ENGINEERING OF THE SMALL CYTOSOLIC RETINOID BINDING PROTEINS
INTO PH-SENSING PROBES AND NOVEL FLUORESCENT PROTEIN TAGS

By

Tetyana Berbasova

A DISSERTATION

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

Chemistry – Doctor of Philosophy

2014

ABSTRACT

ENGINEERING OF THE SMALL CYTOSOLIC RETINOID BINDING PROTEINS INTO PH-SENSING PROBES AND NOVEL FLUORESCENT PROTEIN TAGS

By

Tetyana Berbasova

Reprogramming the biological systems to perform a process other than that dictated by nature illustrates the power of the modern synthetic biology. Protein engineering represents one of the research areas in this field. Crystal structure guided mutagenesis studies allow for a rational protein redesign. Utilizing this *de novo* approach, Cellular Retinoic Acid Binding Protein II (CRABPII) has been reengineered into a Schiff base forming protein. The reaction is performed deep inside the protein cavity and fascilitates a complete ligand closure from the aqueous media.

An initial aim of this project was to design a functioning rhodopsin mimic from CRABPII for the wavelength regulation studies. The strategy combined the principles we learned in the earlier work on CRABPII and human Cellular Retinol Binding Protein II (hCRBPII). The rational modifications of the CRABPII binding cavity provides the effective charge delocalization along the protein-embedded all-*trans*-retinal-PSB that leads to the regulation of absorption. Reengineered CRABPII is capable of spanning the visible spectrum in the range of 474 – 640 nm. Additionally, CRABPII mutants exhibit an extraordinary range of the iminium pK_a values, ranging from 1.8 to 8.1. The latter phenomena allowed for the chromophoric pH-probe design utilizing the protein-bound chromophoric aldehyde,

an isoform of Vitamin A. Moreover, the fusion of the two selected pigments provides a ratiometric protein based pH-sensor.

The aldehydes other than retinal might yield alternative pH-sensing ligands with the engineered CRABPII and hCRBPII mutants. With fluorescence being the most widely used detection method in biology, fluorescent and fluorogenic retinal analogs are the primary targets in the ligand redesign. Synthetic julolidine retinal analog meets most of the criteria for an effective fluorogenic probe with pH-sensing properties upon PSB formation. In the protein-bound form this chromophore accounts for the imine-iminium equilibrium with changes in fluorescence intensity. Only iminium yields a light emitting pigment. Moreover, the wavelength shift in response to the pH changes provided the platform for a single protein ratiometric pH-probe design. This pH-dependent shift of absorption is a result of the more effective charge delocalization along the chromophore after protonation of the anionic carboxylate residue in the iminium region.

Merocyanine retinal analog shows bright fluorescence and penetrates cells easily. These are valuable criteria for live-cell imaging applications. Optimization of the reaction half-life between hCRBPII mutants and merocyanine retinal analog provided an hCRBPII-tag with instantaneous red fluorescence. Designed method comprises a unique fluorescence recovery after the photobleaching when the cells are treated again with merocyanine.

ACKNOWLEDGEMENTS

friend and we often communicate even after she moved to Boston. She is the super-glue that keeps all completely different people together and even arranges a trip to Las Vegas. I am happy I had Ipek Yapici as a Bio-Labmate through most of my time at MSU, because she fixed all the protocols to perfection. But most of all I am grateful for her honesty about everything, even at times when it is hard to accept. I want to thank Elizabeth Santos for being eager to continue my projects, for proofreading my thesis and for being a sweet person with whom I always feel free to talk.

I also want to thank Professor Jim Geiger for the thoughtful discussions at the biomeetings about experiments and about hypothetical explanations. Under his guidance we made a huge progress in the live-cell imaging applications and we designed the experiments possible only with our developed method. I am also very grateful to Meisam Nosrati as a collaborator. With numerous protein structures he solved, we were able to built the new hypothesis to the previously inconclusive results.

I am heartily thankful to the past and present members of the Borhan's lab, for the friendly environment and for all fun parties we had together. I especially should thank Kumar Ashtekar for inspiring me to approach research problems as an opportunity to learn. Being compassionate of his own chemistry, Kumar taught me to choose the most promising part of the project and to be brave to explore the unknown routes of synthetic problems. In addition to being like a synthetic advisor to me, Kumar was always supportive and caring.

I would like to thank Dr. Melinda Frame for her assistance with confocal microscopy; Professor Dan Jones for his help with protein mass spectrometry; and Dr. Lijun Chen for her help with HRMS. I want to thank Babak, Chrysoula, Elizabeth, Kumar and my committee members, Professor Maleczka, Professor Geiger and Professor Walker, for proof-reading my thesis.

Last but not least, I am thankful to my parents, my grandma and my grandfather for their understanding and support. But most of all I am thankful to my brother, Dima Berbasov, for introducing me to the MSU Graduate School and his help during my early days in Michigan. He was the driving force for me going to the graduate school because he insisted on me applying to MSU when I was not sure what to do with my life. Dima is the person who inspired me to become a scientist. I am most thankful to Dima for that.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

xxi

LIST OF SCHEMES

KEY TO SYMBOLS AND ABBREVIATIONS

Symbols

Å	Angstrom
ε	Extinction Coefficient
λ_{max}	Maximal Wavelength
λ_{ex}	Excitation Wavelength
λ_{em}	Emission Wavelength
nm	Nanometer
cm	Centimeter
M	Molar
mM	Millimolar
µM	Micromolar
nM	Nanomolar
mg	Milligram
mmol	Millimole
>	Larger than

Abbreviations of Amino Acids

Ala, A	Alanine
Arg, R	Arginine
Asn, N	Asparagine

Asp, D	Aspartate
Cys, C	Cysteine
Gln, Q	Glutamine
Glu, E	Glutamic acid
His, H	Histidine
Ile, I	Isoleucine
Leu, L	Leucine
Lys, K	Lysine
Met, M	Methionine
Phe, F	Phenylalanine
Pro, P	Proline
Ser, S	Serine
Thr, T	Threonine
Trp, W	Tryptophan
Tyr, Y	Tyrosine
Val, V	Valine

All Other Abbreviations

AM	Acetoxymethyl
amu	Atomic Mass Units
ATP	Adenosine Triphosphate
BCECF	2',7'-Bis-(2-Carboxyethyl)-5-(and-6)-Carboxyfluorescein
BODIPY	Boron-dipyrromethene

BRET	Bioluminescence Resonance Energy Transfer
CCVJ	9-(2-carboxy-2-cyanovinyl)-julolidine
cGMP	Cyclic Guanosine Monophosphate
CRABPII	Cellular Retinoic Acid Binding Protein II
hCRBPII	Human Cellular Retinol Binding Protein II
CRBP$_{nona}$	hCRBPII-Q108K:K40L:T51V:T53C:R58W:T29L:A33W:Q4F:L117E
DCVJ	9-(2,2-dicyanovinyl)-julolidine
DIBAL	Diisobutyl Aluminum Hydride
DIC	Differential Interference Contrast
DMF	Dimethylformamide
ECFP	Enhanced Cyan Fluorescent Protein
ES	Electrospray
ESI	Electrospray Ionization
ESPT	Excited State Proton Transfer
Et$_3$N	Triethylamine
EtOAc	Ethyl Acetate
FAP	Fluorogen-Activated Peptide
FLIM	Fluorescence-Lifetime Imaging Microscopy
FP	Fluorescent Protein
FRET	Förster Resonance Energy Transfer
GDP	Guanosine Diphosphate
GFP	Green Fluorescent Protein

GPCR	G-Protein Coupled Receptor
GTP	Guanosine Triphosphate
HOMO	Highest Occupied Molecular Orbital
HRMS	High-Resolution Mass Spectrometry
HWE	Horner-Wadsworth-Emmons
iLBP	Intracellular Lipid Binding Protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JRA	Julolidine Retinal Analog
K_d	Dissociation Constant
LiHMDS	Lithium Bis(trimethylsilyl)amide
LUMO	Lowest Unoccupied Molecular Orbital
m. pt.	Melting Point
mbar	Milibarr
MCRA	Merocyanine Retinal Analog
mkMCRA	Methyl Ketone Merocyanine Retinal Analog
$MgSO_4$	Magnesium Sulfate
mRFP	Monomeric Red Fluorescent Protein
n.a.	Not Applicable
n.d.	Not Determined
n.r.	Not Ratiometric
Na_2SO_4	Sodium Sulfate

NaH	Sodium Hydride
NaHMDS	Sodium bis(trimethylsilyl)amide
NaOH	Sodium Hydroxide
NIR	Near-Infrared
NMR	Nuclear Magnetic Resonance
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PeT	Photoinduced Electron Transfer
PMT	Photomultiplier Tube
PROPS	Proteorhodopsin Optical Proton Sensor
PSB	Protonated Schiff Base
QY	Quantum Yield
RA	Retinoic Acid
Rt	All-*trans*-retinal
SB	Schiff Base
SNARF	Seminaphthorhodafluor
TBAF	Tetra-*n*-butylammonium fluoride
TLC	Thin Layer Chromatography
TMP	Trimethoprim
TO	Thiazol Orange
UV	Ultraviolet Light
UV-vis	Ultraviolet and Visible Light

WT   Wild Type

YFP   Yellow Fluorescent Protein

CHAPTER I: DESIGN OF pH-SENSING AND FLUORESCENT PROBES: AN OVERVIEW

I.1 Importance of pH in biological systems

A number of critical biological processes are governed by both the inter- and intra-cellular pH of the cytosol and extracellular milieu. These include cell metabolism and growth, ion current through ion channels, the endocytic pathway,[1] synaptic transmission in neurons,[2-3] and cellular contraction in muscles,[4] to name a few. Acidity also plays an integral role in viral pathogenesis, as observed in the low pH endosome activation of viral membrane fusion factors.[5-6] Several studies have shown a correlation between low pH and both tumor invasion and metastasis.[7-11] Many of the genes that are up-regulated in hypoxia and correlate with increased metastatic capacity are also up-regulated by low pH.[12] Additionally, the different cellular compartments have distinct pH values, from highly acidic (pH < 4.5 in late endosomes) to moderately basic (pH > 8 in mitochondria) contributing to the optimal operation of individual metabolic pathways.[13]

Many reports in recent years have focused on investigating the relationship between pH and cancer. It is already well established that tumor cells have a reversed pH gradient across the cell membrane compared to normal cells. They have more alkaline intracellular pH (pH_i) compared to extracellular pH (pH_e). This phenomena creates a tumor specific metabolic microenvironment,

important for tissue development and tumor growth. The reversed pH gradient increases with disease progression.[14]

Tumor cells secrete protons and acidify their environment through membrane-based ion regulators such as Na^+/H^+ exchangers (NHE) and monocarboxylate transporters (MCT). Controlled expression of these genes has revealed that the reversed pH gradient is apparent during the earliest steps of tumor development, favoring cell survival.[14] Moreover, recent reports suggest that protons diffuse from the proximal tumor microenvironment into adjacent normal tissue with further remodeling and invasion.[15]

One of the underlying causes of tumor acidosis is a high rate of glycolysis, leading to lactic acid production and glucose uptake.[16] Glycolysis provides the cell with energy in the form of two molecules of ATP from each molecule of glucose. Presently, up-regulated expression of glycolytic enzymes is a known indicator of tumor aggressiveness. Because cancer develops and expands rapidly, the supporting blood vessels are often limited and disorganized. This leads to a condition of oxygen deficiency in the tumor tissue known as hypoxia. Additionally, perturbation of the mitochondrial respiration and regular pathway of ATP production is observed, which forces the cells to increase glycolysis.[17] Hypoxia, together with acidosis, are two factors always associated with malignant cells despite the oncogene mutation present.[18]

Tumor acidosis is commonly linked to multidrug resistance in anticancer therapy, cell migration (responsible for tumor distribution) and the inhibition of the natural pathway for damaged cells degradation (by natural killer (NK) cells and cytotoxic T lymphocytes).[7] This intrinsic property of tumor cells is being more widely targeted for drug delivery in chemotherapy. Hence, the necessity for highly accurate pH measurement is critical. This would allow for the building of pH specific systems for drug release or drug delivery. Currently, acid-dependent drug release and nanoparticle disassembly are among the most promising strategies.[19-20]

Besides being a hallmark of various cancers there are other acidity related pathological processes that occur in humans. It has been recently suggested that chirality of amyloid fibrils can change based on pH: "normal" or left-handed assembly were formed at pH 2 and higher, while more acidic conditions led to reverse chirality of aggregates.[21] The fibril morphology is linked to cell toxicity and, therefore, acidity can play a vital role in amyloid-associated diseases.

Many proteins can undergo conformational changes with pH perturbation. The presence of titratable protons in amino acid side chains is the main origin of such changes. Protonation and deprotonation events can influence enzymatic activity, modify affinity to substrate, influence protein-protein interactions, and cause protein diffusion or aggregation.[22] The key factors underlying such global changes are the hydrogen bonding networks and/or salt bridges. The naturally occurring amino acids have side chain pK_a values in the biologically relevant pH

range: acidic (aspartate and glutamate, ~4), neutral (histidine, ~6.5) and basic (lysine and arginine, ~10). Therefore, changes in pH are directly reflected in the protonation ability of these residues and the subsequent interactions they are involved in.

For example, most of the hydrolases contain a histidine residue as a part of their catalytic triad. Protonation of the histidine, with a pK_a ~6.5, will have a dramatic effect on its catalytic ability after acidification. A classical example of pH-dependent allosteric regulation is oxygen binding by hemoglobin, which is controlled by an aspartate-histidine interaction.[23] In this case, even a slight change to higher pH breaks the salt bridge and destabilizes the hemoglobin quaternary structure. This leads to oxygen release due to lower affinity after pH-dependent conformational change. Another example of a pH-dependent process is invasion of a host cell by influenza virus, in which a viral protein (haemagglutinine) undergoes a major rearrangement in tertiary structure to facilitate the fusion of viral and host membranes.[24]

This has led to an increased interest in accurate pH measurements in tissues and inside living cells. Among the most common methods of pH determination are NMR, fluorescence spectroscopy and invasive measurements with microelectrodes. Fluorescence-based methods are relatively easier to perform. They have high sensitivity and fluorophore delivery is most often

nondestructive to the cells. For this reason, *in vivo* pH sensors have attracted substantial attention.

For the most part, small organic molecules and nanoparticles are used as general pH-indicators and as intracellular pH sensors.[25] Specific localization in cellular compartments and potential toxicity with these agents, however, is hard to control, plan, or predict, in particular for live-cell imaging applications. It is also particularly important to be able to localize these sensors, as different cellular domains and organelles are known to have variable pH ranges. Alternatively, DNA-encoded protein-based probes do not have the latter disadvantages, but have the potential to be superior agents for pH sensing applications. Genetic triggers for the production of these probes afford temporal control, while their ability to target specific locales provides spatial resolution.

Ideally, pH sensing systems should be independent of probe concentration for accurate measurements that are not altered by intensity output. A general solution to this problem is the use of ratiometric probes, that enable concentration independent pH determinations. This is most often accomplished by having differential absorption of two sensors at the same pH. Protein-based ratiometric pH probes combine all of the aforementioned desirable characteristics.

I.2. Small molecules for intracellular pH determination.

In general an effective pH indicator requires the presence of titratable protons in the desired pH ranges that cause a significant change in spectroscopic

5

properties. Most often this is achieved by the introduction of ionizable groups on

fluorescent molecules. Another approach involves a change in the structure of
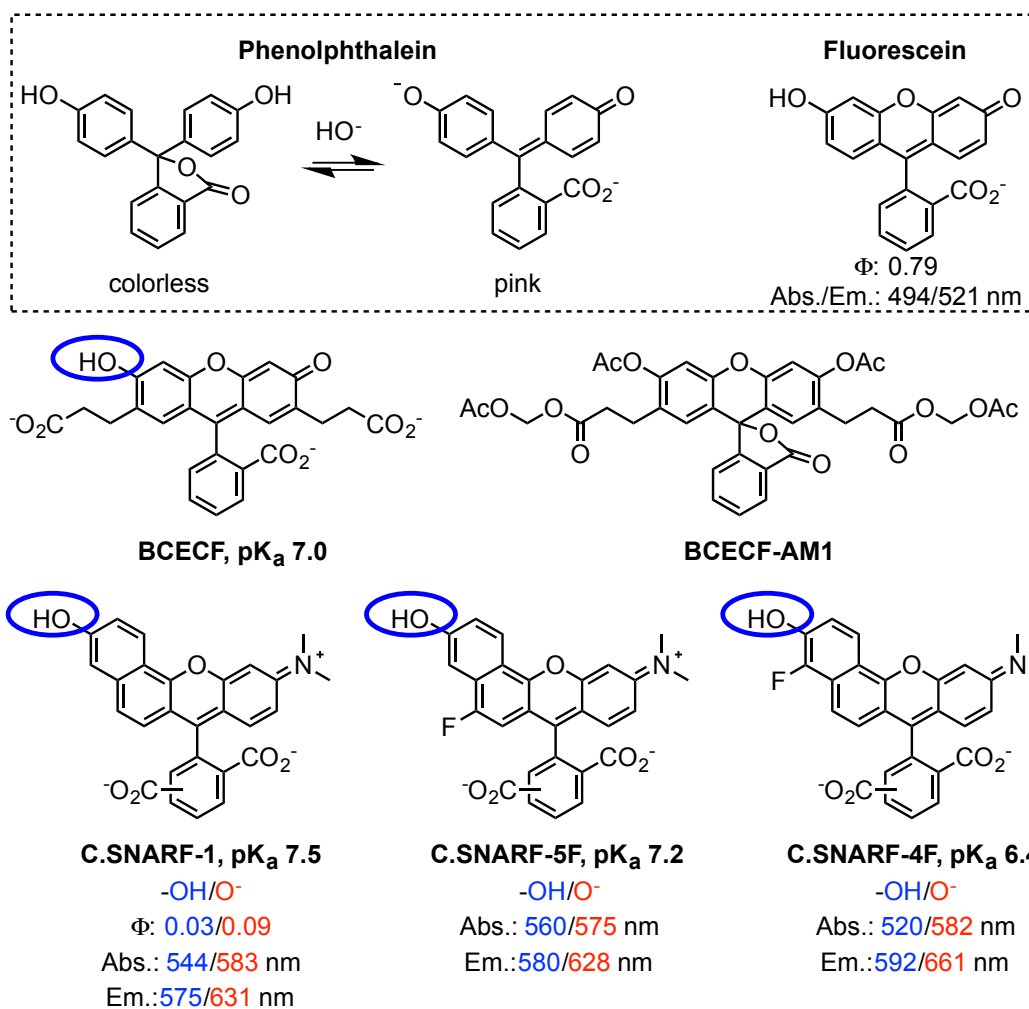
Figure I-1: Structures of phenolphthalein indicator, fluorescein fluorophore and the most widely used intracellular pH-sensors BCECF and SNARF. Ionizable with pH group is circled blue.

indicator. For example, the ring opening of phenolphthalein at basic pH leads

spectroscopic changes. Various types of small molecule based intracellular pH

indicators were reviewed in detail by Han and Burgess in early 2010.[25] The

review also summarizes the practical application and limitations for each group of

compounds. In this section, the most widely used representatives from each

family of pH sensitive molecules are highlighted. Also reports published after 2010, with unique approach to pH-sensor design, are described.

The most widely used intracellular pH indicator is 2′,7′-bis-(2-carboxyethyl)-5-carboxyfluorescein (BCECF). Its cell-permeable form is protected by intracellular cleavable acetoxylmethyl ester (AM) (**Figure I-1**).[26-28] With a pK_a of 7.0 this molecule is suitable for most biologically relevant measurements. Additionally, the presence of charged groups keeps it from leaking out of the cell through the hydrophobic membrane. Moreover, the absorption profile of BCECF changes significantly with pH, which allows its use as a dual excitation probe, for ratiometric measurements, independent of concentration. In this approach, fluorescence intensity ratios, corresponding to excitation at 440 nm and 490 nm, are measured and fit to a standard curve for pH determination.[28]

The second most widely used chromophore is **c**arboxy-**s**emi**na**phto**r**hodamine **f**luorophore C.SNARF-1 (**Figure I-1**). Similar to BCECF, it can be protected by a cleavable acetoxylmethyl (AM) ester, to increase cell permeability. During AM hydrolysis inside the cell, the phenolphthalein-like ring closed form opens to a carboxylate. C.SNARF-1 with pK_a 7.5, was used for accurate pH determination in the cytosol, mitochondria and nucleus, by the emission ratio from excitations at 580 nm and 640 nm.[29-35] A longer excitation wavelength, compared to BCECF, aids in reducing cell damage and minimizes autofluorescence. The introduction of fluorine groups onto SNARF decreases the

7

pKa of the probe (**Figure I-1**).[25] The shared mechanism of structural response to

a change in pH, for both BCEDF and C.SNARF-1, is due to the phenolate

formation from the phenolic form at higher pH, which proceed with a red-shift in

absorption/emission spectra, and an increase in fluorescence intensity. However,

because most of the pathological conditions are acidic, it would be ideal to design

a system that elicits a response upon acidification rather than basification.
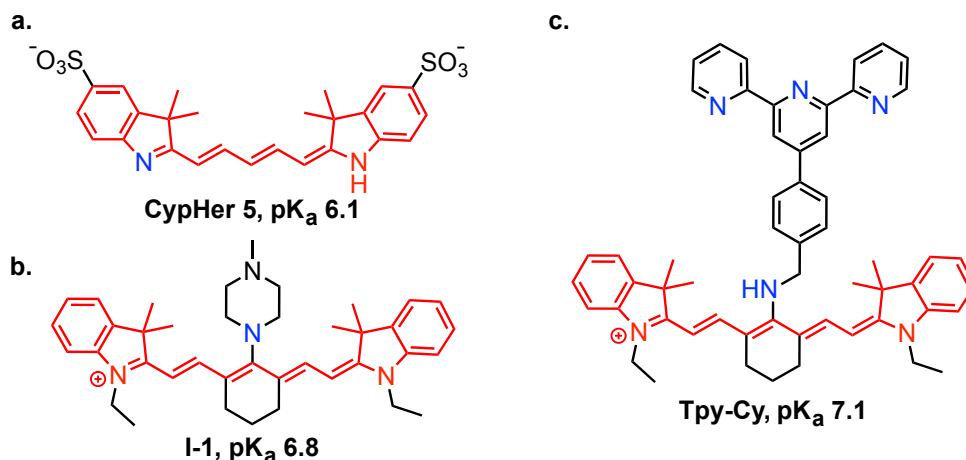
Figure I-2: Structures of pH-sensitive cyanine dyes. Cyanine core is red, titratable groups are blue.

Both BCECF and SNARF share the same fluorescein core (**Figure I-1**).

Fluorescein and its analogs have high quantum yield, good photostability and

tolerate a variety of attached moieties.[25]

Cyanine dyes (**Figure I-2**) constitute another family of fluorophores that

show pH-sensitivity. Characteristic features of this family are large extinction

coefficients and red-shifted spectra, with some approaching the near-infrared

(NIR) region.

Most of the fluorescent cyanine dyes contain a nitrogen-locked indoline ring, creating a resonating positive charge along the polyene. The non-*N*-alkylated indoline has an amine on one side and imine on the other.[36] The resonating positive charge appears after protonation of the imine under acidic conditions. This is reflected in the increased fluorescence and bathochromic shift in λ_{max}. These properties were successfully utilized to design pH sensors with a cyanine core. A commercially available dye of this kind is called **CypHer 5** (**Figure I-2a**).[37-40] It absorbs at 650 nm after protonation, and 450–520 nm in the basic form. Its pK_a, 6.1, allows for the monitoring of receptor internalization in the endosomes.[39]

Another common modification to the cyanine dye is an additional amine group on the polyene chain, while keeping the *N*-alkylated indoline **I-1**[41] (**Figure I-2b**). In such a case, the protonated form provides the positive charge in the middle of the resonating polyene, which extends conjugation and causes the red-shift. The advantage of aminocyanines is a strong fluorescent signal, both before and after protonation of the amine, because the main cyanine core is constantly positively charged. This has led to the construction of dual excitation ratiometric pH probe with NIR fluorescence and pK_a values 5.1–8.7, based on the amine titration.[41]

Another type of pH-sensitive cyanine probes consists of a cyanine fluorophore and a nitrogen containing chromophore, attached to the middle of the

polyene chain. Such an assembly provides the donor/acceptor mechanism for

Förster resonance energy transfer (FRET) or for photoinduced electron transfer

(PeT). **Tpy-Cy** (**Figure I-2c**) is an example of pH sensor with pK$_a$ near neutral,[42]

that proceeds through a PeT quenching mechanism. Protonation of the

terpyridine chromophore changes the absorption profile of this modulator.

Subsequently, quenching through PeT, is interrupted. Thus, the cyanine dye

becomes fluorescent. The dye has shown utility for accurate pH determination in

living cells, in the pH range between 6.7 and 7.9.[42]
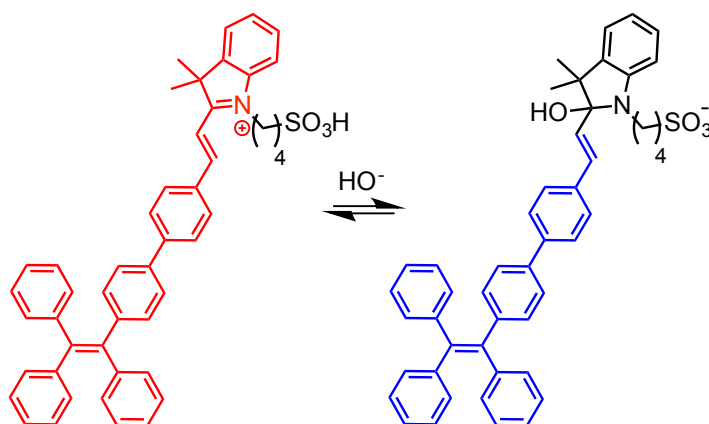
Figure I-3: Structure and chemical response to hydroxide of TPE-Cy dye. Absorption/emission profile changes based on conjugation.

In addition to sensing the proton by direct protonation/deprotonation

events, some cyanine dyes are highly susceptible to nucleophilic addition.

Essentially the iminium can accept a hydroxide ion and break the conjugation of

the chromophore. As a result, the less conjugated chromophore has much more

blue-shifted excitation and emission wavelengths. This property of cyanine was

used to design **TPE-Cy**[43] (**Figure I-3**). After reaction with hydroxide, changes not

only in wavelength were observed, but also the molecule polarity and extent of

aggregation were shown to affect fluorescence intensity. In this way the single molecule probe responds differently in each pH range: (1) it fluoresces at 630 nm if pH < 5, (2) it has a gradual loss of red intensity at pH 5–9, due to deprotonation of the sulfonic acid, with increased hydrophilicity and perturbed aggregation, (3) it fluoresces at 480 nm if pH > 10. Such a unique design of combined reactivity and polarity, provided a probe active over the full pH range of 1–14. TPE-Cy has been used in dual color imaging and "blue-to-red" ratio detection, for accurate pH measurements in live HeLa cells.[44]

BA, pK$_a$ 1.35
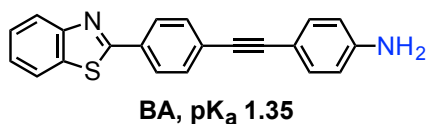
Figure I-4: Structure of benzothiazole aniline fluorophore for sensing low pH.

In a recent report, the intramolecular charge transfer in semi-cyanine was exploited as a pH indicator of extreme acidity (**Figure I-4**).[45] The design included a combination of a benzothiazolphenyl fluorophore and aniline, linked through a triple bond. The resulting 4-((4-(benzothiazol-2-yl)phenyl)ethynyl)aniline (**BA**) had a fluorescence (λ_{max} = 400 nm) increase upon acidification from pH 4.35 to 0.30. The quantum yield was 12% at pH 0.5 or below. Interestingly, **BA** in the amine form exhibited a significant emission wavelength shift from 447 nm in toluene to 562 nm in DMF.

Another group of pH-sensitive fluorophores is known by the name "LysoSensors^{TM}". These molecules were prepared as pH-sensors with fluorescence increase after protonation. They most often localize in the lysosome

because it is an acidic compartment in the cell. However, if other areas have similar acidity then LysoSensors will also fluoresce in those locations. These compounds do not share any specific structural features but typically the protonation of the nitrogen leads to increase in fluorescence. For example, the dye, known as LysoSensor Green or DND-189 (**Figure I-5**), has no fluorescence at the pH above its pK_a. However, it fluoresces with a quantum yield of 0.48 after acidification.[46-47]

Probably the most unique and widely used dye of this kind is LysoSensor[TM] Yellow/Blue, also called DND-160 (**Figure I-5**). It is an oxazole probe (Dapoxyl family of fluorophores), conjugated with pyridine that can undergo protonation. It has a pH-dependent spectral shift of emission wavelength from 464 nm to 542 nm after acidification with no difference in quantum yield between two forms (~0.4). Having a pK_a of approximately 4.5, it has been used for dual emission imaging of lysosomal pH. This dye has good cell permeability and

DND-189, pK$_a$ 5.2
-NH$^+$
Abs.: 443 nm
Em.:505 nm

DND-160, pK$_a$ 4.5
-N/NH$^+$
Abs.: 329/385 nm
Em.:464/542 nm
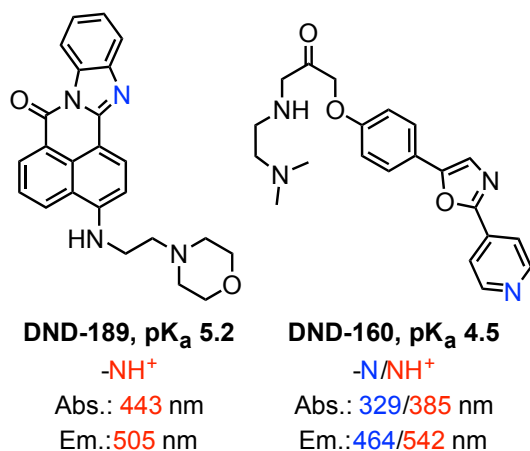
Figure I-5: LysoSensor dyes with pH-sensitive properties.

photostability. However, relatively short excitation wavelength (365 nm) leads to significant autofluorescence and cell damage.[48-56]

The last class of pH-sensitive fluorescent molecules is derived from boron-dipyrromethene (BODIPY), as a fluorescent core, decorated by additional subunits that quench fluorescence in response to pH changes (**Figure I-6**). The BODIPY core itself does not have any group that responds to protonation. However, synthesis of this chromophore family is highly efficient, making it relatively easy to achieve a longer wavelength by increasing conjugation. Additionally, it has high photostability and quantum efficiency. Hence, the modification of the BODIPY dyes for pH determination with orthogonal quencher
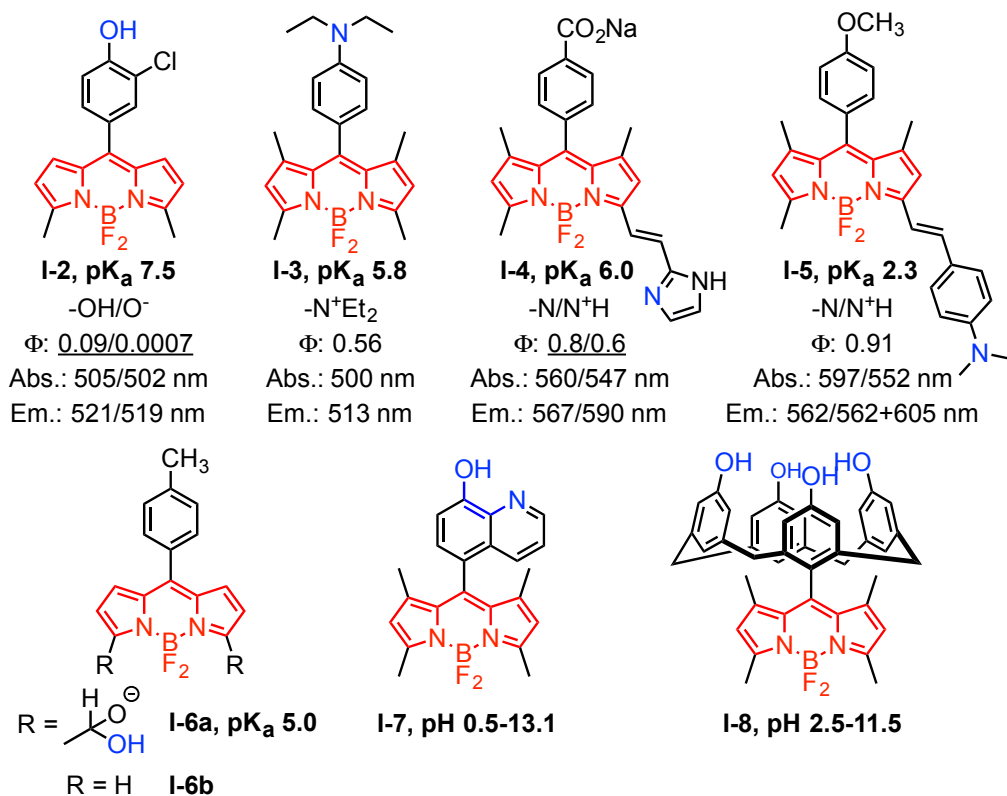
Figure I-6: Fluorescent pH-sensors based on BODIPY dyes. BODIPY core is highlighted red, pH-sensing moiety is highlighted blue.

has sparked interest. In such a case, the wavelengths of absorption and emission do not change significantly. However, fluorescence intensity and, therefore, quantum yield and brightness increase if quenching is interrupted. Hydroxyls (**I-2**)[57] and amines (**I-3**)[58] [59-60] were introduced as titratable groups on quenchers for BODIPY dyes (**Figure I-6**). In the case of compound **I-2**, a 130 times fluorescence increase was recorded during acidification.

A pH-responsive moiety can be introduced as a part of extended conjugation. In this instant, acidity promotes a response in wavelength, rather than emission intensity. In this way, an imidazole ring that exhibits dual pK_a character was introduced to BODIPY through a double bond, as shown in **I-4** (**Figure I-6**). The probe responded with a pK_a of 6.0 for a protonated imidazole and another pK_a, approximately 14.0, for the anionic imidazole.[61] Interestingly, protonation of the tertiary amine, conjugated to the BODIPY core through aromatic ring **I-5**, showed a response in the pH range 0.4–4.4.[62]

An extraordinary approach was presented by acidity-sensing aldehyde conjugated with BODIPY core **I-6a** (**Figure I-6**).[63] Typically, one considers hydroxyl, amine and carboxylate groups for pH-sensing, while an aldehyde is not a prime candidate during the design process. However, comparison of molecule **I-6a** with the respective analog **I-6b**, free from aldehyde, showed different chemical shifts in ^{19}F NMR and ^{11}B NMR. This suggested that there is an

interaction between the fluorine atom and aldehyde that activates the carbonyl for reaction with hydroxide. Indeed, **I-6** exhibited strong green fluorescence at acidic pH, while this fluorescence faded after reaching pH 8.0. There are two suggested mechanisms for the observed proton sensing: (1) protonation of the carbonyl, and as a result decreased PeT quenching between carbonyl and BODIPY, or (2) hemiacetal anion formation at high pH, decreasing the conjugation by one carbonyl, and thus reducing green fluorescence.
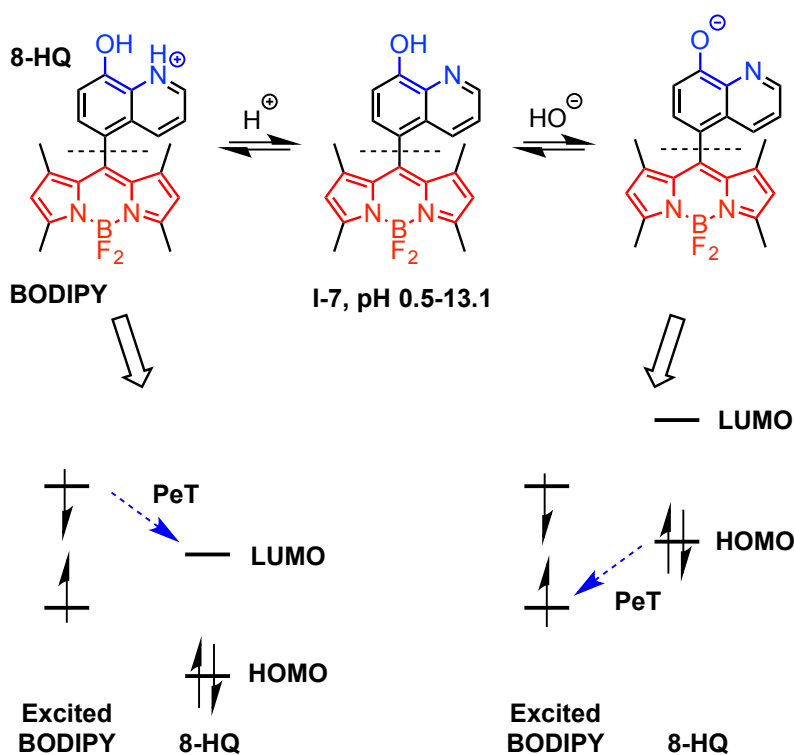
Figure I-7: Mechanism of PeT quenching in **I-7**.

The first example of an 'ON-OFF-ON' type of fluorescent pH sensor came from Jiang's lab and utilized a 8-hydroxyquinoline substitution pattern as a quencher **I-7** (**Figure I-6**), that has two ranges for titration: one between pH 0.5 and 6.8, due to protonation of the pyridyl, and one between pH 6.8 and 13.1, due

to deprotonation of –OH.[64] This unique design with the incorporation of both a weak acid and a weak base into a quencher allowed observation of a fluorescence response in the full pH range. From the crystal structure it was established that hydroxyquinoline is twisted out of conjugation from BODIPY core. The mechanism of quenching was validated by density functional theory calculations, that revealed two different pathways of PeT quenching (**Figure I-7**). With pyridyl protonation, electron transfer is occuring from the excited state BODIPY to the lowest unoccupied molecular orbital (LUMO) of hydroxyquinoline, leading to donor-excited PeT. Additionally, the alkoxide form has PeT from the highest occupied molecular orbital (HOMO) of hydroxyquinoline to the excited state BODIPY. Since hydroxyquinoline participates only in fluorescence quenching, BODIPY wavelength does not change the pH, in either direction.

A fascinating example of pH probe design, based on BODIPY dye, was developed using calix[4]arene as a PeT quencher.[65] This system actually originated from the tight binding between calix[*n*]arene derivatives and metal ions, even at a low concentration of analyte. Calix[4]arene is the most popular building block for host-guest chemistry among all calixarenes, because of its ability to adjust its conical conformation.[66] This conformation is altered during the complexation with a metal. Hence, molecule **I-8** was expected to be a promising metal sensor after further modifications on orthogonal calix[4]arene. However, the authors observed by ^{1}H NMR all of the phenolic protons coming at a single chemical shift suggesting fast proton exchange. Titration from pH 2.53 to 11.54 of

I-8 in water solution showed a constant decrease in fluorescence over the full range. Years later this probe was utilized for sensing the intracellular pH changes during early cell apoptosis.[67]

Since many of the described probes work through ON-OFF fluorescent signal, the perceived OFF state could be the result of the lack of fluorophore at
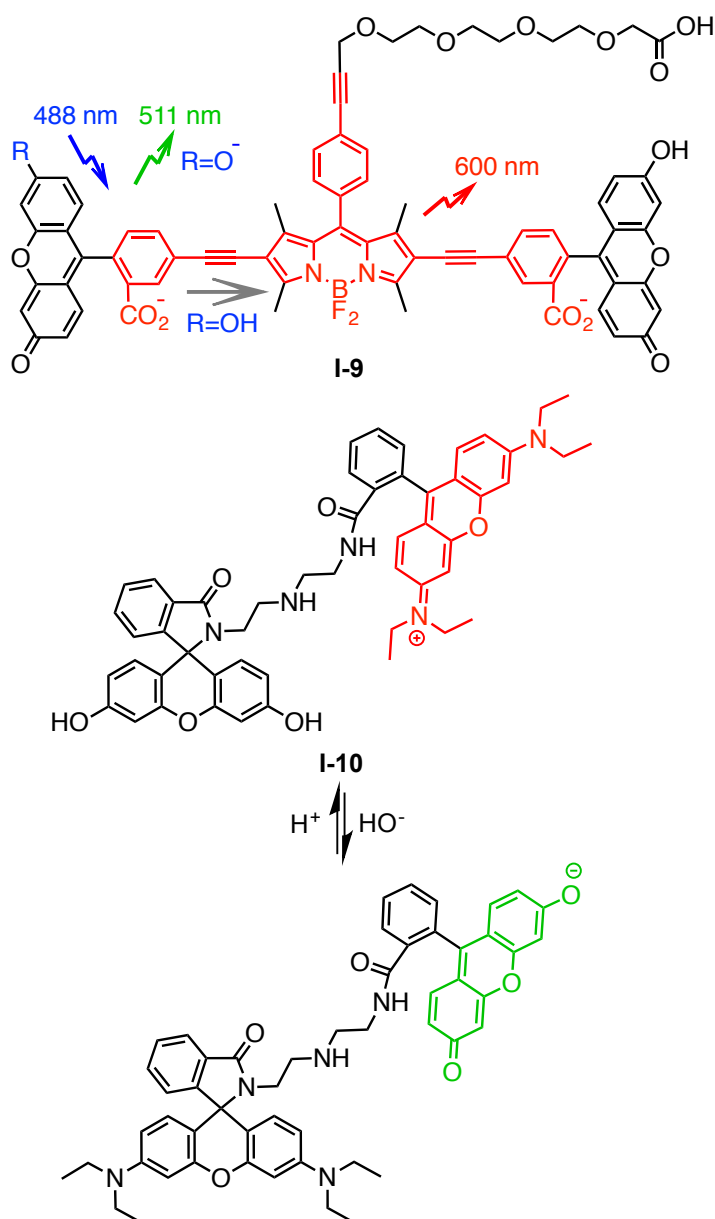
Figure I-8: Structure of energy transfer cassette **I-9** and two-color probe **I-10** that fluoresce red and/or green in response to pH.

17

the targeted location. To address this point the Burgess lab designed an efficient energy transfer cassette **I-9** that fluoresces regardless of its protonation state (**Figure I-8**).[68] The probe **I-9** consists of two fluorescent modules with short and long wavelength as the energy donor and acceptor, respectively. Two xanthane chromophores with short wavelength have pH dependent green fluorescence ($\Phi(O^-/OH)=0.97/0.4$), which is absorbed by BODIPY dye leading to red fluorescence. According to this strategy, after excitation at 488 nm, the probe constantly emits at 600 nm with pH-dependent intensity. However, whether it emits green depends on pH. The fluorescence ratio obtained from red and green channels allowed for the accurate pH determination of cytosol (7.4) and endosomes (5.4) in monkey kidney COS-7 cells. In a follow up study, the same lab used a non-ionizable fluorophore as a donor, to confirm that the recorded pH originates from the xanthane molecules and not elsewhere.[69]

LysoTracker Green or DND-26
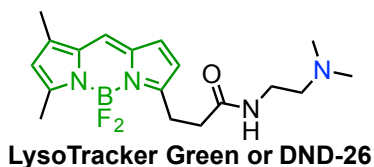
Figure I-9: Chemical structure of Lyso-Tracker Green with fluorophore in green and weak base in blue.

The two-color probe **I-10** (**Figure I-8**) that responds to acid and base with different colors was designed simply by linking two pH-sensitive dyes of different wavelengths.[70] Both dyes respond to pH changes by ring opening and closing mechanism. However, fluorescein lights up green at high pH, while rhodamine

lights up red at low pH. The resulting chameleon dye allows the pH to be accurately recorded over a wide range (4–11) using a ratiometric curve.

As mentioned above the BODIPY dyes themselves have no pH dependent properties. And interestingly, BODIPY dyes are often used as lysosome staining reagents. These dyes are called LysoTrackers (**Figure I-9**). As oppose to LysoSensors, they accumulate in acidic compartment providing a fluorescent signal based on localization instead of ON-OFF emission. Such compounds include quaternary amines with relatively hydrophobic alkane tails as part of the structure.[71-72] The lipophilic weak base penetrates biomembranes, but gets protonated only in acidic compartments. It becomes less hydrophobic and is retained because it is unable to dissociate out of the membrane. Therefore, the dye accumulates in the lysosomes, the most acidic compartment in the cell.

I.3 Genetically encoded pH sensors from the green fluorescent protein (GFP) family.

Specific localization in cellular compartments and potential toxicity with small molecule based pH indicators, is hard to control, plan, or predict, in particular for live-cell imaging applications. Alternatively, DNA-encoded protein-based probes do not have the latter disadvantages. They could yield superior agents for pH sensing applications. Genetic triggers for the production of these probes affords temporal control, while their ability to target specific locales provides spatial resolution.

The wild type green fluorescent protein (GFP) was isolated from jellyfish *Aequorea victoria.*[73] The parameters that control fluorescence, along with the mechanism of chromophore maturation, have been investigated for the past 40 years.[74-75] Since its discovery a number of dynamic processes were imaged with GFP.[76] Directed evolution methods has led to the engineering of probes of various colors and sensing abilities.[77] [78]
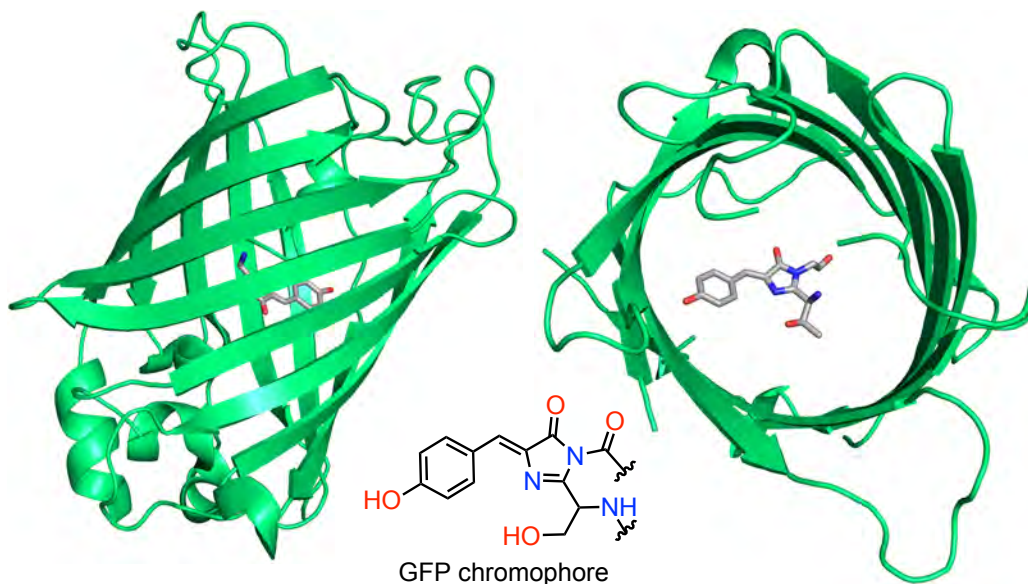
Figure I-10: Crystal structure of green fluorescent protein (GFP) and the chemical structure of its chromophore.

The proton sensing character of the GFP family comes from the pK_a of the *p*-hydroxybenzilydeneimidazolidinone chromophore covered inside the cylindrical beta barrel structure (**Figure I-10**).[77] The chromophore is formed through an autocatalytic cyclization of pre-arranged three consecutive amino acids followed by oxidation with molecular oxygen (**Figure I-11**). The autocatalysed chromophore formation is the most prominent advantage of GFP family because no exogenous reagent is required to generate fluorescent signal. Fluorescence of

GFP depends on the rigidity of the encapsulated chromophore and the anion formation, usually through excited state proton transfer (ESPT). The phenolate form has different spectroscopic properties, as compared to phenolic form. This is the foundation of the pH sensor development.[78]
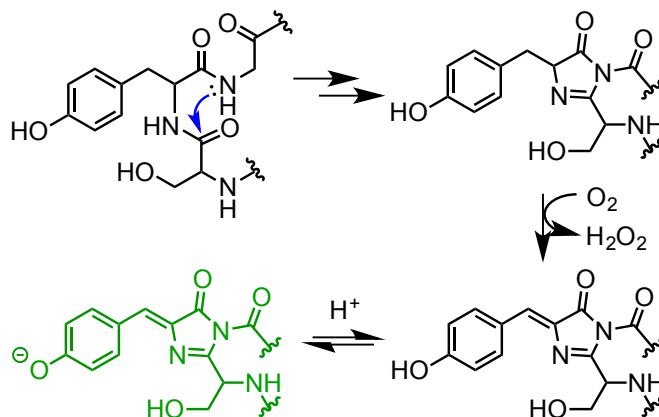
Figure I-11: Autocatalytic chromophore formation in GFP family and protonation/deprotonation event. Chemical structure of fluorescent form is highlighted green.

All pH-sensitive GFP variants can be subdivided into three groups.[79] The first group is based on emission intensity, which functions well as probes for monitoring pH fluctuations inside the cell. The second group has dual excitation and/or emission profiles that have been successfully implemented as ratiometric pH probes. The third group consists of protein fusion chimeras that gain the new acid/base sensitive character upon altering the protein construct. Nevertheless, all GFP-derived probes work based on a protonation/deprotonation event of the hydroxyl group. The spectroscopic properties of the probe vary based on the chromophore's structure or its interactions with the protein environment (**Figure I-12**).

The pioneering results on GFP-based pH sensors came the same year from three different groups. Interestingly, pH dependent fluorescence of wild-type GFP had already been observed 15 years prior to the first pH-sensing report.[79] Kneen and co-workers [80] were the first ones to describe four different mutants of GFP with a consistent increase in emission, as a function of pH. The wavelength of the mutants did not change but the intensity of absorbance at 490 nm and emission at 510 nm both gradually decreased with acidification. At the same time, a new peak with absorbance at 390 nm and a weak emission at 450 nm grew. The pK_a values of 4.8, 5.9, 6.0 and 6.1 were recorded for different mutants, spanning the majority of the physiologically relevant pH range. The shape of the titration curves indicated the involvement of only one proton in the titrated range. The mutant with pK_a 6.0 was successfully localized in the cytosol of LLC-PK1 cells and the mitochondria, Golgi and endoplasmic reticulum of CHO cells. In cytosol, mitochondria and Golgi pH values were measured as 7.4, >7.5 and 6.6, respectively. The pH changes in the endoplasmic reticulum were monitored, but not accurately measured. Because the intensity-based method does not account for protein concentration or photobleaching, it required an *in situ* standard calibration curve, constructed via cell perfusion with ionophores and buffers of known pH.

A month later the Tsien lab reported a blue-shifted GFP variant: enhanced cyan fluorescent protein (ECFP) with an apparent pK_a value of 6.4.[81] ECFP

localization in the Golgi of HeLa cells indicated a pH of 6.58. Although it was a different mutant of GFP, compared to the one reported by Kneen *et al.*, the method employed the same strategy providing a comparable outcome. Ideally, pH-sensing systems should be independent of probe concentration, to allow for accurate measurements that are not skewed by changes in the intensity of the observed signal. A general solution to this problem is the use of ratiometric probes that enable concentration independent pH determinations.

GFP chromophore (weak fluorescence)   GFP chromophore (strong fluorescence)   YFP chromophore   mKeima chromophore

$\lambda_{ex}/\lambda_{em}$=~400/>500 nm $\lambda_{ex}/\lambda_{em}$=~490/>515 nm $\lambda_{ex}/\lambda_{em}$=~515/~530 nm $\lambda_{ex}/\lambda_{em}$=585/610 nm
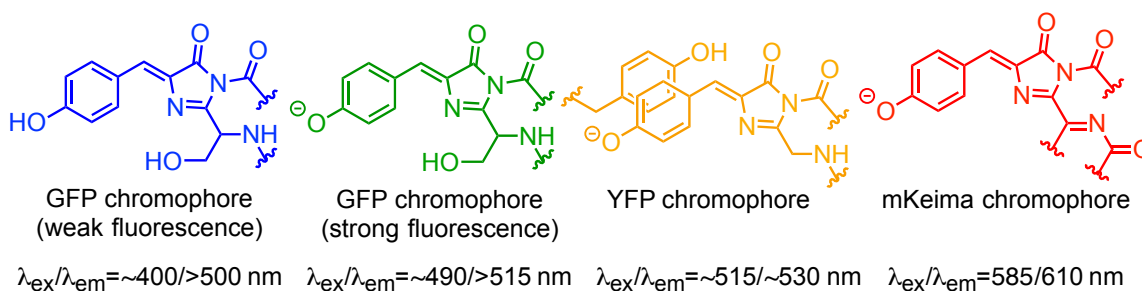
Figure I-12: Chemical structures of chromophores from GFP family utilized as pH sensors.

The breakthrough in the area came two months later, when Miesenböck *et al.*[2] reported the discovery of GFP mutants known as "pHluorin" series. The manuscript described two proteins, each with unique spectra: ratiometric pHluorin and ecliptic pHluorin. Ratiometric pHluorin works by using concentration independent emission ratios at 410 nm and 470 nm ($R_{410/470}$), because both protonated and unprotonated species are fluorescent. This approach requires the *in vitro* construction of a calibration curve. The ratio from an experimental pH determination is subsequently fitted to the curve. The working range for an accurate pH determination is 5.5–7.5, as determined from the standard curve. Determined pH values with compartment-targeted ratiometric pHluorin in HeLa

cells yielded a pH of 7.5 on the cell surface, a pH of 5.51±0.66 in endosomes and a pH of 6.21±0.39 in the Golgi.

As opposed to ratiometric pHluorin, ecliptic pHluorin is in the dark state at a pH below 6. This property made it suitable for visualizing secretion and synaptic transmission. Synaptic vesicles are acidic (pH ~5.0), while vesicle fusion with plasma membrane in neurons causes an instantaneous rise in the pH to approximately 7.4. Therefore, synaptic vesicles expressing ecliptic pHluorin are non-fluorescent in the resting state but are highly fluorescent after initiating a secretory response. The places with elevated fluorescence were slowly diffusing into the cytosol at different time frames of the experiment, as expected in exocytosis. Recorded images at different time lapses had both "on" and "off" spots, indicating vesicle fusion, as well as retrieval. More recently, a report showed the optimization of a ratiometric pHluorin into pHluorin2, with enhanced emission and expression, achieved through mammalianized and folding-induced mutations.[82]

Four ratiometric fluorescent indicators with dual emission using single excitation wavelength, were reported as deGFP.[83] The pK_a values ranging 6.8–8.0 would be suitable for pH measurements close to neutrality. Although the protein is visible in two channels if expressed in PS120 cells, the usefulness of deGFP as a proton sensor in the live-cell imaging has not been demonstrated as of yet.[83] Nevertheless, it is a unique example of a single excitation dual emission protein. It has become a platform for other FP design and has been studied

extensively computationally.[84-85] An accepted mechanism of GFP fluorescence is based on excited state proton transfer (ESPT) that gives green emission after deprotonation in the excited state (**Figure I-13**). As determined from crystal structures of deGFP mutant at pH 9.0 and pH 5.0, the chromophore in acidic conditions does not contain the hydrogen bond network that facilitates ESPT. Different efficiency of ESPT leads to fluorescence of both the protonated (460 nm) and deprotonated (515 nm) states, as a function of pH with single excitation
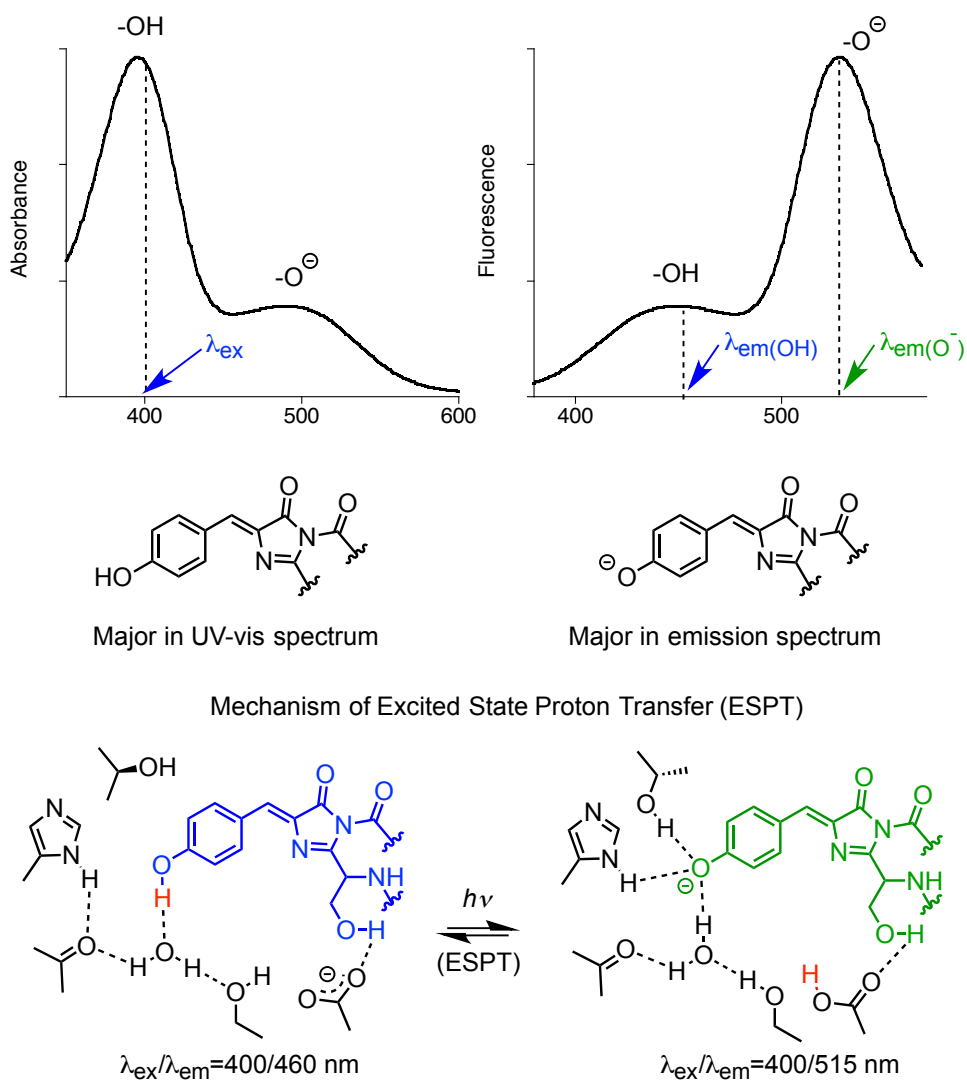
Figure I-13: Excited state proton transfer (ESPT) mechanism for GFP chromophore.

(400 nm).

An example of the dual excitation/dual emission probe from the GFP family is E^{2}GFP reported by Bizzarri *et al*.[86] This GFP mutant has a single absorption peak at pH <5 (430 nm). After basification two peaks (400 nm and 515 nm) were observed. Upon estimating different excitation wavelengths and observing various emission ranges, it was established that E^{2}GFP works best with two selective channels. One ratiometric curve can be obtained using $\lambda_{ex1} = 458$ nm and $\lambda_{ex2} = 488$ nm with the same band pass filter (BP) 500–600 nm that gives a pK_a of 6.9. Another ratiometric curve is generated using one excitation wavelength $\lambda_{ex} = 458$ nm and two band pass filters (475–525 nm and 515–600 nm), yielding a pK_a of 7.5. This system has dual control over pH determination. It was used for observing changes in the pH during cell mitosis in CHO cells. However, a substantial limitation of this GFP is that E^{2}GFP spectroscopic profile depends on the concentration of chloride ions. The same lab moved towards the development of E^{1}GFP,[87] with lower sensitivity to chloride ions. The key step of protein design involved mutagenesis of the residues in the chloride binding site into bulky space filling amino acids.

There are other non-ratiometric intensity-based pH-dependent probes from the GFP family called mCitrine,[88] mtAlpHi,[89] Clover[90] and mWasabi[91] with

respective pK_a values of 5.7, 8.5, 6.2 and 6.5. Officially, mCitrine and Clover were not utilized for measuring pH in the cell or intracellular compartments but their pK_a values of 5.7 and 6.2, respectively, show potential application in the field. Sensor mtAlpHi has a high pK_a relative to other FP. It is a suitable probe for mitochondrial pH determination.[89] The main advantages of mWasabi, over other indicators, are a narrow excitation peak and high quantum yield (0.8).[91]

A ratiometric indicator, with similar characteristics to the ratiometric pHluorin was isolated from the orange seapen *Ptilosarcus gurneyi* and named Pt-GFP.[92] Advantageously, this indicator has a broader pH range and higher fluorescence ratio. It was used for imaging pH response to anoxia and salt-stress in plants.

SypHer is a pH-sensitive protein with the highest pK_a value (8.7) from the entire GFP family.[93] Its ratiometric properties are based on dual excitation (400 nm and 490 nm) and single emission (535 nm). This permitted an accurate pH determination in mitochondria of HeLa cells, in response to calcium ions. The dynamic measurements of pH changes provided the evidence that calcium ions uptake and efflux in mitochondria is anticipated by a pH gradient.

Many red-shifted proteins from the GFP family (λ_{ex} >550 nm, λ_{em} >570 nm) have pK_a values in the appropriate range for proton sensing. For example,

mKate, mKate2, mRuby2 and mTangerine have pK_a values 6.2, 5.4, 5.3 and 5.7, respectively.[94-96] However, not many of them were investigated as pH sensors in live cells. Recently, pHTomato with pK_a 7.8 was used to monitor synaptic activity.[3] Combined with GFP-based calcium indicator pHTomato was used for dual color imaging of spatially resolved presynaptic and postsynaptic neurons. It was shown useful as a red-colored probe for exocytosis. Moreover, combined with optogenetic photostimulating tools, based on channelrhodopsin systems (blue light absorbing ChR2 and green light absorbing VChR1), pHluorine and pHTomato showed individual responses to the light of specific wavelengths. Co-expression of fusion proteins ChR2-pHluorin and VChR1-pHTomato in neurons provided a direct control and readout of neuronal activity triggered by blue and green light, respectively. This was the first example of multiplex control and tracking of distinct circuit pathways.

Thus far, the only ratiometric red fluorescent protein that has been identified, is pHRed derived from mKeima.[97] It is a pH probe with dual excitation (440 nm and 585 nm) and single emission (610 nm). Spectroscopic characterization of pHRed revealed a pK_a of 6.9 and pH-dependent fluorescence lifetime of the sensor. The latter property was used in fluorescence lifetime imaging microscopy (FLIM) for determination of intracellular pH in Neuron2A cells. It has also been applied in simultaneously recording pH and sensing adenosine triphosphate (ATP) in the dual color imaging with green fluorescent

Perceval (GFP-derived ATP sensor). In the same study, researchers showed that the 440 nm absorbing form of purified pHRed can be excited with two-photon excitation, at a single wavelength of 860 nm. This has the potential to become a new tool for imaging thick tissues.

The last subgroup of genetically encoded pH sensors from the GFP family consists of fusion proteins that aim for a better and more reliable ratiometric response. Usually, at least one of the proteins is a pH-sensitive variant with non-ratiometric properties, while the other one serves as a reference point. Genetic fusion of two proteins ensures an equal expression of both reporters.

The first example of fusion protein constructs employed pH-insensitive GFP mutant (GFP$_{UV}$: λ_{ex}=380 nm, λ_{em}=509 nm) as a reference and pH-sensitive GFP or yellow fluorescent protein (YFP) as a reporter.[98] The two probes, with green and yellow emission, were named GFpH and YFpH. They reflect the chosen sensor protein and pH-sensitive character (corresponding pK$_a$ 6.1 and 6.5). GFpH functions as a dual excitation single emission probe, while YFpH has single excitation but dual emission profile. YFpH undergoes pH-dependent FRET from GFP$_{UV}$ (emits at 509 nm) to YFP that is typically excited with 514 nm. The designed fusion proteins have shown utility in cytosol of various cell lines.

Another example of a fusion for FRET, utilized a well-established donor-acceptor pair: cyan fluorescent protein (CFP) and YFP.[99] Chimeras produced in the described work have a general name pHlameleon. Three mutants of YFP were fused with CFP and relative fluorescent intensities were plotted as a function of pH. The products had pK_a values 4.1, 5.8 and 8.3, determined from the FRET ratio. pHlameleons worked well for imaging cytosolic acidification upon lysosomal disruption in breast cancer MCF7 cell line.

A notable example of dual fusion utilized for pH determination in plant cells is dubbed pHusion.[100] It was constructed from enhanced green fluorescent protein (EGFP) and monomeric red fluorescent protein (mRFP). The fundamental concept of this design relies on the unique spectroscopic properties of each reporter. They are well separated and can be employed in two-track imaging. EGFP and mRFP have an excitation and emission wavelength over 100 nm apart that avoids spectral overlap. The fluorescent signals obtained from each track are subdivided at a titrated pH range giving a pK_a 6.0. The probe responded well in the whole plant, including roots. It was possible to follow the apoplastic and cytosolic pH homeostasis upon external pH changes.

The only ratiometric bioluminescent sensor is pHlash.[101] It is also a representative of the fused donor-acceptor pair. However, the donor generates light from luciferase. Fusion of *Renilla* luciferase to the pH independent FP, Venus, utilizes bioluminescence resonance energy transfer (BRET). This probe

does not require excitation by a separate light source. Instead the light comes from the enzyme-catalyzed oxidation of coelenterazine to coelenteramide (**Figure I-14**), which is pH dependent. In this way, excitation of pH-independent Venus depends on pH-dependent chemiluminescence coming from the fused luciferase enzyme. The pH-sensitive nature of the light source was only an observation. The explanation might be in the altered hydrogen-bonding network in the enzymatic reaction. Alternatively, the protonation/deprotonation of the substrate may alter the affinity to the enzyme. In principal, the 475 nm light from reaction excites Venus. The BRET ratio of emission signals at 525 nm to 475 nm is used to generate a concentration independent ratiometric curve. Unfortunately, the data obtained from isolated pHlash and fusion expressed in live yeast cells showed differences in titration ratios. More importantly, the luminescence signal (475 nm) decreased significantly after 20 min of reaction time. Nevertheless, the
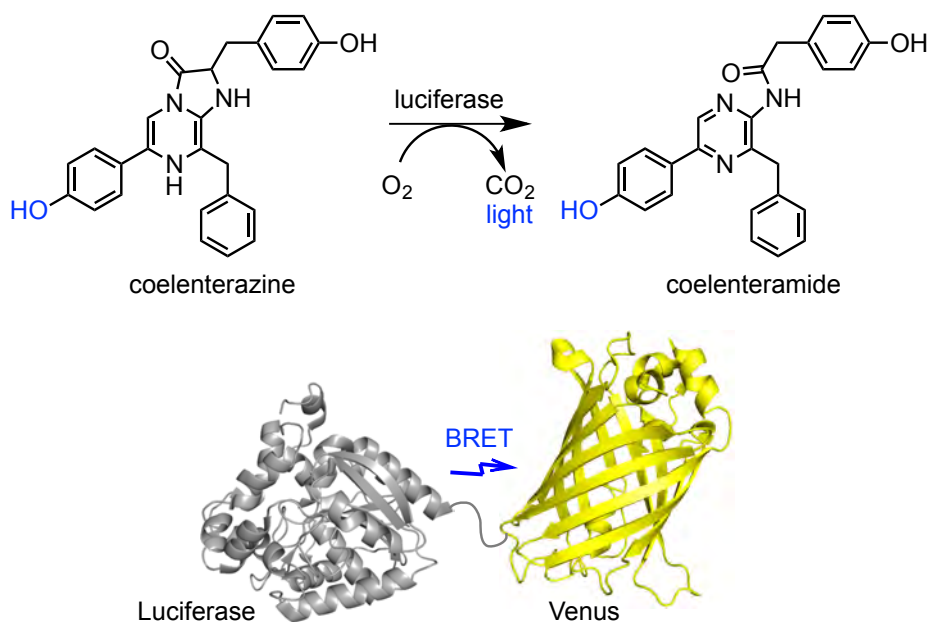
Figure I-14: Design of pHlash from Luciferase-Venus fusion. pH-dependent chemiluminescence of Luciferase is used to excite pH-independent Venus.

approach to create pHlash was different from anything else described in the literature thus far. After further optimization of this recently reported fusion protein, it might find utility as a working pH sensor.

Professor Benčina recently published a review article on pH determination in cellular organelles using GFP mutants.[79] Proteins were targeted to different cellular organelles in mammalian cells, plants, yeast and bacteria. Determined pH values closely matched previously reported ones obtained with microelectrodes, SNARF and BCECF. However, the void in the area remains the determination of pH in lysosomes. The problem is, that GFP probes have relatively high pK_a values, while the pH in the lysosomes is around 5 (in some cases reported <4). Dim fluorescence of the protonated form is another issue in acidic environment. The probes with reverse effect of fluorescence (bright after protonation) would be ideal for this purpose. However, the nature of the fluorophore as a proton acceptor needs to be modified for best outcome.

I.4 Other pH-sensors targetable to specific organelle.

The determination of pH with small molecules has been described many times. However, localization of the probe is always an issue if it is not targeted to specific organelles. Even commercially available tracker dyes give reliable localization only for the first few minutes. After some time they tend to translocate to other places in the living cell. Furthermore, the tracking systems are associated with staining of the whole compartment rather than trafficking specific proteins. Genetically encoded sensor fusions with the protein of interest solve the

localization problem. However, measuring the levels of acidity with the GFP family is not easy due to the dark state of the probe at low pH and a lot of background from autofluorescence. Therefore, specific localization of an acid-sensitive probe is one of the challenging tasks in the area. This section focuses on recent approaches addressing this problem.

KETWWETWWTEWSQPKKKRKV-NH⟍⟋SH

Hydrophobic region + Positive anchor =
Amphiphilic peptide Pep-1
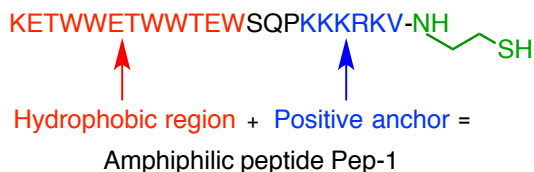
Figure I-15: Amphiphilic peptide Pep-1 that mediates protein delivery inside the cell *via* association with negatively charged membrane followed by penetration through hydrophobic lipid bilayer.

Many reports describe targeting of small molecule fluorescent sensors into the lysosomes. However, in most cases, it is not clear if the probe goes to the lysosome because of some structural features, or if it goes everywhere in the cell, but becomes visible only in lysosomes because of its acidic pH. Usually, the lysosomal localization is confirmed by co-staining with LysoTracker at a different color. If colocolization is achieved, it becomes a fluorescent probe for detection of lysosomal pH. For example, the dye **I-10** (**Figure I-8**) fluoresces red in lysosomes, while compartment with green fluorescence did not correspond to any tested tracker. The determined pH in lysosomes from the ratiometric curve for untreated HeLa cells was 5.7±0.2, while for *N*-acetylcystein or hydrogen peroxide treated cells were 4.9±0.1 and 6.6±0.4, respectively.

An out-of-the-box approach for specific pH-sensor delivery was described for molecule **I-9** (**Figure I-8**). Normally, exogenously added proteins are delivered

to endosomes. However, the Burgess lab discovered that amphiphilic peptide Pep-1 (**Figure I-15**) mediated protein delivery proceeds in a temperature dependent manner. Briefly, the energy transfer cassette **I-9** was conjugated with bovine serum albumin (BSA). The BSA-**I-9** conjugate was added with Pep-1 to COS-7 cells at 37 °C and at 4 °C. Fluorescence localized in endosomes under warm conditions, while it was diffused in the cytosol for cold incubated samples.
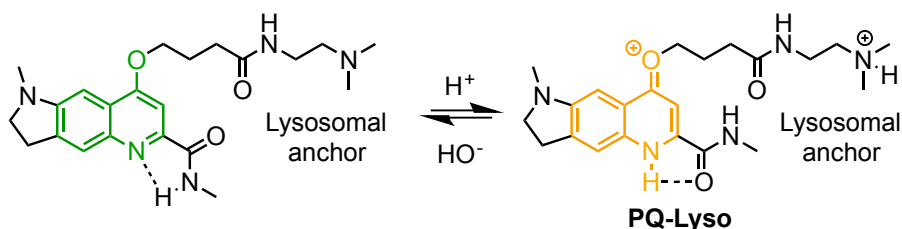
Figure I-16: Design of a ratiometric pH-sensor with lysosomal targeting.

More rationally designed lysosomal targeting utilizes a LysoTracker-like motif, as mentioned for the LysoTracker Green in **Figure I-9**. For instance, in a recent report quinoline dye was derivatized with a lysosomal pH sensor.[102] At first, the authors prepared an amido-quinoline probe with dual emission ratiometric character (pK_a=4.2). Then, the probe was decorated with a lysosomal anchor (dimethyethylamine moiety). The final probe, named PQ-Lyso (**Figure I-16**), facilitated pH measurements in lysosomes, in response to various exogenous factors.

Suppressed fluorescence of aggregated dyes, *via* self-quenching, can be reversed with pH. In this way, a dimer with an acid labile hydrazone linker was used to determine the areas of high acidity in cells and tissues.[103] The strategy is

shown in **Figure I-17a**. The covalent dimer of the cyanine dye is converted into

near infrared fluorescent monomer during acid-mediated hydrolysis of hydrazone.
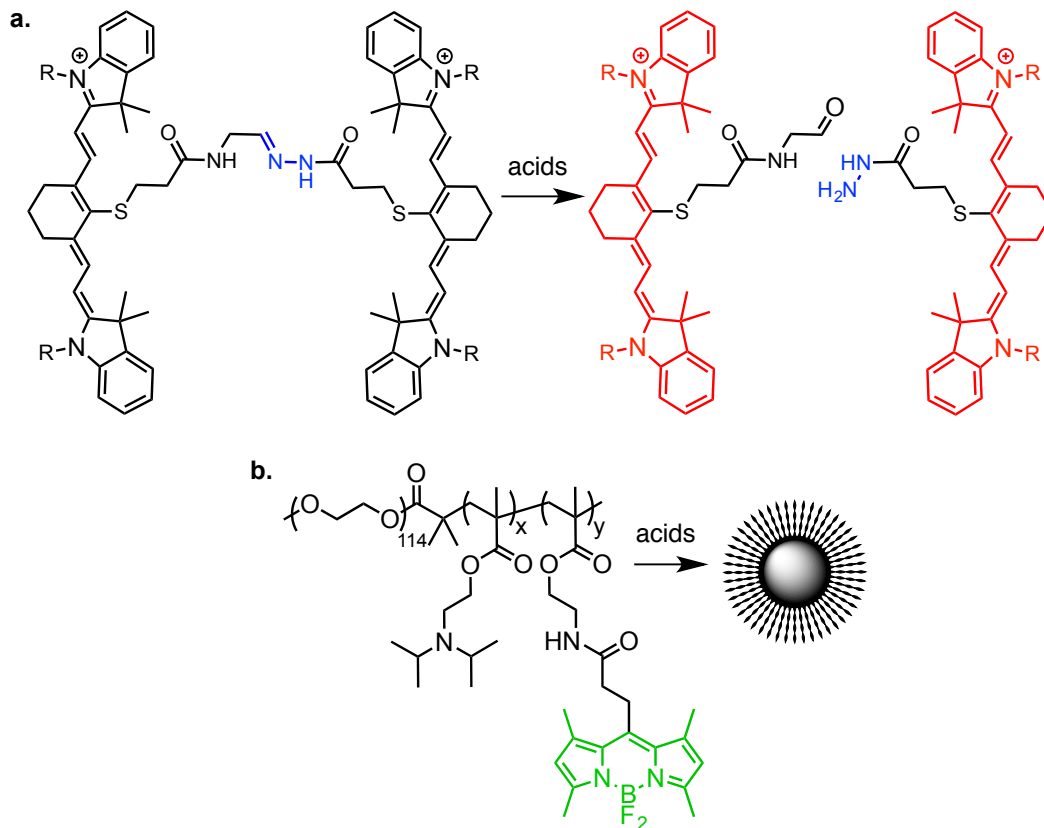
Figure I-17: Acid-promoted fluorescence without direct protonation of chromophore. **a.** Acid-catalyzed cleavage of quenched dimer. **b.** Acidity driven nanoparticle formation.

An example of acid-dependent fluorescence increase with sharp transition

was made from the combination of pH-independent fluorescent dye and a

nanoparticle platform (**Figure I-17b**).[104] Typically, the working range of sensors is

>2 pH units. The larger this range, the harder it is to determine pH accurately. But

the nanotunable method is more sensitive because it depends on the particle

disassembly, instead of the protonation event. As a result, the self-quenching

process is interrupted and emission is activated with a sharp transition (<0.5 pH

units). Although this approach does not utilize specific targeting, fluorescence activation is mostly seen at peri-nuclear regions.

It was described many times that small molecules can be targeted to a specific area, using a rational design of the agent or as an outcome of random discovery. However, as mentioned above, all of these methods are limited to the short-term measurements and not applicable for trafficking a specific protein of interest. The first example of a genetically encoded pH sensor, other than from the GFP family, utilized a receptor-mediated targeting method.[105] The strategy included intracellular expression of single chain antibody that recognizes 4-ethoxymethylene-2-phenyl-2-oxazoline-5-one (phOx) followed by cell staining with a phOx-fluorescein conjugate (**Figure I-18**). Genetic targeting of this antibody to the lumen of the Golgi gave an accurate pH determination (6.25±0.06) using fluorescein in ratio imaging microscopy. Interestingly, this method reported, in early 1999 by Verkman lab, pioneered a more recent high-
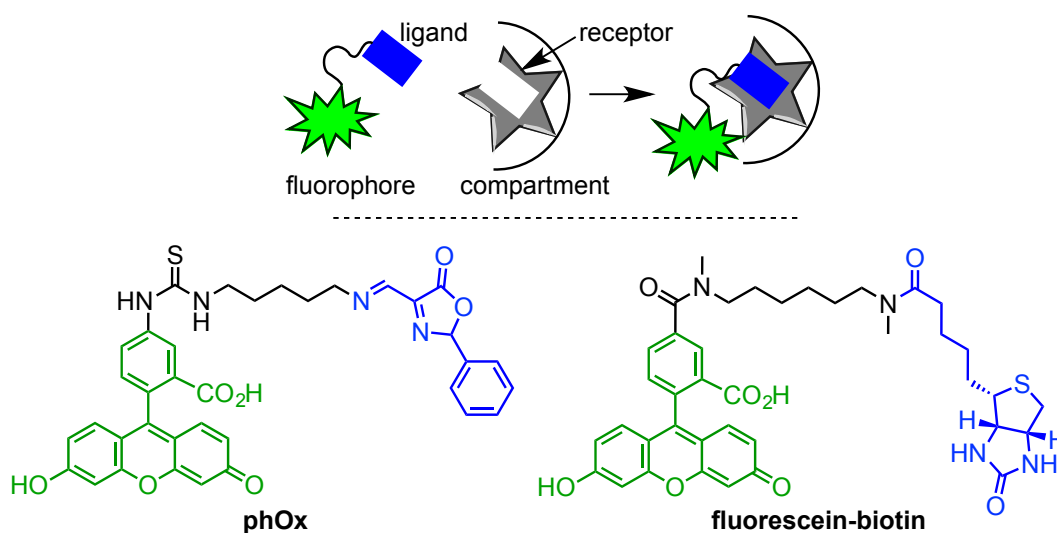
Figure I-18: Specific labeling of genetically encoded receptor protein based on high affinity binding to its ligand conjugated with pH-sensitive fluorescein.

36

impact work on a new series of genetically encoded protein tags. It combines a natural recognition unit attached to the fluorescent molecule (Halo-Tag, SNAP-Tag, CLIP-Tag, TMP-Tag etc), none of which were used as pH-sensors yet. However, there was an example of pH determination, utilizing picomolar affinity between avidin and biotin by localizing avidin in selected compartments and labeling it with the biotin-fluorescein conjugate (**Figure I-18**).[106]
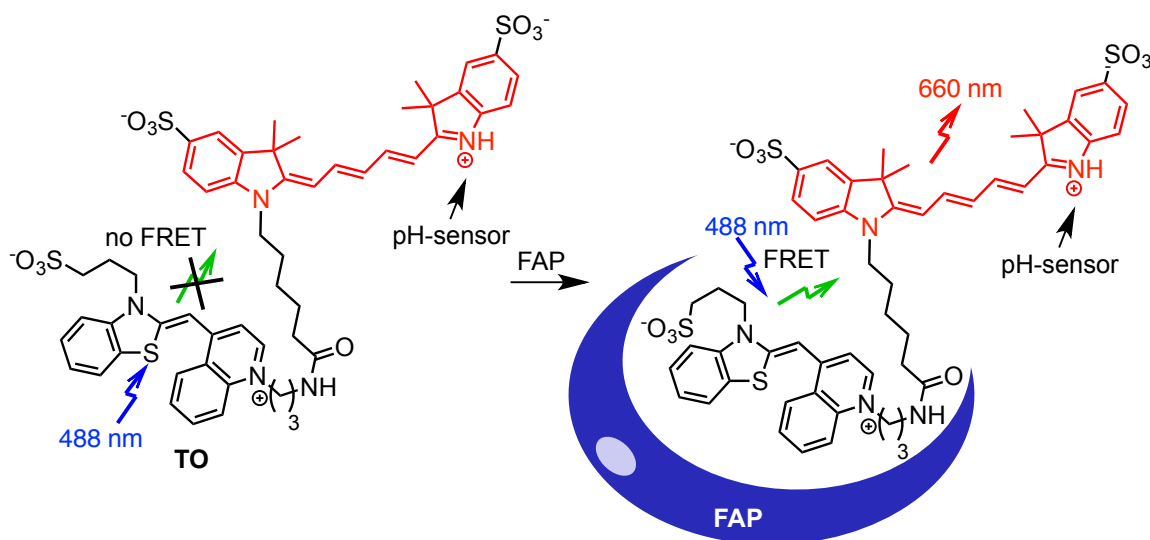
Figure I-19: Design of FRET-based pH-sensor using genetically encoded fluorogen-activating peptide (FAP). FRET takes place only after FAP binds TO while pHsensitive cyanine dye always responds to protonation.

A more recent report on tracking surface proteins through endocytosis utilized genetically encoded fluorogen-activating peptide combined with the pH-sensing cyanine dye.[107] The fluorogen-activating peptide (FAP) is a single chain antibody randomly selected from a large library by screening for enhanced fluorescence of thiazol orange (TO).[108] Fluorescence increased about 2,600-fold because of restricted rapid rotation around the single bond of TO after binding to FAP. The method of imaging the endocytosis used FRET between a TO-

37

derivative (activated by FAP) and a pH-sensitive cyanine dye (**Figure I-19**). FAP fused to adrenergic receptor became a useful tool for dynamic analysis of endosomal trafficking through the cellular membrane and exhibited drug-dependent pH fluctuations in vesicles.
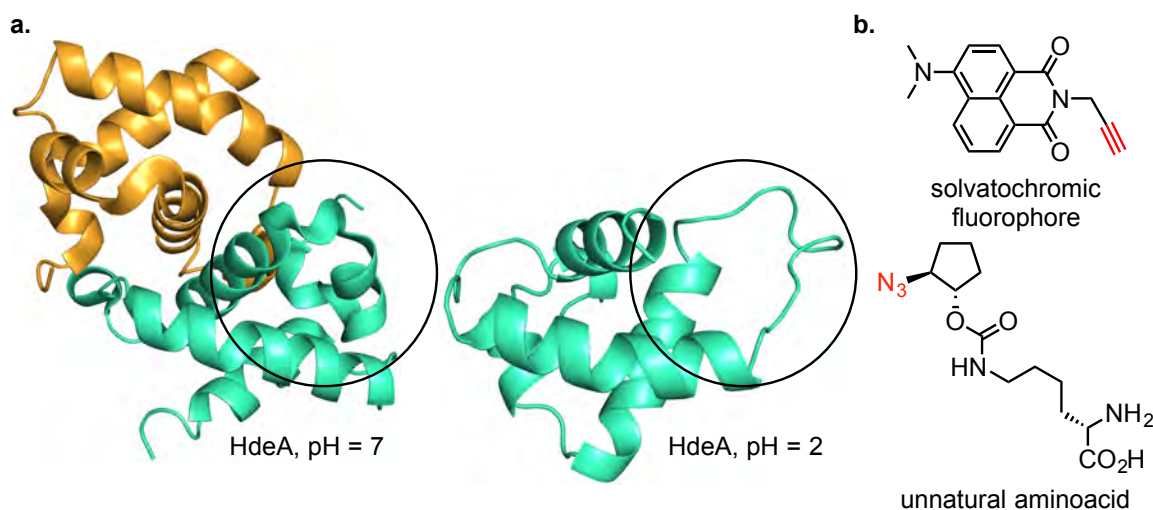
Figure I-20: Solvatochromic fluorophore as a pH-sensor. **a.** Conformational change of HdeA protein at different pH. **b.** Chemical structures of fluorophore and unnatural amino acid for 'click'.

Yang and co-workers presented a different approach to genetically encoded pH-sensing.[109] Their system responds to pH with a significant conformational change of the labeled protein instead of direct proton abstraction from a reporter molecule. The protein HdeA from highly acid resistant bacterial strains was characterized previously as an acid-protecting chaperone essential for bacteria to survive in the acidic environment of the stomach. At normal pH HdeA is a dimer with a hydrophobic region at the interface. However, at pH <3 it is a monomer capable of binding hydrophobic substrate. A solvatochromic fluorophore was used as an alternative substrate. The fluorophore normally fluoresces green in hydrophobic environment, but emits weak red light in

aqueous solution. To incorporate the fluorophore into the protein they used bioothogonal alkyne-azide ligation (**Figure I-20**). An azide-containing unnatural amino acid was introduced into HdeA, *via* amber mutation. The probe proved useful for sensing high acidity in *E. coli* and mammalian cells. However, its broad application is questionable due to high cost and low efficiency of protein expression with unnatural amino acids.

As surmised in this chapter, currently genetically encoded visualization of acidic processes is indeed a challenging task. Many research groups interested in bioorganic chemistry have attempted to develop tools to address this problem. Future solutions would combine the following: 1) the genetically encoded probe needs to be visible at low pH; 2) the protonation/deprotonation event should be reversible and concentration independent; 3) the library of probes needs to account for various pH ranges, including highly acidic ones; and 4) the method should be efficient and widely applicable.

I.5 Retinal and rhodopsin as an inspiration for pH-sensing.

Iminium-containing fluorophores would be interesting to examine in pH-sensing. An iminium has a broad range of pK_a, with representative compounds going as low as 5. Protonation of the imine dramatically changes the electron-withdrawing nature of the substituent and subsequently, is expected to trigger a shift in the spectroscopic properties of the fluorophore. Unfortunately, the iminium can easily hydrolyze to the amine and aldehyde, which makes its use in pH-

sensing relatively challenging. Nevertheless, there are a few reports on steady imine/iminium titrations in the presence of iminium-stabilizing functional groups.

The iminium of retinal, the aldehyde form of vitamin A (**Figure I-21**), can be highly chromophoric, compared to free aldehyde or its imine.[110] The color of retinal-PSB (and subsequently the wavelength of light it absorbs) strongly depends on its environment and is capable of covering the full visible spectra (absorbance range 420 nm–640 nm).[111] Retinylidene-PSB plays a crucial role in color vision by absorbing specific wavelength of light. There are four proteins in the human eye that bind retinal as an iminium or protonated Schiff base (PSB) and become capable of absorbing light. These proteins are known as rhodopsins. They belong to the G-protein coupled receptors (GPCR) that activate the signal transduction in response to a particular event.[112]

Rhodopsin absorbs the light and subsequently undergoes a conformational change that is critical to increase its affinity to α-transducin.[112-114] After α-transducin is bound to rhodopsin, it exchanges guanidine diphosphate (GDP) into guanidine triphosphate (GTP), which triggers dissociation from the other two transducing subunits. This sequence of events activates

all-*trans*-retinal   11-*cis*-retinal   13-*cis*-retinal

380 nm   ~365 nm   >420 nm
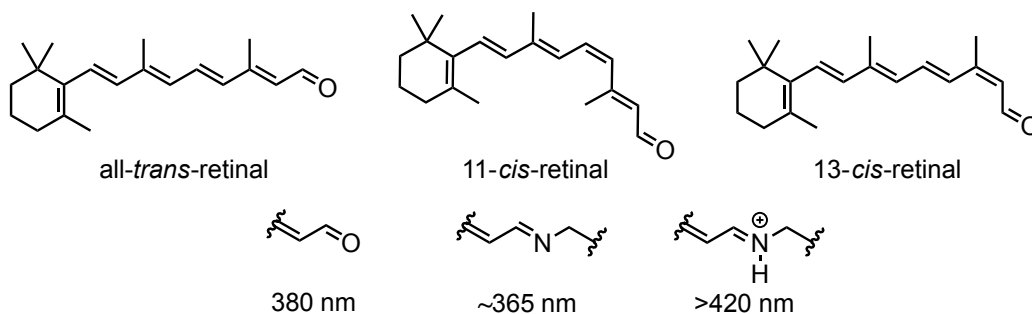
Figure I-21: Chemical structures of retinal isomers and the wavelengths for aldehyde, imine and iminium.

phosphodiesterase to hydrolyze cyclic guanosine monophosphate (cGMP) to guanosine monophosphate (GMP). A decreased concentration of cGMP causes hyperpolarization in photoreceptor cells and creates the signal.
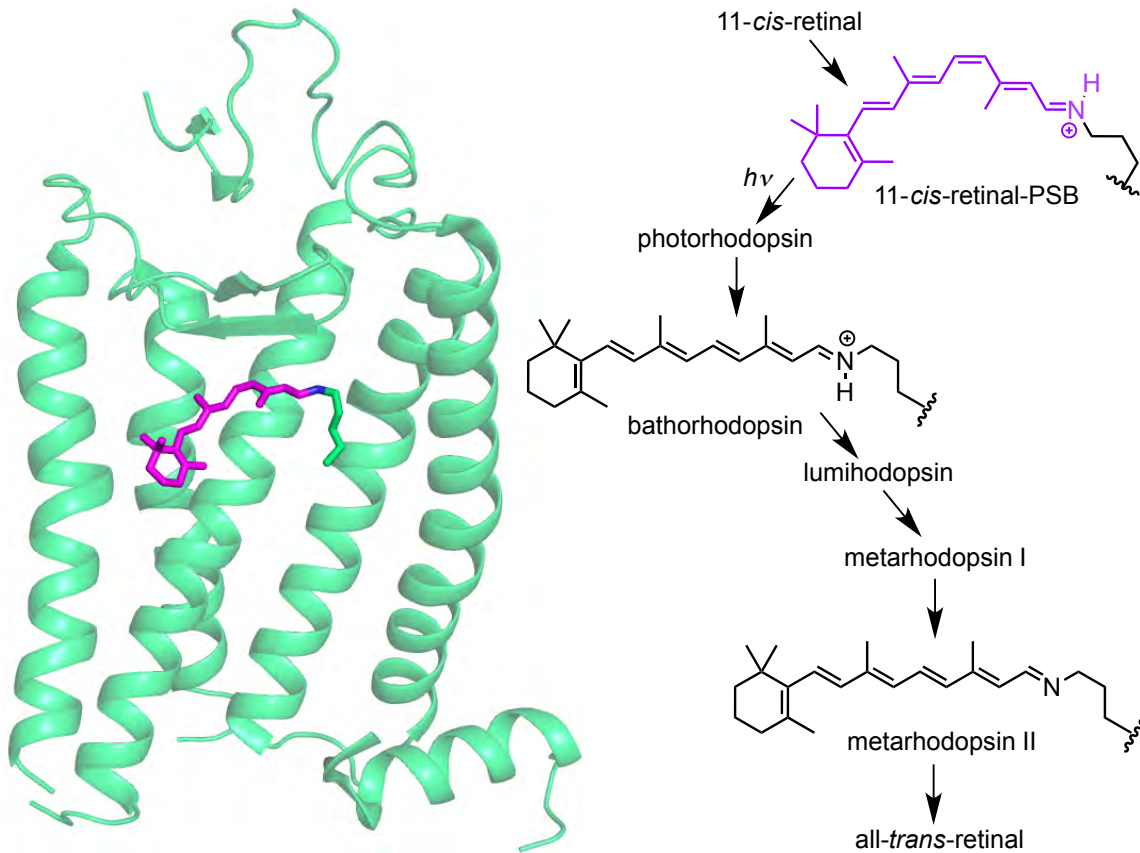
Figure I-22: Crystal structure of bovine rhodopsin and chemical structures of identified intermediates during photoisomerization of retinal. In crystal structure two helices were hidden to highlight bound retinal (magenta).

The crucial conformational change of rhodopsin is completed by photoisomerization of 11-*cis*- to all-*trans*-retinal-PSB (**Figure I-22**).[112] This reaction has multiple intermediates, some of which have even been crystalized. Interestingly, one of the intermediates, metarhodopsin II, is in the deprotonated form. After isomerization, the protein loses its affinity for the ligand and all-*trans*-retinal dissociates out. Ligand-free rhodopsin changes to a conformation capable

of binding 11-*cis*-retinal again and participate in a new cycle, while the 11-*cis*-retinal is regenerated through an enzymatic pathway.

Blue, green and red rhodopsins are named according to the wavelength of light each of them absorbs.[112] There is also rod rhodopsin, responsible for dim
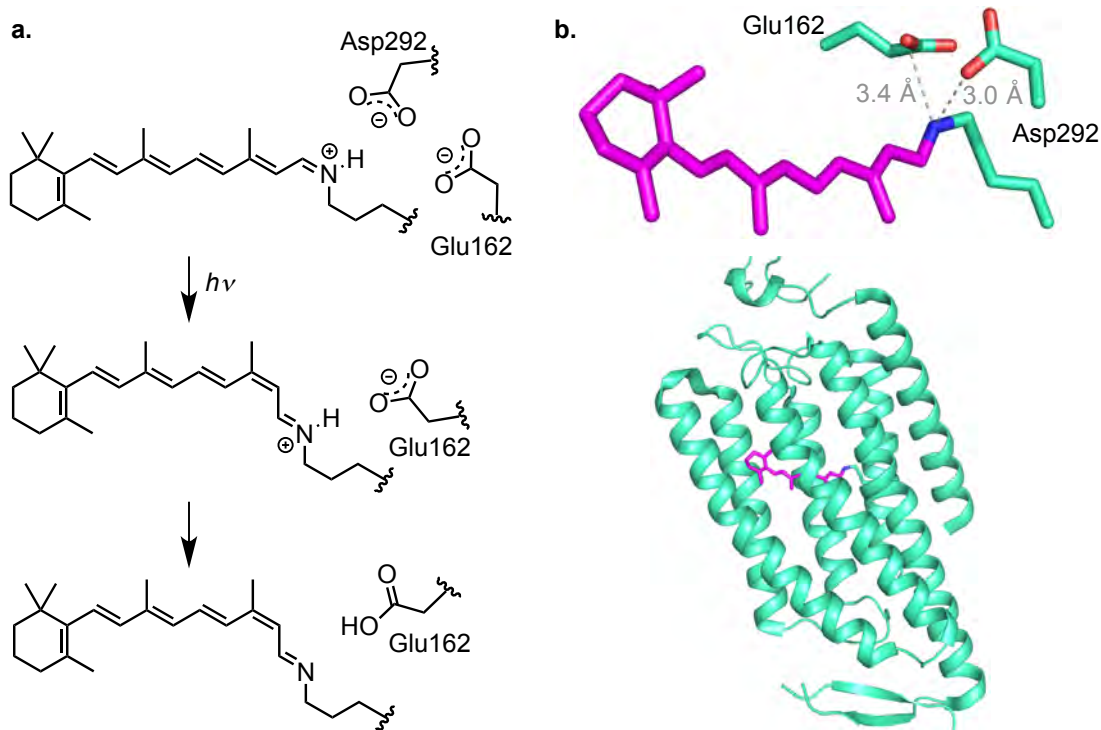
Figure I-23: Proton motion in channelrhodopsin. **a.** Isomerization of retinal. **b.** Crystal structure of channelrhodopsin. If using bacteriorhodopsin numbering Glu162 and Asp292 would be Asp85 and Asp212, respectively.

vision. Besides visual rhodopsins there are other pigments constructed from retinal and its isomers. Archael channelrhodopsin works as a light-gated proton pump that provides controlled motion of flagella to drive microalgae towards or away from light.[115-116] Briefly, it binds all-*trans*-retinal as PSB, which isomerizes to 13-*cis*-retinal-PSB after absorbing a photon of light (**Figure I-23**).[110] As opposed to the visual pigments retinal does not dissociate out of the channelrhodopsin but isomerizes back to the initial all-*trans* form. Isomerization triggers the iminium

region to move. As a result, it transfers the proton from PSB to Glu162 (channelrhodopsin numbering). This proton motion was recently proposed based on the crystal structure of channelrhodopsin.[117]

Not so long ago channelrhodopsins were engineered into optogenetic tools to control the activity of neurons.[115-116] Engineered channelrhodpsin absorbs at 480 nm that allows the initiation of the proton pump with blue light. It is a genetically encoded non-destructive method to create a difference in proton concentration on two sides of the cell membrane. Simply put, blue light irradiation of cells expressing channelrhodopsin causes depolarization and neuronal firing. In essence, the rapidly emerging area of optogenetic tools, represented by channelrhodopsins, composes the proton sensing machinery controlled by light. Therefore, the deprotonation ability of retinylidene is encouraging to investigate for pH-sensing.

Moreover, fluorescence of the retinal-PSB was recently investigated in proton dependent voltage sensing.[118] [119] For this purpose the Cohen lab used mutants of proteorhodopsin, a proton pump from bacteria in oceanic surface waters. At normal pH, proteorhodopsin works similar to channelrhodopsin: isomerization from all-*trans-* *to* 13-*cis*-retinal triggers the proton transport. An engineered mutant for proton sensing showed pH-dependent fluorescence for the protonated state and also had no light-induced proton pumping activity. Weak fluorescence of the retinal-iminium was seven times higher, compared to the imine form. The probe was named proteorhodopsin optical proton sensor

(PROPS).[118] An advantage of this probe is that no additional compound needs to be added exogeneously. Cells already contain retinal since it is one of the essential vitamins and the chromophore in PROPS is reconstituted from an inherently existing chromophore. "ON-OFF" fluorescence of PROPS was compared to pHluorin and showed opposite characteristics in the acidic and basic conditions.

The probe was developed as a voltage sensor accounting for the following: a change in the membrane potential causes the rearrangement of protons near the PSB of PROPS, and subsequently, alters the acid-base equilibrium. However, making a general pH-sensor out of PROPS would be difficult due to solubility issues because proteorhodopsin is a membrane bound protein. Additionally, the pK_a of PROPS is 9.6, which is far from the physiological range. Also, the fluorescence of retinylidene in PROPS is dim with brightness 500 times lower than GFP. On the other hand, an encouraging fact is that recently, this system was reengineered into a brighter version, with fluorescence only ~50 times lower than GFP.[120-121] It makes retinal binding proteins attractive targets for design of novel pH-sensors. However, the high pK_a of the PSB and the solubility of rhodopsins remain to be addressed.

During many decades of rhodopsin research the pK_a of retinylidene-PSB has been investigated thoroughly. The values for different rhodopsins vary from ~9 to ~16, while the pK_a of retinylidene-PSB with butylamine is 7.2 in

PBS/methanol (ratio 1:1).[37, 122-125 126] Metarhodopsin II is in Schiff base (SB) form at physiological pH (**Figure I-24**).[127] Experiments with model compounds and protein mutagenesis provided enough evidence that a counteranion is a key factor that stabilizes the positive charge on the PSB and increases its pK_a. The crystal structures of squid rhodopsin[128] and bacteriorhodopsin[129] with lower pK_a values revealed the absence of a regular carboxylate counterion in direct interaction with PSB (**Figure I-24**).
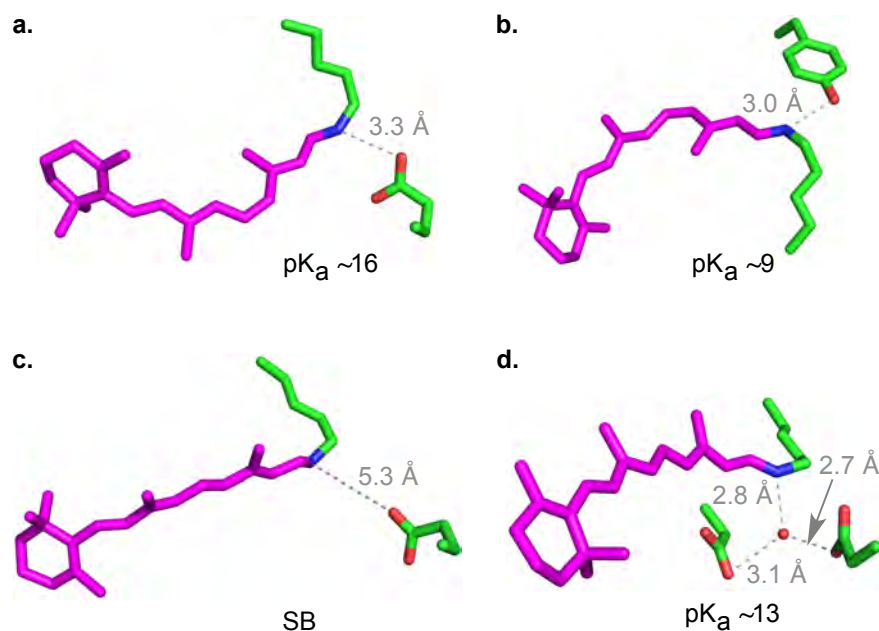
Figure I-24: Retinal-PSB and counterion interactions in crystal structures of **a.** bovine rhodopsin resting state, **b.** squid rhodopsin, **c.** metarhodopsin II and **d.** bacteriorhodopsin.

In order to estimate the pK_a of bovine rhodopsin, Steinberg and co-workers used fluorinated retinal analogs with an inherently lower pK_a.[125] Initial efforts of pK_a determination with retinal were unsuccessful because the protein

loses its structure at pH >11. The retinal analog **I-13** had a pK_a of PSB 5.5 units

lower than the 11-*cis*-retinal-PSB formed with *n*-butylamine (**Figure I-25**).

Comparison of **I-11** and **I-12** showed no impact on the pK_a regardless of the

double bond geometry. The PSB pK_a of **I-13** in bovine rhodopsin was determined

to be 10.2. The natural pigment has a pK_a of ~16 if the difference between two

chromophores is simply added to the obtained value.

There are many possible solutions to modulate the pK_a of rhodopsin.

However, the protein solubility of the membrane bound rhodopsins would be hard

to address. Moreover, investigating the factors that control pK_a would require

high resolution crystal structures, a challenging experiment for any GPCR.

Instead of optimizing rhodopsin, alternative protein systems can be investigated.

Having a handful of proteins engineered to form PSB with all-*trans*-retinal, we

aimed to produce a new series of mutants with wide range of pK_a values.

11-*cis*-retinal-PSB   **I-11**   **I-12**   **I-13**
pK_a 7.2 pK_a 5.0 pK_a 5.0 pK_a 1.8
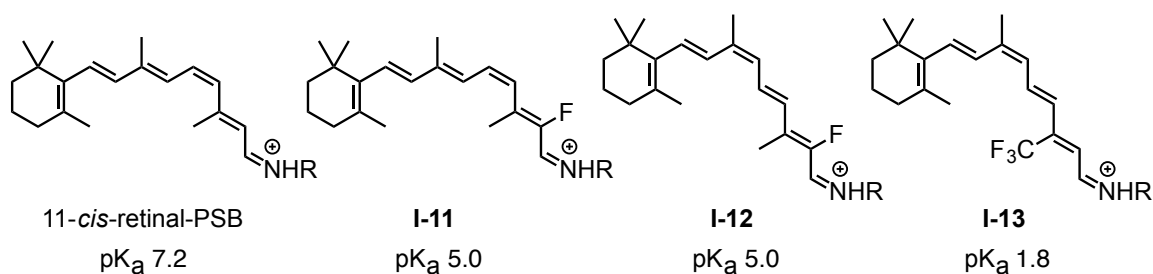
Figure I-25: Retinal analogs with lower pK_a of PSB compared to 11-*cis*-retinal. The pK_a values were measured with *n*-butylamine in 1:1 ethanol-water solution using HCl for pH-titrations. The pK_a of the PSB formed with bovine rhodopsin bound to **I-13** was 10.2.

I.6 Conclusions and significance

As detailed in this chapter, accurately determining the pH in living cells is an important and challenging task. Yet none of the existing systems provide a solution for routine use. pH-sensing molecules are problematic for targeting to specific location. Even if this is achieved, temporal relocation is another obstacle to overcome. Genetically encoded pH-sensors from the GFP family can be easily targeted to different organelles. However, imaging very acidic or an oxygen-free environment becomes a problem due to the nature of GFP. Other genetically encoded proton-sensing methods have their drawbacks, such as cell permeability of the labeling reagent, possibility for ratiometric measurements, availability of the expression platform or bright fluorescence of the indicator. There is much room for improvement, the following thesis describes our efforts in this regard.

REFERENCES

REFERENCES

1.   Scott, C. C.; Gruenberg, J., Ion flux and the function of endosomes and lysosomes: pH is just the start. *Bioessays* **2011,** *33* (2), 103.

2.   Miesenbock, G.; De Angelis, D. A.; Rothman, J. E., Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* **1998,** *394* (6689), 192.

3.   Li, Y.; Tsien, R. W., pHTomato, a red, genetically encoded indicator that enables multiplex interrogation of synaptic activity. *Nature Neuroscience* **2012,** *15* (7), 1047.

4.   Kemp, G.; Boning, D.; Beneke, R.; Maassen, N., Explaining pH change in exercising muscle: Lactic acid, proton consumption, and buffering vs. strong ion difference. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology* **2006,** *291* (1), R235.

5.   Dollery, S. J.; Wright, C. C.; Johnson, D. C.; Nicola, A. V., Low-pH-Dependent Changes in the Conformation and Oligomeric State of the Prefusion Form of Herpes Simplex Virus Glycoprotein B Are Separable from Fusion Activity. *Journal of Virology* **2011,** *85* (19), 9964.

6.   Mould, J. A.; Drury, J. E.; Frings, S. M.; Kaupp, U. B.; Pekosz, A.; Lamb, R. A.; Pinto, L. H., Permeation and activation of the M-2 ion channel of influenza A virus. *Journal of Biological Chemistry* **2000,** *275* (40), 31038.

7.   Parks, S. K.; Chiche, J.; Pouyssegur, J., pH Control Mechanisms of Tumor Survival and Growth. *J. Cell. Physiol.* **2011,** *226* (2), 299.

8.   Castellone, R. D.; Leffler, N. R.; Dong, L.; Yang, L. V., Inhibition of tumor cell migration and metastasis by the proton-sensing GPR4 receptor. *Cancer Lett.* **2011,** *312* (2), 197.

9.   Gatenby, R. A.; Gawlinski, E. T.; Gmitro, A. F.; Kaylor, B.; Gillies, R. J., Acid-mediated tumor invasion: a multidisciplinary study. *Cancer Res.* **2006,** *66* (10), 5216.

10.   Robey, I. F.; Baggett, B. K.; Kirkpatrick, N. D.; Roe, D. J.; Dosescu, J.; Sloane, B. F.; Hashim, A. I.; Morse, D. L.; Raghunand, N.; Gatenby, R. A.; Gillies, R. J., Bicarbonate Increases Tumor pH and Inhibits Spontaneous Metastases. *Cancer Res.* **2009,** *69* (6), 2260.

11.   Rofstad, E. K.; Mathiesen, B.; Kindem, K.; Galappathi, K., Acidic extracellular pH promotes experimental metastasis of human melanoma cells in athymic nude mice. *Cancer Res.* **2006,** *66* (13), 6699.

12.   Fukumura, D.; Xu, L.; Chen, Y.; Gohongi, T.; Seed, B.; Jain, R. K., Hypoxia and acidosis independently up-regulate vascular endothelial growth factor transcription in brain tumors in vivo. *Cancer Res.* **2001,** *61* (16), 6020.

13.   Casey, J. R.; Grinstein, S.; Orlowski, J., Sensors and regulators of intracellular pH. *Nature Reviews Molecular Cell Biology* **2010,** *11* (1), 50.

14.   Cardone, R. A.; Casavola, V.; Reshkin, S. J., The role of disturbed pH dynamics and the Na+/H+ exchanger in metastasis. *Nature Reviews Cancer* **2005,** *5* (10), 786.

15.   Estrella, V.; Chen, T.; Lloyd, M.; Wojtkowiak, J.; Cornnell, H. H.; Ibrahim-Hashim, A.; Bailey, K.; Balagurunathan, Y.; Rothberg, J. M.; Sloane, B. F.; Johnson, J.; Gatenby, R. A.; Gillies, R. J., Acidity Generated by the Tumor Microenvironment Drives Local Invasion. *Cancer Research* **2013,** *73* (5), 1524.

16.   Akram, M., Mini-review on Glycolysis and Cancer. *Journal of Cancer Education* **2013,** *28* (3), 454.

17.   Xu, R. H.; Pelicano, H.; Zhou, Y.; Carew, J. S.; Feng, L.; Bhalla, K. N.; Keating, M. J.; Huang, P., Inhibition of glycolysis in cancer cells: A novel strategy to overcome drug resistance associated with mitochondrial respiratory defect and hypoxia. *Cancer Research* **2005,** *65* (2), 613.

18.   Chiche, J.; Brahimi-Horn, M. C.; Pouyssegur, J., Tumour hypoxia induces a metabolic shift causing acidosis: a common feature in cancer. *Journal of Cellular and Molecular Medicine* **2010,** *14* (4), 771.

19.   Weerakkody, D.; Moshnikova, A.; Thakur, M. S.; Moshnikova, V.; Daniels, J.; Engelman, D. M.; Andreev, O. A.; Reshetnyak, Y. K., Family of pH (low) insertion peptides for tumor targeting. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110* (15), 5834.

20.   Tseng, S. J.; Zeng, Y.-F.; Deng, Y.-F.; Yang, P.-C.; Liu, J.-R.; Kempson, I. M., Switchable delivery of small interfering RNA using a negatively charged pH-responsive polyethylenimine-based polyelectrolyte complex. *Chemical Communications* **2013,** *49* (26), 2670.

21.   Jha, S.; Snell, J. M.; Sheftic, S. R.; Patil, S. M.; Daniels, S. B.; Kolling, F. W.; Alexandrescu, A. T., pH Dependence of Amylin Fibrillization. *Biochemistry* **2014,** *53* (2), 300.

22.   Srivastava, J.; Barber, D. L.; Jacobson, M. P., Intracellular pH sensors: Design principles and functional significance. *Physiology* **2007,** *22*, 30.

23.   Lukin, J. A.; Ho, C., The structure-function relationship of hemoglobin in solution at atomic resolution. *Chemical Reviews* **2004,** *104* (3), 1219.

24.   Bullough, P. A.; Hughson, F. M.; Skehel, J. J.; Wiley, D. C., STRUCTURE OF INFLUENZA HEMAGGLUTININ AT THE PH OF MEMBRANE-FUSION. *Nature* **1994,** *371* (6492), 37.

25.   Han, J.; Burgess, K., Fluorescent Indicators for Intracellular pH. *Chemical Reviews* **2010,** *110* (5), 2709.

26.   Rink, T. J.; Tsien, R. Y.; Pozzan, T., CYTOPLASMIC PH AND FREE MG-2+ IN LYMPHOCYTES. *Journal of Cell Biology* **1982,** *95* (1), 189.

27.   Tsien, R. Y., FLUORESCENT-PROBES OF CELL SIGNALING. *Annual Review of Neuroscience* **1989,** *12*, 227.

28.   O'Connor, N.; Silver, R. B., Ratio imaging: Practical considerations for measuning intracellular Ca(2+) and pH in living cells. In *Digital Microscopy, 3rd Edition*, Sluder, G.; Wolf, D. E., Eds. 2007; Vol. 81, pp 415.

29.   Balut, C.; VandeVen, M.; Despa, S.; Lambrichts, I.; Ameloot, M.; Steels, P.; Smets, I., Measurement of cytosolic and mitochondrial pH in living cells during reversible metabolic inhibition. *Kidney International* **2008,** *73* (2), 226.

30.   Martinez-Zaguilan, R.; Gurule, M. W.; Lynch, R. M., Simultaneous measurement of intracellular pH and Ca-2+ in insulin-secreting cells by spectral imaging microscopy. *American Journal of Physiology* **1996,** *270* (5 PART 1), C1438.

31.   Martinez-Zaguilan, R.; Lynch, R. M.; Martinez, G. M.; Gillies, R. J., Vacuolar-type H+-ATPAses are functionally expressed in plasma membranes of human tumor cells. *American Journal of Physiology* **1993,** *265* (4 PART 1), C1015.

32.   Masuda, A.; Oyamada, M.; Nagaoka, T.; Tateishi, N.; Takamatsu, T., Regulation of cytosol nucleus pH gradients by K+/H+ exchange mechanism in the nuclear envelope of neonatal rat astrocytes. *Brain Research* **1998,** *807* (1-2), 70.

33.   Qian, T.; Nieminen, A. L.; Herman, B.; Lemasters, J. J., Mitochondrial permeability transition in pH-dependent reperfusion injury to rat

hepatocytes. *American Journal of Physiology-Cell Physiology* **1997,** *273* (6), C1783.

34.   vandeVen, M.; Balut, C.; Baron, S.; Smets, I.; Steels, P.; Ameloot, M., Optimization of Mitochondrial and Cytosolic pH Determination in Madin Darby Kidney Cells using CLSM Images. *Biophysical Journal* **2013,** *104* (2), 338A.

35.   Wieder, E. D.; Hang, H. Y.; Fox, M. H., MEASUREMENT OF INTRACELLULAR PH USING FLOW-CYTOMETRY WITH CARBOXY-SNARF-1. *Cytometry* **1993,** *14* (8), 916.

36.   Briggs, M. S.; Burns, D. D.; Cooper, M. E.; Gregory, S. J., A pH sensitive fluorescent cyanine dye for biological applications. *Chemical Communications* **2000,** (23), 2323.

37.   Sheves, M.; Albeck, A.; Friedman, N.; Ottolenghi, M., CONTROLLING THE PKA OF THE BACTERIORHODOPSIN SCHIFF-BASE BY USE OF ARTIFICIAL RETINAL ANALOGS. *Proceedings of the National Academy of Sciences of the United States of America* **1986,** *83* (10), 3262.

38.   Hoischen, D.; Steinmuller, S.; Gartner, W.; Buss, V.; Martin, H. D., Merocyanines as extremely bathochromically absorbing chromophores in the halobacterial membrane protein bacteriorhodopsin. *Angewandte Chemie-International Edition in English* **1997,** *36* (15), 1630.

39.   Adie, E. J.; Kalinka, S.; Smith, L.; Francis, M. J.; Marenghi, A.; Cooper, M. E.; Briggs, M.; Michael, N. P.; Milligan, G.; Game, S., A pH-sensitive fluor, CypHer (TM) 5, used to monitor agonist-induced g protein-coupled receptor internalization in live cells. *Biotechniques* **2002,** *33* (5), 1152.

40.   Ernst, L. A., Mujumdar, R. B., Waggoner, A. S., Method for labeling and detecting materials employing arylsulfonate cyanine dyes. *US5268486* **1993**.

41.   Myochin, T.; Kiyose, K.; Hanaoka, K.; Kojima, H.; Terai, T.; Nagano, T., Rational Design of Ratiometric Near-Infrared Fluorescent pH Probes with Various pK(a) Values, Based on Aminocyanine. *Journal of the American Chemical Society* **2011,** *133* (10), 3401.

42.   Tang, B.; Yu, F.; Li, P.; Tong, L.; Duan, X.; Xie, T.; Wang, X., A Near-Infrared Neutral pH Fluorescent Probe for Monitoring Minor pH Changes: Imaging in Living HepG2 and HL-7702 Cells. *Journal of the American Chemical Society* **2009,** *131* (8), 3016.

43.   Vendrell, M.; Zhai, D.; Er, J. C.; Chang, Y.-T., Combinatorial Strategies in Fluorescent Probe Development. *Chemical Reviews* **2012,** *112* (8), 4391.

44.   Chen, S.; Hong, Y.; Liu, Y.; Liu, J.; Leung, C. W. T.; Li, M.; Kwok, R. T. K.; Zhao, E.; Lam, J. W. Y.; Yu, Y.; Tang, B. Z., Full-Range Intracellular pH Sensing by an Aggregation-Induced Emission-Active Two-Channel Ratiometric Fluorogen. *Journal of the American Chemical Society* **2013,** *135* (13), 4926.

45.   Tan, Y.; Yu, J.; Gao, J.; Cui, Y.; Wang, Z.; Yang, Y.; Qian, G., A fluorescent pH chemosensor for strongly acidic conditions based on the intramolecular charge transfer (ICT) effect. *Rsc Advances* **2013,** *3* (15), 4872.

46.   Huang, X.; Song, L.; Xu, J.; Zhu, G.; Liu, B., Asymmetric Total Synthesis of Leucosceptroid B. *Angewandte Chemie-International Edition* **2013,** *52* (3), 952.

47.   Hornung, V.; Bauernfeind, F.; Halle, A.; Samstad, E. O.; Kono, H.; Rock, K. L.; Fitzgerald, K. A.; Latz, E., Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. *Nature Immunology* **2008,** *9* (8), 847.

48.   Charier, S.; Ruel, O.; Baudin, J. B.; Alcor, D.; Allemand, J. F.; Meglio, A.; Jullien, L., An efficient fluorescent probe for ratiometric pH measurements in aqueous solutions. *Angewandte Chemie-International Edition* **2004,** *43* (36), 4785.

49.   DePedro, H. M.; Urayama, P., Using LysoSensor Yellow/Blue DND-160 to sense acidic pH under high hydrostatic pressures. *Analytical Biochemistry* **2009,** *384* (2), 359.

50.   Diwu, Z. J.; Chen, C. S.; Zhang, C. L.; Klaubert, D. H.; Haugland, R. P., A novel acidotropic pH indicator and its potential application in labeling acidic organelles of live cells. *Chemistry & Biology* **1999,** *6* (7), 411.

51.   Holopainen, J. M.; Saarikoski, J.; Kinnunen, P. K. J.; Jarvela, I., Elevated lysosomal pH in neuronal ceroid lipofuscinoses (NCLs). *European Journal of Biochemistry* **2001,** *268* (22), 5851.

52.   Hurwitz, S. J.; Terashima, M.; Mizunuma, N.; Slapak, C. A., Vesicular anthracycline accumulation in Doxorubicin-selected U-937 cells: Participation of lysosomes. *Blood* **1997,** *89* (10), 3745.

53.   Lin, H. J.; Herman, P.; Kang, J. S.; Lakowicz, J. R., Fluorescence lifetime characterization of novel low-pH probes. *Analytical Biochemistry* **2001,** *294* (2), 118.

54.   Mishra, A.; Krishnamoorthy, G., Photophysical study of 2-(4 '-N,N-dimethylaminophenyl)oxazolo 4,5-b pyridine in different solvents and at various pH. *Photochemical & Photobiological Sciences* **2012,** *11* (8), 1356.

55.   Sillence, D. J., Glucosylceramide modulates endolysosomal pH in Gaucher disease. *Molecular Genetics and Metabolism* **2013,** *109* (2), 194.

56.   Trifonov, R. E.; Rtishchev, N. I.; Ostrovskii, V. A., Influence of medium acidity upon the luminescence properties of 2,5-diphenyl-1,3,4-oxadiazole and 2,5-diphenyl-1,3-oxazole. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **1996,** *52* (14), 1875.

57.   Baruah, M.; Qin, W. W.; Basaric, N.; De Borggraeve, W. M.; Boens, N., BODIPY-based hydroxyaryl derivatives as fluorescent pH probes. *Journal of Organic Chemistry* **2005,** *70* (10), 4152.

58.   Urano, Y.; Asanuma, D.; Hama, Y.; Koyama, Y.; Barrett, T.; Kamiya, M.; Nagano, T.; Watanabe, T.; Hasegawa, A.; Choyke, P. L.; Kobayashi, H., Selective molecular imaging of viable cancer cells with pH-activatable fluorescence probes. *Nature Medicine* **2009,** *15* (1), 104.

59.   He, H.; Ng, D. K. P., A ratiometric near-infrared pH-responsive fluorescent dye based on distyryl BODIPY. *Organic & Biomolecular Chemistry* **2011,** *9* (8), 2610.

60.   Hoogendoorn, S.; Blom, A. E. M.; Willems, L. I.; van der Marel, G. A.; Overkleeft, H. S., Synthesis of pH-Activatable Red Fluorescent BODIPY Dyes with Distinct Functionalities. *Organic Letters* **2011,** *13* (20), 5656.

61.   Boens, N.; Qin, W.; Baruah, M.; De Borggraeve, W. M.; Filarowski, A.; Smisdom, N.; Ameloot, M.; Crovetto, L.; Talavera, E. M.; Alvarez-Pez, J. M., Rational Design, Synthesis, and Spectroscopic and Photophysical Properties of a Visible-Light-Excitable, Ratiometric, Fluorescent Near-Neutral pH Indicator Based on BODIPY. *Chemistry-a European Journal* **2011,** *17* (39), 10924.

62.   Baruah, M.; Qin, W. W.; Flors, C.; Hofkens, J.; Vallee, R. A. L.; Beljonne, D.; Van der Auweraer, M.; De Borggraeve, W. M.; Boens, N., Solvent and pH dependent fluorescent properties of a dimethylaminostyryl borondipyrromethene dye in solution. *Journal of Physical Chemistry A* **2006,** *110* (18), 5998.

63.   Madhu, S.; Rao, M. R.; Shaikh, M. S.; Ravikanth, M., 3,5-Diformylboron Dipyrromethenes as Fluorescent pH Sensors. *Inorganic Chemistry* **2011,** *50* (10), 4392.

64.   Chen, Y.; Wang, H.; Wan, L.; Bian, Y.; Jiang, J., 8-Hydroxyquinoline-Substituted Boron-Dipyrromethene Compounds: Synthesis, Structure, and OFF-ON-OFF Type of pH-Sensing Properties. *Journal of Organic Chemistry* **2011,** *76* (10), 3774.

65.   Baki, C. N.; Akkaya, E. U., Boradiazaindacene-appended calix 4 arene: Fluorescence sensing of pH near neutrality. *Journal of Organic Chemistry* **2001,** *66* (4), 1512.

66.   Joseph, R.; Rao, C. P., Ion and Molecular Recognition by Lower Rim 1,3-Di-conjugates of Calix 4 arene as Receptors. *Chemical Reviews* **2011,** *111* (8), 4658.

67.   Han, F.; Xu, Y.; Jiang, D.; Qin, Y.; Chen, H., A BODIPY-derived fluorescent probe for cellular pH measurements. *Analytical Biochemistry* **2013,** *435* (2), 106.

68.   Han, J.; Loudet, A.; Barhoumi, R.; Burghardt, R. C.; Burgess, K., A Ratiometric pH Reporter For Imaging Protein-dye Conjugates In Living Cells. *J. Am. Chem. Soc.* **2009,** *131* (5), 1642.

69.   Thivierge, C.; Han, J.; Jenkins, R. M.; Burgess, K., Fluorescent Proton Sensors Based on Energy Transfer. *Journal of Organic Chemistry* **2011,** *76* (13), 5219.

70.   Lee, M. H.; Han, J. H.; Lee, J. H.; Park, N.; Kumar, R.; Kang, C.; Kim, J. S., Two-Color Probe to Monitor a Wide Range of pH Values in Cells. *Angewandte Chemie-International Edition* **2013,** *52* (24), 6206.

71.   Deduve, C.; Debarsy, T.; Poole, B.; Trouet, A.; Tulkens, P.; Vanhoof, F., LYSOSOMOTROPIC AGENTS. *Biochemical Pharmacology* **1974,** *23* (18), 2495.

72.   Trapp, S.; Rosania, G. R.; Horobin, R. W.; Kornhuber, J., Quantitative modeling of selective lysosomal targeting for drug design. *European Biophysics Journal with Biophysics Letters* **2008,** *37* (8), 1317.

73.   Shimomura, O.; Johnson, F. H.; Saiga, Y., EXTRACTION, PURIFICATION AND PROPERTIES OF AEQUORIN, A BIOLUMINESCENT PROTEIN FROM LUMINOUS HYDROMEDUSAN, AEQUOREA. *Journal of Cellular and Comparative Physiology* **1962,** *59* (3), 223.

74.   Morise, H.; Shimomur.O; Johnson, F. H.; Winant, J., INTERMOLECULAR ENERGY-TRANSFER IN BIOLUMINESCENT SYSTEM OF AEQUOREA. *Biochemistry* **1974,** *13* (12), 2656.

75.   Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J., PRIMARY STRUCTURE OF THE AEQUOREA-VICTORIA GREEN-FLUORESCENT PROTEIN. *Gene* **1992,** *111* (2), 229.

76.   Ibraheem, A.; Campbell, R. E., Designs and applications of fluorescent protein-based biosensors. *Current Opinion in Chemical Biology* **2010,** *14* (1), 30.

77.   Day, R. N.; Davidson, M. W., The fluorescent protein palette: tools for cellular imaging. *Chemical Society Reviews* **2009,** *38* (10), 2887.

78.   Bizzarri, R.; Serresi, M.; Luin, S.; Beltram, F., Green fluorescent protein based pH indicators for in vivo use: a review. *Analytical and Bioanalytical Chemistry* **2009,** *393* (4), 1107.

79.   Bencina, M., Illumination of the Spatial Order of Intracellular pH by Genetically Encoded pH-Sensitive Sensors. *Sensors (Basel, Switzerland)* **2013,** *13* (12), 16736.

80.   Kneen, M.; Farinas, J.; Li, Y. X.; Verkman, A. S., Green fluorescent protein as a noninvasive intracellular pH indicator. *Biophysical Journal* **1998,** *74* (3), 1591.

81.   Llopis, J.; McCaffery, J. M.; Miyawaki, A.; Farquhar, M. G.; Tsien, R. Y., Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** *95* (12), 6803.

82.   Mahon, M. J., pHluorin2: an enhanced, ratiometric, pH-sensitive green florescent protein. *Advances in bioscience and biotechnology* **2011,** *2.*

83.   Hanson, G. T.; McAnaney, T. B.; Park, E. S.; Rendell, M. E. P.; Yarbrough, D. K.; Chu, S. Y.; Xi, L. X.; Boxer, S. G.; Montrose, M. H.; Remington, S. J., Green fluorescent protein variants as ratiometric dual emission pH sensors. 1. Structural characterization and preliminary application. *Biochemistry* **2002,** *41* (52), 15477.

84.   Meech, S. R., Excited state reactions in fluorescent proteins. *Chemical Society Reviews* **2009,** *38* (10), 2922.

85.   Tonge, P. J.; Meech, S. R., Excited state dynamics in the green fluorescent protein. *Journal of Photochemistry and Photobiology a-Chemistry* **2009,** *205* (1), 1.

86.   Bizzarri, R.; Arcangeli, C.; Arosio, D.; Ricci, F.; Faraci, P.; Cardarelli, F.; Beltram, F., Development of a novel GFP-based ratiometric excitation and emission pH indicator for intracellular studies. *Biophysical Journal* **2006,** *90* (9), 3300.

87. Bizzarri, R.; Nifosi, R.; Abbruzzetti, S.; Rocchia, W.; Guidi, S.; Arosio, D.; Garau, G.; Campanini, B.; Grandi, E.; Ricci, F.; Viappiani, C.; Beltram, F., Green fluorescent protein ground states: The influence of a second protonation site near the chromophore. *Biochemistry* **2007,** *46* (18), 5494.

88. Griesbeck, O.; Baird, G. S.; Campbell, R. E.; Zacharias, D. A.; Tsien, R. Y., Reducing the environmental sensitivity of yellow fluorescent protein - Mechanism and applications. *Journal of Biological Chemistry* **2001,** *276* (31), 29188.

89. Abad, M. F. C.; Di Benedetto, G.; Magalhaes, P. J.; Filippin, L.; Pozzan, T., Mitochondrial pH monitored by a new engineered green fluorescent protein mutant. *Journal of Biological Chemistry* **2004,** *279* (12), 11521.

90. Lam, A. J.; St-Pierre, F.; Gong, Y.; Marshall, J. D.; Cranfill, P. J.; Baird, M. A.; McKeown, M. R.; Wiedenmann, J.; Davidson, M. W.; Schnitzer, M. J.; Tsien, R. Y.; Lin, M. Z., Improving FRET dynamic range with bright green and red fluorescent proteins. *Nature Methods* **2012,** *9* (10), 1005.

91. Ai, H.-W.; Olenych, S. G.; Wong, P.; Davidson, M. W.; Campbell, R. E., Hue-shifted monomeric variants of Clavularia cyan fluorescent protein: identification of the molecular determinants of color and applications in fluorescence imaging. *Bmc Biology* **2008,** *6*.

92. Schulte, A.; Lorenzen, I.; Boettcher, M.; Plieth, C., A novel fluorescent pH probe for expression in plants. *Plant Methods* **2006,** *2*.

93. Poburko, D.; Santo-Domingo, J.; Demaurex, N., Dynamic Regulation of the Mitochondrial Proton Gradient during Cytosolic Calcium Elevations. *Journal of Biological Chemistry* **2011,** *286* (13), 11672.

94. Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N. G.; Palmer, A. E.; Tsien, R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp red fluorescent protein. *Nature Biotechnology* **2004,** *22* (12), 1567.

95. Johnson, D. E.; Ai, H.-w.; Wong, P.; Young, J. D.; Campbell, R. E.; Casey, J. R., Red Fluorescent Protein pH Biosensor to Detect Concentrative Nucleoside Transport. *Journal of Biological Chemistry* **2009,** *284* (31), 20499.

96. Shcherbo, D.; Murphy, C. S.; Ermakova, G. V.; Solovieva, E. A.; Chepurnykh, T. V.; Shcheglov, A. S.; Verkhusha, V. V.; Pletnev, V. Z.; Hazelwood, K. L.; Roche, P. M.; Lukyanov, S.; Zaraisky, A. G.; Davidson, M. W.; Chudakov, D. M., Far-red fluorescent tags for protein imaging in living tissues. *Biochemical Journal* **2009,** *418*, 567.

97.   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* **2011,** *133* (26), 10034.

98.   Awaji, T.; Hirasawa, A.; Shirakawa, H.; Tsujimoto, G.; Miyazaki, S., Novel green fluorescent protein-based ratiometric indicators for monitoring pH in defined intracellular microdomains. *Biochemical and Biophysical Research Communications* **2001,** *289* (2), 457.

99.   Esposito, A.; Gralle, M.; Dani, M. A. C.; Lange, D.; Wouters, F. S., pHlameleons: A Family of FRET-Based Protein Sensors for Quantitative pH Imaging. *Biochemistry* **2008,** *47* (49), 13115.

100.   Gjetting, K. S. K.; Ytting, C. K.; Schulz, A.; Fuglsang, A. T., Live imaging of intra- and extracellular pH in plants using pHusion, a novel genetically encoded biosensor. *Journal of Experimental Botany* **2012,** *63* (8), 3207.

101.   Zhang, Y.; Xie, Q.; Robertson, J. B.; Johnson, C. H., pHlash: A New Genetically Encoded and Ratiometric Luminescence Sensor of Intracellular pH. *Plos One* **2012,** *7* (8).

102.   Li, G.; Zhu, D.; Xue, L.; Jiang, H., Quinoline-Based Fluorescent Probe for Ratiometric Detection of Lysosomal pH. *Organic Letters* **2013,** *15* (19), 5020.

103.   Wang, L.; Zhu, X.; Xie, C.; Ding, N.; Weng, X.; Lu, W.; Wei, X.; Li, C., Imaging acidosis in tumors using a pH-activated near-infrared fluorescence probe. *Chemical Communications* **2012,** *48* (95), 11677.

104.   Zhou, K.; Liu, H.; Zhang, S.; Huang, X.; Wang, Y.; Huang, G.; Sumer, B. D.; Gao, J., Multicolored pH-Tunable and Activatable Fluorescence Nanoplatform Responsive to Physiologic pH Stimuli. *Journal of the American Chemical Society* **2012,** *134* (18), 7803.

105.   Farinas, J.; Verkman, A. S., Receptor-mediated targeting of fluorescent probes in living cells. *Journal of Biological Chemistry* **1999,** *274* (12), 7603.

106.   Wu, M. M.; Llopis, J.; Adams, S.; McCaffery, J. M.; Kulomaa, M. S.; Machen, T. E.; Moore, H. P. H.; Tsien, R. Y., Organelle pH studies using targeted avidin and fluorescein-biotin. *Chemistry & Biology* **2000,** *7* (3), 197.

107.   Grover, A.; Schmidt, B. F.; Salter, R. D.; Watkins, S. C.; Waggoner, A. S.; Bruchez, M. P., Genetically Encoded pH Sensor for Tracking Surface Proteins through Endocytosis. *Angewandte Chemie-International Edition* **2012,** *51* (20), 4838.

108.   Szent-Gyorgyi, C.; Schmidt, B. A.; Creeger, Y.; Fisher, G. W.; Zakel, K. L.; Adler, S.; Fitzpatrick, J. A. J.; Woolford, C. A.; Yan, Q.; Vasilev, K. V.; Berget, P. B.; Bruchez, M. P.; Jarvik, J. W.; Waggoner, A., Fluorogen-activating single-chain antibodies for imaging cell surface proteins. *Nature Biotechnology* **2008,** *26* (2), 235.

109.   Yang, M.; Song, Y.; Zhang, M.; Lin, S.; Hao, Z.; Liang, Y.; Zhang, D.; Chen, P. R., Converting a Solvatochromic Fluorophore into a Protein-Based pH Indicator for Extreme Acidity. *Angewandte Chemie-International Edition* **2012,** *51* (31), 7674.

110.   Ernst, O. P.; Lodowski, D. T.; Elstner, M.; Hegemann, P.; Brown, L. S.; Kandori, H., Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chemical Reviews* **2014,** *114* (1), 126.

111.   Wang, W.; Nossoni, Z.; Berbasova, T.; Watson, C. T.; Yapici, I.; Lee, K. S. S.; Vasileiou, C.; Geiger, J. H.; Borhan, B., Tuning the Electronic Absorption of Protein-Embedded All-trans-Retinal. *Science* **2012,** *338* (6112), 1340.

112.   Kiser, P. D.; Golczak, M.; Palczewski, K., Chemistry of the Retinoid (Visual) Cycle. *Chemical Reviews* **2014,** *114* (1), 194.

113.   Hofmann, K. P.; Scheerer, P.; Hildebrand, P. W.; Choe, H.-W.; Park, J. H.; Heck, M.; Ernst, O. P., A G protein-coupled receptor at work: the rhodopsin model. *Trends in Biochemical Sciences* **2009,** *34* (11), 540.

114.   Shichida, Y.; Matsuyama, T., Evolution of opsins and phototransduction. *Philosophical Transactions of the Royal Society B-Biological Sciences* **2009,** *364* (1531), 2881.

115.   Hegemann, P.; Moeglich, A., Channelrhodopsin engineering and exploration of new optogenetic tools. *Nature Methods* **2011,** *8* (1), 39.

116.   Zhang, F.; Vierock, J.; Yizhar, O.; Fenno, L. E.; Tsunoda, S.; Kianianmomeni, A.; Prigge, M.; Berndt, A.; Cushman, J.; Polle, J.; Magnuson, J.; Hegemann, P.; Deisseroth, K., The Microbial Opsin Family of Optogenetic Tools. *Cell* **2011,** *147* (7), 1446.

117.   Kato, H. E.; Zhang, F.; Yizhar, O.; Ramakrishnan, C.; Nishizawa, T.; Hirata, K.; Ito, J.; Aita, Y.; Tsukazaki, T.; Hayashi, S.; Hegemann, P.; Maturana, A. D.; Ishitani, R.; Deisseroth, K.; Nureki, O., Crystal structure of the channelrhodopsin light-gated cation channel. *Nature* **2012,** *482* (7385), 369.

118.   Kralj, J. M.; Hochbaum, D. R.; Douglass, A. D.; Cohen, A. E., Electrical Spiking in Escherichia coli Probed with a Fluorescent Voltage-Indicating Protein. *Science* **2011,** *333* (6040), 345.

119.   Kralj, J. M.; Douglass, A. D.; Hochbaum, D. R.; Maclaurin, D.; Cohen, A. E., Optical recording of action potentials in mammalian neurons using a microbial rhodopsin. *Nature Methods* **2012,** *9* (1), 90.

120.   Maclaurin, D.; Venkatachalam, V.; Lee, H.; Cohen, A. E., Mechanism of voltage-sensitive fluorescence in a microbial rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110* (15), 5939.

121.   Park, J.; Werley, C. A.; Venkatachalam, V.; Kralj, J. M.; Dib-Hajj, S. D.; Waxman, S. G.; Cohen, A. E., Screening Fluorescent Voltage Indicators with Spontaneously Spiking HEK Cells. *Plos One* **2013,** *8* (12).

122.   Baasov, T.; Sheves, M., ALTERATION OF PKA OF THE BACTERIORHODOPSIN PROTONATED SCHIFF-BASE - A STUDY WITH MODEL COMPOUNDS. *Biochemistry* **1986,** *25* (18), 5249.

123.   Gat, Y.; Sheves, M., A MECHANISM FOR CONTROLLING THE PK(A) OF THE RETINAL PROTONATED SCHIFF-BASE IN RETINAL PROTEINS - A STUDY WITH MODEL COMPOUNDS. *Journal of the American Chemical Society* **1993,** *115* (9), 3772.

124.   Sakmar, T. P.; Franke, R. R.; Khorana, H. G., THE ROLE OF THE RETINYLIDENE SCHIFF-BASE COUNTERION IN RHODOPSIN IN DETERMINING WAVELENGTH ABSORBENCY AND SCHIFF-BASE PKA. *Proceedings of the National Academy of Sciences of the United States of America* **1991,** *88* (8), 3079.

125.   Steinberg, G.; Ottolenghi, M.; Sheves, M., PK(A) OF THE PROTONATED SCHIFF-BASE OF BOVINE RHODOPSIN - A STUDY WITH ARTIFICIAL PIGMENTS. *Biophysical Journal* **1993,** *64* (5), 1499.

126.   Okada, T.; Sugihara, M.; Bondar, A. N.; Elstner, M.; Entel, P.; Buss, V., The retinal conformation and its environment in rhodopsin in light of a new 2.2 angstrom crystal structure. *Journal of Molecular Biology* **2004,** *342* (2), 571.

127.   Choe, H.-W.; Kim, Y. J.; Park, J. H.; Morizumi, T.; Pai, E. F.; Krauss, N.; Hofmann, K. P.; Scheerer, P.; Ernst, O. P., Crystal structure of metarhodopsin II. *Nature* **2011,** *471* (7340), 651.

128.   Murakami, M.; Kouyama, T., Crystal structure of squid rhodopsin. *Nature* **2008,** *453* (7193), 363.

129.   Borshchevskiy, V. I.; Round, E. S.; Popov, A. N.; Bueldt, G.; Gordeliy, V. I., X-ray-Radiation-Induced Changes in Bacteriorhodopsin Structure. *Journal of Molecular Biology* **2011,** *409* (5), 813.

CHAPTER II: REENGINEERING OF CELLULAR RETINOIC ACID BINDING PROTEIN (CRABPII) INTO A pH-RESPONSIVE COLORIMETRIC PROBE

II.1 Introduction

Cellular Retinoic Acid Binding Protein II (CRABPII) belongs to the family of intracellular lipid binding proteins (iLBPs), which have the main function to solubilize and transport hydrophobic molecules in the organism. CRABPII binds and regulates the intracellular concentrations of all-*trans*-retinoic acid (RA) **II-1** (**Figure II-1a**), a metabolite of Vitamin A.[1] After binding, it not only solubilizes RA, but also transports it from the cytosol to the nucleus of the cell as confirmed by CRABPII nuclear localization studies.[2-3] After getting inside the nucleus it transfers RA to the retinoic acid receptor (RAR), which is involved in transcriptional activity of specific genes. CRABPII is a chaperone because it participates in gene regulation through RAR.[2] Like other iLBPs it has two α-helices and two five-stranded β-sheets (**Figure II-1b**).[4-5] CRABPII binds RA with high affinity, K_d = 2 nM.[6] Detailed analysis of the CRABPII crystal structure reveals the presence of hydrophobic residues along the polyene and a tight hydrogen bonding network with RA's carboxylate (**Figure II-1c**), leading to the observed high affinity for binding. By itself CRABPII is an ideal platform for mutagenesis studies because it is a structurally robust protein, and it contains only a few highly conserved amino acids among the iLBP family. With the last two facts in mind our lab initiated the reengineering of CRABPII into a rhodopsin mimic.[7-8] In order to build a rhodopsin mimic the initial goal is to promote the

reaction between the aldehyde group of retinal and the amine group of a lysine residue to form an iminium bond. During CRABPII reengineering, this step involved a ligand switch from the non-covalently bound carboxylic acid to an aldehyde, which forms iminium bond with a properly positioned lysine residue (**Figure II-2a**).

In the first generation rhodopsin mimic, the protonated Schiff base (PSB)

a.

all-*trans*-retinoic acid (RA)
II-1

b.

c.
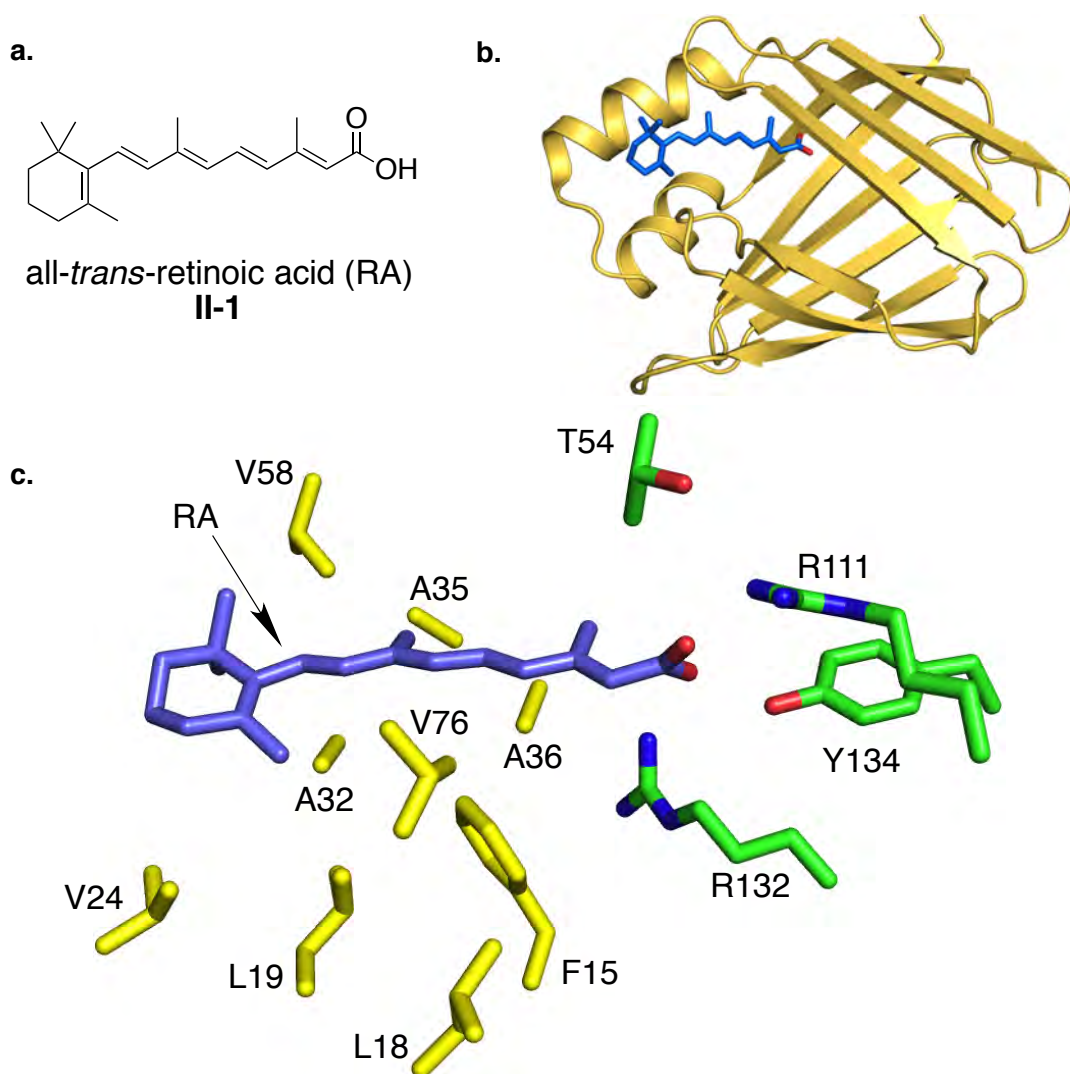
Figure II-1: Crystal structure of WT-CRABPII (PDB ID: 2FR3). **a.** The chemical structure of CRABPII native ligand all-*trans*-retinoic acid (RA); **b.** The tertiary structure of CRABPII bound with RA; **c.** The residues around RA (hydrophobic yellow, hydrophilic green).

was formed through an engineered R132K residue, as confirmed by the crystal structure of the R132K:R111L:L121E triple mutant (**Figure II-2b**).[8] The carboxylate L121E was introduced as a counterion stabilizing the positive charge on the PSB (**Figure II-2c**), while the R111L mutation allowed the ligand to adopt the proper Burgi-Dunitz trajectory for the Lys132 nucleophilic attack to occur. The final CRABPII-R132K:R111L:R121E (CRABPII-KLE) triple mutant bound retinal as PSB with dissociation constant of 1 nM but, unfortunately, it was not capable of modulating the wavelength regardless of mutations at residue Arg59 (435 –
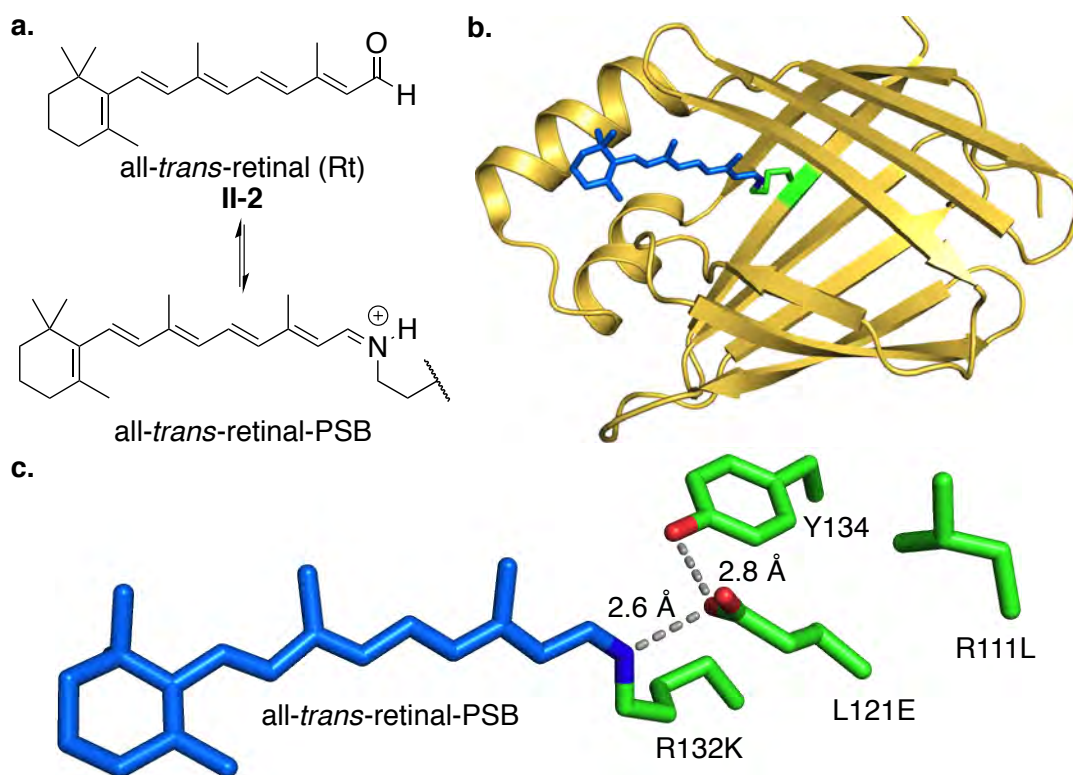
Figure II-2: Crystal structure of CRABPII-R132K:R111L:L121E (PDB ID: 2G7B). **a.** The chemical reaction between all-*trans*-retinal and engineered CRABPII; **b.** The tertiary structure of CRABPII-R132K:R111L:L121E covalently bound with Rt; **c.** The residues involved in successful engineering of CRABPII into a first generation rhodopsin mimic.

450 nm).[9] This failure was attributed to the partial exposure of the chromophore to the highly polar aqueous media, identified from close examination of the crystal structure. If this hypothesis is true, encapsulation of chromophore inside the protein cavity is expected to solve the problem and result in wavelength regulation. The modest wavelength regulation (398 – 424 nm or 1190 – 2731 cm^{-1}) was demonstrated with the two double bonds shorter C15-retinal analog (**Figure II-3a**).[9] When bound to the first generation CRABPII mutants the C15-retinal analog is fully embedded by the protein as evident from the crystal
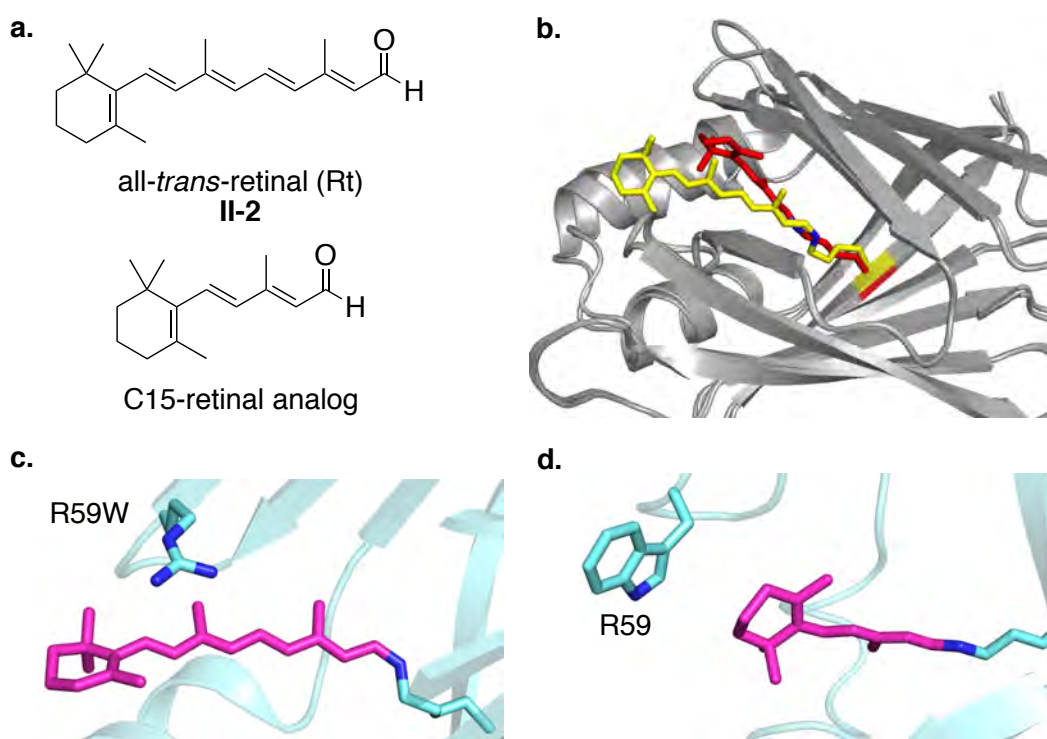
Figure II-3: Comparison of CRABPII mutants binding full-length retinal and C15-retinal analog. **a.** Chemical structures of all-*trans*-retinal and all-*trans*-C15-retinal analog. **b.** Overlayed crystal structures of CRABPII mutants covalently bound with retinal (yellow, PDB ID: 2G7B) and C15-retinal analog (red, PDB ID: 3F8A). The C15-retinal analog is deeper inside the protein cavity as opposed to retinal. **c.** Location of the residue R59W to the C15-retinal analog. **d.** Location of the residue R59 relative to the full-length retinal.

structure (**Figure II-3b**). Because of this encapsulation, the chromophore responds well when residue Arg59 is mutated into amino acids of different polarity, and the wavelength can be regulated (**Figures II-3c and II-3d**).
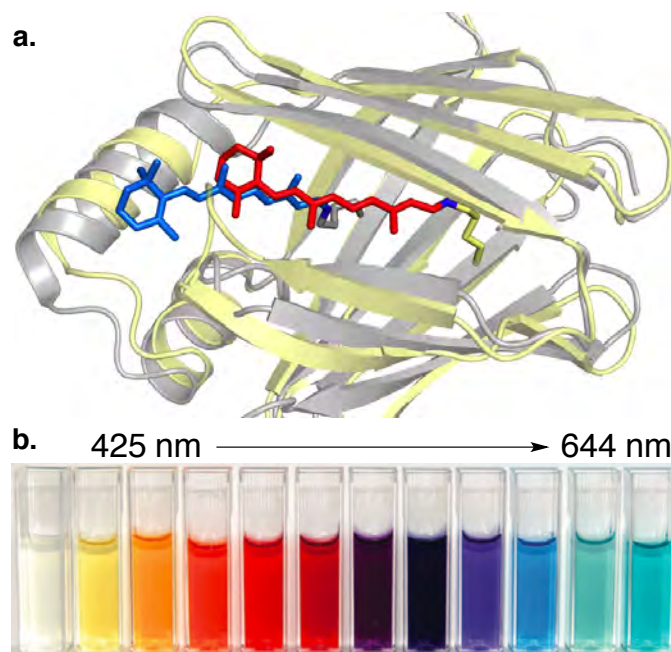
Figure II-4: Wavelength regulation in hCRBPII mutants. **a.** Overlayed crystal structures of hCRBPII-Q108K:K40L:T51V: T53C:R58W:T29L:Y19W:Q4A hepta mutant (protein yellow, retinal red; PDB ID: 4GKC) and CRABPII-KLE mutant (protein grey, retinal blue) covalently bound with retinal. **b.** Visual depiction of the wavelength regulation from 425 nm to 644 nm for protein-embedded retinal.

Furthermore, the wavelength regulation of the full-length retinal chromophore was achieved with an alternate protein system that would better sequester the bound ligand. Another representative from the iLBP family,[10-11] human Cellular Retinol Binding Protein II (hCRBPII), was reengineered into a rhodopsin mimic with a binding site placed about 6 Å deeper as compared to the first generation CRABPII (**Figure II-4a**). A series of hCRBPII mutants provided extraordinary tuning of the protein-embedded all-*trans*-retinal absorption ranging

from 425 to 644 nm (**Figure II-4b**),[12] surpassing the predicted limit for the retinylidene λ_{max}.[13-14] The uniform distribution of electrostatic field around the protein-embedded chromophore in combination with the lack of a counterion was the crucial step towards designing the most red-shifted pigments. Isolating the retinyledene from the bulk medium by enclosing the binding cavity with large hydrophobic amino acids proved to be necessary for the final push in bathochromic shift. On the other hand, localized negative electrostatic potential projected on the retinylidene from the surrounding amino acids gave consistent blue-shift in wavelength among various hCRBPII mutants.[15]

Since encapsulating the bound chromophore was proven essential for wavelength regulation, and in an attempt to verify our hypothesis, the original protein, CRABPII, was revisited and second generation CRABPII mutants were designed. The goal was to reengineer the protein binding cavity to bind retinal further within the pocket, similar to hCRBPII, and examine the principles discovered for wavelength tuning previously. This chapter describes our approach of reengineering the CRABPII binding site, followed by probing the wavelength regulation mechanism. After the initial goal was achieved an inspiring observation led to a new direction of this project: the pK_a values of the PSB provide an ideal platform for a novel protein-based pH-probe design. This approach was utilized to build an absorbance based ratiometric pH-sensor, which allows for an accurate pH determination. Although absorbance based sensors

have limited practical application, this proof-of-principle study introduces a new class of protein based pH-probes.

II.2 Overview of the CRABPII interactions with retinoic acid and retinal

In order to propose the strategy for CRABPII reengineering it is essential to understand the ligand/protein interactions described for this protein earlier. As briefly mentioned in the previous section, CRABPII binds RA through a tight hydrogen bonding network.[5] As evident from the wild-type-CRABPII (WT-CRABPII) crystal structure, the main players of this network are direct interactions of Arg132 and Tyr134 with one of the oxygen atoms of RA and a water mediated hydrogen bonding network from Arg111 and Thr54 to the other oxygen of the native ligand (**Figure II-5a**). Interestingly, the last binding motif was
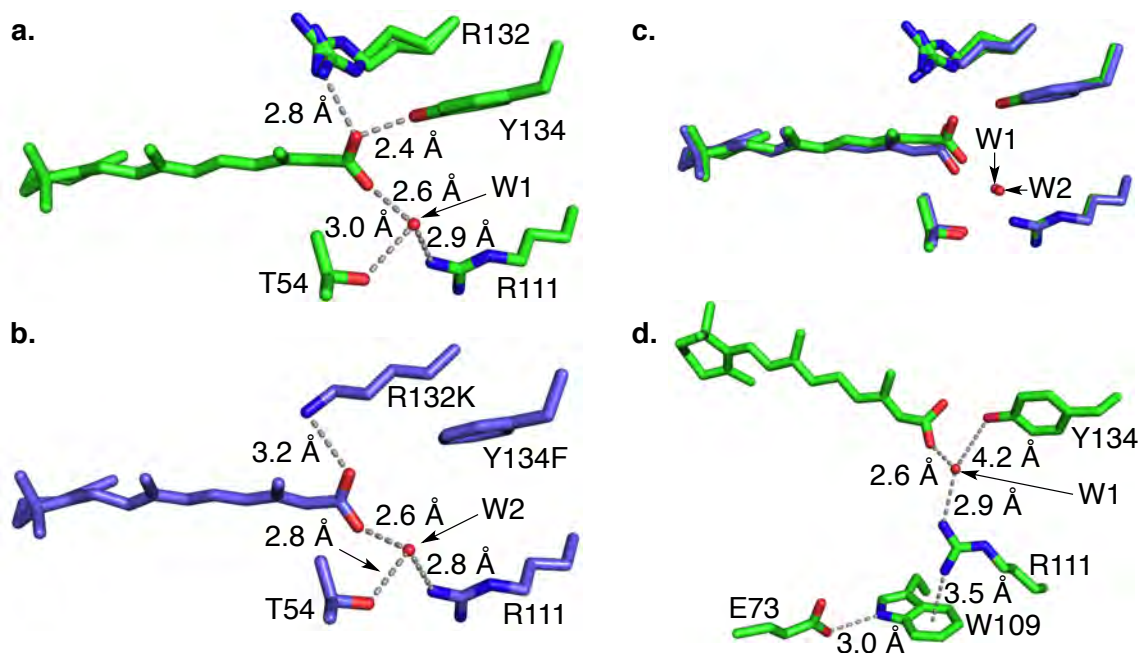
Figure II-5: Interactions of CRABPII with RA. **a.** The residues of WT-CRABPII involved in binding with RA. **b.** Crystal structure of CRABPII-R132K:Y134F (KF) bound with RA (PDB ID: 2G78). **c.** Overlayed crystal structures of WT-CRABPII and KF double mutant. **d.** Spatial arrangement of R111 enforced by W109 and E73.

observed in the crystal structures of CRABPII-R132K:Y134F double mutant (**Figures II-5b and II-6a**).[6-8, 16]

The distances from Arg132 and from the critical water molecule to their respective oxygen atoms of RA in WT-CRABPII are almost the same (2.8 and 2.6 Å, respectively, **Figure II-5a**). However, Arg132 has two possible conformations in the crystal structure as oppose to the rest of the critical residues. Moreover, comparison of WT-CRABPII and CRABPII-R132K:Y134F crystal structures bound with RA identified that Arg111, Thr54 and organized water molecules, W1 and W2, align perfectly between two mutants (**Figure II-5c**). To probe the role of Arg132 in RA binding this residue was mutated into R132K, R132L and R132M, however, the drop in RA binding affinity was not as dramatic as expected.[6] Subsequently, analysis of Tyr134 revealed that the hydroxyl group of the tyrosine also does not contribute much to the binding affinity. Finally, the double mutant CRABPII-R132K:Y134F was crystallized with RA (**Figure II-5b**). Despite two mutations at the binding site the water-mediated interactions between RA,
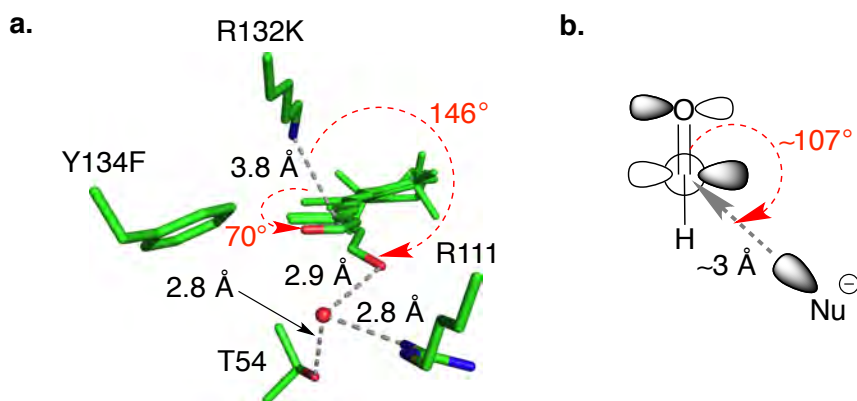
Figure II-6: Burgi-Dunitz trajectory. **a.** Crystal structure of CRABPII-R132K:Y134F mutant bound with Rt: no iminium bond formation due to unfavorable interactions (PDB ID: 2G79). **b.** Classical Burgi-Dunitz trajectory.

Arg111 and Thr54 where in the same conformation as in WT (**Figure II-5c**). Interestingly, the distance between the carboxylate of RA and R132K was 3.2 Å (**Figure II-5b**). All of these studies suggest that the interactions of RA with Arg132 and Tyr134 are relatively flexible and the residues Arg111 and Thr54 contribute the most to the binding affinity (**Table II-1**).

Exploring the latter statement CRABPII-R132K:Y134F mutant was subjected to mutagenesis with R111L and T54V. As expected, both mutations introduced individually and in combination, dramatically reduced affinity of RA to CRABPII (**Table II-1**). Furthermore, other residues interacting with Arg111 proved to be important for locking the necessary orientation for high affinity to RA. For example, Trp109 was proposed to participate in π-cation interaction with Arg111 (**Figure II-5d**). But at the same time Trp109 has interactions with Glu73 and therefore mutations of Glu73 has an impact on affinity (**Table II-1**).[6]

All of the analysis surmised above provided the unprecedented evidence

Table II-1. Binding of all-*trans*-retinoic acid to CRABPII mutants.

Entry	Mutant	K_d
1	WT-CRABPII	2±1.2
2	R132K	65±14
3	R132K:Y134F	100±7.1
4	R132K:R111L	736±114
5	R132K:Y134F:R111L	1000±28
6	R132K:Y134F:T54V	565±79
7	R132K:Y134F:R111L:T54V	900±64
8	R132K:W109L	3196±252
9	R132K:E73A	564±56

that Arg111 is one of the most (if not the most) crucial residues in RA binding. This led to the R111L mutation during the first generation rhodopsin mimic design, which improved PSB formation with retinal by removing the undesired water mediated interaction between Arg111 and the carbonyl oxygen of Rt. When the retinal is trapped by an ordered water molecule into unfavourable confirmation, it is unable to undergo nucleophilic attack from the lysine R132K, as evident from the crystal structure of the CRABPII-R132K:Y134F double mutant that failed to form a Schiff base (**Figure II-6a**). Ideally, the Burgi-Dunitz distance is within 3 Å, in order to be in proximity for a new bond formation, and an attack angle close to 107° in order to avoid repulsion with the π-orbital of the carbonyl, but still provide a significant overlap between the nitrogen lone pair and the antibonding orbital of the carbonyl (**Figure II-6a**).[17-18]

Ironically, Arg111 became the first target to introduce a nucleophilic lysine for engineering the deeper binding site in CRABPII-R111K. The main reason is that an analogous mutation, Q108K, proved to work as a nucleophilic lysine for PSB formation during the engineering of human Cellular Retinol Binding Protein II (hCRBPII) into a rhodopsin mimic.[12]

II.3 Identification of the residues for a new retinal binding site placed deep in the CRABPII cavity

Detailed analysis and *in silico* modeling of the CRABPII binding pocket identified three amino acids placed deep within the binding pocket to cover the full-length chromophore. Putative mutations included R111K, I52K and I63K

because they are located over 10 Å deeper as compared to R132K (**Figure II-7a**). Each of the proposed mutants individually introduced in WT-CRABPII crystal structure was used to build docking models with retinal. The models provided a better estimate about the special arrangement of all-*trans*-retinal relative to lysine. For this purpose each of the highlighted residues was introduced in the apo-WT-CRABPII crystal structure (PDB ID: 2FS6). Removal of the water molecules was necessary to provide more space for the ligand to accommodate itself in the binding cavity. The apo-WT-CRABPII crystal structure was used to generate models because it is slightly more opened than any holo-CRABPII structure and might be more flexible in modeling the ligand binding.[4-5, 19-20] The energy-minimized all-*trans*-retinal with locked geometry along the polyene was used as the ligand. With these settings the docking model of CRABPII with retinal would have higher probability to accommodate the ligand, however it did not account for the possible lysine side chain movement.
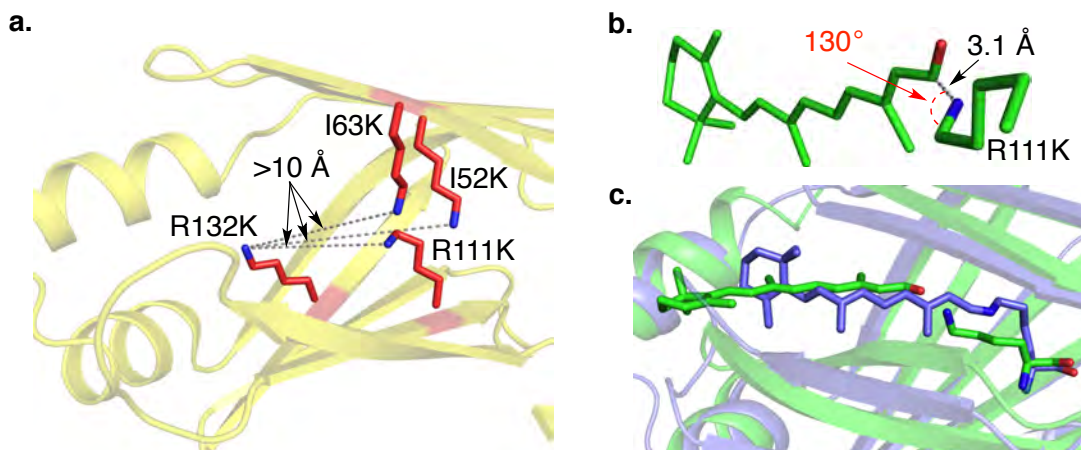
Figure II-7: Detecting the new CRABPII binding site. **a.** Model of the possible nucleophiles in CRABPII deeper pocket and their location relative to the original R132K. **b.** Docking model of CRABPII-R111K mutant with Rt. **c.** Overlay of the crystal structure for hCRBPII-Q108K:K40L (blue) with model of CRABPII-R111K (green). Alignment of R111K and Q108K is shown in stick.

In silico modelling of CRABPII-R111K with Rt suggested the optimum geometry for the putative nucleophilic attack: 3.1 Å distance to the carbonyl carbon with dihedral angle 130° (**Figure II-7b**). But more importantly, the residue R111K aligned with reactive lysine of hCRBPII-Q108K:K40L mutant. At the time when this project was initiated, our lab was already aware of PSB formation with hCRBPII-Q108K:K40L and it inspired us to go after the equivalent residue (Arg111 in CRABPII is analogous to Qln108 in hCRBPII) in CRABPII for placing the reactive lysine (**Figure II-7c**).
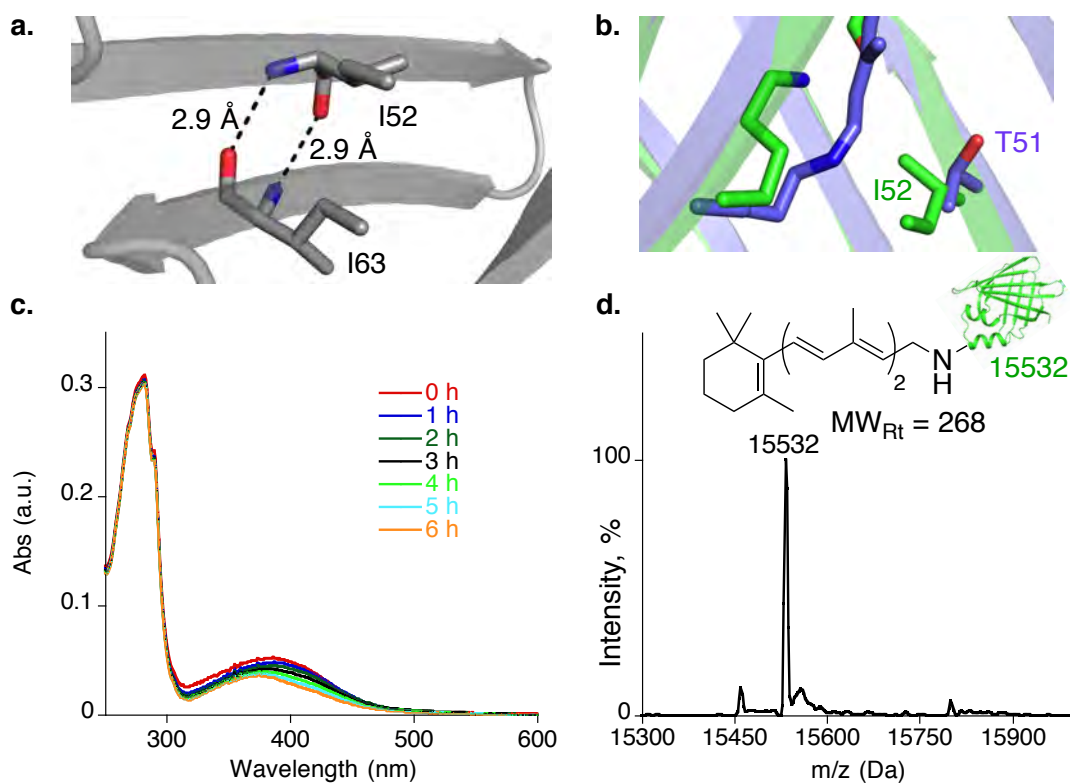
Figure II-8: Choosing a nucleophilic lysine for the new CRABPII binding site. **a.** Interactions between I52 and I63 leads to protein instability if either one is mutated into polar lysine. **b.** Overlay of the CRABPII-R111K model (green) and hCRBPII-Q108K:K40L (blue) crystal structure highlighting the difference between I52 and T51 in respective proteins. **c.** UV-vis of CRABPII-R111K with 0.2 equivalent of retinal over 6 h. **d**. Reductive amination test of CRABPII-R111K with Rt. There is no peak at M+268 corresponding to adduct formation.

All three identified residues I52K, I63K and R111K were targeted for mutagenesis in WT-CRABPII. Unfortunately, only the R111K single mutant could be expressed, purified in significant quantity (20 mg/L) and survived without precipitation for multiple days at 4 $^{\circ}$C. The reason for the folding and solubility problem for the I52K and I63K mutants could be the hydrophobic interactions between residues of the neighboring β-sheets (**Figure II-8a**). Both I63 and I52 are situated in anti-parallel β-sheets keeping the main chain amide bonds within 2.9 Å from one another. At the same time the isoleucine side chains provides hydrophobic packing. Hypothetically, mutagenesis of either one of them into a polar amino acid would result in a disturbed interaction and subsequently in the observed poor protein folding and stability.

Expressed and purified CRABPII-R111K single mutant was incubated with retinal for up to 6 h, however, no peak indicative of PSB formation was recognized from UV-vis. In fact, the spectrum showed only retinal degradation over the time of the experiment (**Figure II-8c**). Mass spectrometry (ES^{+}) verified that no covalent bond was formed. The same mutant was incubated with one equivalent of retinal for 24 h and subsequently subjected to reductive ammination conditions, which would reduce the imine double bond and therefore trap the adduct irreversibly. The product of NaCNBH$_3$-treated CRABPII/Rt complex was analysed by Q-Tof ES^{+} but only the mass corresponding to the protein peak and not a reductively aminated adduct was evident from the acquired data (**Figure II-**

8d). Thus, it was confirmed that R111K single mutant formed neither imine nor iminium with Rt.

Comparison of the CRABPII-R111K model and hCRBPII-Q108K:K40L crystal structure showed that similarly to hydrophobic K40L rationally designed in hCRBPII, CRABPII already has V41 at the corresponding residue. The difference was found at T51 and I52 residues in hCRBPII and CRABPII, respectively (**Figure II-8b**). The hydroxyl group of threonine could hypothetically activate the carbonyl of the Rt for nucleophilic attack from the lysine in hCRBPII. To probe if this difference has a role in PSB formation the I52T mutation was attempted in CRABPII-R111K despite the fact that I52K significantly decreased protein stability before. The CRABPII-R111K:I52T double mutant was expressed in much smaller amounts (1.4 mg/L) and it precipitated during 3 h, while measuring UV-vis with Rt. Moreover, parallel with this study, it was found that in hCRBPII-Q108K:K40L:T51V mutant the PSB formation with retinal is not inhibited and, in fact, T51V mutation makes the protein expression level higher. Thus, the hydrophobic residue at 52 in CRABPII is actually better for protein stability and it is not the right place to promote reaction between the protein and chromophore. Therefore, further CRABPII mutagenesis at the residue I52 was purposely omitted.

II.4. Reengineering of the CRABPII binding pocket

Additional insight for the deeper CRABPII binding site design came from previous work from our lab.[8, 16] Evoking the story about CRABPII-R132K:Y134F

double mutant binding with retinal without covalent bond formation, we proposed a hypothesis for the failed SB/PSB formation between CRABPII-R111K and Rt. The R111K mutation did not dramatically change the nature of the binding pocket because at physiological pH both lysine and arginine are positively charged (pK_a 10.5 and 12.5, respectively) (**Figure II-9a**). The R111K mutation could still accommodate the previously described ordered water molecule between the hydrophilic residues and Rt despite the fact that lysine is shorter than arginine and retinal is less polar than retinoic acid (**Figure II-9b**). The trapped Rt at the original binding site would not be able to slide down the protein cavity at the new binding site. The hydrophilic nature of the protein pocket would not accomodate Rt because of two reasons: (1) the interactions between the carbonyl group of retinal and the polar amino acids in the original location and (2) the presence of hydrophilic amino acids known to interact directly with the carboxylate of retinoic acid, which would not interact favorably with a hydrophobic polyene.

In an attempt to test the effect of the hydrophobic environment on SB
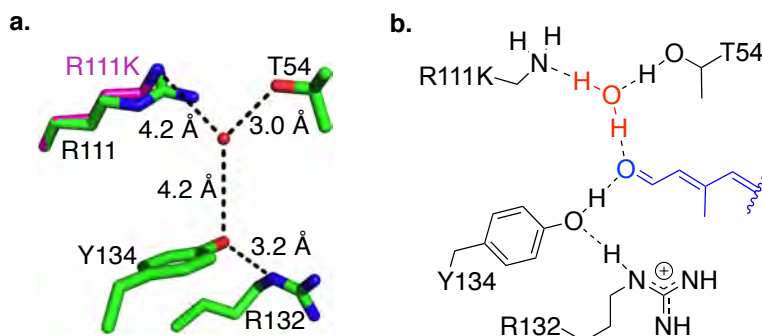
Figure II-9: Hypothetical interactions between R111K and Rt. **a.** Overlay of the WT-CRABPII (green) with CRABPII-R111K (magenta) highlighting the distance to an ordered water from R111K. **b.** Possible mode of binding CRABPII-R111K with Rt.

formation in a completely new binding site, we decided to replace all polar residues known to interact with the native ligand with hydrophobic ones. As mentioned before, positively charged R132 is a crucial residue for retinoic acid binding to CRABPII. Therefore, it was the first candidate to undergo mutagenesis. Double mutant R111K:R132K as described before[6] binds Rt to yield a complex with a λ_{max} = 460 nm (**Table II-2**, entry 4). The absorption wavelength was similar to that observed in R132K:R111L:L121E (**Table II-2**, entry 1) suggesting the PSB formation through R132K. To eliminate this possibility as a negative artifact, R111K:R132L mutant was made. Gratifyingly, introduction of R132L resulted in a protein competent for PSB formation (**Table II-2**, entry 3). Reductive amination of the SB followed by Q-Tof ES^{+} analysis indicates that CRABPII-R111K:R132L forms SB with retinal. However, PSB formation was not complete, even at pH of 5.0. Deconvolution of the UV-vis spectrum of the SB/PSB mixture reveals a significantly red-shifted 527 nm

Table II-2. Binding of all-*trans*-retinal to selected CRABPII mutants.

Entry	Mutant	λ_{max} (nm)	K_d	Red. Am.
1	R132K:R111L:L121E	449	1±5	yes
2	R111K	385	460±34	no
3	R111K:R132L	527[a]	63±5	yes
4	R111K:R132K	468	73±12	yes
5	R111L:R132K	408	567±36	yes
6	R111K:R132L:Y134F	525	54±13	yes
7	R111L:R132L:Y134F	406	7±2	no
8	R111K:R132L:Y134F:T54V	534	5±9	yes

[a]Deconvolution was applied.

absorbance for the PSB. **Figure II-10a** provides an overlay of the recorded UV-vis peak and its deconvolution (PeakFit v4.12) for the CRABPII-R111K:R132L double mutant which was incubated with retinal for 12 h at room temperature at pH=5. The deconvolution was performed using Gaussian distribution to three peaks because after 12 h of incubation retinal could non-specifically bind to the lysines on the protein surface while the acidic pH would lead to PSB formation absorbing at 440 nm in PBS buffer. Lowering the pH did not lead to any observable change in PSB formation. The peak at 527 nm indicates PSB formation for this double mutant. Reductive amination (**Figure II-10b**) for the same double mutant shows a retinal adduct (M+268) peak, which confirms SB formation. Since R132L was determined to be important to promote SB formation it was maintained in most of the following mutants.

It is interesting to note that for the successful reengineering of the previous series of mutants (R132K series) a similar replacement was required but in a
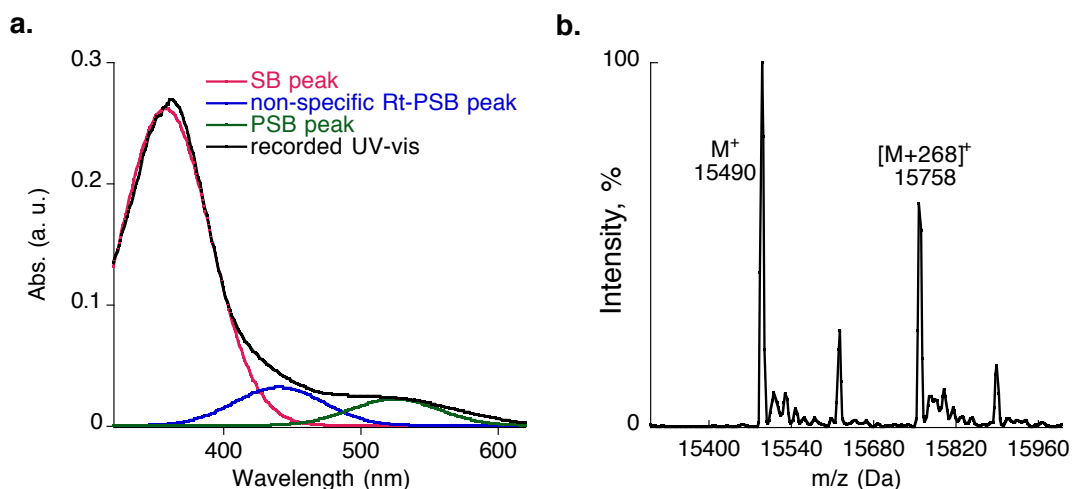
Figure II-10: CRABPII- R111K:R132L binding with Rt. **a.** UV-vis after 12 h at pH 5 and deconvolution data. **b.** Q-Tof ES[+] data of reductively aminated complex.

reverse manner (R132K:R111L) in order to remove the ordered water molecule and improve PSB formation (**Table II-2**, entry 5). Based on previous results (**Table II-2**, entry 4) we also know that the double mutant R132K:R111K can bind retinal as a SB. To evaluate the role and importance of each of these lysines the double mutants R111K:R132K and R111L:R132K were directly compared. The chromophore binding (K_d) was estimated using tryptophan fluorescence quenching. This method is based on the measurement of tryptophan emission at 350 nm which would be quenched by absorbance of any chromophore with UV-vis peak at this wavelength. Rt and Rt-SB absorb at 380 nm and 365 nm, respectively, which can be an effective quencher for tryptophan emission. The amount of fluorescence quenching is expected to be different based on the tryptophan location relative to the chromophore (**Figure II-11a**). As depicted in
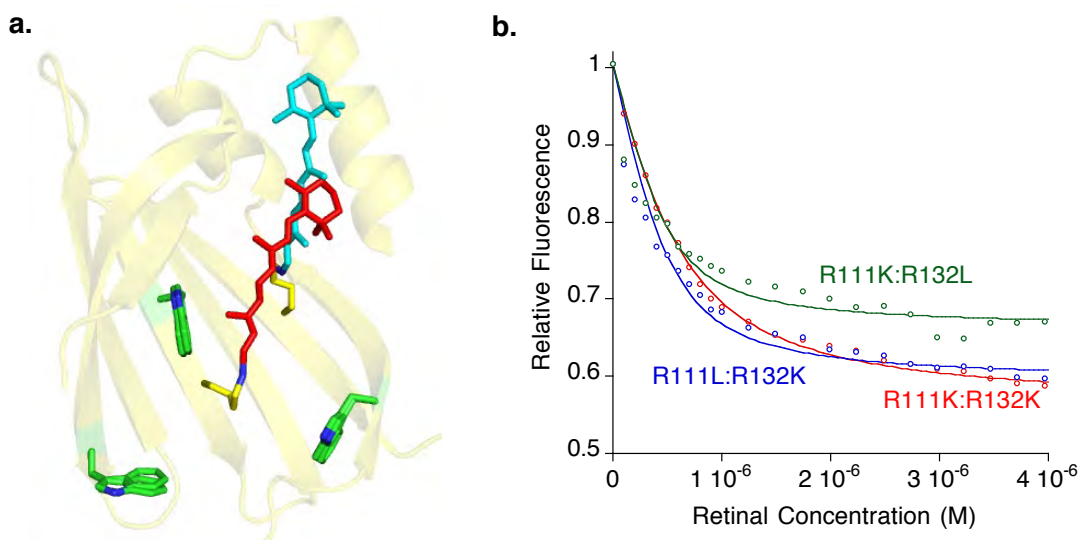
Figure II-11: Dissecting the nucleophilic lysine in SB formation with Rt inside R111K:R132K. **a.** Overlayed SB-Rt through R132K (cyan) and through R111K (red) with all tryptophan residues highlighted based on R132K:R111L:L121E crystal structure and R111K model. **b.** Tryptophan fluorescence quenching of the three depicted mutants.

Figure II-11, decrease in the relative fluorescence intensity for R111K:R132K is similar to the one observed for R111L:R132K, and distinct from R111K:R132L which suggests binding through original R132K in R132K:R111K mutant (**Figure II-11b**). Because there is a possibility that R132 or R132K could interact with the carbonyl of Rt, and thus, not allow binding to R111K, R132 was mutated to non-interacting amino acids for the following studies.
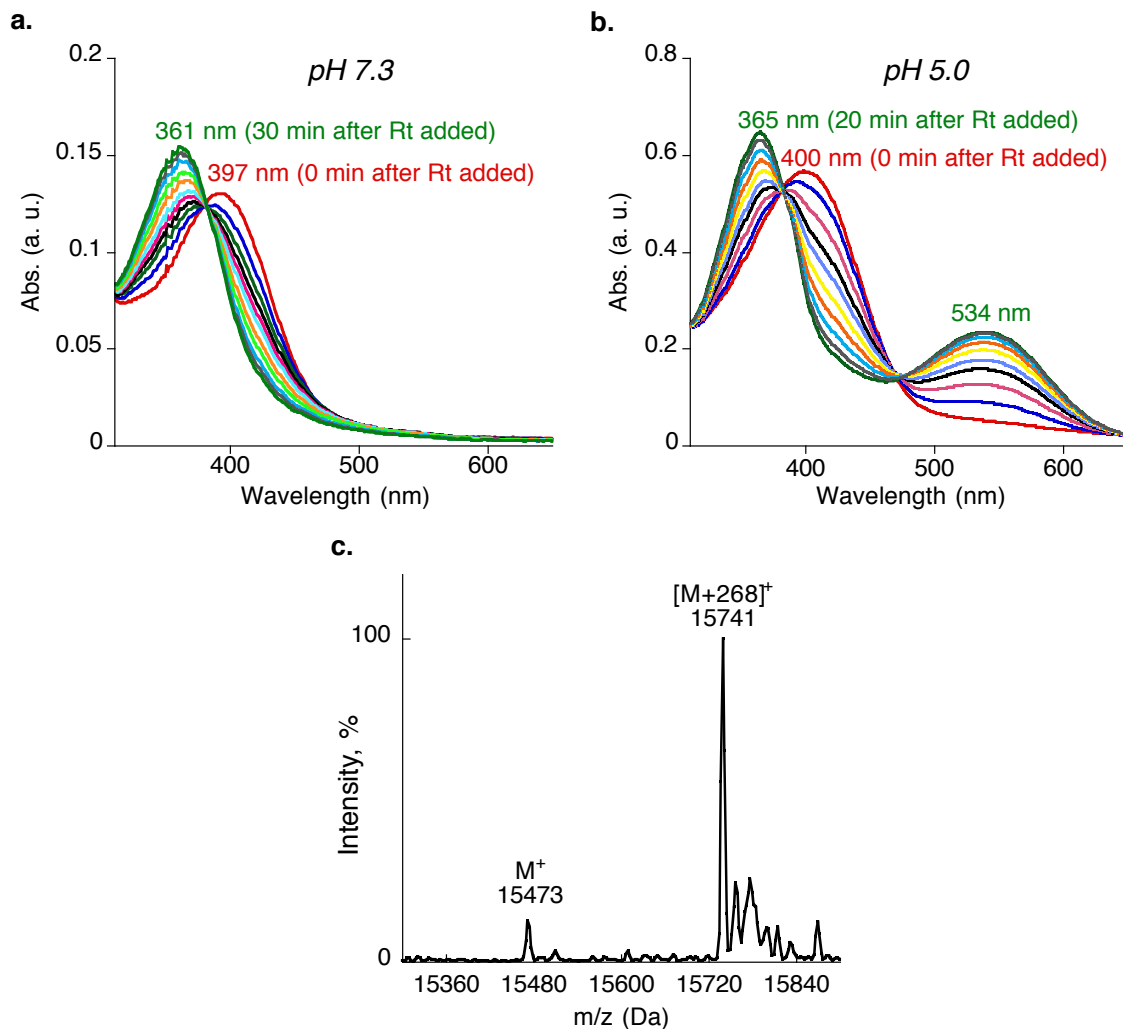
Figure II-12: Spectroscopic properties of R111K:R132L:Y134F:T54V binding with Rt. **a.** UV-vis at pH 7.3 indicates SB formation after 30 min but not PSB. **b.** UV-vis at pH 5.0. **c.** Reductive amination test indicate that most protein forms covalent bond with one equivalent of Rt.

The next polar residue, Y134, was also suggested to participate in the activation of Rt for nucleophilic attack by R132K. In an attempt to redirect binding to R111K we replaced Y134 with phenylalanine resulting in the triple mutant R111K:R132L:Y134F (**Table II-2**, entry 6). The latter mutant gives clear absorption maxima without deconvolution at pH 5.0 and exhibits improved binding affinity to Rt. To prove that iminium forms through R111K the nucleophilic lysine was removed (R111L). The control mutant R111L:R132L:Y134F failed to form an SB or PSB at any pH (**Table II-2**, entry 7). The red-shift of the non-covalently bound Rt relative to its free form (406 nm *versus* 380 nm, respectively) and high affinity of the mutant to the ligand (K_d = 7 ± 2 nM) were the result of tuned hydrophobicity in the binding cavity. The λ_{max} of the non-covalently bound Rt is 406 nm, which is consistent with the wavelength observed in the PSB forming mutants at 0 min after addition of Rt (**Figure II-12a** and **b**).

Finally, removal of the last polar amino acid, T54, involved in retinoic acid binding gave tetra mutant R111K:R132L:Y134F:T54V, with significantly improved binding affinity to Rt (**Table II-2**, entry 8). Moreover, it caused a 9 nm red-shift as compared to the parent triple mutant supporting the previously derived

```
5     NGTWEMESNENFEGYMKALDIDFATRKIAVRLTQTKVID--QDGDNFKTKTTSTFRNYDV   62
      +G W++   +ENFE   +K L ++    RKIAV         ++  Q+GD F   KT++T R   ++
4     SGNWKIIRSENFEELLKVLGVNVMLRKIAVAAASKPAVEIKQEGDTFYIKTSTTVRTTEI   63

63    DFTVGVEFDEYTKSLDNRHVKALVTWEGD-VLVCVQ---KGEKENRGWKQWIEGD-KLYL   117
      +F VG EF+E T  +D R  K+LV WE +  +VC Q    KGE      W + +  D +L L
64    NFKVGEEFEEQT--VDGRPCKSLVKWESENKMVCEQKLLKGEGPKTSWTRELTNDGELIL   121

118   ELTCGDQVCRQVFKKK   133
      +T  D VC +V+ ++
122   TMTADDVVCTRVYVRE   137
```

Figure II-13: Protein sequence alignment of WT-hCRBPII (top line) and WT-CRABPII (bottom line).

hypothesis that more electrostatically neutral protein environment provides not only improved ligand-protein interactions but also the bathochromic shift of the CRABPII bound retinylidene. The tetra mutant formed no PSB at pH 7.3 (**Figure II-12a**), however, SB formation was evident from a 38 nm blue-shift 30 min after incubation with the chromophore as compared to the non-covalently bound Rt when it was just added to the mutant. The PSB formation was clearly observed during UV-vis measurement at pH 5.2 (**Figure II-12b**) indicating that the pK$_a$ of the PSB in CRABPII-R111K:R132L:Y134F:T54V was lower than physiological pH. Most of the protein was converted into covalently bound complex after 24 h incubation with 1 equivalent of ligand, based on reductive amination results (**Figure II-12c**).

The R111K:R132L:Y134F:T54V-CRABPII tetra mutant (KLFV), exhibiting high retinal affinity and PSB absorption at 534 nm (Table 1, entry 5), was chosen as the parent protein for further mutagenesis to produce blue- and red-shifted pigments. The overall strategy for producing the most bathochromically shifted pigments was borrowed from hCRBPII studies: (1) the binding pocket must be isolated as much as possible from the aqueous media; (2) the protonation of the Schiff base must be stabilized in the absence of a directly interacting counteranion; (3) appropriate amino acids along the length of the polyene must be positioned to project a uniform and relatively neutral electrostatic field, leading to maximal conjugation of the resonating positive charge. It is important to mention that the same strategy does not imply repeating identical mutations as

those in hCRBPII because WT-hCRBPII and WT-CRABPII have only 38% sequence identity if aligned in BLAST (**Figure II-13**). The fact that a clear PSB peak was obtained in citric acid buffer at pH 5.0 led to all further wavelength regulation studies in this acidic buffer for CRABPII studies that were built with the new deeper binding pocket.

II.5. Wavelength regulation with CRABPII mutants bound to retinal under acidic conditions.

We began by enclosing the binding cavity to isolate the Rt within the CRABPII binding pocket from the bulk medium and thus enhance the effect of amino acid mutations in maximizing the bathochromic shift. To this end, Arg59, located at the mouth of the binding cavity, was changed to a number of different

Table II-3. Screening the influence of amino acids at the residue R59.

Entry	CRABPII mutant	λ_{max} (nm)	Protein Shift (nm)[a]
1	R111K:R132L:Y134F:T54V	534	0
2	R111K:R132L:Y134F:T54V:R59M	540	6
3	R111K:R132L:Y134F:T54V:R59L	545	11
4	R111K:R132L:Y134F:T54V:R59I	547	13
5	R111K:R132L:Y134F:T54V:R59T	546	12
6	R111K:R132L:Y134F:T54V:R59N	555	21
7	R111K:R132L:Y134F:T54V:R59D	555	21
8	R111K:R132L:Y134F:T54V:R59Q	551	17
9	R111K:R132L:Y134F:T54V:R59E	554	20
10	R111K:R132L:Y134F:T54V:R59F	564	30
11	R111K:R132L:Y134F:T54V:R59W	556	22
12	R111K:R132L:Y134F:T54V:R59Y	561	27

[a]Protein shift with reference to R111K:R132L:Y134F:T54V.

amino acids. Any amino acid in place of R59 caused the red-shift of the pigment as one would expect with removal of the positively charged arginine at the β-ionone ring region (**Table II-3**). The residues with small or hydrophobic side chains gave only 6-13 nm shift (entries 2-5) presumably because they cannot effectively encapsulate Rt. Placement of carboxylates or amides of both sizes (aspartate, asparagine, glutamate, glutamine) provided about the same 17-21 nm shifts (entries 6-9) pointing to the fact that a switch from the positive charge to the negative or partially negative in the β-ionone ring region slightly helped to promote charge delocalization along the polyene lowering the HOMO/LUMO gap of the retinylidene. As expected, the largest red-shifts, 22-30 nm, were observed with large hydrophobic residues (R59F, R59W and R59Y), because the large aromatic amino acids better shield the protein cavity from the bulk medium (entries 10-12). The latter observation was in agreement with our previous hypothesis derived from the data with hCRBPII mutants.
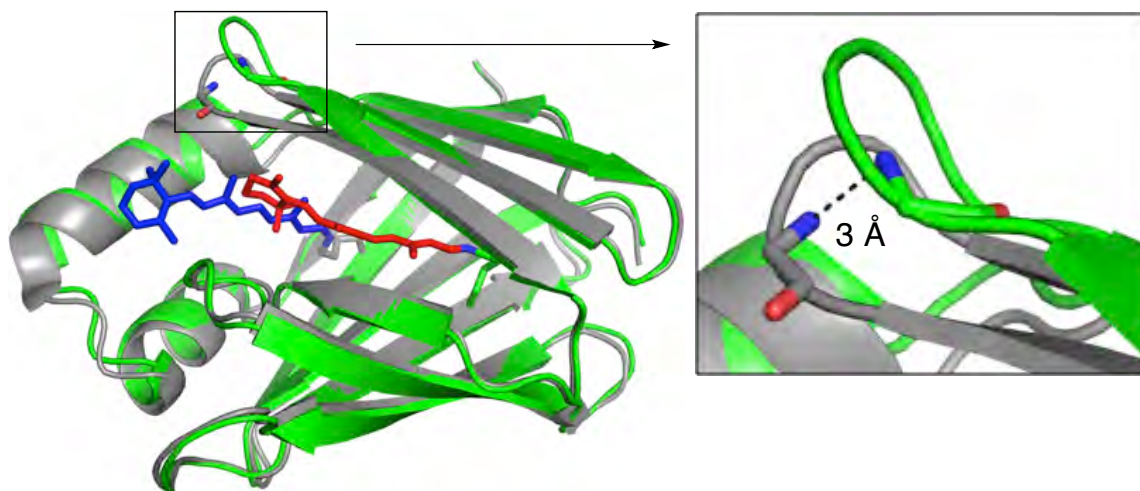
Figure II-14: Crystal structure of CRABPII-R111K:R132L:Y134F:T54V:R59W (protein green, retinal red) aligned with R132K:R111L:L121E (protein grey, retinal blue).

Mr. Meisam Nosrati (Professor Jim Geiger's lab, MSU) was able to obtain the crystal structures of CRABPII-R111K:R132L:Y134F:T54V:R59W bound with retinal at 2.6 Å resolution. Based on this crystal structure there are no doubts left about SB/PSB formation through R111K in the new binding site. Furthermore, having a crystal structure to plan the next mutations greatly facilitated further study of retinylidene wavelength regulation with CRABPII (**Figure II-14**). As shown in **Figure II-14**, the chromophore is located approximately 6 Å deeper inside the protein cavity relative to the first-generation CRABPII (R132K:R111L:L121E-CRABPII). The structure also illustrates that the loop containing amino acids S55–T60 has moved substantially (~3 Å), creating a more open, solvent-accessible binding cavity (**Figure II-14**). To further seal this portal, small residues around the opening were mutated into large aromatic amino acids providing an additional shield from the bulk medium. The 'leak' was identified at A32 and A36 positions that can potentially be mutated into larger hydrophobic residues (**Figure II-15a**). **Table II-4** summarizes the various combinations of mutation of these two amino acids and their effect on the wavelength of the resulting pigment. Replacing either of these alanines with tyrosine using the R111K:R132L:Y134F:T54V:R59Y template did not show any significant change in wavelength (entries 2 and 3). But replacing both of them simultaneously leads to an additional 13 nm bathochromic shift in R111K:R132L:Y134F:T54V:R59Y:A32Y:A36Y hepta mutant (entry 4). *In silico* modeling of this mutant predicted effective packing and encapsulation of the

bound retinal and as a result this mutant is red-shifted (**Figure II-15b**). Interestingly, A32W in combination with A36Y gives much less of a bathochromic shift than A32W alone: 7 nm versus 30 nm for respective hepta and hexa-mutants (entries 5 and 6). This comparison suggested that if too many bulky amino acids are placed at once it might lead to steric clashing which can either change the position of the chromophore or result in a less effective closure of the cavity.

In a parallel study, CRABPII pigments with A32 and A36 mutations added to the R111K:R132L:Y134F:T54V:R59W template showed even more pronounced effect of the chromophore encapsulation as one would expect for a more bulky R59W compared to R59Y. Replacing both alanines at once with two tryptophans or two tyrosines resulted in 19 nm and 23 nm red-shift, respectively (entries 8 and 9). However, in this case the most significant bathochromic shift, 33 nm and 54 nm, came from single point mutations A32Y and A32W, respectively (entries 10 and 11).
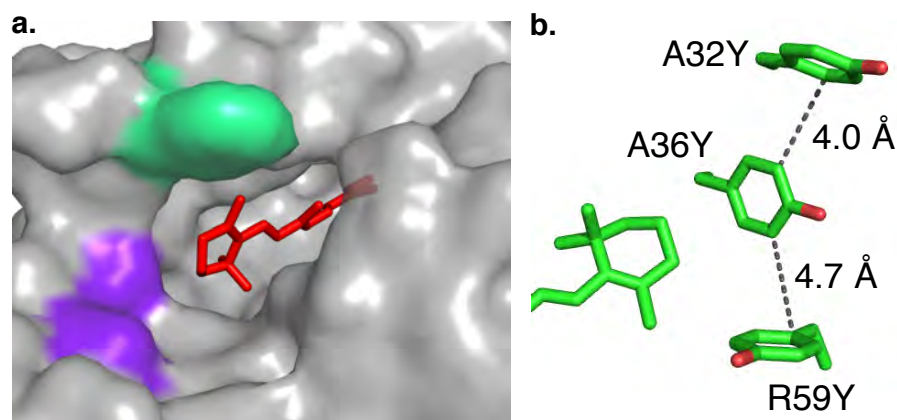
Figure II-15: Identifying the place to enclose the CRABPII cavity. **a.** Solvent exposed retinal out of the protein surface in R111K:R132L:Y134F:T54V:R59W mutant; Ala32 and Ala36 highlighted in purple; R59W highlighted in green. **b.** Model of A32Y and A36Y mutations in addition to R59Y.

The crystal structure of R111K:R132L:Y134F:T54V:R59W:A32W bound with Rt was determined at 2.6 Å resolution (Meisam Nosrati, James Geiger's Lab, MSU) and the chromophore exposure was compared to the R111K:R132L:Y134F:T54V:R59W penta mutant. The chromophore appears to occupy the same position in both mutants. **Figure II-16** shows that the protein

a.

b.
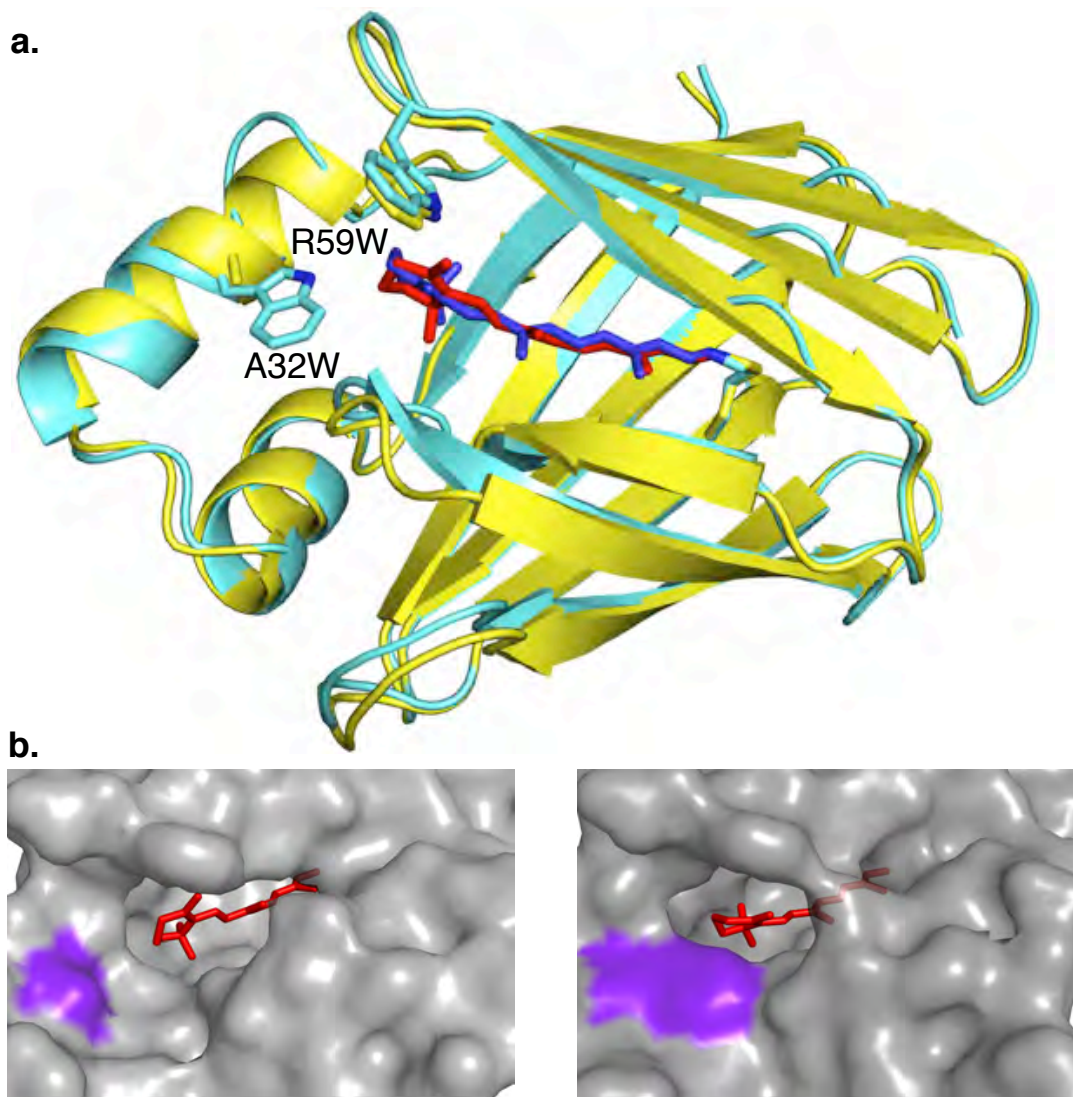
Figure II-16: Crystal structures before and after A32W mutation. **a.** Alignment of two CRABPII crystal strutures: R111K:R132L:Y134F:T54V:R59W (protein yellow, retinal red) and R111K:R132L:Y134F:T54V:R59W:A32W (protein cyan, retinal blue). **b.** Shielding retinylidene before (left) and after (right) A32W.

surface covers the Rt much more effectively in the presence of A32W. It clearly illustrates that both tryptophan residues contribute to encapsulation of the bound chromophore from the bulk medium, leading to the large observed bathochromic shift. The difference of 54 nm shift between R111K:R132L:Y134F:T54V:R59W and R111K:R132L:Y134F:T54V:R59W:A32W mutants is the most pronounced red-shift obtained by one point mutation during the wavelength regulation studies and it clearly indicates the importance of enclosing the protein binding cavity, as was hypothesized earlier with the hCRBPII mutants. It is important to note, that only the wavelength regulation previously probed with equivalent two mutations in hCRBPII (A33W and R58W) (Dr. Wenjing Wang, Babak Borhan's lab, MSU) was successfully translated into CRABPII mutants (A32W and R59W).

Table II-4. Influence of A32 and A36 mutations on binding with 0.5 equiv Rt.

Entry	CRABPII mutant	λ_{max} (nm)	Protein Shift (nm)
1	R111K:R132L:Y134F:T54V:R59Y	561	0[a]
2	R111K:R132L:Y134F:T54V:R59Y:A32Y	562	1
3	R111K:R132L:Y134F:T54V:R59Y:A36Y	563	2
4	R111K:R132L:Y134F:T54V:R59Y:A32Y:A36Y	574	13
5	R111K:R132L:Y134F:T54V:R59Y:A32W:A36Y	568	7
6	R111K:R132L:Y134F:T54V:R59Y:A32W	591	30
7	R111K:R132L:Y134F:T54V:R59W	556	0[b]
8	R111K:R132L:Y134F:T54V:R59W:A32W:A36W	575	19
9	R111K:R132L:Y134F:T54V:R59W:A32Y:A36Y	579	23
10	R111K:R132L:Y134F:T54V:R59W:A32Y	589	33
11	R111K:R132L:Y134F:T54V:R59W:A32W	610	54

[a]Protein shift with reference to R111K:R132L:Y134F:T54V:R59Y. [b]Protein shift with reference to R111K:R132L:Y134F:T54V:R59W.

The crystal structure of R111K:R132L:Y134F:T54V:R59W:A32W that absorbs at 610 nm revealed that the retinylidene could be covered even more effectively. In an attempt to address this issue a series of mutants aimed to sequester the chromophore from the solvent was generated using the R111K:R132L:Y134F:T54V:R59W:A32W template (**Table II-5**). Amino acids surrounding the entrance of the binding cavity were identified from the crystal

Table II-5. Screening amino acids capable of enclosing the binding cavity.

Entry	CRABPII mutant	λ_{max} (nm)	Protein Shift (nm)[a]
1	R111K:R132L:Y134F:T54V:R59W:A32W	610	0
2	R111K:R132L:Y134F:T54V:R59W:A32W:G78W	595	-15
3	R111K:R132L:Y134F:T54V:R59W:A32W:Q74W	607	-3
4	R111K:R132L:Y134F:T54V:R59W:A32W:V58W	598	-12
5	R111K:R132L:Y134F:T54V:R59W:A32W:P39W	603	-7
6	R111K:R132L:Y134F:T54V:R59W:A32W:A36W	575	-35
7	R111K:R132L:Y134F:T54V:R59W:A32W:A36T	605	-5
8	R111K:R132L:Y134F:T54V:R59W:A32W:A35W	610	0
9	R111K:R132L:Y134F:T54V:R59W:A32W:A35F	610	0
10	R111K:R132L:Y134F:T54V:R59W:A32W:T56W	532	-78
11	R111K:R132L:Y134F:T54V:R59W:A32W:T56Y	520	-90
12	R111K:R132L:Y134F:T54V:R59W:A32W:T56Q	590	-20
13	R111K:R132L:Y134F:T54V:R59W:A32W:V76W	608	-2
14	R111K:R132L:Y134F:T54V:R59W:A32W:V76F	610	0
15	R111K:R132L:Y134F:T54V:R59W:A32W:Q45F	601	-9
16	R111K:R132L:Y134F:T54V: ^{75}AA^{76}:R59W:A32W[b]		not isolated
17	R111K:R132L:Y134F:T54V:R59W:A32W:V24W		

[a]Protein shift with reference to R111K:R132L:Y134F:T54V:R59W:A32W.
[b]Between residues 75 and 76 two alanines were introduced.

89

structure (**Figure II-17**). Counter-intuitively, the large hydrophobic tryptophan with G78W, Q74W, V58W and P39W mutations gave moderate blue-shift of 3-15 nm (entries 2-5) and V24W placed between two α-helixes was not isolated (entry 17). If the effect of A36W would be compared to R111K:R132L:Y134F:T54V:R59W: A32W instead of R111K:R132L:Y134F:T54V:R59W (**Table II-5**) the observed 35 nm blue-shift is in agreement with the general trend in wavelength shift and presumably can be attributed to an increased amount of steric clashing between the residues that lead to opening of the protein cavity instead of enclosing it (**Table II-5**, entry 6). To probe if A36W opened the binding pocket due to its large size, the A36T mutation was introduced. Threonine is similar in size to alanine as compared to tryptophan, and therefore, the wavelength is expected to become comparable to parent hexa-mutant, R111K:R132L:Y134F:T54V:R59W:A32W. Indeed, hepta-mutant with A36T absorbed at 605 nm, 5 nm blue-shift, as opposed to the hepta-mutant with A36W, 35 nm blue-shift (**Table II-5**, entries 6 and 7).

Next the influence of the aromatic amino acids facing the protein surface was probed by tryptophan and phenylalanine mutations at A35. Hypothetically, if large aromatic amino acids play any other role besides encapsulation of the chromophore, it should be reflected in the wavelength of CRABPII pigments with A35F or A35W mutations because this residue is relatively close to the Rt. However, the A35 side chain points outside the binding cavity. As expected, both

A35W and A35F are pointing outside the chromophore and result in identical wavelengths as the CRABPII mutant with the original A35 (entries 8 and 9).

A puzzling result was obtained for T56 mutants. Aromatic residues aimed to close the binding pocket gave 78 nm and 90 nm blue-shift for T56W and T56Y, respectively, while only 20 nm blue-shift was observed for T56Q mutation (entries 10-12). According to the Pymol generated model, T56W and T56Y residues clash into the chromophore, while T56Q does not. The protein can accommodate such large residues and still restore Rt binding in two ways: (1) the chromophore moves from its regular location and changes its electrostatic surrounding; or (2) the β-ionone ring can rotate along C6-C7 bond to provide more space for the aromatic amino acid by adjusting the 6-s-*cis* conformation of Rt. The second
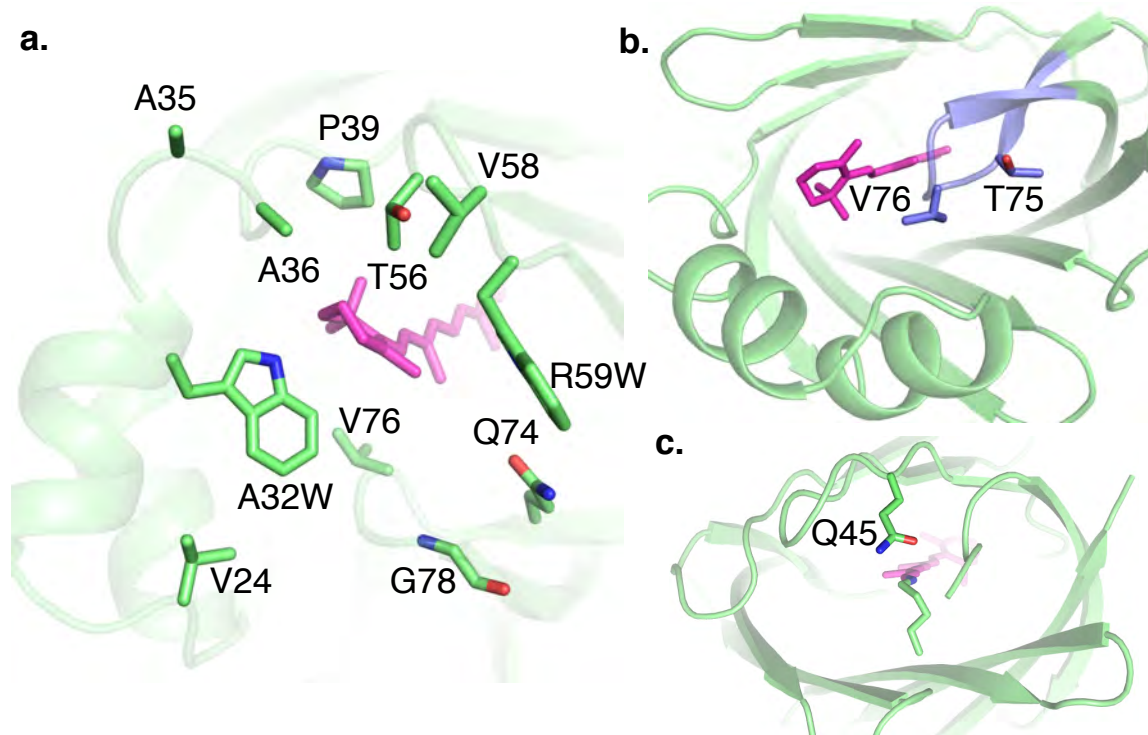
Figure II-17: The residues that potentially can encapsulate the chromophore in addition to R59W and A32W (retinal magenta). **a.** View from the β-ionone ring side. **b.** Location of the loop containing T75-V76 relative to the cut-off. **c.** View from the PSB side.

suggestion would be more logical since the previous work from Garavelli's lab provided computational evidence, supported with experimental results that such conformational change costs 80 nm difference between 6-s-*cis* and 6-s-*trans* forms.[14] Unfortunately, further analysis of the spectroscopic behavior for these mutants is difficult because of the absence of any crystal data.

One of the major structural differences between CRABPII and hCRBPII is the size of the loop containing amino acids T75:V76:D77:G78 (CRABPII numbering). In hCRBPII this loop contains two extra amino acids (lysine and serine), which naturally provides better Rt isolation (T74:**K75:S76**:L77:D78:N79) (**Figure II-13 and II-17b**). To address this issue V76 was mutated into tryptophan and phenylalanine, which did not influence the wavelength (entries 13 and 14). Furthermore, addition of two flexible amino acids between T75 and V76 was attempted to mimic hCRBPII loop. Two alanines were the residues of choice because alanine is a small and flexible residue. However, even addition of only two alanines between residues 75 and 76 (in Table II-5 referred as $^{75}AA^{76}$) did not give enough of soluble protein to characterize its binding with Rt (entry 16).

Finally, looking at the bound ligand from the other side of CRABPII (away from the two helices) identified a possibility to cap the protein from this end as well. An effort to close the cavity from this side was made with the Q45F mutation (**Figure II-17c**), which gave 9 nm blue-shift (entry 15).

Having in hand two crystal structures of CRABPII mutants, screening of other amino acids along the polyene chain and their influence on spectroscopic

Table II-6. Screening of amino acids along the polyene for red-shifted mutants.

Entry	CRABPII mutant	λ_{max} (nm)	Protein Shift (nm)
1	R111K:R132L:Y134F:T54V	534	0[a]
2	R111K:R132L:Y134F:T54V:F15Y	546	12
3	R111K:R132L:Y134F:T54V:L19W	556	22
4	R111K:R132L:Y134F:T54V:R59W	556	0[b]
5	R111K:R132L:Y134F:T54V:R59W:F15Y	557	1
6	R111K:R132L:Y134F:T54V:R59W:L19W	568	11
7	R111K:R132L:Y134F:T54V:R59Y:A32Y:A36Y	574	0[c]
8	R111K:R132L:Y134F:T54V:R59Y:A32Y:A36Y:L19W	574	0
9	R111K:R132L:Y134F:T54V:R59Y:A32Y:A36Y:M93L	578	4
10	R111K:R132L:Y134F:T54V:R59Y:A32Y:A36Y:M123L	578	4
11	R111K:R132L:Y134F:T54V:R59Y:A32Y:A36Y:E73A	586	12
12	R111K:R132L:Y134F:T54V:R59W:A32W	610	0[d]
13	R111K:R132L:Y134F:T54V:R59W:A32W:F15D	520	-90
14	R111K:R132L:Y134F:T54V:R59W:A32W:F15Y	595	-15
15	R111K:R132L:Y134F:T54V:R59W:A32W:L19W	588	-22
16	R111K:R132L:Y134F:T54V:R59W:A32W:M93L	616	6
17	R111K:R132L:Y134F:T54V:R59W:A32W:M93F	618	8
18	R111K:R132L:Y134F:T54V:R59W:A32W:M123L	614	4
19	R111K:R132L:Y134F:T54V:R59W:A32W:E73A	620	10
20	R111K:R132L:Y134F:T54V:R59W:A32W:K38D	610	0
21	R111K:R132L:Y134F:T54C:R59W:A32W	605	-5
22	R111K:R132L:Y134F:T54F:R59W:A32W	603	-7
23	R111K:R132Q:Y134F:T54V:R59W:A32W	618	8
24	R111K:R132E:Y134F:T54V:R59W:A32W	625	15
25	R111K:R132Y:Y134F:T54V:R59W:A32W	619	9

[a]Protein shift with reference to R111K:R132L:Y134F:T54V. [b]Protein shift with reference to R111K:R132L:Y134F:T54V:R59W. [c]Protein shift with reference to R111K:R132L:Y134F:T54V:R59Y:A32Y:A36Y. [d]Protein shift with reference to R111K:R132L:Y134F:T54V:R59W:A32W.

properties of the pigment was initiated. For this purpose four parent mutants were chosen as templates to study the effect of amino acids nearby the polyene chain (**Figure II-18**). At first, two amino acids F15 and L19 were mutated into tyrosine and tryptophan, respectively, mainly because of availability of these primers in the library among screened mutations for the wavelength regulation studies with the first generation CRABPII mutants (in addition to all R59 mutants described above). In the shortest parent mutant R111K:R132L:Y134F:T54V mutations F15Y and L19W individually caused a red-shift, 12 and 22 nm, respectively (**Table II-6**, entries 1-3). However, if added to the R111K:R132L:Y134F:T54V:R59W template, the shift was less significant (**Table II-6**, entries 4-6). Disappointingly, the same two mutations did not have any influence on wavelength in the R111K:R132L:Y134F:T54V:R59Y:A32Y:A36Y series (**Table II-6**, entries 7-9) and actually caused a blue-shift in the R111K:R132L:Y134F:T54V:R59W:A32W series (**Table II-6**, entries 13-15). Probably in the shorter series these two mutations act as a cavity cap to some extend. Since no red-shift in the longer series was observed, the spectroscopic investigation for combination of these mutations was omitted.

Next, inspiration for further red-shift was gained from the hCRBPII crystal structure alignment with CRABPII. Superimposed structures of hCRBPII-Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4A mutant (λ_{max}=613 nm)[12] and CRABPII-R111K:R132L:Y134F:T54V:R59W mutant identified residues L93 and L119 in hCRBPII equivalent to M93 and M123 in CRABPII (**Figure II-19**). Both

M93L and M123L caused a 4 nm red-shift in R111K:R132L:Y134F:T54V:R59Y:A32Y:A36Y series (**Table II-6**, entries 10 and 11) and 6 nm and 4 nm, respectively, in the R111K:R132L:Y134F:T54V:R59W:A32W series (**Table II-6**, entries 16 and 18). The reason is not exactly clear since both methionine and leucine are hydrophobic amino acids, but M93L and M123L in the CRABPII binding pocket led to additional bathochromic shift. The test for increased hydrophobic packing was made with the even larger M93F mutation, which caused even slightly larger red-shift (**Table II-6**, entry 17).

The next mutation, E73A, arose from the hypothesis about interactions between W109 and E73[6] (**Figure II-5d**) near the iminium when retinal is bound through R111K (**Figure II-18**). Being 7.2 Å away from the PSB, E73 has hydrogen bonding interactions with indole nitrogen of W109, which positions
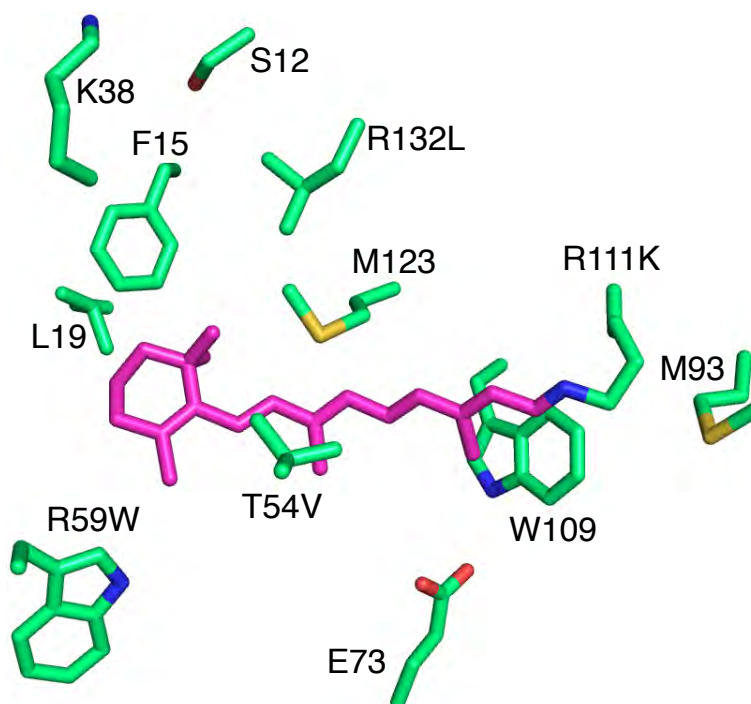
Figure II-18: The residues around Rt described in **Table II-5**.

W109 in the geometry suitable for π-cation interaction with PSB. E73A mutation might remove the negative electrostatic potential projected from the π-cloud on the PSB region and as a result, may cause the red-shift of the complex. To test the influence of such interaction mutating W109 would be ideal. However, W109 mutants (W109F, W109Y, W109L, W109Q and W109A) were not expressed as soluble proteins. The next approach would be to make W109 less polarized through mutagenesis at E73, which is placed more that 7 Å from the chromophore and more than 9 Å from the PSB region. If E73 does not polarize W109, the removal of the carboxylate should not have any effect on the wavelength since it is too far away. The E73A mutation resulted in a bathochromic shift (**Table II-6**, entries 11 and 19) as compared to the corresponding proteins with E73 (**Table II-6**, entries 7 and 12). This result supports the theory that polarized tryptophan 109 affects the wavelength. To prove that there is no effect on spectroscopic properties from amino acids placed more than 7 Å away from the chromophore, K38D mutant was prepared. The K38D mutation, which represents the switch in polarity about 7 Å away from the β-ionone ring, did not change the protein absorption (**Table II-6**, entry 20).
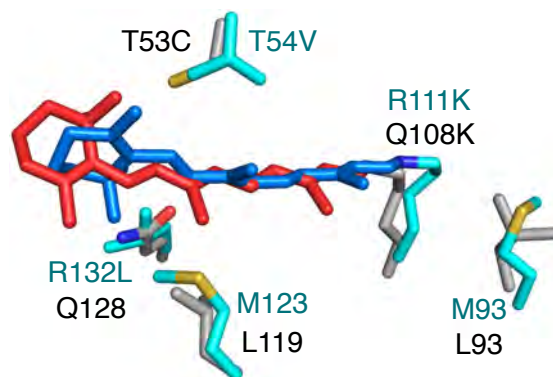
Figure II-19: Alignment of CRABPII (cyan, Rt blue) and hCRBPII (grey, Rt red) bound to Rt as PSB.

Since recreating the hCRBPII environment within the CRABPII cavity (M93L, M123L) resulted in a bathochromic shift, other hCRBPII mimicking residues were also introduced (**Figure II-19**). However, the results were not always consistent between the two proteins. For example, the T53C mutation consistently resulted in a red-shift within the hCRBPII series. When the equivalent mutation (T54C) was applied to the CRABPII mutants, a 5 nm blue-shift was observed (**Table II-6**, entry 21). Residue T54 is in the middle of the polyene chain and mutation in this region of Rt-PSB were shown important for spectral tuning. For instance, W265Y mutation in bovine rhodopsin leads to a 10 nm blue-shift.[21] To further red-shifted CRABPII pigments, the T54F mutation was introduced, placing a phenylalanine in the middle of the polyene. However, the pigment blue-shifted 8 nm (**Table II-6**, entry 22).

Next, we reevaluated the importance of R132L mutations, located 5.5 Å away from the middle of polyene. During CRABPII reengineering the strategy was to mutate R132 into a hydrophobic leucine to allow the Rt to slip down the binding cavity towards R111K. Comparison of hCRBPII-Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4A and CRABPII-R111K:R132L:Y134F:T54V:R59W mutants (**Figure II-19**) showed that equivalent residue to R132L in CRABPII was Q128 in hCRBPII. R132Q mutation caused a red-shift as compared to R132L. However, the shift was not as dramatic as that in the hCRBPII case, where Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q128 absorbs at 590 nm while Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q128L

97

Table II-7. Combining the amino acids for super-red CRABPII pigments.

Entry	CRABPII mutant	λ_{max} (nm)
1	R111K:R132L:Y134F:T54V:R59W:A32W:E73A	620
2	R111K:R132L:Y134F:T54V:R59W:A32W:E73A:M93L	626
3	R111K:R132Q:Y134F:T54V:R59W:A32W:E73A:M93L	630
4	R111K:R132L:Y134F:T54V:R59W:A32W:E73A:M93L:S12D	640
5	R111K:R132L:Y134F:T54V:R59W:A32W:E73A:M93L:S12D:M123L	641
6	R111K:R132E:Y134F:T54V:R59W:A32W:E73A:M93L:S12D	640
7	R111K:R132Q:Y134F:T54V:R59W:A32W:E73A:M93L:S12D	640

has λ_{max} = 532 nm (Dr. Wenjing Wang PhD thesis). Moreover, as opposed to the

hCRBPII case, where Q128E mutant blue-shifts 20 nm (Q108K:K40L:T51V:T53C:R58W:Q128, 585 nm *vs* Q108K:K40L:T51V:T53C: R58W:Q128E, 555 nm) (Dr. Wenjing Wang PhD thesis), R132E mutation in CRABPII produced a 7 nm red-shift when compared to R132Q and a 15 nm red-shift compared to R132L (**Table II-6**, entry 24). The aromatic R132Y mutation caused almost the same effect as R132Q (**Table II-6**, entry 25).

Finally, the residues providing the red-shifted variants in the different series were combined for the final push towards super-red CRABPII pigments. Combination of E73A with M93L and R132Q gave a mutant with λ_{max}=630 nm in citric acid buffer at pH 5.0 (**Table II-7**, entries 1-3). Interestingly, the most red-shifted CRABPII mutant was recorded with the S12D mutation placed about 9 Å from the β-ionone ring region (**Table II-7**, entry 4). However, further addition of

the residues previously contributing to the red-shift did not show any additive effect to surpass 640 nm (**Table II-7**, entries 5-7).

In general, there was no enhancement effect of red-shifting mutations after enclosing the CRABPII cavity with Rt, opposite to what occurred with the previously reported hCRBPII pigments. Moreover, most mutations previously reported to contribute to the wavelength shift were found to bring an opposite effect on the spectroscopic properties: (a) mutations F16Y, R58E, E72L, Q128E all resulted in blue-shift when applied to hCRBPII while the equivalent mutations F15Y (only in short series), R59E, E73A and R132E caused a red-shift in CRABPII series; (b) T53C and T54C had exactly opposite effect in the two systems; (c) as compared to the highly red-shifting Q4F and T51V mutations in hCRBPII, CRABPII has already F3 and I52 as part of its natural sequence. Overall, only R59W and A32W showed the same tendency in the wavelength regulation of CRABPII as R58W and A33W in hCRBPII, with A32W contributing much more than A33W (54 nm *vs* 22 nm, respectively). The wavelength shift is more pronounced in CRABPII because naturally the CRABPII cavity is more open than that of hCRBPII. The next step towards super-red-shifted pigments follows the same strategy in CRABPII as in hCRBPII: overall neutral polarity needs to be projected on the surface of chromophore.[15] Although the strategy is the same, the actual mutations leading to the red-shift were different.

On the other hand, if a blue-shift is desired one could achieve it in the reverse manner by localizing more negative electrostatic potential close to the

PSB. The surrounding of PSB consists of four hydrophobic amino acids: V41, I52, W109 and L121. Three out of these four residues are structurally crucial for a good protein expression (I52 and W109 mutants from the studies above, V41 mutants from Dr. King Sing S. Lee studies). Additionally, based on previous work

Table II-8. Screening L121 mutants with anticipated blue-shift.

Entry	CRABPII mutant	λ_{max} (nm)	Protein Shift (nm)
1	R111K:R132L:Y134F:T54V	534	0[a]
2	R111K:R132L:Y134F:T54V:L121D	474	-60
3	R111K:R132L:Y134F:T54V:L121N	482	-52
4	R111K:R132L:Y134F:T54V:L121Q	492	-42
5	R111K:R132L:Y134F:T54V:L121Y	510	-24
6	R111K:R132L:Y134F:T54V:R59W	556	0[b]
7	R111K:R132L:Y134F:T54V:R59W:L121Q	510	-46
8	R111K:R132L:Y134F:T54V:R59W:L121Y	504	-52
9	R111K:R132L:Y134F:T54V:R59Y:A32W	591	0[c]
10	R111K:R132L:Y134F:T54V:R59Y:A32W:L121Q	538	-53
11	R111K:R132L:Y134F:T54V:R59Y:A32W:L121Y	529	-62
12	R111K:R132L:Y134F:T54V:R59W:A32W	610	0[d]
13	R111K:R132L:Y134F:T54V:R59W:A32W:L121Q	545	-65
14	R111K:R132L:Y134F:T54V:R59W:A32W:L121Y	541	-69
15	R111K:R132L:Y134F:T54V:R59W:A32W:E73A:M93L	626	0[e]
16	R111K:R132L:Y134F:T54V:R59W:A32W:E73A:M93L:L121Q	560	-66
17	R111K:R132L:Y134F:T54V:R59W:A32W:E73A:M93L:L121Y	530	-96

[a]Protein shift with reference to R111K:R132L:Y134F:T54V. [b]Protein shift with reference to R111K:R132L:Y134F:T54V:R59W. [c]Protein shift with reference to R111K:R132L:Y134F:T54V:R59Y:A32W. [d]Protein shift with reference to R111K:R132L:Y134F:T54V:R59W:A32W. [e]Protein shift with reference to R111K:R132L:Y134F:T54V:R59W:A32W:E73A:M93L.

with CRABPII, L121 (4 Å from PSB) can be easily mutated into a number of amino acids of various polarities. Therefore, the main way to get to blue-shifted mutants is to replace L121. At first L121 was mutated into glutamate (L121E) because in the first generation CRABPII mutants same L121E played the role of the counterion and was the most blue-shifted pigment one as compared to shorter aspartate (L121D) or less polar glutamine (L121Q). Upon the L121E mutation a clear shift was observed from 400 nm to 365 nm, indicative of retinal-SB formation (**Figure II-20**). Unexpectedly, further acidification of R111K:R132L:Y134F:T54V:L121E/Rt complex to lower pH gave no indicative PSB peak in UV-vis suggesting an unusually low pK_a of the iminium. That is especially strange since one would imagine that a carboxylate in the PSB region would act as a counteranion.[15, 22-23] However, for a counterion to be effective, it should be positioned in an appropriate geometry to create an interaction with the proton of the iminium. Presumably, L121E mutation does not provide such
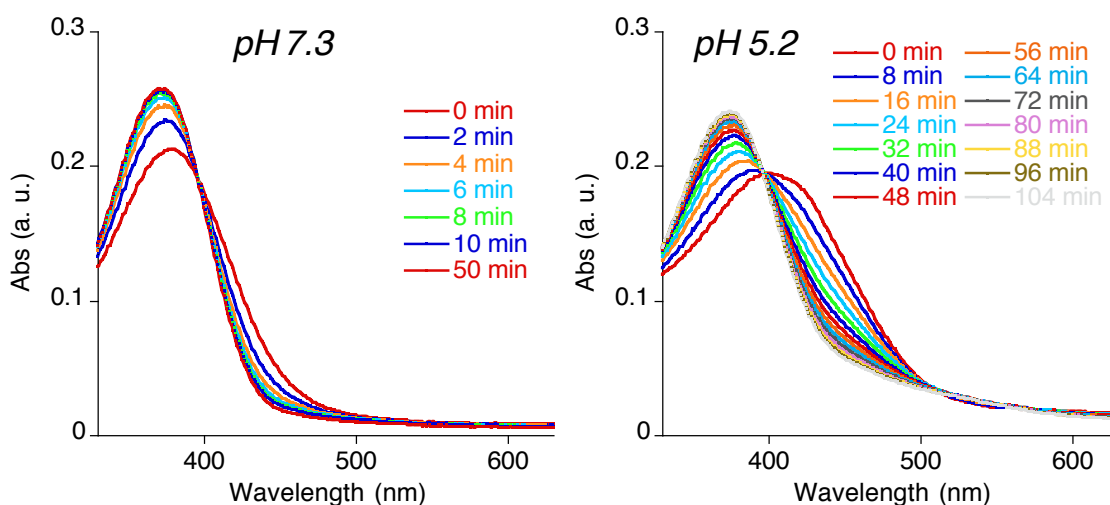
Figure II-20: UV-vis of R111K:R132L:Y134F:T54V:L121E mutant binding with Rt over time at two pH values. Neither pH gave rise to an observable PSB but clear SB formation was seen.

interaction leading to SB formation only. Nevertheless, the fast imine formation between the polyenic aldehyde and the amine of lysine was noted at this point (**Figure II-20**) and carried on as an inspiration for fast PSB-forming protein design essential for successful site-specific labeling, as described in Chapter 4.

Mutations L121D, L121N, L121Q and L121Y near the iminium nitrogen all gave detectable blue-shifted PSB peaks, though L121D mutant required acidification to pH 3 for λ_{max} detection. As one would expect L121D carboxylate caused the strongest blue-shift (60 nm), whereas the amide resulted in smaller protein shifts (L121N and L121Q, 52-42 nm) and the aromatic hydroxyl even smaller (L121Y, 24 nm) (**Table II-8**, entries 1-5). The last two amino acids L121Q and L121Y where screened in the different series of CRABPII mutants to evaluate if the contributions from these polarity changes are consistent among the pigments. Notably, the blue-shift resulting from the L121Q consistently increased in mutants of longer series (**Table II-8**, entries 4, 7, 10, 13 and 16). Interestingly, L121Y showed even stronger hypsochromic enhancement effect
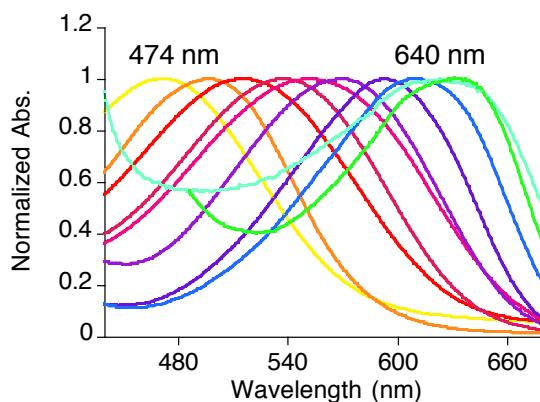
Figure II-21: Overlaid UV-vis of CRABPII mutants bound to Rt. All spectra were acquired at pH 5.2 and normalized to 1.

than L121Q (**Table II-8**, entries 5, 8, 11, 14 and 17).

This enhancement effect is more pronounced in the presence of the residues that enclose the cavity better. Although the enhancement effect was observed for hCRBPII mutants with enclosed chromophore, the spectral shift of enhancement contradicts the data collected for the two systems: enhancement effect in hCRBPII pigments was recorded for the red-shifting mutations.

The normalized UV-vis spectra of some CRABPII mutants bound to the all-*trans*-retinal depict the overall range of absorption for a single chromophore dictated by the protein environment (**Figure II-21**). Although the range 474-640 nm for CRABPII pigments is shorter than previously collected for hCRBPII (425-644 nm), the general trend of the wavelength tuning can be recapitulated from one system into another.

II.6. pH-dependent properties of CRABPII mutants.

The initial goal of CRABPII second generation mutants was to demonstrate wavelength regulation in a protein system that had failed to regulate the wavelength of a bound retinylidene before. This mission was successfully accomplished following principles derived from studies using hCRBPII, with one significant difference between two systems: in general CRABPII mutants had remarkably lower pK_a values (range between 1.8 and 8.1) as compared to hCRBPII mutants (range between 6.5 and 11.2).[12] In addition, the presence of a counteranion (L121E and L121D) was expected to result in hypsochromic shift while also stabilizing the PSB.[8] However, upon addition of these carboxylates the

pKa of the formed PSB was significantly depressed. This drastically different feature of CRABPII second generation mutants as compared to all the rest of the protein systems we engineered so far was used to design a novel protein-based pH-indicator. For this purpose it was crucial to understand the pK_a regulation of the iminium inside the CRABPII mutants and design the experiments where the protein-chromophore pigments could function as reporters of acidity. We therefore set out to determine factors that drive the pK_a changes of CRABPII mutants used for wavelength regulation studies and generate pigments that would respond and change color in the physiologically relevant pH regime.

The pK_a of the retinylidene would dictate the amount of SB and PSB observed at any specific pH value. The protonated form of the retinylidene (PSB) has a broad range of light absorptivity ($\lambda_{max} \geq 420$ nm) dictating the observed color, while the deprotonated form (SB) absorbs consistently at the same λ_{max} outside the visible spectrum (~365 nm) (**Figure II-23**).[24] This property is used as
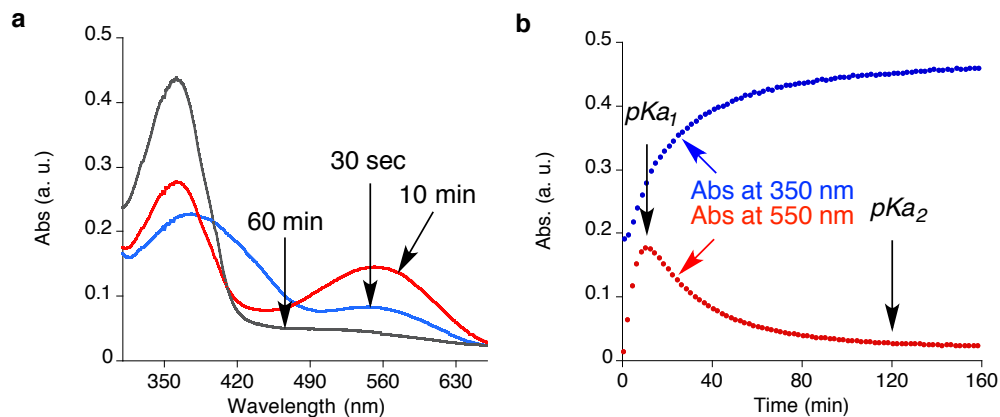
Figure II-22: Time-dependent PSB formation followed by deprotonation. **a**. UV-vis at different time points of experiment. **b**. Plotting SB and PSB in time indicating points of two states.

a rationale for a colorimetric proton sensor design. Moreover, the retinylidene protonation/deprotonation event is utilized by nature in bacteriorhodopsin proton pumps, where isomerization from all-*trans*- to a 13-*cis*-retinylidene double bond triggers the water mediated proton movement from Asp96 to Asp85.[25-29] For this process to work, the iminium pK_a should be higher than the physiological pH in order to provide pH-independent and exclusively light driven proton movement across the membrane.[22] Indeed, while the retilylidene pK_a in solution is 7.2 (Rt + n-butylamine in methanol/water),[30] this value in rhodopsins rises higher than 9 and in the visual pigments it surpasses 16.[23, 26, 28, 30-35] The protein environment dictates these remarkably high pK_a values through different mechanism: (1) iminium needs a properly placed counteranion or long-range interactions with a counteranion mediated by water or polar amino acids; (2) the orientation is important for the interacting polar residues that stabilize the iminium. For the purpose of pH-sensor design, we envisioned a protein–iminium system that is already devoid of a stabilizing counteranion. That, in combination with the increased hydrophobicity in the PSB region should lead to a decrease of the retinylidene pK_a. Since the physiologically relevant pH regime is between pH 4 and 8, the reporters of acidity should function in this pH zone.

PSB, λ_{max} >420 nm   SB, λ_{max} ~360 nm
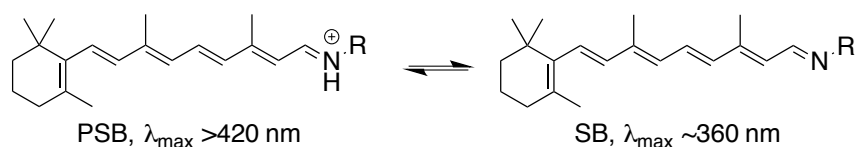
Figure II-23: Iminium deprotonation and respective λ_{max} of two states: PSB vs SB.

105

It is important to note that a lot of mutants described in the previous section exhibit gradual PSB loss and SB increase over time at pH of 5.2 as depicted for R111K:R132L:Y134F:T54V:R59W in **Figure II-22**. Although an interesting property by itself, this time-dependent deprotonation event is unacceptable for a pH-sensor, since the initially colored complexes with PSB (initial form with pK_{a1}) loses color because of some unknown secondary acid-base reaction and became non-colored (time-matured form with pK_{a2}). To avoid complication, while developing a pH-sensor, we chose only CRABPII/Rt complexes with steady pK_a values over time. Overall it appears that the PSB stability depends on mutation at position L121 and A32. All of such mutants are summarized in **Table II-9**.

At a first glance there is no reasonable correlation between the overall polarity of the binding pocket and the observed changes in the pK_a of the iminium because the presence of aspartate or glutamate residues anywhere in the protein binding cavity seem to decrease pK_a values, although one might had expected a reverse effect (**Table II-9**, entries 1, 5, 9, 12 and 22). However, among entries 17–21 there is a reasonable trend with lower pK_a values correlating with the increased hydrophobicity of the residues. For example, the change from glutamine to tyrosine at residue 132 (~5.5 Å from polyene) drops the iminium pK_a by 0.6 units and from tyrosine to leucine by another 0.5 (**Table II-9**, entries 17-

106

Table II-9. pK$_a$ values for CRABPII mutants.

Entry	CRABPII mutant	pK$_a$	λ$_{max}$ (nm)[a]
1	R111K:R132L:Y134F:T54V:L121D	3.0	474
2	R111K:R132L:Y134F:T54V:L121N	4.2	482
3	R111K:R132L:Y134F:T54V:L121Q	5.3	492
4	R111K:R132L:Y134F:T54V:L121Y	4.8/2.6[b]	500
5	R111K:R132L:Y134F:T54V:L121D:R59W	3.9	480
6	R111K:R132L:Y134F:T54V:L121Q:R59W	4.4	505
7	R111K:R132L:Y134F:T54V:L121Y:R59W	2.8	504
8	R111K:R132L:Y134F:T54V:R59W	6.6/2.5[c]	543
9	R111K:R132L:Y134F:T54V:L121D:R59Y:A32Y:A36Y	10.1/2.8[b]	470
10	R111K:R132L:Y134F:T54V:L121Q:R59Y:A32Y:A36Y	5.7	510
11	R111K:R132L:Y134F:T54V:L121Q:R59W:A32Y:A36Y	5.6	510
12	R111K:R132L:Y134F:T54V:L121D:R59W:A32W	2.9	480
13	R111K:R132L:Y134F:T54V:L121Q:R59W:A32W	5.8	545
14	R111K:R132L:Y134F:T54V:L121Q:R59Y:A32W	5.9	538
15	R111K:R132L:Y134F:T54V:L121Y:R59W:A32W	5.1/2.5[b]	530
16	R111K:R132L:Y134F:T54V:L121Y:R59Y:A32W	5.1	529
17	R111K:R132Q:Y134F:T54V:R59W:A32W	8.1	618
18	R111K:R132Y:Y134F:T54V:R59W:A32W	7.5	619
19	R111K:R132L:Y134F:T54V:R59W:A32W	7.0	610
20	R111K:R132L:Y134F:T54V:R59W:A32W:E73A	2.4	606
21	R111K:R132Q:Y134F:T54V:R59W:A32W:M93L:E73A	2.6	607
22	R111K:R132E:Y134F:T54V:R59W:A32W:M93L:E73A:S12D	1.8	610
23	R111K:R132L:Y134F:T54V:L121Q:R59W:A32W:M93L:E73A	4.1	538
24	R111K:R132L:Y134F:T54V:L121Q:R59W:A32W:M93L:E73A:S12D	4.1	560
25	R111K:R132L:Y134F:T54V:L121Y:R59W:A32W:M93L:E73A	3.0	530

[a]λ$_{max}$ correspond to the spectra at the most acidic form during pH titration (pH is 1-2 units below pK$_a$). [b]Both pK$_a$ values come from single acid-base titration suggesting the presence of species with two pK$_a$ values simultaneously. [c]Two pK$_a$ values correspond to two time-dependent forms: one after 10 min and other after 2 h of incubation with Rt.

19). The single mutation E73A, ~9 Å away from the iminium nitrogen drops the pK$_a$ by 4.4 units (entry 20) suggesting that overall polarity rather than direct counterion interaction contribute to the acidity of the PSB.

However, the rest of the mutants that gave a large working range of pK$_a$ values were difficult to rationalize solely based on changes in polarity of the binding pocket. Besides the counter-intuitive impact of carboxylates L121D,
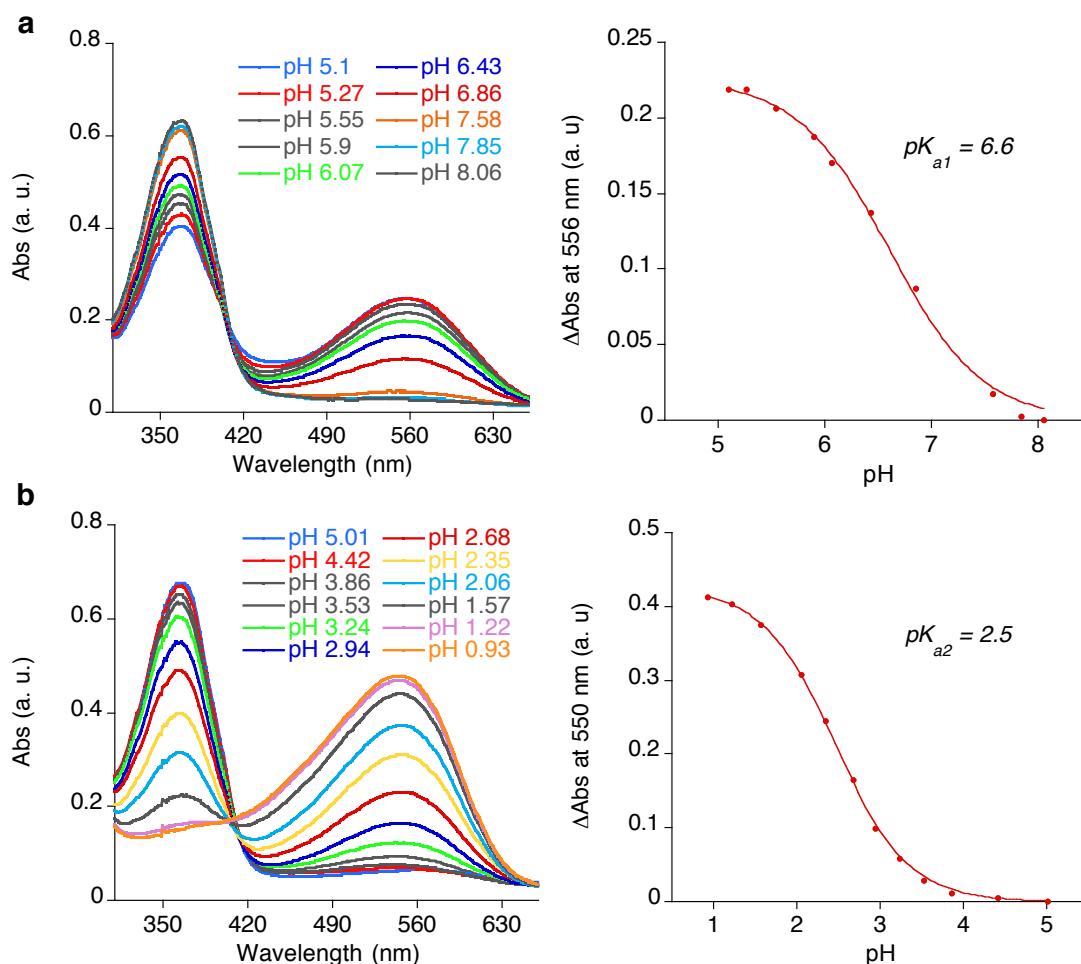
Figure II-24: Two pH titrations of the same R111K:R132L:Y134F:T54V:R59W mutant after **a.** 10 min and **b.** 2 h of incubation with retinal at pH 5. Two states of CRABPII/Rt complex give two different pK$_a$ values for kinetic and thermodynamic product.

R132E, S12D, there was no general trend between pK_a values of L121Q and L121Y mutations. For instance, from entry 3 to entry 6 the pK_a drops, as one would expect for a less polar environment (R59W). However from entry 6 to entries 10 and 14 it increases, although after the A32Y:A36Y and A32W mutations the overall chromophore surrounding is more electrostatically neutral.

The latter observations, namely, no general trend for the retinylidene pK_a changes among most mutants with a physical parameter, remained puzzling until a crystal structure for mutant R111K:R132L:Y134F:T54V:R59W (entry 8) was obtained. Out of all the mutants with time-dependent PSB loss, this mutant behaved well during pH-titrations of both form, enabling the determination of pKa_1 and pKa_2. Upon incubation with retinal, the PSB was fully formed within 10 min (λ_{max} = 556 nm), with a measured pK_{a1} of 6.6 (**Figure II-24a**). The time-dependent reduction in the intensity of PSB correlated well with the increasing
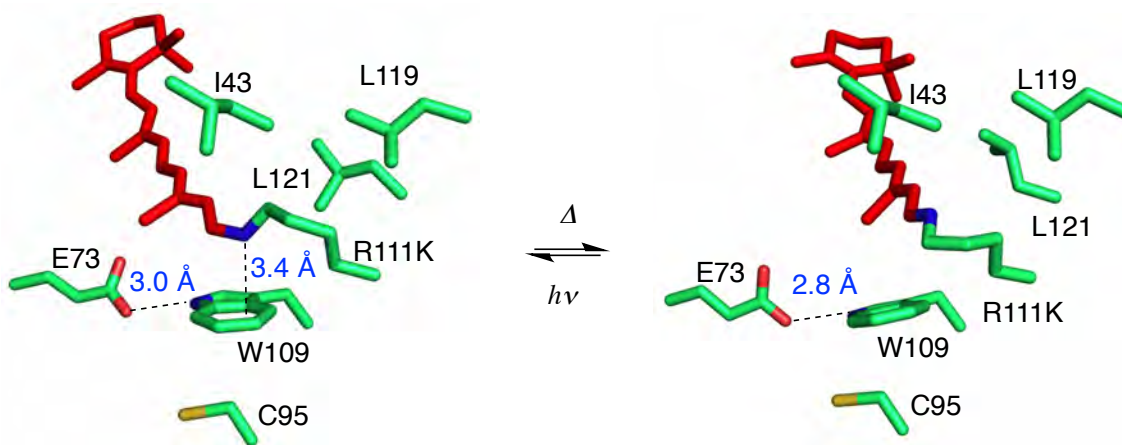
Figure II-25: Two proposed inter-convertible states of kinetic and thermodynamic complex forms based on energy minimized model (top) and crystal structure (bottom) of the same R111K:R132L:Y134F:T54V:R59W mutant.

amount of SB (365 nm), and no complex degradation or free form of Rt (380 nm) was observed (**Figure II-22b**). Within 2 h the conversion was complete, with no apparent PSB present based on UV–vis analysis. Acidification of the mature protein complex did recover the PSB absorption (543 nm), clearly indicating that the protein was not denatured (**Figure II-24b**). Although the wavelength had blue-shifted slightly (13 nm), it does indicate that the overall polarity of the binding pocket had not changed dramatically. Interestingly, titration of the time-matured protein complex gives a pK_{a2} value of 2.5 (entry 8), over four units lower than the initial pK_{a1}. Since the pK_{a2} is lower than the pH of the solution, for time-dependent UV-vis measurements (pH = 5.2) it explains the gradual loss of the PSB peak as depicted in **Figure II-22**.

The slow change in the PSB of R111K:R132L:Y134F:T54V:R59W/Rt complex can only result from local changes in protein environment that destabilized the iminium. The acid-base reactions are fast, and thus the deprotonation event proceeding over 2 h is a result of changes to the chromophore/protein interaction. This is either a result of changes in the structure of the protein, the conformation of the chromophore, or both. Most likely, since the greatest effect is seen in the pK_a of the iminium, this effect is closely related to the interaction of the PSB with its protein environment. This could arise from the geometric isomerization of the iminium (*cis* to *trans*), presumably from an initially observed kinetic product to a more thermodynamically stable state, that

110

greatly alters the interaction of the iminium nitrogen with its surroundings (**Figure II-26**). Since the crystallographic data obtained for this mutant was generated from a crystal growing for about a week, the structure would correspond to the thermodynamically stable state with an apparent pK_{a2}.

Comparison of the iminium region in two CRABPII crystal structures suggested a plausible explanation for the two forms of the mutant. The data for R111K:R132L:Y134F:T54V:R59W bound to Rt exhibit electron density of sufficient quality to enable assignment of the iminium as *trans*. The immediate environment of the imine is lined with hydrophobic residues (Ile43, Leu119, and Leu121), thus leading to the low pK_a (**Figure II-25**). On the other hand, the crystal structure of R111K:R132L:Y134F:T54V:R59W:A32W hexa-mutant with constant high pK_a over time (7.0) shows large electron density in iminium region where either *cis* or *trans* double bond can be built with equal certainty suggesting
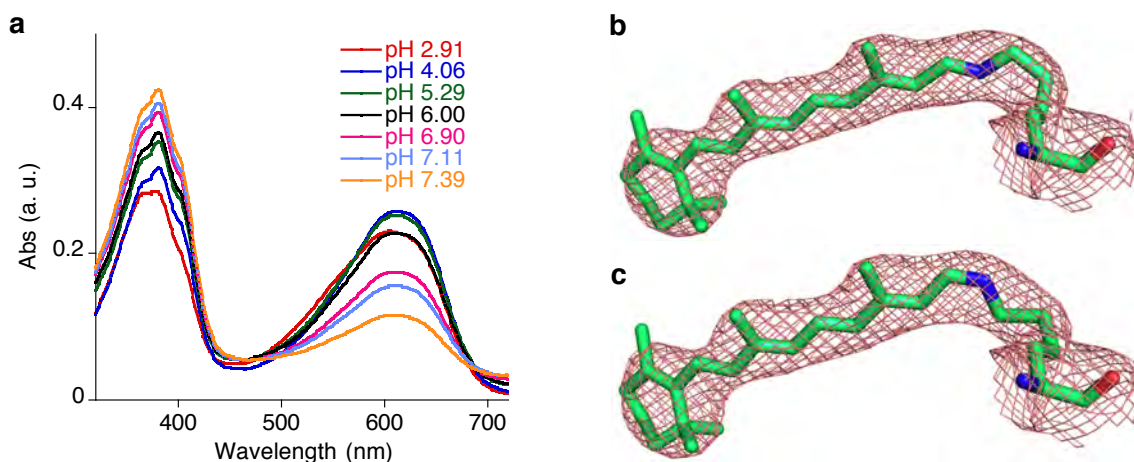
Figure II-26: Data supporting two isomeric iminium forms of R111K:R132L:Y134F:T54V:R59W:A32W mutant by **a.** incomplete PSB formation regardless of pH and by electron density maps refined in **b.** *trans*- and **c.** *cis*-iminium forms.

a mixture of both forms (**Figure II-26b** and **II-26c**). The last suggestion is also supported by the acid titration data with incomplete conversion from SB to PSB even at a very acidic pH (**Figure II-26a**). The putative *cis* iminium of both mutants would orient the proton in a position that would allow for π-cation stabilization via Trp109 (**Figure II-25** and **II-26b**) and lead to higher pK_a of 6.6 for the kinetic form of penta and 7.0 for the hexa-mutant. In fact, this interaction has been indicated previously as a dominant factor in stabilizing the *cis*-iminium with hCRBPII mutants. In such a scenario, the hypothetical kinetic product of R111K:R132L:Y134F:T54V:R59W penta mutant, the *cis*-iminium, when interacting with Trp109, would lose the cationic stabilization upon isomerization to the thermodynamically more stable *trans*-imine. If true, the isomerization reaction should be temperature and light dependent (**Figure II-25**). The secondary deprotonation event was monitored by UV-vis cycles over time as the measurement of putative *cis*-iminium/*trans*-imine isomerization reaction at four different temperatures (4 °C, 16 °C, 23 °C and 37 °C) and, as expected, the reaction slowed down at colder temperatures and was accelerated with heat (**Figure II-27**). On the other hand, UV light irradiation (band pass filter 340–380 nm, SB peak ~365 nm) for 5 min of the time-matured complex (presumably *trans*-imine) recovered the PSB peak. Conversely, exposure to green light (band pass filter 500–600 nm, PSB peak ~550 nm) of the UV-recovered PSB (presumably *cis*-iminium) gave conversion to SB within 12 min as compared to thermal reaction that took 2 h. Remarkably, this seesaw-like motion showed reversibility

in up to 7 cycles (**Figure II-28**). Such an unexpected light controlled SB/PSB interconversion is a new discovery in protein photochemistry and this property has high potential to find a utility as a photoswitchable protein quencher or in artificial proton pump design. All together these results (temperature and light dependent inter-conversion of SB and PSB and crystallographic data of two mutants) support the imine isomerization hypothesis. However, more conclusive support of imine isomerization theory would come from crystallographic data for two separate forms of the same mutant, which is now under detailed investigation
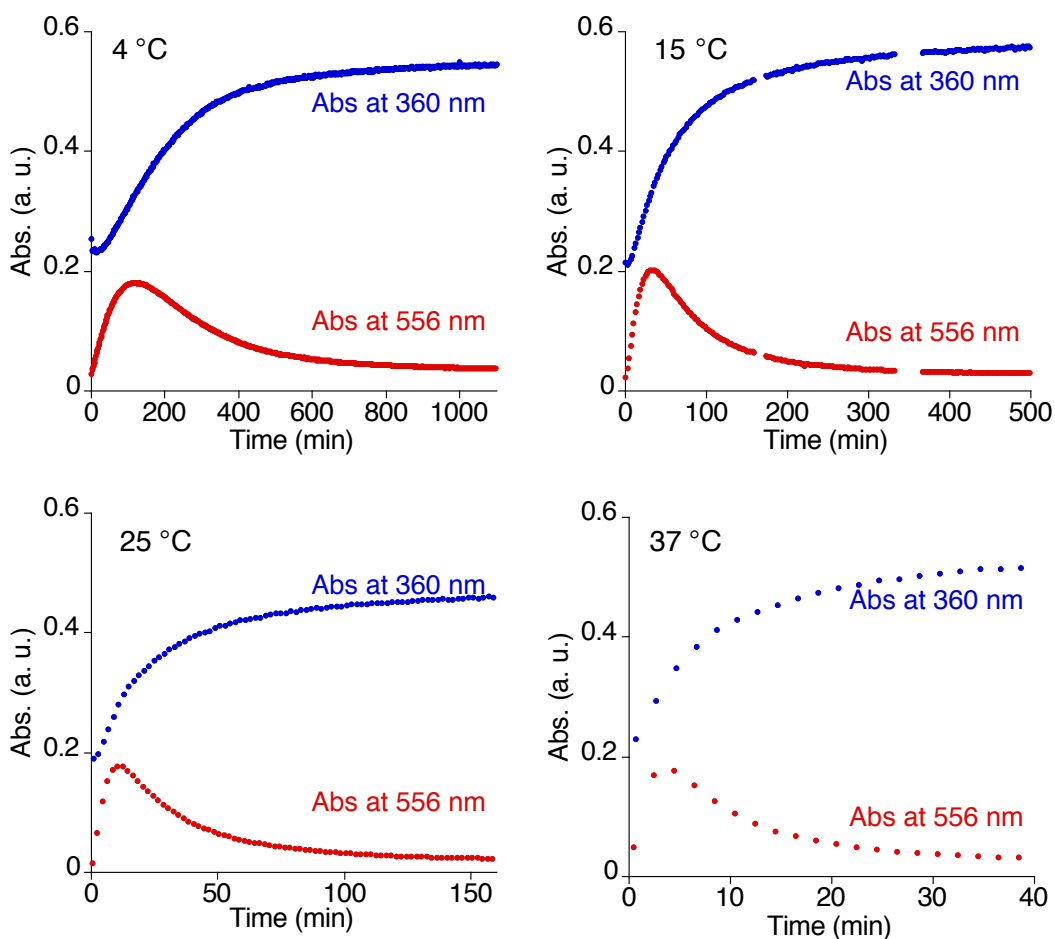
Figure II-27: Formation and loss of PSB (red) and gradual increase of SB (blue) over time for R111K:R132L:Y134F:T54V:R59W mutant showing temperature dependent effect on kinetic and thermodynamic product of the reaction.

by Meisam Nosrati (Professor James Geiger's lab, MSU).

Interestingly, two CRABPII pigments different only in E73A mutation, R111K:R132L:Y134F:T54V:R59W:A32W and R111K:R132L:Y134F:T54V:R59W: A32W:E73A, showed 4.4 units decrease in pK_a value. Since E73 hydrogen bonds with indole nitrogen of W109 (**Figure II-25**), the E73A mutation causes more than just a polarity change in the protein cavity. In the absence of E73, W109 adopts a nonproductive orientation for π-cation interaction with the PSB. Thus, the pK_a of the *trans-* isomer, surrounded by hydrophobic amino acids, is depressed.

On the basis of the latter rationale, we suggest that the higher pK_a variants in **Table II-9** correspond to the *cis*-iminium isomers that are capable of stabilizing the PSB by the protein surroundings. Conversely, the low pK_a mutants putatively adopt the structure of the *trans-* isomer that places the imine in an
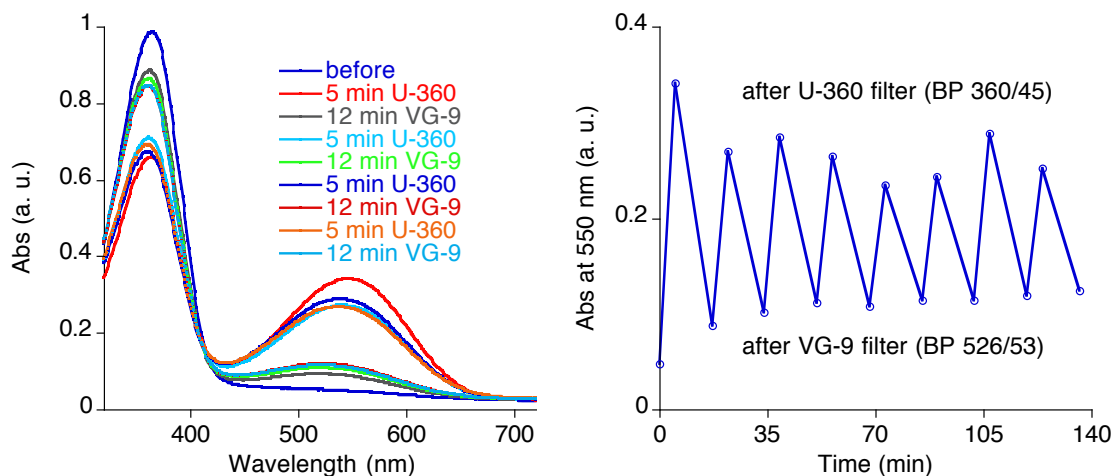
Figure II-28: Reversible light irradiation test and UV-vis measurement of R111K:R132L:Y134F:T54V:R59W mutant with filter BP 360/45 ('U-360') for SB and BP 526/53 ('VG-9') for PSB.

environment that is devoid of cation-stabilizing elements. The suggestion, that both isomeric species can be present at once, is also supported by pH titrations of R111K:R132L:Y134F:T54V: L121Y and R111K:R132L:Y134F:T54V: R59W:A32W:L121Y mutants (**Table II-9**, entries 4 and 15); both of the mutants show two separate pK_a values during the same titration (**Figure II-29**). Two values from the same titration for the last two mutants are comparable to those calculated for kinetic and thermodynamic forms of R111K:R132L:Y134F:T54V: R59W (kinetic ~5 and thermodynamic ~2.5).

Although it is not possible to extend this analysis with absolute certainty to all the protein/Rt complexes, the pigments with stable pK_a can be split into two categories: (1) those exhibiting high pK_a locked in a non-isomerizable *cis*-iminium conformation with A32W (3.2 Å from the β-ionone ring) holding the chromophore from any movement when the protein cavity is enclosed, and (2) those forming
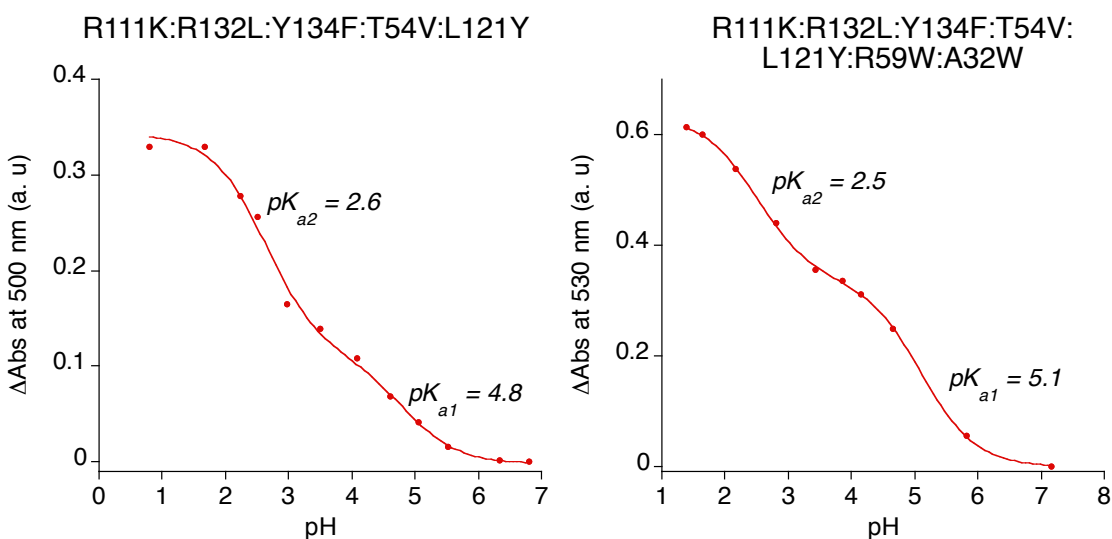
Figure II-29: Acid titrations of two depicted CRABPII mutants containing L121Y fitting a two pK_a curve.

directly low pK_a *trans*-iminum bond with the help from polar group of L121D, L121N, L121Q and in some cases L121Y mutants. Since L121Y is the least polar of the four it could adopt both conformations. For mutants with a pathway for isomerization, we observe a time-dependent change in the iminium pK_a. Alterations that lead to changes in overall polarity within each isomeric regime could also further affect the pK_a, a combination of which provides a rich series of mutant proteins that exhibit a range of pK_a values, extending from 8.1 to 1.8. On the other hand, electrostatic perturbations along the polyene that govern the overall absorption of the bound chromophore (as discussed earlier) provide a series of proteins that are variant not only in the pK_a of the PSB but also in their absorption maximum (**Table II-8**). This unique combination lends itself to the potential use of these protein complexes as chromophoric pH sensors.
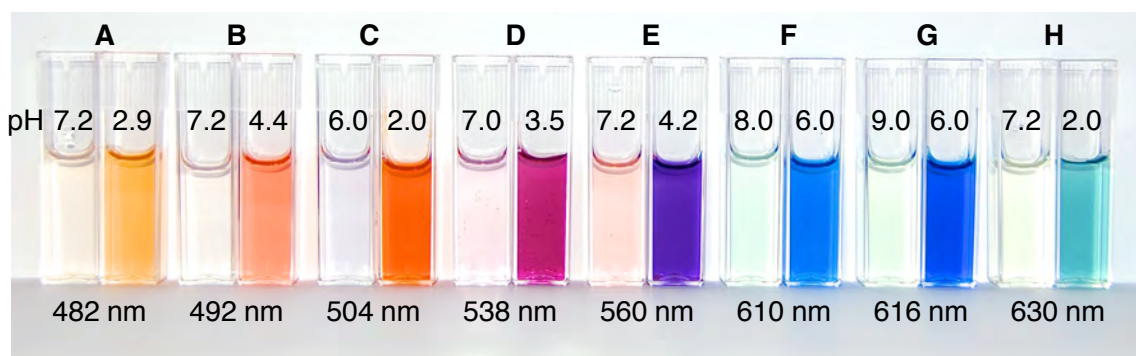
Figure II-30: Colorimetric pH-response of the CRABPII mutants: all solutions are pale yellow at basic pH but produce various colors at acidic pH based on pK_a and wavelength. Mutants: **A.** R111K:R132L:Y134F:T54V:L121N; **B.** R111K:R132L:Y134F:T54V:L121Q; **C.** R111K:R132L:Y134F:T54V:L121Y: R59W; **D.** R111K:R132L:Y134F:T54V:L121Q:R59Y:A32W; **E.** R111K:R132L: Y134F:T54V:L121Q:R59W:A32W:E73A:M93L; **F.** R111K:R132L:Y134F:T54V: R59W:A32W; **G.** R111K:R132Q:Y134F:T54V:R59W:A32W; **H.** R111K:R132Q: Y134F:T54V:R59W:A32W:E73A:M93L.

Figure II-30 illustrates the visual response to pH of some CRABPII mutants in their SB and PSB states. As mentioned earlier, the SB for each mutant absorbs in the same region corresponding to pale yellow color (~365 nm) due to the absence of a resonating positive charge. The protonated states (PSBs), however, absorb at various wavelengths, depending on the effective distribution of the cationic charge along the length of the retinylidene. This is a clear demonstration of the 'on' color state of each mutant at the appropriate pH, dictated by the pK_a of the particular pigment. This library of CRABPII mutants
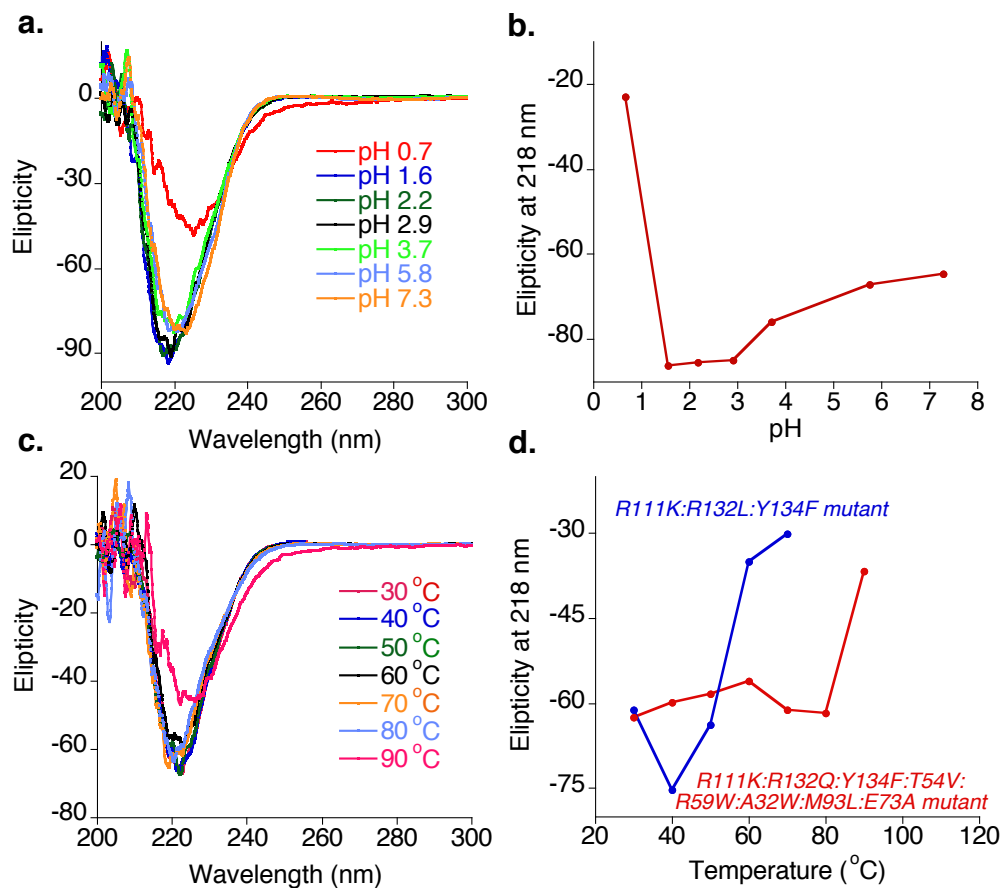
Figure II-31: Stability assays for R111K:R132Q:Y134F:T54V:R59W:A32W: E73A:M93L mutant estimated from CD spectra of β-sheet (218 nm). **a.** Spectra of the mutant titrated between pH 7.3 and 0.7 and **b.** Elipticity value plotted vs pH. **c.** Spectra of the mutant heated between 30 °C and 90 °C and **d.** Elipticity value plotted vs temperature for two indicted mutants for comparison.

117

provides the choice of various mutants, depending on the desired wavelength regime and the pK_a of the complex.

From the list of proteins potentially applicable as colorimetric pH sensors, the R111K:R132Q:Y134F:T54V:R59W:A32W:M93L:E73A CRABPII mutant bears special mention because of its representative protein stability at low pH (this mutant shows utility as a pH sensor in the next section). Exhibiting an apparent pK_a of 2.6 (**Table II-8**, entry 21), titration to pH 1.6 proceeds without denaturation of the protein (**Figure II-31a** and **II-31b**). This CRABP/Rt complex is highly colored (green) at pH 2 (**Figure II-30**), indicating that the protein is properly folded, because denatured protein bound with retinylidene would absorb at ~440 nm, corresponding to the characteristic wavelength of retinal-PSB. Also, the circular dichroic spectra (CD) show little change during acid titration (**Figure II-31b**), further suggesting that the protein retains its native β-sheet folding pattern. Significant changes in the CD spectra are detectable only at pH 0.7 being associated with protein precipitation. We also tested the thermal stability of the protein, which did not lose its tertiary structure until heated to 90 °C incubated with 0.07% detergent (**Figure II-31c**). The CRABPII-R111K:R132L:Y134F triple mutant denatures at 60 °C with the same amount of detergent (**Figure II-31d**). This extreme acid and heat stability is most likely due to enhanced hydrophobic packing in the binding cavity. The stability of acid- and heat- resistant proteins, typically found in acidophilic organisms or in extreme thermophiles, is usually

attributed to increased hydrogen bonding networks and salt bridge interactions, which is in clear contrast to the case reported here.

II.7. Building a ratiometric pH sensor out of two CRABPII mutants.

Genetically encoded pH sensors are of particular interest in the biological world. Practical pH sensors in biology are strongly fluorescent since inherently weak fluorescence cannot overcome issues with low sensitivity especially in the less abundant cellular compartments. The general principle of pH sensitivity described herein with the chromophoric protein systems can be extended to fluorescent systems that have an exchangeable iminium proton. The data below represents the proof-of-principle strategy in the development of CRABPII mutants as ratiometric pH sensors.

Design of a ratiometric system requires the presence of two spectroscopically differential states for a chromophore at different pH. One can suggest that having different SB and PSB peaks would be enough to build a ratiometric pH probe responding directly to the protonation/deprotonation
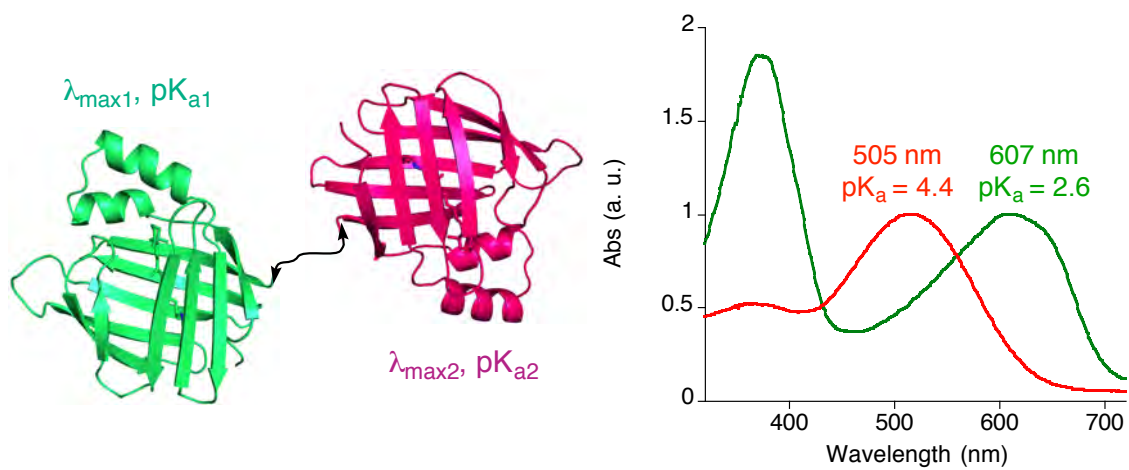
Figure II-32: Principle of building two protein fusion with λ_{max} separated by >100 nm and pK_a over two units apart.

phenomenon. Unfortunately, the peak corresponding to SB (~365 nm) would be oversaturated with the excess amount of all-*trans*-retinal either due to the presence of free form of the chromophore (380 nm) or because of non-specific SB formation with other free amines (~365 nm regardless of environment). Thus, to obtain two different absorption peaks with no overlap in SB region we envisioned a protein fusion constructed from two CRABPII mutants each with distinct λ_{max} and pK_a (**Figure II-32**). Because both mutants have unique absorptions upon protonation of the Schiff base and distinct pK_a values the intensity of absorbance with respect to protonation for each at the same pH would be different. A fusion construct of the two colorimetric proteins would ensure equal expression of each. The two reporters would be chosen based on the approximate pH of the system under investigation. The pK_a of the two proteins must be close enough such that in the pH region of interest both proteins respond colorimetrically to the pH change. In this manner, the ratio of two absorptions at specific wavelengths can be correlated to a standard curve.

The criteria for choosing two proteins for the pH responding pair are: (1) pK_a ~2 units apart; (2) λ_{max} of absorption >100 nm apart; and (3) similar affinity for retinal. The chosen pair (R111K:R132L:Y134F:T54V:L121Q:R59W hexa-mutant; pK_a 4.4, λ_{max} 505 nm, K_d 18 nM, and R111K:R132Q:Y134F:T54V:R59W:A32W:M93L:E73A octa-mutant; pK_a 2.6, λ_{max} 607 nm, K_d 35 nM) meet

120
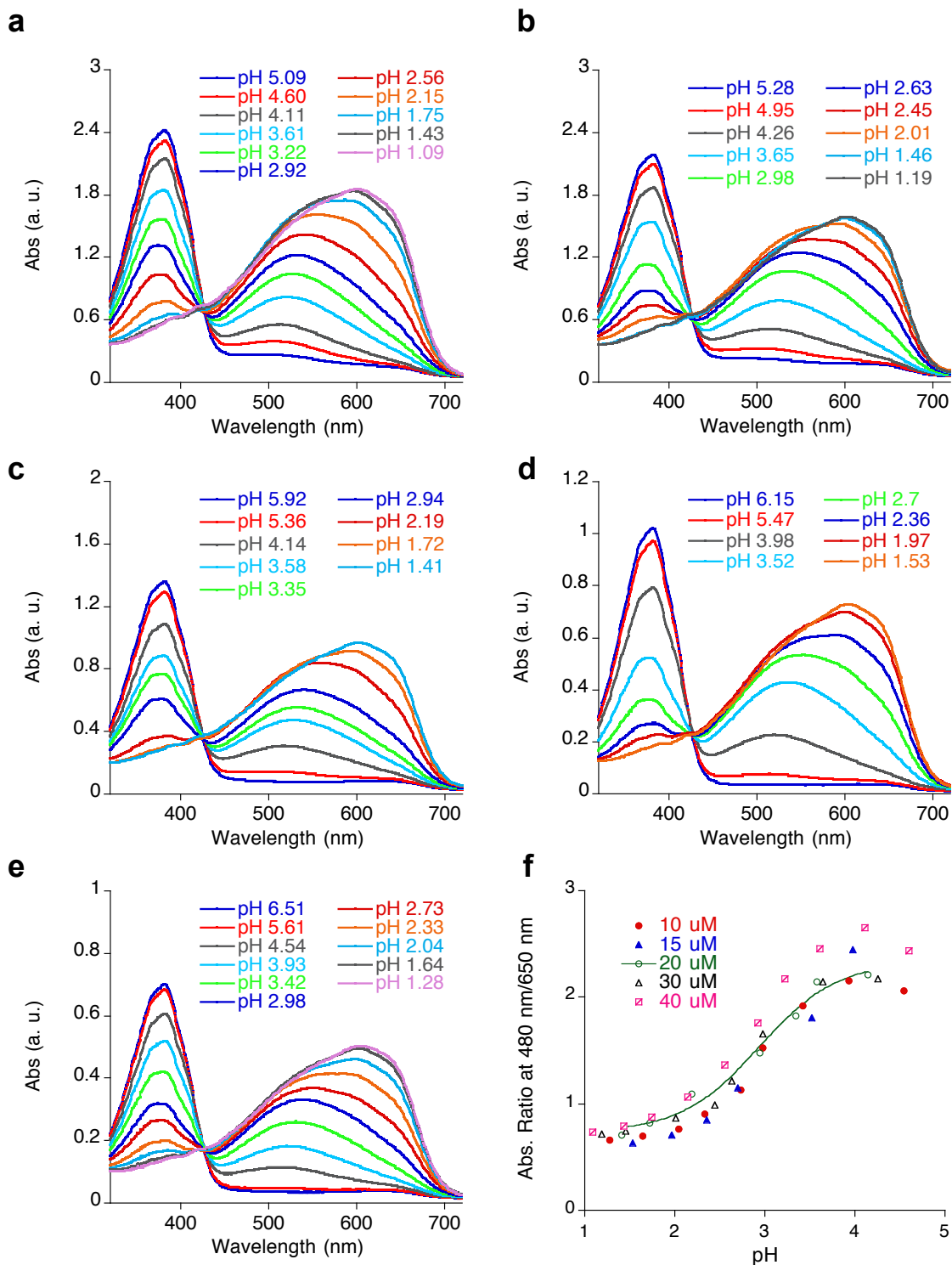
Figure II-33. The pH titrations of 1:1 hexa:octa-mutants at different concentration: **a.** 40 μM, **b.** 30 μM, **c.** 20 μM, **d.** 15 μM and **e.** 10 μM. **f.** The graph of the absorption ratios (480/650) at indicated concentration vs the pH defines the working range of the two-protein ratiometric pH-probe (working range is pH 2-4 with inflection point at 3.0).

all the criteria. The first test of the concept was performed with purified proteins that were mixed together at equal ratios (both mutants, at five different concentrations, from 10 to 40 µM, were independently preincubated with 2 equiv of retinal). The protein pair in citric acid buffer was titrated with a strong acid, and UV-vis spectra were recorded (**Figure II-33a–e**). The ratiometric standard curve was generated by plotting the 480/650 nm absorption ratio as a function of the pH, yielding the plot depicted in **Figure II-33f**. The overlap of the data collected at five different concentrations clearly illustrates the concentration-independent
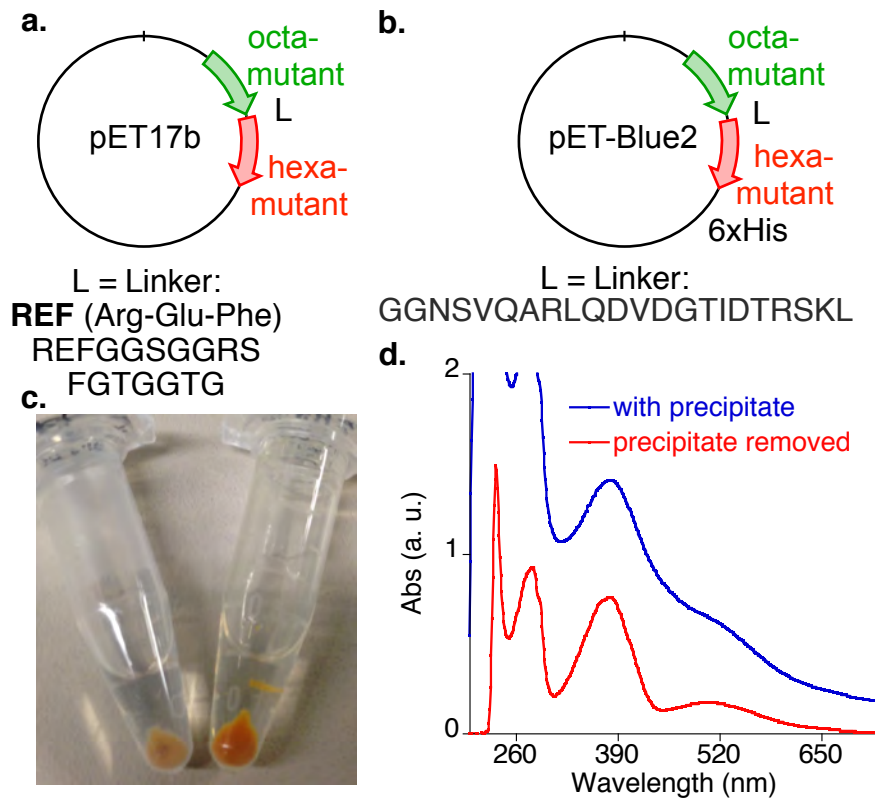
Figure II-34: Design of expression vector for two protein fusion. **a.** Fusion of octa- and hexa-mutants in pET17b vector with three different linkers (bold: the best working linker). **b.** Fusion of the same two mutants in pET-Blue2 vector with a much longer linker and a 6xHis-Tag. **c.** Amount of protein precipitate and non-colored supernatant of the fusion bound to retinal expressed and purified from pET-Blue2. **d.** Spectra of the fusion bound to retinal expressed and purified from pET17b-REF plasmid before and after spinning down the precipitate.

nature of the method.

Building the concentration independent pH-titration curve with the separate pigments in equimolar ratio prompted the construct shown in **Figure II-34**, with the sequence for one protein following the other. All the CRABPII mutants described thus far are cloned into the pET-17b vector and purified by anion exchange chromatography. We initiated cloning of two mutants in the same pET-17b vector (**Figure II-34a**) and in parallel into pET-Blue2 with a 6xHis-Tag sequence (**Figure II-43b**) to ensure reliable purification of at least one of them. The common feature among these constructs is the N-terminal location of the octa-mutant followed by a linker and the hexa-mutant.

The clone of two mutants in pET-Blue2 vector was constructed through two sets of cutting sites: at first the octa-mutant was cloned between NcoI and EcoRI cutting sites of the original vector and after a DNA sequence analysis the hexa-mutant was cloned between HindIII and XhoI cutting sites of the resulting plasmid. The frame was ensured by the proper primer design with addition of two base pairs to EcoRI-CRABPII reverse primer. The final linker had 22 amino acids in length between the two mutants (the amino acid sequence is GGNSVQA**R**LQ**D**V**D**GTI**DTR**S**K**L (**Figure II-34b**). Although long and relatively flexible, this linker contains multiple charged groups (in bold letters) that can potentially perturb protein folding and interfere with the SB formation (for example, non-specific SB with lysines from linker). Expression of protein fusion from this

construct was not high yielding and the Rt binding assay caused complete precipitation within 3 h (**Figure II-34c**).

The gene fusion designed in pET-17b vector was slightly different and in essence more successful. The plasmid already containing the octa-mutant was used as a template. Since there is only ribosomal binding site upstream and multiple cutting sites downstream of the CRABPII gene sequence in pET-17b, the gene encoding hexa-mutant was inserted at the C-terminus. At first, removal of
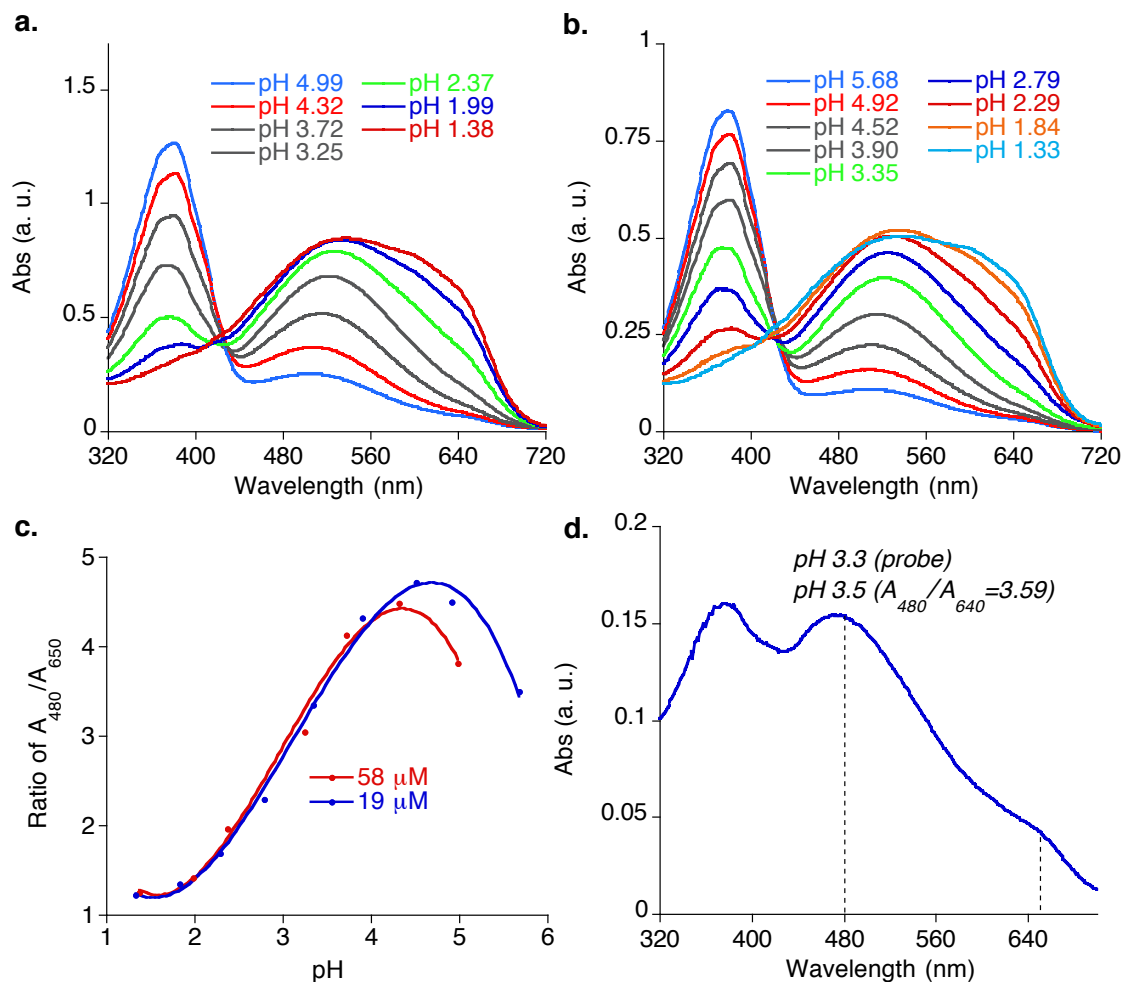
Figure II-35: Building the standard curve from the two protein fusion incubated with two equivalents of retinal. **a.** Acid titration and UV-vis spectra of 58 μM complex at each pH. **b.** Acid titration and UV-vis spectra of 19 μM complex at each pH. **c.** Standard curve from two titrations. **d.** Spectra of the citric acid solution of known pH 3.3 and the value of ratio.

the TGA stop-codon in octa-mutant gene sequence with one round of PCR provided the DNA template suitable for hexa-mutant insertion. After that, the hexa-mutant was inserted between EcoRI and XhoI cutting sites with EcoRI-CRABPII forward primers coding for various linkers and XhoI-CRABPII reverse primer introducing the stop codon. The simplest linkers mounted from mutation of the TGA stop codon into CGA and EcoRI cutting site (GAATTC) had three amino acid sequence Arg-Glu-Phe, R-E-F. In a parallel study, two other linkers were constructed: REFGGSGGRS and FGTGGTG. Since the sequence of CRABPII ends up with **R**136:**E**137 amino acids itself, the last linker was designed by shifting the RE part from the linker to CRABPII (removal of extra Arg-Glu) to provide a non-charged flexible linker. Expression and purification of all three constructs was comparable, but surprisingly binding with retinal gave the least protein precipitation for a construct with the simplest REF linker.

The isolated protein was incubated with 2 equiv of retinal at room temperature in a citrate buffer (10 mM, pH = 5), yielding the spectrum in **Figure II-35d**. The small amount of precipitate was separated by centrifugation. Acid titration of the fusion protein complexed with Rt from pH ~5 to ~1, at two dilutions (58 and 19 μM), gives similar UV-vis spectra as dictated by the pK_a of the two iminium protons bound to each half of the fusion protein. Initially, the absorption with highest intensity was at 510 nm (due to maximum absorption of hexa-mutant which has higher pK_a) but gradually shifted to longer wavelengths as a result of absorbance overlap between hexa- and octa-mutants (**Figure II-35a-b**). The final

125

ratiometric scatter plots at two concentrations have practically overlapping ratios between pH 2 and 4, the working pH range for the chosen protein pairs (**Figure II-35c**). The standard curve was utilized to measure the pH of known buffer solutions to within ±0.2 pH units (3.3 measured with a pH meter compared to 3.5 expected from the ratiometric probe) (**Figure II-35d**). The working range of the probe is identified from the ratiometric scatter plots and in this example, the working pH range is between 2 and 4: octa-mutant with pK_a 2.6 responds to pH changes between 2 and 4, hexa-mutant with pK_a 4.4 responds to pH changes between 3.4 and 5.4. The importance of a fused construct is clear by comparison of experiments with premixed protein solutions and with expressed protein fusion. Although the working range is the same, the ratio values for the fusion are different from the ratio obtained from the experiment with two proteins expressed individually. The latter system illustrates a proof-of-principle example of a ratiometric pH sensor that functions with the same chromophore, yet the chromophore not only absorbs at different wavelengths but also changes color at different pH values. Extension of this methodology into a system capable of in vivo applications with fluorescent chromophores has a great potential to introduce the absolutely novel genetically encoded pH-sensing probes, which will be discussed in details in Chapter 5.

II.8 Conclusion and significance.

An initial goal of this project to redesign the CRABPII binding cavity and perform the wavelength regulation with full length all-*trans*-retinal was achieved.

126

It was necessary to remove the interactions involved in retinoic acid binding to CRABPII in order to get the chromophore to form a SB. The wavelength regulation recipe was inspired by hCRBPII system but not using mutations at the identical residues. In addition to the wavelength regulation pH-dependent behavior of the several mutants was noticed, while the measured iminium pK_a values appeared to span between 1.8 and 8.1 units covering the physiologically important pH range and beyond. The concentration independent system was build from fusion of two selected mutants each with unique spectroscopic properties. Importance of the linker length between two mutants was also investigated giving the priority to the shortest one. Solution of known pH was estimated using standard curve built from the ratio of absorbances and the value within 0.2 units error was determined.

The next step in pH-sensor development would be to make it into an application-oriented probe. Fluorescence-based methods find much broader use in live organisms than the absorbance-based techniques. Even though the fluorescence from retinal has some limited potential with modern microscopy, fluorescent retinal analogs would provide a better choice as chromophores for developing pH-probe. Moreover, the fluorophore can be designed for optimal light penetration through the tissue.

Additionally, while investigating the pH dependent properties of CRABPII mutants, the time-dependent change of the pK_a values was observed in most of the pigments generated with Rt. This was attributed to the putative iminium

isomerization from the *cis-* form with high pK_a to the thermodynamically more preferable *trans-* isomer with low pK_a. All pigments with such properties exhibited pK_a values higher than 6 for the putative kinetic product and lower than 3 for the thermodynamic product. Interestingly, the low pK_a form can be converted into a high pK_a form by irradiation with light at 365 nm, matching the SB wavelength of absorption. The pigments of this kind can be presented as a new class of light-switchable proteins. One of the potential applications of this system would be a quencher system for known fluorescent proteins. Especially attractive is that all components can be genetically encoded without addition of unnatural chromophores because of availability of retinal in living organisms. However, additional studies on the mechanism of the observed pK_a changes are essential to support the hypothesis of iminium isomerization. Protein crystallography and ^{15}N NMR studies might provide some insight on the mechanism of the interconversion.

Another area to explore would be the exceptional heat and acid stability of the CRABPII-R111K:R132Q:Y134F:T54V:R59W:A32W:E73A:M93L mutant. From the first glance, the hydrophobic packing seems to play a crucial role in the observed phenomena. However, comparison of its crystal structure with the less stable mutant from the shorter series might provide an insight on the mechanism. Stability studies of the intermediate mutants might help to understand the factors

that contribute to the observed effect. Finally, crystallization of this mutant at pH < 2 and > 5 would be of particular interest to probe structural differences that could shed light on the stability of the proteins.

II.9 Materials and Methods.

Fluorescence spectra were recorded using a Fluorolog®-3 spectrofluorometer (HORIBA, Ltd.). UV-vis spectra were recorded with a Cary 300 Bio WinUV, Varian spectrophotometer. All-*trans*-retinal was purchased from Sigma-Aldrich, and was used as received. The dissociation constant K_d for protein/retinal binding was determined by fluorescence titration as previously reported.

Site-derected mutagenesis. The CRABPII-pET17b plasmid described before[8] was used for mutagenesis following QuickChange Site-directed Mutagenesis Kit protocol (Agilent Technologies). PCR conditions for amplification of mutants are specified below:

Total Reaction Volume	50 μL
Template (DNA plasmid)	70 ng (x μl)
Primer forward	20 pmol (y μl)
Primer reverse	20 pmol (z μl)
dNTP	1 μL
10× pfu buffer	5 μL
Pfu Turbo (DNApolymerase)	1 μL
DI water	50-x-y-z-7 μL

<div align="center">PCR program</div>

1×	95 ºC	4 min
	95 ºC	30 sec
20×	temperature 3-5 ºC lower than primer melting temperature	1 min
	72 ºC	4 min 30 sec
1×	72 ºC	10 min
1×	25 ºC	10 min

20 units of *Dpn* I restriction enzyme (New England BioLabs®) were added to the PCR reaction for DNA template digestion and the reaction mixture was incubated at 37 ˚C for 1 h. The resulting mixture was transformed into competent cells for DNA amplification and purification.

List of primers:

R111K

Forward: 5'-CCCAAGACCTCGTGGACC<u>AAA</u>GAACTGACCAACGATGGG-3'

Reverse: 5'-CCCATCGTTGGTCAGTTC<u>TTT</u>GGTCCACGAGGTCTTGGG-3'

R132L:Y134F

Forward: 5'-GTGCACC<u>CTG</u>GTC<u>TAC</u>GTCCGAGAG-3'

Reverse: 5'-CTCTCGGAC<u>GTA</u>GAC<u>CAG</u>GGTGCAC-3'

R132Q:Y134F

Forward: 5'- GTTGTGTGCACC<u>CAG</u>GTC<u>TTC</u>GTCCG-3'

Reverse: 5'-CGGAC<u>GAA</u>GAC<u>CTG</u>GGTGCACACAAC-3'

R132Y:Y134F

Forward: 5'- GTTGTGTGCACC<u>TAC</u>GTC<u>TTC</u>GTCCG-3'

Reverse: 5'-CGGAC<u>GAA</u>GAC<u>GTC</u>GGTGCACACAAC-3'

R132E:Y134F

Forward: 5'- GTTGTGTGCACC<u>GAG</u>GTC<u>TTC</u>GTCCG-3'

Reverse: 5'-CGGAC<u>GAA</u>GAC<u>CTC</u>GGTGCACACAAC-3'

T54V

Forward: 5'-CTACATCAAA<u>GTC</u>TCCACCACCGTGCG -3'

Reverse: 5'- CGCACGGTGGTGGA<u>GAC</u>TTTGATGTAG -3'

T54C

Forward: 5'- GGAGACACTTTCTACATCAAA<u>TGC</u>TCC -3'

Reverse: 5'- GGA<u>GCA</u>TTTGATGTAGAAAGTGTCTCC -3'

T54F

Forward: 5'-GACACTTTCTACATCAAA<u>TTC</u>TCC-3'

Reverse: 5'- GGA<u>GAA</u>TTTGATGTAGAAAGTGTC -3'

L121E

Forward: 5'-GATGGGGAACTGATC<u>GAG</u>ACCATGACGGCGGATGAC -3'

Reverse: 5'-GTCATCCGCCGTCATGGT<u>CTC</u>GATCAGTTCCCCATC -3'

L121D

Forward: 5'-GGGGAACTGATC<u>GAC</u>ACCATGACGGCG -3'

Reverse: 5'-CGCCGTCATGGT<u>GTC</u>GATCAGTTCCCC -3'

L121N

Forward: 5'-GAACTGATC<u>AAC</u>ACCATGACG -3'

Reverse: 5'-CGTCATGGT<u>GTT</u>GATCAGTTC -3'

L121Q

Forward: 5'-GAACTGATC<u>CAG</u>ACCATGACG -3'

Reverse: 5'-CGTCATGGT<u>CTG</u>GATCAGTTC -3'

L121Y

Forward: 5'-GAACTGATC<u>TAC</u>ACCATGACG -3'

Reverse: 5'-CGTCATGGT<u>GTA</u>GATCAGTTC -3'

R59D

Forward: 5'-CCTCCACCACCGTG<u>GAC</u>ACCACAGAG-3'

Reverse: 5'-CTCTGTGGT<u>GTC</u>CACGGTGGTGGAGG-3'

R59N

Forward: 5'-ACCGTG<u>AAT</u>ACCACAGAGATTAACTTC -3'

Reverse: 5'- GAAGTTAATCTCTGTGGT<u>ATT</u>CACGGT -3'

R59Q

Forward: 5'-CCTCCACCACCGTG<u>CAG</u>ACCACAGAG-3'

Reverse: 5'-CTCTGTGGT<u>CTG</u>CACGGTGGTGGAGG-3'

R59E

Forward: 5'- CCTCCACCACCGTG<u>GAG</u>ACCACAGAG -3'

Reverse: 5'- CTCTGTGGT<u>CTC</u>CACGGTGGTGGAGG -3'

R59I

Forward: 5'-CCACCGTG<u>ATC</u>ACCACAGAGATTAAC-3'

Reverse: 5'-GTTAATCTCTGTGGT<u>GAT</u>CACGGTGG-3'

R59L

Forward: 5'- CCTCCACCACCGTG<u>TT</u>GACCACAGAG -3'

Reverse: 5'- CTCTGTGGT<u>CAA</u>CACGGTGGTGGAGG -3'

R59F

Forward: 5'- CCTCCACCACCGTG<u>TTC</u>ACCACAGAG -3'

Reverse: 5'- CTCTGTGGT<u>GAA</u>CACGGTGGTGGAGG -3'

R59Y

Forward: 5'- CCTCCACCACCGTG<u>TAC</u>ACCACAGAG -3'

Reverse: 5'- CTCTGTGGT<u>GTA</u>CACGGTGGTGGAGG -3'

R59W

Forward: 5'- CCTCCACCACCGTG<u>TGG</u>ACCACAGAG -3'

Reverse: 5'- CTCTGTGGT<u>CCA</u>CACGGTGGTGGAGG -3'

A32Y

Forward: 5'-GAGGAAGATT<u>TAT</u>GTGGCTGC-3'

Reverse: 5'-GCAGCCAC<u>ATA</u>AATCTTCCTC-3'

A32W

Forward: 5'-GTGATGCTGAGGAAGATT<u>TGG</u>GTGGCTGC-3'

Reverse: 5'-GCAGCCAC<u>CCA</u>AATCTTCCTCAGCATCAC-3'

A36Y

Forward: 5'-GTGGCTGCA<u>TAC</u>TCCAAGCC-3'

Reverse: 5'-GGCTTGGA<u>GTA</u>TGCAGCCAC-3'

A32Y:A36Y

Forward: 5'-GGAAGATT<u>TAT</u>GTGGCTGCA<u>TAT</u>TCCAAGCC-3'

Reverse: 5'-GGCTTGGA<u>ATA</u>TGCAGCCAC<u>ATA</u>AATCTTCC-3'

F15Y

Forward: 5'-CGGAAAAC<u>TAC</u>GAGGAATTGC -3'

Reverse: 5'- GCAATTCCTC<u>GTA</u>GTTTTCCG -3'

F15D

Forward: 5'-CGGAAAAC<u>GAC</u>GAGGAATTGC -3'

Reverse: 5'- GCAATTCCTC<u>GTC</u>GTTTTCCG -3'

L19W

Forward: 5'- CGAGGAATTG<u>TGG</u>AAAGTGCTGGGG-3'

Reverse: 5'- CCCCAGCACTTT<u>CCA</u>CAATTCCTCG -3'

V24W

Forward: 5'-GTGCTGGGG<u>TGG</u>AATGTGATG -3'

Reverse: 5'- CATCACATT<u>CCA</u>CCCCAGCAC -3'

A35W

Forward: 5'-GTGGCT<u>TGG</u>GCGTCCAAGCC -3'

Reverse: 5'- GGCTTGGACGC<u>CCA</u>AGCCAC -3'

A35F

Forward: 5'-GTGGCT<u>TTC</u>GCGTCCAAGCC -3'

Reverse: 5'- GGCTTGGACGC<u>GAA</u>AGCCAC -3'

A36W

Forward: 5'-GTGGCTGCA<u>TGG</u>TCCAAGCC -3'

Reverse: 5'- GGCTTGGA<u>CCA</u>TGCAGCCAC -3'

A36T

Forward: 5'-GTGGCTGCA<u>ACG</u>TCCAAGCC -3'

Reverse: 5'- GGCTTGGA<u>CGT</u>TGCAGCCAC -3'

K38D

Forward: 5'- GCAGCGTCC<u>GAG</u>CCAGCAGTGG-3'

Reverse: 5'- CCACTGCTGG<u>CTC</u>GGACGCTGC -3'

P39W

Forward: 5'- GCGTCCAAG<u>TGG</u>GCAGTGGAG-3'

Reverse: 5'- CTCCACTGC<u>CCA</u>CTTGGACGC -3'

Q45F

Forward: 5'- GTGGAGATCAAA<u>TTC</u>GAGGGAGACAC-3'

Reverse: 5'- GTGTCTCCCTC<u>GAA</u>TTTGATCTCCAC -3'

T54V:T56W

Forward: 5'- GACACTTTCTACATCAAAGTCTCC<u>TGG</u>ACC-3'

Reverse: 5'- GGT<u>CCA</u>GGAGACTTTGATGTAGAAAGTGTC -3'

T54V:T56Q

Forward: 5'- GACACTTTCTACATCAAAGTCTCCC<u>CAG</u>ACC-3'

Reverse: 5'- GGT<u>CTG</u>GGAGACTTTGATGTAGAAAGTGTC -3'

T54V:T56Y

Forward: 5'- CTTTCTACATCAAAGTCTCC<u>TAC</u>ACC-3'

Reverse: 5'- GGT<u>GTA</u>GGAGACTTTGATGTAGAAAG -3'

V58W:R59W

Forward: 5'-CCACCACC<u>TGGTGG</u>ACCACAGAGATTAAC -3'

Reverse: 5'- GTTAATCTCTGTGGT<u>CCACCA</u>GGTGGTGG -3'

G74W

Forward: 5'- GGGGAGGAGTTTGAGGAG<u>TGG</u>AC-3'

Reverse: 5'- GT<u>CCA</u>CTCCTCAAACTCCTCCCC -3'

V76W

Forward: 5'- GGAGCAGACT<u>TGG</u>GATGGGAGG-3'

Reverse: 5'- CCTCCCATC<u>CCA</u>AGTCTGCTCC -3'

V76F

Forward: 5'- GGAGCAGACT<u>TTC</u>GATGGGAGG-3'

Reverse: 5'- CCTCCCATC<u>GAA</u>AGTCTGCTCC -3'

Q78W

Forward: 5'- GACTGTGGAT<u>TGG</u>AGGCCCTG-3'

Reverse: 5'- CAGGGCCT<u>CCA</u>ATCCACAGTC -3'

[75]AA[76]

Forward: 5'- GAGGAGCAGACTGCCGCCGTGGATGGGAGG-3'

Reverse: 5'- CCTCCCATCCACGGCGGCAGTCTGCTCCTC -3'

M93L

Forward: 5'-GAGAATAAACTGGTCTGTGAG-3'

Reverse: 5'-CTCACAGACCAGTTTATTCTC-3'

M93F

Forward: 5'-GAGAATAAATTCGTCTGTGAGC-3'

Reverse: 5'-GCTCACAGACGAATTTATTCTC-3'

M123L

Forward: 5'-GATCCTGACCCTGACGGCGG -3'

Reverse: 5'- CCGCCGTCAGGGTCAGGATC -3'

E73A

Forward: 5'-GGAGTTTGAGGCGCAGACTGTGG-3'

Reverse: 5'-CCACAGTCTGCGCCTCAAACTCC-3'

S12D

Forward: 5'-CTGGAAAATCATCCGAGATGAAAACTTCGAG-3'

Reverse: 5'-CTCGAAGTTTTCATCTCGGATGATTTTCCAG-3'

endCRABPIInoSt

Forward: 5'-CCGAGAGCGAGAGTTCTGC-3'

Reverse: 5'- GCAGAACTCTCGCTCTCGG -3'

NcoI_CRABP_for

5'-CG<u>CCATGG</u>CACCAAACTTCTCTGGCAACTGG-3'

EcoRI_CRABP_rev

5'-CC<u>GAATTC</u>CCTCCCTCTCGGACGAAGAC-3'

CRABP HindIII (for)

5'-GC<u>AAGCTT</u>CCAAACTTCTCTGGCAACTGG-3'

CRABP XhoI (rev)

5'-G<u>CTCGAG</u>TCCCTCTCGGACGAAGAC-3'

EcoRI-CRABP (for)

5'-CC<u>GAATTC</u>ATGCCAAACTTCTC-3'

EcoRI-GGSGGRS-CRABP (for)

5'-CC<u>GAATTC</u>GGTGGTAGTGGTGGTAGAAGTATGCCAAACTTCTCTGGC-3'

EcoRI-GTGGTG-CRABP (for)

5'-CC<u>GAATTC</u>GGTACTGGTGGTACTGGTATGCCAAACTTCTCTGGC-3'

CRABP XhoI St (rev)

5'- GG<u>CTCGAG</u>TCACTCTCGGACGAAGAC-3'

The PCR products (10 µL) were transformed into *E. coli* XL-1 blue competent cells (Novagen®, 100 µL) and grown on Luria-Bertani (LB)-agar plates supplemented with antibiotics (tetracycline: 7.5 µg/mL; ampicillin: 100 µg/mL) at 37 ºC for 18 hours. A single colony was inoculated in LB medium (15 mL) containing 7.5 µg/mL tetracycline and 100 µg/mL ampicillin and incubated at 37 ºC while shaking, for 10 hours. DNA purification was performed using

Promega Wizard® Plus SV Miniprep DNA purification kit (A1330) following the suggested protocols. The average isolated plasmid DNA concentration was 200 ng/μL in 50 μL solution. The plasmid DNA concentration was calculated using Nanodrop (Thermo Scientific). The DNA sequence was verified by the MSU gene sequencing facility with the T7 primer as the initial probe.

Protein Expression and Purification of CRABPII/pET-17b Mutants. The target gene was transformed into *E. coli* BL21(DE3)pLysS competent cells (100 ng of DNA, for 100 μL of cell solution) following standard protocols and the cells were grown on Luria-Bertani (LB)-agar plates supplemented with antibiotics (chloramphenicol: 27 μg/mL; ampicillin: 100 μg/mL) at 37 °C for 12 hours. A single colony was used to inoculate LB medium (10 mL) containing 27 μg/mL chloramphenicol and 100 μg/mL ampicillin and was incubated at 37 °C, while shaking overnight. The grown culture was inoculated into 1 L of LB containing 27 μg/mL chloramphenicol and 100 μg/mL ampicillin and was grown at 37 °C while shaking for an additional 6 hours. The expression was induced with addition of isopropyl-β-D-thiogalactopyranoside (IPTG, Gold Biotechnology, 1 mM) and the culture was shaken at 20 °C for 36 h. The cells were harvested by centrifugation (5000 rpm, 10 min, 4 °C) and resuspended in Tris-binding buffer (10 mM Tris, pH=8.0, 50 mL). The cells were lysed by sonication (Biologics, Inc, power 60%, 3 × 1.5 min). The solution was spun down (5000 rpm, 30 min, 4 °C) and the supernatant was loaded on a FastQ anion exchange column pre equilibrated with Tris-binding buffer (10 mM Tris, pH = 8.0, 100 mL). The column was washed

twice with Tris-binding buffer (2 x 200 mL) and the protein was eluted with Tris-elution buffer (10 mM Tris, 200 mM sodium chloride, pH = 8.0, 100 mL). The eluent was buffer exchanged with Tris-binding buffer using an ultrafiltration cell under nitrogen pressure (~20 psi) (Millipore[TM], regenerated cellulose membrane YM10, NMWL: 10,000). The desired protein was further purified using a BioLogic DuoFlow system (BioRad) equipped with a 15Q anion exchange column, according to the following protocol:

Description		Parameters
Isocratic Flow	pH=8.1, 0% 2M NaCl	10.00 ml, 3.00 ml/min
Linear Gradient	pH=8.1, 0→4% 2M NaCl	20.00 ml, 3.00 ml/min
Isocratic Flow	pH=8.1, 4% 2M NaCl	20.00 ml, 3.00 ml/min
Linear Gradient	pH=8.1, 4→8% 2M NaCl	10.00 ml, 3.00 ml/min
Isocratic Flow	pH=8.1, 8% 2M NaCl	20.00 ml, 3.00 ml/min

The above protocol was optimized for CRABPII mutants. The buffer used is 50 mM Tris, pH is adjusted automatically. The proteins elute with 4% 2M NaCl.

The absorption extinction coefficients (ε) for CRABPII mutants were determined according to the method described by Gill and von Hippel.[36]

General Procedure for Reductive Amination of the Protein/Retinal Complex and Q-TOF ES[+] Analysis. Reductive amination of the CRABPII-retinal complexes non-reversibly traps the protein/chromophore complex. In addition to the protein peak (M), an adduct peak (M+268) is observed if the protein is reductively aminated with retinal in the presence of the reducing reagent

140

(NaCNBH$_3$). During initial reengineering efforts the first four proteins in Table 2 were subjected to reductive amination conditions, verifying SB formation. Only CRABPII-R111K failed to produce the desired M+268 peak.

Briefly, 40-100 μM of CRABPII protein in citric acid buffer (1 mL 0.1 M, pH=5.2) was incubated with 1 equivalent of retinal for 6 h. NaCNBH$_3$ (50 μL of 5 M water solution) was added to the latter solution while stirring for an additional 3 h. The reaction was quickly desalted using FastQ resin and concentrated back to 1 mL with a 3 KDa cut off filter. The adduct was subjected to electrospray MS with desalting column using the protocol below.

The Q-TOF ES+ protocol description

A, %	B, %	Time, min
95	5	1.00
30	70	2.50
30	70	4.50
95	5	4.51
95	5	5.00

Solvent A: water + 0.1% formic acid,

Solvent B: acetonitrile.

UV-vis Measurements. The CRABPII-PSB formation (λ_{max} >440 nm) was followed by UV-vis in phosphate buffer (4 mM NaH$_2$PO$_4$, 16 mM Na$_2$HPO$_4$, 150 mM NaCl, pH=7.3) or in citric acid buffer (50 mM citric acid, pH=5.2). pH was verified every time before recording the spectrum. The experiment was

performed with final protein concentration of 40-80 μM, and 0.5 equivalents of retinal were added (from a stock solution of 1 mM in ethanol) at pH 7.3 or 5.2. Peaks with λ_{max} >450 nm are considered PSB peaks, while deprotonated imine peaks (SB) appear at ~360 nm. Non-covalently bound retinal absorbs at ~380 nm.

PSB pK_a Determination for CRABPII-retinal Complexes. For the pK_a determination we monitored the PSB peak with UV-vis at different pH with acidification/basification. Retinal (0.5 equiv) was added to CRABPII mutants and incubated until Schiff base formation was complete (verified by UV-vis). The solution was acidified with hydrochloric acid (6 M, ~1 µL) for mutants with low pK_a or basified with sodium hydroxide solution (4 M, ~1 µL) for mutants with high pK_a. The corresponding UV-vis spectra are provided. The pK_a value was determined based on the plot of the absorbance change vs the pH, applying the curve fit as described previously for bacteriorhodopsin pK_a determination.[28]

$$\Delta A = \Delta A_0 \Big/ (1 + 10^{[pH-pKa]})$$

The two parameters are: ΔA_o, the total absorbance change of PSB during each acidification in comparison to the most basic point during titration and pK_a, the midpoint of titration.

Time-dependent Changes in pK_a. Time-dependent PSB behavior of the CRABPII mutants was monitored with UV-vis over time at indicated temperature. UV-vis in citric acid buffer (0.2 M citric acid, pH=5.2) was recorded from 800 nm

to 200 nm, every 2 min until the amplitude of the PSB peak remained constant. For each experiment, 1 mL of protein (15 μM) in citrate buffer was pre-incubated at each temperature for 5 min and then degassed. The recording of data started immediately after addition of 0.5 equiv (7.5 μM) of retinal to the protein solution and continued until PSB (λ_{max} >450 nm) remained constant. The absorption of SB (360 nm) and PSB (>450 nm) vs time was plotted, with the first time point adjusted to the UV-vis scanning speed (for example, if the PSB peak data at 556 nm is necessary to plot over time it takes 0.4 min to reach 556 nm from 800 nm with the scanning speed 600 nm/min). Incubation of the solution at 4 ^{o}C reduced the rate of protein chromophore rearrangement from high pK_a to low pK_a form. Thus measurements of the high pK_a for the penta-mutant was performed at 4 ^{o}C.

Titration of equimolar mixture of two selected mutants bound with retinal. Two selected CRABPII mutants (R111K:R132L:Y134F:T54V:L121Q:R59W and R111K:R132Q:Y134F:T54V:R59W:A32W:M93L:E73A at same concentration (60 μM) were separately pre-incubated with retinal (1 equivalent) until SB formation was complete. The two pigment solutions were combined in a 1:1 ratio and diluted subsequently to different concentrations: 40 μM, 30 μM, 20 μM, 15 μM and 10 μM. The mixtures were titrated with acid while recording the UV-vis for each new pH value. The ratio between two peaks was calculated at different wavelengths and was plotted as a function of the pH. The ratio of absorbances at 480 nm/650 nm was found to give the best ratiometric curve, consistent for all protein concentrations.

Titration of two selected fusion mutants bound with retinal. The fusion of two selected CRABPII mutants with REF linker in pET17b was generated as described above. The protein was expressed and purified as previously described for other CRABPII mutants that expressed as a single protein in pET17b vector. Only modification was in salt concentration (10 mM Tris, 300 mM sodium chloride, pH = 8.0, 100 mL) for protein elution. The protocol for BioLogic DuoFlow system equipped with a 15Q anion exchange column was optimized accordingly:

Description		Parameters
Isocratic Flow	pH=8.1, 0% 2M NaCl	10.00 ml, 3.00 ml/min
Linear Gradient	pH=8.1, 0→4% 2M NaCl	10.00 ml, 3.00 ml/min
Isocratic Flow	pH=8.1, 4% 2M NaCl	20.00 ml, 3.00 ml/min
Linear Gradient	pH=8.1, 4→8% 2M NaCl	10.00 ml, 3.00 ml/min
Isocratic Flow	pH=8.1, 8% 2M NaCl	20.00 ml, 3.00 ml/min
Linear Gradient	pH=8.1, 8→15% 2M NaCl	20.00 ml, 3.00 ml/min
Isocratic Flow	pH=8.1, 15% 2M NaCl	20.00 ml, 3.00 ml/min

The protein elutes with 15% 2M NaCl. The purified protein was subjected to Q-TOF ES$^+$ to confirm expression of the fused mutant.

The isolated protein was concentrated to 50 μM and was incubated with 100 μM retinal for 5 h at room temperature (50 mM citric acid, pH=5.2) to generate the mature pH-sensor. Different concentrations of the final pigment was titrated from pH of 5 to 1.5 at different dilutions leading to a ratiometric standard curve.

Molecular Modeling. Docking of retinal was performed using AutoDockTools 4.2.1. Macromolecule model was generated from the crystal structure of WT-CRABPII (PDB: 2FS6) with removed water molecules, ligand retinal was generated in Spartan 08 using B3LYP/6-31G* basis set. Docking was performed into CRABPII-R111K, CRABPII-I63K, CRABPT-I52K, CRABPII-L119K and CRABPII-M93K models.

Molecular modeling was performed on a SGI workstation (octane 2) using the software package InsightII. Crystal structure of R111K:R132L:Y134F:T54V:R59W-CRABPII mutant complexed with retinal refined with the *trans*-imine geometry was used as a template to build the *cis*-imine geometry. The model was heated to 300 K and energy minimized with 1000 iterations using cvff force field.

REFERENCES

REFERENCES

1.   Huang, P.; Chandra, V.; Rastinejad, F., Retinoic Acid Actions through Mammalian Nuclear Receptors. *Chemical Reviews* **2014,** *114* (1), 233.

2.   Schug, T. T.; Berry, D. C.; Shaw, N. S.; Travis, S. N.; Noy, N., Opposing effects of retinoic acid on cell growth result from alternate activation of two different nuclear receptors. *Cell* **2007,** *129* (4), 723.

3.   Sessler, R. J.; Noy, N., A ligand-activated nuclear localization signal in cellular retinoic acid binding protein-II. *Molecular Cell* **2005,** *18* (3), 343.

4.   Chaudhuri, B. N.; Kleywegt, G. J.; Broutin-L'Hermite, I.; Bergfors, T.; Senn, H.; Le Motte, P.; Partouche, O.; Jones, T. A., Structures of cellular retinoic acid binding proteins I and II in complex with synthetic retinoids. *Acta Crystallographica Section D-Biological Crystallography* **1999,** *55*, 1850.

5.   Vaezeslami, S.; Mathes, E.; Vasilelou, C.; Borhan, B.; Geiger, J. H., The structure of apo-wild-type cellular retinoic acid binding protein II at 1.4 A and its relationship to ligand binding and nuclear translocation. *Journal of Molecular Biology* **2006,** *363* (3), 687.

6.   Vasileiou, C.; Lee, K. S. S.; Crist, R. M.; Vaezeslami, S.; Goins, S. A.; Geiger, J. H.; Borhan, B., Dissection of the critical binding determinants of cellular retinoic acid binding protein II by mutagenesis and fluorescence binding assay. *Proteins-Structure Function and Bioinformatics* **2009,** *76* (2), 281.

7.   Crist, R. M.; Vasileiou, C.; Rabago-Smith, M.; Geiger, J. H.; Borhan, B., Engineering a rhodopsin protein mimic. *Journal of the American Chemical Society* **2006,** *128* (14), 4522.

8.   Vasileiou, C.; Vaezeslami, S.; Crist, R. M.; Rabago-Smith, M.; Geiger, J. H.; Borhan, B., Protein design: Reengineering Cellular Retinoic Acid Binding Protein II into a rhodopsin protein mimic. *Journal of the American Chemical Society* **2007,** *129* (19), 6140.

9.   Lee, K. S. S.; Berbasova, T.; Vasileiou, C.; Jia, X.; Wang, W.; Choi, Y.; Nossoni, F.; Geiger, J. H.; Borhan, B., Probing Wavelength Regulation with an Engineered Rhodopsin Mimic and a C15-Retinal Analogue. *Chempluschem* **2012,** *77* (4), 273.

10.   Calderone, V.; Berni, R.; Zanotti, G., High-resolution structures of retinol-binding protein in complex with retinol: pH-induced protein structural changes in the crystal state. *Journal of Molecular Biology* **2003,** *329* (4), 841.

147

11.   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* **2002,** *277* (44), 41970.

12.   Wang, W.; Nossoni, Z.; Berbasova, T.; Watson, C. T.; Yapici, I.; Lee, K. S. S.; Vasileiou, C.; Geiger, J. H.; Borhan, B., Tuning the Electronic Absorption of Protein-Embedded All-trans-Retinal. *Science* **2012,** *338* (6112), 1340.

13.   Andersen, L. H.; Nielsen, I. B.; Kristensen, M. B.; El Ghazaly, M. O. A.; Haacke, S.; Nielsen, M. B.; Petersen, M. A., Absorption of Schiff-base retinal chromophores in vacuo. *Journal of the American Chemical Society* **2005,** *127* (35), 12347.

14.   Rajput, J.; Rahbek, D. B.; Andersen, L. H.; Hirshfeld, A.; Sheves, M.; Altoe, P.; Orlandi, G.; Garavelli, M., Probing and Modeling the Absorption of Retinal Protein Chromophores in Vacuo. *Angewandte Chemie-International Edition* **2010,** *49* (10), 1790.

15.   Wang, W.; Geiger, J. H.; Borhan, B., The photochemical determinants of color vision. *Bioessays* **2014,** *36* (1), 65.

16.   Vasileiou, C.; Wang, W.; Jia, X.; Lee, K. S. S.; Watson, C. T.; Geiger, J. H.; Borhan, B., Elucidating the exact role of engineered CRABPII residues for the formation of a retinal protonated Schiff base. *Proteins-Structure Function and Bioinformatics* **2009,** *77* (4), 812.

17.   Burgi, H. B.; Dunitz, J. D., FROM CRYSTAL STATICS TO CHEMICAL-DYNAMICS. *Accounts of Chemical Research* **1983,** *16* (5), 153.

18.   Burgi, H. B.; Dunitz, J. D.; Lehn, J. M.; Wipff, G., STEREOCHEMISTRY OF REACTION PATHS AT CARBONYL CENTERS. *Tetrahedron* **1974,** *30* (12), 1563.

19.   Kleywegt, G. J.; Bergfors, T.; Senn, H.; Lemotte, P.; Gsell, B.; Shudo, K.; Jones, T. A., CRYSTAL-STRUCTURES OF CELLULAR RETINOIC ACID-BINDING PROTEIN-I AND PROTEIN-II IN COMPLEX WITH ALL-TRANS-RETINOIC ACID AND A SYNTHETIC RETINOID. *Structure* **1994,** *2* (12), 1241.

20.   Vaezeslami, S.; Jia, X.; Vasileiou, C.; Borhan, B.; Geiger, J. H., Structural analysis of site-directed mutants of cellular retinoic acid-binding protein II addresses the relationship between structural integrity and ligand binding. *Acta Crystallographica Section D-Biological Crystallography* **2008,** *64*, 1228.

21.   Lin, S. W.; Kochendoerfer, G. G.; Carroll, K. S.; Wang, D.; Mathies, R. A.; Sakmar, T. P., Mechanisms of spectral tuning in blue cone visual pigments

- Visible and raman spectroscopy of blue-shifted rhodopsin mutants. *Journal of Biological Chemistry* **1998,** *273* (38), 24583.

22.   Kiser, P. D.; Golczak, M.; Palczewski, K., Chemistry of the Retinoid (Visual) Cycle. *Chemical Reviews* **2014,** *114* (1), 194.

23.   Sakmar, T. P.; Franke, R. R.; Khorana, H. G., GLUTAMIC ACID-113 SERVES AS THE RETINYLIDENE SCHIFF-BASE COUNTERION IN BOVINE RHODOPSIN. *Proceedings of the National Academy of Sciences of the United States of America* **1989,** *86* (21), 8309.

24.   Fu, H.-Y.; Lin, Y.-C.; Chang, Y.-N.; Tseng, H.; Huang, C.-C.; Liu, K.-C.; Huang, C.-S.; Su, C.-W.; Weng, R. R.; Lee, Y.-Y.; Ng, W. V.; Yang, C.-S., A Novel Six-Rhodopsin System in a Single Archaeon. *Journal of Bacteriology* **2010,** *192* (22), 5866.

25.   Ernst, O. P.; Lodowski, D. T.; Elstner, M.; Hegemann, P.; Brown, L. S.; Kandori, H., Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chemical Reviews* **2014,** *114* (1), 126.

26.   Baasov, T.; Sheves, M., ALTERATION OF PKA OF THE BACTERIORHODOPSIN PROTONATED SCHIFF-BASE - A STUDY WITH MODEL COMPOUNDS. *Biochemistry* **1986,** *25* (18), 5249.

27.   Chen, D. L.; Wang, G. Y.; Xu, B.; Hu, K. S., All-trans to 13-cis retinal isomerization in light-adapted bacteriorhodopsin at acidic pH. *Journal of Photochemistry and Photobiology B-Biology* **2002,** *66* (3), 188.

28.   Marti, T.; Rosselet, S. J.; Otto, H.; Heyn, M. P.; Khorana, H. G., THE RETINYLIDENE SCHIFF-BASE COUNTERION IN BACTERIORHODOPSIN. *Journal of Biological Chemistry* **1991,** *266* (28), 18674.

29.   Nishikawa, T.; Murakami, M.; Kouyama, T., Crystal structure of the 13-cis isomer of bacteriorhodopsin in the dark-adapted state. *Journal of Molecular Biology* **2005,** *352* (2), 319.

30.   Steinberg, G.; Ottolenghi, M.; Sheves, M., PK(A) OF THE PROTONATED SCHIFF-BASE OF BOVINE RHODOPSIN - A STUDY WITH ARTIFICIAL PIGMENTS. *Biophysical Journal* **1993,** *64* (5), 1499.

31.   Ebrey, T. G., pK(a) of the protonated Schiff base of visual pigments. *Vertebrate Phototransduction and the Visual Cycle, Part A* **2000,** *315*, 196.

32.   Rousso, I.; Friedman, N.; Sheves, M.; Ottolenghi, M., PK(A) OF THE PROTONATED SCHIFF-BASE AND ASPARTIC-85 IN THE

BACTERIORHODOPSIN BINDING-SITE IS CONTROLLED BY A SPECIFIC GEOMETRY BETWEEN THE 2 RESIDUES. *Biochemistry* **1995,** *34* (37), 12059.

33.   Sakmar, T. P.; Franke, R. R.; Khorana, H. G., THE ROLE OF THE RETINYLIDENE SCHIFF-BASE COUNTERION IN RHODOPSIN IN DETERMINING WAVELENGTH ABSORBENCY AND SCHIFF-BASE PKA. *Proceedings of the National Academy of Sciences of the United States of America* **1991,** *88* (8), 3079.

34.   Sheves, M.; Albeck, A.; Friedman, N.; Ottolenghi, M., CONTROLLING THE PKA OF THE BACTERIORHODOPSIN SCHIFF-BASE BY USE OF ARTIFICIAL RETINAL ANALOGS. *Proceedings of the National Academy of Sciences of the United States of America* **1986,** *83* (10), 3262.

35.   Terakita, A.; Koyanagi, M.; Tsukamoto, H.; Yamashita, T.; Miyata, T.; Shichida, Y., Counterion displacement in the molecular evolution of the rhodopsin family. *Nature Structural & Molecular Biology* **2004,** *11* (3), 284.

36.   Gill, S. C.; Vonhippel, P. H., CALCULATION OF PROTEIN EXTINCTION COEFFICIENTS FROM AMINO-ACID SEQUENCE DATA. *Analytical Biochemistry* **1989,** *182* (2), 319.

CHAPTER III: SYNTHESIS OF FLUORESCENT RETINAL ANALOGS AS ALTERNATIVE LIGANDS FOR ENGINEERED PROTEINS

III.1 Introduction

Site-specific protein modifications have shown great utility in modern chemical biology.[1] Recently, a great number of versatile protein labeling techniques have been developed. Typically, the protein used for site-specific labeling is engineered from an enzyme with well-known function and high specificity. The inherent enzymatic activity of the protein is suppressed by mutagenesis and labeling is achieved by the addition of a fluorescent molecule with a binding domain from the natural substrate (alkyl halide for Halo-Tag, benzylguanine for SNAP-Tag, benzylcytosine for CLIP-Tag, trimethoprim for TMP-Tag, etc.).[2-5] In an alternative approach, the short peptide can be labeled with a substrate-fluorophore conjugate, assisted by the engineered enzyme (formyl glycine generating enzyme, transglutaminase, biotin ligase, lipoic acid ligase, farnesyl transferase, etc.).[6-10] A common approach is to redesign the substrate with an appropriate reporter unit, such that its binding affinity to the enzyme and subsequent enzymatic reaction is retained in a single turnover manner, and thus covalently attaching the reporter unit to the protein (**Figure III-1**). The most common reporter unit is a fluorophore that allows direct imaging with high sensitivity. The linker length between the two units often affects protein affinity, solubility and cell penetration. The linker is incorporated into the molecule (usually in the form of a length-adjustable polyether) or it is formed *via* common

bioorthogonal reactions between the recognition unit and the reporter unit. Such reactions include the alkyne-azide 'click' chemistry, Staudinger ligation, inverse-demand Diels-Alder reaction, oxanobornadiene cycloaddition, Pictet-Spengler reaction and other rapidly emerging biocompatible synthetic methods.[11-13 1, 14-17]

A plethora of commercially available fluorophores exists for protein labeling, with several recent reports detailing molecules with improved spectroscopic profiles.[18-25] In general, fluorophores can be divided into groups based on their core structures. These include xanthanes, coumarins, fluoresceins, rhodamines, BODIPY dyes and cyanines, among others.[26] To avoid
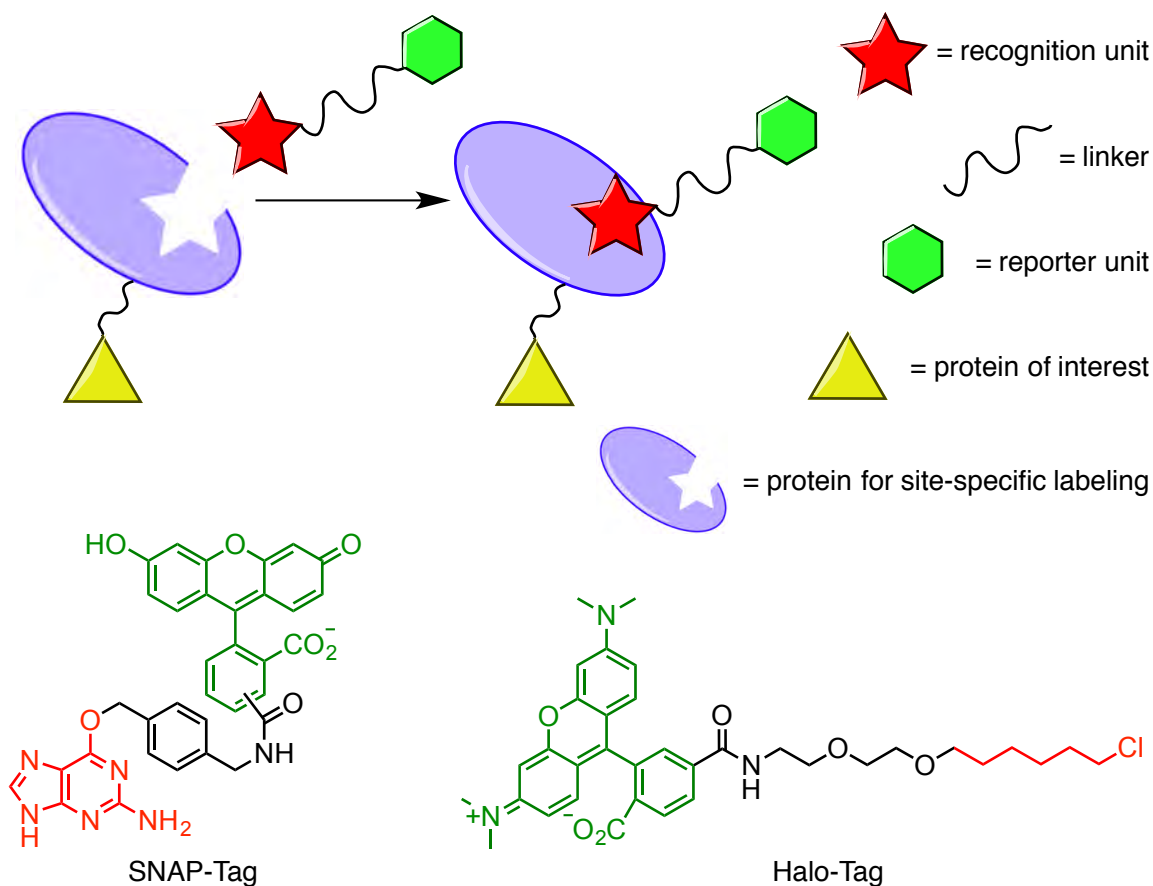
Figure III-1: Typical strategy of the site-specific protein labeling highlighting each unit of designed labeling reagent and two examples of such reagent with parts of the molecule colored identical to the representation.

a constitutively 'on' fluorescence that arises from non-specific fluorophore incorporation, a masking cargo is employed usually in the form of quenchers or intracellular cleavable protecting groups. More recently described reports utilize environment-sensitive fluorophores with spectroscopic properties (excitation and emission wavelength, quantum efficiency, extinction coefficient, etc.) that differ significantly inside and outside the protein cavity.

For the past ten years our lab has been actively working on the development of a site-specific protein labeling method.[27-33] In our approach, all of the units discussed above are incorporated into one small molecule, where the protein specific substrate is also a chromophore. Iminium formation between the aldehyde based chromophore and a reactive lysine residue in the protein cavity leads to fluorescence 'turn on' (**Figure III-2**). The by-product of this reaction, a molecule of water, renders this method highly benign. Specificity arises from the high affinity of the hydrophobic polyene chain for the lipophilic protein pocket, observed in engineered CRABPII and CRBPII that bind all-*trans*-retinal as the Protonated Schiff Base (PSB). In addition to the dual role of the polyene,
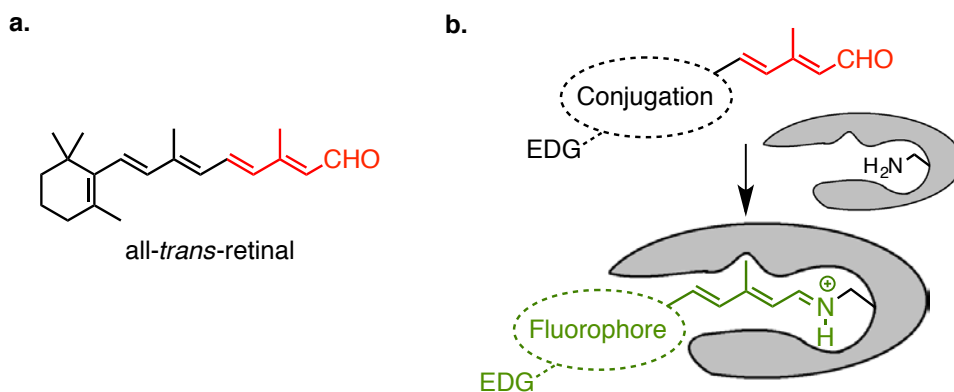
Figure III-2: Site-specific labeling of rhodopsin mimics. **a.** Structure of all-*trans*-retinal recognition unit highlighted in red. **b.** Design of alternative ligands that fluoresce upon PSB formation.

153

conversion of the aldehyde to an iminium in retinal (and its analogs) dramatically changes its spectroscopic properties. The extended conjugation of this polyene lowers the HOMO/LUMO gap, leading to a more red-shifted chromophore. The iminium, with its own distinct spectroscopic profile, is the only fluorescent species observed in imaging experiments, because the aldehyde and imine are non-emissive.

Synthetic chromophores bearing a conjugated aldehyde with a 3-methyl group belong to the family of retinoids.[34] Synthetic retinoids have been used in wavelength and pK_a regulation studies[35] and photo-affinity labeling of rhodopsin,[36] and also in the search for the therapeutics.[34] Synthesis of retinoids is complicated by the sensitivity of these compounds to various factors, such as light and heat. Moreover, even if conjugated double bonds are prepared in a highly stereoselective manner, they can still undergo isomerization if not handled carefully.[34] Nevertheless, the synthesis of retinoids can be accomplished *via* multiple pathways using reported protocols.

Merocyanine retinal analogs have been studied for the wavelength regulation of bacteriorhodopsin.[37] After PSB formation these compounds display a bathochromic shift of over 100 nm, with an increase in extinction coefficient. These compounds, owing to their resonating positive charge within a strong push-pull system, have the potential to become fluorescent in the PSB form. The possible fluorogenic character of the merocyanine retinal analogs in CRABPII

and hCRBPII mutants lend to the idea that merocyanine retinal analogs can be used in protein labeling efforts.

A number of available mutants in our lab were evaluated for PSB formation with fluorogenic retinal analogs. A quick screen of candidates from various CRABPII and hCRBPII series aimed for a system with the following properties: fast reaction time, high quantum efficiency, red-shifted spectra of excitation/emission and stability of the complex. Access to high-resolution crystal structures should be considered for further protein and/or ligand optimization.

The first potentially fluorescent retinal analog synthesized in our lab was merocyanine retinal analog (MCRA) (**Figure III-3**).[38] The same compound was previously described in wavelength regulation studies upon binding to bacteriorhodopsin.[37] The fluorescent properties of MCRA-PSB were elucidated for the first time in our lab, unveiling their high quantum yields and extinction coefficients only after reaction with the protein (**Figure III-3**) (Dr. Kin Sing Stephen Lee, Dr. Wenjing Wang and Ipek Yapici, Professor Babak Borhan's lab,
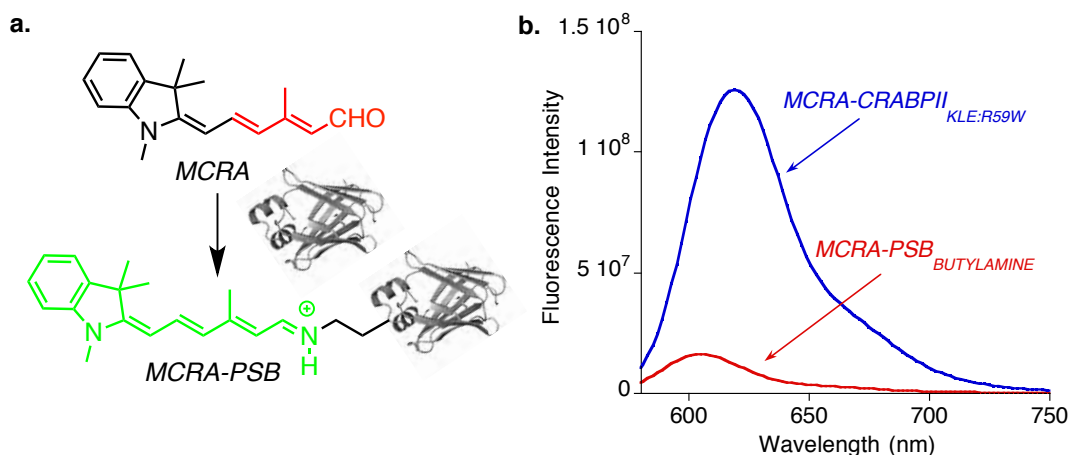
Figure III-3: Labeling of designed proteins with MCRA. **a.** Reaction that triggers fluorescence 'turn-on' (recognition unit red, fluorophore green). **b.** Fluorescence enhancement of protein bound dye relative to PSB with *n*-butyl amine.

155

MSU).

The MCRA-PSB protein complexes absorbed between 565 nm and 605 nm. Even more fascinating is that emission of MCRA-PSB inside the protein is five times higher, as compared to the free amine in solution (**Figure III-3b**). The latter finding was expected to suppress the signal from any non-specific labeling with free intracellular amines.

Moreover, MCRA is small compared to other protein labeling reagents, which might be helpful for the chromophore delivery into the cell. The absence of toxic additives and by-products of the reaction (iminium formation gives elimination of water only) are among other notable advantages of this method.

Finally, if the wavelength of MCRA-protein complex is significantly red-shifted compared to the MCRA-butylamine conjugate, the non-specific fluorescence of PSB with intracellular amines can be eliminated. Taking into account the wavelength regulation studies with retinal, this approach seems as a plausible way to avoid background fluorescence.

Selected CRABPII mutants were successfully imaged in live *E. coli* cells, while hCRBPII showed utility in HeLa and U2OS cells (detailed discussion in Chapter 4). However, MCRA binds as a PSB to both the specifically engineered CRABPII and hCRBPII mutants and free amines, as observed in images.

To understand the nature of non-specific reactivity, *in vitro* binding was measured between MCRA and Bovine Serum Albumin (BSA), a protein standard often used in biochemical assays. BSA is a plasma protein, and doesn't bind

retinal or its analogs. However, it has multiple hydrophobic binding sites for various metabolites, which makes BSA suitable for control experiments for non-specific reactivity. In fact, the absorption/emission spectra and reaction rate of MCRA-BSA complex were comparable with some tested CRABPII and CRBPII mutants. The study of non-specific binding can be approached in two ways: *via* synthesis of alternative ligands that bind our target proteins specifically (described in this Chapter) or by further protein engineering leading to mutants that bind our fluorophores specifically (described in Chapter 4).

In addition to the specificity issues encountered during *in vivo* protein labeling, MCRA is not the ideal ligand for fluorescent pH-sensor design. The mutant with one of the lowest iminium pK_a values for Rt-PSB (CRABPII-R111K:R132Q:Y134F:T54V:R59W:A32W:E73A:M93L, pK_a=2.6) has a pK_a of 8.6 in the MCRA-PSB (**Figure III-4**). Because the physiologically relevant pH range is below 8, MCRA-PSB does not meet this requirement.
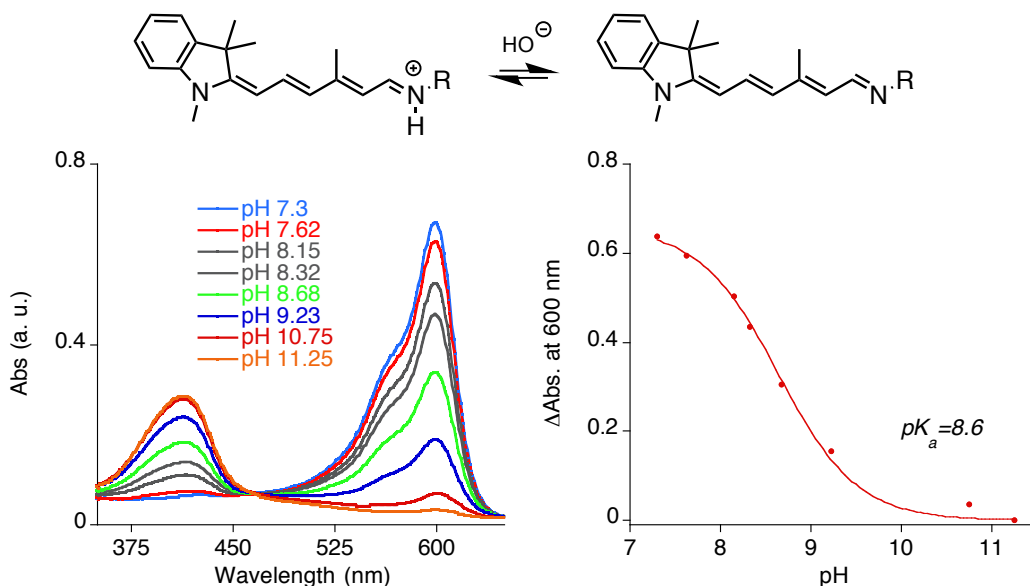
Figure III-4: Base titration of CRABPII-R111K:R132Q:Y134F:T54V:R59W:A32W:M93L:E73A mutant bound with MCRA.

157

The latter concerns about MCRA as a fluorogenic ligand led to the synthesis of a suitable retinal analog for a PSB-forming protein. In this chapter, synthetic approaches toward alternative fluorescent retinal analogs are described. Also an alternative electrophilic trapping in the tandem aza-Payne-like/hydroamination methodology was inspired by an observation during the synthesis of one of the fluorophores with the properties of a Michael acceptor. The synthetic methodology developed the construction of the hexahydrofuro[2,3-c]pyrrole core in one step, from alkyne-containing aziridinol and benzylidenemalononitriles.

III.2 Synthesis of julolidine-containing retinal analogs

Several known fluorophores with high quantum efficiencies contain a tertiary nitrogen in their core. The lone pair of the tertiary amine can participate in conjugation with electron withdrawing groups located on the other end of the molecule, which results in an extended conjugation. If the tertiary amine is in the ring-locked form, the lone pair has a better chance to align with the π^* of the nearby double bond, therefore enhancing delocalization. In this respect, julolidine with a ring-locked nitrogen (**Figure III-5**) is a good candidate to replace the β-ionone ring in a fluorescent retinal analog.
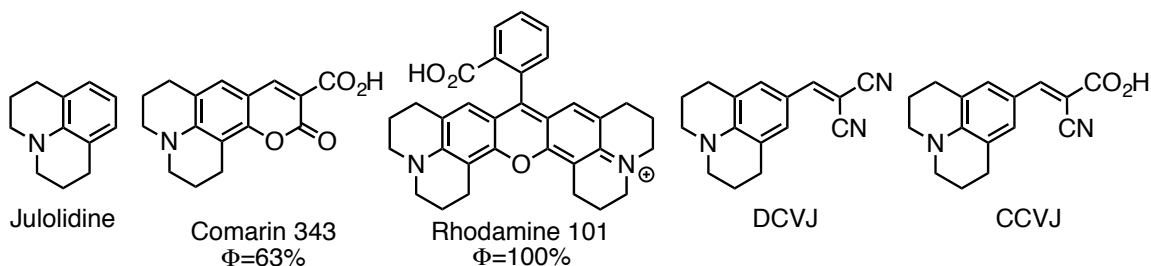
Figure III-5: Structures of known fluorophores with julolidine nucleus.

Usually julolidine-containing fluorophores have high quantum yield, Φ, (**Figure III-5**) and in some cases display solvatochromic absorption/emission profiles.[39] Moreover, the remarkable viscosity-dependent fluorescence of two julolidine chromophores was the main factor that caught our attention. Commercially available molecules, 9-(2-carboxy-2-cyanovinyl)-julolidine (CCVJ) and 9-(2,2-dicyanovinyl)-julolidine (DCVJ), show increased fluorescence intensity in more viscous solutions.[40-44] Molecules of this kind are called molecular rotors because their intrinsic fluorescence depends on rotation of the vinyl group in and out of the plane of julolidine.[45] With an increased viscosity, this rotation becomes less fluent, which leads to increased fluorescence.
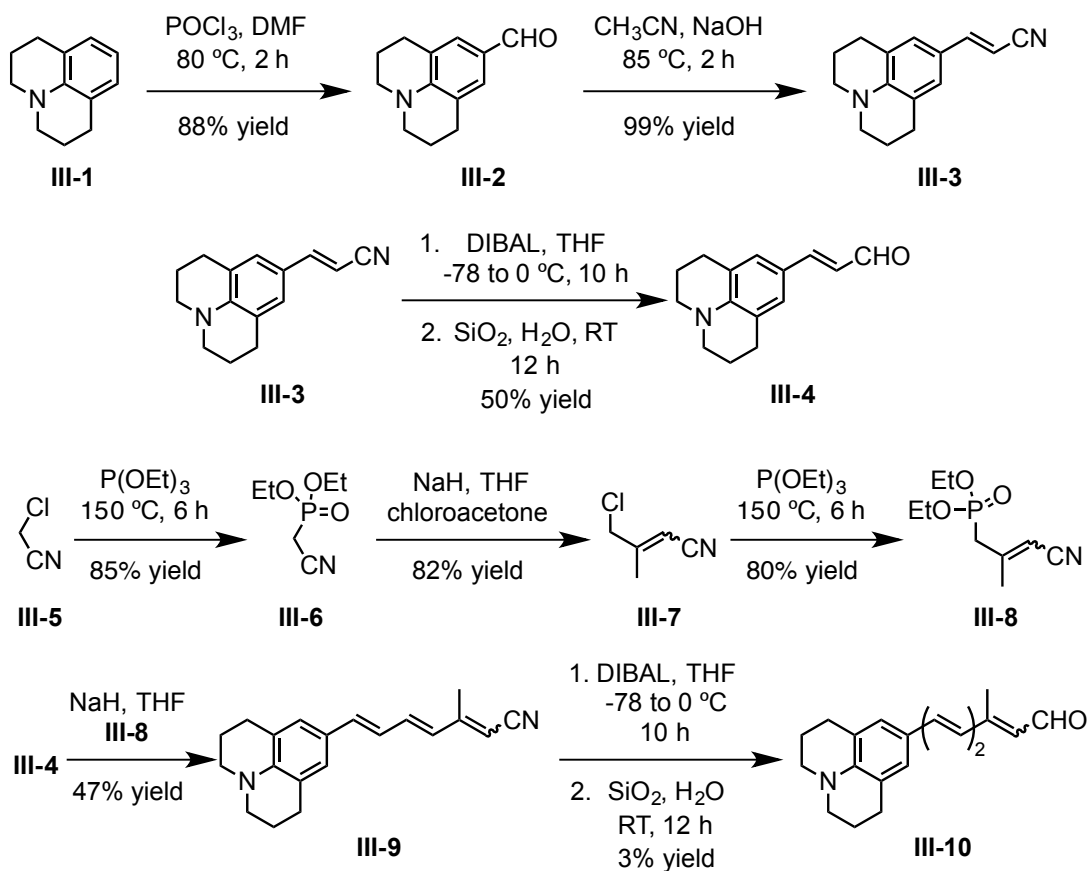
Scheme III-1: Synthesis of julolidine retinal analog **III-10**.

159

Hypothetically, the rotation along the C-C bond would be even more restricted inside the rigid protein cavity, where formation of the iminium may lead to an even more enhanced fluorescence. Findings from Armitage lab have already described an increase in emission intensity of the nitrile-containing cyanine dyes inside the protein.[22, 46] However, in the reported work the protein was selected from a library of random antibodies and labeling was successful only on the yeast surface.

Julolidine retinal analogs substituted with a nitrile group were targeted for synthesis. Additionally, julolidine retinal analogs without a nitrile were synthesized. The synthesis of nitrile-free chromophores is more trivial with repetitive sequences of Horner-Wadwords-Emmons (HWE) olefination and diisobutylaluminum hydride (DIBAL) reduction of the corresponding nitriles.[47]

Aldehyde **III-2** was synthesized from commercially available julolidine **III-1** through the Vilsmeier-Haack reaction in 88% yield.[48] Subsequently, it was
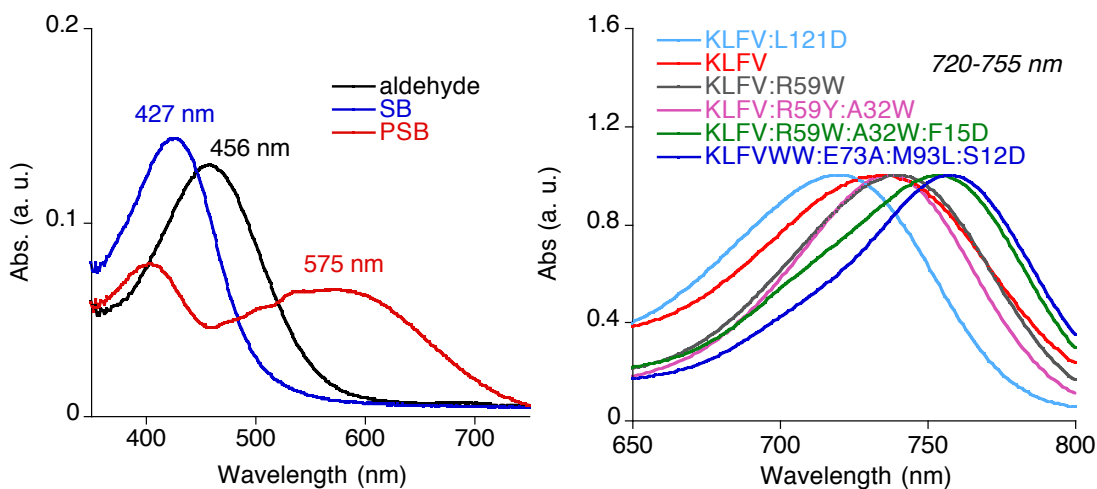
Figure III-6: UV-vis spectra of **III-10** as an aldehyde, SB and PSB with *n*-butylamine (top) and normalized spectra of the CRABPII mutants bound to **III-10** (bottom).

converted quantitatively into the nitrile **III-3** by refluxing in acetonitrile in the presence of NaOH. The DIBAL reduction of this nitrile **III-3** followed by work-up with wet silica provided aldehyde **III-4** in 50% yield.

The synthesis of ylide **III-8** was completed in three steps from chloroacetonitrile **III-5** according to a reported protocol.[47] Phosphonate **III-6** was prepared *via* an Arbuzov reaction, as shown in **Scheme II-1**. HWE reaction of **III-6** with chloroacetone yielded **III-7**. This was further converted to the final ylide **III-8** *via* another Arbuzov reaction, leading to a mixture of two isomers in a ratio 4:3.

In the next step, aldehyde **III-4** and phosphonate **III-8** were converted into nitrile **III-9** under HWE conditions. The mixture of isomers was subjected to DIBAL reduction and hydrolyzed to provide julolidine retinal analog **III-10**.

Absorption spectra of synthesized analog **III-10** and six selected CRABPII mutants, bound with **III-10**, are depicted in **Figure III-6**. The wavelength range for the PSB of **III-10** with selected proteins was 720-755 nm, which is modest compared to the range observed for the Rt-PSB with the same mutants (474-640 nm, **Figure II-22** in Chapter 2). Additionally, excitation beyond 633 nm is not compatible with the lasers currently used for fluorescent microscopy. Analog of the chromophore with one less double bond was hypothesized to be more useful in this respect.
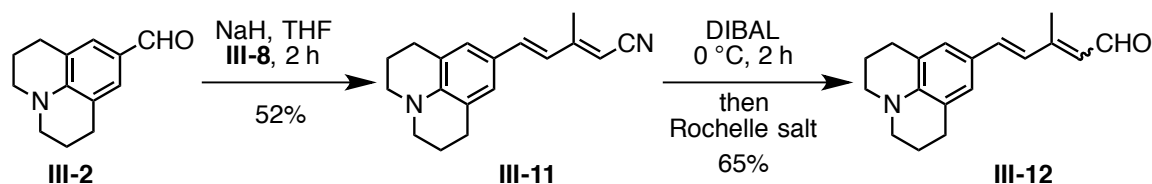
Scheme III-2: Synthesis of julolidine retinal analog **III-12**.

As depicted in **Scheme III-2**, the same julolidine aldehyde **III-2** was directly subjected to olefination with ylide **III-8**, yielding nitrile **III-11**. Recrystallization of the crude product from ethanol gave pure all-*trans* isomer as a bright yellow solid. DIBAL reduction of the all-*trans*- chromophore for 2 h followed up by mild work-up with Rochelle's salt gave good yield for the product. However, the final aldehyde **III-12** was still obtained as the mixture of two isomers, in the ratio 10:3.

Compound **III-12** was tested with a number of CRABPII and hCRBPII mutants. Detailed results of further protein engineering data are described in Chapter 5.

III.3 Synthetic approach to 2-cyano-julolidine retinal analog.

Among all possible ligands capable of PSB formation with our mutants, the main target was julolidine retinal analog with an additional nitrile group in conjugation with the aromatic ring. The challenging part of this synthesis is that the usual sequence of olefination and reduction would not be chemoselective in the presence of two nitriles. Previously published reports on CN-containing molecular rotors were not from the retinal analog family. However, synthetic protocols successfully employed aldol-type condensations. To prepare the CN-
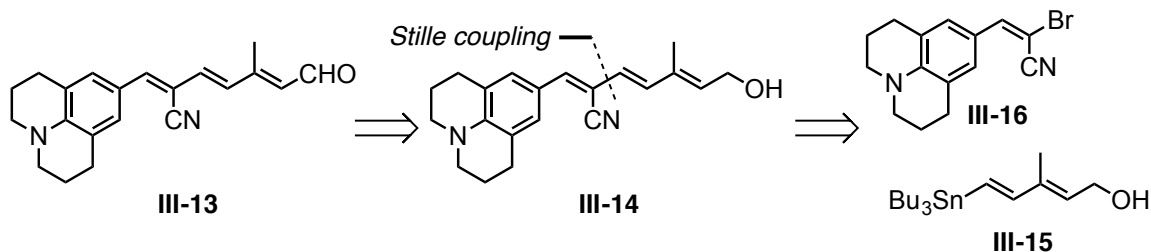
Figure III-7: Retrosynthetic analysis of **III-13** via Pd-catalyzed cross-coupling.

162

julolidine retinal analog, such as **III-13**, an alternative synthetic method had to be considered.

Initial synthetic strategies to access analog **III-13** relied on a variety of palladium-catalyzed Stille coupling methods between a vinyl halide and a vinyl stannane.[49-50] As the bond forming reaction, the Stille coupling was chosen because of having multiple protocols optimized for polyene synthesis.[51-52] The retrosynthetic analysis is depicted in **Figure III-7**, in which two synthons, **III-15** and **III-16**, would provide alcohol **III-14**. Subsequently, oxidation of **III-14** would lead to the targeted aldehyde **III-13**. The synthesis of the prerequisite vinyl tributyltin **III-15** was reported from enynol **III-17**.[53]

As depicted in **Scheme III-3** dibromination of **III-3**, followed by treatment with base gave the desired vinyl bromide **III-16** in 56% yield (two isomers 5:3). Treatment of enynol **III-17** with tributyltin hydride provided the vinyl stannane **III-15** but in much lower yield than the reported protocol.[53] After the bromide **III-16** was recovered under additive-free reaction conditions, the Stille coupling was attempted with moderate heating in the presence of CsF.[54] The reaction proceeded with complete bromide consumption within 2 h. However, the reaction

Entry	Additive	Time	Temperature	Result
1	None	10 h	80 °C	n. r.[a]
2	CsF, CuI	3 h	80 °C	c. m.[b]
3	CsF, CuI	3 h	45 °C	c. m.
4	CsF, CuI	1 h	45 °C	c. m.

[a]No reaction. [b]Complex mixture.
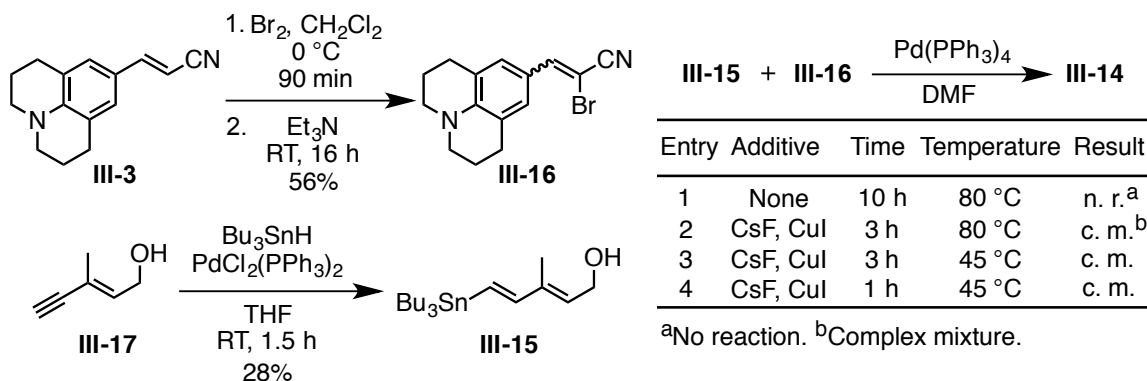
Scheme III-3: Synthesis of **III-14.**

gave a complex mixture of products. Crude NMR analysis did not reveal any trace of the desired product.

An alternative approach to access the same analog **III-13** was inspired by the chemistry of Zincke aldehydes.[55-56] Generally, Zincke aldehydes are prepared from a pyridinium salt and secondary amine, as depicted for picolinium **III-18** and dimethylamine in **Scheme III-4**.[57-58] However, there was no report on successful pyridinium opening with nucleophiles other than secondary amines. Our approach utilized a Morita-Baylis-Hillman zwitterionic intermediate as a nucleophile generated from nitrile **III-3** (**Scheme III-4**). If successful, it would become a one step procedure to the final chromophore with an opportunity to extend the methodology.

The required pyridinium salt was prepared by S_NAr reaction between 4-picoline and 2,4-dinitrochlorobenzene.[59] The pyridinium opening with zwitterionic enolate was evaluated in the presence of various amine bases (**Scheme III-4**). Unfortunately, conditions leading to the desired product were not found.

| III-3 + III-18 | Toluene, RT / Base | III-13 |

Entry	Base	Time	Result
1	DABCO	2 d	n. r.[a]
2	DBU	2 d	n. r.
3	DMAP	2 d	n. r.
4	DABCO	10 d	n. r.

[a]n. r. = no reaction
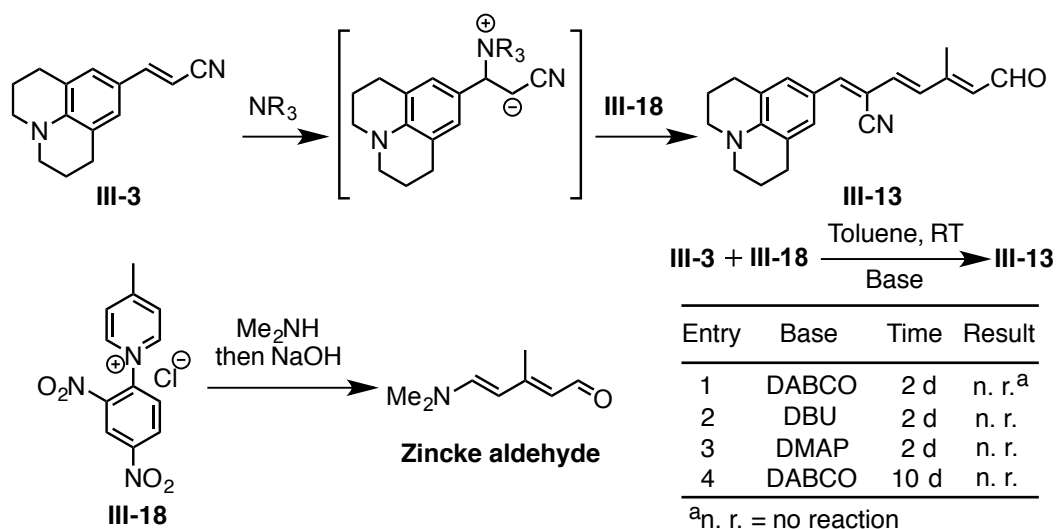
Scheme III-4: Synthesis of **III-13** *via* stabilized anion addition to Zincke salt.

An alternate approach to obtain analog **III-20** (**Scheme III-5**) utilized an addition/elimination protocol known to work for C-C bond formation between enolates and vinyl iodide **III-19**.[60] An attempt was made to form the zwitterionic aza-enolate, which would participate in the conjugate addition with vinyl iodide **III-19**, to give the intermediate **III-21**. Subsequent elimination of ammonium salt was expected to yield the desired product **III-20**. Disappointingly, multiple attempts failed to isolate the chromophore.

Formation of the zwitterionic nucleophile from the vinyl cyanide **III-3** might be unlikely because of the electron-rich character of the double bond that is conjugated with the tertiary amine. Therefore, the synthesis of **III-20** was attempted starting from julolidine aldehyde **III-2**. Ylide **III-22** was anticipated from the reaction of cyanomethylphosphonate **III-6** and vinyl iodide **III-19**. First, the phosphoylide **III-6** was reacted with the vinyl iodide **III-19** until vinyl iodide was consumed, as evident from TLC and ES[+] analysis (m/z 297 amu corresponding to sodium salt was observed, **Scheme III-5**). Aldehyde **III-2** was then added *in situ* to the reaction mixture. Unfortunately, after work-up only nitrile **III-3** was
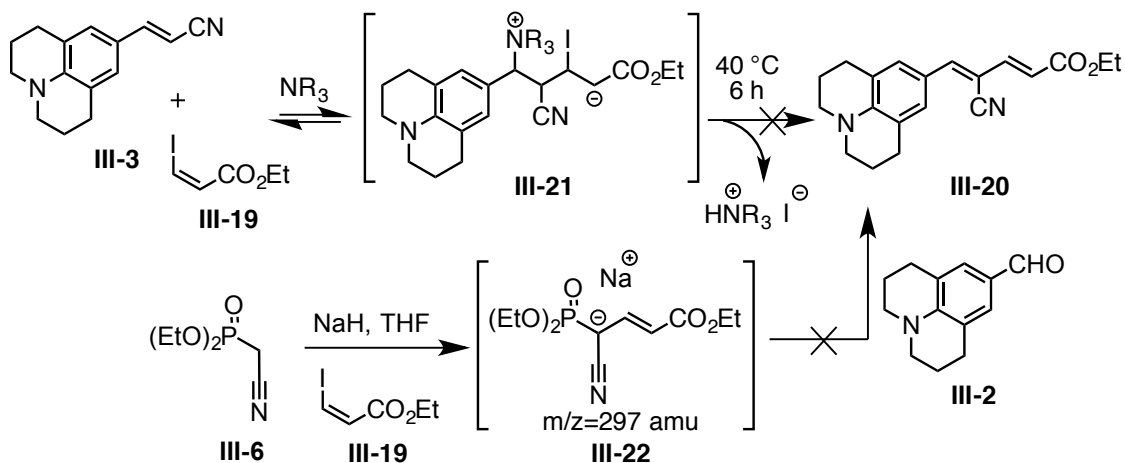
Scheme III-5: Two routes towards synthesis of **III-20**.

165

isolated, the result of HWE of **III-2** with **III-6**. Attempts to synthesize and isolate the corresponding ylide **III-22** were unsuccessful.

Another route involved the formation of **III-23,** hypothetically convertible into aldehyde **III-24**, with commercially available ylide **III-25**. Compound **III-21** can be synthesized in two ways (**Scheme III-6**). DIBAL reduction of dicyanide with 0.5 equivalents of reducing agent gave three different products from TLC. However, only **III-23** was isolated in 25% yield. Formylation of the double bond, starting from nitrile **III-3**, with the Vilsmeier-Haack reagent proceeded exclusively to aldehyde **III-23**. Both reactions gave the product with identical NMR spectra for a single isomer. Disappointingly, reaction between the aldehyde **III-23** and the stabilized ylide **III-25** had no conversion even after 2 days, while multiple attempts of HWE olefination with **III-6** gave a complex mixture of undefined products, with no recovery of the starting materials.

This could be due to the presence of multiple electrophilic centers in the molecule **III-23**. While the aldehyde carbon is electrophilic, the presence of two electron withdrawing groups makes **III-23** also a Michael acceptor. This has also been suggested previously, where molecules with a similar substitution pattern
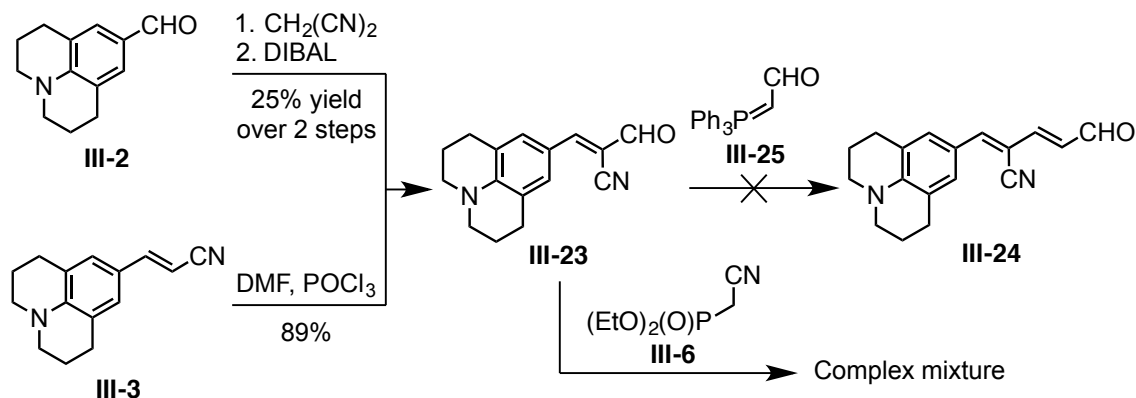
Scheme III-6: Synthetic approach to **III-24**.

166

prefer to undergo 1,4- over 1,2-addition even with strong nucleophiles, such as Grignard reagents.[61-63] Although this approach did not find utility for the synthesis of the desired fluorophore, it inspired a methodology described later in this chapter.

Although not a targeted chromophore, aldehyde **III-23** might have some potential to bind to the PSB-forming mutants, while being less reactive with BSA or butyl amine (**Figure III-8a** and **III-8b**, respectively). RandomLy selected mutants from each series were screened for PSB formation with aldehyde **III-23**.
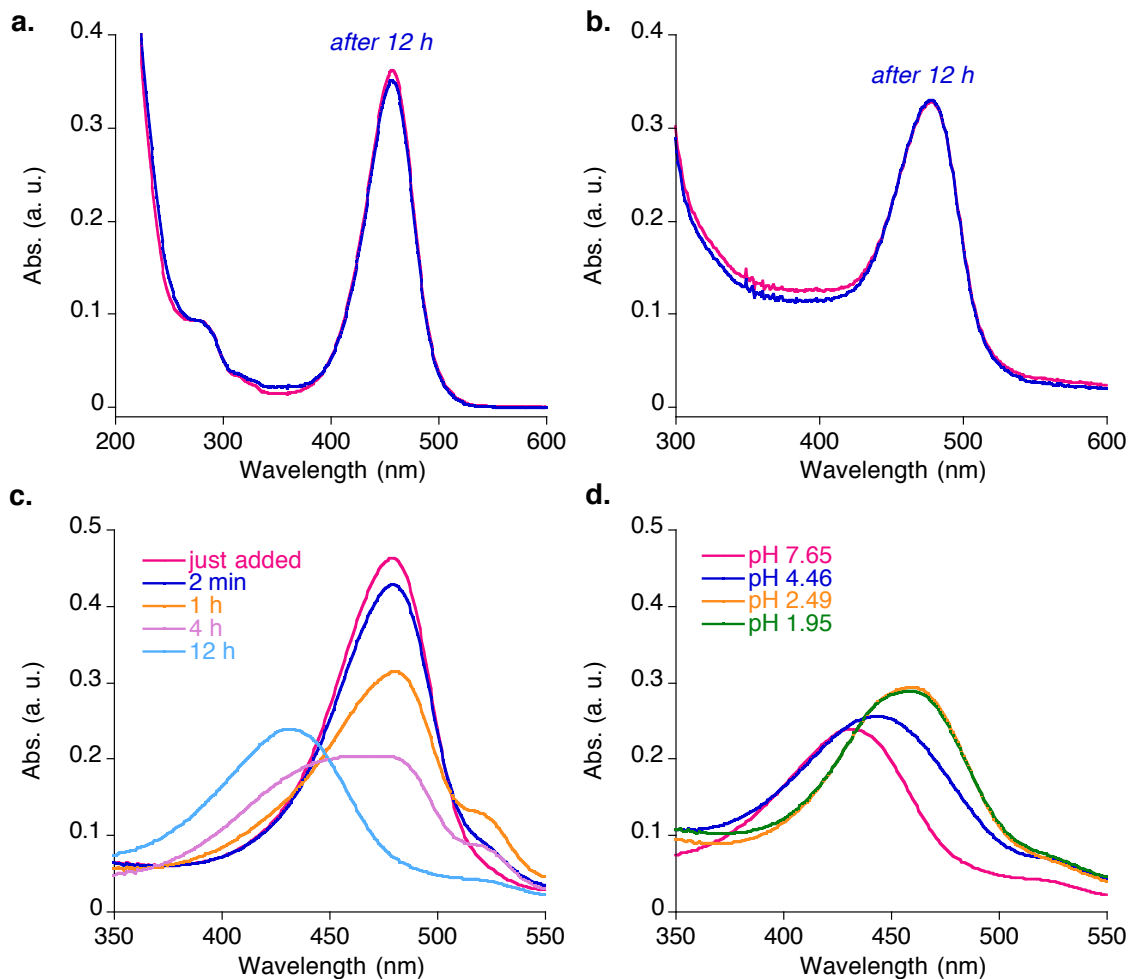
Figure III-8: Binding of **III-23** with **a.** *n*-butyl amine (in ethanol); **b.** bovine serum albumin, BSA; **c.** CRABPII-R132K:R111L:L121E:R59W:A32W mutant and **d.** acid titration of the complex from **c.**

One of the mutants from the CRABPII first generation series (more exposed binding cavity) showed evidence for SB/PSB formation after 12 h of incubation at room temperature (**Figure III-8c**). It is not surprising that only the CRABPII mutant with more accessible nucleophilic lysine, R132K, is capable of the bond formation rather than any other CRABPII or CRBPII mutant with a deeper binding cavity. The large julolidine group might be too bulky to get into the rigid protein cavity. Even though bond formation for CRABPII-R132K:R111L:L121E:R59W:A32W mutant was clearly evident from UV-vis, it formed only a small amount of PSB, relative to SB, even after acidification to pH 2 (**Figure III-8d**).

The final approach taken to synthesize the desired molecule was inspired by a formylation reaction on the double bond, instead of the julolidine aromatic ring. The vinylogous Friedel-Crafts acylation of cyano julolidine **III-3** provided a significant amount of methylketone **III-26**. Ketone **III-26** was subjected to the formylation reaction with POCl$_3$/DMF mixture. Because chlorine and methyl groups have similar Van der Waals radii (1.8 Å and 2.0 Å, respectively), both are expected to bind equally well to the designed retinal binding proteins. The product **III-27** can be considered as a retinal analog (Cl-CJRA) and it became the targeted molecule to be tested for PSB formation (**Figure III-9**). Unfortunately,
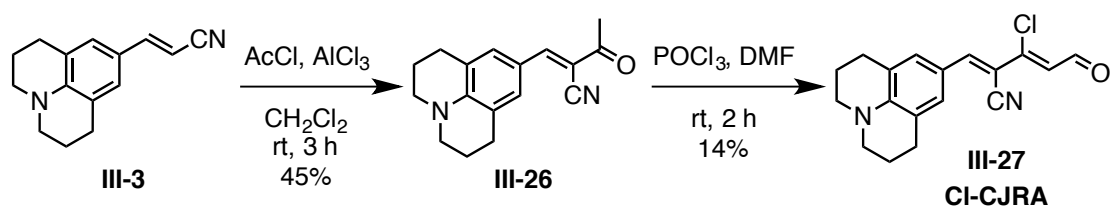
Scheme III-7: Synthesis of chloro-cyano-julolidine retinal analog (Cl-CJRA).

binding of Cl-CJRA with a series of mutants proceeded only to form SB (λ_{max} = 470 nm) and not the PSB (λ_{max} > 605 nm). In the mutants capable of PSB formation, the peak was completely converted to SB over time.

In addition to this, recent reports on spectroscopic properties of a similar fluorophore (α-CN-TO) mentioned the two-fold drop in extinction coefficient, compared to initial thiazol orange (TO) (**Figure III-10**).[22] The MCRA-PSB complexed with proteins have extinction coefficients of $\sim$150,000 $M^{-1} \cdot cm^{-1}$.
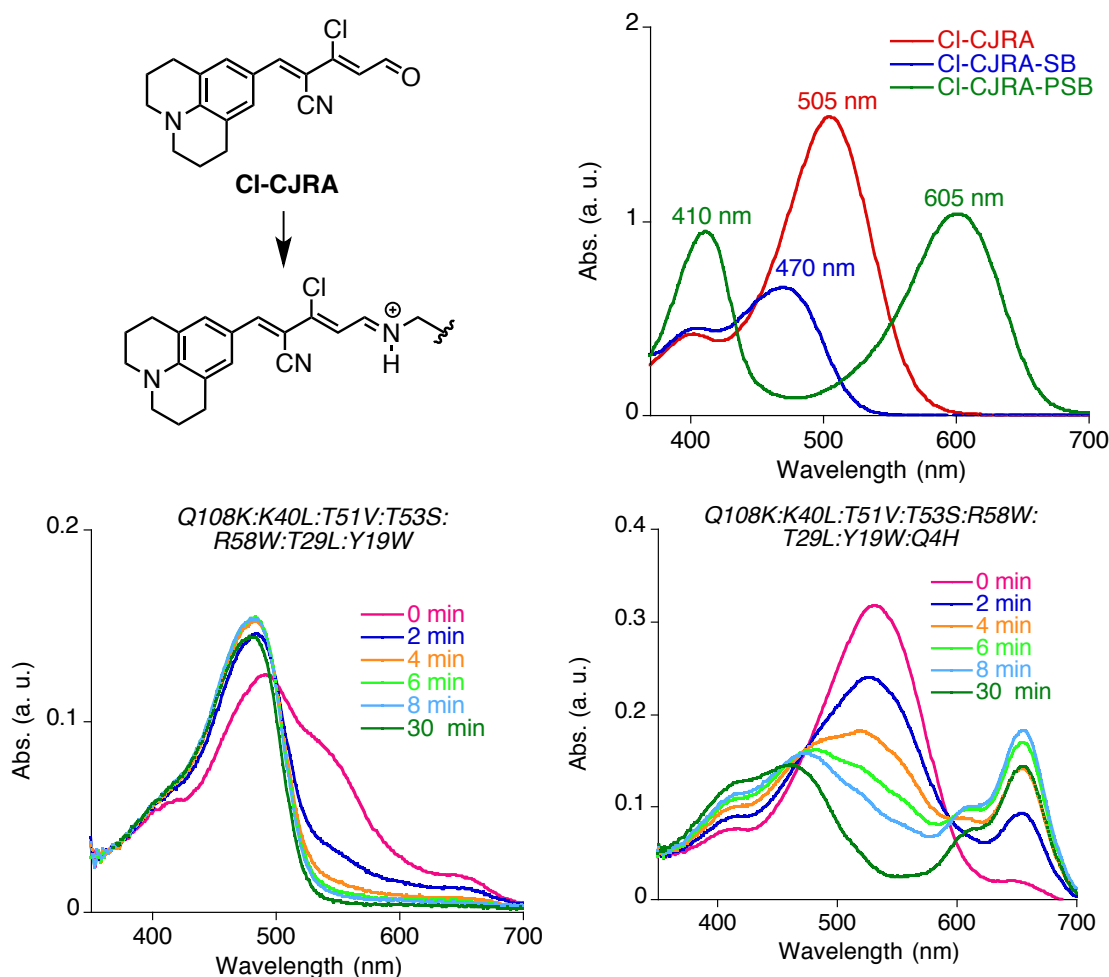
Figure III-9: UV-vis spectra of Cl-CJRA as an aldehyde, SB and PSB with *n*-butylamine (top) and binding spectra of the highlighted hCRBPII mutants (each 2 min).

Considering the approximations by Armitage and coworker, the cyano substituted MCRA would have a substantially reduced ε, predicted to be half of MCRA ($\sim$75,000 M·cm$^{-1)}$. Even so, this would still yield sufficient brightness (multiplication of quantum yield and extinction coefficient). However, the synthesized julolidine retinal analog **III-12** has a significantly lower extinction coefficient than MCRA-PSB (**Figure III-9**). If the structure **III-12** contained a cyanide substitution, the ε for its PSB would be $\sim$23,000 M^{-1}·cm^{-1}. This prediction suggests the formation of a dim complex with CRABPII and hCRBPII mutants.

The disappointingly low pK_a values for **III-23**-PSB and **III-27**-PSB with the tested mutants and prediction about extinction coefficient curtailed further attempts to synthesize cyano-julolidine retinal analogs. In the future, one might promote the protonation in the protein cavity by rational protein engineering. In such a scenario, non-specific binding would be in the SB form, having very

TO
ε = 60,000 cm^{-1}·M^{-1}

α-**CN-TO**
ε = 30,000 cm^{-1}·M^{-1}

MCRA-PSB
ε $\sim$150,000 cm^{-1}·M^{-1}
Φ up to 42%
in CRABPII and hCRBPII mutants

III-12-PSB
ε $\sim$46,000 cm^{-1}·M^{-1}
Φ < 17%
in CRABPII and hCRBPII mutants
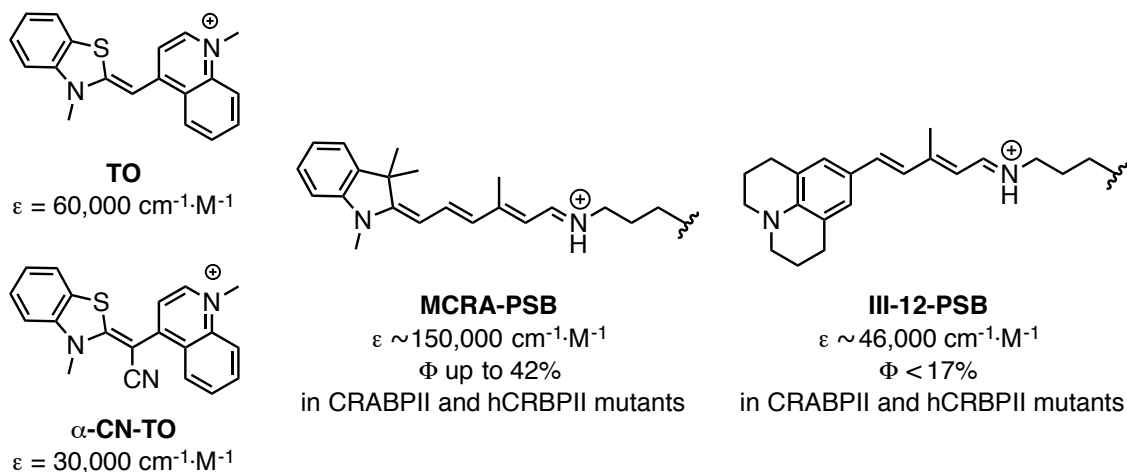
Figure III-10: Difference of extinction coefficient in the dyes with and without cyanide and the values for synthetic compounds.

different spectroscopic properties. Subsequently, only the engineered protein would form the PSB with a unique excitation wavelength. Additionally, Cl-CJRA has two reactive sites: an aldehyde for SB/PSB formation and chlorinated carbon that can undergo Michael addition with concomitant elimination. Michael addition in biology is well-known and usually occurs *via* the reaction of cysteine as a nucleophile. In addition to an increased specificity, it would give an irreversibly trapped adduct, beneficial for complex stability. Furthermore, eliminating the electron-withdrawing -Cl will presumably increase the pK_a of the PSB leading to an increased amount of iminium. If successful, this approach can reduce the background fluorescence in multiple ways: lower the pK_a of PSB for non-specific binding, non-fluorescent state of any chromophore-PSB other than in CRABPII or hCRBPII binding pocket and enhanced specificity to the protein controlled by dual reactivity.

III.4 Synthesis of merocyanine retinal analog and methyl ketone merocyanine retinal analog

Merocyanine retinal analog **III-30** (MCRA) was previously synthesized and tested with the first generation CRABPII mutants by Dr. Kin Sing Stephen Lee (Babak Borhan's lab, MSU). Since then, hundreds of CRABPII and CRBPII mutants have been screened in PSB formation and fluorescence with MCRA. The project was extended further to live-cell imaging experiments. Availability of MCRA limits all ongoing experiments, including crystallographic analysis of holo-mutants, so, more chromophore was synthesized. We also wanted to evaluate a

related chromophore, methyl ketone merocyanine retinal analog **III-31** (mkMCRA), prepared through a common intermediate nitrile **III-29**. The methyl ketone **III-31** should be less reactive than the aldehyde **III-30** during iminium formation. Additionally, it is possible to reengineer the PSB forming mutants to perform a much faster reaction that outcompetes non-specific iminium formation *in vivo*. The synthesis of nitrile **III-29** was achieved in one step from the corresponding aldehyde **III-28** and the ylide **III-8**. Reduction of the nitrile **III-29** with DIBAL gave MCRA, while methylation with methyllithium provided mkMCRA as a mixture of two isomers (3:4) (**Scheme III-8**).

Isolated mkMCRA was enough for preliminary tests of specificity in binding with the engineered proteins. Selected mutants from the CRABPII and hCRBPII series of rhodopsin mimics were incubated with mkMCRA. In parallel, mkMCRA was evaluated in binding with *n*-butylamine, followed by acidification to PSB. The latter reaction displayed no indicative change in the absorption spectra even after 12 h, while MCRA requires only 50 min to reach complete PSB formation (**Figure III-11a** and **11b**). This data confirms the significantly suppressed reactivity of the modified chromophore, which is a strong advantage to outcompete the non-
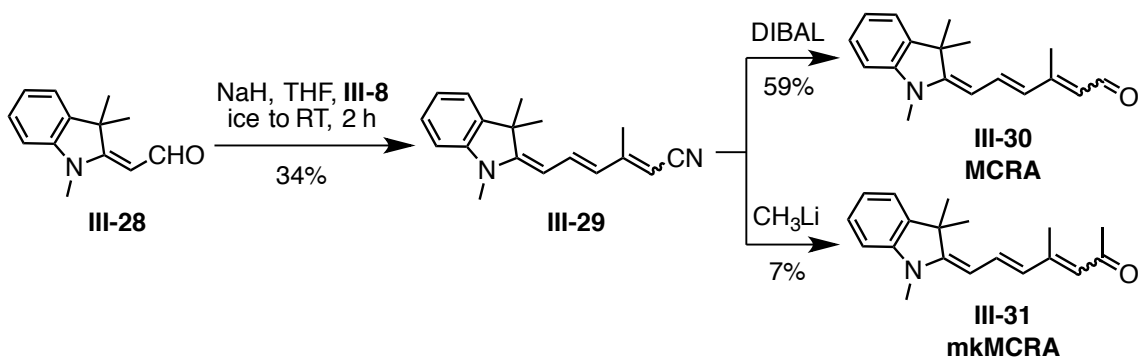
Scheme III-8: Synthesis of merocyanine retinal analog (MCRA) and methyl ketone merocyanine retinal analog (mkMCRA).

172

specific PSB formation.

Selected mutants for initial screening of the specific reaction were CRABPII-R111K:R132L:Y134F:T54V:R59W:L121D and CRBPII-Q108K:K40L:
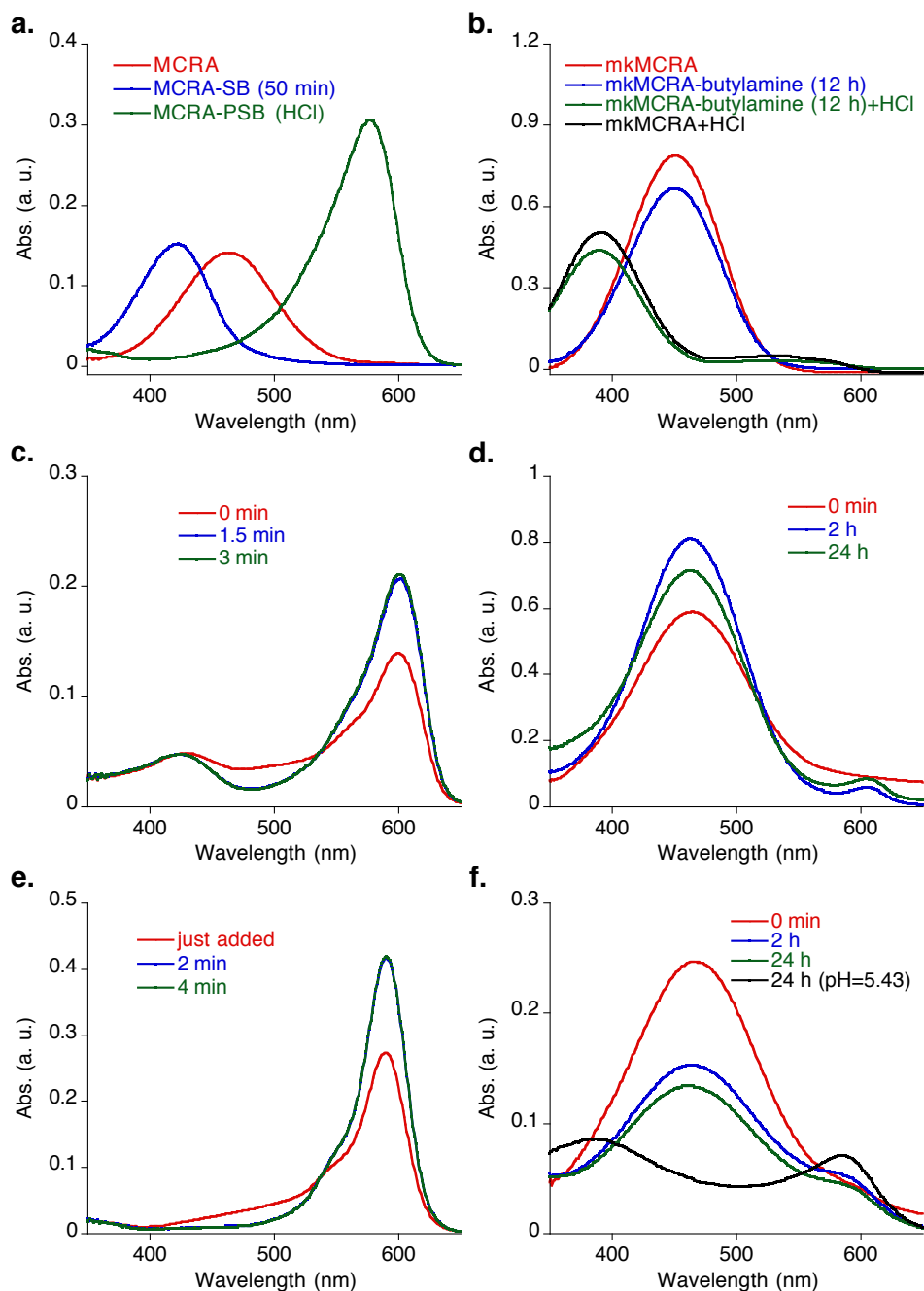
Figure III-11: Comparison in reactivity of MCRA and mkMCRA. **a-b.** Reaction with *n*-butyl amine in ethanol for MCRA (**a**) and mkMCRA (**b**). **c-d.** Reaction with CRABPII-R111K:R132L:Y134F:T54V:R59W:L121D for MCRA (**c**) and mkMCRA (**d**). **e-f.** Reaction of CRBPII-Q108K:K40L:T51V:T53C:R58W: T29L:A33W:Q4F:L117E with MCRA (**e**) and mkMCRA (**f**).

T51V:T53C:R58W:T29L:A33W:Q4F:L117E. These two proteins were compared in iminium formation with both MCRA and mkMCRA. Both proteins form PSB with MCRA within two minutes of incubation at room temperature. However, mkMCRA gave the peak corresponding to PSB only after 2 h (**Figure III-11**). Ideally, observed reactivity inside the mutants compared to no imine formation with *n*-butylamine is a promising starting point to enhance the specificity of protein labeling. But further protein/ligand optimization would be essential to accelerate the binding process.

Before taking any steps towards engineering the protein for efficient mkMCRA binding, this ligand should be evaluated in its ability to penetrate and stain live cells alone. A concentration identical to the optimal MCRA loading and the minimum time for PSB formation were utilized in this experiment. The non-transfected cervical cancer cells, HeLa, were incubated with 0.25 μM of mkMCRA for 20 h and then washed with PBS buffer. Imaging was performed on a two-track laser scanning confocal microscope, where the first track
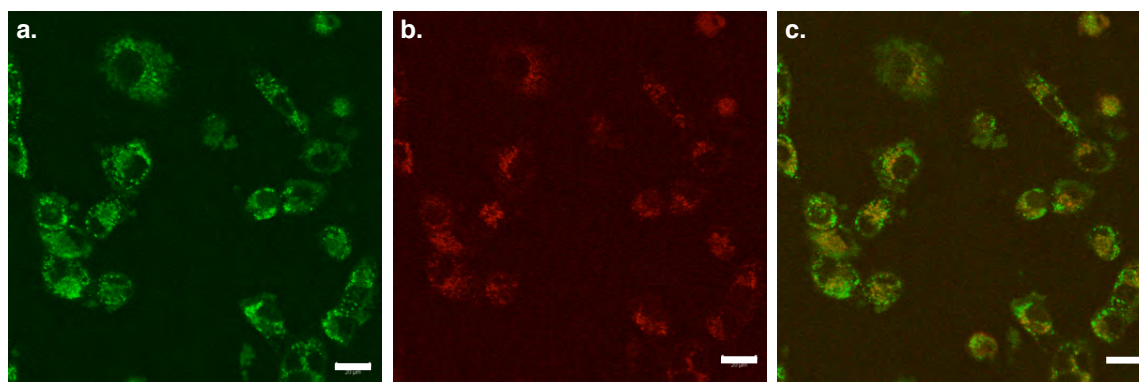
Figure III-12: Confocal images of live HeLa cells stained with mkMCRA for 20 h. **a.** Imaging of unbound mkMCRA (laser 488 nm, filter BP 505-525 nm). **b.** Imaging of mkMCRA-PSB (laser 594 nm, filter LP 615 nm). **c.** Overlaid (a) and (b) has no matching localization. Scale bar, 20 μM.

corresponded to the free ketone (488 nm laser, 505-525 nm BP filter) and the second track was set to the expected mkMCRA-PSB spectroscopic profile (594 nm laser, 615 nm LP filter). Disappointingly, non-transfected cells showed fluorescence in both tracks (**Figure III-12**), even though mkMCRA had no reactivity with *n*-butylamine during the same time period *in vitro*.

This experiment eliminated the need for further optimization of the mkMCRA/protein complex. On the bright side, this test allowed us to explore further ligand optimization. The ideal ligand should react fast with the specifically engineered protein to allow short incubation time, but at the same time it should react slowly with all other intracellular amines. Probably, some functional group with a reactivity between aldehyde and methyl ketone would be a reasonable option. Among the possible chromophore modifications could include: less reactive aldehydes, similar to formylacrylonitrile **III-23**, aldehydes with two reactive centers, like in formylchromones, or more reactive ketones, for example fluoromethyl ketone.

III.5 Formation of the hexahydrofuro[2,3-c]pyrrole using electrophilic trap in the aza-Payne tandem Cyclization/Hydroammination

This project was bourne out from the observations of reactivity of the formylacrylonitrile **III-23** and related compounds. To our surmise, the aldehyde moiety in **III-23** resisted prototypical chemical transformations. As evident from the literature, similar motifs prefer to undergo 1,4-addition in the presence of various nucleophiles. This observation inspired an extension of one of the

175

methodologies developed previously in our lab, a Tandem Aza-Payne/ Hydroamination.[64-65]

As shown in **Figure III-13**, a tandem aza-Payne/hydroamination reaction typically yields densely decorated pyrrolidines with defined relative stereochemistry translated from the starting aziridinol.[65] If CO_2 traps the alkoxide, the furnished carbonate anion opens the aziridine ring.[64] The hypothetical cascade for the new methodology involves the oxa-Michael addition of an aziridinol anion to the electrophile, with subsequent aziridine opening via an

Figure III-13: Tandem aza-Payne/hydroamination reaction (aza-Payne/hydroamination steps red, electrophilic trap green). Dashed box: trapping a latent nucleophile as Michael acceptor by alkoxide intermediate.
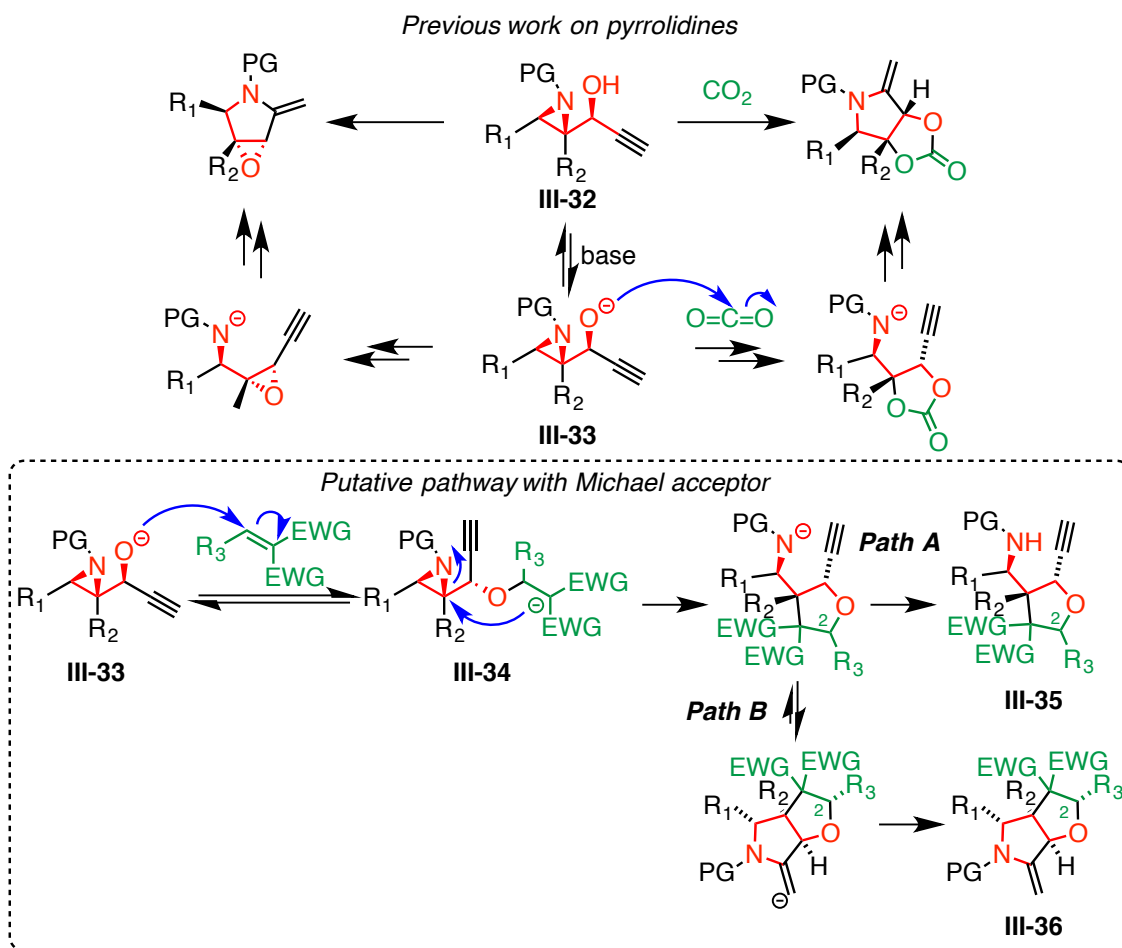
enolate intermediate **III-34** (**Figure III-13**, dashed box). Although the aza-Payne/hydroamination pathway is a highly efficient process, the aziridine-based alkoxide **III-33** outcompetes the intramolecular reaction at high concentration of soluble electrophile. This counterintuitive selectivity is possible because of the formation the thermodynamically more stable five-membered ring, compared to the strained epoxide. In such a case, the C-C bond formation during aziridine opening would be irreversible and would furnish a tetrahydrofurane ring. Furthermore, this intermediate can undergo protonation to yield a pendant tosylamine **III-35** (Path A) or it can exhibit a contemporaneous hydroamination to **III-36** (Path B). This reaction provides a new robust pathway to highly substituted tetrahydrofurans or hexahydrofuro[2,3-c]pyrrole core.

The possibility of accessing the hexahydrofuro[2,3-c]pyrrole core **III-36** was readily tested by treatment of aziridinol **III-37** with commercially available *trans*-β-methyl-β-nitro styrene **III-38**. Aziridinol **III-37**, procured from Kumar Ashtekar (Babak Borhan's lab, MSU) (a precursor for the total synthesis of
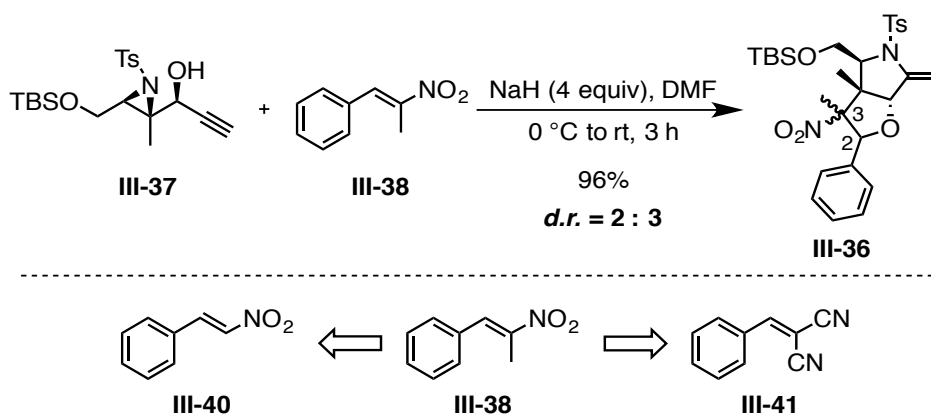
Scheme III-9: Use of *trans*-β-methyl-β-nitro styrene as a latent nucleophile in tandem aza-Payne/hydroamination protocol and potentially better Michael acceptors based on the diastereoselectivity outcome.

Salinosporamide A) was used as a substrate to test this reaction.

The reaction conditions were minimally modified, as compared to the previous report with CO_2 as a latent nucleophile. Only the TBAF additive was eliminated, as the free alcohol was utilized. To ensure that the intermolecular trapping of nitrostyrene **III-38** outcompetes the intramolecular aza-Payne rearranagement, aziridinol **III-37** was added slowly to the reaction mixtute. Considering formation of two new stereogenic carbons, C2 and C3, there are four possible diastereomers. Fortuitously, the product **III-39** was obtained as a mixture of only two diastereomers (ratio 3:2, ^{1}H NMR). It suggested that stereoselectivity of at least one stereocenter was substrate controlled.

When the resonance-stabilized anion intermediate **III-37a** (**Figure III-13**) is formed, the free rotation along C2-C3 bond might epimerize the stereogenic center C3 before the first cyclization. This suggests that a stereocenter at C2 should be controlled by aziridinol **III-37** and racemization at C3. Based on the

Table III-1: Reaction conditions for β-nitrostyrene **III-40**.
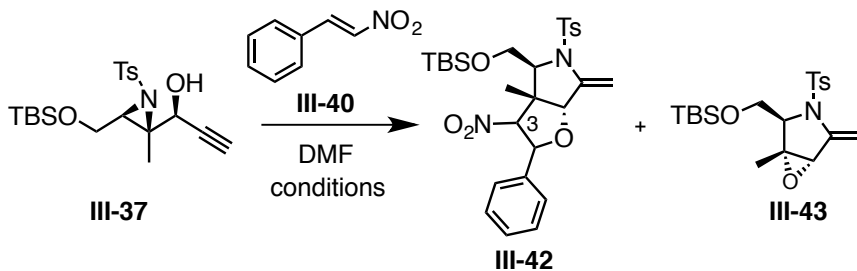
Entry	III-40 (equiv)	Additives	Time	Temperature	Result[a]
1	3	NaH (4 equiv)	3 h	0 °C to rt	**III-43**
2	3	NaH (4 equiv)	3 h	-50 to 0 °C to rt	**III-43**
3	5	NaH (4 equiv)	3 h	0 °C to rt	**III-37**
4	3	Et$_3$N (20 equiv)	18 h	rt	**III-37**

[a]From TLC and crude NMR compared to known **III-37** and **III-43**

178

suggested hypothesis, an enolizable C3 might lead to a single diastereomer, under the reaction conditions giving a thermodynamic product. Thus, nitrostyrene **III-40** as an electrophile was tested next. Also a strong Michael acceptor **III-41**, with no preference for stereoselectivity at C3 was studied.

Unfortunately, nitrostyrene **III-40** did not give the expected product under attempted conditions. Either starting material or the *cis*-fused 3–5 ring system **III-43**, was obtained as a result of the intramolecular reaction (**Table III-1**). Because no nitrostyrene **III-40** was observed in crude NMR, it presumably decomposes spontaneously under reaction conditions and therefore does not yield the desired product.

On the other hand, 2-benzylidenemalononitrile **III-41** (**Table III-2**) gave complete reaction conversion to the desired product. Detailed characterization of the isolated compound identified the tetrahydrofuran **III-44** as a single diastereomer (**Table II-2**, entry 1) formed *via* Path A (**Figure III-13**). Attempts to decrease the equivalents of electrophile led to the epoxyenamine **III-43** over conjugate addition (**Table II-2**, entries 2 and 3). However, consecutive attempts to reproduce the same results gave **III-43** in minor quantities (**Table II-2**, entry 4). Interestingly, a trace amount of benzaldehyde was also isolated. Inconsistency was attributed to the possible varying amounts of water that could get into the mixture during the reaction set-up. The main disadvantage in such a case is hydroxide from NaH that could trigger the retro-aldol reaction upon adding to the electrophile.[66-67] Therefore, decomposition of the electrophile led to the formation

Table III-2: Reaction conditions for 2-benzylidenemalononitrile.

Ts N OH
TBSO ···· (methyl) ≡   **III-37**

III-41
conditions
0 °C to rt
3 h

TBSO— N(Ts) ring with =CH₂, O, methyl   **III-43**

TBSO Ts NH
NC, NC ··· ≡ Ph O   **III-44**

TBSO— N(Ts) ring, NC, NC, Ph, O, =CH₂   **III-45**
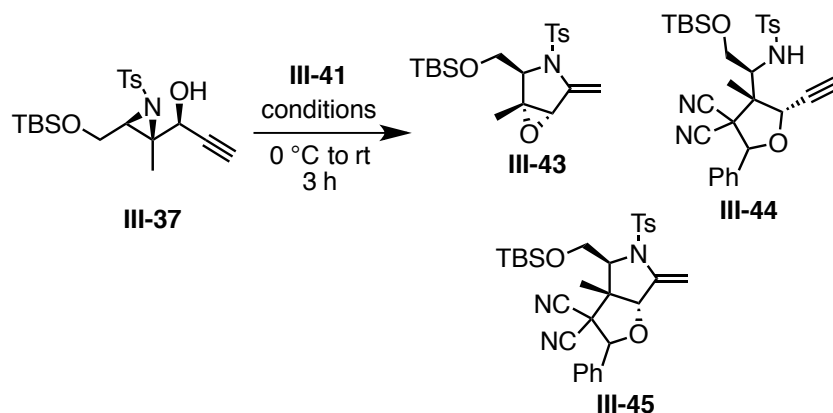
Entry	III-41 (equiv)	Conditions	t, °C	Ratios[a] (III-37/III-43/III-44/III-45)
1	3	NaH (4 equiv), DMF	0→23	0/0/100/0
2	2	NaH (4 equiv), DMF	0→23	0/100/0/0
3	1	NaH (4 equiv), DMF	0→23	0/100/0/0
4[b]	3	NaH (4 equiv), DMF	0→23	0/33/67/0
5	3	NaHMDS (5 equiv), DMF	0→23	100/0/0/0
6	3	LiHMDS (5 equiv), DMF	0→23	100/0/0/0
7	3	NaHMDS (5 equiv), DMF	0→23→60[c]	n.d.[d]
8	3	LiHMDS (5 equiv), DMF	0→23→60[c]	0/100/0/0
9[e]	**3**	**NaH (10 equiv), DMF**	**0→23**	**0/0/100/0**
10	3	NaH (10 equiv), DMSO	0→23	0/100/0/0
11	3	NaH (10 equiv), THF	0→23	0/100/0/0

[a]Ratios determined from ¹H NMR of crude. [b]Irreproducible duplicate of entry 1. [c]Reaction was heated at 60 °C for an additional 2 h after bringing it slowly to rt during 3 h. [d]Not determined as no distinct peaks in NMR were observed. [e]NaH is added as a slurry in DMF *via* dropwise (entries 9-11). Reaction yields for (e) were 70% from two separate replicates compared to 45% obtained in entry 1.

of benzaldehyde. The latter also explains the fact that nitrostyrene **III-40** was consumed without participating in the reaction (summarized in **Table III-1**).
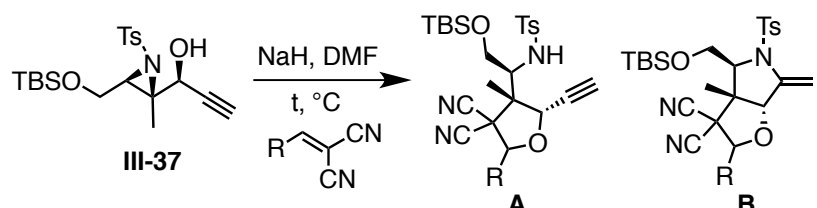
In order to form an alkoxide that immediately traps the Michael acceptor in the intermolecular reaction, the base needs to be added into the reaction mixture slowly. Since solid sodium hydride is hard to add slowly, other soluble bases were used in the reaction after premixing the aziridinol **III-37** and electrophile **III-41**. Surprisingly, use of NaHMDS and LiHMDS at room temperature led to the

recovery of unreacted alcohol that otherwise delivered tetrahydrofuran **III-44** or pyrrolidine **III-43** with NaH as a base (**Table II-2**, entries 5 and 6). Heating the reaction with LiHMDS gave pyrrolidine as a result of an intramolecular reaction (**Table II-2**, entry 8), but the reaction was sluggish with NaHMDS (**Table II-2**, entry 7). Because NaH worked earlier but inconsistently, the base in form of a suspension of NaH in DMF was added slowly into the reaction mixture under inert conditions to furnish **III-44**. This reaction was reproducible in terms of yields and stereoselectivity (**Table II-2**, entry 9). Although hydroamination did not proceed under the described conditions, the desired intermolecular reaction proceeded to the highly decorated tetrahydrofurane ring. The influence of solvent was further probed with the same base, but only pyrrolidine **III-43** was isolated if the reaction was completed in DMSO or THF (**Table II-2**, entries 10 and 11).

Reproducible conditions obtained with a NaH slurry were maintained to explore the substrate scope of the reaction. As summarized in **Table III-3**, various types of substitutions on the benzylidenemalononitrile gave different products. Electrophiles with electron donating groups on the benzene ring delivered hexahydrofuro[2,3-c]pyrroles at a temperature that otherwise furnished tetrahydrofuran with **III-41** (**Table III-3**, entries 1-4 compared to entry 5). Interestingly, *p*-bromo-benzylidenemalononitrile underwent only aza-Payne-reaction, without the concomitant hydroamination, similar to the reaction with **III-41** at room temperature (**Table III-3**, entry 6). To probe whether the hydroamination step can be suppressed in case of electron rich substrates, *p*-

methoxy-benzylidenemalononitrile was subjected to the same reaction at -5 °C. This reaction proceeded for a longer time, but only monocyclized product **A** was isolated (**Table III-3**, entry 7). In order to promote the hydroamination in substrates that are not electronically rich, the reactions were performed at slightly elevated temperatures. Interestingly, electron deficient substrates underwent monocyclization to yield the substituted tetrahydrofuran **A**, while neutral substrates afforded hexahydrofuro[2,3-c]pyrroles **B** (**Table III-3**, entries 8-11). Heating to an ambiguous temperature furnished some amount of the hexahydrofuropyrrole **B** for *p*-chloro-benzylidenemalononitrile **III-52**, while *p*-bromo-benzylidenemalononitrile **III-53** still proceeded without hydroamination to

Table III-3: Substrate scope and reaction outcome based on electrophile.

Entry	R	t, °C	Product	Yield, %
1	*p*-NMe$_2$-C$_6$H$_4$ (**III-46**)	0→23	**B** (**III-58**)	50
2	*p*-OMe-C$_6$H$_4$ (**III-47**)	0→23	**B** (**III-59**)	84
3	*o*-OMe-C$_6$H$_4$ (**III-48**)	0→23	**B** (**III-60**)	84
4	*α*-naphtyl (**III-49**)	0→23	**B** (**III-61**)	71
5	Ph (**III-41**)	0→23	**A** (**III-44**)	70
6	*p*-Br-C$_6$H$_4$ (**III-53**)	0→23	**A** (**III-57**)	71
7	*p*-OMe-C$_6$H$_4$	-10→-5[a]	**A** (**III-54**)	34
8	Ph	0→23→35[b]	**B** (**III-45**)	84
9	*p*-CH$_3$-C$_6$H$_4$ (**III-51**)	0→23→35[c]	**B** (**III-63**)	94
10	*p*-Cl-C$_6$H$_4$ (**III-52**)	0→23→40	**A** (**III-56**)	82
11	*p*-Br-C$_6$H$_4$	0→23→40	**A**	39
12	*p*-Cl-C$_6$H$_4$	0→23→70[d]	**A/B** (**III-56/III-64**)	42/25
13	*p*-Br-C$_6$H$_4$	0→23→70	**A**	94

[a]Reaction continued for 12 h. [b]After 3 h stirring on ice and at rt reaction was heated at specified temperature for additional 2 h (entries 8-13). [c]After bringing reaction slowly to rt during 3 h there were two spots on TLC; heating for additional 2 h converted everything to one product. [d]From crude NMR the ratio of **A** to **B** was 5:4.

A (**Table III-3**, entries 12-13). The observed results were correlated with Hammet

σ–values for *para-* substituents to elucidate the influence of the electronics of the

aromatic substituents at the reacting atoms in the transition states ($\sigma_{p\text{-Cl}}$=0.227,

$\sigma_{p\text{-Br}}$=0.232). Considering that *p*-bromo-benzene is slightly more electron

deficient than the corresponding chloro-analog, the isolation of some

hydroaminated product only in the second case was reasonable (**Table III-3**,

entry 12).

To our delight, the crystal structures of *cis*-fused 5–5 ring systems with

neutral phenyl, electron rich *N,N*-dimethyl- and electron deficient chloro-benzene

were obtained. This provided an unequivocal determination of the

stereochemistry at C-2 of the tetrahydrofuran. Additionally, it confirmed the

retention of stereochemistry from the starting aziridinol, despite the fact that C-C

bond is formed at the quaternary carbon. Moreover, changes in the distance

between electron deficient tosyl amide and newly introduced aromatic substituent

(**Figure III-14**) explains the factors driving the temperature dependent

hydroamination of electronically different substrates.
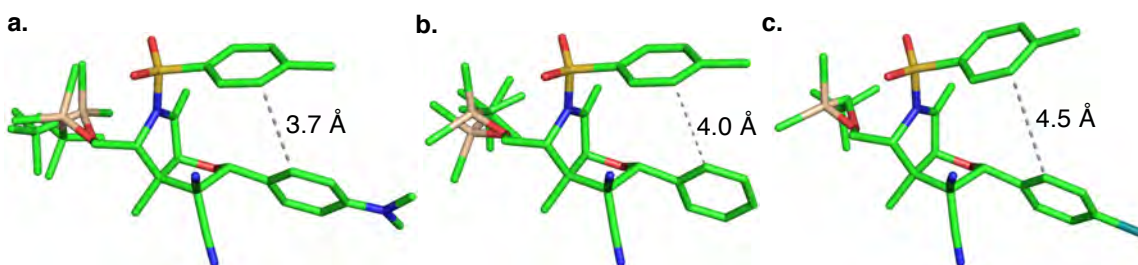
Figure III-14: Comparison of the crystal structures of hexahydrofuro[2,3-c]pyrroles with **a.** *N,N*-dimethylphenyl, **b.** phenyl and **c.** chlorophenyl substituted THF-ring.

The putative π-π interactions existing between the aromatic nuclei may restrict the conformational freedom and orient the *N*-Ts anion in a seemingly better arrangement to ensue the desired *5-exo-dig* ring closure. These π-π interactions may be weaker in the absence of electron donating groups on the aromatic nucleus (phenyl) or may even be less favorable in case of electron withdrawing groups (*p*-chloro) (**Figure III-14**). Hence, heating is necessary to overcome the barrier for hydroamination where the aromatic groups lack electron-donating substituents. The extent of heating during the reaction is dictated by the electronic nature of the substrate: more electron-withdrawing the substituent, higher the temperature.

To verify this hypothesis, we probed the transition states for three electronically biased substrates- a. *p*-OMe-C$_6$H$_4$- substituted, b. phenyl
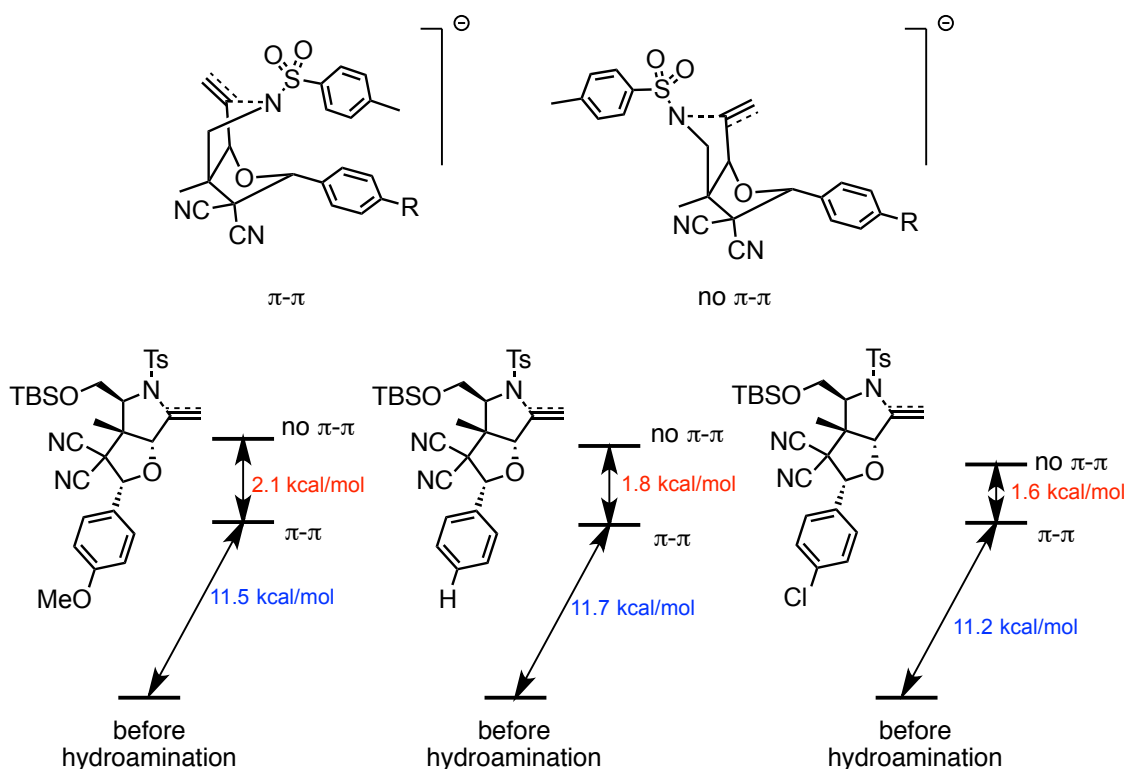
Figure III-15: B3LYP/6-31G* calculations of the transition states.

substituted, and, c. *p*-Cl-C$_6$H$_4$- substituted malononitriles. The transition states for hydroamination of these substrates were evaluated by DFT calculations at the B3LYP/6-31G* level of theory (by Kumar Ashtekar, Professor Babak Borhan's Lab, MSU). As depicted in **Figure III-15**, the transition states that involve the π-π-interactions are energetically more favorable among all three substrates. Furthermore, the strength of these interactions is in accordance to the electronic nature of the aromatic nucleus. For a relative comparison of the individual barriers for hydroamination, the transition states must be probed with an appropriate solvation model. These calculations are a part of ongoing studies.

It is important to mention that proportional scale up of the reaction did not yield the same products. The optimized conditions were established with 0.02 mmol of aziridinol in 0.01 M DMF. Reported yields are based on isolated products. These reactions are reproducible at a small scale, however a scale-up to 0.1 mmol of aziridinol resulted in the mixture of three products corresponding to tetrahydrofurane, 5–5 and 3–5 ring systems. The problems with reaction scale-up could be attributed to the mechanical delivery of non-homogeneous NaH/DMF slurry, which may cause high local concentration of base leading to non-uniformity in the temperature gradient. Addressing this problem needs further optimization in terms of the rate of addition of the NaH/DMF slurry.

To expand the scope of this reaction, several aziridinols with different electronic and steric profiles need to be examined in this tandem Michael addition/aza-Payne/hydroamination. Moreover, since aziridinols are prepared by

Grignard reaction with ethynylmagnesium bromide, which yields the putative alkoxide intermediate, a '*One-Pot*' protocol may be developed wherein the electrophile can be delivered after complete consumption of the aziridinal followed by a gradual increase in the reaction temperature. Additionally, the investigated reaction is an intriguing case study for computational analysis of experimental data from the intermolecular reaction with concomitant substrate controlled second cyclization.

III.6 Conclusions and significance

In summary, various synthetic routes to diverse chromophores were described, in which a few were successfully accomplished and carried forward to biological investigations. The originally targeted cyano julolidine retinal analogs were synthesized in two forms. However, iminium protonation inside the protein was challenging. This study provided chromophores with new routes of potential ways to decrease the non-specific binding. Exploring the secondary interactions between the protein and the chromophore might improve specificity of the labeling reaction. Decreasing the carbonyl reactivity is another potential way to inhibit the non-specific iminium fomation. Finally, the low pK_a value of the probe-PSB with the random intracellular amines is another plausible approach, because of the possibility to cut-off the non-specific SB spectra from the specific PSB.

Misleading reactivity of the synthesized chromophore led to a new reaction discovery, with substrate controlled formation of hexahydrofuro[2,3-c]pyrrole core using *oxa*-Michael addition in the modified tandem aza-Payne/hydroamination.

Although it is not related to the protein engineering and live-cell imaging, the outcome of this methodology becomes quite interesting from both a synthetic and computational standpoint. To improve this methodology for large scale synthesis the choice of base for an alkoxide formation should be explored. Interestingly, this kind of 5-5 bicyclic ring fusion has some similarity to Salinosporomide A after its covalent binding to and inhibition of the proteasome. The biological evaluation of the synthesized compounds should be persued. Additionally, the same type of the Michael acceptors can be examined as an electrophilic trap in the related transformation between aziridinal and activated alkene in the Baylis-Hillman reaction described by Dr. Aman Kulshretha (MSU, Professor Borhan's Lab).

III.7. Materials and methods.

All the glassware was flame-dried or oven-dried dried prior to use. Air/moisture sensitive experiments were performed under an inert atmosphere of dry Nitrogen unless indicated otherwise. Commercially available chemicals were used as received. Julolidine hydrobromide was dissolved in water, basified with NaOH to pH 10 and extracted with EtOAc to get julolidine just prior to use. Malononitrile was refluxed with activated carbon in diethyl ether, filtered while hot thru celite and left for re-crystallization from diethyl ether prior to use. Commercial grade solvents were used for routine purposes without further purification. DMF was distilled under vacuum over Na_2SO_4, and stored over 4 Å molecular sieves covered with aluminum foil. NMR spectra were obtained using either 500 MHz Varian or 600 MHz Inova NMR spectrometer. Column chromatography was

performed using Silicycle (40 – 60 mμ) silica gel. Pre-coated silica gel 60 F$_{254}$ plates were used for analytical TLC and visualized using UV light/*p*-anisaldehyde/PMA as the stain. High-resolution mass spectra (HRMS) were taken on a Waters GCT Premier GC/TOF MS. All compounds with more than one double bond attached to aromatic ring were prepared in the dark room under red light and such compounds and intermediates were stored at -20 °C or -80 °C.

Synthesis of III-2. Round bottom flask (25 mL) was filled with dry DMF (0.8 mL) and brought to 0 °C. Phosphoryl chloride (0.16 mL, 1.78 mmol) was added dropwise to the ice-cold vigorously stirring dry DMF. After 30 min solution of julolidine (0.364 g, 1.78 mmol) in DMF (0.32 mL) was added and the resulting mixture was heated to 80 °C for two hours. The solution was poured on the ice and neutralized to pH 6–8 by addition of saturated sodium acetate and left at room temperature overnight. The desired aldehyde precipitated as a light brown solid. The solid julolidine aldehyde was filtered, washed with water and left under vacuum for 24 h. Yield 0.355 g (88%). Melting temperature: 77–78 °C. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$): δ 1.92–1.96 (m, $J = 6.0$ Hz, 4H), 2.75 (t, $J = 6.0$ Hz, 4H), 3.27 (t, $J = 6$ Hz, 4H), 7.27 (s, 2H), 9.58 (s, 1H). ^{13}C-NMR (125 MHz, CDCl$_3$): δ 21.3, 27.6, 50.0, 120.3, 124.1, 129.4, 147.8, 190.1 ppm.

Synthesis of III-3. A round bottom flask (10 mL) charged with potassium hydroxide (84 mg, 1.5 mmol) and acetonitrile (3 mL) was refluxed for 30 min. Then solution of julolidine aldehyde **III-2** (100 mg, 0.5 mmol) in acetonitrile (0.5 mL) was added at once. The reaction was refluxed for 1 h and the hot solution

was poured on ice. The mixture was then extracted with ethyl acetate (3×10 mL), washed with brine and dried over sodium sulfate. After removal of solvent, pure cyano vinyl julolidine **III-3** was obtained (112 mg, 99%) as bright yellow crystals (m. pt. = 131-132 °C). Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$): δ 1.86–1.91 (m, J = 6.5 Hz, 4H), 2.65(t, J = 6 Hz, 4H), 3.17 (t, J = 6.5 Hz, 4H), 5.40 (d, J = 8.0 Hz, 1H), 6.80 (s, 2H), 7.19 (d, J = 8.0 Hz, 1H) ppm. ^{13}C NMR (125 MHz, CDCl$_3$): δ 21.4, 27.6, 49.8, 87.8, 120.2, 120.4, 120.9, 126.6, 145.3, 150.8 ppm. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{15}$H$_{17}$N$_2$]$^+$ 225.1392, found 225.1395.

Synthesis of III-4. A flame dried 100 mL round bottom flask was charged with **III-3** (250 mg, 1.11 mmol) and dissolved in 50 mL of dry THF. Solution was stirred over dry ice/acetone bath at -78 °C for 20 min followed by dropwise addition of DIBAL (1.33 mL of 2.5 M solution in toluene, 3.33 mmol). The reaction mixture was warmed to 0 °C and left stirring for 10 h. Subsequently a homogeneous mixture of water absorbed on silica was added and stirring was continued overnight at room temperature. Unreacted water was absorbed by MgSO$_4$, the solids were removed by filtration and washed twice with ether. Rotary evaporation of solvents yielded a dark oily material. The product was purified by flash chromatography (silica gel basified with 0.1% triethyl amine, EtOAc/hexane in ratio 1:1) to give 125 mg (50%) of **III-4**. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$): δ 1.93–1.96 (m, J = 6.5 Hz, 4H), 2.73 (t, J = 6.5 Hz, 4 H), 3.24 (t, J = 6.5 Hz, 4H), 6.70 (dd, J = 8.0, J = 16 Hz, 1H), 6.99 (s, 2 H), 7.24 (d, J =

16.0 Hz, 1H), 9.52 (d, $J = 8.0$ Hz, 1H) ppm; ^{13}C-NMR (125 MHz, CDCl$_3$): δ 21.5, 27.7, 49.7, 87.8, 120.8, 121.0, 122.8, 128.2, 154.3, 193.6 ppm.

Synthesis of III-6. The premixed triethyl phosphite (21.98 g, 0.13 mol) and chloroacetonitrile (10 g, 0.132 mol) were heated at reflux for 6 h (180 °C, sand bath). The reaction was brought to room temperature and the main byproduct, chloroethane, was removed under vacuum. The crude oil was subjected to distillation under reduced pressure for purification. The isolated yield of the transparent oil **III-6** was 20 g (85%). Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$): δ 1.26–1.30 (m, 6 H), 2.79–2.83 (dd, $J = 21.0$ Hz, $J = 1.2$ Hz, 2 H), 4.12–4.15 (m, J=6.5 Hz, 4 H) ppm; ^{13}C-NMR (125 MHz, CDCl$_3$): δ 16.03 (d, $J = 6.5$ Hz), 16.1 (d, $J = 143.0$ Hz), 63.62 (d, $J = 6.5$ Hz), 112.5 (d, $J = 12.0$ Hz) ppm.

Synthesis of III-7. Solid NaH (0.316 g, 60% in mineral oil, 7.9 mmol) was transferred at once into the flask with anhydrous THF on ice bath. The mixture was stirred for 10 min and cyanomethylphosphonate **III-6** (1.4 g, 7.9 mmol) dissolved in 3.5 mL THF was added via syringe during 10 min. When solution turned yellow (around 1 h) the chloroacetone was added dropwise to the freshly generated phosphanate anion and was vigorously stirred at room temperature for 6 h. Formation of diethyl phosphate at the bottom of the flask indicated that the reaction was successful. Addition of water dissolved the precipitate and quenched the reaction. This mixture was extracted with diethyl ether, washed with brine and dried over Na$_2$SO$_4$. Purification was done using distillation at 70

ºC/ 30 mbar. Yield: 0.75 g (82%). The *E:Z* ratio 2:1 was assigned based on the previously reported spectra.[68] Spectral data for *E/Z* mixture: ^{1}H NMR (500 MHz, CDCl$_3$): δ 2.05 (d, *J* = 1.7 Hz, 3H, 2.14 (d, *J* = 1.0 Hz, 3H), 4.06 (t, *J* = 0.5 Hz, 2H), 4.26 (s, 2H), 5.28 (s, 1H), 5.50 (dd, *J* = 2.4, 1.2 Hz, 1H) ppm. ^{13}C-NMR (125 MHz, CDCl$_3$): δ 18.9, 20.9, 44.2, 46.8, 98.7, 99.0, 115.1, 115.9, 157.8, 158.1 ppm.

Synthesis of III-8. The protocol is identical to the synthesis of **III-6** with modification in the purification step, which was done using distillation at 180 ºC/30 mbar. Yield: 1.13 g (80%). The *E:Z* ratio 4:3 was assigned based on previously reported spectra.[68] Spectral data for *E/Z* mixture: ^{1}H NMR (500 MHz, CDCl$_3$): δ 1.23–1.33 (m, 12H), 2.06–2.07 (m, 3H), 2.16–2.17 (m, 3H), 2.68 (dd, *J* = 23.5 Hz, 2H), 2.94 (d, *J* = 23.5 Hz, 2H), 4.07–4.15 (m, 8H), 5.24 (d, *J* = 5.5 Hz, 1H), 5.26 (s, 1H) ppm. ^{13}C-NMR (125 MHz, CDCl$_3$): δ 16.3–16.4 (m), 22.4 (d, *J* = 3.5 Hz), 24.3 (d, *J* = 2.0 Hz), 34.8 (d, *J* = 135.0 Hz), 36.5 (d, *J* = 137.0 Hz), 62.5 (t, *J* = 5.5 Hz), 98.9 (d, *J* = 12.5 Hz), 99.3 (d, *J* = 12.5 Hz), 116.3 (d, *J* = 5.5 Hz), 155.4–155.8 (m, *J* = 12.5 Hz) ppm.

Synthesis of III-9. In a 100 mL round bottom flask under the nitrogen NaH (25 mg, 60% in mineral oil, 0.602 mmol) was added at once to 40 mL the ice-cold dry THF. Phosphonate **III-8** (132 mg, 0.602 mmol) was added slowly using a syringe and the reaction mixture was left stirring on ice for 1 h. The solution of vinyl julolidine aldehyde **III-4** (115 mg, 0.506 mmol) in THF (10 mL) was added

dropwise to the reaction mixture. Stirring at room temperature was continued for 6 h (monitored by TLC). The reaction was quenched by addition of brine (20 mL) and the aqueous layer was extracted three times with diethyl ether, dried over $NaSO_4$ and filtered. Purification of the crude was performed by column chromatography (silica gel basified with 0.1% triethyl amine, 1:3 ethyl acetate/hexane; R_f = 0.6) to furnish nitrile **III-9** (70 mg, 47%) as red crystals (two isomers 7:1, m. pt. = 147-157 °C). Spectral data for both isomers: ^{1}H NMR (500 MHz, $CDCl_3$): δ 1.92–1.97 (m, J = 6.5 Hz, 8H), 2.02 (t, J=1.2 Hz, 3H), 2.17 (s, 3H), 2.71–2.74 (m, J = 6.5 Hz, 8H), 3.04–3.06 (t, J = 6.5 Hz, 4H), 3.17–3.19 (t, J = 6.5 Hz, 4H), 4.99 (s, 1H), 5.11 (s, 1H), 6.24 (d, J = 16.0 Hz, 1H), 6.56–6.60 (m, 3H), 6.66–6.76 (m, 4H), 6.87 (s, 2H), 6.89 (s, 2H) ppm; ^{13}C-NMR (125 MHz, $CDCl_3$): δ 16.6, 21.8, 27.7, 49.9, 50.2 93.6, 95.1, 118.5, 121.2, 122.8, 123.7, 126, 127, 128, 137, 138, 143, 157 ppm. MS (ESI) (m/z): [M+H]$^{+}$ calculated for $[C_{20}H_{23}N_2]^{+}$ 291.1861, found 291.3744.

Synthesis of III-10. Nitrile **III-9** (70 mg, 0.24 mmol) dissolved in dry THF (20 mL) and brought to -78 °C was treated with DIBAL (0.3 mL of 2.5 M solution in toluene, 0.75 mmol) and left stirring for 10 h after warming it to 0 °C. After the starting material was consumed, the imine was hydrolyzed with a heterogeneous mixture of water absorbed on silica gel with vigorous stirring overnight at room temperature. After drying the mixture by addition of $MgSO_4$, all solids were filtered and thoroughly washed with diethyl ether. Evaporation of organic solvents

in vacuo gave a dark solid. At first product was purified by column chromatography (silica gel basified with 0.1% triethyl amine, ethyl acetate/hexane, 1:1 (R_f=0.556) and then by HPLC (Zorbax Rx-SIL, 4.6 mm × 25 cm, ethyl acetate/hexane, 3:10, flow rate 2 mL/min, UV detector 440 nm). Aldehyde **III-10** was isolated as an oily mixture of isomers (2 mg, 3%). Spectral data for two major isomers (ratio 10:3): ^{1}H NMR (500 MHz, CDCl$_3$): δ 1.93–1.96 (m, J = 6.5 Hz, 8H), 2.11 (t, J = 1.2 Hz, 3H), 2.28 (s, 3H), 2.72 (t, J = 6.5 Hz, 4H), 3.18 (t, J = 6.5 Hz, 4H), 5.78 (d, J = 8.0 Hz, 1 H), 5.93 (d, J = 8.0 Hz, 1H), 6.33 (d, J = 23.5 Hz, 1H), 6.59–6.66 (m, 3H), 6.85–6.87 (m, 2H), 6.88 (s, 2H), 6.90 (s, 2H), 10.06 (d, J = 8.5 Hz, 1H), 10.18 (d, J = 8.5 Hz, 1H) ppm; ^{13}C-NMR (150 MHz, CDCl$_3$): δ 13.0, 21.8, 27.7, 49.9, 121.2, 123.1, 123.7, 126.1, 128.2, 132.0, 138.0, 138.7, 138.9, 143.6, 155.3, 191.0 ppm. UV-vis (ethanol): λ_{max} = 460 nm. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{20}$H$_{24}$NO]$^+$ 294.1858, found 294.1858.

Synthesis of III-11. To a solution of NaH (120 mg, 60% in mineral oil, 3 mmol) in dry THF (60 mL) ylide **III-8** (597 mg, 2.75 mmol) was added within 10 min at 0 °C. The reaction was warmed to room temperature and solution of julolidine aldehyde **III-2** (502 mg, 2.5 mmol) in THF (5 mL) was added slowly. Reaction continued for additional 1 h, after that brine (20 mL) was added and the aqueous layer was extracted three times with diethyl ether and dried over anhydrous sodium sulfate. After rotary evaporation of the organic solvent the

193

crude mixture was dissolved in 5 mL of ethanol, heated on steam bath and left for recrystallization at -20 °C, overnight. Pure all-*trans*-isomer crystallized as a yellow solid (340 mg, 52%, m. pt. = 122 -125 °C). Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$): δ 1.93 (m, 4H), 2.20 (s, 3H), 2.72 (t, J = 6.5 Hz, 4H), 3.19 (t, J = 6.5 Hz, 4H), 5.13 (s, 1H), 6.57 (d, J = 16.0 Hz, 1H), 6.72 (d, J = 16.0 Hz, 1H), 6.90 (s, 1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 16.6, 21.9, 27.7, 49.9, 94.2, 118.8, 121.2, 122.5, 122.6, 126.5, 136.7, 144.0, 157.6. HRMS (ESI) (m/z): [M+H]$^{+}$ calculated for [C$_{18}$H$_{21}$N$_2$]$^{+}$ 265.1705, found 265.1707.

Synthesis of III-12. Nitrile **III-11** (132 mg, 0.5 mmol) was dissolved in dry dichloromethane (2 mL) and chilled to 0 °C. Solution of DIBAL (1.5 mL, 1M solution in hexane, 1.5 mmol) was added dropwise *via* syringe while stirring on an ice bath. Reaction was left stirring at room temperature for 2 h and TLC verified consumption of the starting material. Reaction was quenched on ice at first with ethyl acetate (10 drops) and then with saturated Rochelle's salt (2 mL). After vigorous stirring for 30 min at room temperature two distinct layers were formed. Reaction was diluted with ethyl acetate (2 mL) and the organic layer was separated from solution of Rochelle salt. The aqueous layer was extracted three times with ethyl acetate (2 mL) and combined organic layers were dried over anhydrous sodium sulfate, filtered and evaporated under reduced pressure. Purification was performed by flash column chromatography (1:4 EtOAc:Hexane) to yield aldehyde **III-12** (86 mg, 65%) as an inseparable mixture of two isomers in ratio 10:3. Spectral data for major isomer: ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.95 (m,

4H), 2.31 (s, 3H), 2.72 (m, 4H), 3.19 (m, 4H), 5.98 (d, J = 8.0 Hz, 1H), 6.64 (d, J = 16.0 Hz, 1H), 6.64 (d, J = 16.0 Hz, 1H), 6.91 (d, J = 16.0 Hz, 1H), 6.94 (s, 2H), 10.07 (d, J = 8.0 Hz, 1H); minor isomer δ 1.95 (m, 4H), 2.13 (s, 3H), 2.72 (m, 4H), 5.78 (d, J = 8.0 Hz, 1H), 6.82 (d, J = 15.5 Hz, 1H), 6.95 (s, 2H), 7.58 (d, J = 15.5 Hz, 1H), 10.26 (d, J = 8.0 Hz, 1H); both isomers ^{13}C-NMR (125 MHz, CDCl$_3$): δ 12.9, 14.1, 21.2, 21.6, 21.9, 27.6, 27.8, 49.8, 49.9, 117.4, 121.1, 121.1, 121.2, 122.8, 125.3, 126.1, 126.6, 126.7, 127.6, 136.8, 137.8, 144.0, 155.6, 155.9, 189.8, 190.8. UV-vis (ethanol): λ_{max} = 445 nm, ε = 30,000 M^{-1}cm^{-1}. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{18}$H$_{22}$NO]$^+$ 268.1701, found 268.1699.

Synthesis of III-15. (If *trans*-3-methyl-2-penten-4-ynol was yellow, it was distilled before reaction, 95 ^{o}C/40 mbar). Oven dried 50 mL round bottom flask with a stir bar was purged with argon and charged with *trans*-3-methyl-2-penten-4-ynol **III-17** (288 mg, 3 mmol), bis(triphenylphosphine) palladium(II) chloride (42 mg, 0.06 mmol) and dry THF (9 mL). The mixture was degassed (liquid nitrogen freeze followed by vacuum, three times) and warmed to 0 ^{o}C. Tributyltin hydride (1050 mg, 3.6 mmol) was added dropwise within 5 min and the reaction was warmed to room temperature. Consumption of **III-17** was monitored by TLC (1:3 EtOAc:Hexane, Rf=0.63). After 1.5 h, the reaction was quenched with water, extracted with diethyl ether (25 mL × 3) and combined organics were dried over sodium sulfate. Solvent was removed on a rotary evaporator and the crude was purified by column chromatography (1:4 EtOAc:Hexane). After purification, 250

mg (28%) of hydrostannylated alcohol **III-15** was isolated as a transparent oil, which turned yellowish under vacuum within 1 h. Spectral data:[53] ^{1}H-NMR (300 MHz, CDCl$_3$) δ 0.85 (m, 15H), 1.30 (m, 7H), 1.48 (m, 6H), 1.77 (s, 3H), 4.29 (d, 2H, J = 6.9 Hz), 5.63 (t, 1H, J = 6.9 Hz), 6.23 (d, 1H, J = 19.2 Hz), 6.53 (d, 1H, J = 19.2 Hz).

Synthesis of III-16. Solution of vinyl cyanide **III-3** (20 mg, 0.089 mmol) in dry CH$_2$Cl$_2$ (5 mL) was treated with Br$_2$ (4.6 μl, 14.28 mg, 0.089 mmol) while stirring on dry ice/acetone bath. Reaction continued at the same temperature for 1 h followed by dropwise addition of dry Et$_3$N (13.5 μl, 10 mg, 0.1 mmol). The solution was brought to room temperature and left stirring overnight. Unreacted bromine was removed by quenching the reaction with saturated Na$_2$S$_2$O$_3$. The organic layer was separated from water, washed with brine and dried over sodium sulfate. The solvent was removed by rotary evaporation and crude compound was purified by column chromatography (2:3 EtOAc:Hexane, R$_f$=0.5). The red crystals of vinyl bromide **III-16** were isolated as a mixture of two isomers 5:3 (15 mg, 56%, m. pt. = 131-132 °C). Spectral data for both isomers: ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.94 (m, 8H), 2.71 (m, 8H), 3.23 (m, 8H), 7.15 (s, 2H), 7.20 (s, 2H), 7.30 (s, 1H), 7.33 (s, 1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 21.3, 27.6, 27.7, 49.9, 117.6, 119.4, 120.4, 120.8, 128.2, 130.0, 144.9, 145.4, 149.0. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{15}$H$_{16}$N$_2$Br]$^+$ 303.0497, found 303.0493.

Synthesis of III-18. Solution of 4-methylpyridine (500 mg, 5.37 mmol) in acetone (5 mL) was treated with 2,4-dinitrochlorobenzene (1091 mg, 5.37 mmol) in a single portion. The mixture was heated to reflux for 12 h, cooled, and the solvent was removed in vacuo. The solid residue was filtered and washed several times with hexanes to afford 1580 mg of *N*-(2,4-dinitrophenyl)-4-methylpyridinium chloride (**III-18**) as a purple tar (99%). Previously published spectral data matched in chemical shifts with the one obtained herein, with the exception of the proton at 8.26 ppm, which was assigned as 1H, although it is 2H.[59] Spectral data: ^{1}H NMR (500 MHz, DMSO): δ 2.77 (s, 3H), 8.26 (d, *J* = 6.5, 2H). 8.38 (d, *J* = 8.5, 1H), 8.94 (dd, *J* = 8.5, 2.5, 1H), 9.10 (d, *J* = 2.5, 1H), 9.21 (d, *J* = 8.5, 2H).

Synthesis of III-19. Sodium iodide (2 g, 13.34 mmol) was added to a solution of ethyl propyolate (600 mg, 6.12 mmol) in acetic acid (3 mL) and the mixture was heated to 70 ºC for 12 h. The reaction was diluted with ether (3 mL) followed by addition of 1N NaOH (10 mL). The aqueous layer was separated and extracted three times with ether. Combined organic layers were washed with brine and dried over sodium sulfate. After solvent evaporation, 1383 mg (100%) iodoacrylate **III-19** was isolated as a transparent oil. Due to the sensitivity of the compound, it was stored under argon at -20 ºC. In case when it turned yellow or brown it was washed with 1 M sodium bisulfate and purity was verified by NMR. Spectral data:[69] ^{1}H NMR (500 MHz, CDCl₃) δ 1.29 (m, 3H), 4.22 (m, 2H), 6.86

(m, 1H), 7.41 (m, 1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 14.1, 60.7, 94.6, 129.9, 164.5.

Synthesis of III-23. Protocol 1. Julolidine methylenemalononitrile for reduction was prepared using condensation between julolidine aldehyde **III-2** (201 mg, 1.0 mmol) and malononitrile (66 mg, 1.0 mmol) in ethanol with a drop of triethylamine. Condensation proceeded for 10 min and was subsequently filtered, washed with ethanol and dried under vacuum for 1 h. The resultant dicyanide (249 mg, 1.0 mmol) was dissolved in dry dichloromethane (2 mL) and cooled to -78 °C. Solution of DIBAL (0.5 mL of 1 M solution in hexane, 0.5 mmol) was added slowly *via* syringe, the flask was placed on ice bath and left stirring for 1 h. The reaction was quenched with Rochelle's salt and extracted with dichloromethane. After rotary evaporation the crude mixture was purified by column chromatography (1:1 EtOAc:Hexane) to yield 60 mg (25%) of pure aldehyde **III-23** as a single isomer (red crystals).

Protocol 2. Flame dried flask charged with DMF (0.4 mL) was kept on ice bath for 10 min. Phosphorus oxychloride (0.1 mL or 160 mg, 1 mmol) was added and the reaction was stirred at 0 °C for 10 min. Solution of cyano julolidine **III-3** (22 mg, 0.1 mmol) in DMF (0.1 mL) was added at once and the reaction was heated to 60 °C for 2 h. The mixture was poured on ice, basified to pH 10, extracted with ethyl acetate, dried over sodium sulfate and evaporated in vacuo. Crude compound was purified by flash chromatography (1:2 EtOAc:Hexane basified with 1% Et$_3$N, R_f = 0.16) to yield 22 mg of cyano julolidine aldehyde **III-**

23 (89%, m. pt. = 144-146 °C). Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.94 (p, J = 6.5 Hz, 4H), 2.72 (t, J = 6.5 Hz, 4H), 3.34 (t, J = 6.5 Hz, 4H), 7.45 (s, 1H), 7.48 (s, 2H), 9.35 (s, 1H); ^{13}C NMR (125 MHz, CDCl$_3$): δ 20.9, 27.4, 50.3, 102.0, 116.9, 118.4, 121.1, 132.1, 148.8, 157.4, 187.8. UV-vis (ethanol): λ_{max} = 480 nm. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{16}$H$_{17}$N$_2$O]$^+$ 253.1341, found 253.1343.

Synthesis of methyl ketone III-26. To a suspension of AlCl$_3$ (198 mg, 1.5 mmol) in dry CH$_2$Cl$_2$ (3 mL) was added acetylchloride (117 mg or 0.1 mL, 1.5 mmol) at 0 °C. After 10 min the solution became yellow and julolidine **III-3** dissolved in CH$_2$Cl$_2$ (1 mL) was added dropwise. Reaction continued for 3 h at room temperature and quenched by pouring on ice. The product was extracted with ethyl acetate (10 mL) twice, aqueous layer was basified to pH 10 and extracted one more time with ethyl acetate (10 mL). Combined organic layers were dried over sodium sulfate and evaporated. Crude compound was purified by column chromatography (1:4 EtOAc:Hexane basified with 1% Et$_3$N, R$_f$=0.25) to yield 120 mg (45%) of red solid **III-26**. Note: there was another distinct by-product on TLC (yellow, R$_f$=0.75), however, it was not isolated in detectable amount. Spectral data of **III-26**: ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.94 (p, J = 6.5 Hz, 4H), 2.46 (s, 3H), 2.72 (t, J = 6.5 Hz, 4H), 3.32 (t, J = 6.5 Hz, 4H), 7.51 (s, 2H), 7.87 (s,

1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 21.0, 27.5, 27.8, 50.2, 99.4, 118.3, 120.2, 120.9, 132.2, 148.1, 152.6, 192.4. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{17}$H$_{18}$N$_2$O]$^+$ 267.1497, found 267.1500.

Synthesis of 3-chloro-4-cyano-5-julolidine retinal analog (Cl-CJRA) **III-27**. Flame dried flask with DMF (2 mL) was kept on ice bath for 10 min. Phosphorus oxychloride (690 mg or 0.42 mL, 4.5 mmol) was added and the reaction was stirred at 0 ºC for 10 min. Solution of ketone **III-23** (120 mg, 0.45 mmol) in DMF (2 mL) was added slowly while stirring on ice and reaction continued at room temperature for 2 h. The mixture was poured on ice, basified to pH 10, extracted with ethyl acetate, dried over sodium sulfate and evaporated under vacuum. The crude compound was purifies by column chromatography (1:19→1:9→1:4 EtOAc:Hexane basified with 1% Et$_3$N) four times (last time with HPLC grade solvents) to yield 25 mg of chloro cyano julolidine aldehyde **Cl-CJRA** (14%). Note: there were three distinct spots on TLC after reaction and isolated spot with R$_f$=0.35 in 1:4 EtOAc:Hexane colored purple on TLC without staining corresponding to the assigned product; two other products were not isolated in pure form for structure determination. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.95 (p, J = 6.5 Hz, 4H), 2.73 (t, J = 6.5 Hz, 4H), 3.32 (t, J = 6.5 Hz, 4H), 6.66 (d, J = 7.2 Hz, 1H), 7.49 (s, 2H), 7.71 (s, 1H), 10.14 (d, J = 7.2 Hz, 1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 21.1, 27.6, 50.2, 97.2, 117.0, 118.9, 121.1,

122.5, 131.5, 147.6, 148.6, 148.7, 190.6. UV-vis (ethanol): λ_{max} = 505 nm, ε = 30,800 M^{-1} × cm^{-1}. HRMS (ESI) (m/z): $[M+H]^{+}$ calculated for $[C_{18}H_{18}N_2OCl]^{+}$ 313.1108, found 313.1107.

Synthesis of III-29 was the same as preparation of **III-11**. Used were aldehyde **III-35** (1020 mg, 5.07 mmol), NaH (243 mg 60% in mineral oil, 6.09 mmol), phosphonate **III-8** (1321 mg, 6.09 mmol) and THF (40 mL). The crude was purified *via* column chromatography (1:4 EtOAc:Hexane) and nitrile **III-34** was isolated as an oily mixture of two isomers in ratio of 4:3 (454 mg, 34%). Spectral data for both isomers: ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.56 (s, 6H), 1.56 (s, 6H), 2.04 (s, 3H), 2.19 (s, 3H), 3.12 (s, 3H), 3.14 (s, 3H), 4.82 (s, 1H), 4.97 (s, 1H), 5.32 (d, *J* = 12.5 Hz, 1H), 5.43 (d, *J* = 12.5 Hz, 1H), 6.03 (d, *J* = 15.5 Hz, 1H), 6.54 (d, *J* = 15.5 Hz, 1H), 6.39 (m, 2H), 6.86 (m, 2H), 7.12-7.18 (m, 6H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 16.6, 19.5, 28.5, 29.1, 29.2, 45.9, 89.7, 91.3, 95.5, 95.8, 106.3, 106.4, 119.5, 120.17, 120.22, 121.1, 121.5, 121.6, 123.3, 127.9, 133.1, 133.6, 138.7, 144.7, 156.9, 157.5, 160.8.

Synthesis of merocyanine retinal analog (MCRA) III-30. The same protocol for the preparation of **III-12** was used for the synthesis of **III-30**. Nitrile **III-29** (80 mg, 0.3 mmol), DIBAL (0.9 mL, 1M solution in hexane, 0.9 mmol) and CH$_2$Cl$_2$ (4 mL) were the amounts used in this reaction. The final product was isolated as a red oil, which turns into a red solid during storage at -78 ºC (48 mg,

59%). Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.58 (s, 6H), 2.31 (s, 3H), 3.16 (s, 3H), 5.42 (d, J = 12.0 Hz, 1H), 5.89 (d, J = 8.5 Hz, 1H), 6.12 (d, J = 15.5 Hz, 1H), 6.66 (d, J = 8.5 Hz, 1H), 6.88 (t, J = 8.5 Hz, 1H), 7.13-7.19 (m, 3H), 7.38 (dd, J = 12.0 Hz, 15.5 Hz 1H), 10.02 (d, J = 8.5 Hz, 1H).

Synthesis of the methyl ketone merocyanine retinal analog (mkMCRA) III-31. A flame dried 25 mL round-bottom flask was placed on ice bath and charged with dry ether (3 mL) and CH$_3$Li (2.5 mL, 1.6 M solution in ether, 3.9 mmol). Solution of nitrile **III-29** (344 mg, 1.3 mmol) in ether (3 mL) was added slowly at 0 ºC. After complete consumption of **III-29** from TLC (30 min) the reaction was quenched with a few drops of ethyl acetate, methanol, and then ice-cold 1N HCl (3 mL dropwise) while stirring on ice. After vigorous stirring for an additional 1 h the aqueous layer was separated from the organics and was extracted with ether (3 x 5 mL). Combined organic layers were dried over sodium sulfate and evaporated. The crude was purified by column chromatography (1:4 EtOAc:Hexane) to give an inseparable mixture of two isomers mkMCRA in ratio 3:4 (25 mg, 7%). Note: when 1.0 equiv of methyllithium was used in the first attempt to synthesize mkMCRA 75% of starting nitrile **III-29** was recovered and no mkMCRA was isolated. Spectral data for both isomers: ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.56 (s, 6H)/1.57 (s, 6H), 2.05 (s, 3H)/2.16 (s, 3H), 2.18 (s, 3H)/2.35 (s, 3H), 3.13 (s, 3+3H), 5.39 (d, J = 12.0 Hz, 1H)/5.51 (d, J = 12.0 Hz, 1H), 5.83 (s, 1H)/6.02 (s, 1H), 6.00 (d, J = 15.5 Hz, 1H), 6.62 (d, J = 8.0 Hz, 2H), 6.85 (m, 2H),
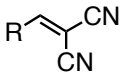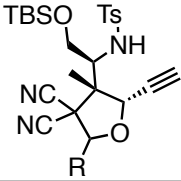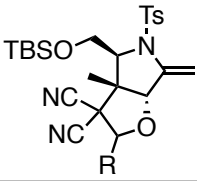
7.11-7.17 (m, 4H), 7.28 (d, J = 15.5 Hz, 1H), 7.31 (d, J = 15.5 Hz, 1H), 7.65 (d, J = 15.5 Hz, 1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 14.1, 14.3, 21.1, 22.7, 24.4, 28.5, 28.6, 29.1, 29.4, 29.7, 30.2, 31.9, 32.0, 45.8, 96.3, 97.3, 106.3, 119.9, 120.0, 120.3, 121.6, 122.4, 122.6, 127.4, 127.8, 133.3, 135.2, 138.9, 144.9, 150.9, 152.9, 159.8, 160.2, 198.0, 198.5. UV-vis (ethanol): λ_{max} = 460 nm, ε = 30,500 M^{-1}cm^{-1}. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{19}$H$_{24}$NO]$^+$ 282.1858, found 282.1869.

Synthesis of III-36. Flame dried 3 mL vial with a stir bar was loaded with NaH (2 mg, 60% in mineral oil 0.08 mmol) and methylnitrocinnamate (13 mg, 0.06 mmol). The solids were purged with argon and dissolved in dry DMF (0.4 mL). Aziridinol **III-37** (8.2 mg, 0.02 mmol) in DMF (0.1 mL) was added at 0 °C *via* syringe during 5 min. Reaction continued on ice slowly melting to room temperature over 1 h and then for an additional 2 h at room temperature. The reaction was quenched with saturated NH$_4$Cl (1 mL), extracted with ethyl acetate (3 x 2 mL) and dried over anhydrous sodium sulfate. The crude product was purified by column chromatography (5:95 EtOAc:Hexane) and isolated as yellow oil (11 mg, 96%). The mixture of two diastereomers (ratio 3:2 from NMR) could not be separated by HPLC. Spectral data for major diastereomer: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.03 (s, 3H), 0.05 (s, 3H), 0.86 (s, 9H), 1.21 (s, 3H), 1.28 (s, 3H), 2.34 (s, 3H), 3.51 (dd, J = 4.0 Hz, 9.8 Hz, 2H), 4.01 (dd, J = 4.0 Hz, 9.8 Hz, 1H), 5.18 (t, J = 2.0 Hz, 1H), 5.40 (s, 1H) 5.59 (s, 1H), 5.70 (s, 1H), 7.19 (3.51 (d, J =

8.5 Hz, 2H), 7.25-7.33 (m, 5H), 7.86 (d, J = 8.5 Hz, 2H); Spectral data for minor

diastereomer: δ 0.01 (s, 3H), 0.03 (s, 3H), 0.86 (s, 9H), 1.33 (s, 3H), 1.37 (s, 3H),

2.37 (s, 3H), 3.31 (dd, J = 4.0 Hz, 9.8 Hz, 2H), 3.93 (dd, J = 4.0 Hz, 9.8 Hz, 1H),

5.36 (t, J = 2.5 Hz, 1H), 5.41 (s, 1H) 5.58 (s, 1H), 6.04 (s, 1H), 7.14 (3.51 (d, J =

8.5 Hz, 2H), 7.25-7.33 (m, 5H), 7.87 (d, J = 8.5 Hz, 2H); ^{13}C-NMR for both (125

MHz, CDCl$_3$): δ -5.7/-5.5, 12.6/12.8, 18.1/18.2, 21.5/21.6, 21.8/22.0, 25.7,

46.8/47.2, 53.9, 56.8, 59.9/60.0, 80.1/80.2, 84.8/85.8, 94.8/96.1, 111.3/111.8,

126.1/126.2, 127.7/127.8, 128.4/128.5, 129.3/129.4, 135.2/136.1, 136.7/137.0,

144.1/144.2, 149.5/150.2. HRMS (ESI) (m/z): [M+H]$^+$ calculated for

[C$_{29}$H$_{41}$N$_2$O$_6$SSi]$^+$ 573.2455, found 573.2455.

Table III-4: Structures of the compounds synthesized and characterized.

R	R⟋CN / CN	TBSO/Ts-NH NC/NC structure	TBSO/Ts-N NC/NC structure
p-NMe$_2$-C$_6$H$_4$	**III-46**	N/A	**III-58**
p-OMe-C$_6$H$_4$	**III-47**	**III-54**	**III-59**
o-OMe-C$_6$H$_4$	**III-48**	N/A	**III-60**
α-naphtyl	**III-49**	N/A	**III-61**
Ph	**III-41**	**III-44**	**III-45**
p-CH$_3$-C$_6$H$_4$	**III-51**	N/A	**III-63**
p-Cl-C$_6$H$_4$	**III-52**	**III-56**	**III-64**
p-Br-C$_6$H$_4$	**III-53**	**III-57**	N/A

N/A - not applicable

General method for preparation of aromatic methylenemalononitriles

III-41 and III-46–III-53. An aromatic aldehyde (5 mmol) and freshly re-crystallized

malononitrile (330 mg, 5 mmol) were dissolved in ethanol (5 mL). After addition of

three drops of triethylamine a precipitate is formed immediately. Condensation

continued for 1 h and the product was filtered and washed with ice-cold ethanol.

After drying the solid in vacuum corresponding benzylidenemalononitriles were isolated in pure form in quantitative yields. Note 1: if reaction with *p*-chlorobenzaldehyde and *p*-bromobenzaldehyde proceeded longer than 1 h the product started decomposing by turning brown and dissolving in ethanol. Note 2: electron deficient benzaldehydes (*p*-cyanobenzaldehyde and *p*-nitrobenzaldehyde) did not give the corresponding benzylidenemalononitriles with this protocol. Note 3: if the reaction was done with not freshly purified malononitrile the yields and purity of products decreased significantly.

III-46. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 3.12 (s, 6H), 6.67 (d, J = 9.5 Hz, 2H), 7.43 (s, 1H), 7.78 (d, J = 9.5 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 40.1, 111.6, 114.9, 115.9, 119.3, 133.8, 154.2, 158.1.

III-47. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 3.90 (s, 3H), 6.99 (d, J = 8.5 Hz, 2H), 7.63 (s, 1H), 7.88 (d, J = 8.5 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 55.8, 78.7, 113.3, 114.4, 115.1, 124.0, 133.4, 158.8, 164.8.

III-48. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 3.91 (s, 3H), 6.97 (d, J = 8.5 Hz, 1H), 7.06 (t, J = 8.5 Hz, 1H), 7.56 (t, J = 8.5 Hz, 1H), 8.16 (d, J = 8.5 Hz, 1H), 8.29 (s, 1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 55.9, 81.5, 111.5, 113.0, 114.3, 120.2, 121.2, 128.9, 136.5, 154.5, 158.9.

III-49. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.59 (d, J = 7.5 Hz, 1H), 7.61 (t, J = 7.5 Hz, 1H), 7.66 (t, J = 8.5 Hz, 1H), 7.93 (d, J = 8.5 Hz, 1H), 8.08 (d,

J = 7.5 Hz, 1H), 8.25 (d, J = 7.5 Hz, 1H), 8.62 (s, 1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 85.1, 112.5, 113.7, 122.3, 125.3, 127.2, 127.5, 128.4, 128.5, 129.4, 131.0, 133.5, 134.9, 157.6.

III-41. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.53 (t, J = 7.5 Hz, 2H), 7.62 (t, J = 7.5 Hz, 1H), 7.76 (s, 1H), 7.89 (d, J = 7.5 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 82.9, 112.5, 113.7, 129.6, 130.7, 130.9, 134.6, 159.9.

III-51. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 2.44 (s, 3H), 7.32 (d, J = 7.5 Hz, 2H), 7.70 (s, 1H), 7.79 (d, J = 7.5 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 22.0, 81.2, 112.8, 114.0, 128.4, 130.3, 130.9, 146.3, 159.7.

III-52. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.50 (d, J = 7.5 Hz, 2H), 7.71 (s, 1H), 7.83 (d, J = 7.5 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 83.4, 112.3, 113.4, 129.3, 130.1, 131.8, 141.2, 158.2.

III-53. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.67 (d, J = 8.0 Hz, 2H), 7.69 (s, 1H), 7.75 (d, J = 8.0 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ 83.6, 112.3, 113.4, 129.7, 129.9, 131.8, 133.1, 158.4.

General method for aza-Payne cyclization and temperature/substrate dependent hydroamination (compounds III-44, III-45, III-54–III-64). Flame dried 3 mL vial with a stir bar was loaded with aziridinol **III-37** (8.2 mg, 0.02 mmol) and corresponding benzylidenemalononitrile, **III-41** or **III-46-III-53** (0.06

mmol). The solids were purged with argon and dissolved in dry DMF (0.2 mL). After 10 min of stirring, a suspension of NaH (5 mg, 60% in mineral oil 0.2 mmol) in DMF (0.3 mL) was added dropwise at 0 °C *via* syringe during 5 min. Reaction continued on ice slowly melting to room temperature over 1 h and then for an additional 2 h at room temperature. If the reaction required additional heat, the vial containing the mixture was transferred to an oil bath at specific temperature for another 2 h. The reaction was quenched with saturated NH$_4$Cl (1 mL), extracted with ethyl acetate (3 x 2 mL) and dried over anhydrous sodium sulfate. The crude product was purified by column chromatography (EtOAc:Hexane) with solvent polarity dictated by the compound.

III-54. Isolated 4 mg, 34%. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.14 (d, J = 2.5 Hz, 6H), 0.93 (s, 9H), 1.49 (s, 3H), 2.4 (s, 3H), 2.67 (d, J = 2.5 Hz, 1H), 3.77 (s, 3H), 3.77 (m, 1H), 4.52 (m, 2H), 4.86 (d, J = 2.5 Hz, 1H), 5.73 (s, 1H), 6.88 (d, J = 7.5 Hz, 2H), 7.32 (d, J = 7.5 Hz, 2H), 7.35 (d, J = 7.5 Hz, 2H), 7.69 (d, J = 77.5 Hz, 2H), 9.52 (s, 1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ -5.6, 18.2, 21.7, 25.9, 31.5, 34.6, 55.2, 62.1, 62.7, 65.3, 74.3, 78.4, 79.4, 85.3, 114.0, 127.4, 127.7, 128.2, 130.3, 132.6, 146.1, 159.6, 159.9. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{31}$H$_{40}$N$_3$O$_5$SSi]$^+$ 594.2458, found 594.2459.

III-44. After column chromatography (5:95 EtOAc:Hexane) compound was purified by HPLC. Isolated 8 mg, 70%. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.09 (d, J = 4.5 Hz, 6H), 0.94 (s, 9H), 1.49 (s, 3H), 2.37 (s, 3H), 2.68 (d, J = 4.5

Hz, 1H), 3.78 (dd, J = 6.5 Hz, 9 Hz, 1H), 4.49-4.88 (m, 2H), 4.88 (d, J = 2.0 Hz, 1H), 5.78 (s, 1H), 7.30-7.37 (m, 5H), 7.45 (d, J = 8.0 Hz, 2H), 7.70 (d, J = 8.0 Hz, 2H), 9.53 (s, 1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ -5.6, 18.3, 21.7, 24.9, 31.6, 55.9, 62.1, 62.7, 65.4, 74.5, 78.4, 79.4, 85.3, 113.9, 126.3, 127.4, 128.6, 128.8, 129.6, 130.3, 132.7, 136.2, 146.1, 159.5. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{30}$H$_{38}$N$_3$O$_4$SSi]$^+$ 564.2352, found 564.2352.

III-56. Isolated 9 mg (82%) if heated at 40 ºC for additional 2 h. Isolated 5 mg (42%) if heated at 70 ºC for additional 2 h. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.14 (d, J = 3.5 Hz, 6H), 0.93 (s, 9H), 1.46 (s, 3H), 2.03 (s, 3H), 2.69 (d, J = 1.5 Hz, 1H), 3.78 (dd, J = 6.5 Hz, 8.5 Hz, 1H), 4.47-4.51 (m, 2H), 4.87 (d, J = 1.5 Hz, 1H), 5.74 (s, 1H), 7.33 (d, J = 8.5 Hz, 4H), 7.39 (d, J = 8.5 Hz, 2H), 7.69 (d, J = 8.5 Hz, 2H), 9.52 (s, 1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ -5.6, 18.2, 21.7, 25.0, 25.9, 55.9, 62.1, 62.6, 65.3, 74.5, 78.2, 79.6, 84.6, 113.7, 127.4, 127.6, 128.9, 130.3, 134.7, 134.8, 146.2, 159.3. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{30}$H$_{37}$N$_3$O$_4$SClSi]$^+$ 598.1963, found 598.1962.

III-57. Isolated 9 mg (71%) if no additional heat applied. Isolated 5 mg (39%) if heated at 40 ºC for additional 2 h. Isolated 12 mg (94%) if heated at 70 ºC for additional 2 h. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.14 (d, J = 3.5 Hz, 6H), 0.93 (s, 9H), 1.45 (s, 3H), 2.39 (s, 3H), 2.69 (d, J = 2.4 Hz, 1H), 3.79 (dd, J = 6.5 Hz, 9 Hz, 1H), 4.48-4.52 (m, 2H), 4.87 (d, J = 2.4 Hz, 1H), 5.73 (s,

1H), 7.33 (d, J = 8.5 Hz, 4H), 7.48 (d, J = 8.5 Hz, 2H), 7.69 (d, J = 8.5 Hz, 2H), 9.52 (s, 1H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ -5.5, 18.3, 21.7, 25.0, 25.9, 55.9, 62.1, 62.7, 65.4, 74.6, 78.2, 79.6, 84.6, 113.7, 122.9, 127.4, 127.9, 130.3, 131.8, 132.7, 133.1, 135.3, 146.2, 159.3. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{30}$H$_{37}$N$_3$O$_4$SBrSi]$^+$ 642.1457, found 642.1458.

III-58. After column chromatography **III-58** was isolated as inseparable mixture with N,N-dimethylbenzaldehyde (retro-aldol product). To remove it the mixture was dissolved in methanol and treated with aqueous NaHSO$_3$ for 1 h. After extraction with ethyl acetate pure **III-58** was isolated (6 mg, 50%) as a yellow oil. The product was dissolved in CH$_2$Cl$_2$ and covered with a thin layer of hexane, yielding yellow crystals after 5 days at 0 ºC (m. pt. = 128-129 °C). Crystal structure of **III-58** was solved by Dr. Richard Staples. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.13 (d, J = 2.5 Hz, 6H), 0.90 (s, 9H), 1.71 (s, 3H), 2.30 (s, 3H), 2.94 (s, 6H), 3.87 (s, 1H), 4.08-4.10 (m, 2H), 4.31 (dd, J = 4.5 Hz, 3 Hz, 1H), 4.61 (s, 1H), 4.76 (s, 1H), 5.36 (s, 1H), 6.61 (d, J = 8.5 Hz, 2H), 6.94 (d, J = 8.5 Hz, 2H), 7.28 (d, J = 8.5 Hz, 2H), 7.89 (d, J = 8.5 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ -5.6, 18.2, 21.6, 25.9, 40.2, 54.5, 56.7, 63.6, 68.0, 77.2, 84.7, 90.9, 98.7, 111.7, 112.8, 119.0, 127.1, 128.0, 129.8, 134.9, 142.6, 145.2, 151.4. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{32}$H$_{43}$N$_4$O$_4$SSi]$^+$ 607.2774, found 607.2775.

III-59. After column chromatography **III-59** was isolated as transparent oil (10 mg, 84%). Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.13 (d, J = 2.5 Hz, 6H), 0.91 (s, 9H), 1.71 (s, 3H), 2.26 (s, 3H), 3.79 (s, 3H), 3.87 (s, 1H), 4.07-4.10 (m, 2H), 4.29 (dd, J = 5.5 Hz, 2.5 Hz, 1H), 4.62 (s, 1H), 4.78 (s, 1H), 5.39 (s, 1H), 6.85 (d, J = 8.5 Hz, 2H), 7.03 (d, J = 8.5 Hz, 2H), 7.26 (d, J = 8.5 Hz, 2H), 7.88 (d, J = 8.5 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ -5.6, 18.1, 21.5, 25.9, 54.4, 55.3, 56.9, 63.6, 67.9, 84.0, 91.1, 99.2, 111.5, 112.6, 113.9, 124.2, 127.4, 128.0, 129.8, 134.9, 142.5, 145.2, 160.8. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{31}$H$_{40}$N$_3$O$_5$SSi]$^+$ 594.2458, found 594.2458.

III-60. After column chromatography **III-60** was isolated as transparent oil (10 mg, 84%). Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.11 (d, J = 4 Hz, 6H), 0.88 (s, 9H), 1.65 (s, 3H), 2.24 (s, 3H), 3.84 (s, 3H), 3.99 (dd, J = 11.5 Hz, 1.5 Hz, 1H), 4.19 (dd, J = 11 Hz, 3.5 Hz, 1H), 4.54 (dd, J = 3.5 Hz, 1.5 Hz, 1H), 4.63 (s, 1H), 4.73 (s, 1H), 4.87 (s, 1H), 5.09 (s, 1H), 6.87 (d, J = 7.5 Hz, 2H), 6.97 (t, J = 7.5 Hz, 1H), 7.21 (d, J = 8.0 Hz, 2H), 7.33 (t, J = 8.0 Hz, 1H), 7.40 (d, J = 8.0 Hz, 1H), 7.91 (d, J = 8.0 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ -5.8, -5.6, 18.1, 18.5, 21.5, 25.8, 52.1, 54.5, 57.3, 63.0, 67.0, 79.9, 90.1, 95.7, 109.9, 111.6, 111.9, 113.2, 120.7, 122.3, 126.8, 128.1, 129.5, 130.6, 135.3, 135.9, 142.0, 144.6, 155.9. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{31}$H$_{40}$N$_3$O$_5$SSi]$^+$ 594.2458, found 594.2458.

III-61. After column chromatography it was isolated as yellow oil (11 mg, 71%). Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.15 (s, 6H), 0.92 (s, 9H), 1.75 (s, 3H), 2.03 (s, 3H), 4.07 (dd, J = 11.5 Hz, 2.5 Hz, 1H), 4.22 (dd, J = 11.5 Hz, 3.5 Hz, 1H), 4.55 (dd, J = 3.5 Hz, 2.5 Hz, 1H), 4.75 (s, 1H), 4.84 (s, 1H), 5.25 (s, 1H), 5.30 (s, 1H), 7.23 (d, J = 8.5 Hz, 2H), 7-48-7.54 (m, 4H), 7.71 (d, J = 8.5 Hz, 1H), 7.86 (d, J = 8.5 Hz, 2H), 7.99 (d, J = 8.5 Hz, 2H), 7.91 (d, J = 8.5 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ -5.7, -5.6, 18.1, 18.6, 21.3, 25.8, 53.0, 57.7, 63.3, 68.2, 80.5, 90.6, 96.9, 111.4, 113.0, 122.6, 125.1, 125.6, 126.0, 126.4, 128.1, 128.5, 129.0, 129.8, 130.5, 133.5, 135.3, 142.1, 155.0. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{34}$H$_{40}$N$_3$O$_4$SSi]$^+$ 614.2509, found 614.2510.

III-45. Reaction was heated at 35 °C for an additional 2 h. After column chromatography **III-45** was isolated as a transparent oil (8 mg, 71%), which formed white crystals after storage (m. pt. = 106-107 °C). Crystal structure of **III-45** was solved by Dr. Richard Staples. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.14 (d, J = 2.5 Hz, 6H), 0.91 (s, 9H), 1.72 (s, 3H), 2.21 (s, 3H), 3.90 (s, 1H), 4.09-4.13 (m, 2H), 4.30 (dd, J = 5.5 Hz, 2.5 Hz, 1H), 4.64 (s, 1H), 4.80 (s, 1H), 5.41 (s, 1H), 7.10 (d, J = 8.5 Hz, 2H), 7.25 (d, J = 8.5 Hz, 2H), 7.32-7.37 (m, 3H), 7.88 (d, J = 8.5 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ -5.6, 18.1, 21.4, 25.9, 54.1, 57.1, 63.6, 67.9, 84.0, 91.2, 99.4, 111.3, 112.5, 125.9, 128.0, 128.6, 129.8, 132.4, 134.9, 142.4, 145.3. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{30}$H$_{38}$N$_3$O$_4$SSi]$^+$ 564.2352, found 564.2352.

III-63. After column chromatography it was isolated as a transparent oil (11 mg, 94%). Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.13 (d, J = 2.5 Hz, 6H), 0.90 (s, 9H), 1.71 (s, 3H), 2.24 (s, 3H), 2.33 (s, 3H), 3.87 (s, 1H), 4.08-4.12 (m, 2H), 4.31 (dd, J = 3.5 Hz, 2.5 Hz, 1H), 4.64 (s, 1H), 4.78 (s, 1H), 5.34 (s, 1H), 6.98 (d, J = 8.5 Hz, 2H), 7.14 (d, J = 8.5 Hz, 2H), 7.25 (d, J = 8.5 Hz, 2H), 7.88 (d, J = 8.5 Hz, 2H); ^{13}C-NMR (125 MHz, CDCl$_3$): δ -5.6, 18.1, 21.3, 21.5, 25.9, 54.2, 56.9, 63.6, 67.9, 84.0, 91.1, 99.0, 111.4, 112.6, 125.8, 127.8, 128.0, 129.3, 129.7, 129.8, 134.8, 139.9, 142.4, 145.3. HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{31}$H$_{40}$N$_3$O$_4$SSi]$^+$ 578.2509, found 578.2508.

III-64. Reaction was heated at 70 ºC for an additional 2 h, yielding **III-56** and **III-64** in a ratio 5:4 from crude NMR. After column chromatography **III-56** was isolated pure (5 mg, 42%) and **III-64** was isolated with 17% of monocyclized product (4 mg of 83% pure, 25% yield). Isolated **III-64** crystallized as a white solid (m. pt. = 138-139 °C) from methanol/dichloromethane 1:1. Crystal structure of **III-64** was solved by Dr. Richard Staples. Spectral data: ^{1}H NMR (500 MHz, CDCl$_3$) δ 0.14 (d, J = 3.5 Hz, 6H), 0.90 (s, 9H), 1.71 (s, 3H), 2.24 (s, 3H), 3.85 (s, 1H), 4.04-4.13 (m, 2H), 4.28 (d, J = 3.5 Hz, 1H), 4.64 (s, 1H), 4.80 (s, 1H), 5.41 (s, 1H), 7.05 (d, J = 8.5 Hz, 2H), 7.26 (d, J = 8.5 Hz, 2H), 7.33 (d, J = 8.5 Hz, 2H), 7.87 (d, J = 8.5 Hz, 2H). HRMS (ESI) (m/z): [M+H]$^+$ calculated for [C$_{30}$H$_{37}$N$_3$O$_4$SClSi]$^+$ 598.1963, found 598.1963.

REFERENCES

REFERENCES

1.   Uttamapinant, C.; Sanchez, M. I.; Liu, D. S.; Yao, J. Z.; Ting, A. Y., Site-specific protein labeling using PRIME and chelation-assisted click chemistry. *Nature Protocols* **2013,** *8* (8), 1620.

2.   Crivat, G.; Taraska, J. W., Imaging proteins inside cells with fluorescent tags. *Trends in Biotechnology* **2012,** *30* (1), 8.

3.   Jing, C.; Cornish, V. W., Chemical Tags for Labeling Proteins Inside Living Cells. *Accounts of Chemical Research* **2011,** *44* (9), 784.

4.   Sahoo, H., Fluorescent labeling techniques in biomolecules: a flashback. *Rsc Advances* **2012,** *2* (18), 7017.

5.   Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohane, R. F.; Urh, M.; Simpson, D.; Mendez, J.; Zimmerman, K.; Otto, P.; Vidugiris, G.; Zhu, J.; Darzins, A.; Klaubert, D. H.; Bulleit, R. F.; Wood, K. V., HatoTag: A novel protein labeling technology for cell imaging and protein analysis. *Acs Chemical Biology* **2008,** *3* (6), 373.

6.   Rashidian, M.; Dozier, J. K.; Distefano, M. D., Enzymatic Labeling of Proteins: Techniques and Approaches. *Bioconjugate Chemistry* **2013,** *24* (8), 1277.

7.   Shi, X.; Jung, Y.; Lin, L.-J.; Liu, C.; Wu, C.; Cann, I. K. O.; Ha, T., Quantitative fluorescence labeling of aldehyde-tagged proteins for single-molecule imaging. *Nature Methods* **2012,** *9* (5), 499.

8.   Howarth, M.; Ting, A. Y., Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nature Protocols* **2008,** *3* (3), 534.

9.   Uttamapinant, C.; White, K. A.; Baruah, H.; Thompson, S.; Fernandez-Suarez, M.; Puthenveetil, S.; Ting, A. Y., A fluorophore ligase for site-specific protein labeling inside living cells. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (24), 10914.

10.   Yao, J. Z.; Uttamapinant, C.; Poloukhtine, A.; Baskin, J. M.; Codelli, J. A.; Sletten, E. M.; Bertozzi, C. R.; Popik, V. V.; Ting, A. Y., Fluorophore Targeting to Cellular Proteins via Enzyme-Mediated Azide Ligation and Strain-Promoted Cycloaddition. *Journal of the American Chemical Society* **2012,** *134* (8), 3720.

11. Hao, Z.; Hong, S.; Chen, X.; Chen, P. R., Introducing Bioorthogonal Functionalities into Proteins in Living Cells. *Accounts of Chemical Research* **2011,** *44* (9), 742.

12. Jewett, J. C.; Bertozzi, C. R., Synthesis of a Fluorogenic Cyclooctyne Activated by Cu-Free Click Chemistry. *Organic Letters* **2011,** *13* (22), 5937.

13. Saxon, E.; Bertozzi, C. R., Cell surface engineering by a modified Staudinger reaction. *Science* **2000,** *287* (5460), 2007.

14. Agarwal, P.; van der Weijden, J.; Sletten, E. M.; Rabuka, D.; Bertozzi, C. R., A Pictet-Spengler ligation for protein chemical modification. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110* (1), 46.

15. Beatty, K. E.; Fisk, J. D.; Smart, B. P.; Lu, Y. Y.; Szychowski, J.; Hangauer, M. J.; Baskin, J. M.; Bertozzi, C. R.; Tirrell, D. A., Live-Cell Imaging of Cellular Proteins by a Strain-Promoted Azide-Alkyne Cycloaddition. *Chembiochem* **2010,** *11* (15), 2092.

16. Liu, D. S.; Tangpeerachaikul, A.; Selvaraj, R.; Taylor, M. T.; Fox, J. M.; Ting, A. Y., Diels-Alder Cycloaddition for Fluorophore Targeting to Specific Proteins inside Living Cells. *Journal of the American Chemical Society* **2012,** *134* (2), 792.

17. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angewandte Chemie-International Edition* **2009,** *48* (38), 6974.

18. Guo, Z.; Park, S.; Yoon, J.; Shin, I., Recent progress in the development of near-infrared fluorescent probes for bioimaging applications. *Chemical Society Reviews* **2014,** *43* (1), 16.

19. Karton-Lifshin, N.; Albertazzi, L.; Bendikov, M.; Baran, P. S.; Shabat, D., "Donor-Two-Acceptor" Dye Design: A Distinct Gateway to NIR Fluorescence. *Journal of the American Chemical Society* **2012,** *134* (50), 20412.

20. MacNevin, C. J.; Gremyachinskiy, D.; Hsu, C.-W.; Li, L.; Rougie, M.; Davis, T. T.; Hahn, K. M., Environment-Sensing Merocyanine Dyes for Live Cell Imaging Applications. *Bioconjugate Chemistry* **2013,** *24* (2), 215.

21. Redy-Keisar, O.; Kisin-Finfer, E.; Ferber, S.; Satchi-Fainaro, R.; Shabat, D., Synthesis and use of QCy7-derived modular probes for the detection and imaging of biologically relevant analytes. *Nature Protocols* **2014,** *9* (1), 27.

22.   Shank, N. I.; Pham, H. H.; Waggoner, A. S.; Armitage, B. A., Twisted Cyanines: A Non-Planar Fluorogenic Dye with Superior Photostability and its Use in a Protein-Based Fluoromodule. *Journal of the American Chemical Society* **2013,** *135* (1), 242.

23.   Shank, N. I.; Zanotti, K. J.; Lanni, F.; Berget, P. B.; Armitage, B. A., Enhanced Photostability of Genetically Encodable Fluoromodules Based on Fluorogenic Cyanine Dyes and a Promiscuous Protein Partner. *Journal of the American Chemical Society* **2009,** *131* (36), 12960.

24.   Shirinian, V. Z.; Shimkin, A. A., Merocyanines: Synthesis and Application. In *Heterocyclic Polymethine Dyes: Synthesis, Properties and Applications*, Strekowski, L., Ed. 2008; Vol. 14, pp 75.

25.   Su, X.; Liptak, M. D.; Aprahamian, I., Water soluble triazolopyridiniums as tunable blue light emitters. *Chemical Communications* **2013,** *49* (39), 4160.

26.   Vendrell, M.; Zhai, D.; Er, J. C.; Chang, Y.-T., Combinatorial Strategies in Fluorescent Probe Development. *Chemical Reviews* **2012,** *112* (8), 4391.

27.   Crist, R. M.; Vasileiou, C.; Rabago-Smith, M.; Geiger, J. H.; Borhan, B., Engineering a rhodopsin protein mimic. *Journal of the American Chemical Society* **2006,** *128* (14), 4522.

28.   Lee, K. S. S.; Berbasova, T.; Vasileiou, C.; Jia, X.; Wang, W.; Choi, Y.; Nossoni, F.; Geiger, J. H.; Borhan, B., Probing Wavelength Regulation with an Engineered Rhodopsin Mimic and a C15-Retinal Analogue. *Chempluschem* **2012,** *77* (4), 273.

29.   Vasileiou, C.; Lee, K. S. S.; Crist, R. M.; Vaezeslami, S.; Goins, S. A.; Geiger, J. H.; Borhan, B., Dissection of the critical binding determinants of cellular retinoic acid binding protein II by mutagenesis and fluorescence binding assay. *Proteins-Structure Function and Bioinformatics* **2009,** *76* (2), 281.

30.   Vasileiou, C.; Vaezeslami, S.; Crist, R. M.; Rabago-Smith, M.; Geiger, J. H.; Borhan, B., Protein design: Reengineering Cellular Retinoic Acid Binding Protein II into a rhodopsin protein mimic. *Journal of the American Chemical Society* **2007,** *129* (19), 6140.

31.   Vasileiou, C.; Wang, W.; Jia, X.; Lee, K. S. S.; Watson, C. T.; Geiger, J. H.; Borhan, B., Elucidating the exact role of engineered CRABPII residues for the formation of a retinal protonated Schiff base. *Proteins-Structure Function and Bioinformatics* **2009,** *77* (4), 812.

32.   Wang, W.; Geiger, J. H.; Borhan, B., The photochemical determinants of color vision. *Bioessays* **2014,** *36* (1), 65.

33.   Wang, W.; Nossoni, Z.; Berbasova, T.; Watson, C. T.; Yapici, I.; Lee, K. S. S.; Vasileiou, C.; Geiger, J. H.; Borhan, B., Tuning the Electronic Absorption of Protein-Embedded All-trans-Retinal. *Science* **2012,** *338* (6112), 1340.

34.   Alvarez, R.; Vaz, B.; Gronemeyer, H.; de Lera, A. R., Functions, Therapeutic Applications, and Synthesis of Retinoids and Carotenoids. *Chemical Reviews* **2014,** *114* (1), 1.

35.   Sheves, M.; Albeck, A.; Friedman, N.; Ottolenghi, M., CONTROLLING THE PKA OF THE BACTERIORHODOPSIN SCHIFF-BASE BY USE OF ARTIFICIAL RETINAL ANALOGS. *Proceedings of the National Academy of Sciences of the United States of America* **1986,** *83* (10), 3262.

36.   Borhan, B.; Souto, M. L.; Imai, H.; Shichida, Y.; Nakanishi, K., Movement of retinal along the visual transduction path. *Science* **2000,** *288* (5474), 2209.

37.   Hoischen, D.; Steinmuller, S.; Gartner, W.; Buss, V.; Martin, H. D., Merocyanines as extremely bathochromically absorbing chromophores in the halobacterial membrane protein bacteriorhodopsin. *Angewandte Chemie-International Edition in English* **1997,** *36* (15), 1630.

38.   Lee, K. S. S., *PhD Thesis* **2010**.

39.   Kim, S.-H.; Lee, S.-Y.; Gwon, S.-Y.; Son, Y.-A.; Bae, J.-S., D-pi-A solvatochromic charge transfer dyes containing a 2-cyanomethylene-3-cyano-4,5,5-trimethyl-2,5-dihydrofuran acceptor. *Dyes and Pigments* **2010,** *84* (2), 169.

40.   Dakanali, M.; Do, T. H.; Horn, A.; Chongchivivat, A.; Jarusreni, T.; Lichlyter, D.; Guizzunti, G.; Haidekker, M. A.; Theodorakis, E. A., Self-calibrating viscosity probes: Design and subcellular localization. *Bioorganic & Medicinal Chemistry* **2012,** *20* (14), 4443.

41.   Haidekker, M. A.; Ling, T. T.; Anglo, M.; Stevens, H. Y.; Frangos, J. A.; Theodorakis, E. A., New fluorescent probes for the measurement of cell membrane viscosity. *Chemistry & Biology* **2001,** *8* (2), 123.

42.   Hawe, A.; Filipe, V.; Jiskoot, W., Fluorescent Molecular Rotors as Dyes to Characterize Polysorbate-Containing IgG Formulations. *Pharmaceutical Research* **2010,** *27* (2), 314.

43.   Iwaki, T.; Torigoe, C.; Noji, M.; Nakanishi, M., ANTIBODIES FOR FLUORESCENT MOLECULAR ROTORS. *Biochemistry* **1993,** *32* (29), 7589.

44.   Lee, C. C.; Hu, A. T., Synthesis and optical recording properties of some novel styryl dyes for DVD-R. *Dyes and Pigments* **2003,** *59* (1), 63.

45.   Kuimova, M. K., Mapping viscosity in cells using molecular rotors. *Physical Chemistry Chemical Physics* **2012,** *14* (37), 12671.

46.   Ozhalici-Unal, H.; Pow, C. L.; Marks, S. A.; Jesper, L. D.; Silva, G. L.; Shank, N. I.; Jones, E. W.; Burnette, J. M.; Berget, P. B.; Armitage, B. A., A rainbow of fluoromodules: A promiscuous scFv protein binds to and activates a diverse set of fluorogenic cyanine dyes. *Journal of the American Chemical Society* **2008,** *130* (38), 12620.

47.   Creemers, A. F. L.; Lugtenburg, J., The preparation of all-trans uniformly C-13-labeled retinal via a modular total organic synthetic strategy. Emerging central contribution of organic synthesis toward the structure and function study with atomic resolution in protein research. *Journal of the American Chemical Society* **2002,** *124* (22), 6324.

48.   Cai, G. L.; Bozhkova, N.; Odingo, J.; Berova, N.; Nakanishi, K., CD EXCITON CHIRALITY METHOD - NEW RED-SHIFTED CHROMOPHORES FOR HYDROXYL-GROUPS. *Journal of the American Chemical Society* **1993,** *115* (16), 7192.

49.   Gallagher, W. P.; Maleczka, R. E., Stille reactions catalytic in tin: A "Sn-F" route for intermolecular and intramolecular couplings. *Journal of Organic Chemistry* **2005,** *70* (3), 841.

50.   Maleczka, R. E.; Lavis, J. M.; Clark, D. H.; Gallagher, W. P., Microwave-assisted one-pot hydrostannylation/Stille couplings. *Organic Letters* **2000,** *2* (23), 3655.

51.   Sorg, A.; Blank, F.; Bruckner, R., Stepwise cross-couplings of a dibromo-gamma-methylenebutenolide as an access to Z-configured alpha-alkenyl-gamma-alkylidenebutenolides. Straightforward synthesis of the antibiotic lissoclinolide. *Synlett* **2005,** (8), 1286.

52.   Tanaka, K.; Katsumura, S., Novel synthesis of the ocular age pigment A2-E: New method for substituted pyridine synthesis via azaelectrocyclization. *Organic Letters* **2000,** *2* (3), 373.

53.   Betzer, J. F.; Delaloge, F.; Muller, B.; Pancrazi, A.; Prunet, J., Radical hydrostannylation, Pd(0)-catalyzed hydrostannylation, stannylcupration

of propargyl alcohols and enynols: Regio- and stereoselectivities. *Journal of Organic Chemistry* **1997,** *62* (22), 7768.

54.   Mee, S. P. H.; Lee, V.; Baldwin, J. E., Stille coupling made easier - The synergic effect of copper(I) salts and the fluoride ion. *Angewandte Chemie-International Edition* **2004,** *43* (9), 1132.

55.   Paton, R. S.; Steinhardt, S. E.; Vanderwal, C. D.; Houk, K. N., Unraveling the Mechanism of Cascade Reactions of Zincke Aldehydes. *Journal of the American Chemical Society* **2011,** *133* (11), 3895.

56.   Steinhardt, S. E.; Silverston, J. S.; Vanderwal, C. D., Stereocontrolled synthesis of Z-dienes via an unexpected pericyclic cascade rearrangement of 5-amino-2,4-pentadienals. *Journal of the American Chemical Society* **2008,** *130* (24), 7560.

57.   Becher, J., SYNTHESIS AND REACTIONS OF GLUTACONALDEHYDE AND 5-AMINO-2,4-PENTADIENALS. *Synthesis-Stuttgart* **1980,** (8), 589.

58.   Vanderwal, C. D., Reactivity and Synthesis Inspired by the Zincke Ring-Opening of Pyridines. *Journal of Organic Chemistry* **2011,** *76* (23), 9555.

59.   Michels, T. D.; Rhee, J. U.; Vanderwal, C. D., Synthesis of delta-Tributylstannyl-alpha,beta,gamma,delta-Unsaturated Aldehydes from Pyridines. *Organic Letters* **2008,** *10* (21), 4787.

60.   Huang, X.; Song, L.; Xu, J.; Zhu, G.; Liu, B., Asymmetric Total Synthesis of Leucosceptroid B. *Angewandte Chemie-International Edition* **2013,** *52* (3), 952.

61.   Caggiano, T. J.; Brazzale, A.; Ho, D. M.; Kraml, C. M.; Trybulski, E.; Chadwick, C. C.; Chippari, S.; Borges-Marcucci, L.; Eckert, A.; Keith, J. C.; Kenney, T.; Harnish, D. C., Estrogen receptor dependent inhibitors of NF-kappa B transcriptional activation-1 synthesis and biological evaluation of substituted 2-cyanopropanoic acid derivatives: Pathway selective inhibitors of NF-kappa B, a potential treatment for rheumatoid arthritis. *Journal of Medicinal Chemistry* **2007,** *50* (22), 5245.

62.   Dong, Z. H.; Liu, X. H.; Feng, J. H.; Wang, M.; Lin, L. L.; Feng, X. M., Efficient Asymmetric Synthesis of 4H-Chromene Derivatives through a Tandem Michael Addition-Cyclization Reaction Catalyzed by a Salen-Cobalt(II) Complex. *European Journal of Organic Chemistry* **2011,** (1), 137.

63.   Erichsen, M. N.; Huynh, T. H. V.; Abrahamsen, B.; Bastlund, J. F.; Bundgaard, C.; Monrad, O.; Bekker-Jensen, A.; Nielsen, C. W.; Frydenvang, K.;

Jensen, A. A.; Bunch, L., Structure-Activity Relationship Study of First Selective Inhibitor of Excitatory Amino Acid Transporter Subtype 1: 2-Amino-4-(4-methoxyphenyl)-7-(naphthalen-1-yl)-5-oxo-5,6,7,8-tetrahydro -4H-chromene-3-carbonitrile (UCPH-101). *Journal of Medicinal Chemistry* **2010,** *53* (19), 7180.

64.   Kulshrestha, A.; Marzijarani, N. S.; Ashtekar, K. D.; Staples, R.; Borhan, B., 3,4-Dihydroxypyrrolidines via Modified Tandem Aza-Payne/Hydroamination Pathway. *Organic Letters* **2012,** *14* (14), 3592.

65.   Schomaker, J. M.; Geiser, A. R.; Huang, R.; Borhan, B., Tetrasubstituted pyrrolidines via a tandem aza-Payne/hydroamination reaction. *Journal of the American Chemical Society* **2007,** *129* (13), 3794.

66.   Bernasconi, C. F.; Fox, J. P.; Kanavarioti, A.; Panda, M., NUCLEOPHILIC-ADDITION TO OLEFINS .15. SOLVENT AND SUBSTITUENT EFFECTS ON THE HYDROLYSIS OF BENZYLIDENEMALONONITRILES IN BASIC ME2SO-WATER SOLUTIONS. *Journal of the American Chemical Society* **1986,** *108* (9), 2372.

67.   Bernasconi, C. F.; Kanavarioti, A.; Killion, R. B., NUCLEOPHILIC-ADDITION TO OLEFINS .12. SOLVENT-INDUCED CHANGE IN THE RATE-LIMITING STEP OF THE HYDROLYSIS OF BENZYLIDENEMALONONITRILE IN ACIDIC ME2SO-WATER SOLUTION. *Journal of the American Chemical Society* **1985,** *107* (12), 3612.

68.   Sovdat, T.; Bassolino, G.; Liebel, M.; Schnedermann, C.; Fletcher, S. P.; Kukura, P., Backbone Modification of Retinal Induces Protein-like Excited State Dynamics in Solution. *Journal of the American Chemical Society* **2012,** *134* (20), 8318.

69.   Paterson, I.; Paquet, T., Total Synthesis and Configurational Validation of (+)-Phorbaside A. *Organic Letters* **2010,** *12* (9), 2158.

CHAPTER IV: REENGINEERING OF HUMAN CELLULAR RETINOL BINDING PROTEIN (HCRBPII) INTO AN INSTANTANEOUSLY FLUORESCENT PROTEIN TAG

IV.1 Currently used protein tags

Site-specific protein modifications have great utility in modern chemical biology. Tagging a protein of interest with a short peptide or an additional protein is commonly used as a purification technique and as a detection method. The most common of the latter are fluorescent proteins (FP), which are perhaps the most useful imaging tool in biology. Besides FP variants, site-specific labeling of a fusion tag used is a rapidly growing research area. These tools are commonly used where FP fail to deliver. The labeling methods compatible with biological conditions include bioorthogonal ligations, where non-native functional groups exhibit selective reactivity (Staudinger ligation,[1] alkyne-azide 'click' chemistry,[2] inverse-demand Diels-Alder reactions,[3] oxanobornadiene cycloaddition,[4] Pictet-Spengler reaction[5]).

Another strategy to label specific proteins evolved from the re-design of the native ligand.[6-10] This strategy requires the conjugate of the natural ligand with a fluorophore. Selective binding of this conjugate to the protein is facilitated by its natural affinity. The tags developed according to this strategy include Halo-tag, TMP-tag, SNAP-tag, CLIP-tag, among others.[6] This labeling strategy requires the conjugate to be in the constantly fluorescent state. The exogenously added conjugate, which binds to the protein tag, does not exclusively reach its destination. Thus, fluorescence is observed at different intracellular locations.

Washing the live cells multiple times with buffer, combined with prolonged incubation time without the labeling reagent, most often allows the elimination of all non-specifically trapped chromophore. However, these steps delay the detection time, which limits the application in real-time imaging.

One of the most effective ways to remove non-specific fluorescence is to provide a quencher that is released upon protein labeling.[11] A highly effective approach for the wash-free SNAP-tag modification was realized by the introduction of a QSY-7 quencher in proximity of a cyanine fluorophore (**Figure IV-1**).[12] The overall size of the exogenously added molecule significantly increased, which required additional reagents to enable cell membrane penetration. Gratifyingly, the cell surface expression of SNAP in HEK293 cells was successfully visualized after 30 min *via* wash-free fluorescence imaging.

Among the ways to approach wash-free labeling is to supply the cells with
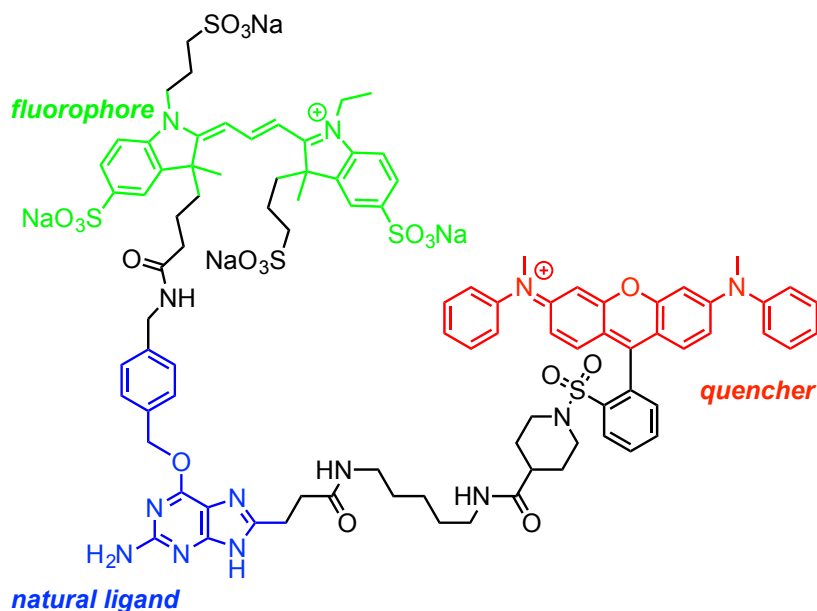
Figure IV-1: The labeling reagent for a wash-free SNAP-tag.

an environment sensitive (solvatochromic) fluorophore.[13-14] In this way, the labeling reagent would have significantly different spectroscopic properties inside of a hydrophobic protein or membrane, as compared to highly polar aqueous solution. For example, the PYP-tag labeling with 7-dimethylaminocoumarine proceeds through the transthioesterification with Cys69 (**Figure IV-2**).[13] This reaction eliminates the quenching moiety, but the cross-reactivity with other thiols is still a potential problem. The environment-sensitive nature of 7-dimethylaminocoumarin solves the cross-reactivity issue. It exhibits enhanced fluorescence inside of the hydrophobic PYP-tag cavity. The authors were able to manipulate the rate of the transthioesterification as well. This tag is the fastest imaging method reported thus far, showing nucleus staining within 6 min of addition of the dye.

Recently, the Johnsson lab described the utility of solvatochromic Nile Red in the SNAP-tag labeling of the cell surface. The authors utilize the enhanced fluorescence of Nile Red in the hydrophobic environment of the plasma membrane, which was directed to the membrane *via* a chemical linker to SNAP-tag. In this approach, the high affinity of the conjugate to SNAP-tag successfully competes with the non-specific staining of the plasma membrane with Nile Red.
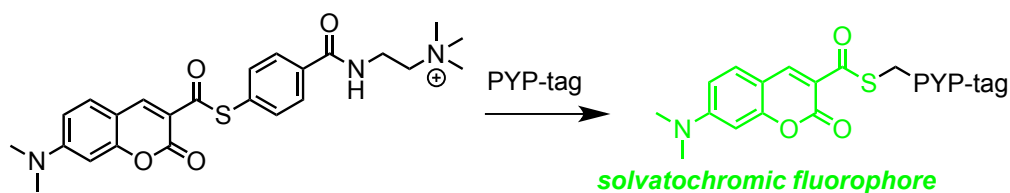
Figure IV-2: The labeling the wash-free PYP-tag.

The utility of molecular rotors with enhanced fluorescence in the planar state is another method to improve signal-to-noise ratio. This method is most commonly utilized for imaging RNA and DNA, making the fluorophores more planar when trapped between grooves. For example, when Thiazol Orange intercalates with DNA, its fluorescence goes up by 550 fold and when Malachite Green binds to a specific RNA aptamer enhancement is 2360 fold (**Figure IV-3**).[15] However, the most notable example is the RNA imaging method with GFP-like chromophores from Jeffrey's lab (**Figure IV-3**).[16-18]

thiazol orange   *malachite green*   *GFP-like*
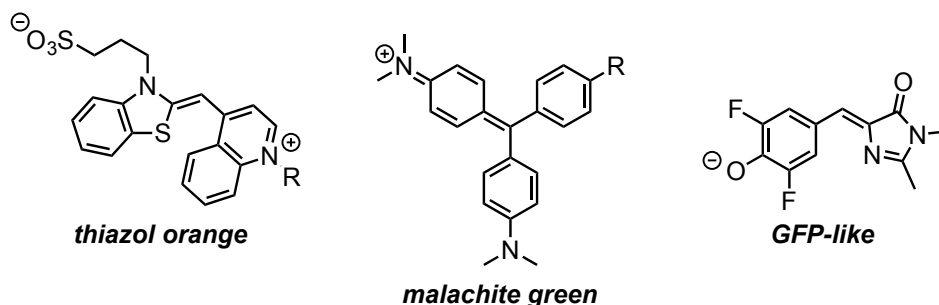
Figure IV-3: Fluorescent rotors.

Not surprisingly, in the area of alternative imaging methods, the most ideal tool would avoid exogenous addition of a labeling reagent and utilize the metabolites already existing in cells. The Cohan lab was able to apply the dim fluorescence of retinal in imaging HEK293 cells. However, this technique is highly demanding to microscope settings.[19] Chlorophyl fluorescence was successfully utilized in the imaging of plants, but the method is most useful for studies of photosynthetic activity.[20] By far the most prominent example of this kind is a bilirubin binding protein extracted from eel muscles.[21-22] The latter protein, UnaG, binds the breakdown product of heme, bilirubin, and 'turns-on' fluorescence in response (**Figure IV-4**). UnaG has sequence and structure similarity with the

fatty acid binding proteins (FABP). For instance, it shares 56% homology with human brain FABP. The crystal structure revealed a compact ligand-protein complex,[22] which is the reason for its highly enhanced fluorescence. UnaG has the potential to become a novel fluorescent sensor of bilirubin since it is a metabolite associated with proper liver function and human waste.

The isolation of UnaG as a FABP-related fluorescent protein aligns well with the direction we took for a fluorescent tag design. The protein folding is similar between CRABPII, hCRBPII and UnaG. In our approach, fluorescence is a result of site-specific protein labeling achieved *via* addition of a small, non-fluorescent molecule to a reengineered CRABPII or hCRBPII protein host. Covalent bond formation converts two non-fluorescent species into a fluorescent tag. The event of iminium formation in the earlier systems is the platform for this design.

The natural function of CRABPII and hCRBPII is to carry an insoluble hydrophobic compound to its destination in the cell. Both systems are small, cytosolic proteins. They are robust to mutagenesis and their targeted expression
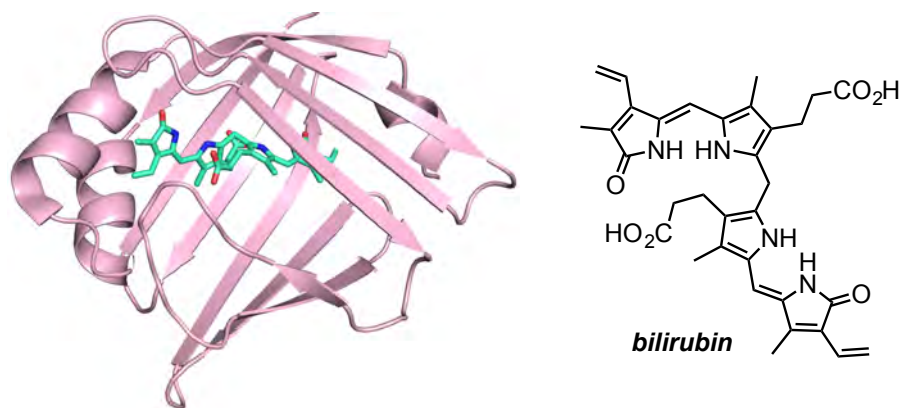
Figure IV-4: Structure of UnaG bound with bilirubin.

225

is not expected to interfere with the cellular machinery. However, CRABPII was reported to have more function than just a carrier protein. For example, it has been shown to participate in ligand transfer to a transcription factor. This might limit the use of CRABPII as a reporter tag of the protein of interest (POI). Therefore, an alternative approach with hCRBPII as a tag is advantageous.

Iminium formation is a reversible reaction, which excludes it from the primary choices of stable conjugates in biology. An alpha-effect is often used in the bioorthogonal ligation to increase the nucleophilicity of the amine. The iminium bond, formed *via* lysine, cannot be trapped irreversibly. However, the high affinity of a lipophilic molecule to the hydrophobic protein cavity can stabilize the complex. Finally, the properly engineered acidic sites inside CRABPII or hCRBPII can speed up the reaction, because iminium formation is often an acid catalyzed process.

IV.2 Role of carboxylate residues in the rate of imine and iminium formation inside CRABPII

During CRABPII reengineering, it was observed that a L121 mutation to carboxylic acid, in proximity of PSB region, could influence the rate of the reaction. As described in Chapter II, the PSB formation with retinal was unsuccessful for the R111K:R132L:Y134F:T54V:L121E mutant even in acidic buffer. However, the peak corresponding to SB formation was observed at 365 nm. The time needed to reach SB maximum for R111K:R132L:Y134F:T54V and R111K:R132L:Y134F:T54V:L121E mutants strongly depends on pH (**Figure IV-**

5). For instance, the tetra-mutant forms SB significantly faster under acidic conditions ($t_{1/2}$ = 18.9 min at pH 5.0 *versus* $t_{1/2}$ = 83.5 min at pH 7.3). On the other hand, reaction with the penta-mutant containing L121E takes place faster at normal pH ($t_{1/2}$ = 14.4 min at pH 5.0 *versus* $t_{1/2}$ = 3.3 min at pH 7.3).

This observation in general can be attributed to acid catalysis during iminium formation. Considering this chemical reaction in general, it is well known that a catalytic proton source accelerates imine formation. Kinetic data for R111K:R132L:Y134F:T54V tetra-mutant in acidic and normal buffer is consistent with the proposed acid-catalyzed reaction. On the contrary, the penta-mutant, R111K:R132L:Y134F:T54V:L121E, contains a carboxylic acid residue, L121E, near the reactive lysine inside of the protein pocket (see crystal structure in **Figure IV-5**). At normal pH, this could be the proton source that activates the carbonyl of retinal for nucleophilic attack with lysine. However, under acidic conditions, the lysine might be protonated as well, which would make it less nucleophilic. Interestingly, the catalysis provided by L121E is more efficient as

| CRABPII mutant | $t_{1/2}$ | |
	pH=5.0	pH=7.3
KLFV*	18.9	83.5 min
KLFV:L121E	14.4 min	3.3 min
KLFV:L121D	8.6 min	8.0 min

*R111K:R132L:Y134F:T54V
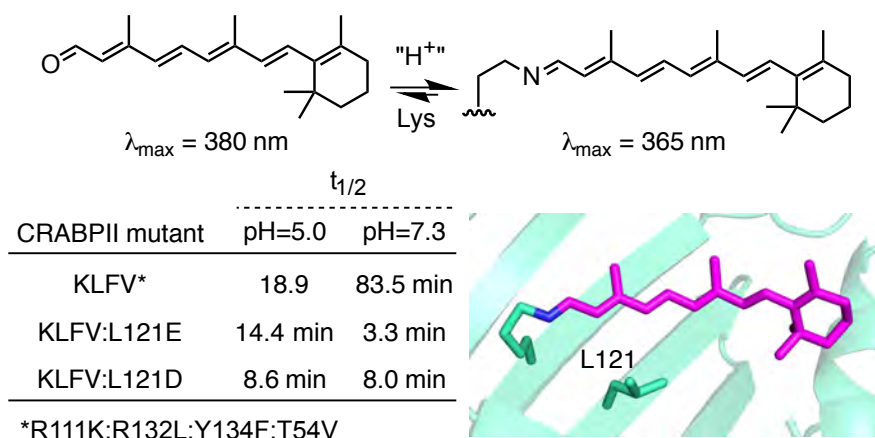
Figure IV-5: Influence of L121E and L121D mutation on the half-life of imine formation between CRABPII mutants and Rt at acidic and normal pH and location of L121 to Rt-PSB.

compared to altering the pH of the solution (**Figure IV-5**). The role of L121E

mutation in the protein can be to: 1) activate the aldehyde in proximity of

nucleophilic lysine or 2) protonate the tetrahedral intermediate, formed after the

first step, essential for water elimination (**Figure IV-6**).

This mechanism is in agreement with a report on accelerated hydrazone

formation, if the carboxylic acid is part of one of the substrates.[23-24] Briefly, the

reaction rate between an aldehyde and a hydrazine increased in the presence of

an acidic substituent on either reagent (**Figure IV-7**). Moreover, the strength of

the acid showed a direct correlation to the rate of the reaction. The rate was the

highest if acid was a part of the hydrazine, because it became not only a proton

donor but also a factor that brought two substrates in close proximity for the

reaction to occur.

In addition to the R111K:R132L:Y134F:T54V:L121E mutant,

R111K:R132L:Y134F:T54V:L121D was also investigated for SB formation in a

time-dependent manner (**Figure IV-5**). The pK_a value for glutamic acid is 4.25

and aspartic acid is 3.86. Inside the hydrophobic protein environment these

values could change significantly, however, the overall trend might be the same.
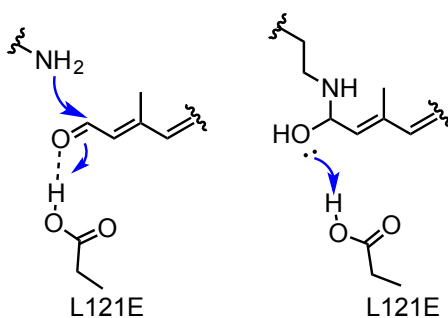
Figure IV-6: Two possible mechanisms of reaction catalysis by L121E.

228

It is reasonable to expect that a slightly more acidic and smaller residue might have a different impact on the rate of the reaction rate. Consistent with previous data R111K:R132L:Y134F:T54V:L121D mutant forms SB faster than R111K:R132L:Y134F:T54V:L121E, but interestingly, pH does not have a significant impact on the reaction (**Figure IV-5**).

The reaction acceleration between CRABPII mutants and retinal was an inspiring observation into the regulation of the rate for PSB formation in general. As described in Chapter III, merocyanine retinal analog (MCRA) was extensively studied in our lab in previous years. However, on average, completion of the reaction takes ~20 min in the CRABPII-R132K:R111L:L121E series (Dr. Kin Sing Stephen Lee and Ipek Yapici, Professor Babak Borhan's lab, MSU) and ~60 min for the hCRBPII-Q108K:K40L:T51V series (Dr. Wenjing Wang, Professor Babak Borhan's lab, MSU) to complete. To understand the factors that control the rate of the reaction would be beneficial for further rate optimization. Moreover, screening the second generation CRABPII mutants for PSB formation with MCRA
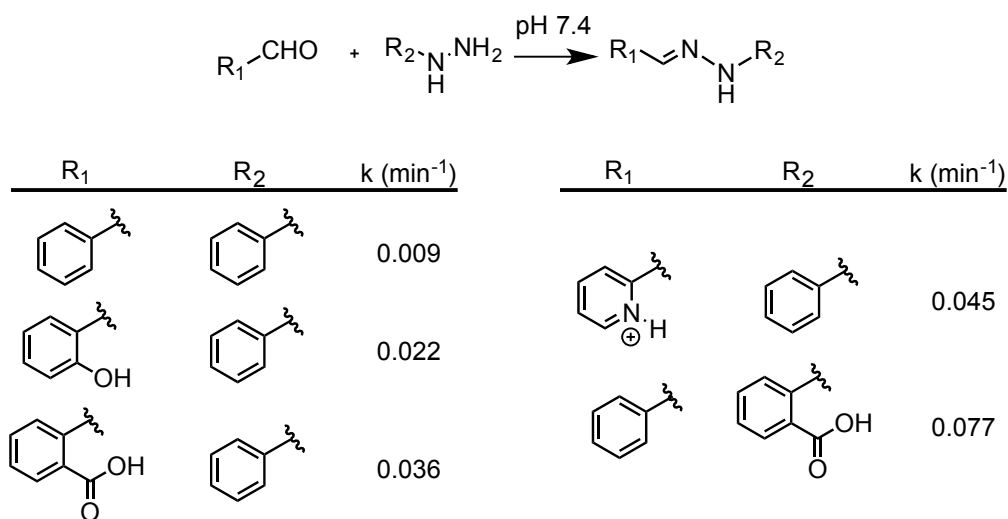
Figure IV-7: Impact of intramolecular acid catalysis on the hydrazone formation.

Table IV-1. Binding of MCRA to selected CRABPII mutants.

Entry	CRABPII mutant	$\lambda_{abs}/\lambda_{em}$ (nm)	QY, %	k (min^{-1})
1	KLFV[a]	594/615	17	0.018
2	KLFV:L121D	592/619	17	0.208
3	KLFV:R59W	596/619	16	0.011
4	KLFV:R59W:L121D	600/629	18	1.785
5	KLFV:R59Y:A32Y:A36Y	595/617	6	0.011
6	KLFV:R59Y:A32Y:A36Y:L121D	595/623	15	0.237
7	KLFV:R59Y:A32W	595/615	8	0.032
8	KLFV:R59Y:A32W:L121D	605/628	31	2.006
9	KLFV:R59W:A32W	598/617	4	0.027
10	KLFV:R59W:A32W:L121D	604/629	20	1.856

[a]KLFV is CRABPII-R111K:R132L:Y134F:T54V.

would be of additional interest because of its exceptional pH-sensitive properties based on the data collected with Rt (Chapter II).

Since the L121D mutation was able to accelerate the reaction independent of pH, CRABPII mutants with L121E and L121D were compared in reactivity with MCRA. The L121D mutation also had an impact on quantum yield (QY), but the wavelengths of absorption and emission were not influenced significantly (**Table IV-1**). In all examples, the L121D mutation accelerated the reaction rate for PSB formation. The smallest impact was in the shortest series, which had a 11-fold increase in rate (**Table IV-1**, entries 1 and 2). The most significant enhancement, a 162-fold increase, was found when L121D mutation was introduced to the R111K:R132L:Y134F:T54V:R59W template (**Table IV-1**, entries 3 and 4). Interestingly, mutants with A32Y and A32W mutations had quantum yields close to the value of MCRA-PSB with butylamine, which mimics non-specific binding
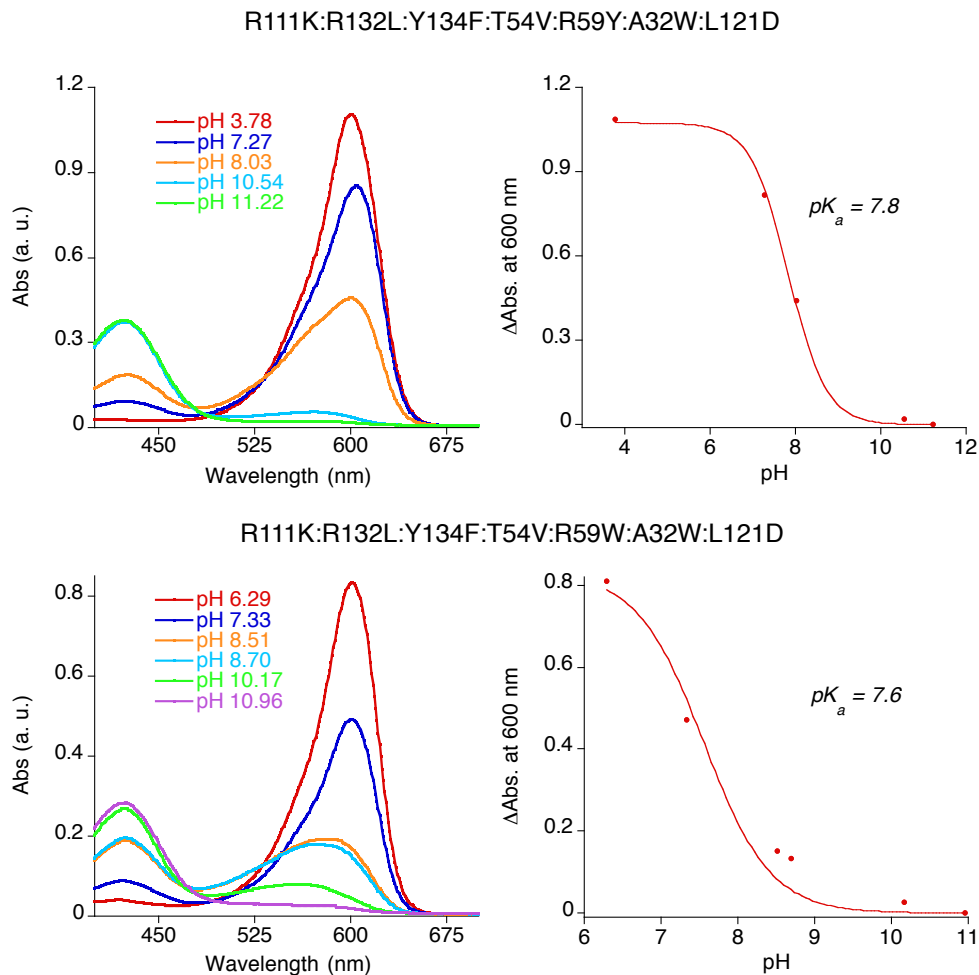
Figure IV-8: UV-vis and pK$_a$ curve fit as a function of pH of two indicated CRABPII mutants bound with MCRA.

(**Table IV-1**, entries 5, 7 and 9). If the L121D mutation was introduced in these series, bright fluorescence was restored (**Table IV-1**, entries 6, 8 and 10).

As mentioned in Chapter II, the L121D mutation counter-intuitively drops the pK$_a$ value of the PSB with retinal. At the same time, the pK$_a$ values for MCRA-PSB in all mutants examined are significantly higher than for the same mutants with retinal. For example, the CRABPII-R111K:R132Q:Y134F:T54V:R59W:A32W:M93L:E73A mutant has a pK$_a$ of 2.6 with retinal (Chapter II, **Table II-6**) but 8.6 with MCRA (Chapter III, **Figure III-4**).

In an attempt to investigate MCRA-CRABPII as a fluorescent reporter of acidity, the pK$_a$ values were measured for two mutants. **Figure IV-8** shows pH titrations of R111K:R132L:Y134F:T54V:R59Y:A32W:L121D and R111K:R132L:Y134F:T54V:R59W:A32W:L121D mutants with pK$_a$ values measured as 7.8 and 7.6, respectively. Those are the lowest numbers for MCRA-PSB to date reported in our lab. The pK$_a$ values are at the high end of the physiologically relevant pH range, which makes them potentially applicable as fluorescent pH-sensors.

IV.3 Live-cell imaging studies of CRABPII mutants labeled with MCRA

Among the tested CRABPII mutants, R111K:R132L:Y134F:T54V:R59Y: A32W:L121D exhibits the highest rate of reaction with MCRA and the highest quantum yield. Additionally, this hepta-mutant can be expressed in good quantities compared to all CRABPII mutants (up to 20 mg/L) and can be stored without precipitation. Considering these observations, this mutant was chosen for expression in eukaryotic cells and live-cell imaging. The gene was amplified and cloned after enhanced green fluorescent protein (EGFP) in the pEGFP-C2 vector, under the cytomegalovirus (CMV) promoter (**Figure IV-9a**). Most known eukaryotic cells are infected with cytomegalovirus and therefore the CMV promoter allows steady expression. Expression of the protein fusion EGFP-CRABPII permits simultaneous dual color imaging of a known fluorescent protein and the CRABPII-MCRA. From the palette of all fluorescent proteins, EGFP was

232

chosen for the fluorescence comparison because its optimal excitation wavelength is over 100 nm away from CRABPII-MCRA complex. This would eliminate the two signals from overlapping ($\lambda_{ex}/\lambda_{em}$: 488/507 and 594/628).

Ultimately, triple fusion, EGFP-CRABPII-RB, was created to demonstrate CRABPII labeling as a fluorescent tag of POI. CRABPII was fused between two functional proteins: 28 kDa enhanced green fluorescent protein (EGFP) at the N-terminus and 110 kDa retinoblastoma protein (RB) at the C-terminus (**Figure IV-9b**). This construct was also created because a similar fusion with hCRBPII, instead of CRABPII, previously showed successful imaging of hCRBPII labeled with MCRA in live cells (Dr. Wenjing Wang, Professor Babak Borhan's lab, MSU). As another control, the same construct was prepared, except with a stop codon after the CRABPII gene sequence (named EGFP-CRABPIIst-RB in **Figure IV-9c**).

RB is a tumor suppressor protein that prevents the cell from replicating damaged DNA.[25] Because of its function to bind the transcription factor E2F, it has a sequence of amino acids close to its C-terminus that localizes RB in the nucleus. Many cancerous cells have dysfunctional mutations in the RB sequence that allows damaged cells to continue proliferation. Osteosarcoma (U2OS) is one
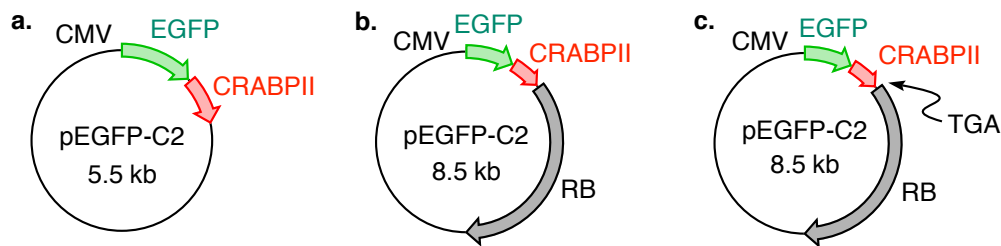
Figure IV-9: DNA plasmids for EGFP-CRABPII fusion expression under CMV promoter.

of the cell lines with RB in its dysfunctional form. Visualization of active RB protein under continuously operating promoter in U2OS or any other cells of this kind is of particular interest because RB inhibits cell replication between G1 and S phase ensuring the single life cycle of the transfected population. In case of EGFP-CRABPII-RB triple fusion, if EGFP fluorescence is localized in the nucleus because of RB, CRABPII is certainly translated.

The NanoJuice transfection of U2OS cells with each described gene gave the steady EGFP signal at the expected location after 36 h. Subsequently, the
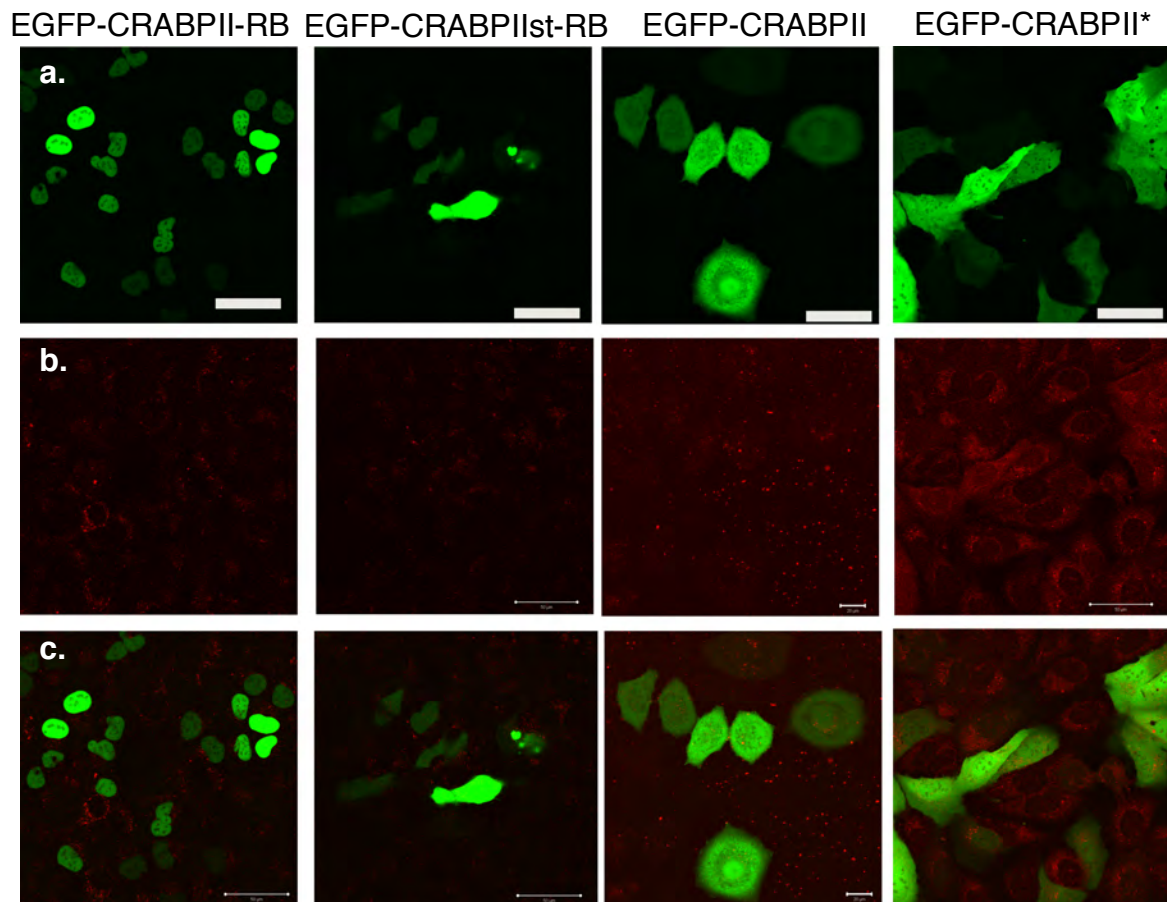
Figure IV-10: Confocal images of labeled of EGFP-CRABPII fusion proteins with MCRA in live U2OS cells irradiated at 488 nm (line **a.**), 594 nm (line **b.**) and both 488 nm and 594 nm (line **c.**). EGFP-CRABPII* was imaged without degradation after wash. All other images are taken after 4 h of degradation after MCRA. Scale bar, 50 μM (line **a.**).

cells were treated with MCRA at an overall concentration of 0.25 μM for 2 h and then incubated for degradation of the excessive chromophore for the next 4 h. Unfortunately, the labeling of either of the fusions expressed in live cells was unsuccessful (**Figure IV-10**), while the same labeling conditions worked before for the same hCRBPII constructs (Dr. Wenjing Wang, Professor Babak Borhan's lab, MSU). Various optimization attempts involved modifications in: concentration (5 μM, 1 μM, 0.5 μM), incubation time (1 h, 30 min, 10 min, 1 min) and the wash/degradation period (4 h, 2 h, 1 h and no wash). None of these conditions gave evidence for the co-localization of EGFP and MCRA labeled CRABPII mutant.

A possible explanation for the unsuccessful CRABPII utility as an imaging tool might be related to its natural function. Reports from the Noy lab provide substantial evidence that CRABPII transfers retinoic acid to the retinoic acid receptor (RAR) through protein-protein interaction.[26-30] Although CRABPII hepta-mutant does not favor retinoic acid binding, the native residues on the protein surface remain capable of interactions with RAR. CRABPII hepta-mutant might be loaded with the intracellular retinal or retinoic acid that promotes association of CRABPII and RAR and makes the ligand exchange impossible. It is also possible that the mutant itself has interactions with RAR, that encloses the binding cavity and prevents binding of any ligand. Finally, the inconsistency between the data with protein purified from *E. coli* and with CRABPII imaging with

235

MCRA in live U2OS cells could be related to protein folding differences in prokaryotic and eukaryotic systems.

IV.4 CRABPII-inspired hCRBPII reengineering into a fast PSB-forming protein

Facing the problem of *in vivo* CRABPII labeling, even after designing a fast binding bright mutant *in vitro,* has potentially three solutions. From one side, if the surface residues of CRABPII that are suggested to participate in interactions with RAR are mutated, the new protein would be free from other functions besides MCRA binding. The second approach involves choosing a cell line deficient in CRABPII and/or RAR. Finally, changing the protein from retinoic acid chaperone (CRABPII) to hCRBPII, which has already been shown to work, has the potential to yield an efficient protein labeling method. We took the third approach, choosing to optimize the already existing hCRBPII mutants in reaction with MCRA.
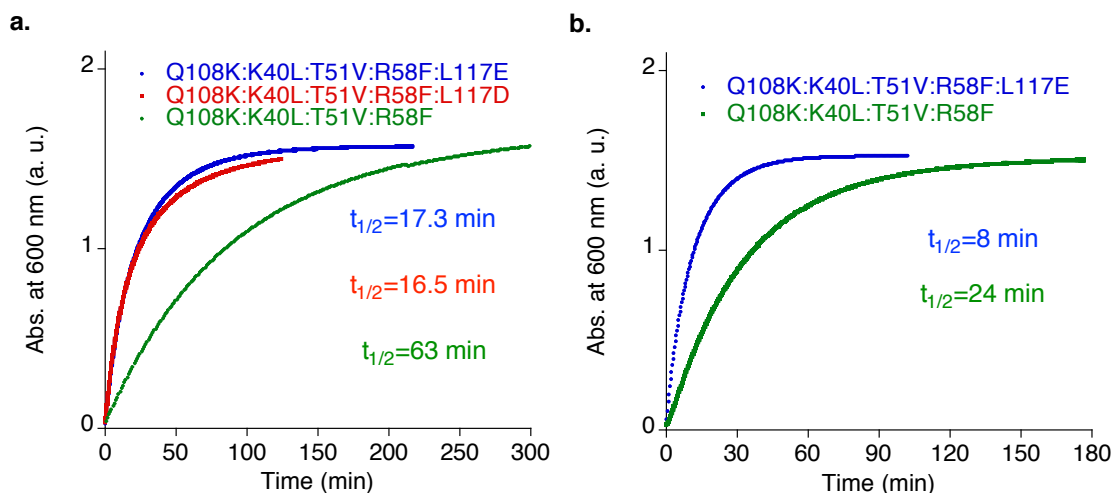
Figure IV-11: Kinetics of MCRA-PSB formation and reaction half-lives with indicated hCRBPII mutants at RT (**a.**) and 37 °C (**b.**)

Optimization started with the hCRBPII-Q108K:K40L:T51V:R58F mutant as a template that was previously optimized for imaging (Dr. Wenjing Wang, Professor Babak Borhan's lab, MSU). In short, the Q108K:K40L:T51V:R58F mutant has high quantum efficiency and brightness (**Table IV-2**, entry 1). *In vitro* PSB formation with MCRA proceeds in 4 h at room temperature (RT) and 2 h at 37 °C. The optimal *in vivo* labeling of this mutant in U2OS cells (37 °C) takes 2 h. This is followed by thorough washing and 6-8 h of additional chromophore-free incubation at 37 °C.

An equivalent residue to L121 in CRABPII is L117 in hCRBPII. Investigation into the impact of two anionic residues, mutations L117D and L117E were investigated with Q108K:K40L:T51V:R58F as the starting template. The rate of the reaction for both penta-mutants was four times faster as compared to the parent tetra-mutant (**Table IV-2**, entries 1-3). The PSB formation was completed within 1 h at RT with a reaction half-life of ~17 min (**Figure IV-11a**), while at physiological temperature (37 °C) half labeled complex was obtained in 8 min (**Figure IV-11b**).

Table IV-2. Binding of MCRA to selected hCRBPII mutants.

Entry	hCRBPII Mutant	$\lambda_{abs}/\lambda_{em}$ (nm)	QY, %	k (min^{-1})
1	Q108K:K40LT51V:R58F	603/623	36	0.011
2	Q108K:K40LT51V:R58F:L117D	596/617	30	0.042
3	Q108K:K40LT51V:R58F:L117E	593/615	36	0.040
4	Q108K:K40L:T51V:T53C:R58W:T29L: A33W:Q4F	600/629	19	0.222
5	Q108K:K40L:T51V:T53C:R58W:T29L: A33W:Q4F:L117D	590/613	30	2.095
6	Q108K:K40L:T51V:T53C:R58W:T29L: A33W:Q4F:L117E	590/611	37	2.744

The penta-mutant with L117E has a slightly higher quantum yield than with L117D (**Table IV-2**, entries 2 and 3). Therefore, the L117E mutation was used to test the influence of rate on hCRBPII labeling in live cells. *In vitro* kinetic data for PSB formation suggests that incubation at 37 °C can be decreased from 2 h to 20-30 min (**Figure IV-11b**) to achieve maximum specific labeling and less non-specific labeling in live cells. The L117E mutation was introduced into the plasmid containing EGFP and hCRBPII-Q108K:K40L:T51V:R58F (CRBP$_{KLVF}$) fusion proteins. Both constructs (CRBP$_{KLVF}$ and CRBP$_{KLVF:L117E}$) were transiently transfected into U2OS cells for comparison of labeling efficiency with shorter incubation time. The cells expressing the fusion proteins were treated with 0.25 μM MCRA for 20 min at 37 °C, washed with warm PBS buffer and incubated for
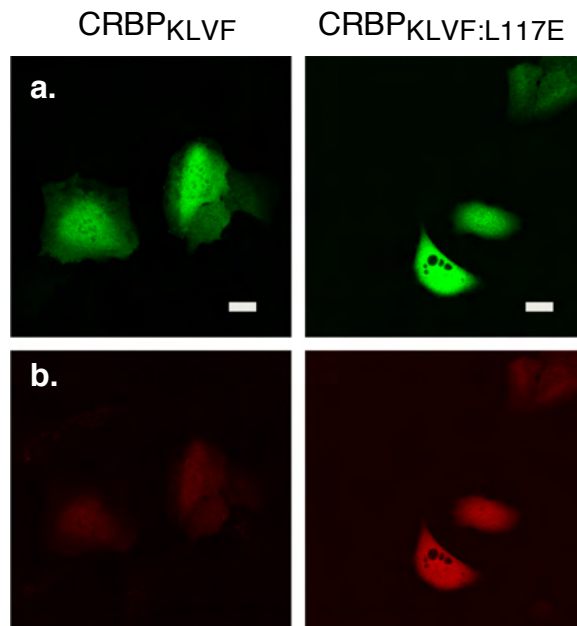
Figure IV-12: Confocal images of indicated EGFP-CRBPII constructs in U2OS cells incubated for 20 min with 0.25 μM MCRA and washed for 2 h. Line (**a.**) excited at 488 nm; line (**b.**) excited at 594 nm. Scale bar, 20 μM.

an additional 2 h. This additional incubation is necessary to degrade any chromophore trapped non-specifically in the intracellular compartments.

As expected, the shorter incubation time was not enough to achieve efficient labeling of the slower binding CRBP$_{KLVF}$. Fortuitously, reaction with the faster binding protein gave much brighter fluorescence within the same incubation time (**Figure IV-12**). However, 20 min incubation with MCRA was also enough time for the chromophore to bind to intracellular amines as a PSB and fluoresce in undesired locations. Since the reaction proceeded for a shorter time period, the degradation step could be also decreased from 6-8 h to 2 h because less MCRA is trapped non-specifically.

Shorter incubation and degradation times are encouraging for the fast PSB forming proteins as functional fluorescent tags. However, the incubation time is still relatively long, yielding significant background fluorescence. To estimate the rate of PSB formation with non-specific protein, we collected the kinetic data with
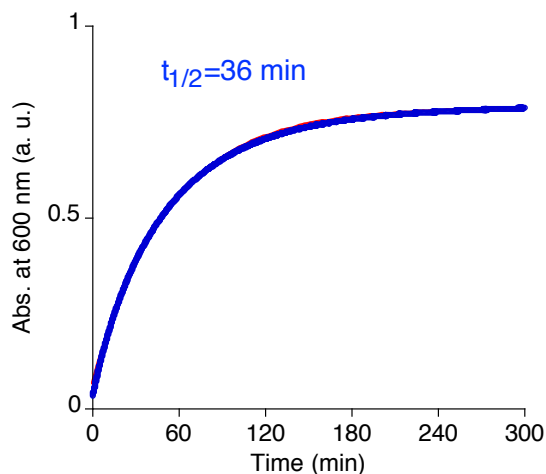
Figure IV-13: Kinetic data of PSB formation between MCRA and BSA.

native bovine serum albumin (BSA). Surprisingly, the PSB formation between MCRA and BSA was faster than between MCRA and CRBP$_{KLVF}$ initially used in imaging. In fact, the reaction half-life with BSA at RT was 36 min (**Figure IV-13**), and even shorter at 37 °C. This explains the significant background fluorescence observed after only 20 min incubation of live cells with MCRA.

Considering the BSA kinetic date, the specific reaction of MCRA with hCRBPII mutants should be complete within minutes to overcome non-specific reactivity. The data obtained with CRABPII-R111K:R132L:Y134F:T54V:R59Y:A32W:L121D meets this requirement, having a reaction half-life of 20 sec (**Table IV-1**, entry 8). Therefore, our strategy was to alter the hCRBPII binding pocket to more closely resemble that of the CRABPII
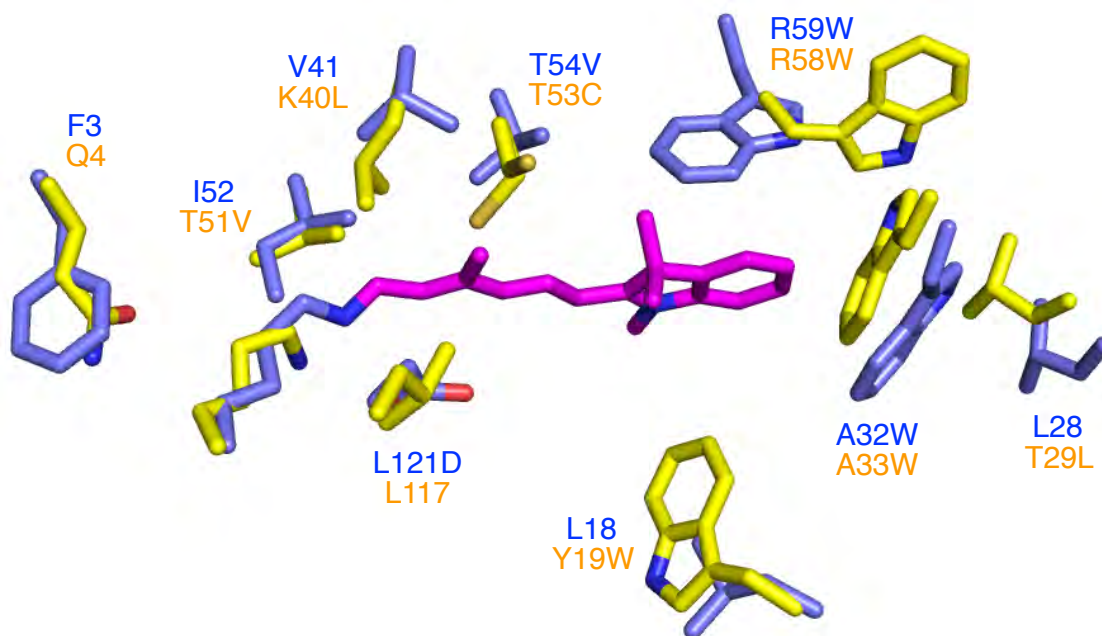
Figure IV-14: Overlayed crystal structures of CRABPII-R111K:R132L:Y134F: T54V:R59W:A32W:L121D (blue) bound with MCRA (magenta) and hCRBPII-Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:A32W (yellow) bound with retinal (not shown).

240

fast binders. We were fortunate to obtain a crystal structure of CRABPII-R111K:R132L:Y134F:T54V:R59W:A32W:L121D, bound with MCRA (**Figure IV-14**). This mutant has a reaction half-life of 22 sec (**Table IV-1**, entry 10). Comparison of its crystal structure with hCRBPII-Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:A33W mutant (PDB ID: 4EDE) showed major differences at residues Q4, Y19W and L117 (hCRBPII numbering). From these observations, hCRBPII-Q108K:K40L:T51V:T53C:R58W:T29L:Q4F:A33W mutant (Dr. Wenjing Wang, Professor Babak Borhan's lab, MSU) was chosen as a template to introduce L117D and L117E mutations (**Table IV-2**, entries 4-6). The template itself showed much faster PSB formation, as compared to the previously used CRBP$_{KLVF}$. However, the quantum yield was low (**Table IV-2**, entries 1 and 4). Both nona-mutants had an even more accelerated reaction as compared to parent octa-mutant. Interestingly, in the longer series, L117E showed faster rate and higher quantum efficiency than L117D (**Table IV-2**, entries 5 and 6). A reaction half-life of 15 sec at RT was maintained for hCRBPII-Q108K:K40L:T51V:T53C:R58W:T29L:Q4F:A33W:L117E nona-mutant (CRBP$_{nona}$).

Table IV-3. Influence of T53V and R58Y mutations on hCRBPII binding to MCRA.

Entry	hCRBPII Mutant	$\lambda_{abs}/\lambda_{em}$ (nm)	QY, %	k (min^{-1})
1	Q108K:K40L:T51V:T53V:R58W:T29L: A33W:Q4F	605/625	17	0.071
2	Q108K:K40L:T51V:T53V:R58W:T29L: A33W:Q4F:L117E	593/613	30	0.26
3	Q108K:K40L:T51V:T53C:R58Y:T29L: A33W:Q4F	610/630	14	0.129
4	Q108K:K40L:T51V:T53C:R58Y:T29L: A33W:Q4F:L117E	595/619	20	0.46

Since the fastest PSB-forming CRABPII mutant contained R59Y and T54V residues (CRABPII numbering), equivalent mutations were also investigated in the hCRBPII system (**Table IV-3**). Counterintuitively, both T53V and R58Y mutations decreased the reaction rate. Moreover, the quantum efficiencies were significantly worse than for $CRBP_{nona}$. On the positive side, the most red-shifted mutant bound with MCRA was discovered. Q108K:K40L:T51V:T53C:R58Y: T29L:A33W:Q4F forms MCRA-PSB with its absorption λ_{max} at 610 nm and emission at 630 nm (**Table IV-3**, entry 3).

IV.5 Building DNA plasmids of hCRBPII fused to localization peptides and proteins for imaging in intracellular compartments

Showing that a hCRBPII mutant can be successfully fused between two other proteins (EGFP and RB) without perturbing their individual functions was an important milestone in the imaging experiments. The follow-up experiment involved the use of hCRBPII for imaging, without EGFP fluorescence, as a control. To address this point, Dr. Wenjing Wang cloned hCRBPII-Q108K:K40L:T51V:R58F in place of EGFP for imaging. Unfortunately, no red fluorescence was observed in the cytosol after MCRA addition. This could be attributed either to a protein expression problem or to ligand binding.

Aiming to demonstrate the use of hCRBPII without EGFP, we decided to switch the vector from EGFP-C2 to another that is unrelated to the FP family, but applicable for eukaryotic cell expression. The hypothesis was that hCRBPII might

need a well-folding protein or peptide in the N-terminus in order to have steady expression. We used the pFLAG-CMV2 vector for cloning hCRBPII mutants after the FLAG epitope tag (DYKDDDDK). In addition, FLAG-tag can also be used for staining with fluorescent antibodies during Western blot or Immunostaining experiments to confirm protein expression, even if labeling fails.

The plasmids previously validated for expression in the most widely used HeLa cells were used. The amyloid-1β was replaced with $CRBP_{nona}$ using the
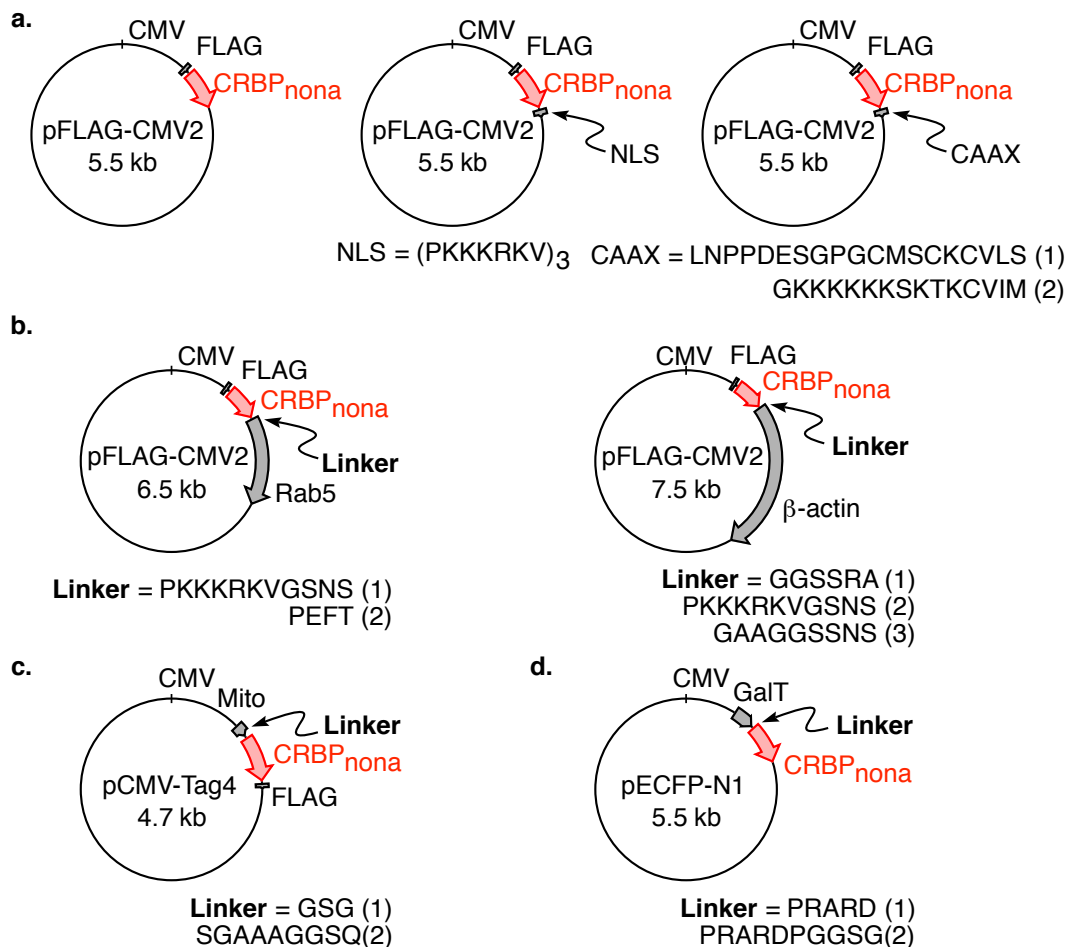
Figure IV-15: Plasmid maps designed and cloned for expression and localization of hCRBPII mutant in eukaryotic cells; **a.** Localization with short peptides in pFLAG-CMV2 vector, **b.** Localization with proteins in pFLAG-CMV2 vector, **c.** Localization with mitochondrial peptide in pCMV-Tag4 vector and **d.** Localization with Golgi-associated protein in pECFP-N1 vector.

plasmid pFLAG-CMV2-AML1B[31] (Addgene#: 12504). In parallel, CRBP$_{nona}$ with

nuclear localization peptide (NLS) was generated from three sequential repeats

of PKKKRKV peptide (**Figure IV-15a**).[32] This final 21 amino acid peptide placed

at the C-terminus of any protein is known to transport the expressed fusion into

the nucleus by importin α. The final protein localization is estimated to be 95% in

the nucleous of HeLa cells.

Another short localization peptide, analogous to the NLS signal, was

placed in the same expression vector. The plasma membrane localization

peptide, known as the CAAX motif, requires the crucial cysteine. This signal is

derived from the Ras protein and the membrane localization is achieved through

farnesylation of cysteine that ends up being associated with the membrane, due

to its lipophilicity.[33] Various sequences of the CAAX motif were utilized in different

reports, but the common sequence requires cysteine (C) followed by two aliphatic

amino acids (AA) and one variable (X). Two CRBP$_{nona}$-CAAX fusions were

prepared for comparison of membrane localization efficiency:

LNPPDESGPGCMSCKCVLS[34] and GKKKKKKSKTKCVIM (**Figure IV-15a**).

Besides expression with short peptides, the usefulness of the hCRBPII-tag

method was confirmed by tracking the protein of interest (POI). For this purpose,

we fused CRBP$_{nona}$ with Rab5 (25 kDa), a GTPase compartmentalized in the

membrane of early endosomes.[35-36] [37] Its functions include regulation of the

membrane transport, motility and intracellular distribution of early endosomes. To

construct the vector containing CRBP$_{nona}$-Rab5 fusion, Rab5 was amplified from

mRFP-Rab5[35] plasmid (Addgene#: 14437) and cloned into the C-terminus of

CRBP$_{nona}$ in the same pFLAG-CMV2 vector. Two types of fusion were prepared,

with longer linkers between hCRBPII and Rab5: PKKKRKVGSNS, and shorter,

PEFT (**Figure IV-15b**), respectively.

As another POI, β-actin (42 kDa) was cloned in the same fusion order and

in the same vector. The gene encoding the β-actin sequence was amplified from

the Lap2-β-actin fusion (Addgene#: 34839) previously described by Ting's lab in

imaging engineered lipoic acid ligase as a fluorophore ligase.[34] The linkers

between CRBP$_{nona}$ and β-actin were GGSSRA (same as in Lap2-β-actin),

PKKKRKVGSNS (same as in CRBP$_{nona}$-Rab5) and GAAGGSSNS (**Figure IV-15b**).

The broad application of new imaging tools depends on how general the

proposed system behaves. One of the important points to consider is the

evidence of the hCRBPII-tag working at either end of the POI. Because the

pFLAG-CMV2 vector has the FLAG-tag before the multicloning cutting site,

alternative vectors should be used. We chose pCMV-Tag4, previously described

as pCMV-Tag4-hL3MBTL2 (Addgene#: 28232) (study on malignant brain tumor

(MBT) protein that binds methylated histone residues).[38] This vector has the

FLAG-tag after the multicloning cutting site useful for expression control, similar

to pFLAG-CMV2. Gene sequences for the short localization peptides can be

introduced at the N-terminus of CRBP$_{nona}$, when generating an insert during the PCR amplification. The mitochondrial localization peptide (mito), the same as in the commercially available EGFP vectors (MSVLTPLLLRGLTGSARRLPVPRAKIHSL)[39] was introduced with two rounds of PCR. The clones constructed had short and flexible linkers between the peptide CRBP$_{nona}$: GSG and SGAAAGGSQ (**Figure IV-15c**).

Finally, the gene fusion constructs with amino acids 1-60 of human β-1,4-galactosyltransferase (GalT) were prepared from GalT-ECFP plasmid (Addgene#: 11937).[40] GalT transports the overall expressed chimera into the Golgi of live cells. This protein is always placed at the N-terminus of the fluorescent reporter to preserve its function. For the imaging experiments, the ECFP gene was replaced with the CRBP$_{nona}$ sequence with PRARDPGGSG and PRARDP linkers between GalT and hCRBPII for comparison (**Figure IV-15d**).
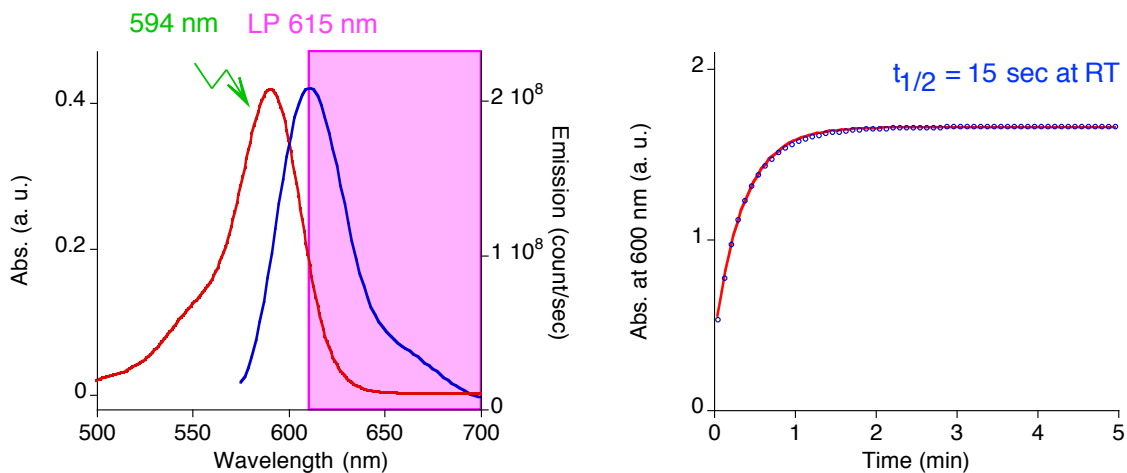
Figure IV-16: Spectroscopic profile and kinetic data for CRBP$_{nona}$ with MCRA.

IV.6 Live-cell imaging of labeled hCRBPII with laser scanning confocal microscopy

Having in hand the CRBP$_{nona}$ protein with a 15 sec half-life for *in vitro* reaction with MCRA, fast labeling was tested in the live HeLa (cervical cancer) cells. An excitation wavelength of 594 nm was used with emission long pass filter at 615 nm in order to match the spectroscopic profile of MCRA-CRBP$_{nona}$ (**Figure IV-16**). Initially, we had to decide on the incubation and wash time for the nona mutant expressed in the cytosol. The reaction is complete within 1 min at 23 °C (**Figure IV-16**). Thus, addition of MCRA to the cells for 1 min should be long enough. The previously used concentration (0.25 μM) was used, followed by wash and imaging after 0 min, 30 min and 60 min. There was no significant difference in background fluorescence among images acquired at the different time points after the wash step (**Figure IV-17**).

For an equivalent comparison of CRBP$_{nona}$ with CRBP$_{KLVF}$ the latter mutant was cloned into the same vector, pFLAG-CMV2, with cytosolic distribution and nuclear localization. In contrast to the fast binding mutant, the slowly reacting CRBP$_{KLVF}$ expressed in the cytosol, does not light up with the same efficiency after 1 min incubation with MCRA. The background fluorescence intensity directly after washing is almost the same as that of specifically labeled cells (**Figure IV-18a**). Additional incubation time leads to degradation of the non-specifically trapped MCRA within 30 min (**Figure IV-18b**), which is not significantly different

from the degradation observed for 60 min (**Figure IV-18c**). However, fluorescence intensity for all of these images is much weaker, as compared to the same set of pictures acquired with the fast binding nona-mutant (**Figure IV-17**).

Next, we compared the labeling efficiency of these two hCRBPII mutants localized in the nucleus using the NLS signal peptide. The fast binding $CRBP_{nona}$ mutant localized in the nucleus has fluorescence as strong as the one localized in cytosol after 1 min incubation with chromophore (**Figure IV-19a**). Although, in this case, MCRA needs to penetrate not only the cellular membrane, but also enter the nucleus before forming the PSB with hCRBPII. In contrast, the slow reaction between MCRA and $CRBP_{KLVF}$ directly reflects in its poor imaging data.
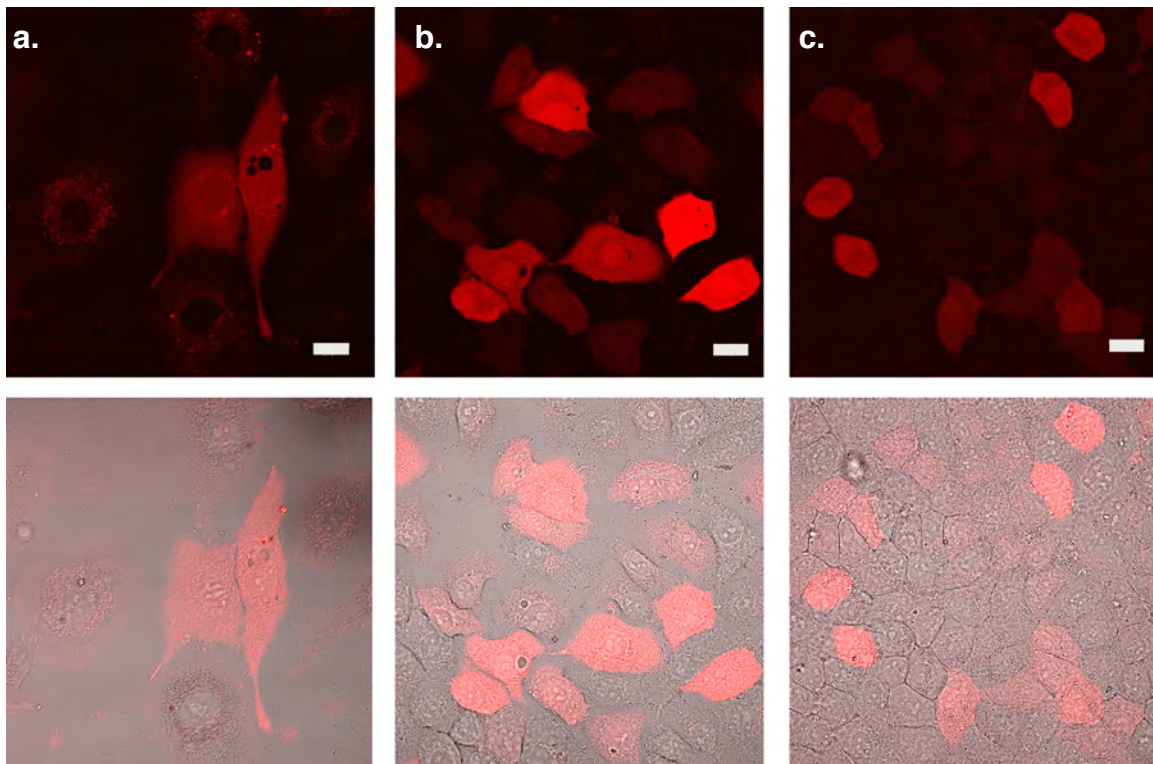
Figure IV-17: Comparison of $CRBP_{nona}$ staining with MCRA in cytosol of HeLa cells after 1 min incubation and wash followed by degradation for **a.** 0 min, **b.** 30 min, **c.** 60 min. Overlay with DIC image is provided below the confocal image. Scale bar, 20 μm.

Hardly any fluorescence was noticed in the nucleus after the cells were supplied with MCRA for 1 min (**Figure IV-19b**). In fact, the HeLa cells expressing CRBP$_{KLVF}$ and HeLa cells not transfected had analogous images under the described conditions that were ideal for fast reacting hCRBPII.

To determine the minimum necessary time from chromophore addition to imaging, we added MCRA to the cells transfected with CRBP$_{nona}$-NLS plasmid directly on the confocal imaging platform and washed the excess (**Figure IV-20**). Since the focus should be at the same spot before and after labeling, it is hard to predict if the selected area contains expressed CRBP$_{nona}$. Therefore, the images were first taken at random areas of the transfected cells, and luckily, in five
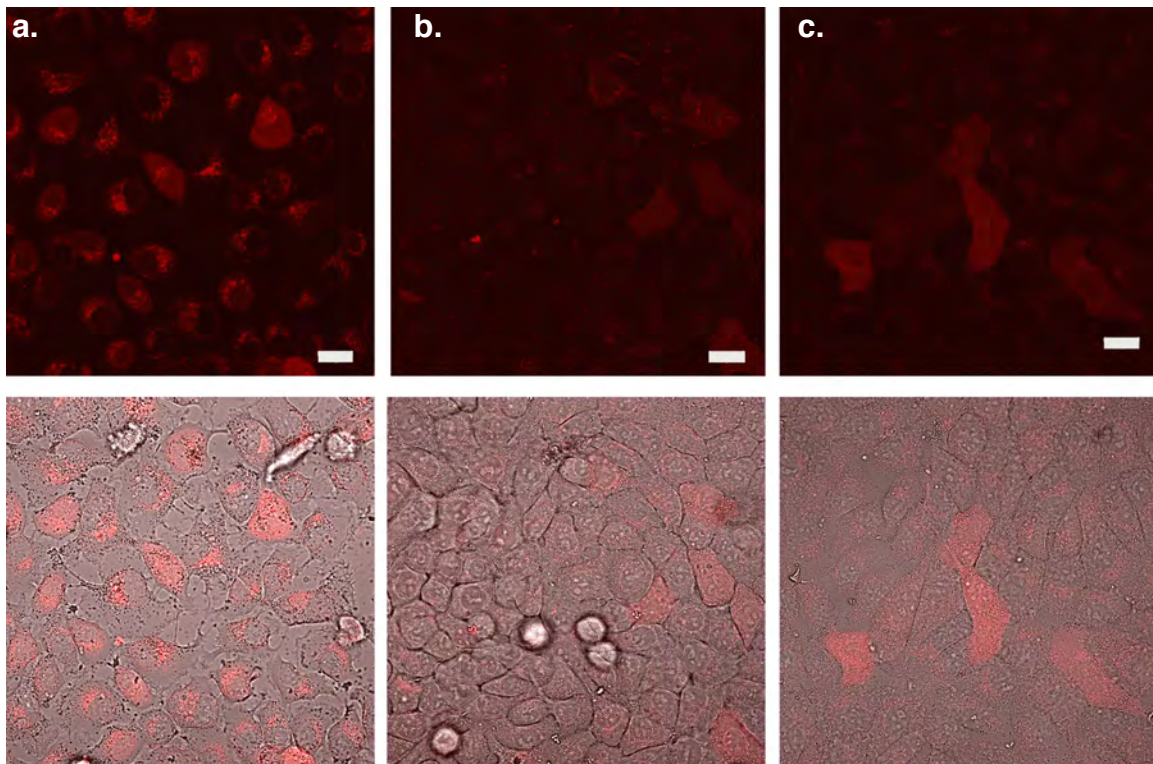
Figure IV-18: Staining CRBP$_{KLVF}$ with MCRA in cytosol after 1 min incubation and wash followed by degradation for 0 min (**a.**), 30 min (**b.**) and 60 min (**c.**). Overlay with DIC image is provided below confocal image. Scale bar, 20 μm.

attempts, the proper images were collected. As shown in **Figure IV-20** before MCRA addition, the cells did not fluoresce under 594 nm laser light (long pass filter 615 nm). Subsequently, MCRA (250 nM) in PBS buffer, warmed to 37 °C, was added to the cells and washed after 10 sec with chromophore-free PBS buffer. Immediately following this, three pictures were acquired, one minute apart for comparison. Confocal images acquired 1 min (right after washing), 2 min and 3 min after wash did not show a significant difference in labeling efficiency or signal-to-noise ratio (**Figure IV-20**). This experiment proved that MCRA readily penetrates the cell and nuclear membrane and binds to hCRBPII mutant within

CRBP$_{nona}$-NLS	CRBP$_{KLVF}$-NLS	no transfection
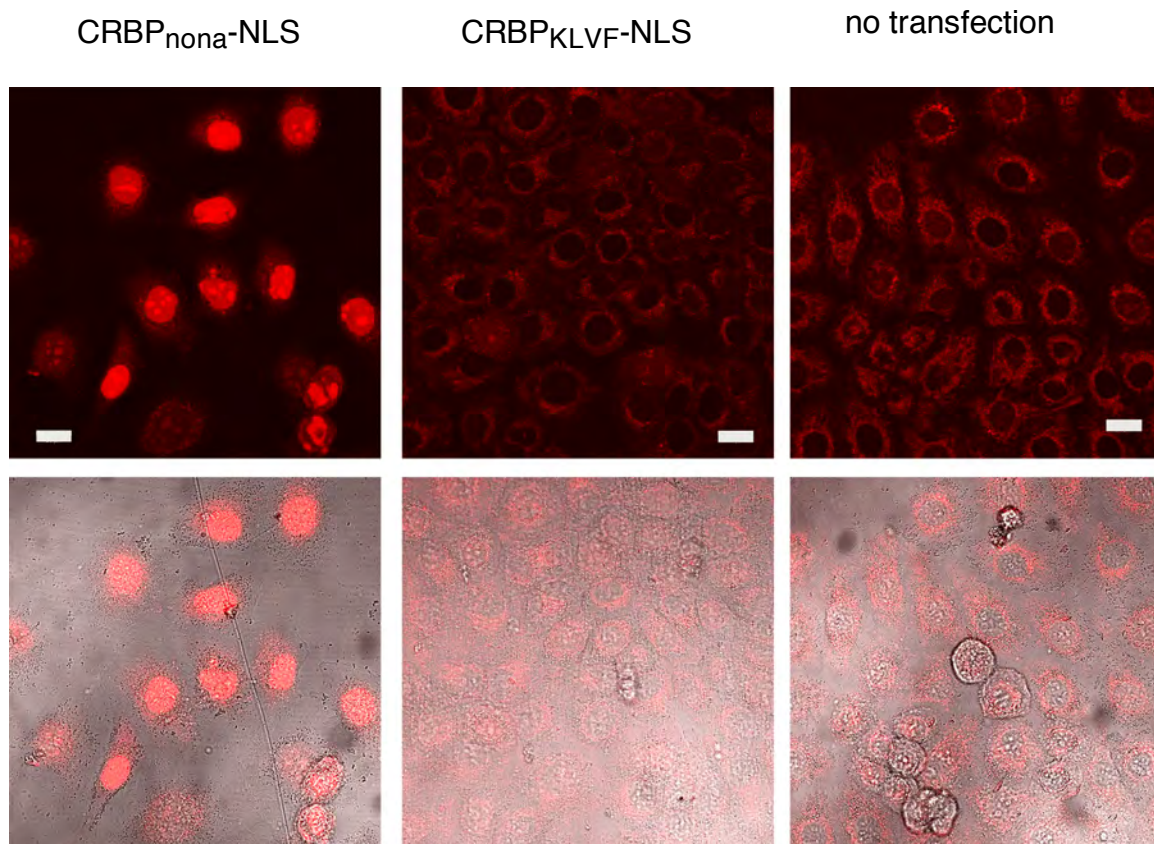
Figure IV-19: Confocal images of specified HeLa cells after labeling with MCRA for 1 min and washed with PBS buffer. Excitation, 594 nm. Overlay with DIC image is provided below the confocal image. Scale bar, 20 μm.

seconds. It is clear evidence that the complex of MCRA and CRBP$_{nona}$ can be

visualized immediately following chromophore addition in abundant cellular

compartments, represented here by the nucleus of HeLa cells.

Next, CRBP$_{nona}$ localization in the plasma membrane of HeLa cells was

tested in confocal imaging. Two plasmids were constructed: CRBP$_{nona}$-CAAX-(1)

and CRBP$_{nona}$-CAAX-(2) (**Figure IV-15**). The only difference is the amino acid

sequence of the farnesylation peptide, LNPPDESGPGCMSCKCVLS and

GKKKKKKSKTKCVIM, respectively. The localization pattern was also correlated

to EGFP-CAAX (CAAX: GKKKKKKSKTKCVIM) fluorescence in HeLa cells that

does not require exogenously added chromophore. In contrast to CRBP$_{nona}$-NLS,

both CRBP$_{nona}$-CAAX fusion proteins required an additional 2 h degradation of
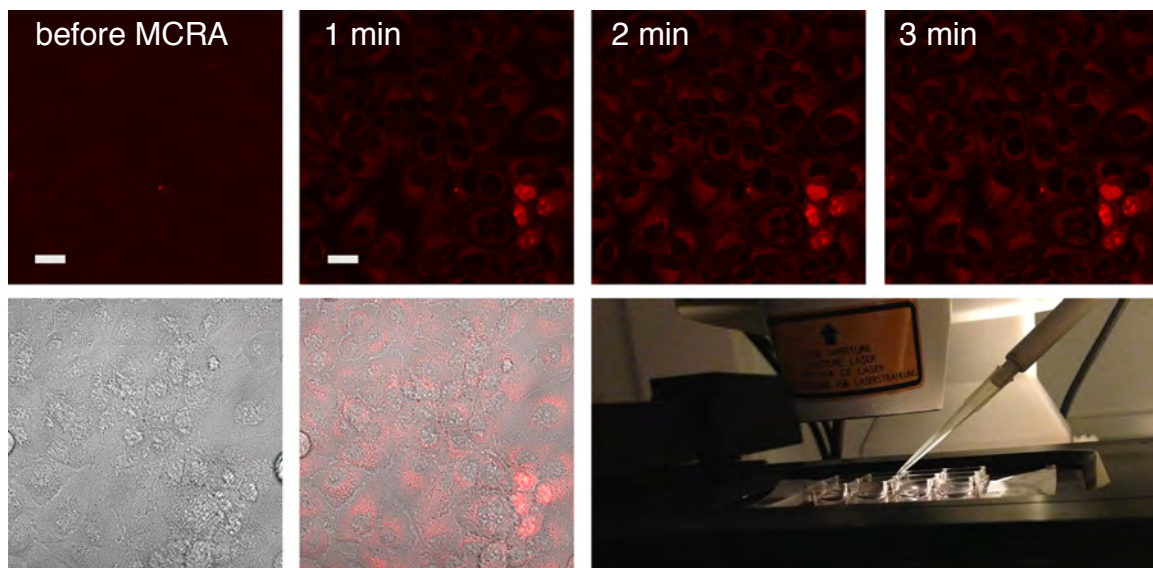
Figure IV-20: Confocal images of CRBP$_{nona}$-NLS expressed in HeLa cells directly on inverted objective before and after addition of MCRA for 10 sec as shown on bottom photo. Overlay with DIC image is provided below confocal image. Scale bar, 20 μm.

the excess chromophore in order to distinguish between specific and non-specific fluorescence. The construct CRBP$_{nona}$-CAAX-(2) had more membrane-like distribution and it correlated better to the EGFP-CAAX pattern in HeLa cells than the CRBP$_{nona}$-CAAX-(1) (**Figure IV-21**). Experiments completed 48 h and 24 h after transfection showed that the fusion is localized in the membrane better after 24 h.

Ultimately, we showed the usefulness of hCRBPII-tag technology in tracking a protein of interest. As described in the previous section, CRBP$_{nona}$ was fused with Rab5 and β-actin. The localization of red fluorescence from CRBP$_{nona}$-Rab5 was identical to mRFP-Rab5 (**Figure IV-22**). Additionally, the pattern was similar to the previously published data for mRFP-Rab5 in HeLa cells. Since Rab5 regulates the attachment of the early endosomes to the microtubules, the images had some tubulin-like fluorescent pattern that was also consistent with images of mRFP-Rab5, as a positive control. The latter fusion does not rely on external chromophore addition to fluoresce under 594 nm excitation laser, and hence, it is a positive control for solely protein distribution in the cell.

On the down side, multiple attempts to image the CRBP$_{nona}$-β-actin protein fusion, after MCRA addition was unsuccessful, regardless of the linker size or the chromophore incubation time. Also Mito-CRBP$_{nona}$ and GalT-CRBP$_{nona}$ with fusion at the N-terminus of hCRBPII appeared exactly like

CRBP$_{nona}$-CAAX-(1)   CRBP$_{nona}$-CAAX-(2)   EGFP-CAAX-(2)
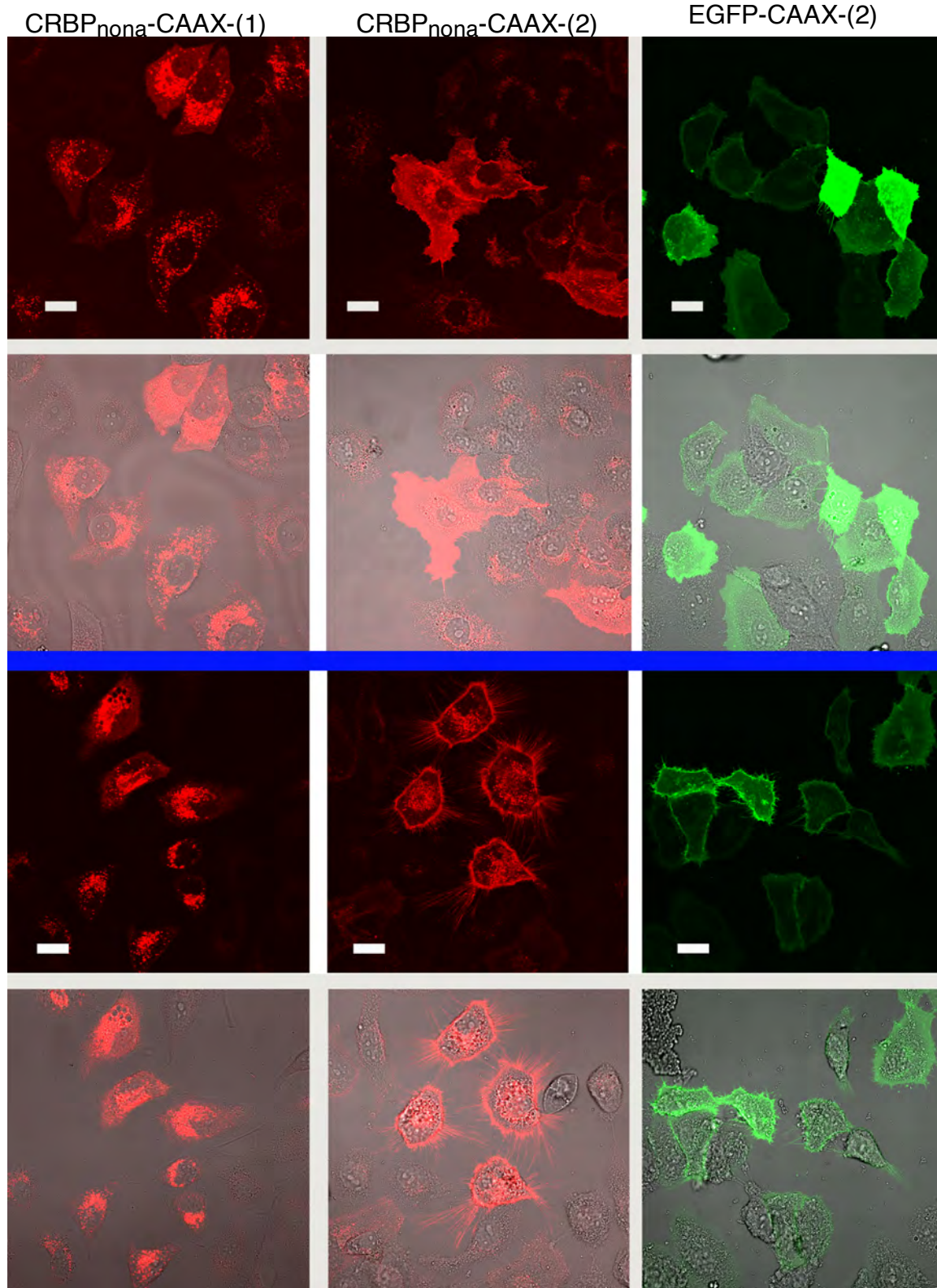
Figure IV-21: Confocal images of CRBP$_{nona}$ and EGFP localized in the plasma membrane of HeLa cells with two types of farnesylation peptides. Blue line separates the data from two experiments (48 h and 24 h). Overlay with DIC image is provided below the confocal image. Scale bar, 20 μm.

untransfected HeLa cells after incubation with MCRA (**Figure IV-24**). This data

can suggest two possibilities: 1) the protein fusion is not expressed properly, or 2)

the localized protein fusion has a difficult time binding MCRA in the targeted

compartment. To understand the origin of the labeling problem we designed

another control experiment, where the FLAG-tag of this fusion was stained with

the fluorescent antibody that binds the FLAG sequence, anti-FLAG-

AlexaFluor488. None of the CRBP$_{nona}$-β-actin (FLAG in N-terminus) and Mito-

CRBP$_{nona}$ (FLAG in C-terminus) constructs showed positive staining with the

antibody. In contrast, CRBP$_{nona}$-NLS (FLAG in N-terminus) fusion that always

works in labeling reaction showed no issues in giving fluorescence from the

mRFP-Rab5   CRBP$_{nona}$-Rab5
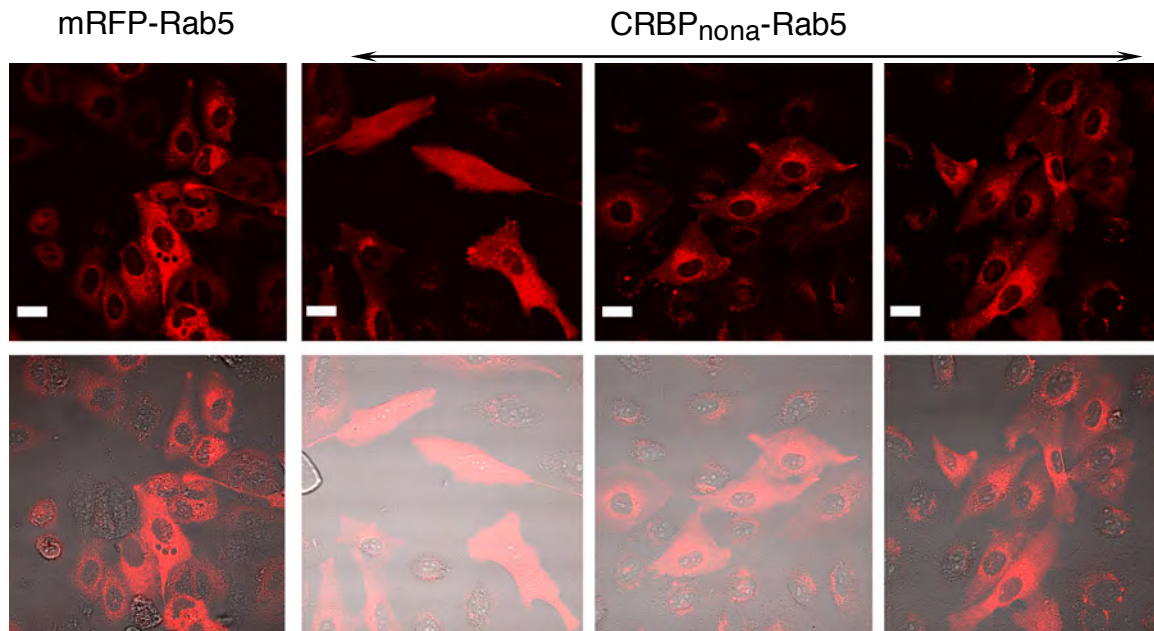
Figure IV-22: Confocal images of mRFP and CRBP$_{nona}$ localized in the endosomes of HeLa cells using Rab5. Overlay with DIC image is provided below the confocal image. Scale bar, 20 μm.

nucleus, after staining with the same antibody, under identical conditions (**Figure IV-23**). This data suggested that the problem is with the expression of CRBP$_{nona}$-β-actin and Mito-CRBP$_{nona}$ in HeLa cells, rather than the labeling reaction or chromophore delivery.

GalT constructs had no FLAG-tag sequence to perform this control experiment. Therefore, localization and labeling of CRBP$_{nona}$ in the Golgi remained inconclusive. The difficulty in distinguishing specific Golgi labeling and non-specific MCRA distribution is that the expected fluorescence of the Golgi looks similar to the non-specific binding of the chromophore (one set of the representative images is in **Figure IV-24**). This experiment proves that
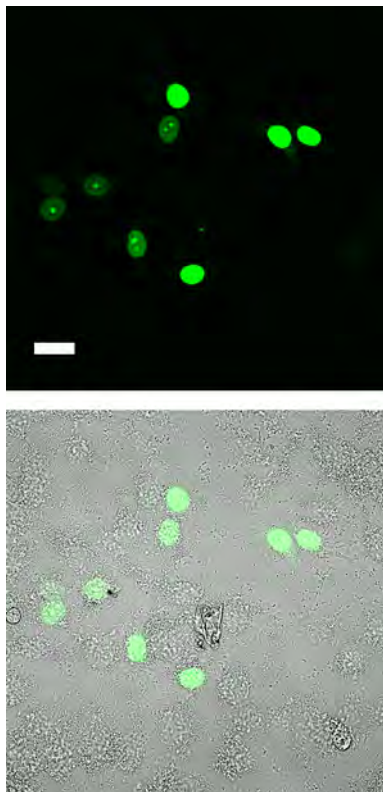
Figure IV-23: Confocal images of fixed HeLa cells after expression of CRBP$_{nona}$-NLS and staining with anti-FLAG-AlexaFluor488. Scale bar, 20 μm.

255

alternative chromophores other than MCRA with significantly lower background fluorescence are necessary for a more reliable imaging with hCRBPII-tag. However, control data on labeling and colocalization of GalT-CRBP$_{nona}$-EGFP and GalT-EGFP-CRBP$_{nona}$ proteins are still necessary to confirm successful expression of the full-length fusion in order to understand if expression or reaction is a problem.

All the images described up to this point were collected using HeLa cells, because it is the most widely used cell line in imaging, and many images for fluorescence comparison and trouble shooting are available. Besides HeLa cells we briefly tested expression and labeling efficiency of CRBP$_{nona}$-NLS and CRBP$_{nona}$-CAAX in osteosarcoma cells (U2OS), monkey kidney fibroblast-like
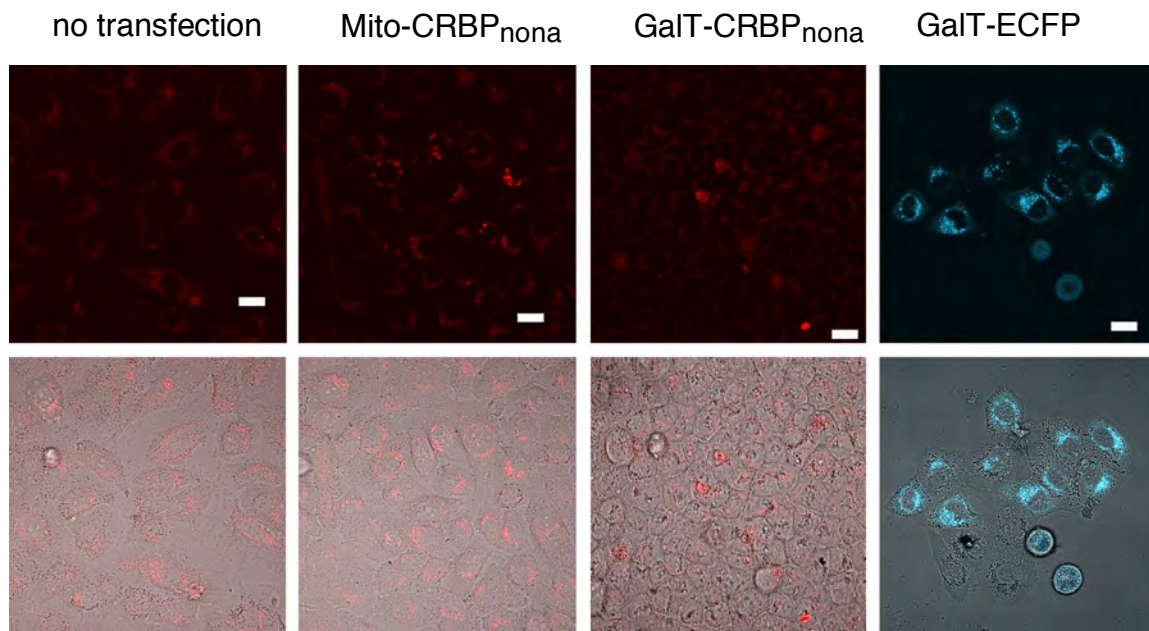
Figure IV-24: Confocal images of HeLa cells with highlighted constructs. Imaging of ECFP: excitation 594 nm, emission band pass filter 475–525 nm. Overlay with DIC image is provided below confocal image. Scale bar, 20 μm.

cells (COS-7) and kidney embryonic cells (HEK293). Both U2OS and COS-7 cells readily uptake MCRA under the same conditions as HeLa (1 min incubation, PBS wash and immediate imaging), providing nuclear localization with red fluorescence. Unfortunately, red fluorescence was observed in other intracellular areas (**Figure IV-25**). Interestingly, HEK293 cells had less background fluorescence than HeLa cells and a better localization pattern, under the same conditions for both constructs (**Figure IV-25**).

However, HEK293 cells tend to detach from the plate and an additional coating with fibronectin was essential to keep the cells on the imaging slide. Fibronectin is a glycoprotein that binds the extracellular components like collagen
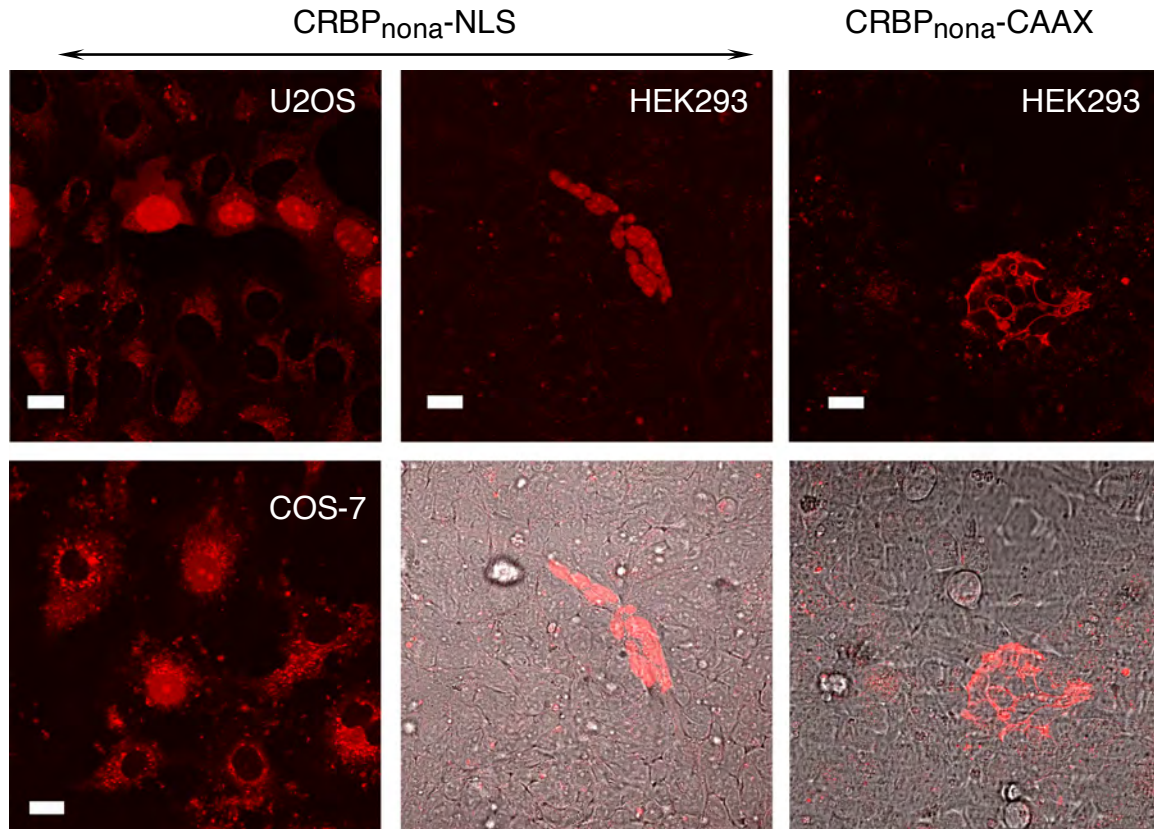
Figure IV-25: Confocal images of CRBP$_{nona}$ localized in indicated cell lines after 1 min MCRA addition, wash and direct imaging. Scale bar, 20 μm.

and fibrin, therefore, facilitating better eukaryotic cell attachment. Even when the slide is coated with fibronectin, extra care should be taken during wash steps after adding transfection reagents and chromophore. It is also important to note that sometimes the red fluorescence comes from the coated plate with fibronectin where no cells are attached. It is likely that, the surface lysine residues of fibronectin are capable of binding MCRA that leads to fluorescence. If a better method is found to attach the HEK293 cells, it will be preferable to use HEK293 cells in imaging experiments.

IV.7 Control experiments of hCRBPII labeling system

If one expects to introduce hCRBPII mutant as a fluorescent protein tag, a few standard control experiments should be performed. Non-specific red fluorescence observed after MCRA addition to non-transfected cells originates from PSB formation of the fluorogenic MCRA with available amines. MCRA-PSB is the only fluorescent species observed with an excitation wavelength at 594 nm, because MCRA-SB and free MCRA absorb more than 100 nm away (λ_{max} = 425 nm and λ_{max} = 492 nm, respectively). Besides MCRA, natural porphyrins can absorb in the same wavelength regime. However, the red fluorescence from untreated cells is not nearly as bright.

Because of the non-specific reactivity in some intracellular compartments, one needs to identify the origin of this red fluorescence. Traditionally, this is done through colocalization experiments with fluorescent compounds designed to go to

specific organelles. These compounds are known as trackers and sensor dyes. Many of them are commercially available (see **Chapter I** for description). They work based on the change in the protonation state, dictated by the pH difference in the cytosol and the specific organelle. This pH change triggers different events in two types of dyes: the trackers change membrane permeability, and the sensors change their spectroscopic properties. Localization of these dyes is temporary in live cells. Thus, delayed imaging may give inaccurate results. When choosing a dye for a colocalization experiment, one should avoid spectral overlap of the tracker or sensor and the chromophore under investigation.

LysoTracker Green and MitoTracker Green were chosen for colocalization with MCRA in non-transfected HeLa cells. To define the nucleus of the stained cells, Hoechst 33342 was also added. Colocalization between LysoTracker and MCRA-PSB was greater as compared to MitoTracker and MCRA-PSB (**Figure IV-26**); Pearson's coefficients were 0.64 and 0.45, respectively. This suggested that MCRA-PSB is generated in lysosomes, which is consistent with iminium formation in acidic media. Moreover, lysosomal localization of MCRA-PSB is consistent with the mode of function for lysotracker. Neutral MCRA easily penetrates the lysosomal membrane, but does not dissociate out after amine and/or iminium protonation.

Cytotoxicity studies of any new labeling reagent are also important, since the externally added compounds might lead to cell death and false results. Moreover, MCRA is an aldehyde, a functional group that is often used in protein
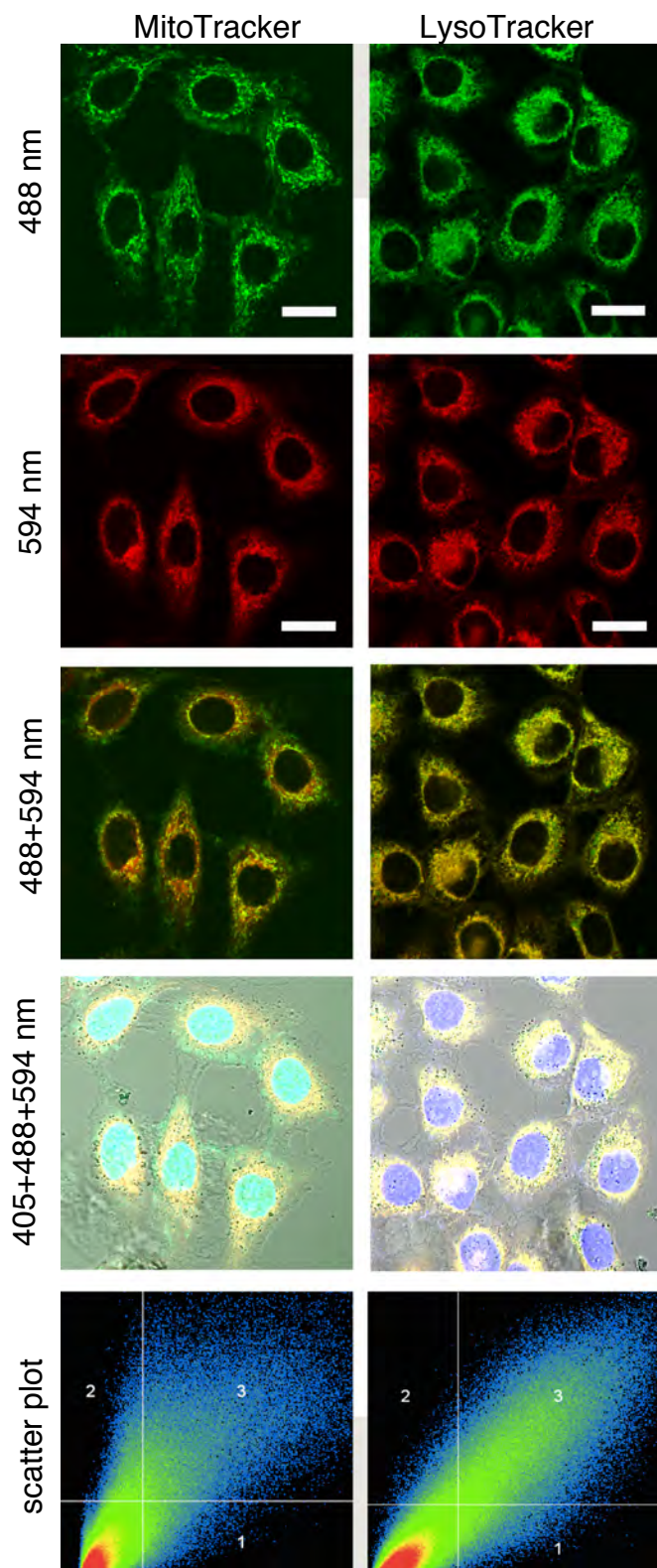
Figure IV-26: Confocal images and colocalization scatter plots of HeLa cells coincubated with MCRA and MitoTracker Green or LysoTracker Green. See Materials and methods for details. Scale bar, 20 μm.

crosslinking and cell preservation. We did not observe any adverse effects at the MCRA concentration used for imaging studies described above. Nevertheless, the cell viability assay of MCRA was conducted to demonstrate the broad applicability of the designed system.

To address this point, we used the MTT assay, a colorimetric test for cell viability. Briefly, healthy cells contain NAD(P)H oxidoreductases that effectively reduce the tetrazolium MTT dye to formazan. It triggers the progression of purple color that can be quantified with UV-vis. The more alive cells that are present in solution, the higher absorbance intensity of the purple dye is observed. The measurements acquired at various concentrations of a test compound, with constant amount of MTT, are normalized to a negative control with MTT dye.

The MTT assay revealed that over 94% of the HeLa cells survived a MCRA concentration range of 0.1–1 μM (**Figure IV-27**). Increasing the chromophore concentration caused an increased cell death, with only 7% of the
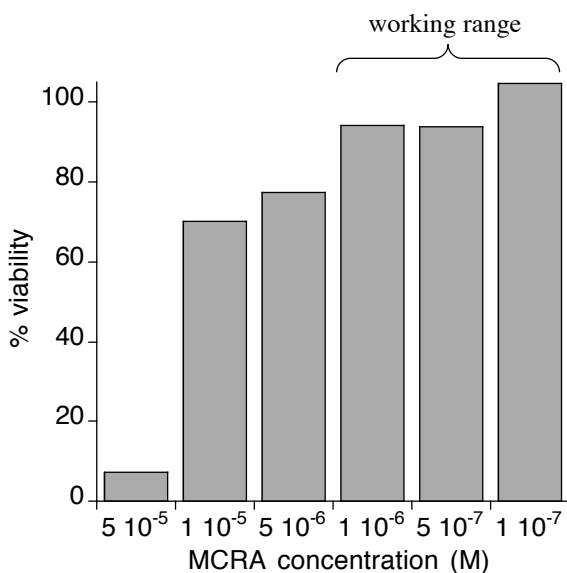
Figure IV-27: Viability of HeLa cells >94% at the working concentration range of MCRA from MTT assay.

cells remaining alive at MCRA concentration of 50 μM. MCRA concentrations

kept below 1 μM are advisable to retain highest cell viability.

To obtain the optimal molarity for visualization, three concentrations of

MCRA were tested in HeLa cells for CRBP$_{nona}$ labeling (**Figure IV-28**). The laser

power and gain/amplifier offset were set to the same parameters so an unlabeled

comparison could be made. The lowest concentration, 125 nM, gave significantly

lower background fluorescence. However, the visual intensity of the labeled

protein was slightly lower than in the two other experiments. The background

fluorescence was much higher after reaction with 500 nM, as MCRA compared to

250 nM. Therefore, 250 nM concentration of MCRA gives the best signal-to-noise
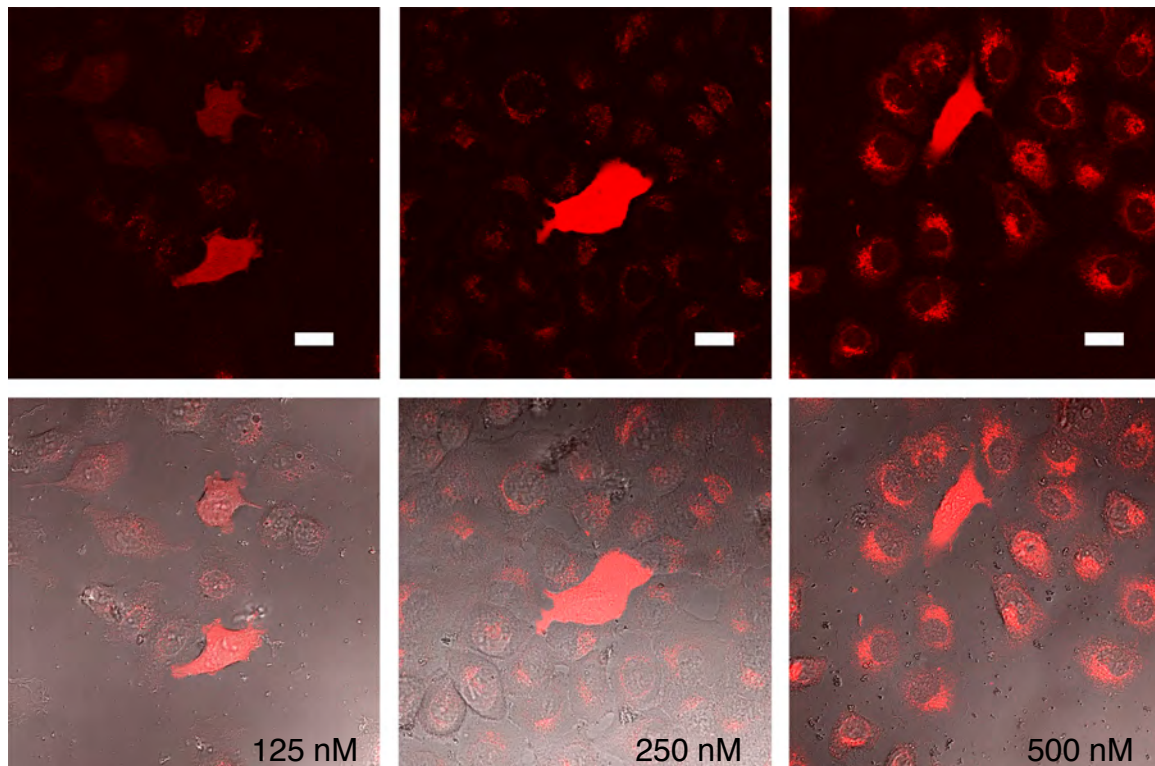
Figure IV-28: Confocal images of CRBP$_{nona}$ expressed in cytosol of HeLa cells incubated at indicated MCRA concentration for 1 min. Overlay with DIC image is provided below the confocal image. Scale bar, 20 μm.

ratio after one minute of reaction (**Figure IV-28**).

Prolonged imaging of cells might interfere with interpretation of results, especially when cells are imaged at room temperature, while their ideal growth temperature is 37 °C. If no protein trafficking is required and one can capture the specific process for later visualization, the imaging of fixed cells can be
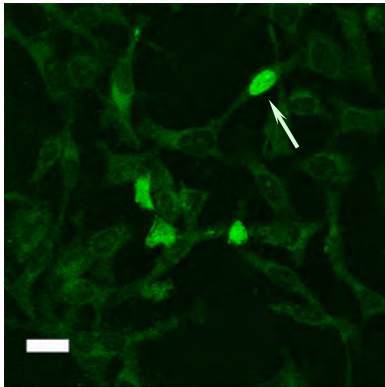
Pseudocolored green signal corresponds to the fluorescent anti-FLAG labeling. Arrow indicates the cell with successfully labeled CRBP$_{nona}$-NLS (nucleus, fixed HeLa cells)
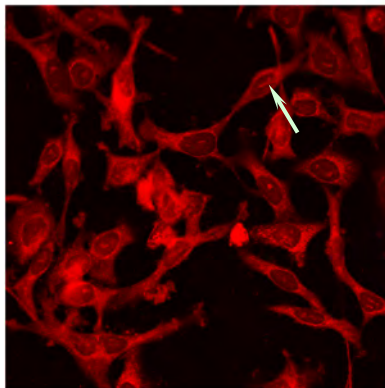
Pseudocolored red signal corresponds to the MCRA labeling. Arrow indicates the spot where MCRA would become the most fluorescent in case of successful labeling reaction.
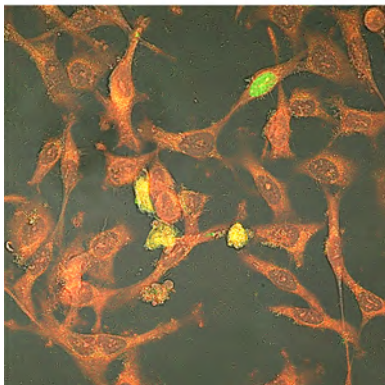
Overlay of two top images and DIC.

Figure IV-29: Confocal images of fixed HeLa cells transfected with CRBP$_{nona}$-NLS and incubated with anti-FLAG (green) and MCRA (red). Scale bar, 20 μm.

considered. Fixation is the process of mechanical preservation of the cells. Fixed cells look the same as live ones. But all biochemical transformations are terminated. There are many protocols for cell fixation. For example, formaldehyde fixation is often recommended to preserve the tissue that contains fluorescent proteins. On the other hand, if lysine reactivity needs to be preserved for the next experiment one would avoid formaldehyde fixation because it crosslinks most intracellular amines.

After trying various protocols, the methanol fixation was found to be the most effective. In two individual experiments, MCRA was incubated with HeLa cells expressing $CRBP_{nona}$-NLS for 1 min, before and after fixation. In both cases, the cells were also costained with the fluorescent FLAG-recognizing antibody for control of the protein expression and localization. If cells were labeled after MCRA addition, all red fluorescence corresponding to MCRA-PSB disappeared (data not shown). However, when cells were incubated with MCRA after fixation, excessive amount of background was observed that showed no colocolization with green fluorescence from anti-FLAG (**Figure IV-29**).

It is possible that the MCRA, soluble in methanol, was simply washed out in the first case. If this is true, an optimized formaldehyde fixation may work better for the fixed-cell imaging of protein with MCRA. In the second experiment, membrane permeability might be perturbed with methanol and, consequently, the chromophore may penetrate directly through the pores in the membrane, instead of endocytosis. The latter suggestion is even more convincing as evident by the

non-specific red fluorescence being distributed all over the cells, while before it had a specific dot-like pattern. To reach through the nucleus and other organelles, MCRA has to penetrate multiple membranes. This was not occurring in live cells. Although the experiment was unsuccessful for the fixed-cell specific
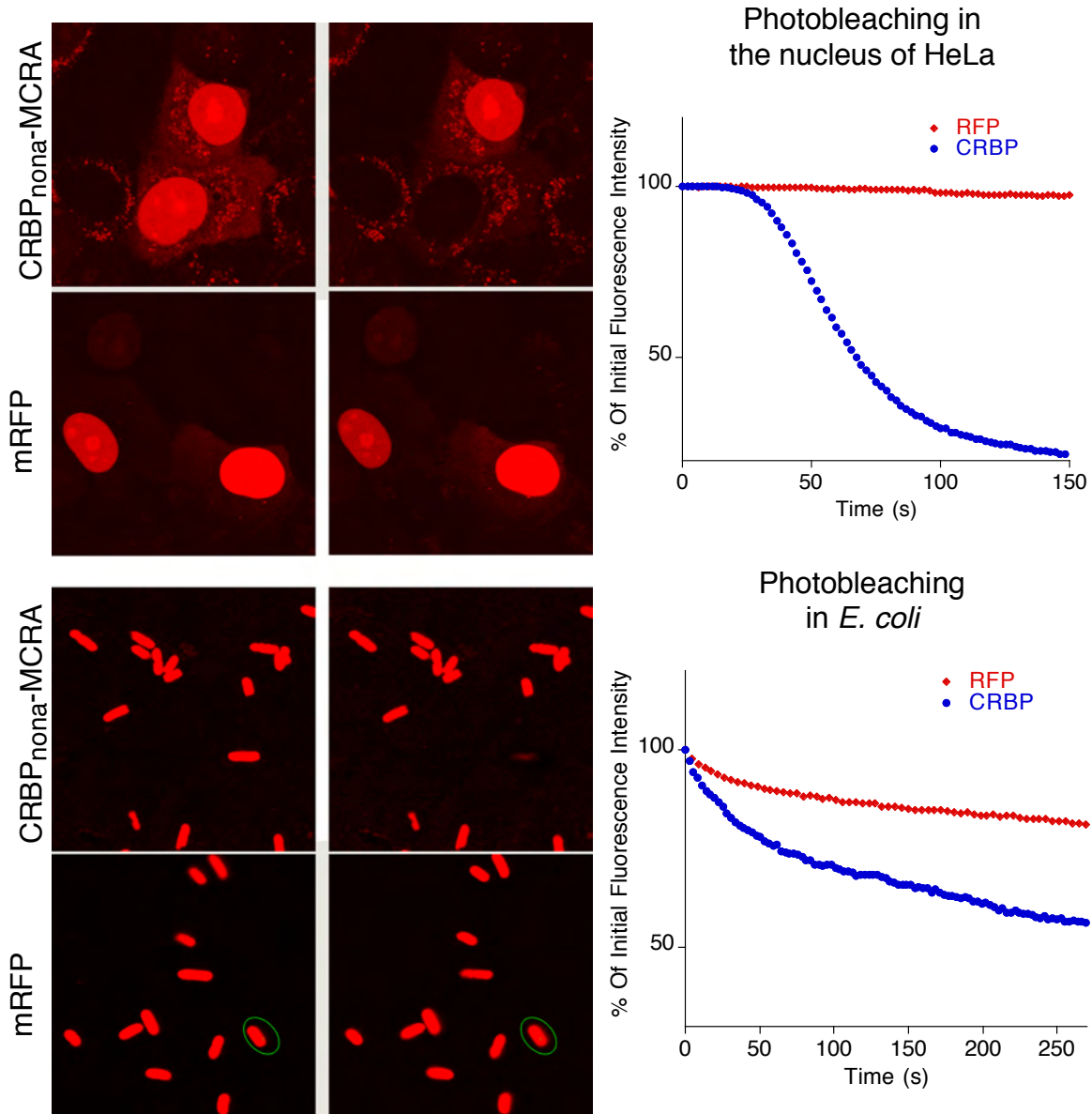
Figure IV-30: Photobleaching experiments of CRBP$_{nona}$ bound with MCRA and mRFP in comparison. Data sets collected for CRBP$_{nona}$-NLS and mRFP-NLS in HeLa and CRBP$_{nona}$ and mRFP in *E. coli*. See Main text and Materials and methods for details.

labeling, it suggested some insight to the mechanism of chromophore penetration in the live-cell experiments.

Finally, one should always check the photostability of the newly developed fluorescent probe, relative to a known standard. The photobleaching data for CRBP$_{nona}$ was acquired in two separate experiments, using HeLa and *E. coli* expression platform. The bleaching time was compared to mRFP that absorbs in a similar wavelength regime. For experiments in eukaryotic cells, both proteins were targeted to the nucleus to provide a more consistent area for photobleaching. Continuous irradiation at maximum laser power (594 nm, 2.0 mW) for 150 seconds photo-bleached CRBP$_{nona}$-MCRA with emission decreasing to 20% from the original one (**Figure IV-30**). At identical conditions mRFP sustained more than 95% of its original fluorescence. Experiments in *E. coli* showed faster decay for CRBP$_{nona}$-MCRA fluorescence than for mRFP. However, fluorescence of CRBP$_{nona}$-MCRA decayed only to 50% as opposed to 80% of mRFP (**Figure IV-30**).

IV.8 Photobleaching and unique recovery of hCRBPII fluorescence

The general advantages of any labeling reaction with a fluorophore in comparison to the GFP family of proteins would be valid for hCRBPII-tag. These claims usually include oxygen-free expression, small protein size, possibility for multiple color variants with one protein and no protein toxicity if overexpressed in

the live cells. Compared to other methods, our labeling reagent is much smaller in size and can easily penetrate inside the cell without additives.
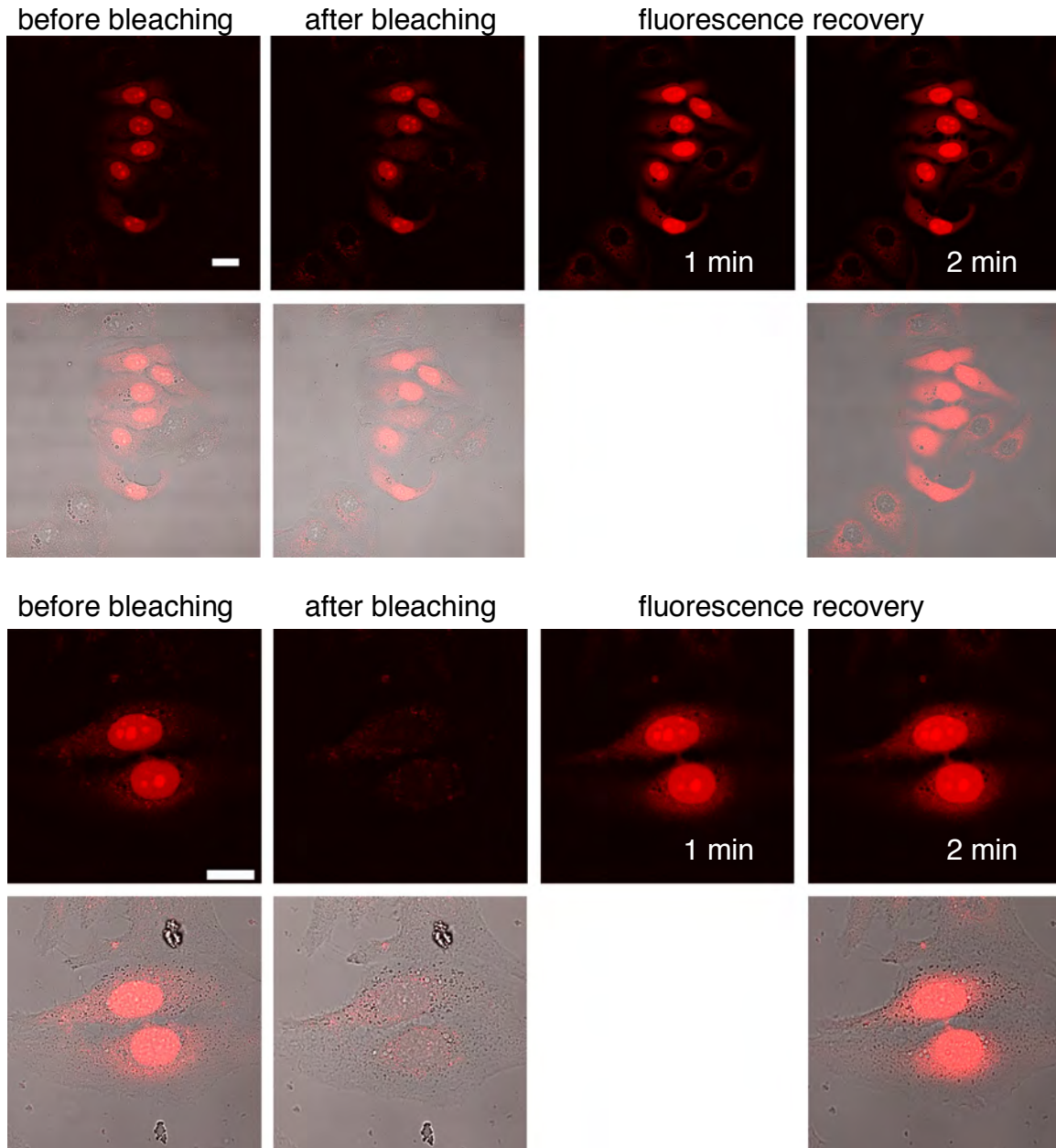
Figure IV-31: Photobleaching and fluorescence recovery experiments with second aliquot of MCRA added to HeLa cells expressing $CRBP_{nona}$-NLS. Two experiments utilized different magnification. After 1 min of recovery the DIC image was not acquired in order to get an image after 2 min (one image recording takes 1 min using the same confocal settings). Scale bars, 20 μm.

Searching for one unique application of the designed labeling system, we focused on the most distinctive advantages of hCRBPII-tag relative to the other labeling methods. The rapid binding of MCRA to $CRBP_{nona}$ might find many applications. As an example, if fluorescence should be switched on at a specific time point, this system shows potential. One might also find a positive outcome in reversibility of iminium formation. For example, if two ligands of significantly different affinities are capable of binding to hCRBPII mutant, a ligand exchange might be possible. If photobleaching of MCRA proceeds with the same mechanism as the cyanine dyes, singlet oxygen would oxidize MCRA. The

Successfully labeled cell

Cell nucleus after photobleaching

Confocal images of the same photobleached cell without new MCRA addition after 1 min and 5 min.
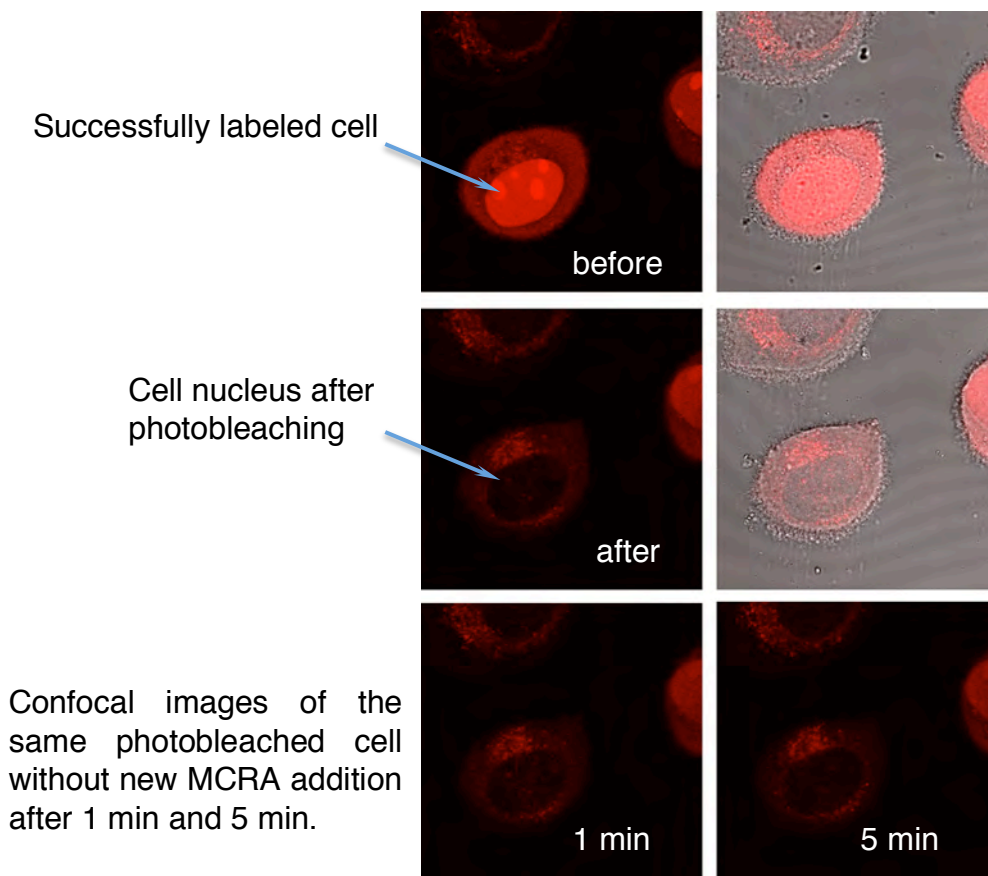
Figure IV-32: Confocal images before and after photobleaching followed by no fluorescence recovery without new addition of MCRA (HeLa cells).

photobleached product is more hydrophilic,[41] and thus the likelihood of it staying in the hydrophobic hCRBPII cavity would be unlikely. Therefore, MCRA might outcompete the PSB of the photo-product.

We combined both of the described advantages and utilized them to make a probe with unique fluorescence recovery. All fluorophores tend to undergo photobleaching, although with different rates. In the case of fluorescent protein tags, this photobleaching event would be irreversible regardless of the fluorophore source. Most of the fluorescent proteins from the GFP family cannot be recovered. There are only a few exceptions of photoswitchable probes that can switch on and off under light irradiation of a precise wavelength. Protein systems, such as TMP-tag, SNAP-tag, Halo-tag, which operate based on the site-specific labeling, always employ an irreversible reaction. Therefore, if the fluorophore is bleached, the new round of protein expression and labeling would be required.

The cyanine dyes are not among the fluorophores that show outstanding photostability. They often suffer from singlet oxygen degradation, induced by the strong light irradiation. Cyanine dye formation inside the protein is novel, but the data suggest a relatively low photostability of the probe. It is likely that after photobleching the degraded product dissociates out of the protein binding pocket, because of the low affinity to the protein. Therefore, a new equivalent of MCRA can bind to the same protein again and recover the red fluorescence.

This experiment was conducted in HeLa cells with $CRBP_{nona}$-NLS fusion (**Figure IV-31**). The cells were labeled and visualized the same as before (first column). Subsequently, the nucleus of selected cells was photobleached to the dark state and imaged again (second column). Irradiation time depends on selected area and initial signal intensity, and thus duration of photobleaching is not the same for all experiments. Following photobleaching, the cells were supplied with 250 nM of MCRA, in warm PBS, for 10 sec and then washed with warm chromophore-free PBS once. Imaging data after 1 min and 2 min from recovery appeared the same. Control experiment without new MCRA addition after photobleaching did not show fluorescence recovery even after 5 min (**Figure IV-32**).

After successful fluorescence recovery, we hypothesized that it might be possible to cycle this process multiple times. Three rounds of photobleaching and recovery are summarized in **Figure IV-33**. This worked as expected, with most of the signal recovered. Unfortunately, multiple additions of the fluorophore led to an increase in background fluorescence. Moreover, the cells started shrinking, suggesting that the increased stress was detrimental to cell viability (**Figure IV-33**, round 3).
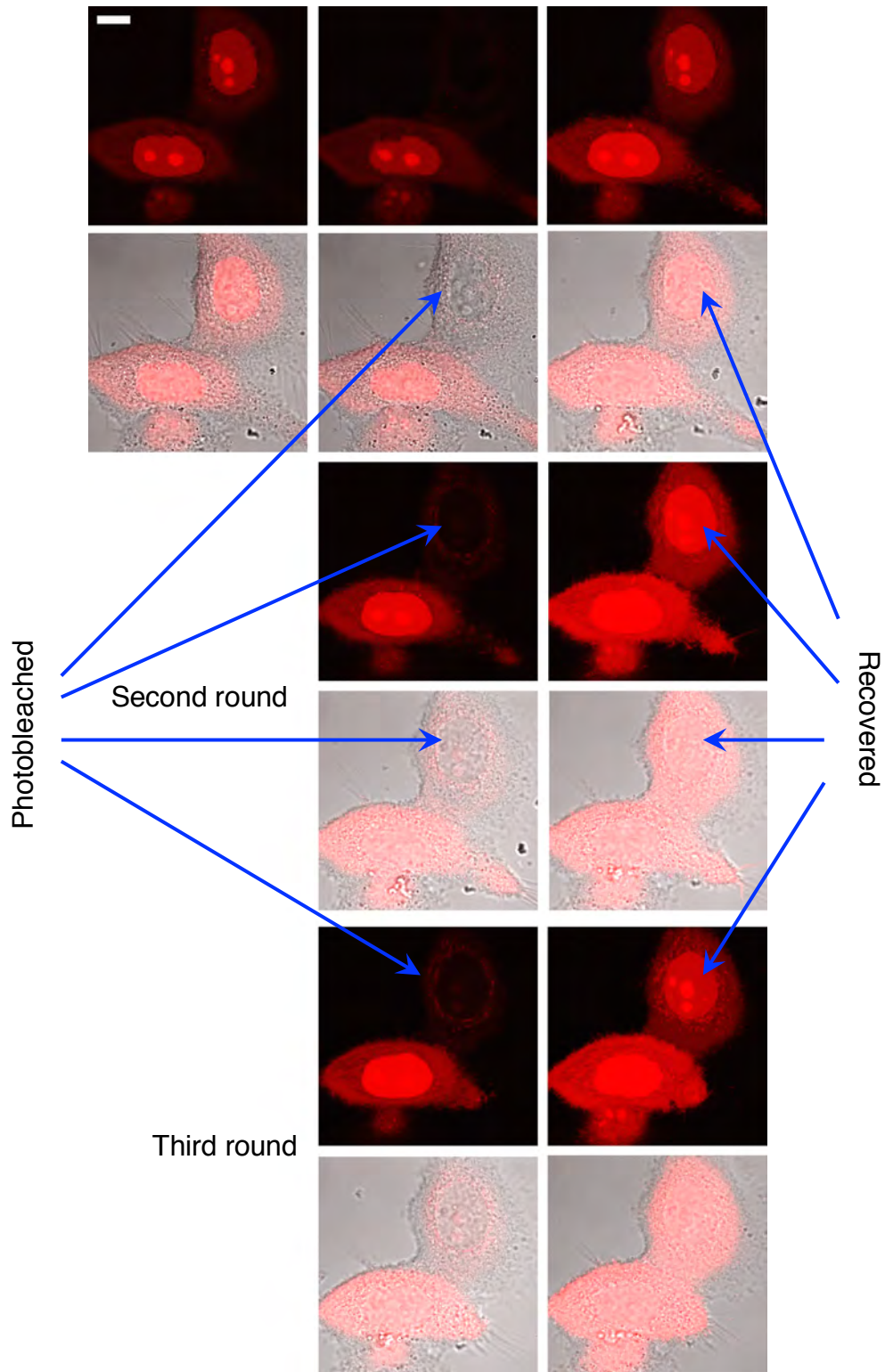
Figure IV-33: Confocal images of three rounds of photobleaching and fluorescence recovery by addition of 0.25 μM MCRA to HeLa cells expressing CRBP$_{nona}$-NLS. One "round" includes photobleaching and 1 min recovery. For details see Main text and Materials and methods. Scale bar, 10 μm.

IV.9 Conclusion and significance

In summary, a full set of optimized experiments of a protein tag with red fluorescence based on hCRBPII labeling with MCRA has been presented. This method requires less than a minute to complete the labeling reaction and 'lights up' immediately after MCRA addition. This is the fastest protein labeling method reported to date. It potentially can be used for tracking proteins with a short life-time or for the imaging of fast protein expression induced by an exogenous factor. In control experiments we optimized the chromophore addition and concentration to achieve the lowest background fluorescence. We compared the non-specific fluorescence with the known localization dyes to understand the origin of the background signal, showing that most of the non-specific fluorescence is due to accumulation of the fluorophore in the lysosome. Toxicity of MCRA is negligible at working concentrations. Although MCRA bound hCRBPII tags photobleach faster than mRFP under extreme conditions, they are photostable to normal imaging conditions. However, the photobleached pigment was capable of binding another molecule of MCRA, which restored fluorescence, thus providing a unique property not shared with other fluorescent tags. This probe can be recovered and imaged immediately multiple times, and might find utility in imaging systems that would benefit from defining a fluorescent 'pulse'.

IV.10 Materials and methods:

The hCRBPII mutants were obtained from previously cloned plasmids in pET17b. CRABPII and hCRBPII mutants were obtained according to the same protocol described in Chapter II. Primers for hCRBPII mutagenesis were:

L117E

5'-CAAGCTGTAC<u>GAG</u>GAGCTGACC-3'

5'-GGTCAGCTC<u>CTC</u>GTACAGCTTG-3'

L117D

5'-CAAGCTGTAC<u>GAC</u>GAGCTGACC-3'

5'-GGTCAGCTC<u>GTC</u>GTACAGCTTG-3'

R58Y

5'-CTAGCACATTC<u>TAC</u>AACTATGATGTG-3'

5'-CACATCATAGTT<u>GTA</u>GAATGTGCTAG-3'

T51V:T53V

5'-CAAG<u>GTA</u>AAA<u>GTC</u>ACTAGCACATTC-3'

5'-GAATGTGCTAGT<u>GAC</u>TTT<u>TAC</u>CTTG-3'

Preparation of other plasmids. All restriction sites for cloning were chosen with sticky ends (HindIII, BamHI, EcoRI, XhoI, NotI). For all PCR amplifications described herein Phusion High-Fidelity DNA polymerase was used according to the protocol:

Total Reaction Volume	50 μL
Template (DNA plasmid)	100 ng (x μl)
Primer forward	50 pmol (y μl)
Primer reverse	50 pmol (z μl)
dNTP	1 μL
5× pfu buffer	10 μL
Phusion (DNApolymerase)	0.5 μL
DI water	50-x-y-z-7 μL

PCR program

1×	98 ºC	30 sec
	98 ºC	10 sec
35×	52 ºC	20 sec
	72 ºC	15 sec per kb
1×	72 ºC	10 min
1×	4 ºC	10 min

The DNA fragment of hCRBPII mutants, β-actin or Rab5 were amplified by PCR with appropriate primers. NLS and CAAX sequences required two rounds of PCR amplification with annealing complimentary oligos. For example, the method of annealing the complementary oligos was used for cloning hCRBPII-NLS into pFLAG-CMV2 vector. The primers in the first PCR were:

hCRBPII_HindIII _for

5'-GCAAGCTTACGAGGGACTTCAATGGAACC-3'

NLS1_hCRBPII_rev

5'-CTTCTTGGGGTCCACCTTCCTCTTCTTCTTGGGGTCAGCCCTGCTCTT
CTTTTTGAACAC-3'

Amplified and purified fragment was used in the second PCR as a template with primers:

hCRBPII_HindIII_for

5'-GCAAGCTTACGAGGGACTTCAATGGAACC-3'

EcoRI_NLS2_rev

5'-GAATTCACACCTTCCTCTTCTTCTTGGGGTCCACCTTCCTCTTCTTCTTG
GGGTCCAC-3'

where the underlined sequence represents restriction sites, grey sequence is complimentary to hCRBPII gene and blue sequence is complimentary between two primers.

The final fragment was purified by Wizard® SV Gel and PCR Clean-Up System (Promega) from 1% agarose gel in 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 using T4 DNA Ligase (New England BioLabs). Ligated product was transformed into *E. coli* XL-1 blue competent cells and grown on LB-agar plates supplemented with antibiotics at 37 ºC for 18 hours. Colonies (4-8) were inoculated in LB medium (15 mL) and incubated at 37 ºC

while shaking, for 10 hours. DNA purification was performed using Promega Wizard® Plus SV Miniprep DNA purification kit following the suggested protocol. The DNA sequence was verified with the corresponding *sequencing primers* (for example, plasmids in pFLAG-CMV2 with seq_pFlag-CMV2 primer, 30 pmol) by the MSU gene sequencing facility.

Cloning primers:

CRABPII_XhoI _for

5'-CCCTCGAGGCCAAACTTCTCTGGCAACTGGAAAATC-3'

CRABPII_HindIII_rev_RB

5'-GGAAGCTTCCCTCTCGGACGAAGACAAGGGTGCACAC-3'

CRABPII_HindIII_rev_noRB

5'-GGAAGCT**TCA**CTCTCGGACGAAGACAAGGGTGCACAC-3'

hCRBPII_HindIII _for

5'-GCAAGCTTACGAGGGACTTCAATGGAACC-3'

hCRBPII_EcoRI _rev_Stop

5'-CCGAATTCTCACTTCTTTTTGAACAC-3'

hCRBPII_EcoRI _rev(1NLS) (no localization)

5'-CCGAATT**CTA**GCCTACTTTTCTTTTTTTTTTGGGCTTCTTTTTGAACACTT GACGGC-3'

NLS1_hCRBPII_rev

5'-CTTCTTGGGGTCCACCTTCCTCTTCTTCTTGGGGTCAGCCCTGCTCTT CTTTTTGAACAC-3'

EcoRI_NLS2 _rev

5'-GAAT**TCA**CACCTTCCTCTTCTTCTTGGGGTCCACCTTCCTCTTCTTCTTG GGGTCCAC-3'

CAAX1_hCRBPII_rev(1)

5'-CACTCTCATCAGGAGGGTTCAGCTTCTTCTTTTTGAACAC-3'

EcoRI_CAAX2_rev(1)

5'-GAAT**TCA**CAGGAGAGCACACACTTGCAGCTCATGCAGCCGGGGGCCAC TCTCATCAG-3'

CAAX1_hCRBPII_rev(2)

5'-CTTTTTACCCTTCTTTTTGAACACTTGACGG-3'

EcoRI_CAAX2_rev(2)

5'-CC<u>GAATT</u>**CTA**CATAATTACACACTTTGTCTTTGACTTCTTTTTCTTCTTTTT TACCCTT-3'

hCRBPII_EcoRI _rev_noStop

5'-CC<u>GAATTC</u>CGACTTCTTTTTGAACAC-3'

CRBP_PKKKRKVGSNS_Rab5(for) (for PCR to align the fusion after cloning

hCRBPII-1NLS-stop-Rab5)

5'-CAAAAAGAAGGCCAAAACAAAAACAAAAGGAGGC<u>GG</u>AATTCCGATGGC-3'

CRBP_PKKKRKVGSNS_Rab5(rev)

5'-GCCATCGGAATT<u>CC</u>GCCTCCTTTTGTTTTTGTTTTGGCCTTCTTTTTG-3'

hCRBPII_EcoRI _rev_GAAGGSS

5'-CG<u>GAATTC</u>GAGCTGCCTCCTGCAGCTCCCTTCTTTTTGAAC-3'

Rab5(for)EcoRI (amplified from the sequence aligned up to EcoRI)

5'-CG<u>GAATTC</u>CGATGGCTAG-3'

Rab5(rev)BamHI

5'-CG<u>GGATCC</u>**TCA**GTTACTACAACACTG-3'

β-actin(for)EcoRI

5'-CC<u>GAATTC</u>CATGGATGATGATATC-3'

β-actin(rev)BamHI

5'-CC<u>GGATCC</u>**CTA**GAAGCATTTGC-3'

Mito1_CRBP(for)

5'-CGGCGGCTCCCAGTGCCGCGCGCCAAGATCCATTCGTTGATGACGAGGG

AC-3'

Mito1GGS_CRBP(for)

5'-CGGCGGCTCCCAGTGCCGCGCGCCAAGATCCATTCGTTGGGTTCCGGT

ATGACGAGGGAC-3'

BamHI_Mito2(for)

5'-<u>GGATCC</u>ATGTCCGTCCTGACGCCGCTGCTGCTGCGGGGCTTGACAGGCT

CGGCCCGGCGG-3'

Mito1_SGAAAGGSQ_CRBP(for) (amplified from SGAAAGGSQ_CRBP)

5'-CAGGCTCGGCCCGGCGGCTCCCAGTGCCGCGCGCCAAGATCCATTCGT

TGTCCGGAGCCG-3'

BamHI_kozakMito2(for)

5'-<u>GGATCC</u>CTCGCCACC**ATG**TCCGTCCTGACGCCGCTGCTGCTGCGGGGCT

TGACAGGCTCG-3'

hCRBPII_XhoI_rev_noStop

5'-GCTCGAGCTTCTTTTTGAACAC-3'

BamHI_CRBP_GalT(for)

5'-CGGGATCCAATGACGAGGGACTTCAATGG-3'

BamHI_CRBP_GGSG_GalT(for)

5'-CGGGATCCAGGGGGGTTCCGGTATGACGAGGGACTTC-3'

NotI_CRBP_GalT(rev)

5'-CCGCGGCCGC**TCA**CTTCTTTTTGAACACTTGAC-3'

mRFP(for)HindIII

5'-GACGACAAGCTTATGGCCTCCTCC-3'

NLS1-mRFP(rev) (for second round of PCR **EcoRI_NLS2 _rev** was used)

5'-CTTCTTGGGGTCCACCTTCCTCTTCTTCTTGGGGTCAGCCCTGCTGGCGC

CGGTGGAGTG-3'

Sequencing primers:

pEGFP_seq(for) (sequencing of any plasmid with EGFP/ECFP/EYFP region)

5'-GAGTTCGTGACCGCCGCCGGGATC-3'

seq_pFlag-CMV2 (sequencing of any plasmid in pFLAG-CMV2)

5'-GGGAGGTCTATATAAGCAGAGCTCG-3'

CMVend_seq (sequencing of any plasmid with CMV promoter region)

5'-GGTCTATATAAGCAGAGCTGGTTTAG-3'

seq_Tag4(for) (sequencing of any plasmid in pCMV-Tag4)

5'-CCCTCACTAAAGGGAACAAAAGCTG-3'

Kinetic measurements were acquired using UV-vis absorbance at 600 nm for MCRA-PSB or at 355 nm for Rt-SB. All experiments were carried at 23 °C unless otherwise noted (temperature controller). The protein and chromophore concentrations were consistent among all experiments. In general, protein (20 μM in PBS buffer, 20 mM pH 7.3 or in citric acid buffer, 100 mM pH 5.0, as per description in text) and MCRA (10 μM in ethanol) were mixed and the spectrum was recorded immediately (1 point per 0.5 sec). Collected data points were processed as a scatter plot in KaleidaGraph 4.1.3 and fitted to the first-order reaction kinetics. The data points were fitted to equation:

$$A = A_0 \times (1 - e^{-kt}) + c,$$

where A_0 is the final absorbance value after the complex formation is complete; A is absorbance value at each recorded time-point; k is the rate constant; t is each time point when A_0 is acquired; and c is a free constant that accounts for the time delay from MCRA addition to the first data point. Since the molar absorptivity and the length of the cuvette are constant throughout the measurement, the concentration is proportional to absorption according to the Beer-Lambert law. Therefore, absorbance is directly used in the equation.

Quantum yield measurements. The quantum yields were determined by comparing the proportional gradients of the unknown sample and known standard using the following equation:

$$QY_x = QY_{std} \times \left(\frac{grad_x}{grad_{std}}\right) \times \left(\frac{\eta_x}{\eta_{std}}\right)^2$$

where "grad" is equal to the slope of a plot relating the integrated emission (y axis) to the absorbance (x axis) using a minimum of three different concentrations per sample; and η is the refractive index of the solvent used for the fluorescence readings.

In all experiments three or four concentrations of sample were prepared by dilution of the most concentrated one (1 mL for UV-vis and 3 mL for fluorescence). The absorption of the most concentrated sample should not exceed 0.1 at 565 nm. Fluorescence and absorbance spectra were acquired in parallel. All emission spectra were excited at 565 nm and collected from 575 nm to 800 nm. An integrated emission curve was plotted against absorption value at 565 nm. The data points were fitted to a linear equation and the slope was used in equation described above. Dyes Oxazine-170 and Oxazine-1 dissolved in ethanol were used as fluorescent standards. PBS buffer (20 mM, pH = 7.3) was used for dilution of protein samples. The temperature was maintained 23 ˚C throughout the experiments.

Mammalian cell culture. All cells (HeLa, U2OS, HEK293, COS-7) were cultured in Dulbecco's modified Eagle medium (DMEM, Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS, BioWest) and 1x Penicillin-Streptomycin-Glutamine (PSG, GIBCO). The tissue cultures were maintained at 37 °C under 5% CO_2, 10% O_2 and 70% H_2O atmosphere. For imaging the cells

were plated on the sterile 8-well ibiTreat µ-Slide (ibidi) in serum-free and antibiotic-free colorless 1640-RPMI medium. The slide can be reused multiple times without coating by thorough wash with bleach, water and ethanol followed by UV-light sterilization.

Transfection protocol and labeling reaction. The cells were washed with warm PBS buffer and transfected with expression plasmids (300-500 ng per 1.0 cm^2 well) using NanoJuice Transfection Kit and serum-free/antibiotic-free DMEM (optimized reaction conditions: NanoJuice Core Transfection Reagent 0.1 µL, NanoJuice Transfection Booster 0.2 µL, DMEM 10 µL, plasmid 1 µg). After transfection (24-48 h) the cells were incubated with 250 nM MCRA in warm serum-free/antibiotic-free DMEM or Dulbecco's PBS (Sigma-Aldrich) for 1 min. The media was removed and the cells were washed twice with warm PBS and supplied with colorless 1640-RPMI medium. Each experiment had its own degradation time if necessary, specified in figures.

Confocal imaging. Cell microscopy was performed using inverted laser scanning confocal microscope (LSM510Meta, Carl Zeiss, Jena, Germany) equipped with diode, argon and HeNe lasers. Either 40x or 63x oil-immersed objective was used. MCRA-PSB/mRFP (594 nm excitation, 594 nm primary dichroic, 545 nm secondary dichroic, >615 nm emission), EGFP/LysoTracker/MitoTracker (488 nm excitation, 488 nm primary dichroic, 490 nm secondary dichroic, 505-530 nm emission), ECFP (458 nm excitation, 458 nm primary dichroic, 545 nm secondary dichroic, 475-525 nm emission), Hoechst

33342 (405 nm excitation, 405 nm primary dichroic, 545 nm secondary dichroic, >475 nm emission) and DIC images were collected and analyzed using Axiovert 200M software. Kalman averaging 8 was applied in all confocal images. Fluorescence in each experiment was normalized to the same intensity adjusting the gain and amplifier offset. All the images have pseudocolor.

Photobleaching. For photobleaching experiments HeNe laser power was set to 100% (2.0 mW). The same area was chosen for mRFP and CRBP$_{nona}$ irradiation in both eukaryotic (HeLa) and prokaryotic (BL21(DE3)pLysS) cells. Collected datasets were normalized to 100% by emission intensity, and decay scatter plots were generated in KaleidaGraph. In photobleaching and fluorescence recovery experiments the slide was fixed to the confocal stage by scotch tape to avoid movement during wash.

Colocalization experiments of MCRA-PSB with LysoTracker Green and MitoTracker Green. The mounted HeLa cells were supplied with the following mixtures: for lysosomal comparison 200 μL DMEM, LysoTracker Green 50 nM, Hoechst 33342 1 μg/mL and MCRA 250 nM, for mitochondrial comparison 200 μL DMEM, MitoTracker Green 50 nM, Hoechst 33342 1 μg/mL and MCRA 250 nM. The mixtures were incubated for 5 min at 37 °C. The colorless medium was changed once and the cells were subjected for imaging immediately.

Methanol fixation. HeLa cells were plated on sterile 6-well plate with glass slides placed at the bottom (single use only). Transfection with selected plasmids was done 48 h prior to experiment. Methanol (12 mL) was cooled to -20 °C in a

freezer for 1 h. The cells were treated with methanol (2 mL per well or as needed to cover the glass slide) at -20 °C freezer for 20 min. After that the cells were washed three times with PBS buffer (sterile Dulbecco's PBS, Sigma-Aldrich) and rehydrated with ice-cold Dulbecco's PBS buffer supplied with 10% glycerol (20 min at 4 °C, 2 mL per well). Fixed and rehydrated cells can be treated as needed (for *Immunofluorescence* see protocol below; for fixed cell imaging of labeled hCRBPII the cells were supplied with 125 nM MCRA in ice cold PBS for 1 min in amount 2 mL per well and washed three times with chromophore-free PBS). The fixed cells can be stored as long as needed at 4 °C (the longest time tried for this study was 2 weeks) only if the glass slide is under PBS buffer.

Immunofluorescence. HeLa cells were plated on 6-well plate with glass slides placed at the bottom. Transfection with selected plasmids containing FLAG-tag was done 48 h prior to experiment (for details see *Transfection protocol*). The cells were fixed with methanol at -20 °C for 20 min (for details see *Methanol fixation*). Fixed and rehydrated cells were permeabilized with Tween-20 (0.2% by volume in Dulbecco's PBS, 2 mL per well) for 2 h at 4 °C. Detergent was removed by thorough wash with Dulbecco's PBS (as needed to eliminate soapy character, usually three time) and the cells were incubated with blocking solution overnight at 4 °C (dry milk dissolved in Dulbecco's PBS, 10% by weight). Blocking reagent was removed and the cells were washed with ice-cold PBS. After that the cells were treated with FLAG antibody conjugated with Alexa Fluor 488 for 1.5 h at 4 °C. For each FLAG-tag staining experiment three

concentrations of antibody were tested in three separate wells because binding efficiency strongly depends on protein expression levels (dilutions 1:400, 1:200 and 1:100 antibody to PBS, 0.5 mL each). The cells were washed and kept in PBS at 4 °C, protected from light, until imaging.

Cytotoxicity of labeling reaction. HeLa cells were analyzed in 96-well plates using MTT[42] assay with Adriamycin as a positive control and ethanol as a negative control. Ethanol was also a solvent for Adriamycin and MCRA. Six concentrations of MCRA ranging between 50 μM and 50 nM were added to the cells in serum-free/antibiotic-free DMEM for 24 h at 37 °C. Absorbance at 550 nm was recorded using BioRad plate reader. The average of four data sets for each concentration was calculated.

Binding control of isolated EGFP-CRBP$_{nona}$ with MCRA. When working with the protein fusion it is important to remember that the rate of the reaction inside the protein can be sufficiently influenced by the designed construct. Therefore, we did a test experiment with isolated EGFP-CRBP$_{nona}$ from bacterial expression. EGFP-CRBP$_{nona}$ was cloned into pET22b vector for bacterial expression and His-Tag at C-terminus. Insert fragment was amplified in two rounds of PCR: 1) EGFP was amplified with forward GG<u>CATATG</u>ACGAGGGACCAG and reverse CCGGACTTGTACAGC primers, CRBP$_{nona}$ was amplified with forward GCTGTACAAGTCCGGAGCCGCTGCAGGAGGCAGCCAAATGACGAGGTTC

285

and reverse GCTCGAGCTTCTTTTTGAACAC primers; 2) purified products from (1) were combined and amplified with the same EGFP forward primer and CRBP reverse primer to generate the fused insert. The final amplification was ligated between NdeI and XhoI cutting sites of pET22b vector. The protein was expressed in BL21 cells at 23 °C for 12 h and purified on Ni-column. The final fusion of EGFP and hCRBPII mutant was incubated with 0.5 equiv of MCRA. Although the isolated hCRBPII nona-mutant binds within 1 min at room temperature, it takes 10 min for the fusion of the same mutant with EGFP to form PSB with MCRA.
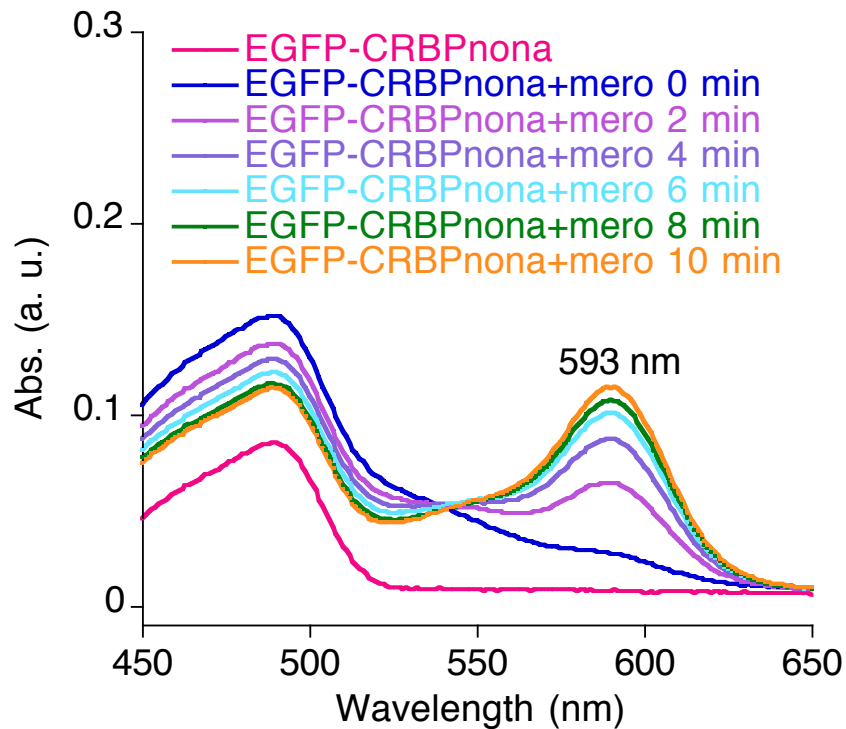
Figure IV-34: Binding of the purified EGFP-CRBP$_{nona}$-6His fusion with MCRA.

REFERENCES

REFERENCES

1.   Saxon, E.; Bertozzi, C. R., Cell surface engineering by a modified Staudinger reaction. *Science* **2000,** *287* (5460), 2007.

2.   Milles, S.; Tyagi, S.; Banterle, N.; Koehler, C.; VanDelinder, V.; Plass, T.; Neal, A. P.; Lemke, E. A., Click Strategies for Single-Molecule Protein Fluorescence. *Journal of the American Chemical Society* **2012,** *134* (11), 5187.

3.   Liu, D. S.; Tangpeerachaikul, A.; Selvaraj, R.; Taylor, M. T.; Fox, J. M.; Ting, A. Y., Diels-Alder Cycloaddition for Fluorophore Targeting to Specific Proteins inside Living Cells. *Journal of the American Chemical Society* **2012,** *134* (2), 792.

4.   Debets, M. F.; Van Berkel, S. S.; Dommerholt, J.; Dirks, A. J.; Rutjes, F. P. J. T.; Van Delft, F. L., Bioconjugation with Strained Alkenes and Alkynes. *Accounts of Chemical Research* **2011,** *44* (9), 805.

5.   Agarwal, P.; van der Weijden, J.; Sletten, E. M.; Rabuka, D.; Bertozzi, C. R., A Pictet-Spengler ligation for protein chemical modification. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110* (1), 46.

6.   Jing, C.; Cornish, V. W., Chemical Tags for Labeling Proteins Inside Living Cells. *Accounts of Chemical Research* **2011,** *44* (9), 784.

7.   Mizukami, S.; Hori, Y.; Kikuchi, K., Small-Molecule-Based Protein-Labeling Technology in Live Cell Studies: Probe-Design Concepts and Applications. *Accounts of Chemical Research* **2014,** *47* (1), 247.

8.   Rashidian, M.; Dozier, J. K.; Distefano, M. D., Enzymatic Labeling of Proteins: Techniques and Approaches. *Bioconjugate Chemistry* **2013,** *24* (8), 1277.

9.   Romanini, D. W.; Cornish, V. W., PROTEIN LABELLING Playing tag with proteins. *Nature Chemistry* **2012,** *4* (4), 248.

10.   Sahoo, H., Fluorescent labeling techniques in biomolecules: a flashback. *Rsc Advances* **2012,** *2* (18), 7017.

11.   Hori, Y.; Kikuchi, K., Protein labeling with fluorogenic probes for no-wash live-cell imaging of proteins. *Current Opinion in Chemical Biology* **2013,** *17* (4), 644.

12.	Sun, X.; Zhang, A.; Baker, B.; Sun, L.; Howard, A.; Buswell, J.; Maurel, D.; Masharina, A.; Johnsson, K.; Noren, C. J.; Xu, M.-Q.; Correa, I. R., Jr., Development of SNAP-Tag Fluorogenic Probes for Wash-Free Fluorescence Imaging. *Chembiochem* **2011,** *12* (14), 2217.

13.	Hori, Y.; Norinobu, T.; Sato, M.; Arita, K.; Shirakawa, M.; Kikuchi, K., Development of Fluorogenic Probes for Quick No-Wash Live-Cell Imaging of Intracellular Proteins. *Journal of the American Chemical Society* **2013,** *135* (33), 12360.

14.	Suh, E. H.; Liu, Y.; Connelly, S.; Genereux, J. C.; Wilson, I. A.; Kelly, J. W., Stilbene Vinyl Sulfonamides as Fluorogenic Sensors of and Traceless Covalent Kinetic Stabilizers of Transthyretin That Prevent Amyloidogenesis. *Journal of the American Chemical Society* **2013,** *135* (47), 17869.

15.	Szent-Gyorgyi, C.; Schmidt, B. A.; Creeger, Y.; Fisher, G. W.; Zakel, K. L.; Adler, S.; Fitzpatrick, J. A. J.; Woolford, C. A.; Yan, Q.; Vasilev, K. V.; Berget, P. B.; Bruchez, M. P.; Jarvik, J. W.; Waggoner, A., Fluorogen-activating single-chain antibodies for imaging cell surface proteins. *Nature Biotechnology* **2008,** *26* (2), 235.

16.	Paige, J. S.; Nguyen-Duc, T.; Song, W.; Jaffrey, S. R., Fluorescence Imaging of Cellular Metabolites with RNA. *Science* **2012,** *335* (6073), 1194.

17.	Paige, J. S.; Wu, K. Y.; Jaffrey, S. R., RNA Mimics of Green Fluorescent Protein. *Science* **2011,** *333* (6042), 642.

18.	Strack, R. L.; Jaffrey, S. R., New approaches for sensing metabolites and proteins in live cells using RNA. *Current Opinion in Chemical Biology* **2013,** *17* (4), 651.

19.	Maclaurin, D.; Venkatachalam, V.; Lee, H.; Cohen, A. E., Mechanism of voltage-sensitive fluorescence in a microbial rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110* (15), 5939.

20.	Gorbe, E.; Calatayud, A., Applications of chlorophyll fluorescence imaging technique in horticultural research: A review. *Scientia Horticulturae* **2012,** *138*, 24.

21.	Kumagai, A.; Ando, R.; Miyatake, H.; Greimel, P.; Kobayashi, T.; Hirabayashi, Y.; Shimogori, T.; Miyawaki, A., A Bilirubin-Inducible Fluorescent Protein from Eel Muscle. *Cell* **2013,** *153* (7), 1602.

22.   Broichhagen, J.; Trauner, D., Bilirubin in a New Light. *Angewandte Chemie-International Edition* **2013,** *52* (52), 13868.

23.   Crisalli, P.; Kool, E. T., Importance of ortho Proton Donors in Catalysis of Hydrazone Formation. *Organic Letters* **2013,** *15* (7), 1646.

24.   Kool, E. T.; Park, D. H.; Crisalli, P., Fast Hydrazone Reactants: Electronic and Acid/Base Effects Strongly Influence Rate at Biological pH. *Journal of the American Chemical Society* **2013,** *135* (47), 17663.

25.   Dick, F. A.; Rubin, S. M., Molecular mechanisms underlying RB protein function. *Nature Reviews Molecular Cell Biology* **2013,** *14* (5), 297.

26.   Manor, D.; Shmidt, E. N.; Budhu, A.; Flesken-Nikitin, A.; Zgola, M.; Page, R.; Nikitin, A. Y.; Noy, N., Mammary carcinoma suppression by cellular retinoic acid binding protein-II. *Cancer Research* **2003,** *63* (15), 4426.

27.   Schug, T. T.; Berry, D. C.; Shaw, N. S.; Travis, S. N.; Noy, N., Opposing effects of retinoic acid on cell growth result from alternate activation of two different nuclear receptors. *Cell* **2007,** *129* (4), 723.

28.   Schug, T. T.; Berry, D. C.; Toshkov, I. A.; Cheng, L.; Nikitin, A. Y.; Noy, N., Overcoming retinoic acid-resistance of mammary carcinomas by diverting retinoic acid from PPAR beta/delta to RAR. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** *105* (21), 7546.

29.   Sessler, R. J.; Noy, N., A ligand-activated nuclear localization signal in cellular retinoic acid binding protein-II. *Molecular Cell* **2005,** *18* (3), 343.

30.   Shaw, N.; Elholm, M.; Noy, N., Retinoic acid is a high affinity selective ligand for the peroxisome proliferator-activated receptor beta/delta. *Journal of Biological Chemistry* **2003,** *278* (43), 41589.

31.   Biggs, J. R.; Peterson, L. F.; Zhang, Y.; Kraft, A. S.; Zhang, D.-E., AML1/RUNX1 phosphorylation by cyclin-dependent kinases regulates the degradation of AML1/RUNX1 by the anaphase-promoting complex. *Molecular and Cellular Biology* **2006,** *26* (20), 7420.

32.   Kalderon, D.; Roberts, B. L.; Richardson, W. D.; Smith, A. E., A SHORT AMINO-ACID SEQUENCE ABLE TO SPECIFY NUCLEAR LOCATION. *Cell* **1984,** *39* (3), 499.

33.   Wright, L. P.; Philips, M. R., CAAX modification and membrane targeting of Ras. *Journal of Lipid Research* **2006,** *47* (5), 883.

34. Uttamapinant, C.; White, K. A.; Baruah, H.; Thompson, S.; Fernandez-Suarez, M.; Puthenveetil, S.; Ting, A. Y., A fluorophore ligase for site-specific protein labeling inside living cells. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (24), 10914.

35. Vonderheit, A.; Helenius, A., Rab7 associates with early endosomes to mediate sorting and transport of semliki forest virus to late endosomes. *Plos Biology* **2005,** *3* (7), 1225.

36. Zerial, M.; McBride, H., Rab proteins as membrane organizers. *Nature Reviews Molecular Cell Biology* **2001,** *2* (2), 107.

37. Sherer, N. M.; Lehmann, M. J.; Jimenez-Soto, L. F.; Ingmundson, A.; Horner, S. M.; Cicchetti, G.; Allen, P. G.; Pypaert, M.; Cunningham, J. M.; Mothes, W., Visualization of retroviral replication in living cells reveals budding into multivesicular bodies. *Traffic* **2003,** *4* (11), 785.

38. Trojer, P.; Li, G.; Sims, R. J., III; Vaquero, A.; Kalakonda, N.; Boccuni, P.; Lee, D.; Erdjument-Bromage, H.; Tempst, P.; Nimer, S. D.; Wang, Y.-H.; Reinberg, D., L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell* **2007,** *129* (5), 915.

39. Rizzuto, R.; Brini, M.; Pizzo, P.; Murgia, M.; Pozzan, T., CHIMERIC GREEN FLUORESCENT PROTEIN AS A TOOL FOR VISUALIZING SUBCELLULAR ORGANELLES IN LIVING CELLS. *Current Biology* **1995,** *5* (6), 635.

40. Cole, N. B.; Smith, C. L.; Sciaky, N.; Terasaki, M.; Edidin, M.; LippincottSchwartz, J., Diffusional mobility of Golgi proteins in membranes of living cells. *Science* **1996,** *273* (5276), 797.

41. Wang, M.; Holmes-Davis, R.; Rafinski, Z.; Jedrzejewska, B.; Choi, K. Y.; Zwick, M.; Bupp, C.; Izmailov, A.; Paczkowski, J.; Warner, B.; Koshinsky, H., Accelerated Photobleaching of a Cyanine Dye in the Presence of a Ternary Target DNA, PNA Probe, Dye Catalytic Complex: A Molecular Diagnostic. *Analytical Chemistry* **2009,** *81* (6), 2043.

42. Mosmann, T., RAPID COLORIMETRIC ASSAY FOR CELLULAR GROWTH AND SURVIVAL - APPLICATION TO PROLIFERATION AND CYTO-TOXICITY ASSAYS. *Journal of Immunological Methods* **1983,** *65* (1-2), 55.

V.1 CRABPII mutants coupled with Julolidine Retinal Analog III-12

Engineered CRABPII mutants, displaying a pH-dependent color change, inspired the design of a genetically encoded pH-sensor. In this system the equilibrium between colored and colorless states is controlled by the retinylidene-PSB deprotonation. This phenomenon led to the development of a novel colorimetric pH-sensitive system, based on the CRABPII/chromophore complex. Furthermore, the proper choice of two mutants allowed for the development of a concentration-independent system, based on the ratio of absorbances.[1] However, fluorescence-based pH-probes find more applications.

Fluorescence is one of the most sensitive detection methods. Although retinal fluorescence was exploited in imaging of a single HEK293 cell, it is far from being a universal tool in confocal microscopy.[2-3] Its limited application is attributed to its dim emission as a result of its low quantum yield. If one wants to design a pH-sensor for broad use, a strong emission signal should be one of the primary considerations.

Considering the latter, we decided to switch from the chromophoric ligand, retinal, to potentially fluorescent retinal analogs. As described in Chapter IV, our attempt to use MCRA was not successful as a pH-sensitive chromophore due to its high pK_a. Julolidine retinal analog (JRA) **III-12** (synthesis described in Chapter III) was hypothesized to suit this purpose with utility in pH-sensing application

because of weaker push-pull nature due to the nitrogen conjugation through the aromatic ring.

The spectroscopic characterization of JRA and its imine/iminium forms are summarized in **Figure V-1**. The chromophore absorbs at 446 nm in ethanol. After SB formation with butylamine is completed the peak blue-shifts to 408 nm. Acidification with HCl gives JRA-PSB with λ_{max} at 570 nm. The shift of 162 nm in absorption wavelength is more significant for the protonation event, as compared to the 75 nm shift observed with retinal. This complex has a Stokes shift of 95 nm (λ_{max} = 665 nm) in ethanolic solution (excitation at 565 nm). A quantum yield of 0.3% was determined for JRA-PSB in ethanol (excited at 565
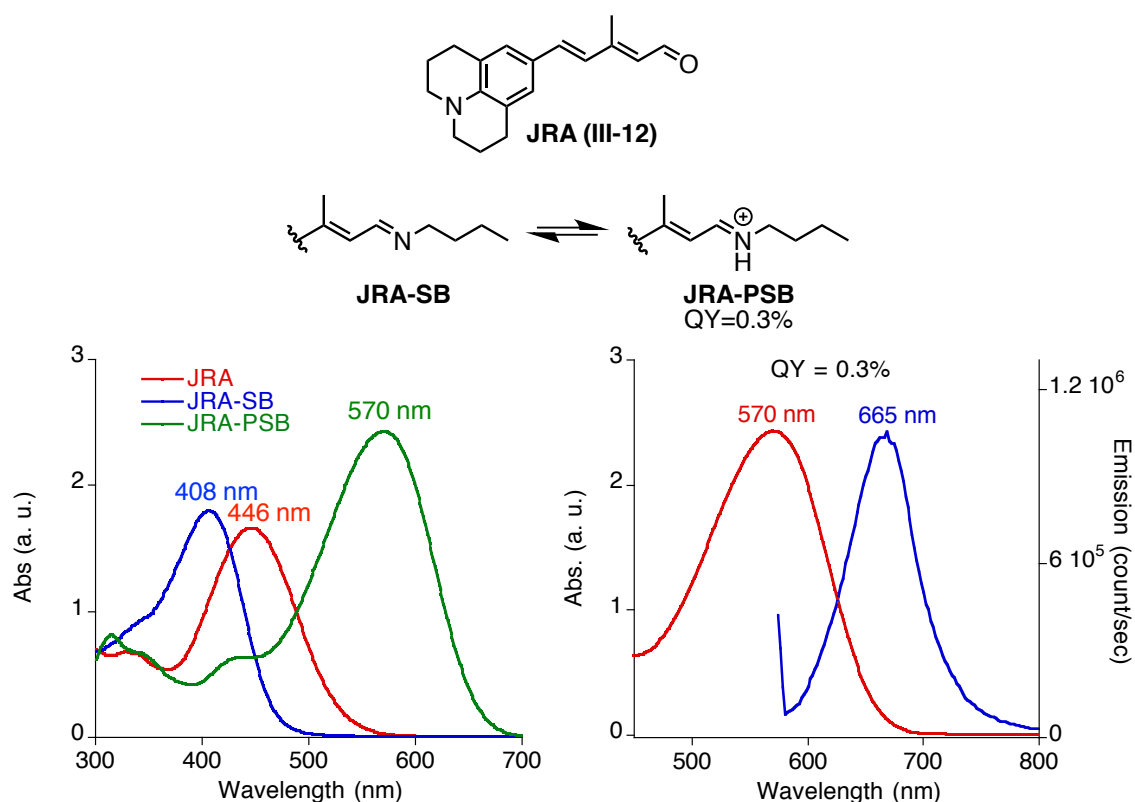
Figure V-1: Chemical structure and spectroscopic properties of julolidine retinal analog (JRA) in ethanol.

nm, using Oxazine 1 as a standard). Low quantum efficiency of the iminium with butylamine should lead to a low background fluorescence when incubated with cells labeling reactions.

Among CRABPII mutants in the library, the first one tested was R111K:R132L:Y134F:T54V:R59Y:A32W:L121D hepta-mutant because it had already been cloned into the vector for eukaryotic cell expression. Additionally, this mutant has one of the lowest pK_a values for the PSB with all-*trans*-retinal and MCRA, suggesting that the imine may not be protonated at neutral pH.
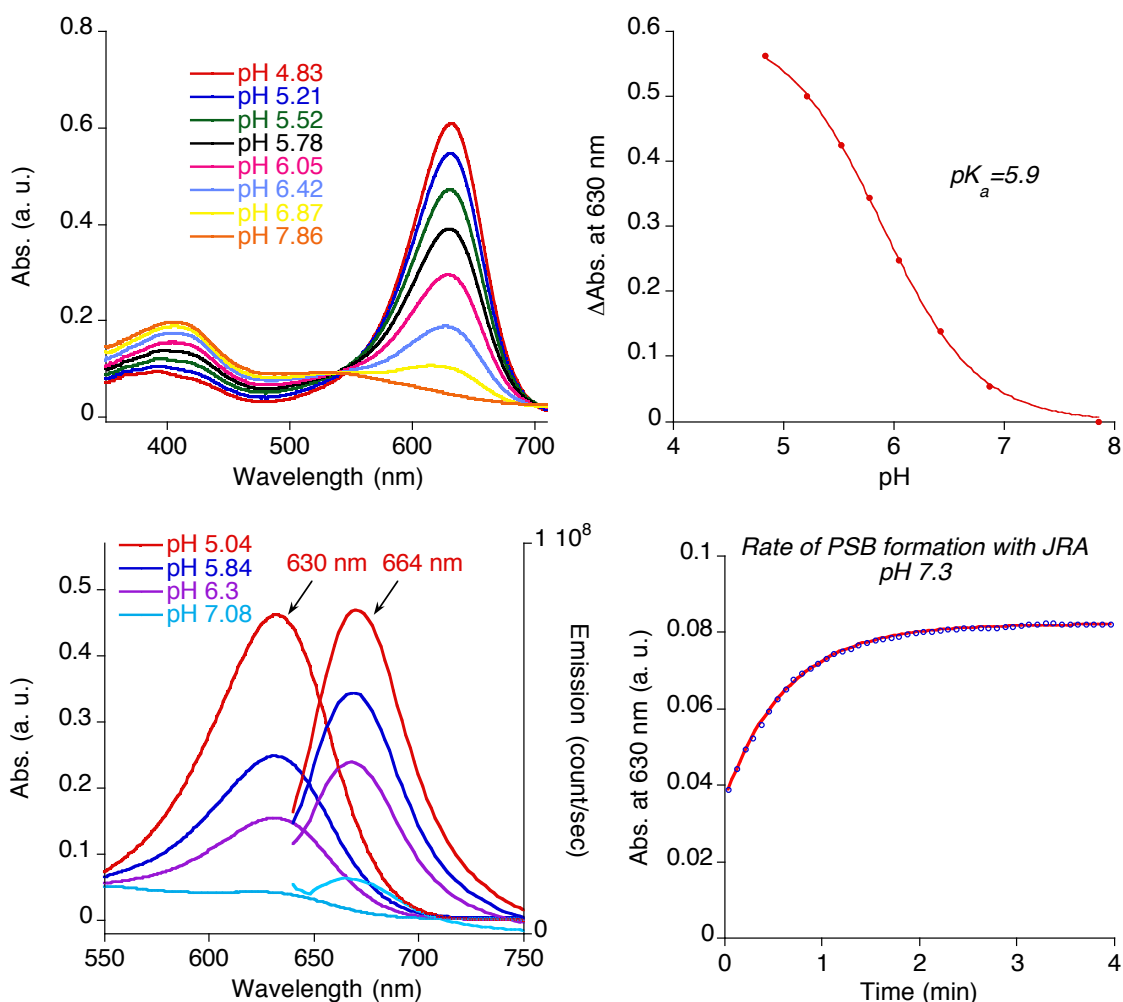
Figure V-2: Spectroscopic characterization of complex between JRA and CRABPII-R111K:R132L:Y134F:T54V:R59Y:A32W:L121D.

JRA and hepta-mutant complexation led to only a small amount of PSB, with the dominant peak corresponding to the SB. An acid-base titration of this mutant bound with JRA gave an iminium pK_a of 5.9 (**Figure V-2**). The most prominent response during the titration was found to occur in the physiologically relevant pH range, 5.0–7.0. An increasing absorbance intensity at 630 nm, in response to acidification, was observed in the proportionally elevated emission at 664 nm. The quantum yield of this mutant was 12% (excited at 565 nm, using Oxazine 1 as a standard). PSB formation was completed within 2 min at RT in phosphate buffer, similar to the reaction of the same mutant with MCRA.

This spectroscopic data suggests that JRA is an ideal candidate for observing pH changes. Especially attractive was the 10-fold fluorescence increase in response to acidification, that is unlike any other genetically encoded fluorescent pH-probe. Very likely, screening of other mutants would provide more candidates of different wavelength, pK_a and quantum efficiencies. This would allow for a ratiometric pH-sensor design using a two-protein fusion strategy.
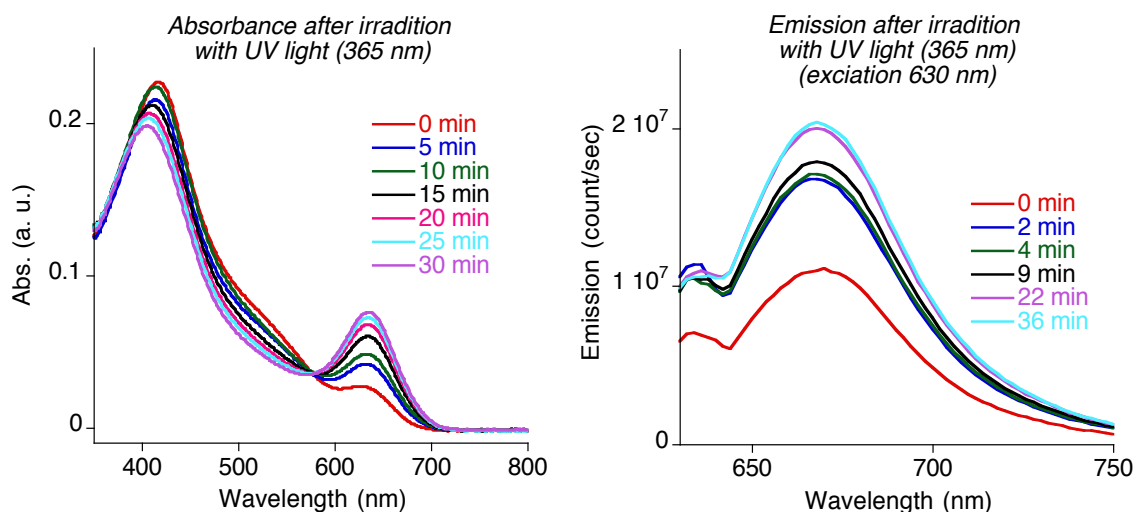
Figure V-3: Light irradiation of JRA and CRABPII complex.

However, CRABPII labeling in eukaryotic cells was found to be problematic with no observable fluorescence indicative of a successful reaction. Therefore, we turned to the hCRBPII system for optimization of the pK_a, applying the principles we learned about pK_a regulation with all-*trans*-retinal and JRA in CRABPII.

Another notable discovery was observed during analysis of the CRABPII hepta-mutant bound with JRA. Irradiation of the matured complex, with UV light of 360 ± 20 nm, increased the amount of PSB without altering the wavelength (**Figure V-3**). Notably, the wavelength of light during irradiation did not correspond to any form of the chromophore (aldehyde, SB or PSB, see **Figure V-1**). The intensity change was also reflected in fluorescence. However, the process was not reversible with red light (LP 600 nm).

V.2 Mutagenesis of hCRBPII aimed to decrease iminum pK_a

Our study in understanding the factors that control the pK_a of the PSB in CRABPII mutants bound with all-*trans*-retinal, led to two important observations. Firstly, hydrophobicity of the binding cavity is unlikely to favor the protonated state leading to the low pK_a values observed. Secondly, based on our spectroscopic data in combination with crystallographic analysis, the environment of the PSB region can be strongly influenced by the geometry of the iminium.[1]

Before applying these two principles with the hCRBPII system with a new ligand, it is essential to understand if the julolidine retinal analog behaves similar to Rt. As such, the pK_a of JRA-PSB in existing hCRBPII mutants were measured

and compared to Rt-PSB. For this purpose, we selected two hCRBPII[4] mutants

that showed the lowest pK$_a$ values with retinal:

Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4W:F16W, pK$_a$ = 6.0, and

Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4W:Q128E, pK$_a$ = 6.1 (Dr.

Wenjing Wang, Professor Babak Borhan's lab, MSU). Similarly, both complexes

with JRA showed pK$_a$ values of 6.8 and 6.4, respectively (**Table V-1**, entries 1

and 2). Another mutant, with less acidic Rt-PSB, was also chosen for pH-titration.

Table V-1. Mutants of hCRBPII designed to decrease the pK$_a$ of JRA-PSB.

Entry	hCRBPII Mutant	λ_{max}^{a} (nm)	pK$_a^{a}$
1	Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4W:F16W	665 (588)	6.8 (6.0)
2	Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4W:Q128E	660 (590)	6.4 (6.1)
3	Q108K:K40L:T51V:R58F:	646 (561)	8.8 (8.7)
4	Q108K:K40L:T51V:R58F:L117E	585	9.8
5	Q108K:K40L:T51V:R58F:L117E:Q4F	594	10.4
6	Q108K:K40L:T51V:R58F:L117E:Q4F:W106F	600	9.6
7	Q108K:K40L:T51V:R58F:L117E:Q4F:Q128E	604	9.2
8	Q108K:K40L:T51V:R58F:L117E:Q4F:Q38E	621	8.4
9	Q108K:K40L:T51V:R58F:L117E:Q4F:Q38A	624	7.8
10	Q108K:K40L:T51V:R58F:L117E:Q4F:Q38F	636	7.9
11	Q108K:K40L:T51V:R58F:L117E:Q4F:Q38A:Q128E	615	9.7
12	Q108K:K40L:T51V:R58F:L117E:Q4F:Q38A:Q128L	645	5.8
13	Q108K:K40L:T51V:R58F:L117E:Q4F:Q38F:Q128L	642	7.6
14	Q108K:K40L:T51V:T53C:R58W:T29L:A33W:Q4F:L117D	565	10.6
15	Q108K:K40L:T51V:T53C:R58W:T29L:A33W:Q4F:L117E	595	10.6

[a]The λ_{max} and the pK$_a$ values in parenthesis are for the mutant bound with all-*trans*-retinal.

The Q108K:K40L:T51V:R58F tetra-mutant complexed with retinal (pK_a of 8.7) was chosen because it displays a pK_a value similar to that of most retinal-hCRBPII mutants.[4] JRA-PSB of this tetra-mutant gave a pK_a of 8.8 (**Table V-1**, entry 3). At the same time, the wavelength regulation of these three mutants displayed the same trend as the Rt-PSB, although more red-shifted.

The spectroscopic data for these three hCRBPII mutants suggests that JRA might behave similarly to retinal in the protein pocket. Additionally, the principles learned during Rt-PSB pK_a regulation might be applicable in the case of using JRA as a ligand. We focused on the Q108K:K40L:T51V:R58F tetra-mutant as a template to study the pK_a regulation of JRA-PSB, because it demonstrated a stable pK_a with Rt-PSB over the time of experiment as opposed to the other two mutants. This eliminates the putative time-dependent conformational change of the bound ligand and provides data exclusively on the pK_a of the complex in one conformation. Moreover, if this project was extended to live-cell pH determination, this mutant would be easier to apply because it has already been cloned into the proper expression vectors.

Q108K:K40L:T51V:R58F tetra-mutant also gave a well-diffracting crystal bound with JRA (Meisam Nosrati, Professor Geiger's Lab, MSU). The location of JRA in the hCRBPII mutant was similar to retinal (**Figure V-4**). Both chromophores adopt a *cis*-iminium geometry, highlighted by water-mediated

interactions with Q4. Because JRA is significantly shorter than retinal, it is not surprising that the R58F residue rotated inside the cavity to better enclose it. Interestingly, two highly ordered water molecules, Wat1 and Wat2, found in most crystal structures of hCRBPII mutants, were absent in the JRA-hCRBPII structure (**Figure V-4**). Cyclohexane ring of julolidine occupied this space. This means that the effect of amino acid polarity projected onto JRA, but not polarized water molecules, is solely responsible for pK_a changes, because the water molecules are absent.

Using the Q108K:K40L:T51V:R58F tetra-mutant as a template, the L117E mutation was introduced (for location of L117 see **Figure V-5a** and **V-5b**). As one would expect, a carboxylate in the PSB region effectively stabilized the positive charge, increasing the pK_a to 9.8 (**Table V-1**, entry 4). Next, we wanted to alter the protonation of the iminium by removal of water-mediated interactions with Q4 (for location of Q4 see **Figure V-5**). The Q4F mutation was chosen due to high protein expression levels. Interestingly, the pK_a of the PSB in the hexa-mutant
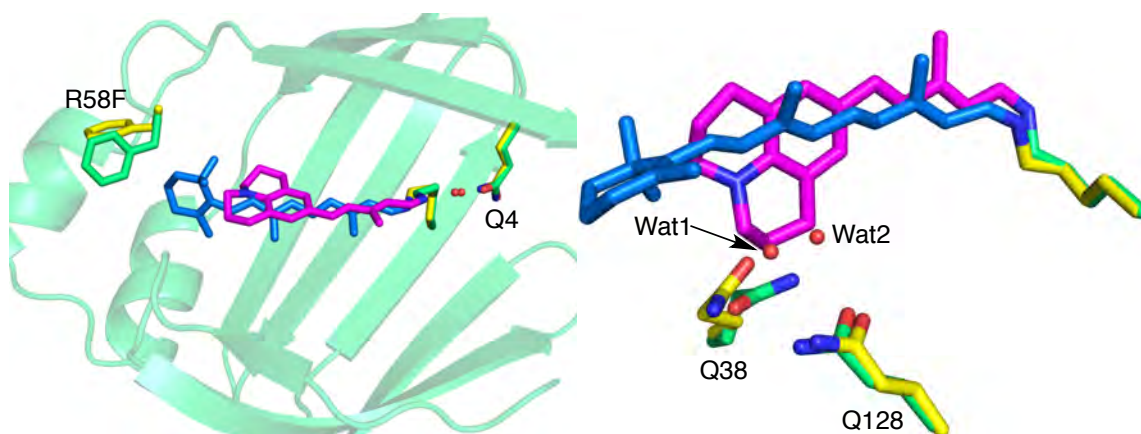
Figure V-4: Crystal structures of hCRBPII-Q108K:K40L:T51V:R58F mutant bound with JRA (magenta) and all-*trans*-retinal (blue).

299

increased to 10.4 (**Table V-1**, entry 4). It is possible that the Q4F mutation changed the iminium geometry making the interaction between PSB and L117E more favorable. Previously, we had already observed the *trans*-imine geometry in hCRBPII mutants containing Q4R, Q4A and Q4H mutations, when bound with all-*trans*-retinal (**Figure V-5b**). Therefore, it would be reasonable to expect the *trans*-iminium bond for the complex of JRA and hexa-mutant. The λ_{max} values corresponding to tetra-, penta- and hexa-mutants (**Table V-1**, entries 3-5) were in
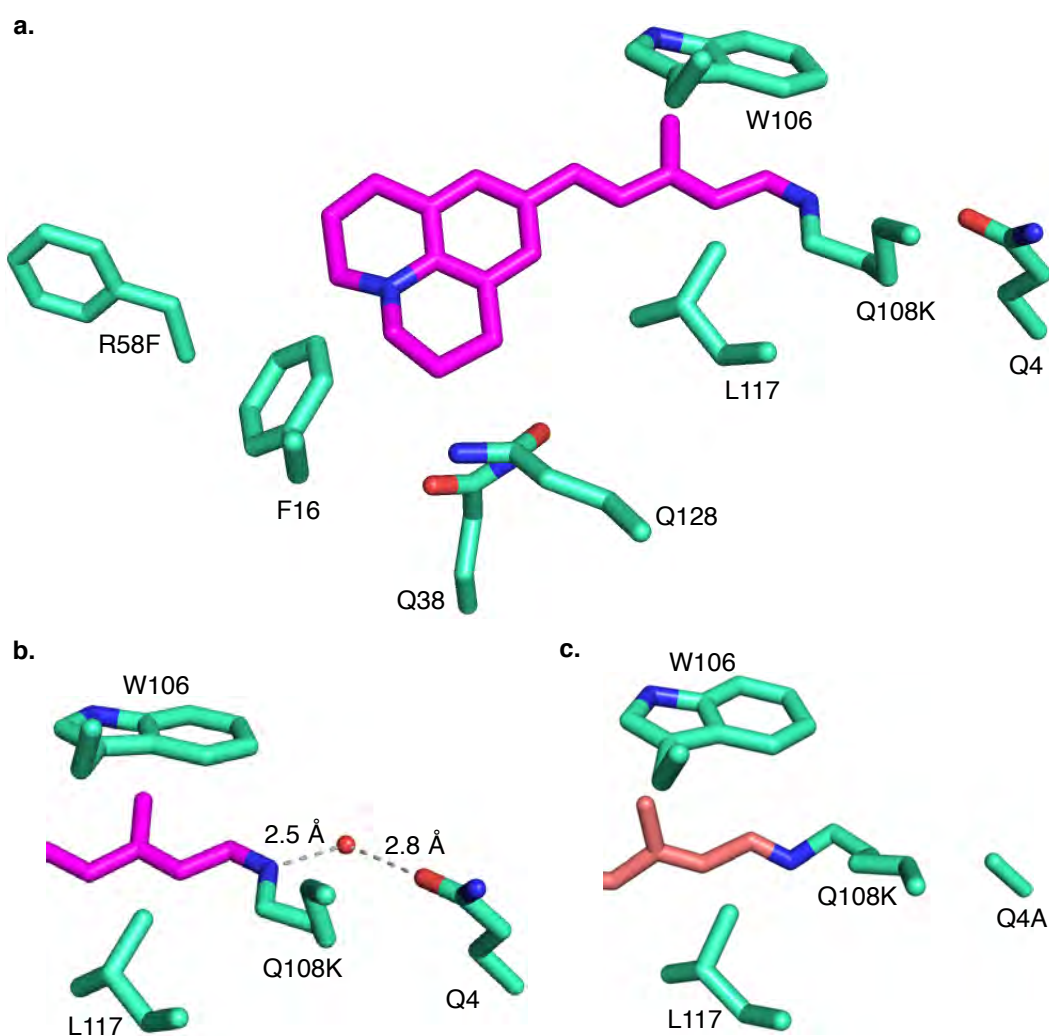
Figure V-5: a. Crystal structures of hCRBPII-Q108K:K40L:T51V:R58F mutant bound with JRA, **b.** The same crystal structure zoomed on the water-mediated interactions between Gln4 and PSB, **c.** Crystal structure of hCRBPII-Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4A mutant bound with Rt.

agreement with the wavelength regulation theory. If the positive charge is more distributed then a more red-shifted wavelength is anticipated.

Previously, based on the crystal structures, we hypothesized that W106 in hCRBPII and W109 in CRABPII could participate in π-cation interactions that lead to a higher pK_a of the complex[1, 5] (see **Figure V-5b** for location of W106 in hCRBPII relative to JRA-PSB). Removal of the polarizable tryptophan would provide a good control experiment for its impact on protonation. Mutations at W109 to Ala, Phe, Tyr, Gln and Leu in CRABPII resulted in insoluble protein expression (see Chapter II for details). However, the equivalent W106F mutation in hCRBPII was not detrimental to protein stability. Q108K:K40L:T51V:R58F:L117E:Q4F:W106F hepta-mutant was expressed in ~40 mg/L. As expected, the pK_a of this mutant decreased to 9.6 (**Table V-1**, entry 6).

After studying mutations in the iminium region, we focused on the residues near the julolidine group. As mentioned above, one of the julolidine rings occupied the space, that otherwise was filled with two water molecules trapped between Q38 and Q128. Mutations at either of these residues were detrimental for the high pK_a of Rt-PSB in earlier work (Dr. Wenjing Wang, Professor Borhan's Lab, MSU). In targeting a decreased pK_a of JRA-PSB in the hCRBPII series, two residues, Q38 and Q128, were mutated. Introduced individually, Q128E and Q38E decrease the pK_a from 10.4 (**Table V-1**, entry 5) to 9.2 and 8.4 (**Table V-1**,

entries 7 and 8), respectively. Placing hydrophobic residues Q38A and Q38F
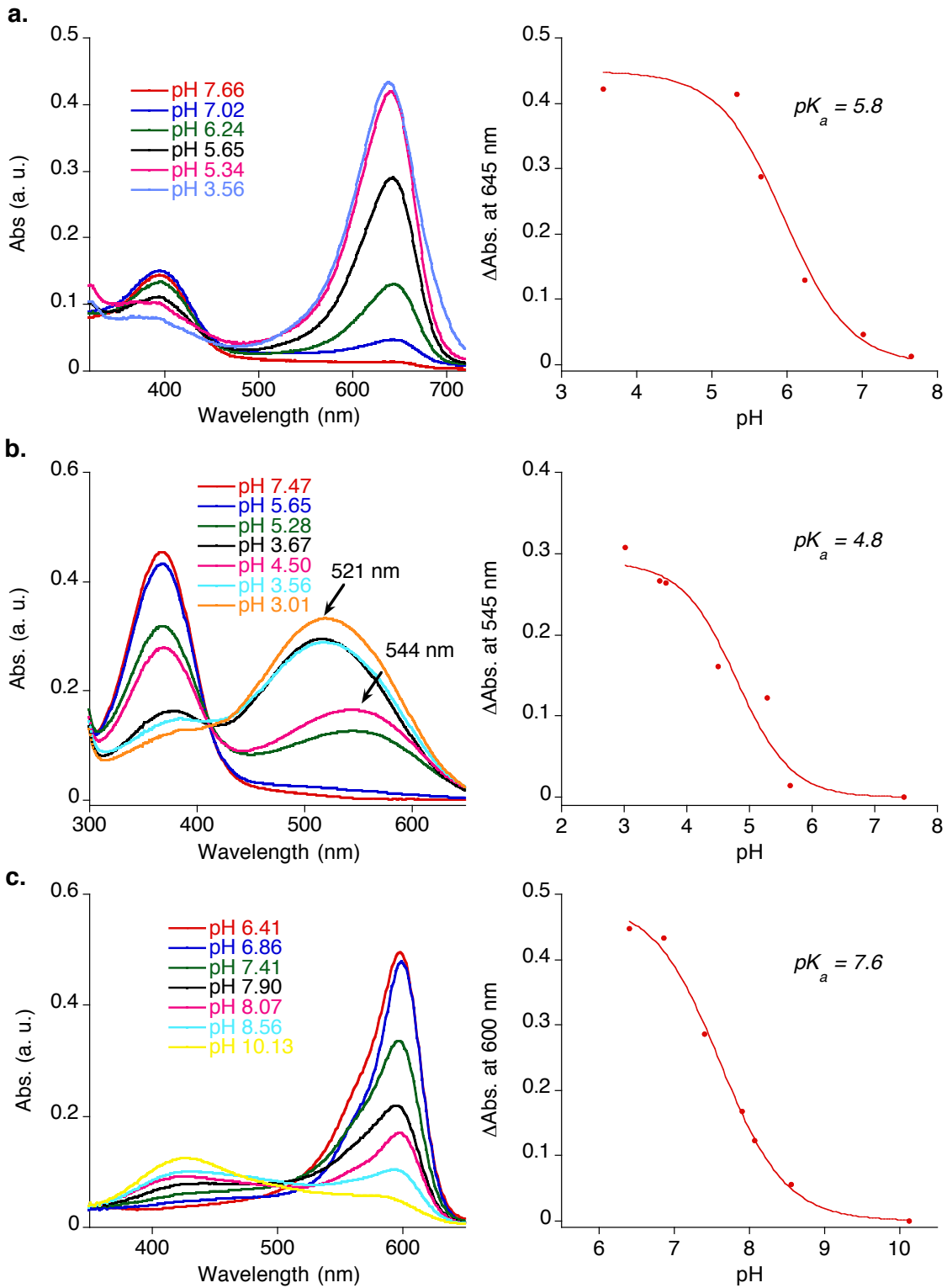
Figure V-6: pH-titrations of hCRBPII-Q108K:K40L:T51V:R58F:L117E:Q4F: Q38A:Q128L mutant bound with JRA, (**a.**), Rt (**b.**) and MCRA (**c.**).

decreased the pK_a even further to 7.8 and 7.9 (**Table V-1**, entries 9 and 10), respectively. The mutation Q128E combined with Q38A restored the pK_a to 9.7 (**Table V-1**, entry 11), but Q38A:Q128L yielded a pK_a of 5.8, the lowest pK_a for a JRA-PSB (**Figure V-6a**) among all tested mutants (**Table V-1**, entry 12). Interestingly, combination of Q128L and the large hydrophobic residue Q38F yielded a pK_a value of 7.6, almost two units higher than Q38A:Q128L. (**Table V-1**, entry 13).

One of the octa-mutants, Q108K:K40L:T51V:R58F:L117E:Q4F:Q38A: Q128L, was also used to measure the pK_a with all-*trans*-retinal and MCRA, exhibiting pK_a values of 4.8 and 7.6, respectively (**Figure V-6a** and **V-6b**). Consistent with the data for JRA-PSB, these were the lowest pK_a values among all tested hCRBPII mutants for each corresponding ligand. Interestingly, comparison of hCRBPII and CRABPII mutants displaying similar pK_a values among all three ligands allowed for the identification of three pairs of residues similar in size and polarity at equivalent positions: Q4F and F3, Q38A and P39, Q128L and R132L (for comparison in size see the model of hCRBPII with Q4F, Q38A and Q128L and the crystal structure of CRABPII-R111K:R132L:Y134F:T54V:R59W **Figure V-7**).

Additionally, the wavelength regulation between Q38 and Q128 mutants showed also a distinguishable trend. Acidic residues, Q38E and Q128E, both

triggered a red-shift from 594 nm to 621 nm and 604 nm, respectively (**Table V-1**, entries 7 and 8). Most probably, the carboxylate in proximity of the ring-locked tertiary amine can promote positive charge delocalization from the PSB. Furthermore, hydrophobic residues Q38A and Q38F gave a more significant red-shift, 624 nm and 636 nm, respectively (**Table V-1**, entries 9 and 10). Combining Q38A:Q128L and Q38F:Q128L yielded an even more pronounced bathochromic shift to ~645 nm (**Table V-1**, entries 12 and 13). The Q108K:K40L:T51V:R58F:L117E:Q4F:Q38A:Q128E octa-mutant displayed absorbance maxima at 615 nm (**Table V-1**, entry 11), with an absoption value between the hepta-mutants with only Q38A (624 nm) or Q128E (604 nm) mutations. The wavelength shift 'recipe' for JRA-PSB is consistent with Rt-PSB: an overall neutral protein environment that leads to a more effective charge distribution across the chromophore, leading to the red-shift.

Finally, we measured the pK_a of JRA-PSB in two nona-mutants that react rapidly with MCRA, Q108K:K40L:T51V:T53C:R58W:T29L:A33W:Q4F:L117D and
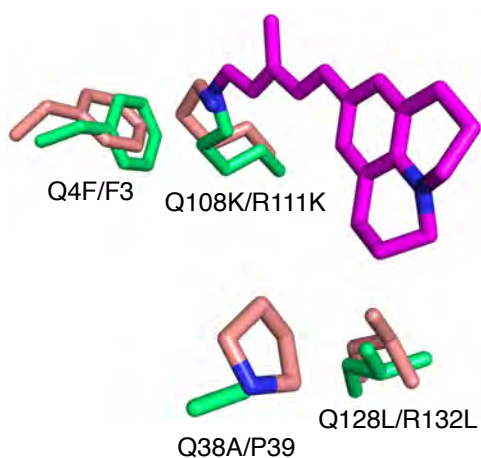
Figure V-7: Alignment of CRABPII crystal structure and hCRBPII model highlighting equivalent Q4F, Q38A and Q128L mutations (hCRBPII numbering).

Q108K:K40L:T51V:T53C:R58W:T29L:A33W:Q4F:L117E. As expected, both mutants gave a pK_a similar to that of Q108K:K40L:T51V:R58F:L117E:Q4F because of the presence of both Q4F and carboxylate at L117 (**Table V-1**, entries 14 and 15). Counter-intuitively, the shorter side chain, L117D, gave a pK_a value identical to that of L117E. The λ_{max} of the nona-mutant with L117D was the most blue-shifted among all hCRBPII mutants, surpassing the λ_{max} of JRA-PSB with butylamine by 5 nm.

V.3 Role of L117E residue as a second titratable proton

An exciting observation was made during the UV-vis measurements of JRA bound to Q108K:K40L:T51V:R58F:L117E. The absorbance spectra of the penta-mutant had two distinct λ_{max}'s at pH 5.0 and at pH 8.0; the shift was observed during the titration from pH 5.0 to pH 8.0 (**Figure V-8**). However, the tetra-mutant without the L117E residue, Q108K:K40L:T51V:R58F, displayed a
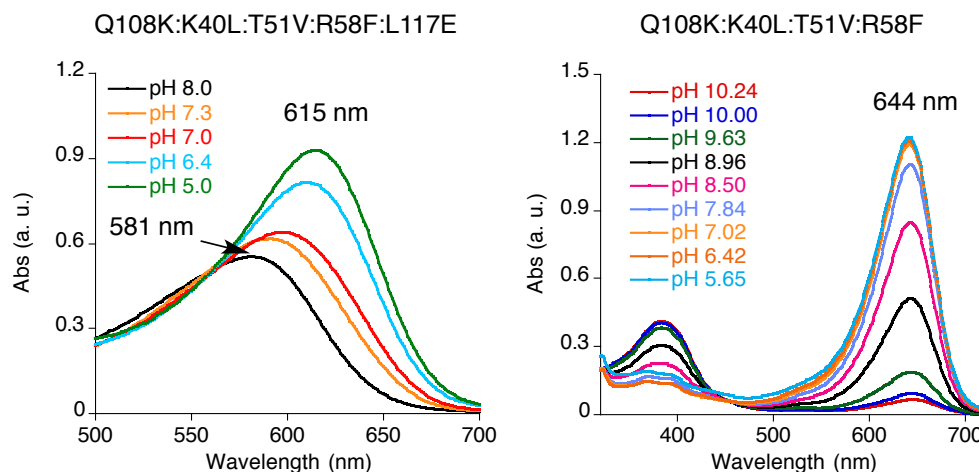
Figure V-8: Response to the pH changes in the λ_{max} of two JRA-bound proteins different only in L117E mutation.

single λ_{max} at 644 nm in titration from pH 5.6 to pH 10.2 (**Figure V-8**). Hence, the origin of the pH-dependent wavelength shift was hypothesized to arise because of the L117E carboxylate residue and its protonation state.

Interestingly, the L117 residue is located 4.4 Å away from the double bond in the iminium region, as determined by the crystal structure of the tetra-mutant (**Figure V-9a**). Having a carboxylate in the iminium region of the penta-mutant localizes the positive charge on the PSB, leading to a blue-shifted complex with λ_{max} at 581 nm at pH 7.3. This is a 63 nm blue-shift in comparison to Q108K:K40L:T51V:R58F tetra-mutant at the same pH (**Figure V-8**). Acidification of the penta-mutant JRA-PSB results in the protonation of the carboxylate at residue L117E, which leads to the less stabilized positive charge in the iminium region (**Figure V-9b**). The latter phenomenon affects the wavelength of the protein-chromophore complex by red-shifting the λ_{max} to 615 nm. This is in
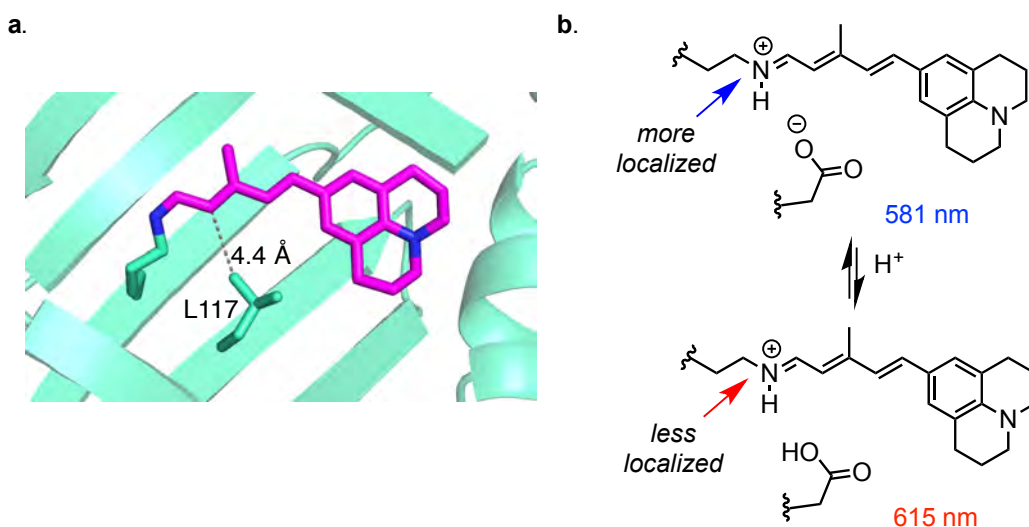
Figure V-9: a. Crystal structure of Q108K:K40L:T51V:R58F tetra-mutant highlighting the L117 residue and its distance to JRA; **b.** Putative mechanism of the red-shift in response to protonation of the nearby amino acid.

agreement with the mechanism that was previously proposed for the wavelength regulation of retinylidene: a more neutral electrostatic environment around the chromophore provides a more red-shifted Rt-PSB.

The pH-induced wavelength shift inspired the design of a novel pH-sensor. A single protein/chromophore complex displayed two λ_{max} values more than 30 nm apart, dependent on the pH. This also allowed for the excitation at two distinct wavelengths. The emission spectra at two excitation wavelengths, at varying pH, can be integrated to develop a ratiometric curve as a function of pH. If the titration curve is indeed concentration independent, this approach will be advantageous over others because both pH-dependent forms correspond to the same chromophore, JRA-PSB. An intrinsically high pK_a value of the iminium is key to avoiding loss of protonation and thus its color, while the acidic residue is titrated. This approach is highly advantageous over currently used genetically encoded pH-sensors because of the small size of the ratiometric probe (133 amino acids) and strong fluorescence of both pH-dependent forms.

V.4 Single protein ratiometric pH probes: *in vitro* characterization

Design of a single protein ratiometric pH-sensor involves the introduction of a second titratable proton in the form of an ionizable amino acid in the proximity of JRA-PSB. The protonation/deprotonation event of this titratable residue can influence the electrostatic environment projected onto the chromophore, thus, promoting the pH-dependent wavelength shift. While a carboxylate in the PSB region displayed this phenomena, we should not be

limited to only altering residues in this chromophore zone. Based on our previous work with all-*trans*-retinal as a ligand, any zone along the chromophore can lead to a wavelength shift, induced by the projected electrostatic potential onto the chromophore.

Taking this into account, it was hypothesized that placing amino acids with either a positively charged protonated side chain or a negatively charged deprotonated residue could provide pH-dependent wavelength regulation. In such a scenario, we can consider placing two charged residues simultaneously at the ends of the chromophore. The proper combination of multiple charged residues

Table V-2. hCRBPII mutants that show a pH-dependent wavelength shift with JRA.

Entry	hCRBPII Mutant[a]	λ_{max}[b] (nm)	λ_{max}[c] (nm)	pH[d]	E.R.[e]
1	KLV:R58F:L117E	581	615	7-9	4.05
2	KLV:R58F:L117E:Q4F	594	630	6-8	3.07
3	KLV:R58F:L117E:Q4F:W106F	598	644	4.5-7	1.89
4	KLV:R58F:L117E:Q4F:E72W	618	652	n.r.[f]	n.r.
5	KLV:R58F:L117E:Q4F:W106F:E72W	599	639	3-8	1.80
6	KLV:R58F:L117E:Q4F:Q128E	594	636	5-8	1.97
7	KLV:R58F:L117E:Q4F:Q38E	619	651	4-6	1.16
8	KLV:R58F:L117E:Q4F:Q38A	584	624	7-9	n.d.[g]
9	KLV:R58F:L117E:Q4F:Q38F	635	643	n.r.	n.r.
10	KLV:R58F:L117E:Q4F:Q38A:Q128E	603	633	6-8	1.40
11	KLV:R58F	644	644	n.r.	n.r.
12	KLV:T53C:R58W:T29L:A33W:Q4F:L117D	555	630	6-8	5.56
13	KLV:T53C:R58W:T29L:A33W:Q4F:L117E	594	630	6-8	2.12
14	KLV:R58F:L117E:Q4A	595	630	5-10	2.89

[a]KLV, Q108K:K40L:T51V. [b]λ_{max} of absorbance at basic pH. [c]λ_{max} of absorbance at acidic pH. [d]pH range of different ratios of emission, excited at 594 nm and at 633 nm. [e]The highest ratio of two collected emission spectra achieved with excitation wavelengths at 594 nm and at 633 nm. [f]n.r., non-ratiometric emission during pH titration. [g]n.d., not determined.

along the JRA-PSB might allow for a more substantial change in the UV-vis between the two forms, yielding better spectral resolution. Alternatively, it can also provide pH probes of various working ranges based on the pK_a value of the side chain.

Searching for other hCRBPII mutants displaying a pH-dependent wavelength shift, we focused on using a secondary titratable proton in the form of a carboxylate near the chromophore. One of the main criteria was a relatively high pK_a value of the iminium. Screened mutants are summarized in **Table V-2**.

On average, the wavelength shift was around 30 nm between acidic and basic solutions of the protein-JRA complex. A control experiment using the mutant without the carboxylate residue showed a constant UV-vis spectrum over the entire titrated range (**Table V-2**, entry 11). The W106F mutation surprisingly enhanced the wavelength shift between the two forms, from 36 nm to 46 nm (**Table V-2**, compare entries 2 and 3) and from 34 nm to 40 nm (**Table V-2**, compare entries 4 and 5). One of the mutants with two carboxylate residues at opposite ends of the protein-bound chromophore, Q108K:K40L:T51V:R58F:L117E:Q4F:Q128E displayed an enhanced wavelength shift during pH-titration of 42 nm (**Table V-2**, entry 6). However, two of the tested mutants of this kind displayed a shift similar to that with the single carboxylate, 32 nm and 30 nm (**Table V-2**, entries 7 and 10). Interestingly, the more hydrophobic mutants with Q38A and Q38F gave opposite results (**Table V-2**, entries 8 and 9). When the first mutant gave an enhanced shift of 40 nm, the second mutant

shifted only 8 nm over the entire pH range. The most interesting result was obtained for the protein containing a L117D mutation rather than L117E: the λ_{max} was shifted 75 nm during the titration from pH 5 to pH 8, as opposed to 36 nm for L117E (**Table V-2**, entries 12 and 13).

Having in hand a set of mutants capable of a pH-dependent wavelength shift, we compared the fluorescence ratios of the two forms. We used two
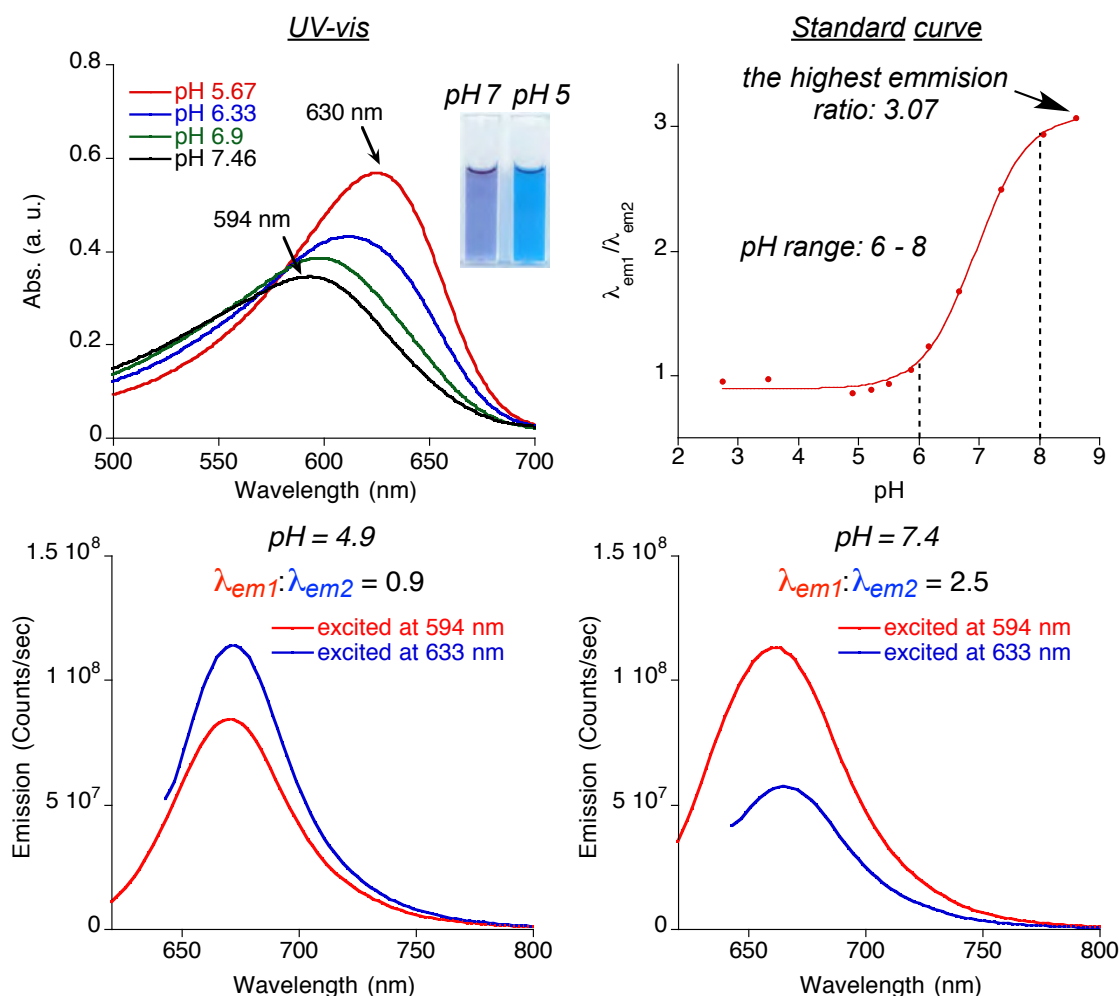

Figure V-10: Ratiometric pH response of the Q108K:K40L:T51V:R58F: L117E:Q4F hexa-mutant. During acidification fluorescence intensity of λ_{em1} decreases but λ_{em2} increases. The ratio of two integrated emissions as a function of pH gives sigmoidal pK_a curve with a working pH range 6–8.

excitation wavelengths that covered a common range between all tested complexes and that was compatible with settings available on the laser scanning confocal microscope aiming, for future application *in vivo*.

The typical spectroscopic behavior of these mutants is summarized in **Figure V-10**. Emission at each pH point was recorded after excitation at λ_{ex1}=594 nm (λ_{em1}=600–800 nm) and at λ_{ex2}=633 nm (λ_{em2}=640–800 nm). The emission spectra were integrated and the ratio, $\lambda_{em1}/\lambda_{em2}$, was plotted as a function of pH. Interestingly, pH-titration influenced only the absorption, but not the emission λ_{max}. Gratifyingly, most of the tested mutants had a standard curve with a sigmoidal shape, similar to a typical pK$_a$ curve (**Table V-2**, mutants with 'n.r.' did not provide sigmoidal curves).

After evaluating the ratiometric nature of the fluorescent complexes, we looked at the difference in the working pH ranges among the tested hCRBPII
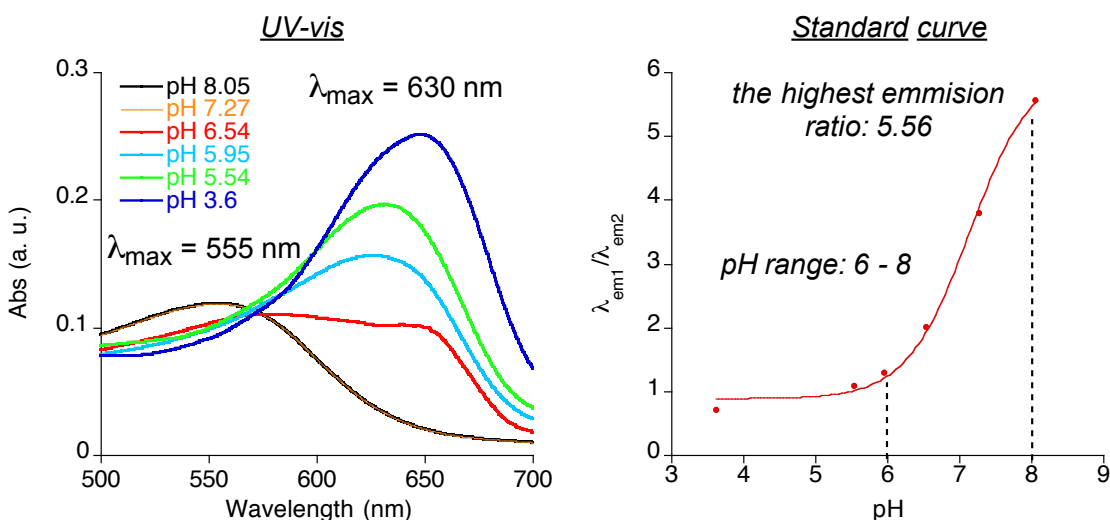
Figure V-11: UV-vis spectra during pH titration of the Q108K:K40L:T51V:T53C:R58W:T29L:A33W:L117D:Q4F nona-mutant and ratiometric emission based standard curve.

mutants (**Table V-2**, column pH). An example of a pH-range determination is highlighted on the sigmoidal curve in **Figure V-10** (dashed lines). While all the mutants showed a shift in absorption during titration, the working pH ranges were different. We were able to obtain the mutants that work at acidic pH values (**Table V-2**, entries 3, 5 and 7), neutral pH values (**Table V-2**, entries 2, 6, 10, 12 and 13), and basic pH values (**Table V-2**, entries 1 and 8).

Besides having a working range, a good pH probe should exhibit a high ratio value to develop a reliable standard curve for a more accurate pH determination. As one would expect, the mutant with the highest wavelength shift, 75 nm, gave the highest ratio of the two fluorescent forms, 5.56, because of the pronounced separation between the two excitation peaks (**Table V-2**, entry 12 and **Figure V-11**). Additionally, a relatively high ratio was obtained for Q108K:K40L:T51V:R58F:L117E and Q108K:K40L:T51V:R58F:L117E:Q4F mutants bound with JRA, 4.05 and 3.07, respectively (**Table V-2**, entries 1 and 2).
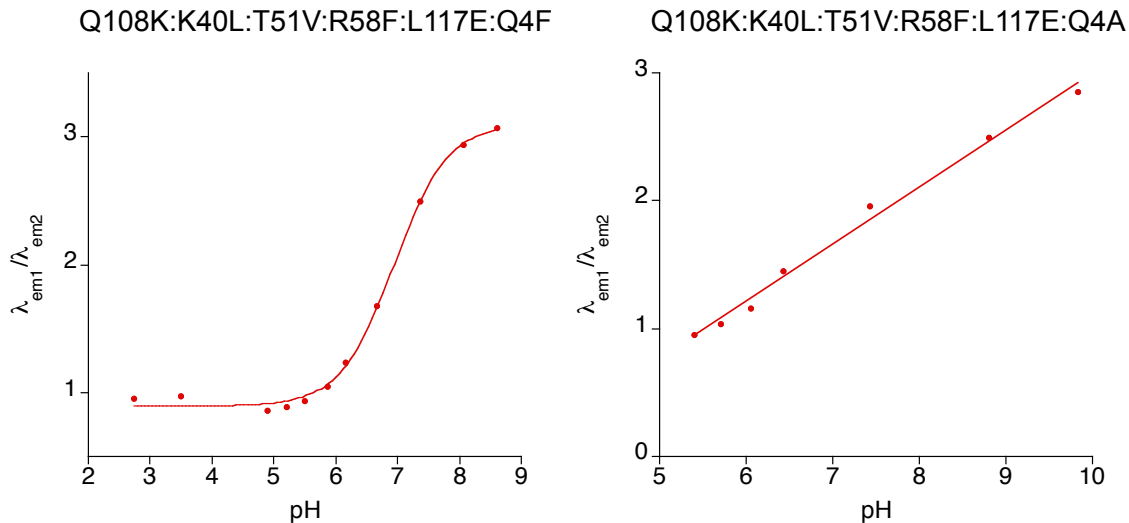
Figure V-12: Standard curves of the highlighted hCRBPII mutants bound with JRA.

These results make these mutants the top candidates for *in vivo* pH determination.

It is important to mention that none of the hCRBPII proteins with the Q4F mutation gave quality crystals with any of the ligands. However, the Q4F mutation does provide excitation maxima more compatible with confocal experiments. Because the Q108K:K40L:T51V:T53C: R58W:T29L:Y19W:Q4A mutant had already been crystallized with all-*trans*-retinal, we prepared and evaluated a pH-sensor candidate containing the Q4A mutation. Mutant Q108K:K40L:T51V:R58F:L117E:Q4A had almost the same UV-vis spectra as Q108K:K40L:T51V:R58F:L117E:Q4F (**Table V-2**, entries 2 and 14). However, ratiometric fluorescence measurements gave a line rather than a sigmoidal curve, over a pH range 5 to 9 (**Figure V-12**). While this mutant showed the expected pH dependent wavelength shift and was expected to yield good crystals, crystals of sufficient quality for analysis were not obtained (Camille Watson, Professor Geiger's Lab, MSU).

V.5 Fluorescence of hCRBPII complexes with JRA

The quantum efficiency of the emission-based probe dictates the brightness of the "glowing" complex. Based on our work with MCRA as a ligand, the cyanine dye formed with the protein can have a high quantum yield with emission in the red region of the spectrum. JRA-PSB can also be considered a type of cyanine dye, with a resonating positive charge between the two nitrogen lone pairs. However, JRA-PSB requires disruption of aromaticity in order to

313

resonate, providing a significant difference between the spectroscopic properties of MCRA and JRA after PSB formation.

For example, the wavelength regulation of MCRA-PSB did not surpass 40

Table V-3. Quantum yields of hCRBPII mutants bound with JRA.

| Entry	hCRBPII Mutant	λ_{max}[a] (nm)	λ_{em}[b] (nm)	Φ[c] (%)
1	KLV:T53C:R58W:T29L:A33W:Q4F:L117D	565	660	18
2	KLV:R58F:L117E	581	657	14
3	KLV:R58F:L117E:Q4F	594	664	10
4	KLV:T53C:R58W:T29L:A33W:Q4F:L117E	594	666	10
5	KLV:R58F:L117E:Q4A	595	665	7
6	KLV:R58F	644	675	6

[a]λ_{max} of absorbance in the PBS buffer, pH=7.3. [b]λ_{max} of emission at the same pH. [c]Quantum yields calculated using 565 nm excitation based and Oxazine 1 as a standard.

nm in all three series of PSB-forming mutants (two CRABPII binding pockets and hCRBPII). But JRA-PSB showed a more significant wavelength regulation, over 100 nm, among the tested hCRBPII mutants (**Figure V-13**). The absorbance range of JRA-PSB was smaller, 105 nm, in comparison to retinal bound proteins, 221 nm, suggesting that the lone pair of the tertiary amine was participating in resonance, but not as efficiently as in MCRA. In any case, the ability to regulate the wavelength of JRA-PSB in hCRBPII mutants permits the observation of pH-dependent wavelength shift.

Although absorption was regulated in the range of 555-665 nm, emission maxima spanned from 657 to 675 nm (**Figure V-13**). Even the direction of the shift was not consistent between absorbance and fluorescence. In some cases, a

red-shifted absorbance corresponded to a blue-shifted fluorescence, and vice versa (**Table V-3**, entries 1 and 2).

A similar phenomenon was previously observed with the Rt-PSB in hCRBPII mutants,[4] as well as with the photoactive yellow protein.[6] The latter study suggested that the width of the excited state surface is different between mutants. This influences the probability of transition between vibrational energy levels.[4, 6]
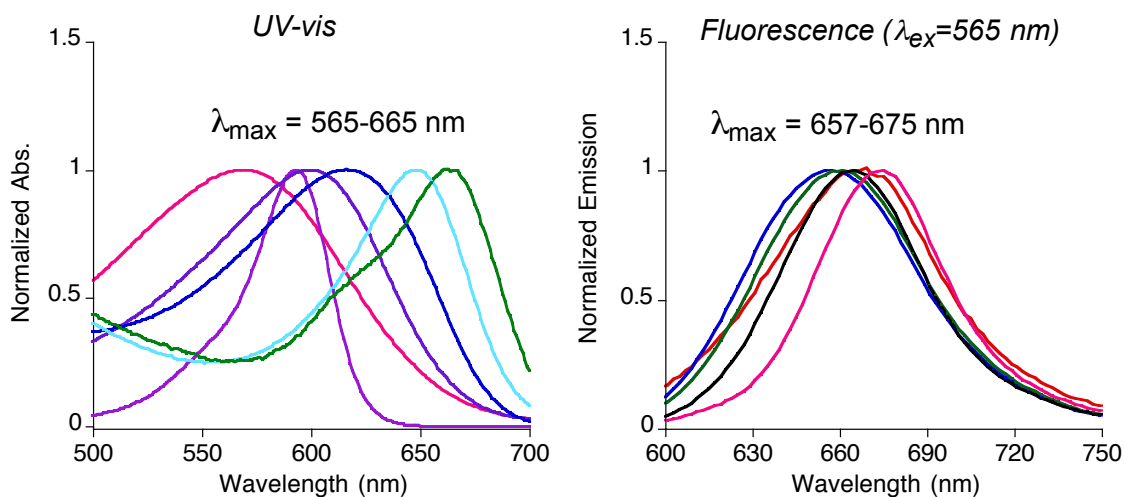
Figure V-13: UV-vis and fluorescence spectra of hCRBPII mutants bound with JRA selected from **Tables V-1–V-3**. All spectra were acquired in PBS buffer, pH=7.3.

Interestingly, the quantum yields of various mutants showed a diminished quantum efficiency with an increased λ_{max} of absorption. With that said, the most blue-shifted mutant had the highest quantum yield, 18% (**Table V-3**, entry 1). Alternatively, the lowest quantum yield, 6%, was found for the most red-shifted complex (λ_{max} = 644 nm, **Table V-3**, entry 6). Determination of quantum yields with even more red-shifted mutants gave inconsistent results, probably due to the

absorbance value being almost 0 at 565 nm. However, both mutants showed a significantly increased quantum yield as compared to JRA-PSB with butylamine, 0.3% (**Figure V-1**).

V.6 Attempted *in vivo* imaging of hCRBPII mutants with JRA

Based on the collected *in vitro* data, Q108K:K40L:T51V:T53C:R58W: T29L:A33W:Q4F:L117D nona-mutant would be an ideal candidate to evaluate for *in vivo* pH-sensing. Mutation L117D was introduced into the existing plasmid (pCMV2-FLAG without localization) with Q108K:K40L:T51V:T53C:R58W:T29L:A33W:Q4F:L117E nona-mutant for live-cell imaging in the cytosol. However, HeLa cells transfected with the plasmid containing the latter hCRBPII mutant did not show any cytosolic distribution of the fluorescent JRA-PSB. Imaging was performed with 543 nm, 594 nm and 633 nm lasers and emission was collected using a long pass filters of 615 nm and 635 nm.
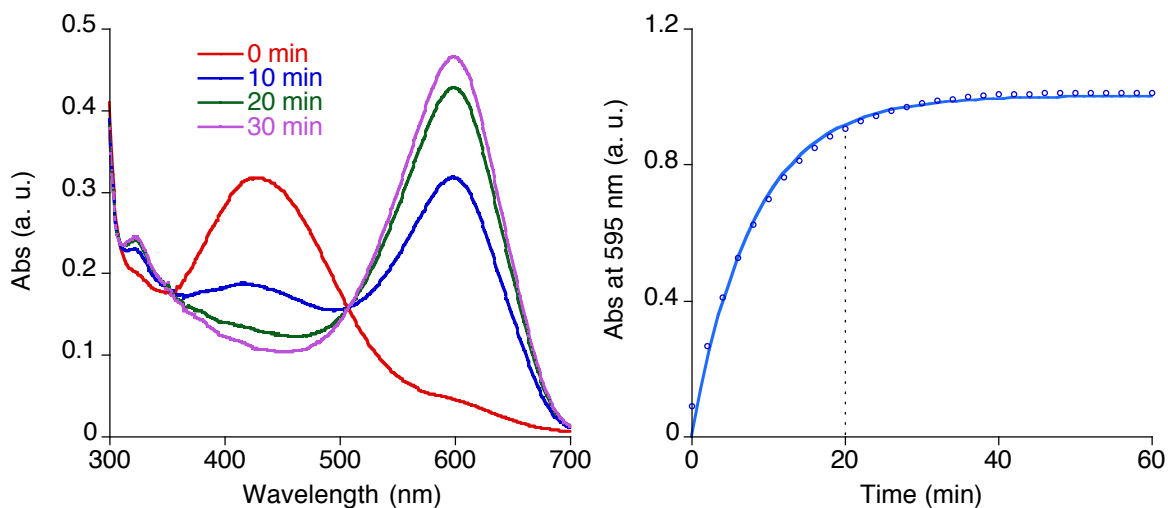
Figure V-14: Kinetic data of CRBP$_{nona}$ binding with JRA in PBS buffer, pH = 7.3.

Subsequently, Q108K:K40L:T51V:T53C:R58W:T29L:A33W:Q4F:L117E mutant (CRBP$_{nona}$) was used, as it had already worked with MCRA. Additionally, JRA with this mutant provided good characteristics for a potential pH-sensor: good quantum yield for JRA-PSB, pH-dependent wavelength shift (two excitation wavelengths) and matching available confocal lasers. The maximum ratiometric
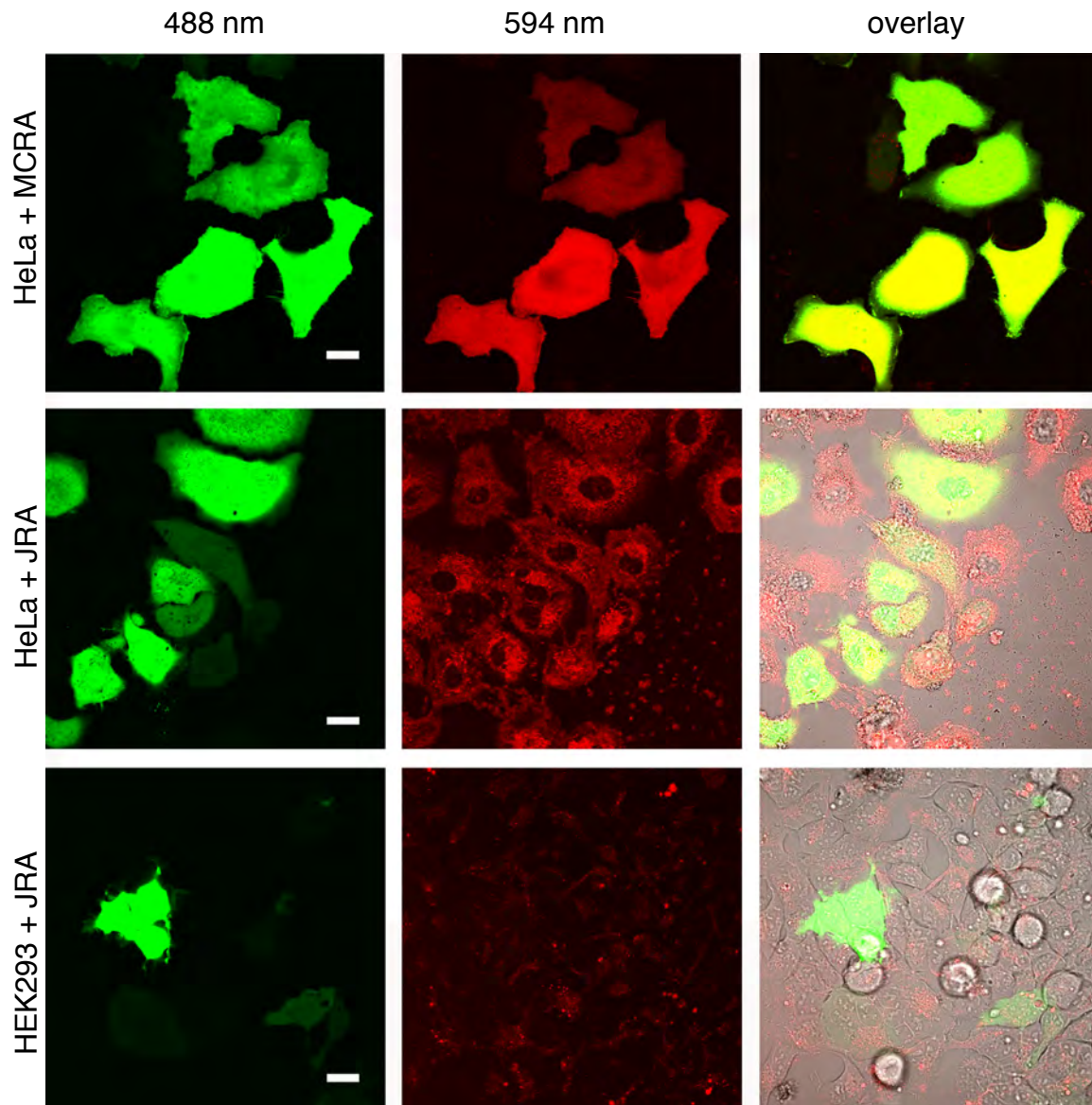
Figure V-15: Confocal images of EGFP-CRBPnona fusion labeled with MCRA and JRA. The last column is an overlay of first two columns and DIC. Watch the images from left to right. Scale bar, 20 μm.

response was 2.12, which is good enough for initial testing.

One potential problem in imaging JRA-PSB bound protein is its emission wavelength, ~665 nm; this is at the edge of the visible spectrum and the human eye may have difficulties observing its fluorescence, especially if the brightness is low. For easy determination of the fluorescence area, we prepared the fusion
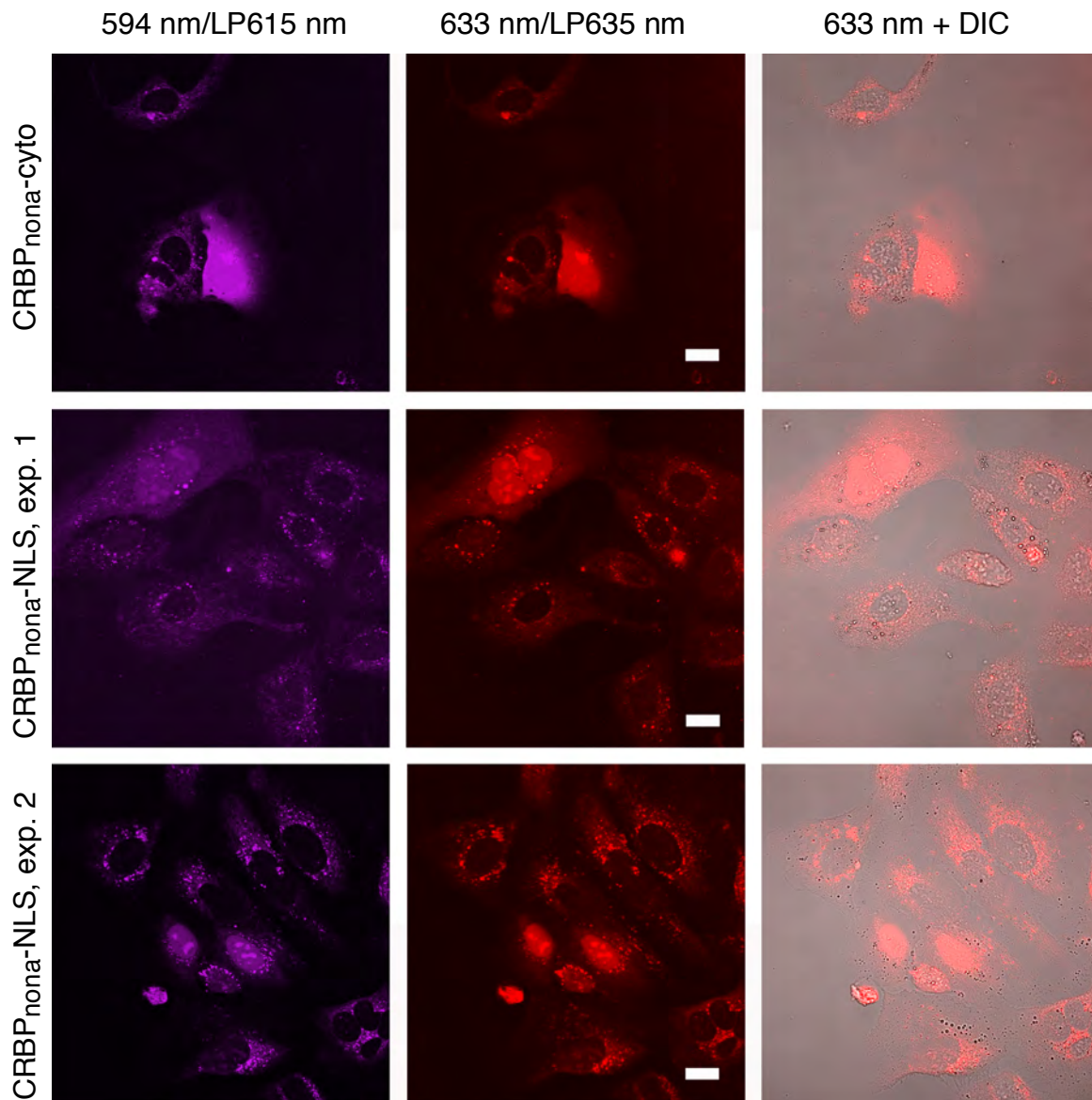
Figure V-16: Confocal images of CRBP$_{nona}$ in the cytosol and in the nucleus of U2OS cells labeled with JRA, 10 μM. Experiment in the nucleus was repeated twice, indicated as "exp. 1" and "exp. 2". For details see Main text and Materials and Methods. Follow images from left to right. Scale bar, 20 μm.

construct of EGFP-CRBP$_{nona}$ in the pFLAG-CMV2 vector. Screening as a positive control, MCRA was used as a control fluorophore. The same construct was evaluated for its reactivity with JRA at a chromophore concentration of 10 μM. The incubation time was 20 min to match the JRA binding time to this mutant (**Figure V-14**). Unfortunately, only EGFP emission was observed, while JRA fluorescence was not visible (**Figure V-15**). In a parallel control experiment employing MCRA, the cytosol of transfected HeLa cells was highly fluorescent in both imaging tracks corresponding to EGFP and MCRA-PSB. An experiment with HEK293 cells instead of HeLa showed no sign of the evenly distributed specific red fluorescence after JRA incubation (**Figure V-15**).

Changing the cell line to U2OS showed potential in solving the non-specific binding problem. At a high concentration of JRA (10 μM) the CRBP$_{nona}$ and CRBP$_{nona}$-NLS constructs showed the expected signal (**Figure V-15**). The cells with evenly distributed fluorescence in the cytosol and nucleus were clearly different from the non-transfected cells. The experiment proved reproducible under the same conditions (**Figure V-16**).

It is important to note that the observed fluorescence was invisible when viewed through the oculars of the microscope. Selected spots were successfully found only through the PMT-detector with the laser power set to 30% (MCRA requires 10% power of the same laser). The dim fluorescence of JRA-PSB in CRBP$_{nona}$ (ε = 46,000 M^{-1} $\times$ cm^{-1}, QY = 10 %) relative to MCRA-PSB (ε =
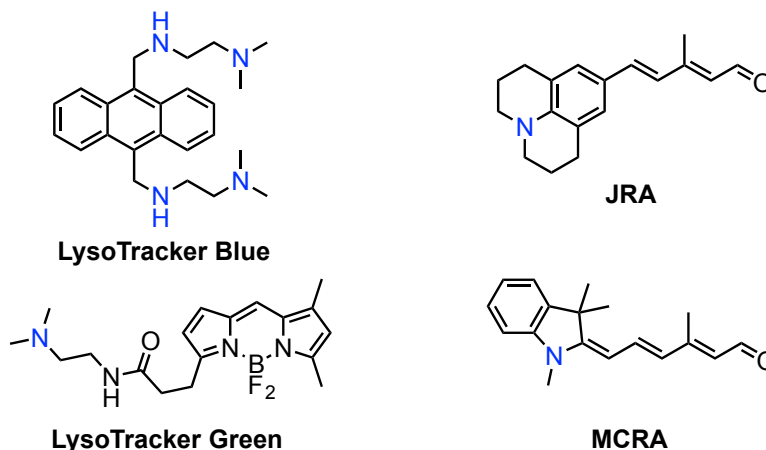
Figure V-17: Chemical structures of the known lysosomal fluorophores and their similarity with designed ligands. Tertiary amines are highlighted blue.

$160{,}000 \ M^{-1} \times cm^{-1}$, QY = 37 %) can explain the reason that successful imaging required increased laser power.

Due to the low signal-to-noise shown in **Figure V-16** (for imaging attributed to the non-specific reaction of the JRA) additional optimization was essential. Various adjustments to the initial protocol did not yield optimal imaging conditions. The following is an extremely brief summary of the optimization experiments undertaken. Decreasing the amount of the chromophore (5 μM, 2.5 μM and 1.25 μM) gave only non-specific fluorescence when imaged immediately. No fluorescence was observed after 3 h of additional JRA-free incubation at the same concentrations. Repeating the initial conditions (10 μM), but with additional JRA-free degradation time (1–5 h) resulted in a loss of specifically labeled cells.

From the observed confocal images, it appears that JRA localizes and undergoes PSB formation in some specific intracellular compartment. The high concentration of the chromophore (10 μM) can oversaturate the compartment,

which triggers further JRA distribution. Subsequently, if JRA escaped from the compartment into the cytosol, it may bind with $CRBP_{nona}$ and 'light-up'.

In fact, this mechanism would match the proposed pathway for LysoTracker localization into the lysosomes *via* protonation of the amine group (see **Chapter I**). Taking this into account it is hypothesized that the tertiary amine of synthetic fluorophores can act in the same manner, undergoing protonation in acidic pH (**Figure V-17**). The chromophore might penetrate the cell through endocytosis and accumulate in the lysosomal vesicles. PSB formation in these vesicles would lead to observed background fluorescence.

If the amine gets protonated inside the lysosomes, its pK_a should be around 5. The pK_a of an amine on an aromatic ring can vary between 4 and 8, depending on the substituents. Both JRA and MCRA are aromatic tertiary amines. However, MCRA showed the possibility of traveling to various locations in the cell as opposed to JRA.

V.7 Conclusions and significance

A fluorescent pH-dependent protein probe was developed using julolidine retinal analog as a fluorogenic ligand. The probe functions based on fluorophore protonation/deprotonation event with fluorescence 'turn-on' under acidic conditions. Alternatively, an approach, that avoids changes in the chromophore protonation state, was also developed. The latter works based on the excitation wavelength change due to the protonation state of a nearby acidic amino acid,

321

which changes the electrostatic potential projected on the chromophore. As a result, this probe is constantly fluorescent regardless of pH. Gratifyingly, two excitation wavelength give consistant change in emission ratio as a function of pH. This phenomenon leads to a ratiometric pH-probe.

The julolidine-based system did not function well in live-cell imaging experiments. One of the ways to overcome this problem is to adjust chromophore delivery. For instance, it might be possible to promote chromophore penetration into the cytosol through the cellular membrane. This is possible to achieve at different incubation temperatures or in the presence of additives. Additives can either permeabilize the membrane or form micelles that would facilitate the chromophore delivery. However, any of these methods would have substantial limitations in developing a general application for JRA. Most likely, design of different fluorescent ligands would be more advantageous in imaging and in pH-sensing. For instance, amine-based fluorogenic compounds should have substantially low pK_a value to avoid its protonation and undesired localization in the intracellular compartments. Alternatively, amine-based probes can be simply avoided.

V.8 Materials and methods

Designed hCRBPII mutants were obtained from the previously described plasmid in pET17b. Mutations were made according to the protocol described in Chapter II for CRABPII. Primers for mutagenesis were the following:

L117E

5'-CAAGCTGTAC<u>GAG</u>GAGCTGACC-3'

5'-GGTCAGCTC<u>CTC</u>GTACAGCTTG-3'

L117D

5'-CAAGCTGTAC<u>GAC</u>GAGCTGACC-3'

5'-GGTCAGCTC<u>GTC</u>GTACAGCTTG-3'

Q108K:W106F

5'- GAACCGCGGC<u>TTC</u>AAGAAGTGG-3'

5'- CCACTTCTT<u>GAA</u>GCCGCGGTTC-3'

Q128E

5'- CAGGTGTGCCGT<u>GAG</u>GTGTTCAAAAAG-3'

5'- CTTTTTGAACAC<u>CTC</u>ACGGCACACCTG-3'

Q128L

5'-CAGGTGTGCCGT<u>CTG</u>GTGTTCAAAAAG-3'

5'-CTTTTTGAACAC<u>CAG</u>ACGGCACACCTG-3'

K40L:Q38E

5'-CGTCTCACT<u>GAG</u>ACGCTGGTTATTG-3'

5'-CAATAACCAGCGT<u>CTC</u>AGTGAGACG-3'

K40L:Q38A

5'- CGTCTCACT<u>GCG</u>ACGCTGGTTATTG -3'

5'- CAATAACCAGCGT<u>CGC</u>AGTGAGACG -3'

K40L:Q38F

5'-CAAGCTGTAC<u>TTC</u>GAGCTGACC-3'

5'-GGTCAGCTC<u>GAA</u>GTACAGCTTG-3'

Q4A

5'-GACGAGGGAC<u>GCG</u>AATGGAACC-3'

5'-GGTTCCATT<u>CGC</u>GTCCCTCGTC-3'

Q4F

5'-GACGAGGGAC<u>TTC</u>AATGGAACC-3'

5'-GGTTCCATT<u>GAA</u>GTCCCTCGTC-3'

E72W

5'-GTAGAGTTTGAC<u>TGG</u>TACACAAAGAGC-3'

5'-GCTCTTTGTGTA<u>CCA</u>GTCAAACTCTAC-3'

UV-vis pH-titrations. The protein in PBS buffer (pH = 7.3) was incubated with 0.5 equivalents of JRA until PSB formation was completed (20 min – 5 h time specific for each mutant). The concentration of JRA was verified based on the absorbance value and extinction coefficient in ethanol (ε=30,000 $M^{-1}cm^{-1}$, Chapter III). The solution was acidified with citric acid buffer to a pH of ~5 and then titrated with 2 M NaOH. UV-vis measurements were collected at each pH point. The titration was continued until the PSB was converted to SB. Absorbance values at the λ_{max} were plotted as a function of pH. The obtained scatter plot was fitted to a sigmoidal curve for pK_a determination, using the same

equation as in Chapter II.

In the case of evaluating the pH-dependent UV-vis shift, the complex was first basified to a pH of 8-9 and titrated with HCl until the peaks started overlapping.

Fluorescence pH-titrations. The proteins with a pH-dependent UV-vis shift were evaluated for ratiometric emission using a Fluorolog-3 fluorimeter. A solution of protein-JRA complex in PBS buffer (3 mL with absorbance no more than 0.2) was titrated from pH ~8 and to pH ~4 using 1 M citric acid buffer (pH = 3.5). At each pH point spectra were collected: one excited at 594 nm and one excited at 633 nm. The first emission was collected from 600 nm to 800 nm and the second emission was collected from 640 nm to 800 nm. The final peaks were integrated and divided. The obtained ratio was plotted as a function of pH and fitted to the following equation:

$$FR_x = {FR_{max}}/{(1 + 10^{pH-pKa})} + FR_{min}$$

where *FR* is a ratio of integrated emission spectra excited at 594 nm and at 633 nm. The derivation is as follows:

Assuming the wavelength shift is purely a result of carboxylate titration, we can estimate the next equilibrium:

$$\text{-COO}^- + \text{H}^+ \rightleftharpoons \text{-COOH}$$
594 nm   633 nm
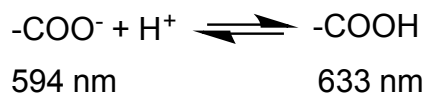
We are using the ratio of carboxylate to carboxylic acid, therefore:

$$pH = pKa + \frac{[-COO^-]}{[-COOH]}$$

325

The highest fluorescence ratio, FR_{max}, corresponds to maximum carboxylate ([-COO⁻]>>[-COOH]). When glutamate is converted to glutamic acid the ratio of emissions changes, with respect to pH, giving FR_x. At each pH point the ratio of concentrations is:

$$\frac{[-COO^-]}{[-COOH]} = \frac{FR_{max} - FR_x}{FR_x}$$

Placing this ratio into the Henderson-Hasselbalch equation yielding the following equation:

$$pH = pKa + \frac{FR_{max} - FR_x}{FR_x}$$

Solving this equation for FR_x yields:

$$FR_x = \frac{FR_{max}}{(1 + 10^{pH-pKa})}$$

The latter equation needs to be adjusted for the minimal fluorescence ratio, FR_{min}, because FR_x is never reaches 0. Therefore, FR_{min} is subtracted from FR_x.

Quantum yield measurements. The quantum yields were determined using a linear fit as described in Chapter IV.

Mammalian cell culture and transfection protocol. All cells were cultured and transfected with the gene of interest according to the protocol described in Chapter IV.

Labeling reaction. Two days after transfection the cells were incubated with 10 μM JRA, that was dissolved in supplements-free DMEM, for 20 min at 37 °C. The media was removed and the cells were washed twice with warm PBS and supplied with colorless 1640-RPMI medium. Imaging was performed within the next 20 min.

Confocal imaging. Cell microscopy was performed using inverted laser scanning confocal microscope (LSM510Meta, Carl Zeiss, Jena, Germany) equipped with diode, argon and HeNe lasers. Oil-immersed objective (40x) was used. JRA-PSB was imaged using two tracks: 1) 594 nm excitation, 594 nm primary dichroic, 545 nm secondary dichroic, >615 nm emission, and 2) 633 nm excitation, 633 nm primary dichroic, 545 nm secondary dichroic, >635 nm emission. Both lasers required 30% power in order to observe the signal. EGFP was imaged using the same parameters as in Chapter IV. All images were analyzed using Axiovert 200M software. Kalman averaging 8 was applied in all confocal images. Fluorescence in each experiment was normalized to the same intensity adjusting the gain and amplifier offset. All the images are pseudocolored.

REFERENCES

REFERENCES

1.   Berbasova, T.; Nosrati, M.; Vasileiou, C.; Wang, W.; Sing, K.; Lee, S.; Yapici, I.; Geiger, J. H.; Borhan, B., Rational Design of a Colorimetric pH Sensor from a Soluble Retinoic Acid Chaperone. *Journal of the American Chemical Society* **2013,** *135* (43), 16111.

2.   Maclaurin, D.; Venkatachalam, V.; Lee, H.; Cohen, A. E., Mechanism of voltage-sensitive fluorescence in a microbial rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110* (15), 5939.

3.   Kralj, J. M.; Hochbaum, D. R.; Douglass, A. D.; Cohen, A. E., Electrical Spiking in Escherichia coli Probed with a Fluorescent Voltage-Indicating Protein. *Science* **2011,** *333* (6040), 345.

4.   Wang, W., *PhD Thesis* **2012**.

5.   Wang, W.; Nossoni, Z.; Berbasova, T.; Watson, C. T.; Yapici, I.; Lee, K. S. S.; Vasileiou, C.; Geiger, J. H.; Borhan, B., Tuning the Electronic Absorption of Protein-Embedded All-trans-Retinal. *Science* **2012,** *338* (6112), 1340.

6.   Philip, A. F.; Nome, R. A.; Papadantonakis, G. A.; Scherer, N. F.; Hoff, W. D., Spectral tuning in photoactive yellow protein by modulation of the shape of the excited state energy surface. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (13), 5821.