGT C-3’

A33D: 5’-ACC CGC AAG ATT **GAT** GTA CGT CTC-3’

A33E: 5’-CGC AAG ATT **GAG** GTA CGT CTC AC-3’

A33H: 5’-ACC CGC AAG ATT **CAC** GTA CGT CTC-3’

A33K: 5’-GCC ACC CGC AAG ATT **AAA** GTA CGT CTC-3’

A33M: 5’-CC ACC CGC AAG ATT **ATG** GTA CGT CTC AC-3’

A33R: 5’-ACC CGC AAG ATT **CGT** GTA CGT CTC-3’

A33S: 5’-C CGC AAG ATT **AGC** GTA CGT CTC AC-3’

A33W: 5’-CGC AAG ATT **TGG** GTA CGT CTC AC-3’

A33Y: 5’-ACC CGC AAG ATT **TAC** GTA CGT CTC ACT-3’

Q38E:K40L: 5’-GTA CGT CTC ACT **GAG** ACG CTG GTT ATT GAT CAA-3’

Q38F:K40L: 5’-GTA CGT CTC ACT **TTT** ACG CTT GTT ATT GAT C-3’

Q38H:K40E: 5’-GCA GTA CGT CTC ACT **CAC** ACG GAG GTT ATT G-3’

Q38H:K40L: 5’-GTA CGT CTC ACT **CAC** ACG CTG GTT ATT GAT-3’

Q38P: 5’-GTA CGT CTC ACT **CCG** ACG AAG GTT ATT G-3’

Q38Q: 5’-GTA CGT CTC ACT **CAG** ACG AAG GTT ATT GAT-3’

Q38Y:K40E: 5’-GCA GTA CGT CTC ACT **TAC** ACG GAG GTT ATT G-3’

Q38Y:K40L: 5’-GCA GTA CGT CTC ACT **TAC** ACG CTG GTT ATT GAT-3’

K40A: 5’-CTC ACT CAG ACG **GCT** GTT ATT GAT CAA-3’

K40C: 5’-CTC ACT CAG ACG **TGC** GTT ATT GAT CAA G-3’

Q38F:**K40C**: 5’-GTA CGT CTC ACT TTT ACG **TGT** GTT ATT GAT-3’

K40D: 5’-CTC ACT CAG ACG **GAT** GTT ATT GAT CAA GAT GG-3’

Q38F:**K40D**: 5’-GTA CGT CTC ACT TTT ACG **GAC** GTT ATT GAT CAA-3’

K40E: 5’-CTC ACT CAG ACG **GAG** GTT ATT GAT CAA-3’

Q38F:**K40E**: 5’-GTA CGT CTC ACT TTT ACG **GAA** GTT ATT GAT CAA-3’

K40H: 5’-CTC ACT CAG ACG **CAC** GTT ATT GAT CAA-3’

Q38F:**K40H**: 5’-GTA CGT CTC ACT TTT ACG **CAT** GTT ATT GAT-3’

K40K: 5’-CGT CTC ACT CAG ACG **AAG** GTT ATT GAT CAA GAT-3’

Q38F:**K40K**: 5’-GTA CGT CTC ACT TTT ACG **AAG** GTT ATT-3’

Q38F:**K40L**: 5’-CGT CTC ACT TTC ACG **CTG** GTT ATT GAT C-3’

Q38F:**K40M**: 5’-GTA CGT CTC ACT TTC ACG **ATG** GTT ATT-3’

K40N: 5’-CTC ACT CAG ACG **AAC** GTT ATT GAT CAA G-3’

Q38F:**K40N**: 5’-CTC ACT TTT ACG **AAC** GTT ATT GAT CAA GAT GG-3’

Q38F:**K40P**: 5’-C ACT TTT ACG **CCG** GTT ATT GAT CAA GAT GG-3’

K40Q: 5’-CTC ACT CAG ACG **CAG** GTT ATT GAT CAA G-3’

Q38F:**K40Q**: 5’-GTA CGT CTC ACT TTT ACG **CAA** GTT ATT GAT-3’

Q38F:**K40R**: 5’-GTA CGT CTC ACT TTT ACG **CGA** GTT ATT GAT-3’

K40S: 5’-CTC ACT CAG ACG **TCG** GTT ATT GAT CAA GAT GG-3’

Q38F:**K40S**: 5’-GTA CGT CTC ACT TTT ACG **TCT** GTT ATT GAT-3’

K40T: 5’-CTC ACT CAG ACG **ACG** GTT ATT GAT CAA G-3’

Q38F:**K40T**: 5’-GTA CGT CTC ACT TTT ACG **ACA** GTT ATT-3’

K40Y: 5’-CTC ACT CAG ACG **TAC** GTT ATT GAT CAA GAT GG-3’

Q38F:**K40Y**: 5’-GTA CGT CTC ACT TTT ACG **TAC** GTT ATT-3’

K40L:**I42E**: 5’-ACG CTG GTT **GAA** GAT CAA GAT GGT-3’

K40H:**I42F**: 5’-ACG CAT GTT **TTC** GAT CAA GAT GGT GAT-3’

K40L:**I42K**: 5’-CAG ACG CTG GTT **AAG** GAT CAA GAT GGT G-3’

K40H:**I42P**: 5’-ACG CAT GTT **CCC** GAT CAA GAT GGT-3’

K40H:**I42W**: 5’-ACG CAT GTT **TGG** GAT CAA GAT GGT GAT-3’

Q44K: 5’-CT CAG ACG CTG GTT ATT GAT **AAG** GAT GGT GAT AAC-3’

F49K: 5’-GAT CAA GAT GGT GAT AAC AAG **AAG** GTA AAA AGC AC-3’

T51C:T53S: 5’-GGT GAT AAC TTC AAG **TGC** AAA AGC ACT AGC AC-3’

T51E:T53E: 5’-GGT GAT AAC TTC AAG **GAG** AAA GAG ACT AGC-3’

T51M:T53A: 5’-GGT GAT AAC TTC AAG **ATG** AAA GCG ACT AGC-3’

T51M:T53S: 5’-GGT GAT AAC TTC AAG **ATG** AAA AGC ACT AGC AC-3’

T51T:T53S: 5’-GGT GAT AAC TTC AAG **ACA** AAA AGC ACT AGC AC-3’

T51T**:T53A**: 5’-C TTC AAG ACA AAA **GCG** ACT AGC ACA TTC CG-3’

T51V:**T53C**: 5’-C TTC AAG GTA AAA **TGC** ACT AGC ACA TTC CAC-3’

T53D: 5’-C TTC AAG ACA AAA **GAT** ACT AGC ACA TTC CG-3’

T53E: 5’-TTC AAG ACA AAA **GAG** ACT AGC ACA TTC-3’

T51V:**T53E**: 5’-AAC TTC AAG GTA AAA **GAG** ACT AGC ACA TTC CAC AAC-3’

T51V:**T53H**: 5’-AAC TTC AAG GTA AAA **CAC** ACT AGC ACA TTC CAC AAC-3’

T51V:**T53S**: 5’-AAC TTC AAG GTA AAA **AGC** ACT AGC ACA TTC CAC AAC-3’

T51V:**T53Y**: 5’-AAC TTC AAG GTA AAA **TAC** ACT AGC ACA TTC CAC AAC TAT-3’

T51V:T53S:**S55E**: 5’-AAG GTA AAA AGC ACT **GAG** ACA TTC CAC AAC TAT GAT-3’

T51V:T53S:**S55H**: 5’-AAG GTA AAA AGC ACT **CAC** ACA TTC CAC AAC TAT GAT-3’

T51V:T53S:**S55S** :5’-GTA AAA AGC ACT **AGC** ACA TTC CAC AAC TAT GAT GTG-3’

T51V:T53S:**S55Y**: 5’-AAG GTA AAA AGC ACT **TAC** ACA TTC CAC AAC TAT GAT GTG-3’

R58A: 5'-CT AGC ACA TTC **GCG** AAC TAT GAT GTG-3'

R58C: 5’-GC ACT AGC ACA TTC **TGC** AAC TAT GAT GTG G-3’

R58E: 5’-CT AGC ACA TTC **GAG** AAC TAT GAT GTG-3’

R58H: 5’-CT AGC ACA TTC **CAC** AAC TAT GAT GTG-3’

R58K: 5’-AGC ACA TTC **AAG** AAC TAT GAT GTG-3’

R58L: 5’-CT AGC ACA TTC **CTG** AAC TAT GAT GTG-3’

R58M: 5’-CC ACT AGC ACA TTC **ATG** AAC TAT GAT GTG G-3’

R58Q: 5’-CT AGC ACA TTC **CAG** AAC TAT GAT GTG-3’

R58R: 5’-C ACT AGC ACA TTC **CGC** AAC TAT GAT GTG GAT TTC A-3’

R58S: 5’-ACC ACT AGC ACA TTC **TCA** AAC TAT GAT GTG GAT-3’

R58T: 5’-CC ACT AGC ACA TTC **ACG** AAC TAT GAT GTG G-3’

R58W: 5’-CT AGC ACA TTC **TGG** AAC TAT GAT GTG-3’

R58Y: 5’-CT AGC ACA TTC **TAC** AAC TAT GAT GTG-3’

Y60L: 5’-TTC CGC AAC **CTG** GAT GTG GAT-3’

V62Y: 5’-C CGC AAC TAT GAT **TAC** GAT TTC ACT GTTG G-3’

V62F: 5’-C CGC AAC TAT GAT **TTC** GAT TTC ACT GTT GG-3’

V62Y: 5’-C CGC AAC TAT GAT **TAC** GAT TTC ACT GTT GG-3’

F64E: 5’-AAC TAT GAT GTG GAT **GAA** ACT GTT GGA GTA GAG-3’

S76D: 5’-TAC ACA AAG **GAC** CTG GAT AAC CGG-3’

S76E: 5’-TAC ACA AAG **GAG** CTG GAT AAC CGG-3’

S76G: 5’-GAG TAC ACA AAG **GGC** CTG GAT AAC CGG-3’

S76H: 5’-TAC ACA AAG **CAC** CTG GAT AAC-3’

S76K: 5’-GAG TAC ACA AAG **AAG** CTG GAT AAC-3’

S76R: 5’-TAC ACA AAG **CGA** CTG GAT AAC-3’

S76V: 5’-GAG TAC ACA AAG **GTC** CTG GAT AAC-3’

L77A: 5’-C ACA AAG AGC **GCA** GAT AAC CGG C-3’

L77C: 5’-G TAC ACA AAG AGC **TGC** GAT AAC CGG CAT G-3’

L77D: 5’-G TAC ACA AAG AGC **GAT** GAT AAC CGG-3’

L77E: 5’-G TAC ACA AAG AGC **GAG** GAT AAC CGG-3’

L77F: 5’-G TAC ACA AAG AGC **TTC** GAT AAC CGG CAT G-3’

L77H: 5’-ACA AAG AGC **CAC** GAT AAC CGG CAT-3’

S76G:**L77I**: 5’-C ACA AAG GGC **ATC** GAT AAC CGG CAT G-3’

L77K: 5’-ACA AAG AGC **AAA** GAT AAC CGA-3’

L77M: 5’-G TAC ACA AAG AGC **ATG** GAT AAC CGG C-3’

L77R: 5’-ACA AAG AGC **CGG** GAT AAC CGG-3’

L77S: 5’-C ACA AAG AGC **TCA** GAT AAC CGG C-3’

L77W: 5’-G TAC ACA AAG AGC **TGG** GAT AAC CGG CAT G-3’

L77Y: 5’-G TAC ACA AAG AGC **TAC** GAT AAC CGG CAT G-3’

L93E: 5’-TGG GAA GGT GAT GTC **GAA** GTG TGT GTG CAA AAG-3’

L93K: 5’-GG GAA GGT GAT GTC **AAG** GTG TGT GTG-3’

Q108A: 5’-GGC TGG AAG **GCC** TGG ATT GAG G-3’

Q108L: 5’-GGC TGG AAG **CTG** TGG ATT GAG G-3’

Q108Q: 5’-C GGC TGG AAG **CAA** TGG ATT GAG G-3’

Q108A:**I110K**: 5’-GGC TGG AAG GCC TGG **AAG** GAG G-3’

Q108Q:**I110K**: 5’-GGC TGG AAG CAG TGG **AAA** GAG GGG GAC-3’

I110K: 5’-G AAG CAA TGG **AAG** GAG GGG GAC AAG-3’

L115E: 5’-GGG GAC AAG **GAG** TAC CTG GAG C-3’

L117C: 5’-GAC AAG CTG TAC **TGT** GAG CTG ACC TGT-3’

L117D: 5’-C AAG CTG TAC **GAC** GAG CTG ACC-3’

L117E: 5’-C AAG CTG TAC **GAG** GAG CTG ACC-3’

L117H: 5’-GAC AAG CTG TAC **CAC** GAG CTG-3’

L117M: 5’-GAC AAG CTG TAC **ATG** GAG CTG ACC TG-3’

L117Q: 5’-C AAG CTG TAC **CAG** GAG CTG ACC-3’

L117S: 5’-GAC AAG CTG TAC **AGC** GAG CTG ACC TG-3’

L117T: 5’-C AAG CTG TAC **ACA** GAG CTG ACC-3'

L117Y: 5’-GAC AAG CTG TAC **TAC** GAG CTG ACC TGT G-3’

L119E: 5’-G CTG TAC CTG GAG **GAG** ACC TGT GGT GAC-3’

L119D: 5’-G CTG TAC CTG GAG **GAT** ACC TGT GGT GAC-3’

Q128E: 5’-CAG GTG TGC CGT **GAG** GTG TTC AAA AAG-3’

Q128F: 5’-GTG TGC CGT **TTT** GTG TTC AAA-3’

Q128H: 5’-CAG GTG TGC CGT **CAC** GTG TTC AAA AAG-3’

Q128M: 5’-GTG TGC CGT **ATG** GTG TTC AAA-3’

Q128S: 5’-CAG GTG TGC CGT **AGC** GTG TTC AAA AAG-3’

Q128T: 5’-CAG GTG TGC CGT **ACC** GTG TTC AAA AAG-3’

Q128Y: 5’-CAG GTG TGC CGT **TAC** GTG TTC AAA AAG AAG-3’

F130C: 5’-CAG GTG TGC CGT CAA GTG **GAG** AAA AAG AAG-3’

F130E: 5’-GC CGT CAA GTG **TGC** AAA AAG AAG TG-3’

F130H: 5’-GC CGT CAA GTG **CAC** AAA AAG AAG TG-3’

F130Q: 5’-G TGC CGT CAA GTG **CAG** AAA AAG AAG TG-3’

F130S: 5’-GC CGT CAA GTG **TCC** AAA AAG AAG TG-3’

F130T: 5’-GC CGT CAA GTG **ACC** AAA AAG AAG TG-3’

F130Y: 5’-CAG GTG TGC CGT CAA GTG **TAC** AAA AAG AAG-3’

V41E *(CRABPII)*: 5’-CC AAG CCA GCA **GAG** GAG ATC AAA CAG G-3’

V41I *(CRABPII)*: 5’-GCG TCC AAG CCA GCA **ATC** GAG ATC AAA C-3’

V41Q *(CRABPII)*: 5’-CC AAG CCA GCA **CAG** GAG ATC AAA CAG-3’

T54D *(CRABPII)*: 5’-C TAC ATC AAA **GAC** TCC ACC ACC GTG C-3’

R59E *(CRABPII)*: 5'-C ACC ACC GTG **GAG** ACC ACA GAG-3'

L121N *(CRABPII)*: 5'-GG GAA CTG ATC **AAC** ACC ATG ACG GCG-3'

L121Q *(CRABPII)*: 5'-GAA CTG ATC **CAG** ACC ATG ACG GCG-3'

L121W *(CRABPII)*: 5'-GGG GAA CTG ATC **TGG** ACC ATG ACG-3'

To digest the original template DNA, the crude PCR products were subjected to 20 units of DpnI enzyme (New England Biolabs), and were incubated at 37 °C for one h. The resulting solution (15 µL) was then added to *E. coli* XL-1 Blue competent cells (Novagen, 50 µL), and incubated on ice for at least 30 min. Subsequently, the solution was spread gently 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 20 h.

A single and well-isolated colony was then picked from the agar plate and inoculated into 10 mL LB media containing 100 µg/mL ampicillin and 12.5 µg/mL tetracycline. LB media was prepared by adding 10 g tryptone, 10 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 LB media was shaken at 37 °C for approximately 20 h. Cells were harvested via centrifugations at 5000 rpm for 10 min. PCR-amplified DNA was purified by a Promega Wizard Plus SV miniprep DNA purification system (A1330) following the manufacturers' directions with the exception of using 45 µL of Nuclease-Free water for DNA elution instead of the recommended 100 µL to yield more concentrated DNA product.

The isolated plasmids' concentration was measured via Thermo Scientific NanoDrop™ 1000 Spectrophotometer; the average concentration was 300 ng/µL in the 45 µL solution. A sample containing at least 700 ng (about 8 µL) of purified plasmid was transferred into another Eppendorf tube and sequenced by The Research Technology Support Facility at Michigan State University using a primer corresponding to the T7 promoter for all pET-17b plasmid.

IV.2 hCRBPII and CRABPII expression and purification in pET-17b expression plasmids

The target gene (100 ng of DNA for 100 µL cell solution) was added to thawed BL21(DE3) pLysS competent cells (Invitrogen™) *E. coli* competent cells on ice and incubated for 30 min, and subsequently, the cells were 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.

In order to grow a cell culture, 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_2PO_4 and 12.54 g of K_2HPO_4 in 90 mL DI water. Before inoculation, the solutions were mixed. The inoculated culture was shaken at 37 °C at 250 rpm until optical density (OD) at 600 nm was in the range of 0.75 - 0.95; this typically takes 7 to 9 h. Overexpression was induced by adding 1 mL of 1 M isopropyl-β-D-

thiogalactopyranoside (IPTG, Gold Biotechnology), resulting in final concentration of 1 mM. The culture was then shaken at 23 °C for 20 h at 225 rpm.

The cells were then harvested by centrifugation (5000 rpm, 12 min, 4 °C). The supernatant was discarded, and the cells were resuspended in 50 mL Tris-binding buffer (10 mM Tris, pH=7.8-8.0). The cells were then lysed by sonication (Model 300 V/T Ultrasonic Homogenizer, Biologics Inc, power 60%, 3 min), and treated with DNAase (300 units/ 50 mL suspension) and $MgCl_2$ (0.12 mmol/ 50 mL suspension). The solution was again centrifuged to separate the pellet and supernatant (5000 rpm, 40 min, 4°C). All further protein purification was also conducted at 4 °C.

The protein in the supernatant was purified as follow: The supernatant was loaded onto an anion exchange column (Q Sepharose ™ Fast Flow resin, GE Healthcare, column diameter: ~ 4 cm; height: ~10 cm), pre-equilibrated with the Tris buffer (10 mM Tris, pH= 7.8-8.0). After binding the protein to the Fast Q anion exchange resin, the column was washed twice with the Tris buffer (100 mL).

The protein was eluted with Tris-elution buffer (100 mL, 10 mM Tris, 200 mM sodium chloride, pH=8.0). The eluent from the Fast Q anion exchange column was 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, filter code: YM10). The protein was first concentrated to ~ 35 mL and then diluted to 150 mL with Tris buffer. This solution was concentrated again to ~ 30 mL.

Further 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 IV-2**. The pH at all steps was set to 8.1% B corresponds to the percent salt, where 100% is equivalent to 1 mM NaCl.

Table IV-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

Table IV-3. 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

Next, all tubes with 40 mM, 80 mM, and 150 mM of NaCl, which have higher intensity for 280 nm peak (green) in comparison to 260 nm peak (purple), were collected. Then the solution was 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 size exclusion is shown in **Table IV-3**. At all steps, the pH was set to 8.1.

IV.3 Protein characterization

IV.3.1 UV/Vis and fluorescence spectroscopy

Spectroscopic characterizations of purified proteins were carried out using a Cary 300 Bio UV-Visible spectrophotometer (Varian) using 1-cm, 1.0-mL quartz micro cuvettes (Starna Cells). Fluorescence spectroscopy was performed on a Fluorolog®-3 spectrofluorometer (Jobin Yvon, Horiba Scientific) with 1-cm, 3.5-mL quartz cuvettes, or 1-cm, 1.0-mL quartz micro cuvettes (Starna Cells). An entrance slit of 2 nm and an exit slit of 2 nm was used for all measurements. For all experiments, Protein samples (20 μM) were incubated with ligand (0.5 equiv) in 2X PBS buffer and incubated at room temperature until Schiff base (SB) or protonated Schiff base (PSB) formation was complete. This was verified by the protein/chromophore complex UV-Vis spectrum.

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.

IV.3.2 Extinction coefficient determination

The extinction coefficients of the proteins were measured at 280 nm, as previously described by Gill and Von Hippel.[3] The theoretical extinction (ε^{Theor}) coefficient is calculated based on the following formula:

$$\varepsilon^{Theor} = a \times \varepsilon^{Trp} + b \times \varepsilon^{Tyr} + c \times \varepsilon^{Cys}$$

where *a*, *b* and *c* are the numbers of tryptophans, tyrosines, and cysteine residues, respectively. The extinction coefficients of the three residues were determined at 280 nm previously (ε^{Trp} = 5,690 $M^{-1}\cdot cm^{-1}$, ε^{Tyr} = 1,280 $M^{-1}\cdot cm^{-1}$, ε^{Cys} = 120 $M^{-1}\cdot cm^{-1}$).

The protein absorption was measured at 280 nm 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.

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

The concentration of the protein can be measured via the Lambert-Beer's equation. where *b* is the cuvette path length, and *c* is the concentration of the protein.

$$A_{native} = \varepsilon^{exp} \times b \times c$$

The extinction coefficients of all purified hCRBPII and CRABPII monomers described in this thesis are listed in **Table IV-4**. All proteins are hCRBPII mutants unless otherwise noted. For the proteins forming a stable dimer, the extinction coefficient of the dimer is calculated as well. (The extinction coefficient values are upon binding **ThioPhenol**).

Table IV-4. Extinction coefficients of hCRBPII and CRABPII mutants.

Protein	$\varepsilon_{exp\ (280\ nm)}$	
	Monomer	Dimer
Q108K:K40L:T51V:T53S	30,888	
Q108K:K40L:T51V:T53S:R58W	33,984	
Q108K:K40L:T51V:T53S:R58H	28,504	
Q108K:K40L:T51V:T53S:R58E	28,960	
Q108K:K40L:T51V:T53S:R58L	29,079	
Q108K:K40L:T51V:T53S:R58Q	29,013	
Q108K:K40L:T51V:T53S:R58K	28,640	
Q108K:K40L:T51V:T53S:R58C	26,929	
Q108K:K40L:T51V:T53S:R58S	28,562	
Q108K:K40L:T51V:T53S:R58Y	28,466	

Table IV-4 (cont'd)

Q108K:K40L:Y19W:T51V:T53S:R58W	36,880
Q108K:K40L:Y19W:T51V:T53S:R58H	34,571
Q108K:K40L:Y19W:A33W:T51V:T53S:R58W	43,042
Q108K:K40L:Y19W:A33W:T51V:T53S:R58H	37,958
Q108K:K40L:Y19W:A33W:T51V:T53S:R58W:L117E	42,130
Q108K:K40L:Y19W:A33W:T51V:T53S:R58H:L117E	38,408
Q108K:K40L:Q4F:Y19W:T51V:T53S:R58W: L117D	38,135
Q108K:K40L:Q4F:Y19W:T51V:T53S:R58H:L117D	31920
Q108K:K40L:Y19W:T51V:T53S:R58W:L117E	35721
Q108K:K40L:Y19W:T51V:T53S:R58H:L117E	32192
Q108K:K40E:Q4F:Q38F:T53A:R58H	24,391
Q108K:K40E:Q4F:Q38F:T53A:R58L	27,681
Q108K:K40L:Y19W:A33W:T51V:T53S:R58W:L77K:L117E	39, 330
Q108K:K40L:Y19W:A33W:T51V:T53S:R58W:S76R:L117E	39,763
Q108K:K40L:Y19W:A33W:T51V:T53S:R58W:S76H:L117E	38, 892
Q108K:K40L:Y19W:A33W:T51V:T53S:R58W:S76K:L117E	39,995
Q108K:K40L:Y19W:T29H:A33W:T51V:T53S:R58W:L117E	40,434
Q108K:K40L:Q4F:Y19W:T29K:T51V:T53S:R58W:L117D	35,743
Q108K:K40L:Q4F:Y19W:T51V:T53S:R58W:L77H:L117D	36,213
Q108K:K40L:Q4F:Y19W:M20K:T51V:T53S:R58W:L117D	35,893
Q108K:K40L:Q4F:Y19W:M20H:T51V:T53S:R58W:L117D	36,187
Q108K:K40L:Q4F:Y19W:T51V:T53S:R58W:S76H:L117D	37,121

Table IV-4 (cont'd)

Q108K:K40L:Q4F:Y19W:A33H:T51V:T53S:R58W:L117D	34,862
Q108K:K40L:Q4F:Y19W:A33E:T51V:T53S:R58W:L117D	35,628
Q108K:K40L:Q4F:Y19W:T51V:T53S:R58W:S76E:L117D	36,986
Q108K:K40L:Q4F:Y19W:T29D:T51V:T53S:R58W:L117D	37,334
Q108K:K40L:Q4F:Y19W:T29E:T51V:T53S:R58W:L117D	34,761
Q108K:K40L:A33H:T51V:T53S:R58H	25,741
Q108K:K40L:A33H:T51V:T53S:R58W	30,869
Q108K:K40L:Y19W:A33H:T51V:T53S:R58W:L117E	39,391
Q108K:K40L:Y19W:A33H:T51V:T53S:R58H	29,874
Q108K:K40L:A33E:T51V:T53S:R58H	27,154
Q108K:K40L:A33S:T51V:T53S:R58H	27,480
Q108K:K40L:Y19W:A33S:T51V:T53S:R58H	34,205
Q108K:K40L:Y19W:A33E:T51V:T53S:R58H	34,827
Q108K:K40L:A33Y:T51V:T53S:R58H	29,287
Q108K:K40L:Y19W:A33Y:T51V:T53S:R58H	33,034
Q108K:K40L:Y19W:A33H:T51V:T53Y:R58H	33,718
Q108K:K40L:Y19W:A33W:T51V:T53H:R58H	38,143
Q108K:K40L:Y19W:A33W:T51V:T53Y:R58H	38,497
Q108K:K40L:F16Y:A33H:T51V:T53S:R58H	29,244
Q108K:K40L:F16Y:Y19W:A33Y:T51V:T53S:R58H	35,708
Q108K:K40L:F16Y:T51V:T53S:R58H	30,701
Q108K:K40L:F16Y:Y19W:A33W:T51V:T53S:R58H	34,602

Table IV-4 (cont'd)

Q108K:K40L:Y19W:A33H:T51V:T53H:R58H	29,754	
Q108K:K40L:Y19W:A33W:T51V:T53E:R58H	39,205	
Q108K:K40L:T51V:T53S:R58H:L117E	26,669	
Q108K:K40L:A33H:T51V:T53S:S55E:R58H	32,006	53,816
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53S:R58H	34,158	
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53Y:R58H	32,919	
Q108K:K40L:Q38E:T51V:T53S:R58H	25,760	
Q108K:K40L:F16E:A33H:T51V:T53S:R58H	29,098	
Q108K:K40L:F16E:T51V:T53S:R58H	26,727	
Q108K:K40L:F16H:Y19W:A33W:T51V:T53S:R58H	35,303	
Q108K:K40L:F16H:Y19W:A33Y:T51V:T53S:R58H	38,777	
Q108K:K40L:F16E:Y19W:A33W:T51V:T53S:R58H	39,943	
Q108K:K40L:A33H:Q38Y:T51V:T53S:R58H	27,958	
Q108K:K40L:F16Y:A33H:I42E:T51V:T53S:R58H		54,179
Q108K:K40L:F16Y:Y19W:A33W:T51E:T53S:R58H	39,061	
Q108K:K40L:F16E:Y19W:A33Y:T51V:T53S:R58H	34,119	
Q108Q:K40L:T51V:T53S:I110K	30,045	
Q108K:K40L:Q38H:T51V:T53S:R58H	25,095	
Q108K:K40L:T51V:T53E:R58H	26,199	
Q108K:K40L:T51V:T53Y:R58H	28,178	
Q108K:K40L:T51V:T53S:S55Y:R58H		54,120
Q108K:K40L:T51V:T53S:S55E:R58H	32,005	53,376

Table IV-4 (cont'd)

Q108K:K40L:Y19W:A33H:T51V:T53E:R58H	33,433	
Q108K:K40L:T51V:T53S:S55H R58H		50,732
Q108K:K40L:F16W:A33H:T51V:T53S:R58H	32,566	
Q108K:K40L:Y19W:A33H:Q38E:T51V:T53E:R58H	37,341	
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53E:R58H	31,938	
Q108K:K40L:F16Y:A33H:T51V:T53E:R58H	27,444	
Q108K:K40L:Y19W:A33Y:T51V:T53E:R58H	33,338	
Q108K:K40L:A33H:T51V:T53E:R58H	28,006	
Q108K:K40L:F16Y:Y19W:A33H:Q38E:T51V:T53S:R58H	31,488	
Q108K:K40L:Q38E:T51V:T53E:R58H	27,454	
Q108K:K40L:F16Y:Y19W:A33W:T51V:T53E:R58H	38,611	
Q108K:K40L:F16Y:A33H:Q38E:T51V:T53S:R58H	29,734	
Q108K:K40L:F16Y:Q38E:T51V:T53S:R58H	28,978	
Q108K:K40L:F16Y:A33H:T51V:T53S:R58H:L117T	27,933	
Q108K:K40L:Q38E::T51V:T53S:R58H:L117T	27,053	
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53S:R58H:L117T	28,008	
Q108K:K40L:F16Y:Y19W:A33Y:Q38E:T51V:T53S:R58H	36,571	
Q108K:K40L:Q4T:F16Y:Y19W:A33H:T51V:T53S:R58H	32,281	
Q108K:K40L:F16Y:T51V:T53S:R58H:L117Y	29,375	
Q108K:K40L:F16Y:A33H:T51V:T53S:R58H:L117Y	30,800	
Q108K:K40L:Y19W:A33H:T51V:T53E:R58H:L117Q	30,670	
Q108K:K40L:Y19W:A33H:T51V:T53E:R58H:L117H	32,650	

Table IV-4 (cont'd)

Q108K:K40L:Q38E:T51V:T53S:R58H:L117H	26,340
Q108K:K40L:F16Y:A33H:T51V:T53S:R58H:L117C	27,969
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53S:R58H:L117C	32,803
Q108K:K40L:Q38E:T51V:T53S:R58H:L117C	25,190
Q108K:K40L:Q4S:F16Y:A33H:T51V:T53S:R58H	28,225
Q108K:K40L:Q4Y:Y19W:A33H:T51V:T53E:R58H	28,601
Q108K:K40L:F16Y:A33H:T51V:T53S:R58H:L117S	28,601
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53S:R58H:L117S	31,516
Q108K:K40E:Y19W:A33Y:T51V:T53S:R58H	32,310
Q108K:K40L:F16Y:A33Y:T51V:T53S:R58H	30,012
Q108K:K40L:Y19W:A33H:T51V:T53E:R58H:L117C	31,567
Q108K:K40L:F16Y:T51V:T53S:R58H:L117C	30,063
Q108K:K40E:F16Y:T51V:T53S:R58H	28,681
Q108K:K40L:Y19W:A33Y:T51V:T53S:R58H:L117C	33,809
Q108K:K40L:Y19W:A33H:T51V:T53S:R58H:L117C	28,398
Q108K:K40L:A33H:T51V:T53S:R58H:L77H	27,535
Q108K:K40L:A33H:T51V:T53S:R58H:L77Y	28,331
Q108K:K40L:T51V:T53S:R58H:F16Y:L77Y	27,506
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53S:R58H:L117E	32,378
Q108K:K40L:Y19W:A33Y:T51V:T53S:R58Y	36,299
Q108K:K40L:F16Y:T51V:T53S:R58Y	31,297
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53S:R58Y	34,308

Table IV-4 (cont'd)

Q108K:K40L:F16Y:A33Y:T51V:T53S:R58Y	31,774
Q108K:K40T:F16Y:Y19W:A33H:T51V:T53S:R58H:L117C	33,233
Q108K:K40Y:F16Y:Y19W:A33H:T51V:T53S:R58H:L117C	34,618
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53S:R58H:L117M	31,761
Q108K:K40D:F16Y:Y19W:A33H:T51V:T53S:R58H:L117C	32,171
Q108K:K40Q:F16Y:Y19W:A33H:T51V:T53S:R58H:L117C	33,434
Q108K:K40E:Q4F:F16Y:Q38F:T53S:R58H	29,635
Q108K:K40C:Q4F:F16Y:Y19W:A33H:T51V:T53S:R58H:L117C	32,514
Q108K:K40N:Q4F:F16Y:Y19W:A33H:T51V:T53S:R58H:L117C	32,606
Q108K:K40H:Q4F:F16Y:Y19W:A33H:T51V:T53S:R58H:L117C	31,227
Q108K:K40S:Q4F:F16Y:Y19W:A33H:T51V:T53S:R58H:L117C	33,162
Q108K:K40E:Q4F:F16Y:Q38F:T53A:R58H	24,829
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53S:R58H:L77Y:L117C	34,691
Q108K:K40E:Q4F:A33H:Q38F:T53A:R58H	26,772
Q108K:K40E:Q4F:F16Y:A33H:Q38F:T53A:R58H	30,458
Q108K:K40E:Q4F:F16Y:Q38F:T53A:R58Y	31,173
Q38P	26,828
Q108K:K40E:Q4F:Y19W:Q38F:T53A:R58H	34,783
Q108K:K40E:Q4F:F16Y:Y19W:Q38F:T53A:R58H	40,509
Q108K:K40E:Q4F:T29Y:Q38F:T53A:R58H	29,354
Q108K:K40E:Q4F:F16Y:T29Y:Q38F:T53A:R58H	31,156
Q108K:K40E:Q4F:Q38F:T53A:R58H:L77M	30,509

Table IV-4 (cont'd)

Q108K:K40E:Q4F:F16Y:Q38F:T53A:R58H:L77M	34,440	
Q108K:K40E:Q4F:F16Y:A33M:Q38F:T53A:R58H	30,877	
Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53S:R58H:L117C	34,288	
Q108K:K40E:Q4F:F16Y:A33C:Q38F:T53A:R58H	31,598	
Q108K:K40E:Q4F:F16Y:Q38F:T51M:T53A:R58H	28,998	
Q108K:K40E:Q4F:Q38F:T51M:T53A:R58H	28,519	
Q108K:K40L:F16Y:Y19W:A33H:T51M:T53A:R58H:L117C	33,412	
Q108K:K40L:F16Y:Y19W:A33H:T51M:T53S:R58H:L117C	34,361	
Q108K:K40E:Q4F:A33M:Q38F:T53A:R58H	28,116	
Q108K:K40E:Q4F:F16Y:Q38Y:T53A:R58H	31,815	
Q108K:K40E:Q4F:F16Y:M20Y:Q38F:T53A:R58H	29,132	
Q108K:K40E:Q4F:M20Y:Q38F:T53A:R58H	29,096	
Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51C:T53S:R58H:L117C	33,99	
Q108K:K40E:Q4F:F16Y:T29Y:Q38Y:T53A:R58H	32,318	
Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53C:R58H:L117C	32,744	
Q108K:A28C:L36C:T51D:F57H	28,122	
Q108K:K40E:Q4F:A33H:Q38F:T53A:R58H:L77Y	28,955	
Q108K:K40D:Q4F:Q38F:T51V:T53A:R58L:V62E	26,357	53,118
Q108K:K40D:Q4F:Q38F:T53A:R58L:V62E	25,820	56,982
Q108K:K40L:Q4F:Q38F:I42E:T53A:R58L	29,326	
Q108K:K40D:Q4F:Q38F:T53A:R58L	28,682	56,298
Q108K:K40D:Q4F:Q38F:T53A:R58H	29,197	

Table IV-4 (cont'd)

Q108K:K40H:Q4F:Q38F:T53A:R58H	28,664
Q108K:K40K:Q4F:Q38F:T53A:R58H	28,439
Q108K:K40N:Q4F:Q38F:T53A:R58H	27,224
Q108K:T51D	28,500
Wild Type hCRBPII	33,068
Wild Type CRABPII	19,127
Q108K:K40E:Q4F:F16Y:Q38H:T53A:R58H	30,840
Q108K:K40E:Q4F:F16Y:Q38Y:T53A:R58H:L117E	35,000
Q108K:K40R:Q4F:Q38F:T53A:R58H	26,242
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53C:R58H:L77Y:L117C	35,025
Q108K:K40L:A33H:T51V:T53S:R58H:L77Y:L117C	29,103
Q108K:K40L:F16Y:Y19W:T29Y:T51V:T53C:R58H:L77Y:L117C	33,813
Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53C:R58H:L77Y:L117C	43,453
Q108K:K40L:Q38E:T51V:T53S:R58L:L117T	27,263
Q108K:K40L:F16Y:Y19W:T29F:A33H:T51V:T53C:R58H:L77Y:L117C	36,463
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53C:R58H:L77F:L117C	36,747
Q108K:K40L:F16Y:Y19W:T29F:A33H:T51V:T53C:R58H:L117C	38,713
Q108K:K40L:F16Y:Y19W:A33H:T51V:T53C:R58H:L77W:L117C	40,083
Q108K:K40L:F16Y:Y19W:T29W:A33H:T51V:T53C:R58H:L117C	38,757
Q108K:K40L:Y19W:A33H:T51V:T53C:R58H:L77Y:L117C	33,205
Q108K:K40L:Q4F:Q38F:T53A:R58H	28,241
Q108K:K40L:Q4F:Q38F:T53A:R58L	31,014

Table IV-4 (cont'd)

Q108K:K40E:Q4F:Q38F:T53A:R58L:V62E	26,787
Q108K:K40L:Q4F:T29L:A33W:T51V:T53C:R58W:L117E	36,239
Q108A:K40L:Q44K:T51V:T53S	23,406
Q108K:K40L:Q4E:Q38E:T51V:T53S:R58H	25,764
Q108K:K40L:Q4F:Y19W:T51V:T53S:R58W:L117D:L77D	39,341
Q108K:K40L:Y19W:T29R:A33W:T51V:T53S:R58W:L117E	45,051
Q108K:K40L:Q4F:Y19W:T51V:T53S:R58W:S76R:L117D	46,135
T53D	33,678
T53E	26,630
L119D	22,592
R58A	30,005

IV.3.3 pK_a measurements of hCRBPII/chromophore complexes

For pK_a measurement, protein (20 μM in PBS) was incubated with ligand (0.5 equiv) at room temperature (23 °C) until Schiff base (SB) or protonated Schiff base (PSB) formation was complete. This was verified by UV-Vis spectrum. The solution was then titrated with acid (1 M NaOAc, pH=4) for mutants with low pK_a or base (1 M NaOH) for mutants with high pK_a in ~ 0.5 pH units, and the absorption spectra were recorded at each point. The λ_{max} of the protein/ligand complex versus pH was plotted. a polynomial fit of the data (3rd power) was applied for pK_a determination:

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

The two parameters are: ΔA, the total absorbance change of the PSB (or SB for some proteins) during each point of titration in comparison to the starting absorption, and pK_a, the midpoint of titration. It should be noted that a constant is included to account for the deviation from zero absorbance intensity of the deprotonated PSB. pH values were recorded with an accumet™ Basic pH meter (Fisher Scientific) equipped with a PerpHect™ ROSS™ Micro Combination pH electrode (Thermo Scientific Orion).

IV.3.4 hCRBPII/chromophore binding Kinetic measurements

Binding kinetics were measured at 23 °C using a Cary temperature controller. hCRBPII mutant 20 μM in PBS 2x buffer (pH=7.4) was mixed with 0.5 equivalent of the chromophore. The spectra were recorded immediately after mixing the protein and chromophore, and the absorbance intensity was plotted as a function of time.

IV.3.4.A. Pseudo-first-order binding rate measurement

Proteins with low pK_a (pK_a < 9) forming both SB and PSB under the kinetic measurement conditions were fit to a pseudo-first-order rate equation, as the chromophore concentration could not be determined accurately. The fit for the pseudo-first-order rate equation is shown in the equation below,

$$A = A_0 \times (1 - e^{-kt}) + c$$

where A is the absorbance value at each recorded time point, A_0 is the final absorbance value (after the completion of the reaction), k is the pseudo-first-order

rate constant, *t* is the time elapsed, and *c* is a varying constant which accounts for any time delay from the addition of the chromophore to the point recording was started. This equation was rewritten in KaleidaGraph in the following format, where m2 is the rate constant, and the half-life ($t_{1/2}$) of the reaction can be calculated by ln(2)/*k*.

$$y = m1 \times (1 - e^{-m2 \times m0}) + m3$$

IV.3.4.B. Second-order binding rate measurement

For proteins with high imine pK_a (pK_a > 9), the majority of the protein complex is in PSB form, and chromophore concentration can be determined confidently. The product concentration can be calculated at each point by Beer's Law using absorbance intensity at each time point and the extinction coefficient of the hCRBPII/ligand complex. Then, the concentration of product (hCRBPII/ligand complex) versus time was plotted.

The data were fit to a second-order rate equation as shown in the equation below, which was derived for product formation following the second-order rate equation previously.[4,5]

$$y = m3 - \frac{1}{(m2 \times m0) + m1}$$

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), and m2 is the rate constant *k*. The half-life ($t_{1/2}$) of the reaction can be calculated by $1/(k^*[A]_0)$, wherein 10 μM was assumed for $[A]_0$.[6]

IV.3.5. Absolute fluorescent quantum yield measurements

Absolute fluorescence quantum yields (Φ) were recorded at room temperature on a Quantaurus-QY C11347-11 (Hamamatsu Photonics K.K., Japan) equipped with a Xenon lamp and a monochromator as excitation light source, an integration sphere, and a multichannel back-thinned CCD detector. All samples were diluted with PBS.2X solution or corresponding organic solvents (A < 0.1). Recorded values are average numbers (n = 5).

IV.4 Cloning to mammalian expression vectors

IV.4.1 General cloning protocol

The DNA fragment was amplified using Pfu Turbo DNA Polymerase (Agilent) with the appropriate primers depending on the restriction site to be introduced to the template plasmid. All restriction sites for cloning were chosen with sticky ends. PCR conditions are specified in **Table IV-5** using a Bio-Rad iCycler thermal cycler.

The PCR amplified DNA fragment was purified by Wizard® SV Gel and PCR Clean-Up System (Promega) from 1% agarose gel in an amount of 20-50 ng/μL. The fragment was digested with proper enzymes and ligated to a similarly prepared plasmid (50 ng/μL). Ligation between 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* XL-1 blue competent cells and grown on LB-agar plates supplemented with antibiotics (100 μg/mL ampicillin, 7.5 μg/mL tetracycline) at 37 °C for 20 hours. Colonies (3-6) were

inoculated in LB medium (10 mL) containing the proper amount of antibiotics (100 µg/mL ampicillin, 7.5 µg/mL tetracycline) and incubated at 37 °C while shaking, for 20 hours.

Table IV-5. PCR cycling conditions for cloning.

	PCR Program	Time (min)
1x	97 °C	1:00
1x	95 °C	3:00
15x	95 °C	00:30
	72°C	00:45
	72 °C (temp decreases after cycle 1 by 1 °C by every cycle)	03:40
20x	95 °C	00:30
	3-5 °C Below Primer Tm	00:45
	72 °C	03:40
1x	72 °C	08:00
1x	4 °C	10:00

Reactant	volume
DNA (Template Plasmid)	100 ng (x µl)
Primer forward	50 pmol (y µl)
Primer reverse	50 pmol (z µl)
10 mM dNTP	1 µL
DMSO	1 µL
50 mM $MgCl_2$	1 µL
5× Pfu buffer	5 µL
Pfu Turbo DNA Polymerase	1.5 µL
DI Water	50 – x – y – z – 9.5 µL

DNA purification was performed using the Promega Wizard® Plus SV Miniprep DNA purification kit (A1330) following the suggested protocol. The DNA sequence was verified with the corresponding sequencing primers by the MSU gene sequencing facility. Sequencing primers used are shown below.

CMV end_Seq: 5'-GGT CTA TAT AAG CAG AGC TGG TTT AG-3'

midGFP: 5'-CGT GCT GCT GCC CGA CAA CC-3'

IV.4.2 Sequences of plasmids described in this thesis

Plasmid 1: HindIII-EGFP-NotI-Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53C:R58H:L117C-EcoRI-3NLS-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACCAGAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTATGAGGGCTGGATGAAGGCCCTGGATA
TTGATTTTGCCTACCGCAAGATTCACGTACGTCTCACTCAGACGCTGGTTATTGATCAA
GATGGTGATAACTTCAAGGTAAAATGCACTAGCACATTCCACAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACTGTGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGAATTCGCTGACCCCAAGAAGAAGA
GGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGTGA
AAACATCGATAGATCTGATATC

Plasmid 2: HindIII-EGFP-NotI-

Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53C:R58H:L117C-EcoRI-NES-

Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACCAGAAT
GGAACCTGGGAGATGGAGAGTAATGAAAACTATGAGGGCTGGATGAAGGCCCTGGATAT
TGATTTTGCCTACCGCAAGATTCACGTACGTCTCACTCAGACGCTGGTTATTGATCAAG
ATGGTGATAACTTCAAGGTAAAATGCACTAGCACATTCCACAACTATGATGTGGATTTC
ACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGGC
ACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAACC
GCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACTGTGAGCTGACCTGTGGTGAC
CAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGAATTCGAGCTTGCCGAGAAACTTGC
CGGGCTTGACATAAATTGAGGATCCCGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCT

Plasmid 3: HindIII-EGFP-NotI-

Q108K:K40L:F16Y:Y19W:T29Y:A33H:T51V:T53C:R58H:L117C-EcoRI-Linker

(whole cell localization)-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA

AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGGGCGGCCGCATGACGAGGGACCAGAATGGAACCTGGGAGATGGA
GAGTAATGAAAACTATGAGGGCTGGATGAAGGCCCTGGATATTGATTTTGCCTACCGCA
AGATTCACGTACGTCTCACTCAGACGCTGGTTATTGATCAAGATGGTGATAACTTCAAG
GTAAAATGCACTAGCACATTCCACAACTATGATGTGGATTTCACTGTTGGAGTAGAGTT
TGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGGCACTGGTCACCTGGGAAG
GTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAACCGCGGCTGGAAGAAGTGG
ATTGAGGGGGACAAGCTGTACTGTGAGCTGACCTGTGGTGACCAGGTGTGCCGTCAAGT
GTTCAAAAAGAAGTGCGAATTCATCGATAGATCTGATATCGGTACCAGTCGACTCTAGA
GGATCCCGGGTGGCATCCCTGTGACCCCTCCCCA

Plasmid 4: HindIII-EGFP-NotI-Q108K:K40E:Q4F:Q38F:T53A:R58H-EcoRI-3NLS-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGGGCGGCCGCATGACGAGGGACTTCAATGGAACCTGGGAGATGGA
GAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATATTGATTTTGCCACCCGCA
AGATTGCAGTACGTCTCACTTTTACGGAAGTTATTGATCAAGATGGTGATAACTTCAAG
ACAAAAGCCACTAGCACATTCCACAACTATGATGTGGATTTCACTGTTGGAGTAGAGTT
TGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGGCACTGGTCACCTGGGAAG
GTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAACCGCGGCTGGAAGAAGTGG
ATTGAGGGGGACAAGCTGTACCTGGAGCTGACCTGTGGTGACCAGGTGTGCCGTCAAGT
GTTCAAAAAGAAGTGCGAATTCGCTGACCCCAAGAAGAAGAGGAAGGTGGACCCCAAGA
AGAAGAGGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGTGAAACATCGATAGATCTGA

Plasmid 5: HindIII-EGFP-NotI- Q108K:K40E:Q4F:Q38F:T53A:R58H-EcoRI-NES-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC

TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGGGCGGCCGCATGACGAGGGACTTCAATGGAACCTGGGAGATGGA
GAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATATTGATTTTGCCACCCGCA
AGATTGCAGTACGTCTCACTTTTACGGAAGTTATTGATCAAGATGGTGATAACTTCAAG
ACAAAAGCCACTAGCACATTCCACAACTATGATGTGGATTTCACTGTTGGAGTAGAGTT
TGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGGCACTGGTCACCTGGGAAG
GTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAACCGCGGCTGGAAGAAGTGG
ATTGAGGGGGACAAGCTGTACCTGGAGCTGACCTGTGGTGACCAGGTGTGCCGTCAAGT
GTTCAAAAAGAAGTGCGAATTCGAGCTTGCCGAGAAACTTGCCGGGCTTGACATAAATTG
AGGATCCCGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCT

Plasmid 6: HindIII-EGFP-NotI-Q108K:K40E:Q4F:Q38F:T53A:R58H-EcoRI-

Linker (whole cell localization)-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGGGCGGCCGCATGACGAGGGACTTCAATGGAACCTGGGAGATGGA
GAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATATTGATTTTGCCACCCGCA
AGATTGCAGTACGTCTCACTTTTACGGAAGTTATTGATCAAGATGGTGATAACTTCAAG
ACAAAAGCCACTAGCACATTCCACAACTATGATGTGGATTTCACTGTTGGAGTAGAGTT
TGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGGCACTGGTCACCTGGGAAG
GTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAACCGCGGCTGGAAGAAGTGG
ATTGAGGGGGACAAGCTGTACCTGGAGCTGACCTGTGGTGACCAGGTGTGCCGTCAAGT
GTTCAAAAAGAAGTGCGAATTCATCGATAGATCTGATATCGGTACCAGTCGACTCTAGA
GGATCCCGGGTGGCATCCCTGTGACCCCTCCCC

Plasmid 7: HindIII-EGFP-NotI-Q108K:K40E:Q4F:Q38F:T53A:R58L-EcoRI-3NLS-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACTTCAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATA
TTGATTTTGCCACCCGCAAGATTGCAGTACGTCTCACTTTTACGGAAGTTATTGATCAA
GATGGTGATAACTTCAAGACAAAAGCCACTAGCACATTCCTGAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACCTGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGAATTCGCTGACCCCAAGAAGAAGA
GGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGGACCCCAAGAAGAAGAGGAAGGTG**TGA**
AAACATCGATAGATCTGATATCGGTACCAGTCGA

Plasmid 8: HindIII-EGFP-NotI-Q108K:K40E:Q4F:Q38F:T53A:R58L-EcoRI-NES-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACTTCAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTTTGAGGGCTACATGAAGGCCCTGGATA

TTGATTTTGCCACCCGCAAGATTGCAGTACGTCTCACTTTTACGGAAGTTATTGATCAA
GATGGTGATAACTTCAAGACAAAAGCCACTAGCACATTCCTGAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACCTGGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGAATTCGAGCTTGCCGAGAAACTTG
CCGGGCTTGACATAAATTGAGGATCCCGGGTGGCATCCCTGTGACCCCTCCCCAGTGCC

Plasmid 9: HindIII-EGFP-NotI-Q108K:K40E:Q4F:Q38F:T53A:R58L-EcoRI-

Linker (whole cell localization)-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGGGCGGCCGCATGACGAGGGACCAGAATGGAACCTGGGAGATGGA
GAGTAATGAAAACTATGAGGGCTGGATGAAGGCCCTGGATATTGATTTTGCCTACCGCA
AGATTCACGTACGTCTCACTCAGACGCTGGTTATTGATCAAGATGGTGATAACTTCAAG
GTAAAATGCACTAGCACATTCCACAACTATGATGTGGATTTCACTGTTGGAGTAGAGTT
TGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGGCACTGGTCACCTGGGAAG
GTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAACCGCGGCTGGAAGAAGTGG
ATTGAGGGGGACAAGCTGTACTGTGAGCTGACCTGTGGTGACCAGGTGTGCCGTCAAGT
GTTCAAAAAGAAGTGCGAATTCATCGATAGATCTGATATCGGTACCAGTCGACTCTAGA
GGATCCCGGGTGGCATCCCTGTGACC

Plasmid 10: HindIII-EGFP-NotI-

Q108K:K40L:F16Y:Y19W:A33H:T51V:T53C:R58H:L77Y:L117C-EcoRI-3NLS-

Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC

GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACCAGAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTATGAGGGCTGGATGAAGGCCCTGGATA
TTGATTTTGCCACCCGCAAGATTCACGTACGTCTCACTCAGACGCTGGTTATTGATCAA
GATGGTGATAACTTCAAGGTAAAATGCACTAGCACATTCCACAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCTACGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACTGTGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGAATTCGCTGACCCCAAGAAGAAGA
GGAAGGTGGACCCCAAGAAGAAGAGGAAGGTGGACCCCAAGAAGAAGAGGAAGGTG**TGA**
AAACATCGATAGATCTGATATCGGTACCAGTCGACTCTAGAGGATC

Plasmid 11: HindIII-EGFP-NotI-

Q108K:K40L:F16Y:Y19W:A33H:T51V:T53C:R58H:L77Y:L117C-EcoRI-NES-

Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTCCGGAGCCGCTGCAGGAGGCGGCCGCATGACGAGGGACCAGAA
TGGAACCTGGGAGATGGAGAGTAATGAAAACTATGAGGGCTGGATGAAGGCCCTGGATA
TTGATTTTGCCACCCGCAAGATTCACGTACGTCTCACTCAGACGCTGGTTATTGATCAA
GATGGTGATAACTTCAAGGTAAAATGCACTAGCACATTCCACAACTATGATGTGGATTT
CACTGTTGGAGTAGAGTTTGACGAGTACACAAAGAGCTACGATAACCGGCATGTTAAGG
CACTGGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAAC
CGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACTGTGAGCTGACCTGTGGTGA
CCAGGTGTGCCGTCAAGTGTTCAAAAAGAAGTGCGAATTCGAGCTTGCCGAGAAACTTG
CCGGGCTTGACATAAAT**TGA**GGATCCCGGGTGGCATCCCTGTGACCCCTCCCCAGTGCC

Plasmid 12: HindIII-EGFP-NotI-

Q108K:K40L:F16Y:Y19W:A33H:T51V:T53C:R58H:L77Y:L117C-EcoRI-Linker

(whole cell localization)-Stop-BamHI (pFlag-CMV2)

AAGCTTATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCC
TGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG
ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGGGCGGCCGCATGACGAGGGACCAGAATGGAACCTGGGAGATGGA
GAGTAATGAAAACTATGAGGGCTGGATGAAGGCCCTGGATATTGATTTTGCCACCCGCA
AGATTCACGTACGTCTCACTCAGACGCTGGTTATTGATCAAGATGGTGATAACTTCAAG
GTAAAATGCACTAGCACATTCCACAACTATGATGTGGATTTCACTGTTGGAGTAGAGTT
TGACGAGTACACAAAGAGCTACGATAACCGGCATGTTAAGGCACTGGTCACCTGGGAAG
GTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGAGAACCGCGGCTGGAAGAAGTGG
ATTGAGGGGGACAAGCTGTACTGTGAGCTGACCTGTGGTGACCAGGTGTGCCGTCAAGT
GTTCAAAAAGAAGTGCGAATTCATCGATAGATCTGATATCGGTACCAGTCGACTC**TAG**A
GGATCCCGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGT

IV.5 Mammalian cell culture and transfection

All cell lines (HeLa and COS-7) were cultured in Dulbecco's Modified Eagle's Medium (DMEM, containing 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.

Cells were seeded 2 d before transfection on an ibidi μ-Slide 8 well coverslip (with ibiTreat). Transient transfection was performed at ~70% confluency with 0.25 µg of plasmid DNA (per well) using Genjet Ver. II (purchased from SignaGen) according to the manufacturer's protocol. The media was replaced with complete serum/antibiotics containing medium after 5 h. After about another 20 h incubation, cells were washed twice with Dulbecco's phosphate-buffered saline (DPBS, supplemented with calcium chloride and magnesium chloride; Sigma-Aldrich) and incubated in Phenol red-free RPMI 1640 medium (Sigma-Aldrich). Prior to confocal imaging, the fluorophore (in DMSO solution) was diluted to the specified concentration noted in text with pre-heated (37 °C) media and was added to the cells. For mutants that need more time to form the complex, the chromophore was added 2 h or the night before imaging. All washing steps were omitted.

IV.6 General methods for confocal imaging

Confocal images were acquired using an inverted laser scanning confocal microscope; Olympus FluoView 1000 spectral-based laser scanning confocal microscope configured on an Olympus IX81 automated inverted microscope platform, equipped with blue diode laser (405 nm), the Argon gas laser (458 nm, 488 nm, and 514 nm), the green diode laser (559 nm) and the red Helium-Neon gas laser (635 nm) sources. Differential Interference Contrast (DIC) images were acquired using a µ-Slide DIC lid (ibidi). All images were taken with Olympus UPlanFL N 40x/1.30 oil and PlanApo N 60x/1.40 oil objectives, and Olympus Fluoview version 4.2a software was used to process images.

hCRBPII/**ThioPhenol** and hCRBPII/**CyThioPhenol** were imaged using multiple settings, as indicated in the main text. 1) For the hydroxyl moiety single ESPT of **ThioPhenol**: 515 nm excitation, excitation dichroic mirrors (DM) 405-440/515 nm and bandpass 620 nm-720 nm. 2) For the hydroxyl moiety single ESPT of **CyThioPhenol**: 559 nm excitation, excitation dichroic mirrors (DM) 405/488/559 and bandpass 620 nm-720 nm. 3) For the double ESPT of **ThioPhenol**: 405 nm excitation, excitation dichroic mirrors (DM) 405/488 nm and bandpass 560 nm-660 nm. 4) For the imine bond single ESPT of **CyThioPhenol**: 405 nm excitation, excitation dichroic mirrors (DM) 405/488 nm, and bandpass 480 nm-580 nm. 5) For the green fluorescence of **EGFP**: 488 nm excitation, excitation dichroic mirrors (DM) 405/488 nm and bandpass 500 nm-550 nm.

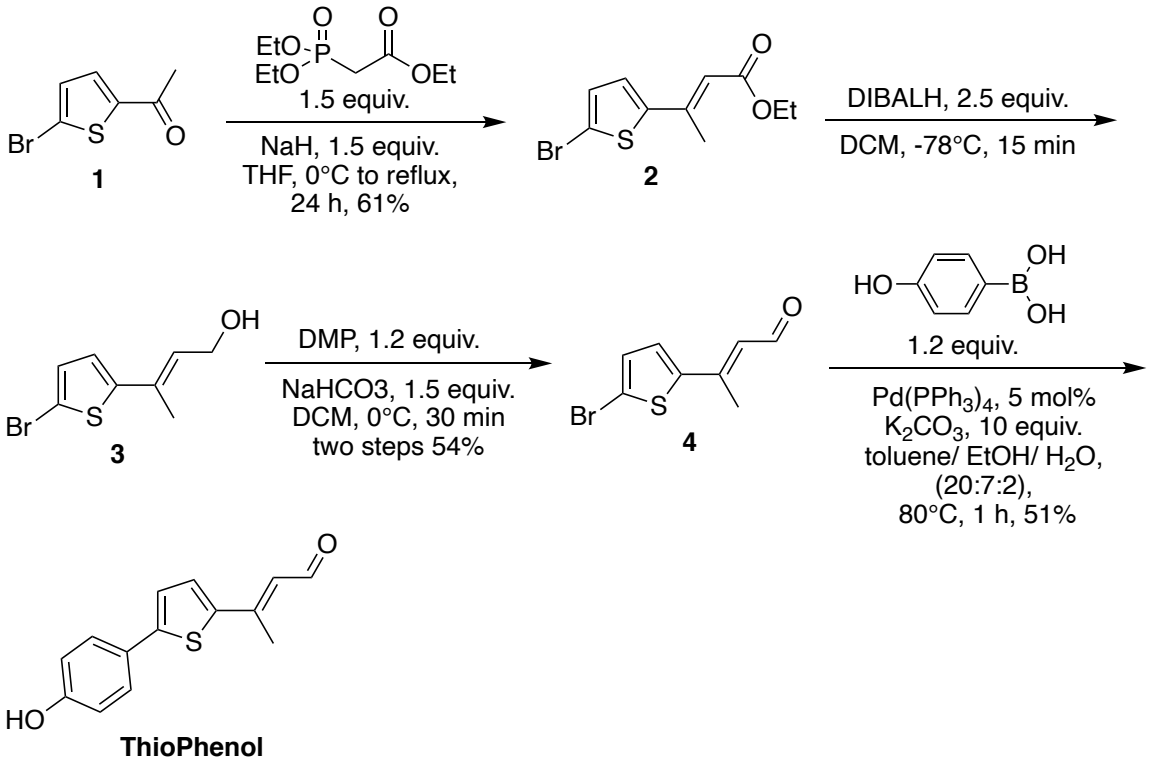
DIC images were acquired using a μ-Slide DIC lid (ibidi). Kalman averaging was applied each time prior to each imaging experiment. Fluorescence in each experiment was normalized to the same intensity, including the same laser intensity, gain, and amplifier offset. All images are pseudocolored with green, red, or cyan colors.

The following section describes the synthesis procedures of several chromophores discussed in chapters II and III, synthesized and purified by Mr. Mehdi Moemeni.

IV.7 General synthetic procedures

Commercially available starting materials were obtained from Sigma-Aldrich and were used without further purification unless specified. All moisture sensitive reactions were carried out in flame-dried or oven-dried glassware under an atmosphere of nitrogen or argon. Unless otherwise mentioned, solvents were purified as follows: tetrahydrofuran (THF) and diethyl ether (Et_2O) were distilled freshly from the classical sodium/benzophenone ketyl still pot; dichloromethane (DCM), acetonitrile, and toluene were dried over CaH_2 and freshly distilled prior to use; dimethylsulfoxide (DMSO), dimethylformamide (DMF), and triethylamine (Et_3N) were distilled from CaH_2 and stored over activated molecular sieves. Chemical shifts were reported relative to the residual solvent peaks. (^{1}H-NMR: δ 7.26 ppm for $CDCl_3$, δ 3.31 ppm for CD_3OD, δ 2.50 ppm for DMSO-d_6, 2.05 ppm for Aceton-d_6 respectively. ^{13}C-NMR: δ 77.16 ppm for CDCl3, δ 49.00 ppm for CD_3OD, δ 39.52 ppm for DMSO-d_6, 206.68 and 29.92 ppm for Aceton-d_6 respectively.) Analytical thin layer chromatography (TLC) was performed with pre-coated silica gel 60 F_{254} plates (Analtech, Inc.) Compounds in TLC were visualized upon UV irradiation and various staining techniques, i.e., p-anisaldehyde, potassium permanganate, phosphomolybdic acid in ethanol. Silica gel flash column chromatography was performed with Silicycle 40-60 Å (30 ~ 75 μM) silica gel.

IV.7.1 Synthesis of Thiophenol



ethyl (*E*)-3-(5-bromothiophen-2-yl)but-2-enoate (2)

Triethyl phosphonoacetate (1.98 mL, 10 mmol) was added dropwise under nitrogen to a suspension of NaH (60% on mineral oil, 400mg, 10 mmol) in dry THF (5mL) at 0 °C and the resulting mixture was stirred for another 30 min at the same temperature. A solution of 2-acetyl-5-bromothiophene **1** (1.35 g, 6.6 mmol) in THF (5 mL) was added dropwise to the mixture and heated to reflux for 24 h. After cooling, the reaction was quenched with saturated NH_4Cl and extracted three times with ethyl acetate which dried on Na_2SO_4. After evaporating the solvent, the product was purified via column chromatography by using hexane/ ethyl acetate (97:3) as eluent to give compound **2** as a yellow solid (1.1 g, 61%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.07 (d, *J* = 4.0 Hz, 1H), 7.01 (d, *J* = 4.0 Hz, 1H), 6.13 (q, *J* = 1.2 Hz, 1H), 4.21 (q, *J* = 7.2 Hz, 2H), 2.55 (d, *J* = 1.2 Hz, 3H), 1.32 (t, *J* = 7.1 Hz, 3H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 166.51, 146.77, 146.71, 130.88, 126.96, 114.65, 114.41, 60.03, 16.69, 14.33.

(*E*)-3-(5-bromothiophen-2-yl)but-2-en-1-ol (3)

To the compound **2** (819 mg, 3mmol) in dry DCM (10 mL) at -78 °C was added DIBALH (1M in hexane, 7 mL, 7 mmol) dropwise. After 15 min, the reaction was completed, and methanol (5 mL) was added dropwise to quench the reaction. saturated solution Potassium sodium tartrate (15mL) was added with DCM (10 mL). The mixture was stirred for 2 h and extracted with DCM (3 × 10 mL) and the combined organic layers were dried over Na_2SO_4. Evaporation of solvent gave compound **3** as a yellow solid and it used in next step without further purification.

(*E*)-3-(5-bromothiophen-2-yl)but-2-enal (4)

To compound **3** (559 mg, 2.4 mmol) in dry DCM (10 mL) at 0 °C, Dess-Martin periodinane (DMP) (1.22 g, 2.88 mmol) and $NaHCO_3$ (302 mg, 3.6 mmol) was added. Reaction was stirred at this temperature for 30 min and was quenched by addition of saturated sodium thiosulfate (10 mL) and saturated sodium bicarbonate (10 mL). The mixture was extracted with DCM (3 × 10 mL) and the combined organic layers were dried over Na_2SO_4. After evaporating the solvent, it

was purified via column chromatography by using hexane/ ethyl acetate (96:4) as eluent to give compound **4** as a yellow solid (374mg, two steps 54%).

^{1}H NMR (500 MHz, DMSO-d_6) δ 10.05 (d, J = 7.6 Hz, 1H), 7.52 (d, J = 4.0 Hz, 1H), 7.33 (d, J = 4.1 Hz, 1H), 6.23 (dq, J = 7.7, 1.2 Hz, 1H), 2.51 (s, 3H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 190.47, 148.71, 145.67, 131.37, 128.21, 124.44, 116.73, 15.56.

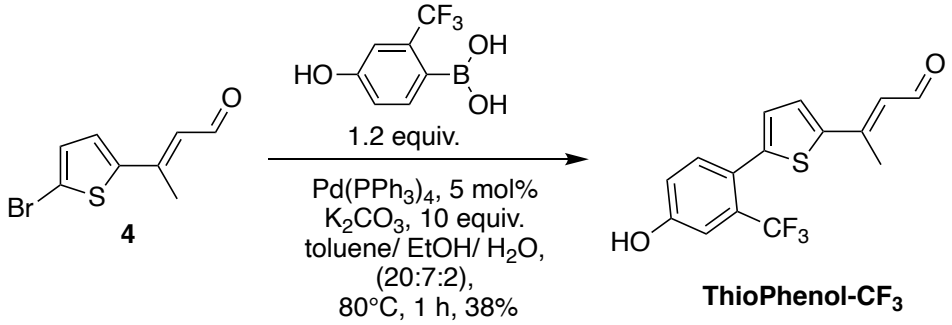
ThioPhenol

To the mixture of compound **4** (230 mg, 1.0 mmol), $Pd(PPh_3)_4$ (58 mg, 0.05 mmol), and toluene (20.0 mL), a suspension of (4-hydroxyphenyl)boronic acid (165.6 mg, 1.20 mmol) in ethanol (7 mL) and a solution of K_2CO_3 (1.38 g, 10 mmol) in water (2.0 mL) were added. The mixture was heated to 80 °C in an argon atmosphere for one hour. After cooling to room temperature, the mixture was filtered through Celite, and the organic phase was washed with water. The organic layers were then dried over Na_2SO_4 and concentrated under vacuum. Column chromatography with hexane/ ethyl acetate (9:1) of the residue afforded a mixture of *trans*/*cis* (84:16) isomers of ThioPhenol that were subjected to recrystallization with hexane/ ethyl acetate (7:3) to give pure *trans*-ThioPhenol as orange crystals (124 mg, 51%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 10.042 (*d, J*= 7.5, 1H), 9.844 (*s*, 1H), 7.637 (*d, J*= 4, 1H), 7.534 (*d, J*= 8.5, 2H), 7.402 (*d, J*= 4, 1H), 6.812 (*d, J*= 9, 2H), 6.264 (*d, J*= 7.5, 1H), 2.542 (*s*, 3H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 191.643, 158.669, 150.549, 148.104, 141.318, 131.303, 127.682, 124.403, 123.857, 123.338, 116.442, 15.906.

IV.7.2 Synthesis of ThioPhenol-CF_3

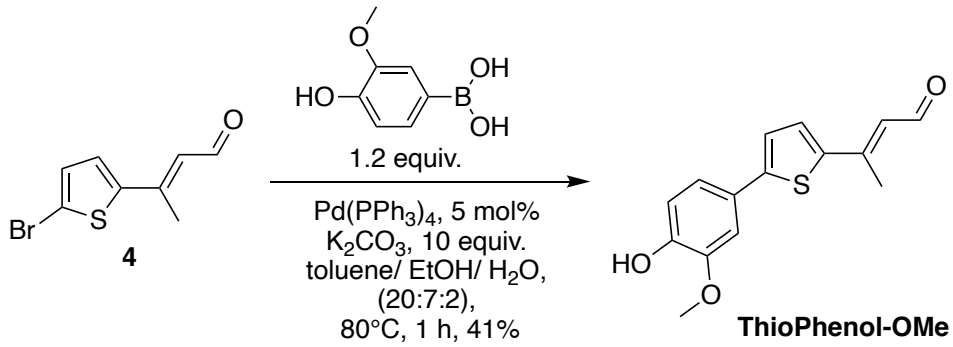


The synthesis of ThioPhenol-CF_3 was performed according to the general procedure described above (with same scale). Pure ThioPhenol-CF_3 was obtained as orange crystals (118 mg, 38%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 10.12 (d, J = 7.9 Hz, 1H), 7.41 (d, J = 3.9 Hz, 1H), 7.36 (d, J = 8.5 Hz, 1H), 7.26 (d, J = 3.9 Hz, 1H), 7.09 – 7.04 (m, 2H), 6.47 (dq, J = 7.9, 1.2 Hz, 1H), 2.61 (d, J = 1.1 Hz, 3H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 191.26, 156.28, 150.87, 144.33, 143.92, 134.41, 130.41, 130.26, 130.17, 129.05, 128.38, 124.49, 124.03, 122.30, 120.47, 119.05, 118.36, 114.02, 15.82.

IV.7.3 Synthesis of ThioPhenol-OMe$_2$

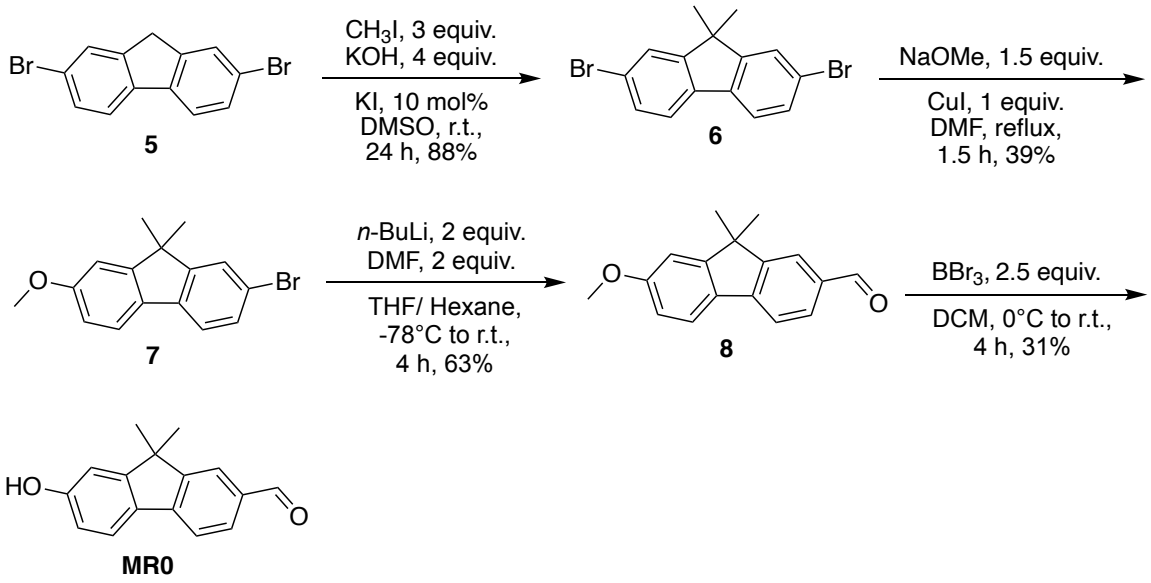


The synthesis of ThioPhenol-OMe$_2$ was performed according to the general procedure described above (with same scale). Pure ThioPhenol-OMe$_2$ was obtained as orange crystals (112 mg, 41%).

^{1}H NMR (500 MHz, Chloroform-d) δ 10.13 (d, J = 7.8 Hz, 1H), 7.41 (d, J = 3.9 Hz, 1H), 7.21 – 7.17 (m, 2H), 7.09 (d, J = 2.0 Hz, 1H), 6.96 (d, J = 8.3 Hz, 1H), 6.45 (dq, J = 7.9, 1.1 Hz, 1H), 5.74 (s, 1H), 3.98 (s, 3H), 2.59 (d, J = 1.1 Hz, 3H).

^{13}C NMR (126 MHz, Chloroform-d) δ 190.52, 149.78, 148.42, 146.80, 146.39, 142.22, 129.25, 126.01, 123.75, 123.23, 119.61, 114.94, 108.52, 56.04, 15.72.

IV.7.4 Synthesis of MR0



2,7-dibromo-9,9-dimethyl-9*H*-fluorene (6)

2,7-Dibromo-*9H*-fluorene **5** (7.77 g, 24 mmol) was added to a stirring solution of potassium hydroxide (5.38 g, 96 mmol) and potassium iodide (0.4 g, 2.4 mmol) in dimethylsulforxide (40 mL). Then Iodomethane (4.48 mL, 72 mmol) was added dropwise, and the mixture was stirred at room temperature for 24 hours. After completion, the reaction was poured into 250 mL water and extracted with DCM three times and the combined organic layers were dried over Na_2SO_4. After evaporating the solvent, the product was purified via flash column chromatography by using DCM as eluent to give compound **6** (7.43 g, 88%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.57 – 7.52 (m, 4H), 7.46 (dd, J = 8.1, 1.7 Hz, 2H), 1.47 (s, 6H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 155.25, 137.15, 130.32, 126.20, 121.47, 121.45, 47.32, 26.86.

2-bromo-7-methoxy-9,9-dimethyl-9*H*-fluorene (7)

Sodium methoxide was prepared by adding sodium (96.5 mg, 4.2 mmol) into 1.5 mL of anhydrous methanol under nitrogen. When the sodium disappeared, compound **6** (1 g, 2.8 mmol) in 15 mL of dry DMF and copper iodide (532 mg, 2.8 mmol) was added to the above solution and heated to reflux for 1.5 hour. The reaction mixture was poured into ice water and then extracted with ethyl acetate (3 × 15 mL). The combined organic layers were washed with 1 M HCl and brine and dried over Na_2SO_4. The resulting crude was purified by column chromatography (hexane/ ethyl acetate 98:2) to give compound **7** as a white solid (330 mg, 39%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.60 (d, *J* = 8.3 Hz, 1H), 7.51 (d, *J* = 1.8 Hz, 1H), 7.49 (d, *J* = 8.0 Hz, 1H), 7.43 (dd, *J* = 8.0, 1.8 Hz, 1H), 6.95 (d, *J* = 2.4 Hz, 1H), 6.89 (dd, *J* = 8.3, 2.4 Hz, 1H), 3.88 (s, 3H), 1.46 (s, 6H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 159.93, 155.26, 155.22, 138.35, 131.03, 130.00, 125.97, 120.91, 120.55, 119.75, 112.77, 108.53, 55.56, 47.08, 27.13.

7-Methoxy-9,9-dimethyl-9*H*-fluorene-2-carbaldehyde (8)

To a solution of compound **7** (439 mg, 1.45 mmol) in dry THF, *n*-BuLi (1.81 mL of 1.6 M solution in hexane, 2.9 mmol) was added dropwise at -78 °C under nitrogen. The reaction mixture was stirred for 1 hour at the same temperature, and

an orange suspension was formed. DMF (212 mg, 2.9 mmol) was added dropwise, and the formed solution was stirred for additional 2 hour at -78 °C. Then the reaction mixture was warmed to room temperature and stirred for 1 hour. The reaction was quenched with 2 M HCl aq., and the solution was extracted three times with ethyl acetate. The combined organic phase was dried with Na_2SO_4 and evaporated. The crude was purified by flash chromatography (hexane/ ethyl acetate 96:4) to yield compound **8** as a white solid (230 mg, 63%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 10.03 (s, 1H), 7.94 (d, *J* = 1.4 Hz, 1H), 7.84 (dd, *J* = 7.8, 1.4 Hz, 1H), 7.75 (d, *J* = 7.8 Hz, 1H), 7.71 (d, *J* = 8.3 Hz, 1H), 7.00 (d, *J* = 2.3 Hz, 1H), 6.94 (dd, *J* = 8.4, 2.4 Hz, 1H), 3.91 (s, 3H), 1.50 (s, 6H).

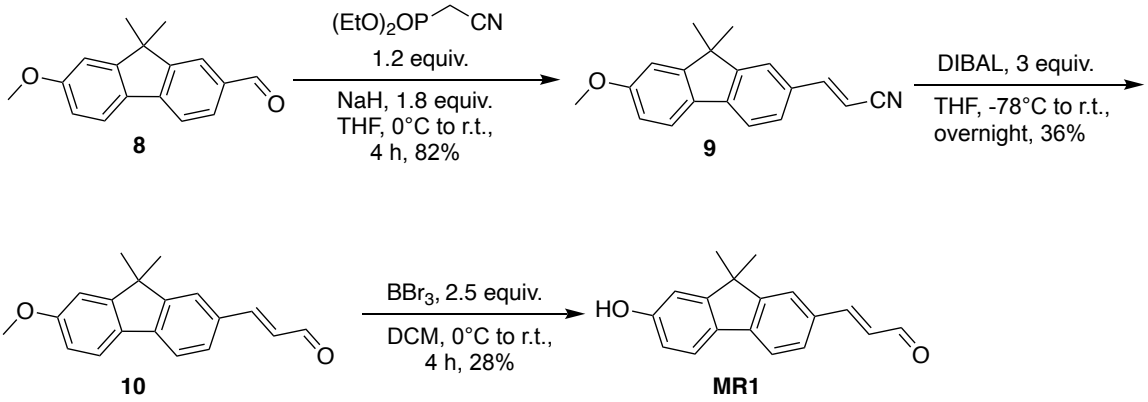
MR0

To a solution of compound **8** (50 mg, 0.2 mmol) in 3 mL dry DCM, BBr_3 (0.5 mL of 1 M solution in DCM, 0.5 mmol) was added dropwise at 0 °C and let the reaction warm to room temperature under nitrogen. After 4 h, the reaction was quenched by ice water and extracted three times by DCM. The solvent evaporated and the product purified by flash chromatography (hexane/ ethyl acetate 88:12) to yield **MR0** as a yellow solid (14.8 mg, 31%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 10.03 (s, 1H), 7.93 (d, *J* = 1.4 Hz, 1H), 7.84 (dd, *J* = 7.8, 1.5 Hz, 1H), 7.75 (d, *J* = 7.8 Hz, 1H), 7.67 (d, *J* = 8.2 Hz, 1H), 6.95 (d, *J* = 2.3 Hz, 1H), 6.87 (dd, *J* = 8.2, 2.4 Hz, 1H), 1.50 (s, 6H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 192.28, 157.52, 156.95, 153.76, 139.31, 134,53, 130.95, 130.65, 122.82, 122.48, 119.31, 114.79, 110.09, 46.87, 26.97.

IV.7.5 Synthesis of MR1



(*E*)-3-(7-methoxy-9,9-dimethyl-9*H*-fluoren-2-yl)acrylonitrile (9)

A solution of diethyl (cyanomethyl)phosphonate (72 mg, 0.41 mmol) in 2 mL tetrahydrofuran was added to a stirred suspension of NaH (60% on mineral oil, 25 mg, 0.62 mmol) in 2 mL tetrahydrofuran at 0 °C. The mixture was stirred at this temperature for 30 min. Then a solution of compound **8** (85 mg, 0.34 mmol) in 3 mL tetrahydrofuran was added. The reaction was stirred at ambient temperature for 4 hours. The reaction mixture was then poured into cold water and extracted with ethyl acetate. The organic phase was dried over Na_2SO_4, then the solvent evaporated, and the product purified by flash chromatography (hexane/ ethyl acetate 94:6) to yield compound **9** (76mg, 82%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.65 (dd, J = 9.9, 8.1 Hz, 2H), 7.50 – 7.37 (m, 3H), 6.98 (d, J = 2.3 Hz, 1H), 6.92 (dd, J = 8.4, 2.3 Hz, 1H), 5.89 (d, J = 16.6 Hz, 1H), 3.90 (s, 3H), 1.49 (s, 6H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 160.56, 156.35, 153.96, 151.00, 142.68, 131.37, 130.82, 127.29, 121.62, 121.12, 119.62, 118.74, 113.10, 108.54, 94.36, 55.58, 46.90, 27.14.

(*E*)-3-(7-methoxy-9,9-dimethyl-9*H*-fluoren-2-yl)acrylaldehyde (10)

To a solution of compound **9** (44 mg, 0.16 mmol) in 3 mL dry tetrahydrofuran was added DIBAL (1 M in hexane, 0.5 mL) at -78 °C under argon. The reaction mixture was then allowed to warm to room temperature and was stirred overnight. Cold methanol was added dropwise to quench the reaction. The mixture was treated with saturate solution of Rochelle's salt and extracted with dichloromethane. The organic phase was dried over Na_2SO_4, then the solvent evaporated, and the product purified by flash chromatography (hexane/ ethyl acetate 95:5) to yield compound **10** as yellow solid (16 mg, 36%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 9.72 (d, *J* = 7.7 Hz, 1H), 7.69 – 7.64 (m, 2H), 7.62 – 7.51 (m, 3H), 7.00 (d, *J* = 2.4 Hz, 1H), 6.93 (dd, *J* = 8.4, 2.4 Hz, 1H), 6.78 (dd, *J* = 15.8, 7.7 Hz, 1H), 3.90 (s, 3H), 1.51 (s, 6H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 193.78, 160.60, 156.48, 153.96, 153.58, 142.86, 131.87, 130.93, 128.56, 127.32, 122.32, 121.67, 119.66, 113.12, 108.55, 55.57, 46.89, 27.16.

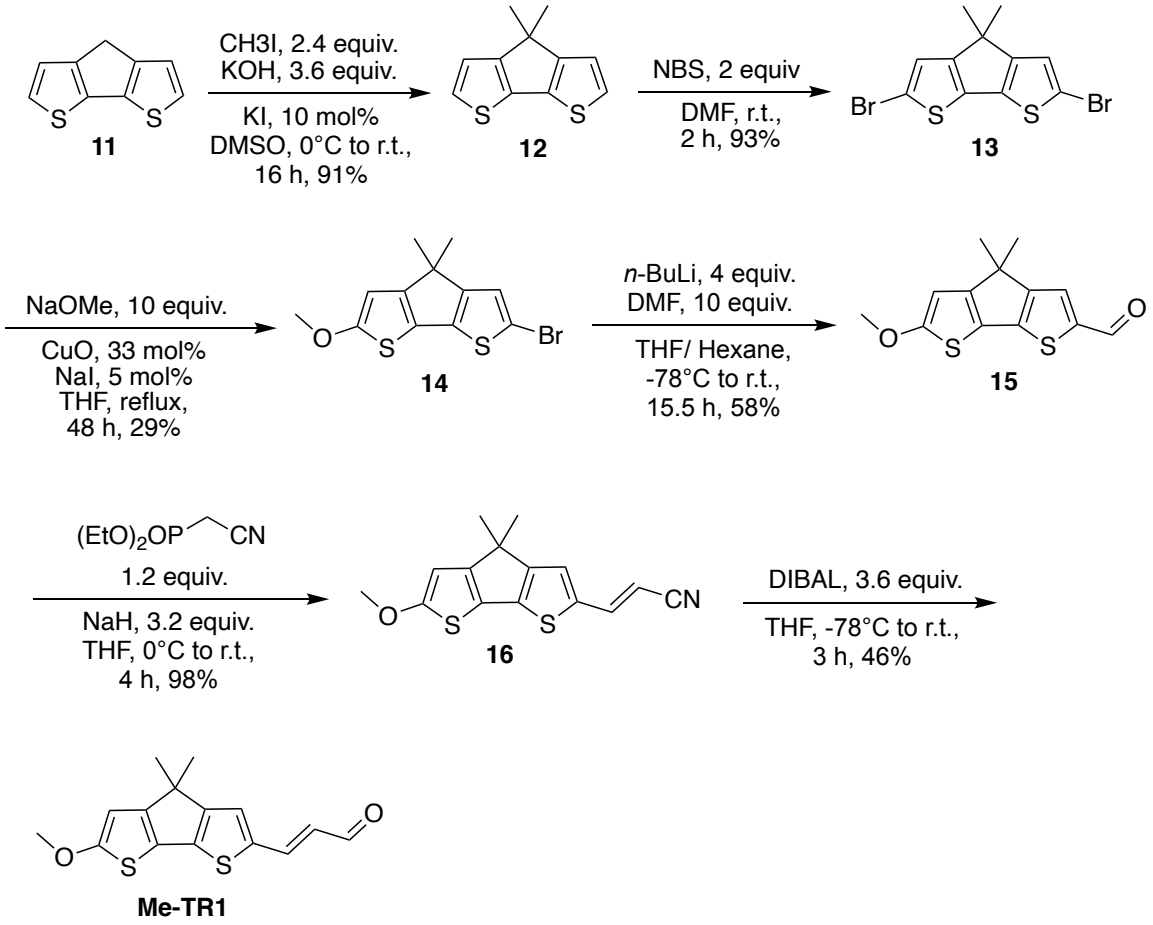
MR1

To a solution of compound **10** (52 mg, 0.2 mmol) in 3 mL dry DCM, BBr_3 (0.5 mL of 1 M solution in DCM, 0.5 mmol) was added dropwise at 0 °C and let the reaction warm to room temperature under nitrogen. After 4 h, reaction was quenched by ice water and extracted by DCM. The solvent evaporated and the product purified by flash chromatography (hexane/ ethyl acetate 9:1) to yield **MR1** as a yellow solid (14.7 mg, 28%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 9.72 (d, J = 7.8 Hz, 1H), 7.68 – 7.52 (m, 5H), 6.95 (d, J = 2.3 Hz, 1H), 6.86 (dd, J = 8.2, 2.3 Hz, 1H), 6.79 (dd, J = 15.8, 7.8 Hz, 1H), 1.49 (s, 6H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 194.16, 156.83, 156.72, 154.05, 153.89, 142.95, 131.79, 130.91, 128.65, 127.18, 122.37, 121.86, 119.61, 114.68, 110.09, 46.83, 27.10.

IV.7.6 Synthesis of Me-TR1



4,4-dimethyl-4*H*-cyclopenta[2,1-*b*:3,4-*b'*]dithiophene (12)

Potassium hydroxide (1.12 g, 20 mmol) was added to a solution of 4*H*-cyclopenta[2,1-*b*:3,4-*b'*]dithiophene **11** (1 g, 5.6 mmol), KI (93 mg, 0.56 mmol) and CH_3I (1.9 g, 13.4 mmol) in DMSO (33 mL) at 0 °C. Then the reaction was stirred at room temperature for 16 hours. After completion the reaction, organic phase was extracted with diethyl ether and washed multiple times with water to decrease the amount of DMSO. Then, combined organic layers were dried over Na_2SO_4 and after evaporating the solvent, the product was purified via flash alumina (activated) column chromatography with hexane as eluent to give compound **12** (1.04g, 91%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.15 (d, J = 4.9 Hz, 1H), 6.99 (d, J = 4.9 Hz, 1H), 1.45 (s, 3H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 155.93, 137.24, 124.72, 122.17, 38.92, 25.24.

2,6-dibromo-4,4-dimethyl-4*H*-cyclopenta[2,1-*b*:3,4-*b'*]dithiophene (13)

A solution of **12** (1g, 4.85 mmol) and NBS (1.75 g, 9.83 mmol) in 35 mL dry DMF (distilled and dried on molecular sieve) was stirred in dark at room temperature for 2 hours. The reaction mixture was poured into 50 mL of saturated sodium thiosulfate aqueous solution, and the product was extracted three times with hexane. The combined organic layers were washed with saturated sodium chloride solution and dried over Na_2SO_4 and filtered. The filtrate was concentrated and purified via flash alumina (activated) column chromatography by using hexane as eluent to give compound **13** as white solid (1.64 g, 93%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.00 (s, 2H), 1.41 (s, 6H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 158.30, 135.13, 123.96, 111.37, 46.42, 24.82.

2-bromo-6-methoxy-4,4-dimethyl-4*H*-cyclopenta[2,1-*b*:3,4-*b'*]dithiophene (14)

Sodium methoxide was prepared by adding sodium (150 mg, 6.5 mmol) into 3 mL of anhydrous methanol under nitrogen. When the sodium disappeared, a solution of compound **13** (236 mg, 0.65 mmol) in 2 mL of dry THF, copper oxide (17.8 mg, 0.22 mmol) and a catalytic amount of sodium iodide was added to the above solution and heated to reflux for 48 hours. The reaction mixture was poured

into water and then extracted with DCM. The organic layer was washed with brine and dried over Na_2SO_4. The crude was purified by alumina (activated) column chromatography (hexane/ ethyl acetate 99:1) to give compound **14** (59mg, 29%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 6.98 (s, 1H), 6.23 (s, 1H), 3.93 (s, 3H), 1.39 (s, 6H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 167.77, 156.08, 155.65, 136.31, 123.79, 120.06, 108.55, 99.36, 60.84, 46.76, 24.91.

6-methoxy-4,4-dimethyl-4*H*-cyclopenta[2,1-*b*:3,4-*b'*]dithiophene-2-carbaldehyde (15)

To a solution of compound **14** (438 mg, 1.39 mmol) in dry THF, *n*-BuLi (2.2 mL of 2.5 M solution in hexane, 5.56 mmol) was added dropwise at -78 °C under nitrogen. The reaction mixture was stirred for 1.5 hour at the same temperature, and an orange suspension was formed. DMF (1 mL) was added dropwise, and the formed solution was stirred for an additional 2 hour at -78 °C. Then the reaction mixture was warmed to room temperature and stirred for 12 hours. The reaction was quenched with water and was extracted three times with DCM. The combined organic phase was dried with Na_2SO4 and evaporated. The crude was purified by flash alumina (activated) column chromatography (hexane/ ethyl acetate 8:2) to yield pure compound **15** (212 mg, 58%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 9.77 (s, 1H), 7.57 (s, 1H), 6.28 (s, 1H), 3.98 (s, 3H), 1.46 (s, 6H).

(*E*)-3-(6-methoxy-4,4-dimethyl-4*H*-cyclopenta[2,1-*b*:3,4-*b'*]dithiophen-2-yl)acrylonitrile (16)

A solution of diethyl (cyanomethyl)phosphonate (155 mg, 0.87 mmol) in 4 mL tetrahydrofuran was added to a stirred suspension of NaH (60% on mineral oil, 94 mg, 2.35 mmol) in 4 mL tetrahydrofuran at 0 °C. The mixture was stirred at this temperature for 30 min. Then a solution of compound **15** (193 mg, 0.73 mmol) in 4 mL tetrahydrofuran was added. The reaction mixture was stirred at ambient temperature for 4 hours. The reaction mixture was then poured into cold water and extracted with ethyl acetate. The organic phase was dried over Na_2SO_4, then the solvent evaporated, and the product purified by flash alumina (activated) column chromatography (hexane/ ethyl acetate 85:15) to yield compound **16** as a yellow solid (205 mg, 98%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.40 (dd, J = 16.0, 0.6 Hz, 1H), 7.07 (s, 1H), 6.25 (s, 1H), 5.44 (d, J = 16.0 Hz, 1H), 3.96 (s, 3H), 1.41 (s, 6H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 170.34, 159.91, 157.99, 143.47, 141.78, 136.49, 124.78, 119.70, 119.27, 99.34, 89.06, 60.69, 46.24, 24.86.

Me-TR1

To a solution of compound **16** (145 mg, 0.50 mmol) in 9 mL dry tetrahydrofuran was added DIBAL (1 M in hexane, 1.8 mL) at -78 °C under argon. The reaction mixture was then allowed to warm to room temperature and was stirred for 3 hours. Cold methanol was added dropwise to quench the reaction. The mixture was treated with saturate solution of Rochelle's salt and extracted with

dichloromethane. The organic phase was dried over Na_2SO_4, then the solvent evaporated, and the product purified by flash alumina (activated) column chromatography (hexane/ ethyl acetate 8:2) to yield **Me-TR1** as yellow solid (66 mg, 46%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 9.56 (d, *J* = 7.8 Hz, 1H), 7.53 (dd, *J* = 15.4, 0.6 Hz, 1H), 7.20 (s, 1H), 6.41 (d, *J* = 15.4 Hz, 1H), 6.26 (s, 1H), 3.97 (s, 3H), 1.44 (s, 6H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 192.57, 170.64, 160.19, 158.41, 145.85, 143.76, 137.48, 125.72, 123.40, 119.93, 99.36, 60.68, 46.20, 24.92.

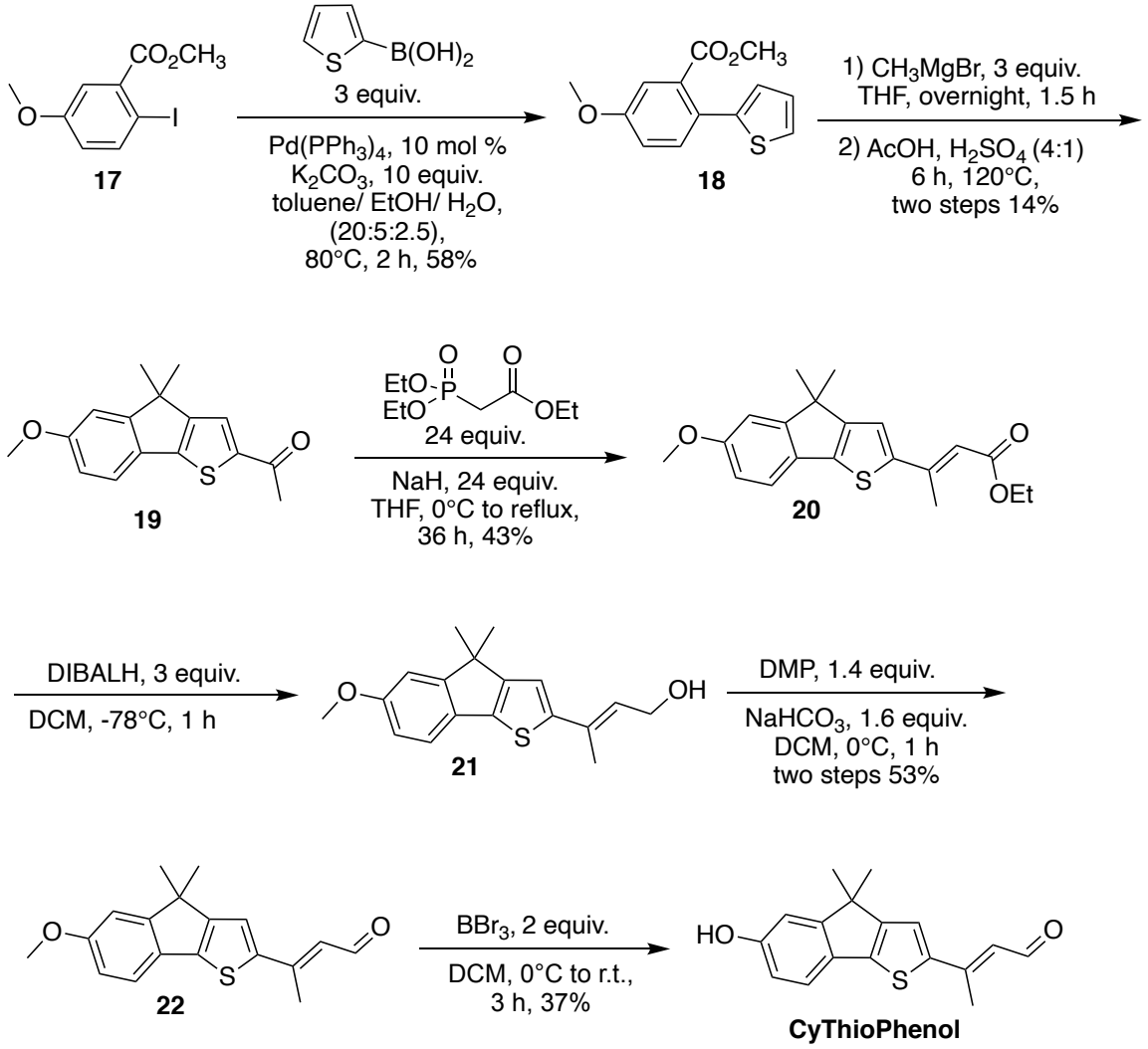
IV.7.7 Synthesis of CyThioPhenol

methyl 5-methoxy-2-(thiophen-2-yl)benzoate (18)

To the mixture of methyl 2-iodo-5-methoxybenzoate **17** (292 mg, 1.0 mmol), $Pd(PPh_3)_4$ (116 mg, 0.1 mmol), and toluene (20.0 mL), a solution of thiophen-2-ylboronic acid (384 mg, 3 mmol) in ethanol (5 mL) and K_2CO_3 (1.38 g, 10 mmol) in water (2.5 mL) were added. The mixture was heated to 80 °C in an argon atmosphere for 2 hours. After cooling to room temperature, the mixture was filtered through Celite, and the organic phase was washed with water. The organic layers were then dried over Na_2SO_4, filtered, and concentrated under vacuum. Column chromatography with (hexane/ ethyl acetate 19:1) of the residue afforded compound **18** as a colorless oil (143 mg, 58%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.41 (d, J = 8.5 Hz, 1H), 7.31 (dd, J = 5.1, 1.2 Hz, 1H), 7.26 (d, J = 2.8 Hz, 1H), 7.07 – 7.01 (m, 2H), 6.97 (dd, J = 3.5, 1.2 Hz, 1H), 3.87 (s, 3H), 3.74 (s, 3H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 168.86, 159.04, 141.99, 132.53, 127.09, 126.58, 125.86, 125.33, 117.17, 114.21, 55.57, 52.29.



1-(6-methoxy-4,4-dimethyl-4*H*-indeno[1,2-*b*]thiophen-2-yl)ethan-1-one (19)

In a flame dried and under argon atmosphere, compound **18** (248 mg, 1.0 mmol) was dissolved in 5 mL dry THF and methyl magnesium bromide (1M in THF, 3.5 mL, 3.5 mmol) were added dropwise over 1.5 hours at room temperature. The reaction mixture was stirred under argon overnight at same temperature. The reaction was quenched with 1M HCl and extracted by DCM. The organic layer was then dried over Na_2SO_4 and concentrated under vacuum. The tertiary alcohol product was used for the next step without further purification. A mixture of 4:1 glacial acetic acid and 98% sulfuric acid (7 mL) was added, and the mixture heated at 120 °C for 6 hour. After cooling the reaction, it was quenched with 25 M solution of ammonia and extracted with DCM three times. The organic layers were then dried over Na_2SO_4, filtered, and concentrated under vacuum. Column chromatography with (hexane/ DCM 5:1) of the residue afforded pure compound **19** (38 mg, 14%).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.61 (d, J = 0.7 Hz, 1H), 7.46 (d, J = 8.3 Hz, 1H), 6.95 (d, J = 2.4 Hz, 1H), 6.86 (dd, J = 8.3, 2.3 Hz, 1H), 3.88 (s, 3H), 2.57 (d, J = 0.7 Hz, 3H), 1.48 (s, 6H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 190.70, 159.82, 159.16, 157.11, 149.03, 144.92, 128.34, 126.28, 121.40, 112.45, 109.29, 55.62, 45.91, 26.58, 26.23.

ethyl (*E*)-3-(6-methoxy-4,4-dimethyl-4*H*-indeno[1,2-*b*]thiophen-2-yl)but-2-enoate (20)

Triethyl phosphonoacetate (1.98 mL, 10 mmol) was added dropwise under nitrogen to a suspension of NaH (60% on mineral oil, 400 mg, 10 mmol) in dry THF (5mL) at 0 °C and the resulting mixture was stirred for another 30 min at the same temperature. A solution of compound **19** (170 mg, 0.62 mmol) in 2 mL THF was added dropwise to the reaction mixture and heated to reflux. The reaction followed by TLC and after 24 hours, the reaction was not complete. Then the reaction mixture was cooled to 0 °C and NaH (60% on mineral oil, 200 mg, 5 mmol) and triethyl phosphonoacetate (0.99 mL, 5 mmol) were added to the mixture and heated to reflux for 12 hours. After completion of the reaction, the mixture was brought to room temperature and quenched with saturated NH_4Cl and extracted three times with ethyl acetate. The organic layers were then dried over Na_2SO_4 and concentrated under vacuum. Column chromatography with (hexane/ ethyl acetate 96:4) of the residue afforded a mixture of *trans*/*cis* (2:1) isomers of compound **20** (91 mg, 43%).

(*E*)-3-(6-methoxy-4,4-dimethyl-4*H*-indeno[1,2-*b*]thiophen-2-yl)but-2-en-1-ol (21)

To the mixture of *cis*/*trans* isomers of compound **20** (127 mg, 0.37 mmol) in dry DCM (10 mL) at -78 °C was added DIBALH (1M in hexane, 1.1 mL, 1.1 mmol) dropwise. After 1 hour, the reaction was complete, and methanol (5 mL) was added dropwise to quench the reaction. Then 15 mL of saturated solution of potassium sodium tartrate was added with 10 mL of DCM. The mixture was stirred for 2 h and extracted with DCM (3 × 10 mL) and the combined organic layers were dried over

Na_2SO_4. Evaporation of solvent gave compound **21** as a yellow solid and it used in next step without further purification.

(*E*)-3-(6-methoxy-4,4-dimethyl-4*H*-indeno[1,2-*b*]thiophen-2-yl)but-2-enal (22)

To compound **21** (used directly from last step) in dry DCM (10 mL) at 0 °C, Dess-Martin periodinane (DMP) (207 mg, 0.5 mmol) and $NaHCO_3$ (50 mg, 0.6 mmol) were added. The reaction mixture was stirred at this temperature for 1 hour and after completion of reaction it was quenched by saturated sodium thiosulfate (10 mL) and saturated sodium bicarbonate (10 mL). The mixture was extracted with DCM (3 × 10 mL) and the combined organic layers were dried over Na_2SO_4. After evaporating the solvent, the product was purified via column chromatography by using (hexane/ ethyl acetate 9:1) as eluent to give compound **22** as a mixture of *trans*/*cis* (2:1) isomers (58 mg, two steps 53%).

CyThioPhenol

To a solution of compound **22** (89 mg, 0.3 mmol, mixture of *trans*/*cis* 2:1) in 3 mL dry DCM, BBr_3 (0.6 mL of 1 M solution in DCM, 0.6 mmol) was added dropwise at 0 °C and let it warm to room temperature under nitrogen. After 3 h, reaction was quenched by ice water and extracted by DCM. The solvent was evaporated, and it purified by flash chromatography (hexane/ ethyl acetate 9:1) to yield trans **CyThioPhenol** as an orange solid (31.5 mg, 37%)**.**

^{1}H NMR (500 MHz, Acetone-d_6) δ 10.13 (d, J = 7.7 Hz, 1H), 8.63 (s, 1H), 7.72 (s, 1H), 7.39 (d, J = 8.1 Hz, 1H), 7.01 (d, J = 2.2 Hz, 1H), 6.83 (dd, J = 8.2, 2.3 Hz, 1H), 6.35 (dq, J = 7.7, 1.1 Hz, 1H), 2.64 (d, J = 1.1 Hz, 3H), 1.47 (s, 6H).

^{13}C NMR (126 MHz, Acetone-d_6) δ 189.67, 158.97, 157.73, 157.42, 150.50, 144.84, 143.71, 127.31, 122.93, 122.03, 120.66, 114.17, 110.49, 45.75, 25.37, 15.06.

REFERENCES

REFERENCES

1 Wang, W. In PhD Thesis. *Michigan State University* (2012).

2 Wang, W. *et al.* Tuning the electronic absorption of protein-embedded all-trans-retinal. *Science* **338**, 1340-1343, doi:10.1126/science.1226135 (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. *et al.* "Turn-On" Protein Fluorescence: In Situ Formation of Cyanine Dyes. *Journal of the American Chemical Society* **137**, 1073-1080, doi:10.1021/ja506376j (2015).

5 Yapici, I. In PhD Thesis. *Michigan State University* (2015).

6 Santos, E. M. In PhD Thesis. *Michigan State University* (2017).