PROTEIN DESIGN: REENGINEERING OF CELLULAR RETINOL BINDING PROTEIN II (CRBPII) INTO A RHODOPSIN MIMIC, FUNCTIONALIZATION OF CRBPII INTO A FLUORESCENT PROTEIN TAG AND DESIGN OF A PHOTOSWITCHABLE PROTEIN TAG

By

Wenjing Wang

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Chemistry

2012

ABSTRACT

PROTEIN DESIGN: REENGINEERING OF CELLULAR RETINOL BINDING PROTEIN II (CRBPII) INTO A RHODOPSIN MIMIC, FUNCTIONALIZATION OF CRBPII INTO A FLUORESCENT PROTEIN TAG AND DESIGN OF A PHOTOSWITCHABLE PROTEIN TAG

By

Wenjing Wang

The intrinsic mechanism for wavelength regulation observed in color rhodopsins has been under intense study since the last century. One single chromophore, retinal, was found to absorb over a wide range of the visible spectrum, from 420 nm to 560 nm, depending on which rhodopsin it is bound to. Different model compound studies, rhodopsin mutagenesis studies and computational studies have been carried out to understand what causes the spectral differences. However, this question is still not conclusively answered, due to lack of crystal structures of the color rhodopsins and rhodopsin mutants. Our lab has engineered a small cellular protein, Cellular Retinoic Acid Binding Protein II, into a rhodopsin mimic that can bind all-*trans*-retinal as a protonated Schiff base. Further studies demonstrated that full sequestration of the chromophore from the bulk solvent is critical for spectral tuning. Therefore, a second generation rhodopsin mimic using Cellular Retinol Binding Protein II (CRBPII) was engineered and mutagenesis was carried out to study the cause-effect relationships of different stereo-electronic effects on wavelength regulation.

We were able to regulate the wavelength over an unprecedented range (474 nm to 644 nm), surpassing the existing limits. Electrostatic calculations based on the high resolution crystal structures of CRBPII mutants revealed that electrostatic interactions are playing the major role in the spectral tuning observed.

Fluorescent protein tags have been widely applied in microbiological studies to study the protein expression level, protein localization, protein-protein interactions and some important biological events. GFP and GFP like proteins have been greatly developed, along with some other fluorescent protein tags. However, in the fluorescence palette, bright photostable red fluorescent and near-IR are still lacking. Due to the robustness of CRBPII mutants, we wanted to functionalize them into red fluorescent protein tags or near-IR fluorescent protein tags by using appropriate chromophores. A merocyanine analogue of retinal, has proved to be suitable to be used along with CRBPII mutants as red fluorescent protein tags both in prokaryotic and eukaryotic systems.

Azobenzene has been widely applied in material science and chemical biological systems, in order to achieve photoswitchable properties. We want to design a protein tag that can bind specifically to the *trans*-isomer of the azobenzene derivatives. Upon light irradiation, the azobenzene derivative will isomerize to *cis*. Consequently, the protein tag will lose its affinity to the *cis*-isomer and dissociate. This photoswitchable protein tag can be used for light-controlled protein purification. Phage display was applied to evolve this protein tag from the phage library generated based on WT-CRABPII.

Dedicated to my beloved parents and husband, for their love and support.

ACKNOWLEDGEMENTS

I want to express my heartily gratefulness to my Ph.D advisor, Professor Babak Borhan, for his excellent guidance in my graduate research and continuous support throughout the years. I really appreciate his encouragement that helped me build confidence and criticism that helped me improve. My growth into an independent researcher would be impossible without his training and instructions. He has been a very caring and fun advisor as well. I not only learned a lot as a scientist during the past five and half years, but also lived a fascinating life.

Besides Babak, the other person that helped me a lot in my graduate school studies is Chryssoula vasileiou. She has been a great instructor and friend. She guided me into my major Ph.D research projects-rhodopsin mimic engineering, when I first joined the group and had zero background in chemical biology. She helped me through my first few group meeting presentations, student seminar and second year oral exam. I really appreciate the effort she put into, as I was a terrible speaker and presenter at that point. Without her, I could not have survived my first two years of graduate school studies. And I greatly appreciate the care she has for everyone in the lab and generousness in every aspect. The Chicago trip was very much fun, however it is such a shame that we could not finally go to Las Vegas together.

I would like to thank the BioLab members in Borhan's lab, King Sing Stephen Lee, Tanya Berbasova, Ipek Yapici and Camille Watson. I am very

grateful to have such wonderful persons as my Lab-roomates. Room 606 has been a very friendly place to work in and to discuss scientific problems and a bunch of nonsense. I want to thank Tanya for being a very sweet and good company either drunk or sober; Ipek for being a stubborn but a very good friend and all the dramas that she shared with me; Camille for her craziness, a lot of shopping that we did together and for the crystallization work she did.

I want to extend my thanks to Professor Jim Geiger, “the Maverick”, for creating the magic recipe for crystallization. It has been a lot of fun having Geiger meetings with him. I am also very grateful to have Rafida Nossoni as the collaborator. I greatly appreciate the hard work she has done and many crystal structures that she resolved.

I am heartily thankful to have to the present and past members in Borhan’s group, for their friendliness and help over the years, and all the fun that we had together: playing group soccer, summer trips and lots of parties and dinners. I especially need to thank Mercy Anyika, for being a good friend and all the fun activities that she brought me to. I also want to thank Roozbeh Yoosefi for sharing the hood when I was doing synthesis; Calvin Grant for being very entertaining and nice; Atefeh Garzan for her sweetness.

I also want to thank Satyaki (Bill Henry’s Lab) for the generous gift of pEGFP-RB vector and U2-OS cell lines and showing me the transfection in mammalian cells. I want to thank Dr. Melinda Frame for her assistance in confocal microscopic fluorescent pictures. I also want to thank Dr. Bruce

Hammock for allowing me to work in his lab for two weeks, Hee Joo Kim for showing me phage display and Sing for accommodating me for two weeks and taking me to all the nice restaurants in Davis.

I would like to express my sincere gratitude to Babak, Chryssoula and my committee members, Prof. Jackcson, Prof. Walker and Prof. Geiger, for proof-reading my thesis.

Last but not the least, I want to express my thanks to my beloved grandma, my parents and my younger brother for being very understandable and their support of my career. I want to thank my dearest husband for his support and caring all the time.

TABLE OF CONTENTS

List of Tables

List of Figures

Key to Symbols and Abbreviations

Å	Angstrom
ε	Extinction coefficient
λ_{max}	Maximal wavelength
cm	Centimeter
M	Molar
μM	Micromolar
>	Larger than
>>	Much larger than

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
ABCR	ATP-binding cassette transporter
Amp	Ampicillin
bR	Bacteriorhodopsin
cGMP	Cyclic Guanosine Monophosphate
Clm	Chloramphenicol
CRABPII	Cellular Retinoic Acid Binding Protein II
CRBPII	Cellular Retinol Binding Protein II
Da	Dalton
DMF	Dimethylformamide
DMSO	Dimehtylsulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
Equiv	Equivalent
EDTA	Ethylenediaminetetraacetic acid

FPLC	Fast Protein Liquid Chromatography
FQ	Fast-Q, quaternary ammonium resin
GDP	Guanosine Diphosphate
GTP	Guanosine Triphosphate
GPCR	G-Protein Coupled Receptor
h	Hour
HPLC	High Performance Liquid Chromoatography
iLBP	Intracellular Lipid Binding Protein
IPTG	Isopropylthiogalactoside
K_d	Dissociation constant
LB	Luria bertani
L	Liter
LRAT	Lecithin: retinal acyl transferase
MALDI-TOF	Matrix Assisted Laser Desorption Ionization-Time of Flight
Mero1	Merocyanine analogue of retinal 1
Mero2	Merocyanine analogue of retinal 2
mL	Milliliter
mM	Milli-molar
mW	Milliwatt
µL	Microliter
$nBuNH_2$	n-Butylamine
n.d.	Not determined

nm	Nanometer
nM	Nanomolar
NMR	Nuclear Magnetic Resonance
Ni-NTA	Nickel-nitrilotriacetic acid
$NiSO_4$	Nickel sulfate
PAGE	PolyAcryamide Gel Electrophoresis
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PDE	Phosphodiesterase
PSB	Protonated Schiff base
QY	Quantum yield
RA	Retinoic acid
Rt	Retinal
rt, RT	Room temperature
RPM	Revolutions per minute
SB	Schiff base
SDS	Sodium dodecyl sulfate
t	Time
$t_{1/2}$	Maturation half time
THF	Tetrahydrofuran
UV	Ultraviolet light
vis	Visible light

CRBPII mutants

WT	Wild type protein
KL	Q108K:K40L
KLV	Q108K:K40L:T51V
KLVF	Q108K:K40L:T51V:R58F
NaCl	Sodium chloride
Tris	Tris(hydroxymethyl)aminomethane
HCl	Hydrogen chloride
NaH_2PO_4	Sodium monophosphate

Chapter I. Introduction

I.1 How vertebrate color vision works

Ever since I learned about chemistry and molecular biology, I have always been amazed by how nature could achieve function through intricate, but accurate molecular mechanisms. It is natural for us to take for granted our normal physical functions, such as movement, memory and vision, while each of these capabilities actually involves inexplicably delicate chemistry. As the science has progressed, scientists have unraveled the mechanisms of many processes in biology. Nonetheless, a number of systems have not revealed their detailed mechanistic underpinnings.

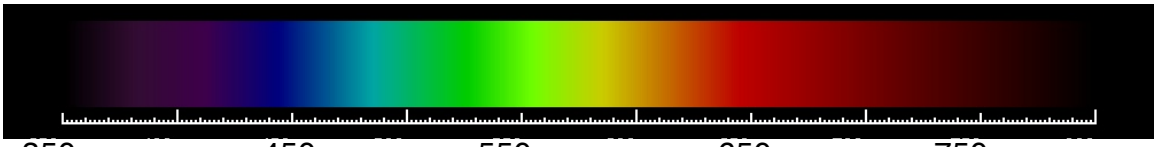


Figure I-1: Visible spectrum. "For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation."

How can we see colors? In simple words, how do we distinguish between different wavelengths in the visible spectrum? To answer this question, first we need to remember what are colors and what makes them different. From a middle school physics textbook, we learned that light exists in wave and particle form. Different colors of light have different wavelengths and correspond to different energies, according to the Planck-Einstein equation: $E = h\nu = h\frac{c}{\lambda}$. The rainbow colors, red, orange, yellow, green, cyan, blue, violet, are intrinsically

photons of different energy, corresponding to different wavelengths as illustrated in **Figure I-1**.

Energy, in the form of light, is absorbed by the eye's photoreceptors and is transformed into chemical energy that leads to the perception of color. The photoreceptor was first discovered in the rod cells of frog retinas by Franz Boll in 1877 as a purple pigment, later verified by Kühne in 1878 and finally named

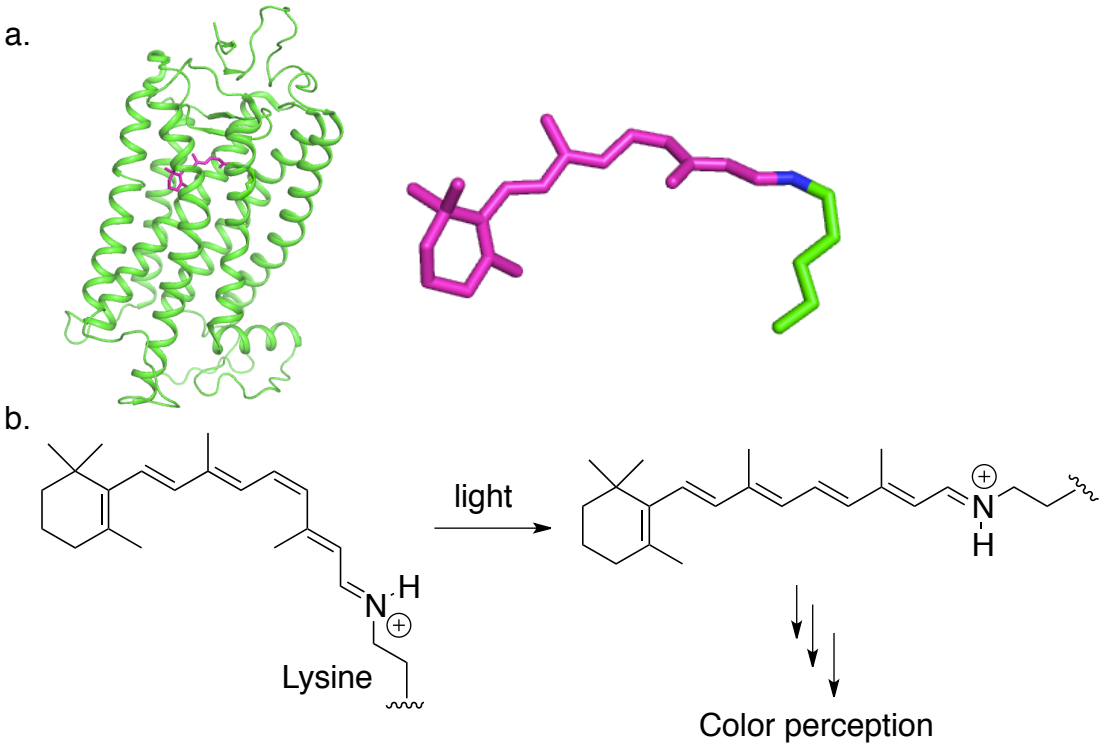


Figure I-2: Illustration of rhodopsin and isomerization. 11-cis-retinal is covalently bound to one of the lysine residues in the binding pocket as an iminium.

rhodopsin.[5] Rhodopsin is made of two components, one of which is a 7-α-helix transmembrane protein, called opsin, and the other one is a chromophore, identified as 11-*cis*-retinal.[6] The chromophore is covalently bound to one of the

lysine residues in the binding pocket of opsin, as a protonated Schiff base (**Figure I-2**).[7] Rhodopsin is highly sensitive to light; when exposed to light, the color is bleached very fast.[8] This is because upon light irradiation, 11-*cis*-retinylidene isomerizes on a femtosecond time scale to all-*trans*-retinylidene (**Figure I-2**). As a result of the latter isomerization and resulting change in shape, rhodopsin no longer binds the retinal, the imine bond is hydrolyzed, and the chromophore dissociates from the rhodopsin.[8]

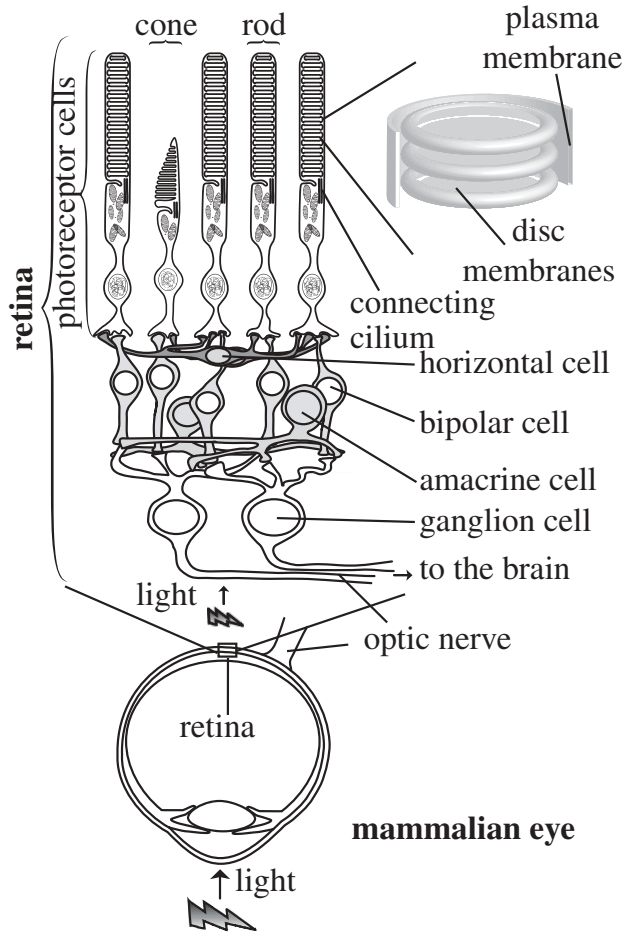


Figure I-3: Illustration of mammalian eye.[2]

Four photo-receptors are present in the retina of human eyes, rod opsin from rod cells, mainly responsible for dim light vision, and three color rhodopsins, blue, green and red rhodopsins from cone cells, which are concentrated in the fovea of the retina.[9] Blue rhodopsin is most sensitive to blue light, while green rhodopsin is most

sensitive to green light and red rhodopsin is most sensitive to red light. The absorption spectra of the three color rhodopsins in single cone cells were recorded in the 1960s when microspectrophotometers were introduced and used to measure the difference spectra of a single cone cell in the rod free area of the human fovea.[10]

Roughly speaking, there are 100 milliion rod cells and 6 million cone cells in the human retina.[2] Rod cells are more sensitive to light than cone cells, and are mainly responsible for dim light vision.

Detailed mechanistic studies of how light activation of rhodopsins leads to color vision have been performed. As illustrated in **Figure I-3,**[2] 11-*cis*-retinal combines with opsin to generate rhodopsin with the chromophore covalently bound as 11-*cis*-retinylidene. Different types of rhodopsins have their own intrinsic acuity to different light, depending on their absorption spectra. Upon light absorption, 11-*cis*-retinylidene isomerizes to all-*trans*-retinylidene with a quantum efficiency of around 67%.[11] Simply put, for every three photons absorbed by rhodopsin, two of them lead to signal transduction. Nature has perfected rhodopsin's structure to make the eye highly sensitive to light.

Upon retinylidene isomerization, the conformation of the rhodopsin changes to a large degree, leading to its activated form, which has high affinity for the α subunit of transducin (**Figure I-4**),[12] which belongs to the G protein family.

G proteins are a family of guanosine nucleotide binding proteins, and they are involved in transducing signals from outside the cell into changes inside the cell.[13] Transmembrane bound G Protein Coupled Receptors (GPCR) are necessary for G proteins to initiate signal transduction.[14] G proteins are made of three subunits, α, β, and γ. When the α subunit is bound with GDP (Guanosine DiPhosphate), the G protein is not active. It is activated when GTP (Guanosine TriPhosphate) displaces GDP, changing the conformation of the complex, resulting in the dissociation of the α subunit.

Upon binding to the activated rhodopsin, the α subunit of transducin is prompted to exchange GDP to GTP (**Figure I-4**).[15] As a result, the GTP-bound α subunit of transducin dissociates from the βγ complex. The α subunit is now free to bind with the two γ subunits of PDE (cyclic nucleotide phosphodiesterase), activating PDE to hydrolyze the phosphodiester bond of cGMP (cyclic Guanosine MonoPhosphate). This leads to a decrease of cGMP concentration, which closes the cGMP gated channel and causes a hyperpolarization in the photoreceptor cells and finally creates a signal for the associated neuron across the synaptic gap.[2, 16]

The signal from photoactivation of rhodopsin is amplified through two enzymatic reactions. Each rhodopsin activates ~100 G protein, each of which activates one PDE enzyme molecule, which can each hydrolyze ~1,000 cGMP molecules.[17] This explains why our eyes are sensitive to light.

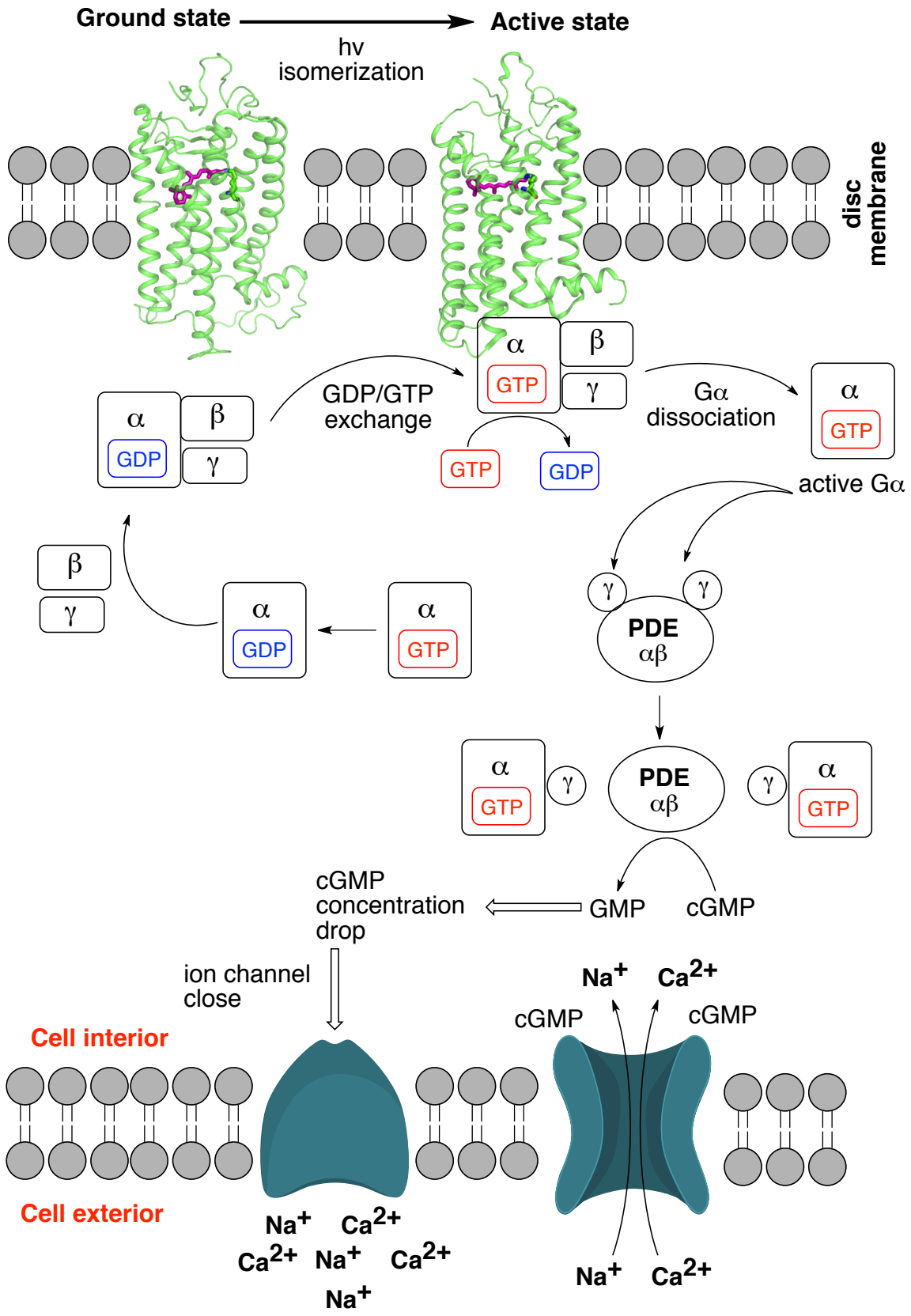


Figure I-4: Detailed diagram of photo transduction in the photoreceptor cells.

Rhodopsin, as a G protein coupled receptor, is an interesting target for crystallographers, not only for understanding color vision, but also for understanding GPCR mechanisms and the folding of membrane bound proteins.[18] This is even more pressing because more than 60% of pharmaceutical targets are membrane bound proteins.[19] The first crystal structure of dark adapted bovine rhodopsin was obtained by Palczewski, et al, in 2000, of a resolution of 2.8 Å.[20] It was the first crystal structure of a visual rhodopsin and revealed the folding of rhodopsin and the ligand binding site interactions. As shown in **Figure I-5**, rhodopsin has seven α-helices. 11-*cis*-retinal is covalently bound to Lys296 as a *trans*-imine. The acidic residue Glu113 forms a salt bridge with the iminium to stabilize the protonated state of the chromophore.

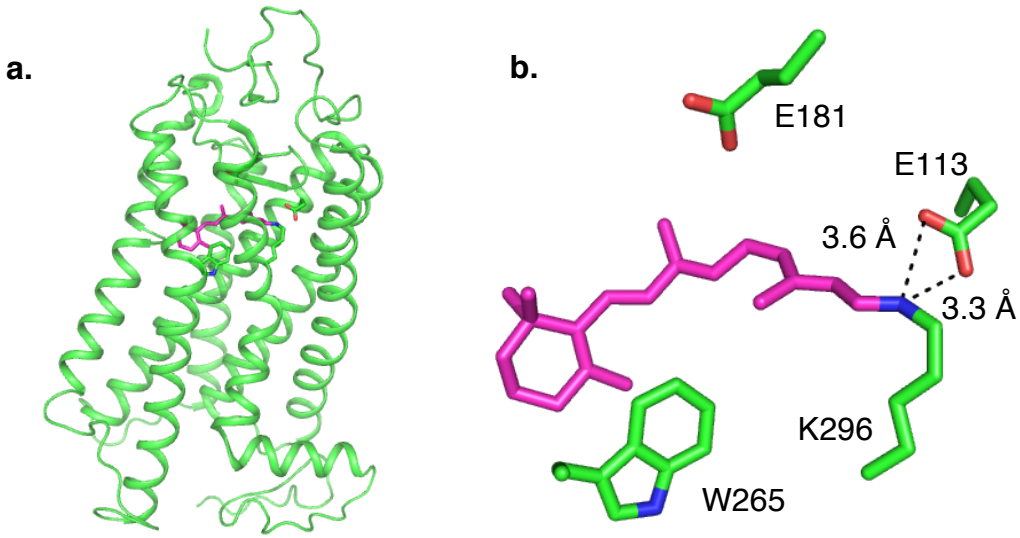


Figure I-5: **a.** Crystal structure of bovine rhodopsin. (PDB entry: 1F88) The protein has 7-alpha-helices, and the chromophore is sitting vertical to the rod, parallel to the membrane. **b.** Zoomed-in picture of the chromophore binding site with the counter anion highlighted.

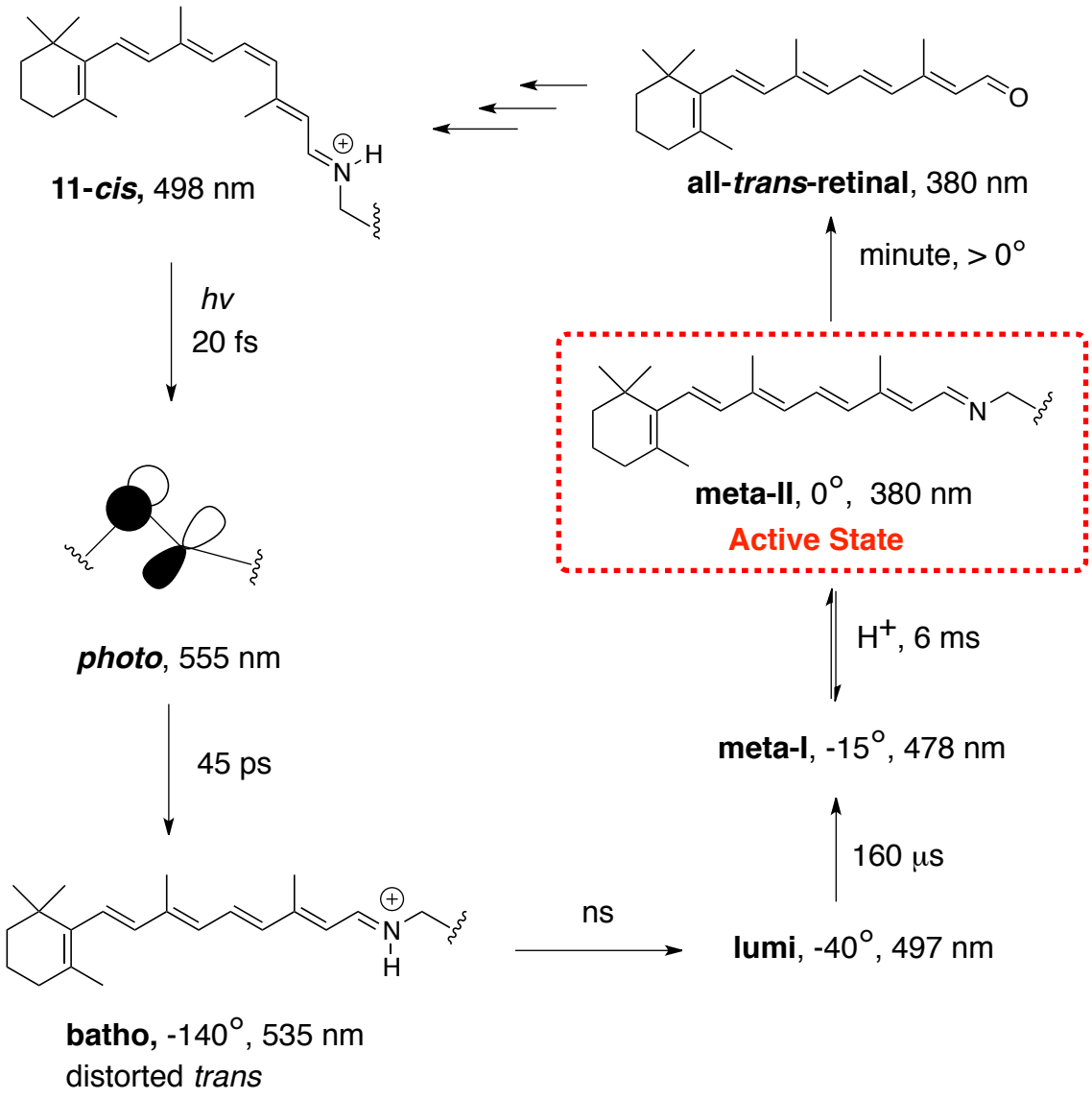


Figure I-6: Detailed transduction pathway of bovine rhodopsin.

The crystal structure mentioned above represents the ground state of rhodopsin. All the intermediates from the photocycle of rhodopsin have been extensively studied and some intermediates have been crystallized also. Each photo-intermediate can be trapped at different low temperatures and different visual transduction intermediates were assigned according to the different absorption spectra (**Figure I-6**).[21]

The isomerization from *cis* to *trans* isomer happens in a time scale of picoseconds.[22] The high efficiency of the isomerization is partially because of the highly distorted nature of 11-*cis*-retinal, due to the steric hindrance between the C10 hydrogen and C13 methyl group. Upon light activation, 11-*cis*-retinylidene isomerizes to the *trans*-isomer to relieve the steric hindrance. Within such a short time frame of isomerization, the protein can not adapt its conformation in response to the change of the chromophore structure; therefore a highly twisted all-*trans*-isomer is formed first. This intermediate is called bathorhodopsin, with λ_{max} at 535 nm. The crystal structure of bathorhodopsin was obtained by exposure of bovine rhodopsin crystals to 488 nm light at 95 K to

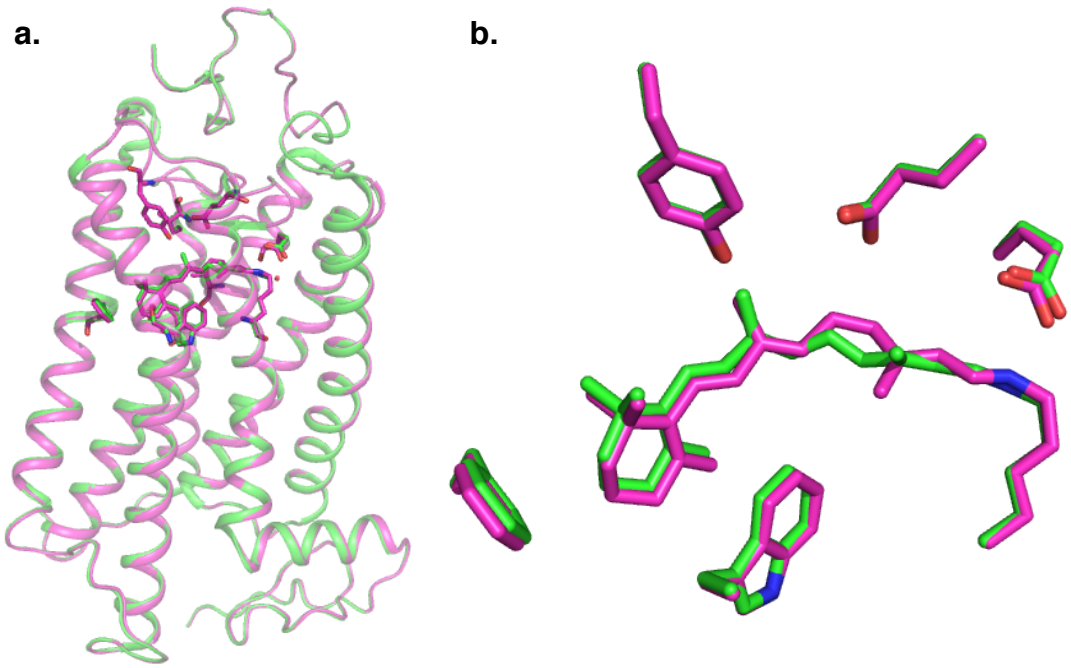


Figure I-7: Crystal structure of bathorhodopsin. Cartoon (a) and (b) the chromophore binding site of overlaid crystal structure of ground state bovine rhodopsin (magenta) (PDB entry: 1U19) and bathorhodopsin (green) (PDB entry: 2G87).

trap the batho intermediate.[23]

As shown in **Figure I-7,** the overlaid crystal structures of ground state rhodopsin and bathorhodopsin show that the backbone and the side chains of the protein barely move. Only the 11-*cis* double bond is isomerized to a highly twisted *trans*-bond, and in this form, it traps most of the energy absorbed from the light. This energy is what drives the subsequent conformational changes of the protein. The bathochromic shift observed in bathorhodopsin is due to the highly twisted nature of the double bond, which increases the ground state energy dramatically and lowers the energy gap between ground state and excited state, leading to the observed red shift of bathorhodopsin.

Meta II rhodopsin, a photo intermediate beyond the batho stage, is recognized as the active state for G protein activation, and is important in understanding the function and mechanisms of GPCR proteins. Very recently, two groups have published the crystal structure of meta II rhodopsin.[24, 25] Both of these groups were able to obtain the crystal form of meta II bound with a short C-terminal peptide derived from Gα of transducin. These crystal structures can show the interactions between the activated rhodopsin and Gα of the transducin, and more importantly the conformational change in the protein that leads to the activation of rhodopsin.

As shown in **Figure I-8a**, all-*trans*-retinal is bound in a very different way as compared to 11-*cis*-retinal due to their significant conformational differences, inducing the transmembrane helix 3 to move out to open up a binding site for the

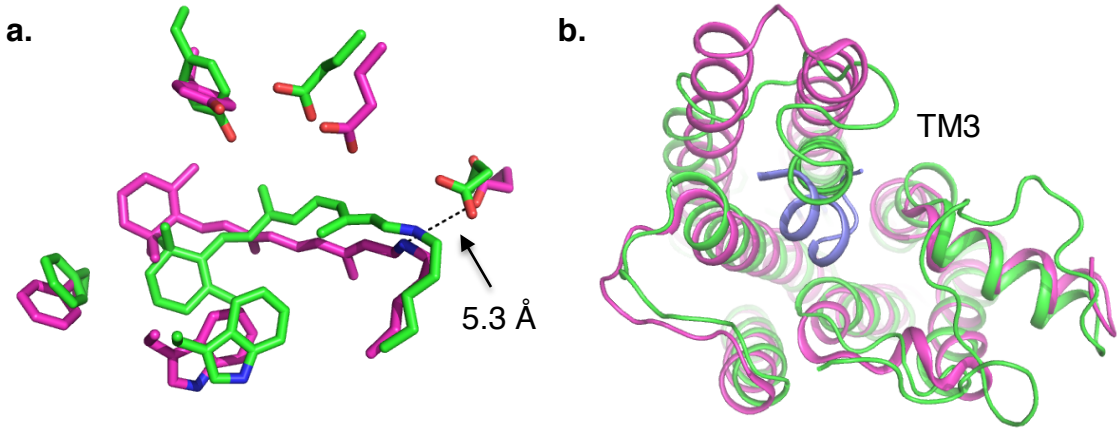


Figure I-8: Crystal structure of meta II. The chromophore binding site (a) and cartoon (b) of the overlaid crystal structures of ground state bovine rhodopsin (green) (PDB entry: 1U19) and meta II (magenta) (PDB entry: 3PQR). The short peptide of Gα of transducin is shown in purple.

Gα peptide **(Figure I-8b)**. This crystal structure also explains the drop in pK_a for the retinal protonated Schiff base (ret-PSB) in meta II rhodopsin. As a result of the isomerization and conformational change of the protein, the counteranion moves 5.3 Å away from the Schiff base nitrogen and also adopts an unsuitable angle for optimum hydrogen bonding via a bridged water molecule.

It has been observed that the active form of the photoactivated rhodopsin intermediate, meta II, has a longer life time in rod rhodopsin than any of the other pigmented rhodopsins.[26] Therefore, rod rhodopsin could activate more G proteins than blue, green or red rhodopsin. This is another reason why rod rhodopsin is more sensitive to dim light than cone color opsins, besides the fact that there are more rod rhodopsin than cone rhodopsin.[2]

The instability of meta II rhodopsin prompts the retinal to dissociate from the protein, releasing all-*trans*-retinal. For the visual cycle to start again, the

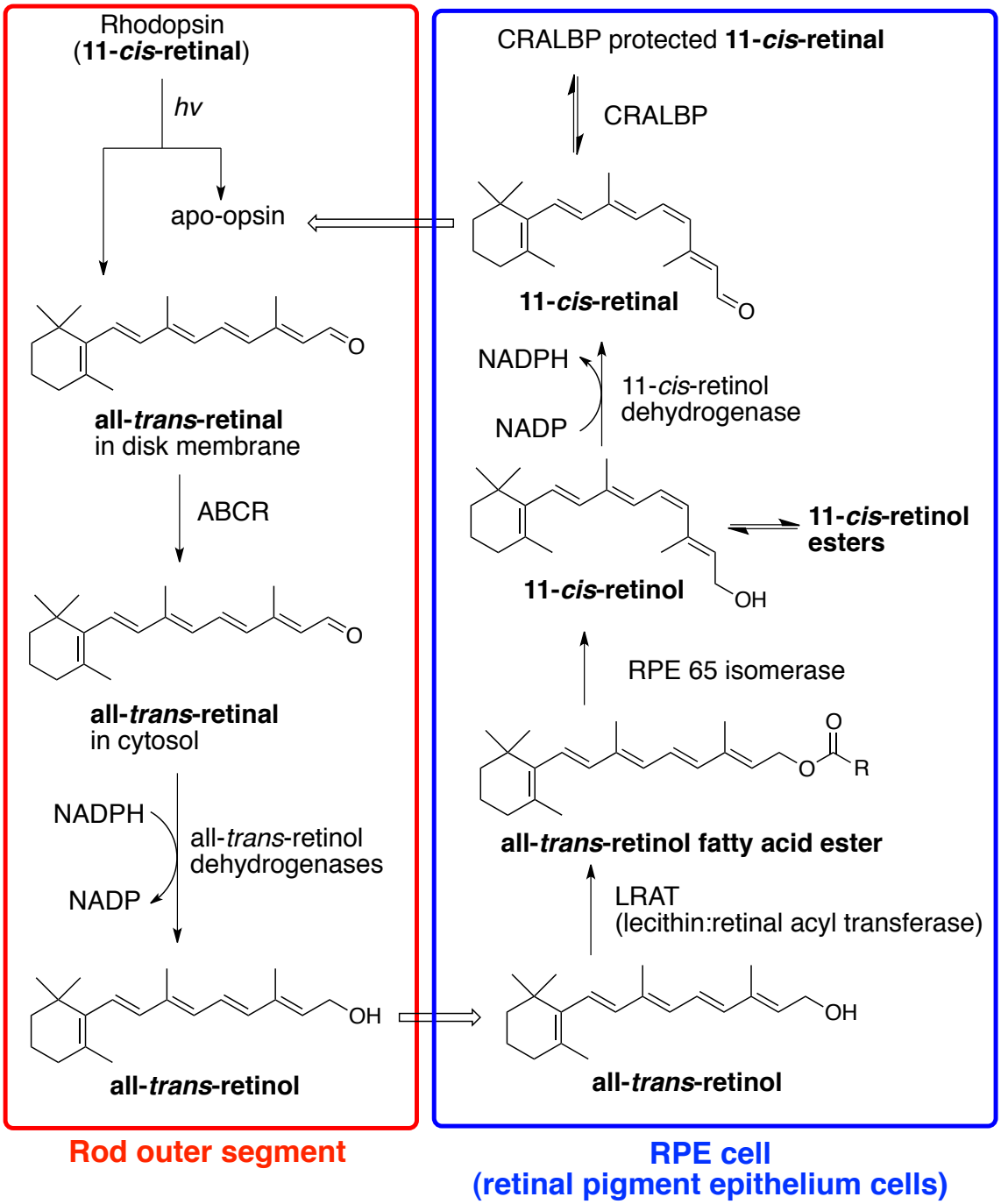


Figure I-9: 11-*Cis*-retinal photocycle regeneration pathway in human retina.

photopigment must be regenerated. For that to happen, all-*trans*-retinal must be isomerized to 11-*cis*-retinal.

Figure I-9 summarizes the major pathway for regeneration of 11-*cis*-retinal in retinal pigment epithelium cells.[27] After photoisomerization, the

chromophore dissociates from the rhodopsin embedded in the disk membrane of the outter segment of rod cells and is transported to the cytosolic solution by an ATP-binding cassette transporter (ABCR). All-*trans*-retinal is first reduced to all-*trans*-retinol by all-*trans*-retinol dehydrogenases and is translocated to the retinal pigment epithelium. Lecithin: retinal acyl transferase (LRAT) catalyzes the esterification of all-*trans*-retinol with palmitoyl acid.[28] All-*trans*-retinol ester is isomerized and at the same time hydrolyzed to 11-*cis*-retinol by RPE65, a protein found in retinal epithelium cells.[29] The hydrolysis of the ester bond provides the energy necessary to compensate for the isomerization from the *trans* to *cis* isomer. After that, the 11-*cis*-retinol is oxidized to the aldehyde form and is sequestered by 11-*cis*-retinal binding protein, CRALBP, before it is transferred back to the outer segment of rod cells to regenerate rhodopsin, after binding to opsin.

I.2 Wavelength regulation studies on model compounds

Color vision has been under intense study since the mid 20th century. Starting from finding a red pigment in human retinal by Kühne (1877), trichromacy was not fully demonstrated till the 1960s when microspectrophotometry was developed, which enabled UV-vis recording of single rod and cone cells. There are three types of cone rhodopsins in the retina of human eyes, each absorbing at: short wavelength (~420 nm), medium

wavelength (~530 nm) and long wavelength (~560 nm).[30, 10] Collectively, they constitute the foundation of human color vision. Different vertebrate species exhibit varied cone photoreceptors as a result of adaptation to their environment for survival.

The chromophore identified for vertebrate cone receptors is usually 11-*cis*-retinal. However, 11-*cis*-dehydroretinal, which has one more double bond added to the conjugated polyene system as compared to 11-*cis*-retinal, is found in some invertebrate species.[31, 32] Rhodopsins using 11-*cis*-dehydroretinal are called porphyropsin.

Deep water fish tend to have more red shifted cone photoreceptors; for some of them the absorption reaches up to 630 nm, attributed to the 11-*cis*-dehydroretinal.[31] This is an evolutionary result, as light with a longer wavelength is more penetrable in water. For some species, rod rhodopsin can accomodate both chromphores, 11-*cis*-retinal and 11-*cis*-dehydroretinal.[33] In these cases, the photoreceptor ratio changes with season, but the overall number of total photopigments remains constant.[34] For some amphibian species like frogs, the photoreceptor changes from phophyropsin (520 nm) to rhodopsin (500 nm), when metamorphosis occurs, as a result of evolution to adapt to different environments.[35]

The wide-range of absorptions observed in rhodopsins bound with the same chromophore, 11-*cis*-retinal, has been under intense investigations. 11-*cis*-retinal absorbs at 380 nm in aqueous solution, and when it forms a Schiff base (SB) with n-butylamine, it blue shifts to 360 nm. When the SB is protonated, it red shifts 80 nm to 440 nm. As all three color rhodopsins are much more red shifted than 380 nm, it indicates that 11-*cis*-retinal forms a protonated Schiff base in the rhodopsins.

Compared to the PSB of retinal in aqueous solution, rod, green and red rhodopsins are much more red shifted. This red shift, also known as opsin shift,[36] has raised a lot of interest. Different model studies have been performed and general hypotheses have been proposed to explain this opsin shift, before genetic manipulation of rhodopsin was possible.

Based on a general knowledge of polyene systems, more delocalized π electrons can lead to a more conjugated system and a red shift. It is generally believed that through protein-chromophore interactions, the positive charge of the iminium can be delocalized along the polyene and contribute to a large portion of the opsin shift observed.

Red rhodopsin is 120 nm more red shifted than retinal-PSB in aqueous solution. This has spawn great interest and curiosity in determing the maximum red shift possible for retinal-PSB. In 1971, Blatz showed that the conjugated cation derived from dehydration of all-*trans*-retinol (320 nm) was highly red shifted (600 nm) compared to the neutral form of the molecule (**Figure I-10**).[37]

This dramatic red shift is the result of the highly delocalized positive charge along the polyene. This seems to represent the most delocalized positive charge in a polyene system and suggested the possible longest wavelength of retinal-PSB to be ~600 nm.

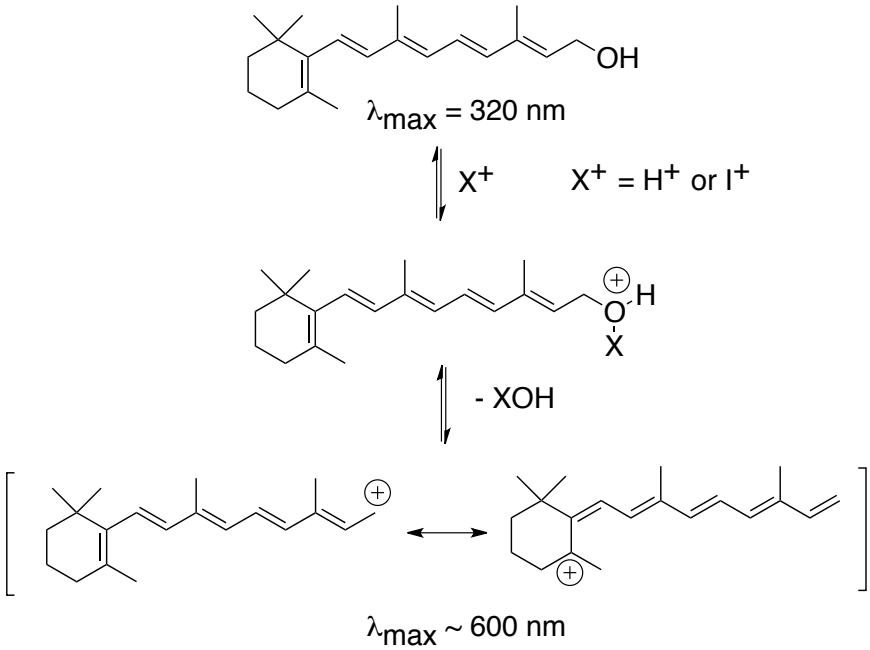


Figure I-10: Chemical conversion of all-*trans*-retinol into retinene cation results in a dramatic red shift.

Since retinylic carbocation is not retinylidine after all, there were more studies later to predict the likely absorption maximum for retinylidine. In order to achieve the maximum delocalization of positive charge along the polyene, the counteranion was moved at least 10 Å away from the iminium nitrogen. Blatz performed a model study of retinal-PSB with counteranions of different radii.[38] As expected, counteranions with larger radii generated more red-shifted absorption maxima. In addition, Blatz and coworkers could extrapolate the

distance at which the counteranion had negligible electrostatic effect on the stabilization of the positive charge and they predicted the long-wavelength limit to be around 580 nm.[38] Honig and Ebrey also predicted the isolated PSB to be absorbing around 600 nm.[39]

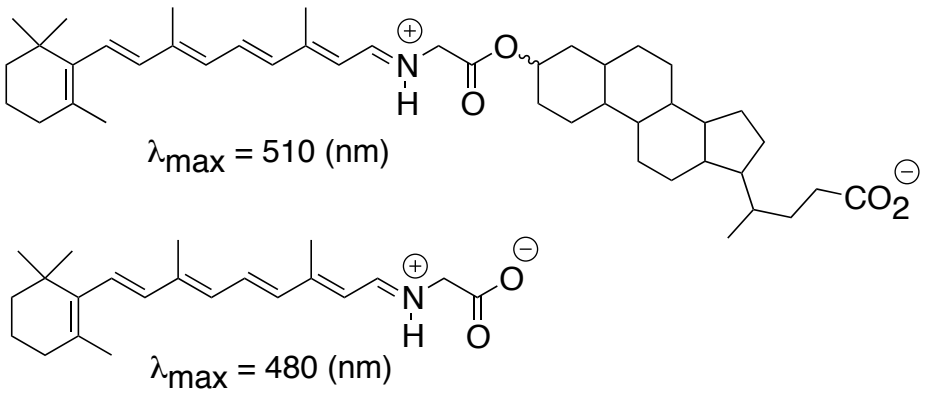


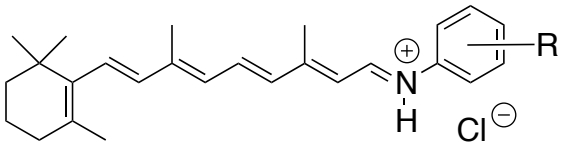
Figure I-11: Model compounds for studying the counteranion effect on the absorption maxima of retinal-PSB.

Recently the spectrum of the retinal-PSB formed with n-butylamine in vacuum was obtained and the absorption maximum was determined to be 610 nm.[40, 41] In vacuum, since there is no counteranion to stabilize the positive charge of the iminium, the positive charge can be delocalized to the largest extent and cause the most red shifted absorption. The experimental value coincides with the previous studies and shows the importance of the positive charge delocalization along the polyene for bathochromic shift. That would explain why no absorption maximum surpassing 610 nm has yet been reported for stable retinal-PSB species in nature.

The important role of the counteranion in modulating the absorption maxima of the retinal-PSB has been realized. Weakening the interactions of the counteranion with the iminium could encourage positive charge delocalization and lead to red shift. Sheves and coworkers have shown that a red shift resulted when the counteranion was moved further away from the iminium site using rigid framework compounds that place the counteranion at predetermined distance from the iminium (**Figure I-11**).[42]

Electronic interactions between the opsins and retinal-PSB could play important roles in the opsin shifts observed in color rhodopsins. As early as 1967, Rosenberg used retinal-PSB models to study electronic inductive effects on the absorption profile of the chromophore. As **Table I-1** shows, electron withdrawing groups, which could destabilize the positive charge on the imine

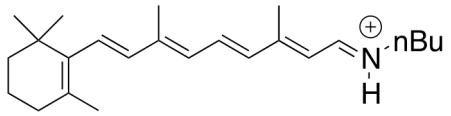
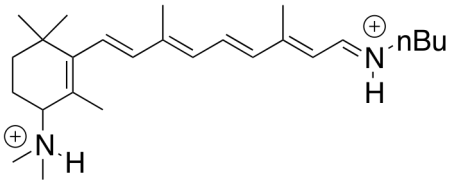
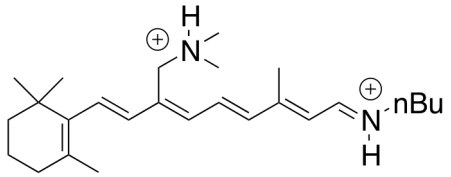
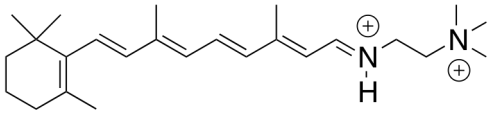
Table I-1: Inductive effects on the absorption of retinal-PSB.



-R	Para substitution		Meta substitution
	λ_{max} in EtOH (nm)	λ_{max} in $CHCl_3$ (nm)	λ_{max} in EtOH (nm)
$-OCH_3$	505	522	508
-H	504	522	504
-Cl	512	530	514
-CN	533	564	-
$-NO_2$	543	574	522

nitrogen through inductive effects, result in red shift.[43] This study proves that the

Table I-2: Effects of different placement of positive charge on the absorption maxima of retinal-PSB.

Chromophore	λ_{max} (nm)				
	EtOH	TFE	TFIP	CH_2Cl_2 (1 eq of TFA)	CH_2Cl_2 (1 M TFA)
	440	467	492	448	513
	423	431	442	426	461
	419	419	428	423	455
	455	508	536	468	538

more destabilized the positive charge on iminium, the more red shifted the absorption is. The author suggested that the protein environment of rhodopsin could have similar effects. By inductive or field effects on the positive charge of the iminium, different degrees of opsin shift could result.[43]

Sheves' group has also shown that electronic effects between retinal-PSB and its surrounding environment can have a significant impact on the

absorptions, through model compound studies.[44] As shown in **Table I-2**, the most red shifted model compound is the one with the positively charged ammonium group placed on the side of the PSB, which can promote the charge delocalization across the polyene due to positive charge repulsion. On the contrary, when the positively charged ammonium is placed near the middle of the polyene or the ionone ring, blue shift results, possibly due to reduced charge delocalization. This suggests to us that different electronic environments created by the rhodopsins could effect the retinal-PSB in a similar way, to either promote or inhibit the positive charge delocalization and result in different opsin shifts.

Along the same lines, the point charge theory was proposed to account for the opsin shift.[36] According to this theory, besides counteranion, there is another negatively charged residue along the polyene to interact with the delocalized positive charge and increase the delocalization along the polyene.

The idea of point charge was postulated because a counteranion in close proximity to the iminium is necessary to stabilize the retinal-PSB in order to achieve a high pK_a in rhodopsin. Therefore, removal of counteranion could not be used to explain the dramatic opsin shift observed in green rhodopsin (530 nm) and red rhodopsin (560 nm). To compensate for the stabilization of positive charge on the iminium, a negative point charge along the polyene could stabilize the delocalized positive charge and lead to red shift.

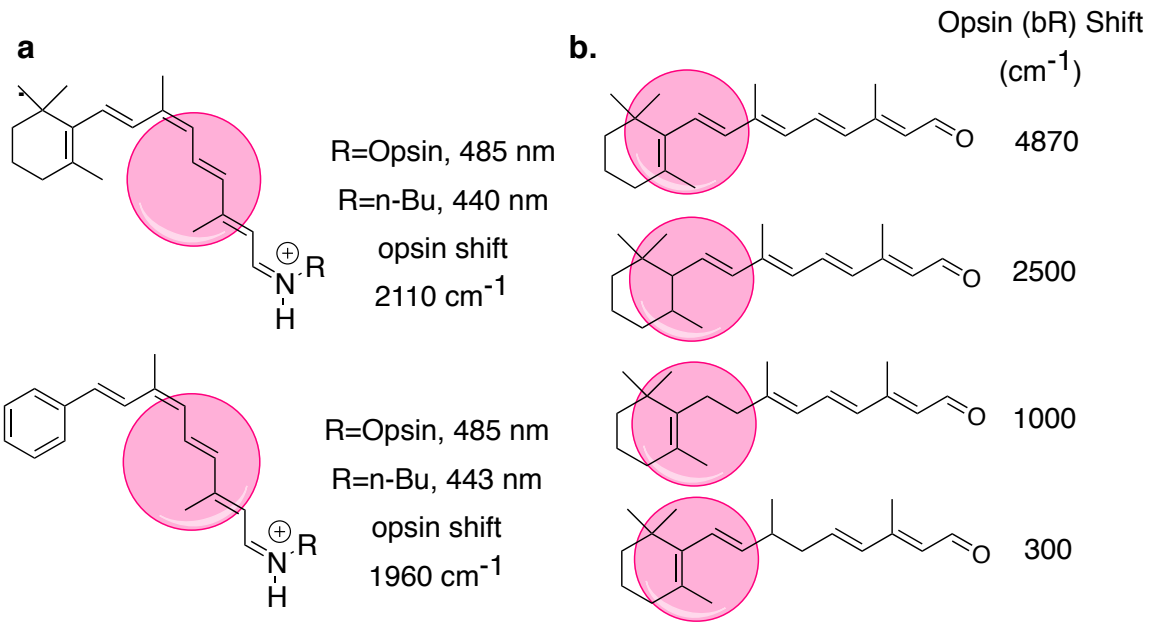


Figure I-12: Model compounds studied to support the point charge theory. **a.** Opsin shift with different retinal analogue. **b.** bR opsin shift with different dihydro-retinal analogues.

Nakanishi and Honig proposed that the location of the point charge for bovine rhodopsin and bacteriorhodopsin were different.[36, 45] In bovine rhodopsin, the point charge was localized in the middle of the polyene, while in bacteriorhodopsin the point charge was localized near the ionone ring region. This was proposed as one of the reasons why bacteriorhodopsin (bR), 560 nm, was more red shifted than bovine rhodopsin, 500 nm.

As shown in **Figure I-12a**, the change of chemical structure of the ionone ring for 9-*cis*-retinal resulted in a similar degree of opsin shift as compared to 9-*cis*-retinal, when bound with bovine rhodopsin.[36] This agreed with the hypothesis that the point charge was localized near the middle region of the polyene, which could explain the similar opsin shift in spite of the drastic change

in the ring structure.[36] As shown in **Figure I-12b**, for bacteriorhodopsin (bR), the most pronounced opsin shift was observed when the conjugated system is extended to the ionone ring region, possibly due to the closest interaction with the negative dipole in the ionone ring region.[46]

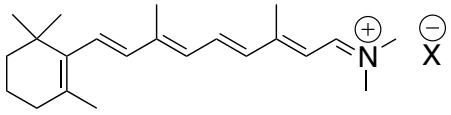
The point charge theory has lost favor, as mutagenesis on most of the charged amino acids do not cause large shifts, except the counteranion.[47] The crystal structures of bovine rhodopsin and bacteriorhodopsin also show that there are no negatively charged amino acid residues in the binding pocket in close proximity to the chromophore besides the counteranion.[20, 44, 48, 49] The point charge theory might still stand if we consider negative dipoles instead of negative charges.

Besides electronic effects from permanent dipoles, polarizability could play an important role in the opsin shift as well. As early as the 1970s, Irving proposed that induced dipoles could affect the spectra of retinal-PSB to a large extent.[50] This is because the charge distributions of retinal-PSB are very different in the ground state and excited state, which could affect the polarity of a polarizable environment to a large extent. Upon electronic excitation, the positive charge localized on the iminium is transferred toward the ionone ring. In 1976, Rich Mathies and Lubert Stryer determined the dipole moments of the ground state and excited state of retinal protonated Schiff base and observed the negative charge transfer from the ionone ring region toward the PSB region upon

electronic excitation.[51] In other words, the positive charge localized on the PSB end is transferred to the ionone ring end.

Permanent dipoles of the protein side chains are oriented to stabilize the ground state of retinal-PSB to the largest extent possible to reach energy minimum for the ground state. When the retinal-PSB is excited, as a result of different charge distribution of retinal-PSB, the permanent dipoles can not have favorable electrostatic interactions with the excited state. Besides, reorientation of permanent dipoles can not take place within the time frame of electronic excitation from ground state to excited state (10^{-15} sec). However, polarizable residues could generate inducible dipoles upon electronic excitation to have optimal electrostatic interaction with the excited state.

Table I-3: Solvent effects on retinal-PSB with different counteranions.



Solvent	Dielectric constant	λ_{max} (nm)		
		$X^- = ClO4^-$	$X^- = I^-$	$X^- = Br^-$
Ethyl ether	4.33	451	445	442
Methanol	32.36	453	448	444
Benzene	2.28	474	455	451
Chloroform	4.80	481	485	477
Dichloromethane	9.08	496	489	484

This hypothesis was supported by the fact that polarizable solvents, such as benzene, chloroform and dichloromethane, yielded more red shifted species (**Table I-3**).[52]

I.3 Mutagenesis studies on visual rhodopsins

With the development of genetics and molecular cloning, it became possible to compare different genetic sequences of visual rhodopsins. Jeremy Nathans and colleagues for the first time sequenced the genes encoding blue, green and red opsins in 1986.[53] As shown in **Table I-4,** blue rhodopsin is very different from green and red rhodopsin, while red and green rhodopsins are different in only 15 amino acids, indicating that the two are evolutionarily closely related.

Table I-4: Sequence identity (below the diagonal) and sequence homology (above the diagonal) between different visual rhodopsins.

	Percentage			
	Rhodopsin	Blue	Green	Red
Rhodopsin	100	75	73	73
Blue	42	100	79	79
Green	40	43	100	99
Red	41	44	96	100

Note: Sequence identity refers to the percentage of identical amino acids in sequence alignment. Sequence homology represents the percentage of conserved amino acids with the same functionally equivalent phiscochemcial properties.

Later in 1991 Neitz and coworkers aligned the genes encoding eight cone pigments with varied absorption from 530 nm to 562 nm and found that three amino acids likely accounted for the 30 nm absorption difference between green and red rhodopsin. Moreover, it was suggested that the effect of these three positions was additive.[54]

a.

Mutation	λmax (nm)	Protein shift (cm^{-1})
WT-bRho	500	0
A164S	502	75
F261Y	510	400
A269T	514	550
F261Y/A269T	520	775
F261Y/A164S	512	400
A269T/A164S	514	550

b.

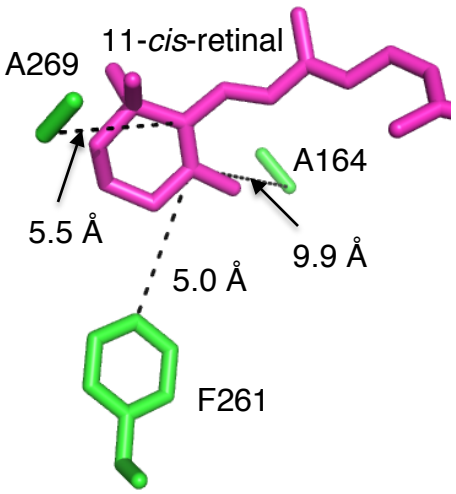


Figure I-13: Mutagenesis on bovine rhodopsin. **a.** Table of mutagenesis on bovine rhodopsin. **b**. Crystal structure of bovine rhodopsin with the three positions for mutagenesis highlighted.

This finding was further supported through mutagenesis studies on rhodopsin by Oprian's group and Sakmar's.[54] Oprian's group mutated all 15 positions, where green rhodopsin is different from red rhodopsin according to sequence alignment.[55] The conclusion was that seven amino acids were

responsible for the different absorption spectra between red and green rhodopsins, three of which had been proposed by Neitz in 1991.

Sakmar's group also made mutations on the three corresponding positions in bovine rhodopsin, which was proposed by Neitz, and indeed red shift was observed as listed in **Figure I-13**. With the crystal structure of rhodopsin resolved,[20] we can see that A269, F261 and A164 are located in the ionone ring region. Mutation of hydrophobic residues into polar residues in the ionone ring region could increase the favorable electrostatic interaction of the protein with the excited state of the chromophore and lead to a red shift. However, these three mutations only account for 20 nm of the opsin shift, not explaining the 120 nm opsin shift observed in red rhodopsin.

We can also conclude from the crystal structures why A164S has less perturbation on the absorption of retinal-PSB, as compared to A269T and F261Y. A164 is 9.9 Å away from the conjugated system of the retinal-PSB, almost twice the distances of A269 and F261 from the chromophore. As electric force is in inverse relationship with the square of the distance between the two interacting points, a longer distance leads to much weakened electrostatic interactions. However, it is surprising to see that A269 and F261 still have such an effect in red shifting the absorption of retinal-PSB, considering they are around 5 Å away from the conjugated system of the chromophore. This is probably due to the low dielectric constant in the hydrophobic binding pocket of bovine rhodopsin, which allows the electric field from the dipole to have a long range effect.

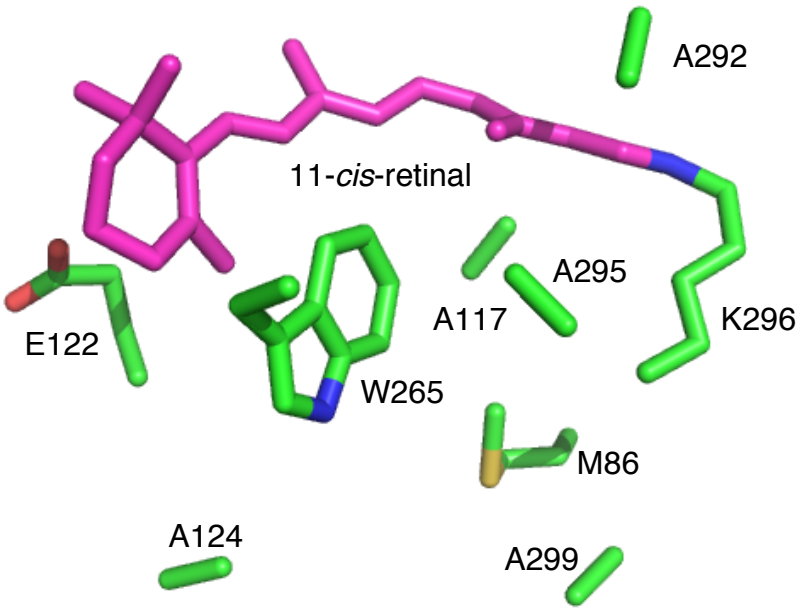


Figure I-14: Bovine rhodopsin crystal structure (PDB entry: 1U19) with retinal and some residues highlighted. These positions were important for blue-shifting of bovine rhodopsin from 500 nm to 438 nm.

This mutagenesis studies of bovine rhodopsin suggest the importance of polarity change of the protein environment on the spectral tuning of the retinal-PSB. It might also indicate the necessity to have an enclosed hydrophobic binding pocket for more efficient wavelength regulation.

Along the same lines, Sakmar's group mutated 9 amino acids that could blue shift the rod rhodopsin from 500 nm to 438 nm.[56] As shown in **Figure I-14**, the applied mutations were E122L, W265Y, A124T, A299C, M86L, A295S, A117G, A292S and G90S. The results agreed with the general trend that removal of negative polarity in the ionone region together with introduction of negative polarity in the PSB region could lead to blue shift, as it stabilized the positive charge in the PSB region for the ground state while destabilizing the

positive charge in the ionone region for the excited state, increasing the energy gap and thus resulting in blue shift.

However, no crystal structures of the mutants were solved to verify that the blue shift was simply a result of the electronic effect and not due to conformational change of the protein. Trp265 was found to be important for tight packing interactions with retinal and is conserved in both green and red rhodopsins, while in blue rhodopsin it is replaced with a tyrosine.[57] Among the mutations Sakmar made, W265Y has the most dramatic effect in blue shifting bovine rhodopsin. This tells the signicant role that Trp265 plays in the red shift observed in the long wavelength rhodopsins. However, without the crystal structure of mutant W265Y, it is not clear how the W265Y mutation leads to the blue shift.

The role of the counteranion has also been well studied using the actual rhodopsin system. First, Oprian's group identified the residue that was likely to be the counteranion, by mutation of every glutamate and aspartate residues thought to be buried in the membrane into neutral glutamine and asparagine residues. It was found that only E113Q abolished the protonated Schiff base formation by significantly lowering the pK_a.[58]

Sakmar's group later showed that the red shift caused by Glu113 mutations also depends on the radius of anions in the solution as shown in **Table I-5.**[59] The studies revealed that replacement of negatively charged Glu113 by

the neutral residue glutamine or the hydrophobic residue alanine, or just shortening of the carbon chain using aspartate could result in up to 30 nm red shift. This is likely due to less stabilization of the positive charge on the iminium, which encourages charge delocalization along the polyene.

Table I-5: Effect of solute anions on absorption maximum of rhodopsin E113 mutants.

Solution Anion	λ_{max} (nm)				
	Rho	E113D	E113N	E113Q	E113A
Fluoride	499	505	515	508	486
Chloride	499	510	520	496	506
Bromide	499	504	512	493	500
Iodide	501	510	519	504	507
Formate	499	507	520	488	496
Acetate	498	501	522	488	496
Perchlorate	498	508	520	500	510
Tartrate	497	509	522	510	525
Citrate	497	507	524	516	526
Benzoate	497	509	524	513	528
Chloroacetate	498	504	514	494	488
Dichloroacetate	498	504	514	501	509
Trichloroacetate	498	503	519	506	512
Deionized water	500	513	520	515	528

In the case of E113A, the solute anion plays a critical role in the absorption maximum. It is likely that the cavity created by E113A could only accommodate small solute anions to stabilize the PSB. Therefore, large solute anions cause more red shift, as they can not be accommodated in the binding cavity to be close to the iminium. For instance, tartrate, citrate and benzoate all

result in a larger red shift than smaller solute anions such as Fluoride and Chloride. However, no crystal structure was obtained to support this hypothesis.

Similar phenomena have been observed with some cone visual pigments, where incubation of cone visual pigments with different solute anions results in different absorption maxima. This was proposed to be due to different hydrogen bonding networks with different solute anions in the PSB region.[31] However, without crystal structures, it is hard to picture the binding mode and analyze its effect.

To sum up, mutagenesis studies in visual rhodopsin greatly helped understand more about wavelength regulation. However, to fully understand the protein-chromophore interactions and to study the cause and effect relationship in wavelength regulation of retinal-PSB, it is necessary to obtain crystal structures of the WT protein and its mutants as well. This way, it would be more conclusive to assign the cause-effect relationships, either due to conformational change or electronic factors, or a combination of both.

I.4 Mutagenesis studies on microbial rhodopsins

Visual rhodopsins initiated the field of wavelength regulation studies. It was found later that rhodopsin systems are not unique to higher organisms. Similar retinal-bound rhodopsins are found in lower organisms also such as archaea and bacteria. These microbial rhodopsins are used for functions other than vision, such as phototaxis, and light activated proton and chloride-

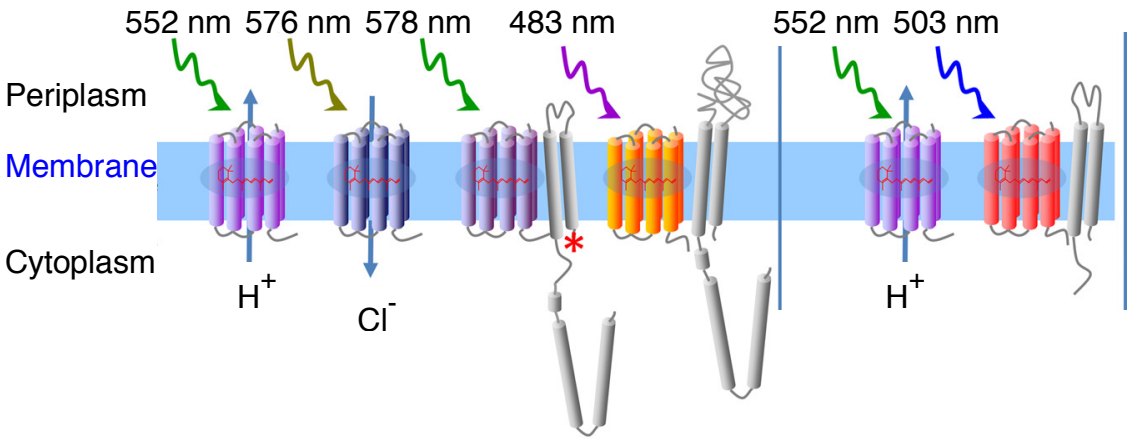


Figure I-15: Illustration of six different rhodopsins found in a single achaeon.[4]

pumping.[60] Different types of rhodopsins could be found in a single organism with varied absorption wavelength, with an extreme case reported recently, where six different rhodopsins were found in a single archaeon.[4] The absorptions for the six rhodopsins found range from 483 nm to 578 nm. A diagram illustrating different light-dependent functions of these six rhodopsins is shown in **Figure I-15**.

Microbial rhodopsins provide a good platform to study wavelength regulation, as the crystal structures of sensory rhodopsin II, bacteriorhodopsin and halorhodopsin have been obtained.[48, 61, 62] A number of tryptophan residues were found to line the polyene in the binding pocket of the three latter microbial rhodopsins. This observation brought attention to the function of these electron rich, polarizable aromatic residues in wavelength regulation.

Mutations of the tryptophan residues in **Figure I-16** to phenylalanines usually resulted in a blue shift.[63] This might be because tryptophan could

stabilize the delocalized positive charge through π-cation interactions to encourage positive charge delocalization. Besides, polarizable tryptophan residues could also stabilize the positive charge, which is transferred to the ionone ring in the excited state of the chromophore through induced dipoles, to decrease the energy gap and lead to red shift. In addition, flat tryptophan residues could also contribute to a red shift through planarization of the polyene system due to tight packing. Although the crystal structure of WT-bacteriorhodpsin is available now, mutants made on tryptophan for wavelength regulation studies are not crystallized yet to help better understand the contribution to the opsin shift.

The two methionine residues, M118 and M145 **(Figure I-16)**, close to the ionone ring were also found critical for the red shift observed in bacteriorhodpsin.[64] This could be due to tight packing between the chromophore and the methionine residues or the polarizability of the sulfur group. In this case, the crystal structures of mutants of these two methionine residues are extremely important in analyzing their effects on wavelength regulation.

The counteranion effect was studied via mutagenesis of the counteranion in bacteriorhodopsin. Mutation of Asp85 into neutral residues leads to a red shift as expected. It is interesting to observe a red shift at pH=2 for bacteriorhodopsin.[65] It was believed that this is because the counteranion Asp85 is protonated at pH=2, which leads to destabilization of the positive charge on the iminium and red shift. However, bacteriorhodopsin has a complex

hydrogen bonding network in the PSB region. Crystal structures are important to unravel the effect of the change in the hydrogen bonding network on wavelength regulation, and the pK_a of the PSB.

Although the crystal structures of both sensory and bacteriorhodopsins are available, the factors causing the absorption difference between sensory rhodopsin II (500 nm) and bacteriorhodopsin (560 nm) are still not conclusive. Ten amino acids in sensory rhodopsin II, within 5 Å from the chromophore, were replaced with the corresponding amino acids in bacteriorhodopsin, in order to red shift it from 500 nm to 560 nm, but this was not successful. A maximum red shift of 24 nm was achieved.[66] This result indicates that residues further away from the chromophore could have an effect on the spectral profile. There could

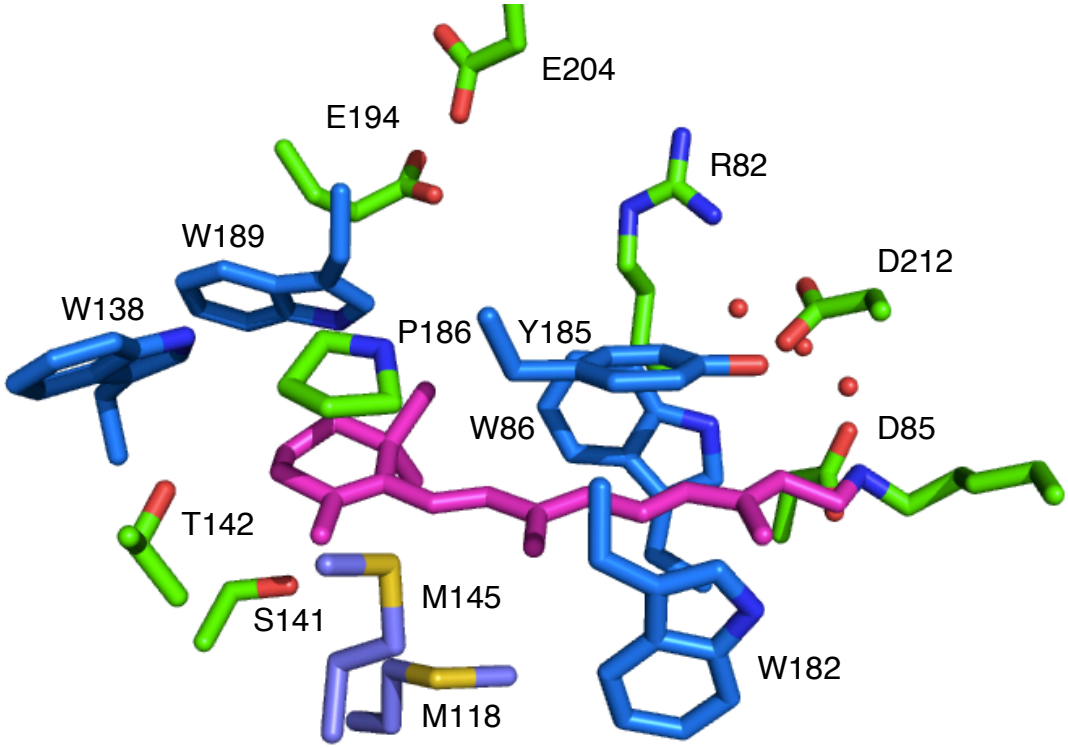


Figure I-16: Binding cavity of bacteriorhodpsin (PDB entry:1C3W).

be some long-distance electrostatic interactions, as rhodopsin is embedded in the hydrophobic membrane and is in a relatively low dielectric environment. Or even residues outside the binding pocket could have structured effects on the binding cavity of the protein.

In summary, different mutagenesis studies on microbial rhodopsins did facilitate the investigation of wavelength regulation; however, the limitation is that the crystal structures of these rhodopsin mutants with different spectral characteristics are not available to help understand the effects of the mutations precisely. Although there are a few crystal structures available for microbial rhodopsins, still the system is not ideal for studying the effects of protein-chromophore interactions, due to lack of crystal structures of the mutants.

I.5 Wavelength regulation due to conformational change

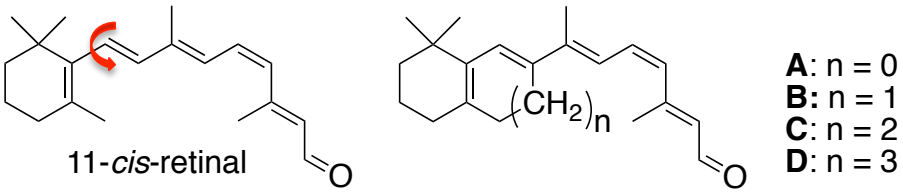
The conformation of the chromophore has been suggested to play an important role in the different absorption maxima of retinal-PSB observed in rhodopsins. The more planar the chromophore is, the better π-orbital overlap can be achieved and the more red shifted the absorption will be. On the contrary, twisting along the single bonds is expected to reduce the degree of conjugation achieved and lead to blue shift. Due to different protein-chromophore packing, different conformations could be obtained in different rhodopsin systems. However, the crystal structures of the color rhodopsins are not available to

compare the conformations of the chromophore in different rhodopsins and assign its contribution to the spectral differences.

One of the major conformational differentiations is believed to occur along the C6-C7 single bond. For the 6-s-*cis* conformer, due to steric hindrance between C5-methyl and C8-H, the plane of the ionone ring double bond is twisted from the plane of the polyene, resulting in less conjugation than the 6-s-*trans* conformer. Bovine rhodopsin adopts a 6-s-*cis* conformation, while the microbial rhodopsins adopt a 6-s-*trans* conformation. This was proposed to be the reason why microbial rhodopsins are usually more red shifted than visual rhodopsins, due to better planarity along the C6-C7 single bond.

Different ring-locked retinal analogues were studied in bovine rhodopsin

Table I-6: Characterization of different ring-locked retinal analogues.



Chromophore	λ_{max} (nm)			Opsin shift (cm^{-1})
	aldehyde	PSB	Rhodopsin	
A	422	506	539	1200
B	416	496	546	1850
C	386	457	503	2000
D	374	440	483	2000
11-*cis*-retinal	377	440	498	2650

systems.[67] As shown in **Table I-6**, the planarization along C6-C7 is important for red shift, when comparing the absorptions of the aldehyde and the PSB formed in solution for the ring-locked compounds with the unlocked. The ring locked analogues **A** and **B**, which are most planar, exhibit the most red shifted absorption in solution. However, they also resulted in the least opsin shift when bound to rhodopsin, which suggested that part of the opsin shift observed in rhodopsin bound with 11-cis-retinal, results from planarization of the 11-*cis*-retinal due to protein-chromophore packing. That is why, for the analogues **A** and **B**, which are already planar in solution, less planarization resulted and a smaller opsin shift was observed.

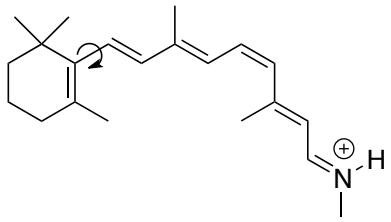
Garavelli's group studied the absorptions of all-*trans*-retinal-PSB and a few retinal analog PSB in vacuum, showing that the 6-s-*trans* conformer of retinal-PSB in vacuum absorbs at 610 nm and the 6-s-*cis* conformer absorbs at 530 nm. The drastic difference between the absorptions of 6-s-*cis* and 6-s-*trans* is due to the highly twisted conformation along C6-C7.[41]

It is clear that the conformation of the chromophore is important in the study of protein-chromophore interactions. Therefore, crystal structures of all the protein-chromophore complexes should be obtained to dissect the contribution of conformational change in wavelength tuning.

I.6 Modern computational studies on wavelength regulation

Different hypotheses regarding wavelength regulation contribute greatly to the understanding of spectral tuning, but still no single conclusion can be reached to fully explain wavelength regulation. With the recent developments in computational simulation and methods, there is a great interest in modeling wavelength regulation, in order to qualify or even quantify the absorption maxima, given the crystal structures of the protein-chromophore complexes.

Sakurai's group tried to understand the 10 nm red shift observed in halorhodopsin upon chloride binding, which was counter-intuitive in the first place, as introduction of a negative charge in the PSB region could better stabilize the PSB and lead to a blue shift.[68] For the calculation, they set ionizable groups in the protein interior to be charged if ion pair formation is possible and all the ionizable groups exposed to aqueous medium to be neutral, as the dielectric screening effect of the solvent cancels out the charges of the ionized residues. If only the electronic state of the chromophore was considered, a blue shift resulted from the calculation. However, after taking into account the induced polarization of the protein environment as a result of the positive charge transfer of the chromophore upon electronic excitation to the excited state, a red shift was observed in the calculation. This was a demonstration of the decisive role of electronic polarization in wavelength regulation.



$\theta = 0°$ λ_{max} = 673 nm (1.84 eV)
$\theta = 68°$ λ_{max} = 545 nm (2.27 eV)
$\theta = 90°$ λ_{max} = 535 nm (2.31 eV)

Figure I-17: Computational analysis of gas phase 11-*cis*-retinal-PSB absorption with different torsion of the C6-C7 bond as highlighted in the picture.

Olivucci and Garavelli computed the vertical excitation energy of 11-*cis*-retinal-PSB from S_0 to S_1 in gas phase and predicted the absorption maximum to be 545 nm (2.48 eV) with the ionone ring 68° twisted in the 6-s-*cis* conformation.[69] They also calculated the absorption of 11-*cis*-retinal PSB if the θ equals 90°, when the ionone ring double bond is completely out of the plane of the polyene and has no overlap with the conjugated π system of the polyene. The value of 535 nm matches very well with the experimental data using 5,6-dihydroretinal-PSB in gas phase obtained recently.[67] Accordingly, they suggested that in red and green rhodopsin the chromophore was more planar, while in blue rhodopsin the chromophore was highly twisted along the C6-C7 single bond.

The suggestion of a highly twisted angle along C6-C7 in blue Rhodpsin was reiterated by Trabanino and coworkers in 2006.[70] They also suggested that the conformational twisting of the 11-*cis*-retinal PSB was the major factor causing the blue shift observed between green and blue rhodopsins. In addition, dipolar

side chains in the binding pocket contributed to the red shift observed for red rhodopsin compared to green rhodopsin.

The retinal-PSB absorption in gas phase (610 nm) set a new starting point for computational studies. Volker Buss and coworkers performed the calculation of rhodopsin absorptions starting from retinal-PSB in vacuum and began building in the counteranion, followed by the surrounding protein residues.[71] The calculation showed that introduction of the counteranion contributed the most blue shift from 610 nm in vacuum to 486 nm, close to the absorption of rhodopsin. Addition of the protein surrounding residues into the calculation led to a slight red shift through dipolar interactions with the chromophore. Conformational change of the chromophore was calculated to play a minor role in wavelength tuning in rhodopsin. Therefore, they claimed that it is not the binding pocket, but the counteranion that is dictating the absorption maxima of retinal-PSB in rhodopsin. This statement is questionable, as mutagenesis of the counteranion in bovine rhodopsin resulted in a 30 nm red shift.[59]

Olivucci and Ryde performed calculations at the CASPT2//CASSCF level (second-order multiconfigurational perturbation theory) for WT rhodopsin and two mutants and isorhodopsin in QM/MM structures based on two crystal structures.[72] Although they showed the importance of the dipoles or polarizable residues in affecting the electronic characteristics of the chromophore, they also showed that different calculations models (Amber 1994 and 2003 force fields)

could predict different results, up to 16 KJ/mol (~1300 cm^{-1}) differences in excitation energy. The calculation of retinal-PSB absorption in the protein environment is still difficult, as various factors, could strongly perturb the electronic profile of retinal-PSB. A slight change in the conformation of the side chains in the protein cavity or the chromophore might result in significantly different values.

In conclusion, retinal-PSB is a complicated system, especially inside a protein environment, since the dynamics of the surrounding residues and also polarizability of the protein binding pocket are critical, but difficult at this stage to consider computationally. Since crystal structures of rhodopsins available are limited, it is hard to compare rhodopsin systems with only 70% homology. Computational studies on protein systems with only a few amino acid differences, but with well differentiated spectra, could greatly simplify the computational problem and provide a platform to develop methods and probe the important factors for quantifying the electronic excitation energy.

I.7 Strategies for spectral tuning in a rhodopsin mimic

Although there have been a large number of studies on model compounds, rhodopsin mutageneses, and computational investigations on wavelength regulation of retinal-PSB, the factors that are crucial in spectral tuning have remained largely unknown. This is mainly due to the lack of crystal

structures of rhodopsins with various absorption profiles and the mutants of rhodopsins designed to study wavelength regulation.

As a membrane bound protein, rhodopsins are hard to express and crystallize, which greatly hampers the study of wavelength regulation. Rhodopsins are also highly sensitive to light, making their preparation even more tedious. A more flexible strategy is to find a small cellular protein and engineer it into a rhodopsin mimic that can bind retinal as a protonated Schiff base. Such a system might offer a better platform to study wavelength regulation, with easily crystallized mutants. In addition, if it is possible to recapitulate the wavelength regulation observed in rhodopsins with a completely heterologous protein, the critical factors for spectral tuning can be conclusively mapped out.

Based on the previous studies on wavelength regulation of retinal-PSB, two important electronic issues were considered for wavelength tuning in the rhodopsin mimic. One of them is the different charge distribution in the ground state and excited state. This characteristic is important for design and analysis of the electrostatic interactions between the protein and the retinal-PSB as it absorbs light.

Figure I-18a shows the electronic profile of the all-*trans*-retinylidene carbons in the ground state. C15, the carbon directly linked to the iminium nitrogen has the largest concentration of positive charge. Generally speaking, carbons with odd numbers **(Figure I-18a)** have positive charge character, while carbons with even numbers are slightly negative. This is due to the resonance

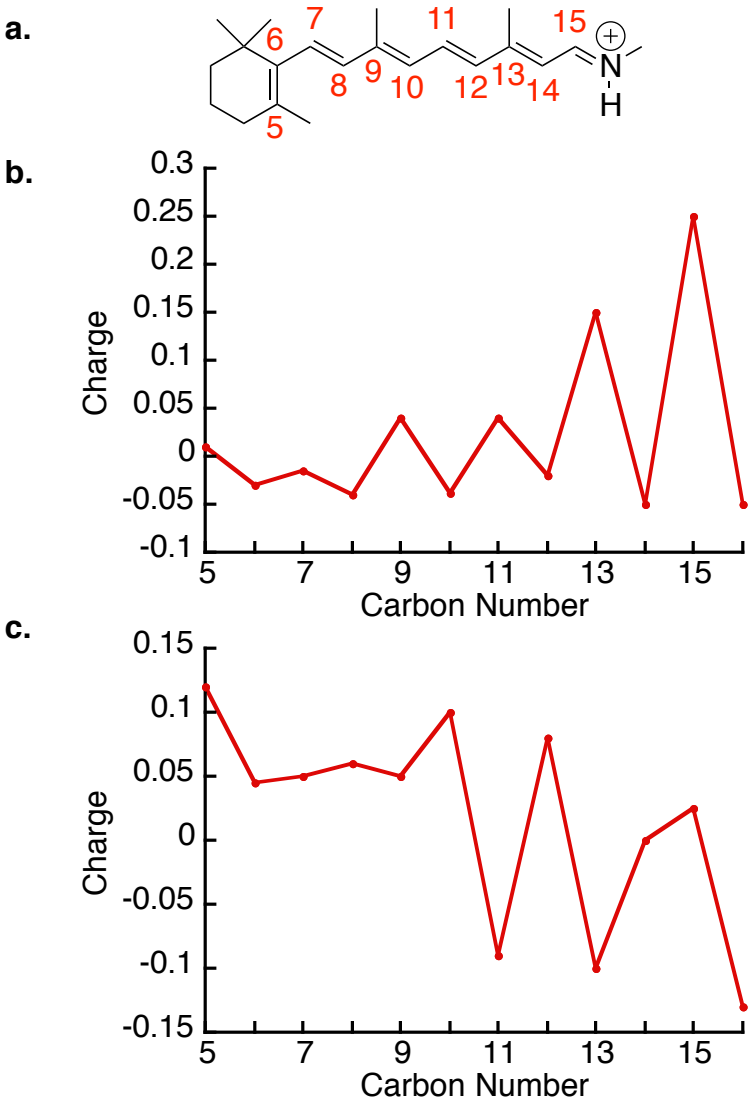


Figure I-18: Charge distribution on retinal-PSB. **a**. All-*trans*-retinal PSB with the conjugated polyene carbons numbered. Calculated charged state of carbon atom in all-*trans*-retinal-PSB ground state (**b**) and excited state (**c**).[3]

structure of retinylidene to delocalize the positive charge on the iminium. The amount of positive charge decreases as the carbon is further away from the iminium.

In the excited state, the electronic picture is very different (**Figure I-18c**). The positive charge is transferred exclusively toward the other end of the retinal-

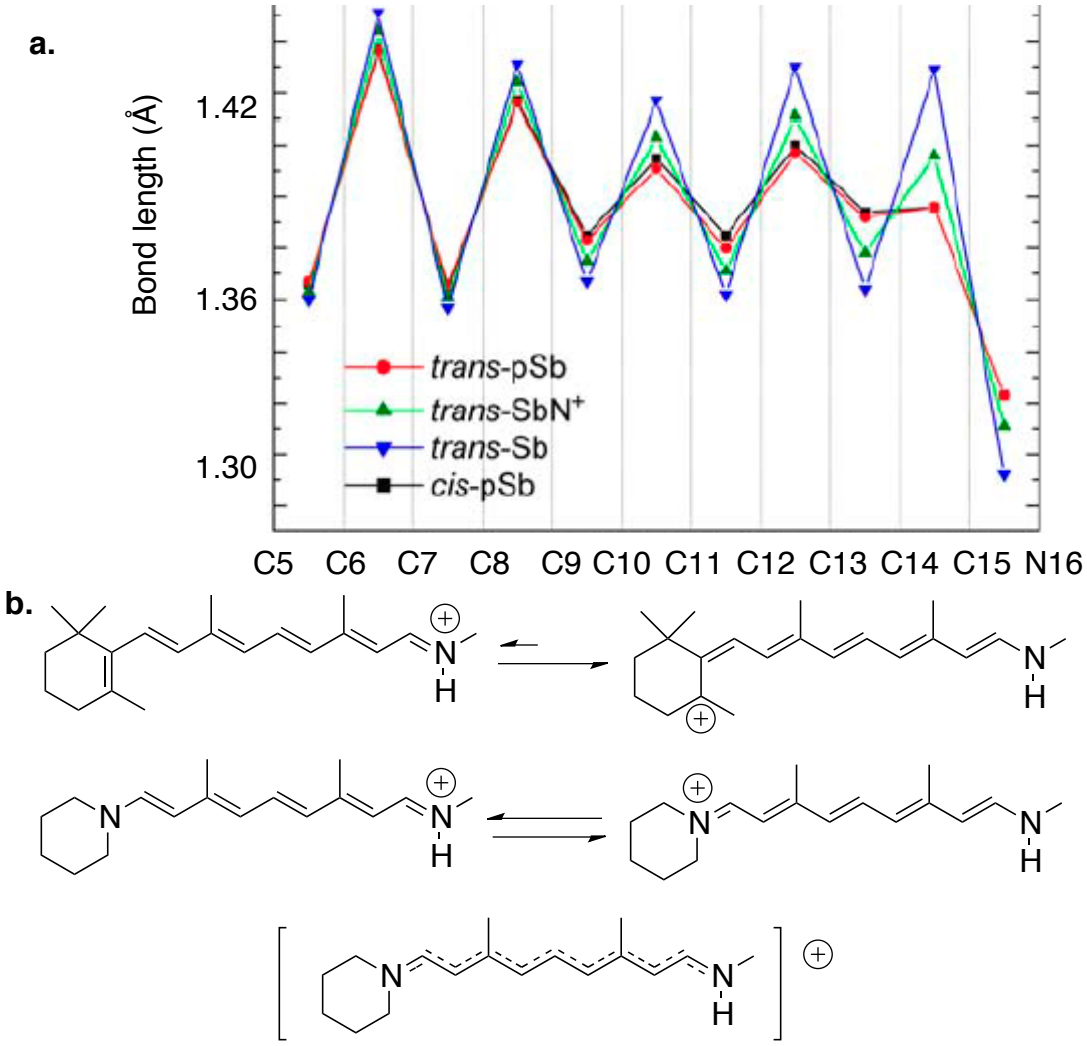


Figure I-19: **a.** Calculated bond distance for different retinal compounds.[3] **b.** Resonance structures of retinal-PSB and cyanine dye. Cyanine dyes are highly delocalized with equal contribution from both resonance structures.

PSB, with positive charge on carbons 5 to 9, while a negative charge on C11 and C13 and the nitrogen.

To decrease the vertical energy gap between the ground state and excited state, one can increase the negative dipoles in the ionone ring region to interact favorably with the positively charged carbons towards the ionone ring in the excited state. Attemptively, one can decrease the negative dipoles in the PSB

region to interact unfavorably with the positively charged carbons towards the iminium end in the ground state.

The second characteristic of the retinal-PSB is their alternating single bond/ double bond character. Buss' group performed calculations on the bond distance of all-*trans*-retinal SB and PSB in solution, as well as the retinal-PSB in rhodopsin. As shown in **Figure I-19a**, SB, which is most blue shifted, has the most obvious alternating single-double bond character, evident from the bond distance of single bonds and double bonds. When the Schiff base gets protonated, it red shifts from 360 nm to 440 nm and the single/double bond alternation becomes less significant than that in SB, concluded from the bond distances up to C9-C10. For rhodopsin, the alternating single/double bond character is least distinguished, resulting in the largest red shift.

The diminishing character of alternating single-double bond is a result of charge delocalization through resonance structures. As shown in **Figure 19b**, resonance structures of retinal-PSB shows that the more stable the alternate resonance structure of the retinal-PSB, the more delocalized the positive charge will be and as a result, a smaller degree of single bond/double bond alternation will be observed. The extreme case is cyanine dyes, where the alternate resonance structure is almost as stable as the original structure with the positive charge localized on the iminium (**Figure 19b**). As a result, the positive charge is fully delocalized along the conjugated system and little alternate single-double bond character is observed. As expected, a high degree of red shift results as

well; consistent with the fact that cyanine dyes are usually much more red shifted than retinal, with similar number of double bonds.[73]

Therefore, in order to cause red shift, one could try to delocalize the positive charge along the polyene by stabilizing the alternate resonance structures through favorable electrostatic interactions. This could be accomplished by positioning negative dipoles close to the ionone ring region of the chromophore.

The majority of wavelength regulation studies on rhodopsin, model compound studies and recent computational studies seem to suggest the same idea: the overall electrostatic interaction with the chromophore is critical in spectral tuning. This can be due to either differential electrostatic interactions with the ground state and excited state of the chromophore or through charge delocalization by electrostatic interactions that stabilize the alternate resonance structures.

Matthias Ullmann and his coworkers provided a straightforward calculation to qualify the interaction of the protein with the chromophore.[74] They performed calculations to obtain the overall electrostatic potential projected on the Van der Waals surface of the chromophore to read out the spectral tuning for sensory rhodopsin II (500 nm), bacteriorhodopsin (560 nm) and halorhodpsin (570 nm). Simply put, they took the partial charge from each point of the protein and projected the electrostatic field on the Van der Waals surface of the chromophore. The result is the sum of the electric field from the protein, and

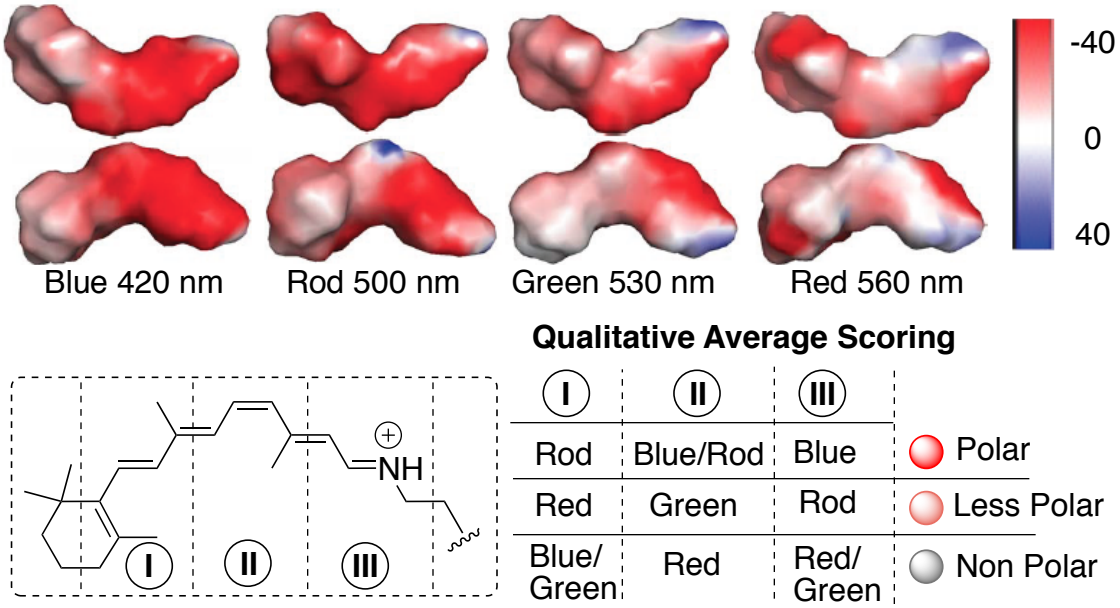


Figure I-20: Electrostatic potential calculation. Electrostatic potential calculation (APBS suite) of blue, rod, green, and red opsin (calculations of blue, green, and red are based on available homology models) projected on the Van der Waals surface of retinal (the electrostatic potential of the retinylidene chromophore was set to zero as to only illustrate the electrostatic contribution of the protein).[1] The chromophore is divided into three segments; the qualitative average score for each segment represents the overall electrostatic potentials that lead to the wavelength regulation of each opsin.

generally describes the electrostatic interaction between the chromophore and the protein environment. However, a requirement for this calculation is to have a high resolution crystal structure of the retinal-bound protein, as the conformation of each side chain is critical in the read out of the electrostatic potential projected on the surface of the chromophore.

In our lab, Dr. Lee performed similar calculations on the three color rhodopsin models and bovine rhodopsin crystal structures.[1] Consistent results were obtained for the color rhodopsins as shown in **Figure I-20**. A remarkable correlation between absorption spectrum and the electrostatic potential felt by the

chromophore can be readily discerned. While the blue rhodopsin clearly has the most negative electrostatic potential near the Schiff base, as expected (local stabilization of iminium ion), it also has the least negative potential on the ionone ring side of the chromophore. On the other hand, the electrostatic potential around the Schiff base becomes less negative for red and green opsins, while it becomes more negative on the ionone ring side as the wavelength becomes longer. This indicates that relatively subtle local changes in the electrostatic potential can have a significant effect on the absorption profile.

This electrostatic calculation will be used as a general guide and for analysis of the spectral tuning in the future wavelength regulation studies on rhodopsin mimic.

I.8 Understanding pK_a regulation of retinal-PSB

Another interesting fact about the protonated Schiff base in a rhodopsin system is their extremely high pK_a values observed for the iminium (**Figure I-21**). The pK_a for the retinal-PSB is estimated to be higher than 16 in bovine rhodopsin and ~12 and ~13 respectively for sensory rhodopsin II and

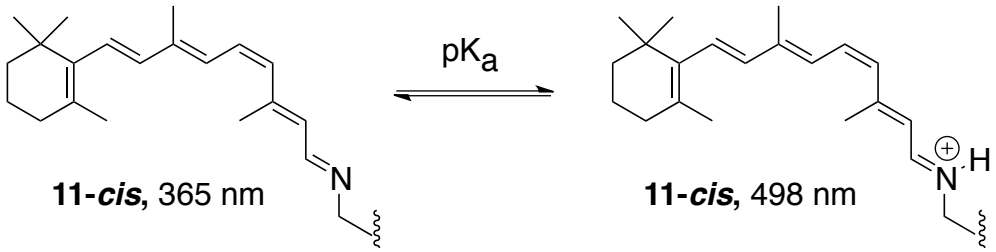


Figure I-21: Absorption of Schiff base and protonated Schiff base of 11-*cis*-retinal.

bacteriorhodopsin,[75, 76] while the pK_a for retinal-PSB with *n*-Butylamine in buffer is ~6.5. Considering that the rhodopsin binding pocket is relatively hydrophobic, charge formation could be suppressed. It is interesting to understand what causes this dramatic increase of pK_a in rhodopsins.

Even before the rhodopsin crystal structure was available, it was realized that a counter anion should be present in the rhodopsin to stabilize the positive charge on iminium through salt bridge formation and charge compensation. Scheiner showed that in order for the retinal-PSB and carboxylic acid pair to form a salt bridge, a polar environment was important due to better solvation of the ion pair.[77]

The trajectory of the counteranion carboxylic acid toward the PSB is critical as well as the electronic effects of the protein environment. Sheves showed that a rigidified angle between the carboxylic acid and the retinal-imine could have an effect on the pK_a of retinal-PSB through model structure studies.[78] Sheves also showed through model compounds that by placing a non-conjugated positive charge in close proximity to the PSB, the stability of the protonated state of the PSB could be disturbed and the pK_a would decrease.[44]

Through calculations on a few simple model structures, Sheves, et al. concluded that the best trajectory to stabilize the positive charge on an iminium was to place the carboxylic acid on the same plane as the iminium and have a CNO angle of 108° (from the imine C-N to the oxygen of the carboxylic acid,

illustrated in **Figure I-22a**).[79] In other words, the protonated state of the iminium is most preferred when the imininium proton is placed along the axis of the negative dipole of the counteranion. This trajectory makes sense intuitively as well. In this way, there will be optimal hydrogen bonding interactions between the iminium hydrogen and both oxygens of the carboxylic acid.

The bovine rhodopsin crystal structure shows that the counteranion adopts a similar conformation to the suggested optimal trajectory toward the iminium. As shown in **Figure I-22b**, the carboxylic acid of E113 is almost on the same plane as that of the imine and has a CNO angle of 104°, so that both of the carboxylic oxygens from E113 could have optimal interactions with the retinal-PSB. In addition, the dipole of the carboxylic is exactly pointing towards where the iminium hydrogen is supposed to be. Besides, Glu113 was also stabilized by other hydrogen bonding interactions as shown in **Figure I-22b**.

The crystal structure of bacteriorhodopsin showed structured a hydrogen bonding network in the PSB region of the binding pocket as shown in **Figure I-**

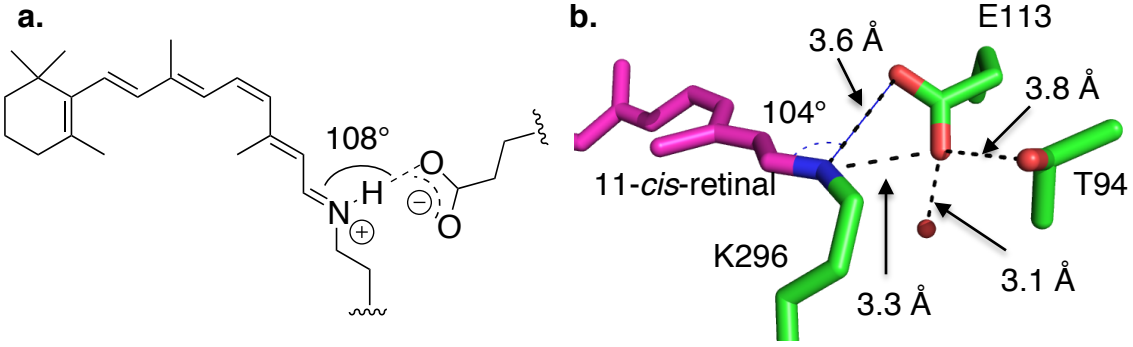


Figure I-22: The importance of counteranion position. **a.** Illustration of CNO angle for optimal pK_a of retinal-PSB. **b.** Zoom in of the interaction between E113 and retinal-PSB in bovine rhodopsin crystal structure.

23. Different from bovine rhodopsin, where there is a direct interaction between the counteranion and the PSB, in bacteriorhodopsin, the counteranion interacts with the PSB indirectly via a water molecule. Although there are two acidic residues close to the PSB, D85 was found to be the major counteranion, as mutagenesis at only position 85 caused a red shift and a dramatic decrease of the pK_a (but not D212).

This could be understood from the hydrogen bonding trajectory and the dipole of the two acidic residues, as the dipole of D85 was pointing towards where the iminium hydrogen is, while D212 was pointing away from the imine. In addition, D212 could not form good hydrogen bonding with the bound water molecule, as the water molecule is not on the same plane of the carboxylate of Asp212.

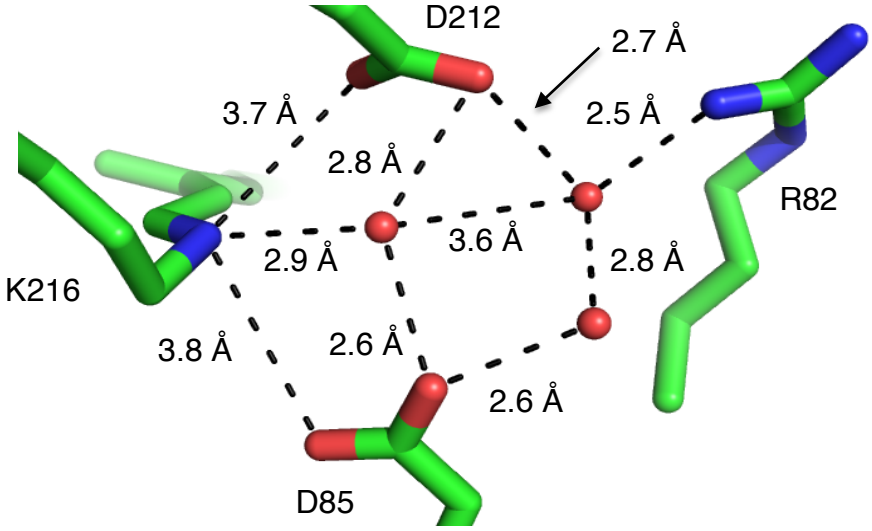


Figure I-23: Hydrogen bonding network in the PSB region of bacteriorhodopsin. The red dots represent water molecules (PDB etnry: 1C3W).

References

References

1. Lee, K. S. S. Ph.D Thesis, Michigan State University, East Lansing, MI, 2011.

2. Shichida, Y.; Matsuyama, T., Evolution of opsins and phototransduction. *Phil. Trans. R. Soc. B* **2009,** *364* (1531), 2881-2895.

3. Sekharan, S.; Weingart, O.; Buss, V., Ground and excited states of retinal schiff base chromophores by multiconfigurational perturbation theory. *Biophys. J.* **2006,** *91* (1), L7-L9.

4. 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. *J. Bacteriol.* **2010,** *192* (22), 5866-5873.

5. Wald, G., The photochemistry of vision. *Doc. Ophthalmol.* **1949,** *3*, 94-137.

6. (A) Hubbard, R., *Cis-trans* isomers of vitamin-a and retinene in the rhodopsin system. *Fed. Proc.* **1952,** *11* (1), 233-233; (B) Hubbard, R.; Wald, G., *Cis-trans* isomers of vitamin-a and retinene in the rhodopsin system. *J. Gen. Physiol.* **1952,** *36* (2), 269-315; (C) Hubbard, R.; Wald, G., *Cis-trans* isomers of vitamin-a and retinene in vision. *Science* **1952,** *115* (2977), 60-63; (D) Dieterle, J. M.; Robeson, C. D., Crystalline neoretinene-B. *Science* **1954,** *120* (3110), 219-220.

7. Wang, J. K.; Mcdowell, J. H.; Hargrave, P. A., Site of attachment of 11-*cis*-retinal in bovine rhodopsin. *Biochemistry* **1980,** *19* (22), 5111-5117.

8. Wald, G., The biochemistry of vision. *Annu. Rev. Biochem.* **1953,** *22*, 497-526.

9. Mustafi, D.; Engel, A. H.; Palczewski, K., Structure of cone photoreceptors. *Prog. Retin. Eye Res.* **2009,** *28* (4), 289-302.

10. Brown, P. K.; Wald, G., Visual pigments in single rods + cones of human retina - direct measurements reveal mechanisms of human night + color vision. *Science* **1964,** *144* (361), 45.

11. Waddell, W. H.; Crouch, R.; Nakanishi, K.; Turro, N. J., Quantitative aspects of photochemistry of isomeric retinals and visual pigments. *J. Am. Chem. Soc.* **1976,** *98* (14), 4189-4192.

12. Stryer, L.; Hurley, J. B.; Fung, B. K. K., Transducin - An amplifier protein in Vision. *Trends Biochem. Sci.* **1981,** *6* (9), 245-247.

13. Gilman, A. G., G-Proteins - Transducers of receptor-generated signals. *Annu. Rev. Biochem.* **1987,** *56*, 615-649.

14. 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 Biochem. Sci.* **2009,** *34* (11), 540-552.

15. (A) Vuong, T. M.; Chabre, M.; Stryer, L., Millisecond activation of transducin in the cyclic-nucleotide cascade of vision. *Nature* **1984,** *311* (5987), 659-661; (B) Guy, P. M.; Koland, J. G.; Cerione, R. A., Rhodopsin-stimulated activation deactivation cycle of transducin - kinetics of the intrinsic fluorescence response of the alpha-subunit. *Biochemistry* **1990,** *29* (30), 6954-6964.

16. Graykeller, M. P.; Detwiler, P. B., The calcium feedback signal in the phototransduction cascade of vertebrate rods. *Neuron* **1994,** *13* (4), 849-861.

17. (A) Eigenthaler, M.; Nolte, C.; Halbrugge, M.; Walter, U., Concentration and regulation of cyclic-nucleotides, cyclic-nucleotide-dependent protein-kinases and one of their major substrates in human platelets - estimating the rate of camp-regulated and cgmp-regulated protein-phosphorylation in intact-cells. *Eur. J. Biochem.* **1992,** *205* (2), 471-481; (B) Miki, N.; Keirns, J. J.; Marcus, F. R.; Freeman, J.; Bitensky, M. W., Regulation of cyclic nucleotide concentrations in photoreceptors - ATP-dependent stimulation of cyclic nucleotide phosphodiesterase by light. *P. Natl. Acad. Sci. Usa* **1973,** *70* (12), 3820-3824; (C) Lamb, T. D., Gain and kinetics of activation in the g-protein cascade of phototransduction. *P. Natl. Acad. Sci. Usa* **1996,** *93* (2), 566-570.

18. Salon, J. A.; Lodowski, D. T.; Palczewski, K., The significance of g protein-coupled receptor crystallography for drug discovery. *Pharmacol. Rev.* **2011,** *63* (4), 901-937.

19. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L., Opinion - How many drug targets are there? *Nat. Rev. Drug Discov.* **2006,** *5* (12), 993-996.

20. Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M., Crystal structure of rhodopsin: a g protein-coupled receptor. *Science* **2000,** *289* (5480), 739-745.

21. Nakanishi, K., Recent bioorganic studies on rhodopsin and visual transduction. *Chem. Pharm. Bull.* **2000,** *48* (10), 1399-1409.

22. (A) Doukas, A. G.; Junnarkar, M. R.; Alfano, R. R.; Callender, R. H.; Kakitani, T.; Honig, B., Fluorescence quantum yield of visual pigments - evidence for subpicosecond isomerization rates. *P. Natl. Acad. Sci-Biol.* **1984,** *81* (15), 4790-4794; (B) Buchert, J.; Stefancic, V.; Doukas, A. G.; Alfano, R. R.; Callender, R. H.; Pande, J.; Akita, H.; Baloghnair, V.; Nakanishi, K., Picosecond absorption and fluorescence kinetic-studies of bovine rhodopsin with a fixed 11-ene. *Biophys. J.* **1983,** *41* (2), A338-A338.

23. Nakamichi, H.; Okada, T., Crystallographic analysis of primary visual photochemistry. *Angew. Chem. Int. Ed. Engl.* **2006,** *45* (26), 4270-4273.

24. 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-U137.

25. Standfuss, J.; Edwards, P. C.; D'antona, A.; Fransen, M.; Xie, G.; Oprian, D. D.; Schertler, G. F. X., The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **2011,** *471* (7340), 656-660.

26. Imai, H.; Terakita, A.; Tachibanaki, S.; Imamoto, Y.; Yoshizawa, T.; Shichida, Y., Photochemical and biochemical properties of chicken blue-sensitive cone visual pigment. *Biochemistry* **1997,** *36* (42), 12773-12779.

27. Travis, G. H.; Golczak, M.; Moise, A. R.; Palczewski, K., Diseases caused by defects in the visual cycle: retinoids as potential therapeutic agents. *Annu. Rev. Pharmacol. Toxicol.* 2007; Vol. 47, Pp 469-512.

28. Moiseyev, G.; Crouch, R. K.; Goletz, P.; Oatis, J.; Redmond, T. M.; Ma, J. X., Retinyl esters are the substrate for isomerohydrolase. *Biochemistry* **2003,** *42* (7), 2229-2238.

29. Kiser, P. D.; Golczak, M.; Lodowski, D. T.; Chance, M. R.; Palczewski, K., Crystal structure of native rpe65, the retinoid isomerase of the visual cycle. *P. Natl. Acad. Sci. Usa* **2009,** *106* (41), 17325-17330.

30. Wald, G., Human vision and the spectrum. *Science* **1945,** *101* (2635), 653-658.

31. Kleinschmidt, J.; Harosi, F. I., Anion sensitivity and spectral tuning of cone visual pigments in situ. *P. Natl. Acad. Sci. Usa* **1992,** *89* (19), 9181-9185.

32. Suzuki, T.; Makinotasaka, M., Analysis of retinal and 3-dehydroretinal in the retina by high-pressure liquid-chromatography. *Anal. Biochem.* **1983,** *129* (1), 111-119.

33. Tsin, A. T. C.; Liebman, P. A.; Beatty, D. D.; Drzymala, R., Rod and cone visual pigments in the goldfish. *Vision. Res.* **1981,** *21* (6), 943-946.

34. Crescitelli, F., The natural-history of visual pigments - 1990. *Prog. Retin. Res.* **1991,** *11*, 1-32.

35. Tsin, A. T. C.; Alvarez, R. A.; Fong, S. L.; Bridges, C. D. B., Conversion of retinol to 3,4-didehydroretinol in the tadpole. *Comp. Biochem. Phys. B* **1985,** *81* (2), 415-421.

36. Motto, M. G.; Sheves, M.; Tsujimoto, K.; Baloghnair, V.; Nakanishi, K., Opsin shifts in bovine rhodopsin and bacteriorhodopsin - comparison of 2 external point-charge models. *J. Am. Chem. Soc.* **1980,** *102* (27), 7947-7949.

37. Blatz, P. E.; Baumgart.N; Balasubr.V; Balasubr.P; Stedman, E., Wavelength regulation in visual pigment chromophore - large induced bathocromic shifts in retinol and related polyenes. *Photochem. Photobiol.* **1971,** *14* (4), 531-&.

38. Blatz, P. E.; Mohler, J. H.; Navangul, H. V., Anion-induced wavelength regulation of absorption maxima of schiff-bases of retinal. *Biochemistry* **1972,** *11* (5), 848-&.

39. Honig, B.; Greenberg, A. D.; Dinur, U.; Ebrey, T. G., Visual-pigment spectra - implications of protonation of retinal schiff-base. *Biochemistry* **1976,** *15* (21), 4593-4599.

40. 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. *J. Am. Chem. Soc.* **2005,** *127* (35), 12347-12350.

41. 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. *Angew. Chem. Int. Ed. Engl.* **2010,** *49* (10), 1790-1793.

42. Livnah, N.; Sheves, M., Model compounds can mimic spectroscopic properties of bovine rhodopsin. *J. Am. Chem. Soc.* **1993,** *115* (1), 351-353.

43. Rosenber.B; Krigas, T. M., Spectral shifts in retinal schiff base complexes. *Photochem. Photobiol.* **1967,** *6* (10), 769-&.

44. Baasov, T.; Sheves, M., Alteration of pka of the bacteriorhodopsin protonated schiff-base - a study with model compounds. *Biochemistry* **1986,** *25* (18), 5249-5258.

45. Honig, B.; Dinur, U.; Nakanishi, K.; Baloghnair, V.; Gawinowicz, M. A.; Arnaboldi, M.; Motto, M. G., External point-charge model for wavelength regulation in visual pigments. *J. Am. Chem. Soc.* **1979,** *101* (23), 7084-7086.

46. Nakanishi, K.; Baloghnair, V.; Arnaboldi, M.; Tsujimoto, K.; Honig, B., An external point-charge model for bacteriorhodopsin to account for its purple color. *J. Am. Chem. Soc.* **1980,** *102* (27), 7945-7947.

47. Beppu, Y.; Kakitani, T., Theoretical-study of color control mechanism in retinal proteins .1. role of the tryptophan residue, tyrosine residue and water molecule. *Photochem. Photobiol.* **1994,** *59* (6), 660-669.

48. Luecke, H.; Schobert, B.; Lanyi, J. K.; Spudich, E. N.; Spudich, J. L., Crystal structure of sensory rhodopsin ii at 2.4 angstroms: insights into color tuning and transducer interaction. *Science* **2001,** *293* (5534), 1499-1503.

49. Luecke, H.; Schobert, B.; Richter, H. T.; Cartailler, J. P.; Lanyi, J. K., Structure of bacteriorhodopsin at 1.55 angstrom resolution. *J. Mol. Biol.* **1999,** *291* (4), 899-911.

50. (A) Irving, C. S.; Byers, G. W.; Leermake.Pa, Effect of solvent polarizability on absorption spectrum of all-trans-retinylpyrrolidiniminium perchlorate. *J. Am. Chem. Soc.* **1969,** *91* (8), 2141-&; (B) Irving, C. S.; Byers, G. W.; Leermake.Pa, Spectroscopic model for visual pigments - influence of microenvironmental polarizability. *Biochemistry* **1970,** *9* (4), 858-&.

51. Mathies, R.; Stryer, L., Retinal has a highly dipolar vertically excited singlet-state - implications for vision. *P. Natl. Acad. Sci. Usa* **1976,** *73* (7), 2169-2173.

52. Baasov, T.; Sheves, M., Model compounds for the study of spectroscopic properties of visual pigments and bacteriorhodopsin. *J. Am. Chem. Soc.* **1985,** *107* (25), 7524-7533.

53. Nathans, J.; Thomas, D.; Hogness, D. S., Molecular-genetics of human color-vision - the genes encoding blue, green, and red pigments. *Science* **1986,** *232* (4747), 193-202.

54. Neitz, M.; Neitz, J.; Jacobs, G. H., Spectral tuning of pigments underlying red-green color-vision. *Science* **1991,** *252* (5008), 971-974.

55. Asenjo, A. B.; Rim, J.; Oprian, D. D., Molecular determinants of human red/green color discrimination. *Neuron* **1994,** *12* (5), 1131-1138.

56. Lin, S. W.; Kochendoerfer, G. G.; Carroll, H. 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. *J. Biol. Chem.* **1998,** *273* (38), 24583-24591.

57. Nakayama, T. A.; Zhang, W. M.; Cowan, A.; Kung, M., Mutagenesis studies of human red opsin: trp-281 is essential for proper folding and protein-retinal interactions. *Biochemistry* **1998,** *37* (50), 17487-17494.

58. Zhukovsky, E. A.; Oprian, D. D., Effect of carboxylic-acid side-chains on the absorption maximum of visual pigments. *Science* **1989,** *246* (4932), 928-930.

59. 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. *P. Natl. Acad. Sci. Usa* **1991,** *88* (8), 3079-3083.

60. (A) Lin, J. Y., A User's guide to channelrhodopsin variants: features, limitations and future developments. *Exp. Physiol.* **2011,** *96* (1), 19-25; (B) Suzuki, D.; Irieda, H.; Homma, M.; Kawagishi, I.; Sudo, Y., Phototactic and chemotactic signal transduction by transmembrane receptors and transducers in microorganisms. *Sensors-Basel* **2010,** *10* (4), 4010-4039; (C) Purcell, E. B.; Crosson, S., Photoregulation in prokaryotes. *Curr. Opin. Microbiol.* **2008,** *11* (2), 168-178; (D) Sasaki, J.; Spudich, J. L., Signal transfer in haloarchaeal sensory rhodopsin-transducer complexes. *Photochem. Photobiol.* **2008,** *84* (4), 863-868.

61. Kolbe, M.; Besir, H.; Essen, L. O.; Oesterhelt, D., Structure of the light-driven chloride pump halorhodopsin at 1.8 angstrom resolution. *Science* **2000,** *288* (5470), 1390-1396.

62. (A) Vogeley, L.; Sineshchekov, O. A.; Trivedi, V. D.; Sasaki, J.; Spudich, J. L.; Luecke, H., Anabaena sensory rhodopsin: a photochromic color 0 sensor at 2.0 angstrom. *Science* **2004,** *306* (5700), 1390-1393; (B) Pebaypeyroula, E.; Rummel, G.; Rosenbusch, J. P.; Landau, E. M., X-Ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science* **1997,** *277* (5332), 1676-1681.

63. Mogi, T.; Marti, T.; Khorana, H. G., Structure-function studies on bacteriorhodopsin .9. substitutions of tryptophan residues affect protein-retinal interactions in bacteriorhodopsin. *J. Biol. Chem.* **1989,** *264* (24), 14197-14201.

64. Greenhalgh, D. A.; Farrens, D. L.; Subramaniam, S.; Khorana, H. G., Hydrophobic amino-acids in the retinal-binding pocket of bacteriorhodopsin. *J. Biol. Chem.* **1993,** *268* (27), 20305-20311.

65. Subramaniam, S.; Marti, T.; Khorana, H. G., Protonation state of asp (glu)-85 regulates the purple-to-blue transition in bacteriorhodopsin mutants arg-82- ala and asp-85- glu - the blue form is inactive in proton translocation. *P. Natl. Acad. Sci. Usa* **1990,** *87* (3), 1013-1017.

66. Shimono, K.; Iwamoto, M.; Sumi, M.; Kamo, N., Effects of three characteristic amino acid residues of pharaonis phoborhodopsin on the absorption maximum. *Photochem. Photobiol.* **2000,** *72* (1), 141-145.

67. Wada, A.; Tsutsumi, M.; Inatomi, Y.; Imai, H.; Shichida, Y.; Ito, M., Retinoids and related compounds. part 26. synthesis of (11 z)-8,18-propano- and methano-retinals and conformational study of the rhodopsin chromophore. *J. Chem. Soc. Perk. T. 1* **2001,** (19), 2430-2439.

68. Sakurai, M.; Sakata, K.; Saito, S.; Nakajima, S.; Inoue, Y., Decisive role of electronic polarization of the protein environment in determining the absorption maximum of halorhodopsin. *J. Am. Chem. Soc.* **2003,** *125* (10), 3108-3112.

69. Cembran, A.; Gonzalez-Luque, R.; Altoe, P.; Merchan, M.; Bernardi, F.; Olivucci, M.; Garavelli, M., Structure, spectroscopy, and spectral tuning of the gas-phase retinal chromophore: the beta-ionone "handle" and alkyl group effect. *J. Phys. Chem. A* **2005,** *109* (29), 6597-6605.

70. Trabanino, R. J.; Vaidehi, N.; Goddard, W. A., III, Exploring the molecular mechanism for color distinction in humans. *J. Phys. Chem. B* **2006,** *110* (34), 17230-17239.

71. Sekharan, S.; Sugihara, M.; Buss, V., Origin of spectral tuning in rhodopsin - it is not the binding pocket. *Angew. Chem. Int. Ed. Engl.* **2007,** *46* (1-2), 269-271.

72. Soederhjelm, P.; Husberg, C.; Strambi, A.; Olivucci, M.; Ryde, U., Protein influence on electronic spectra modeled by multipoles and polarizabilities. *J. Chem. Theory Comput.* **2009,** *5* (3), 649-658.

73. Derguini, F.; Caldwell, C. G.; Motto, M. G.; Baloghnair, V.; Nakanishi, K., Bacteriorhodopsins containing cyanine dye chromophores - support for the external point-charge model. *J. Am. Chem. Soc.* **1983,** *105* (3), 646-648.

74. Kloppmann, E.; Becker, T.; Ullmann, G. M., Electrostatic potential at the retinal of three archaeal rhodopsins: implications for their different absorption spectra. *Proteins* **2005,** *61* (4), 953-965.

75. Steinberg, G.; Ottolenghi, M.; Sheves, M., Pk(A) of the protonated schiff-base of bovine rhodopsin - a study with artificial pigments. *Biophys. J.* **1993,** *64* (5), 1499-1502.

76. Druckmann, S.; Ottolenghi, M.; Pande, A.; Pande, J.; Callender, R. H., Acid-base-equilibrium of the schiff-base in bacteriorhodopsin. *Biochemistry* **1982,** *21* (20), 4953-4959.

77. Scheiner, S.; Duan, X. F., Effect of intermolecular orientation upon proton-transfer within a polarizable medium. *Biophys. J.* **1991,** *60* (4), 874-883.

78. Gat, Y.; Sheves, M., A mechanism for controlling the pKa of the retinal protonated schiff-base in retinal proteins - a study with model compounds. *J. Am. Chem. Soc.* **1993,** *115* (9), 3772-3773.

79. Scheiner, S.; Hillenbrand, E. A., Modification of pKa values caused by change in H-bond geometry. *P. Natl. Acad. Sci. Usa* **1985,** *82* (9), 2741-2745.

Chapter II. Engineering the 2nd generation rhodopsin mimic using CRBPII

II.1 The first generation rhodopsin mimic, an introduction

The wavelength regulation observed in rhodopsin systems has attracted a lot of interest since the mid 20th century. One single chromophore, 11-*cis*-retinylidine, is responsible for the wide range of absorptions from ~420 nm to ~570 nm.[1, 2, 3] Color visual pigments provide a good example for studying protein-chromophore interactions that lead to different spectral characteristics of the chromophore. Semi-rational site directed mutagenesis of rhodopsins has been used as the major tool to decipher the factors responsible for causing the opsin shift. Irrational screening methods have also been developed recently to select for rhodopsin mutants with different absorptions, using *E. coli* capable of synthesizing endogenous retinal.[4] It was found that both the residues surrounding the chromophore and outside the binding pocket could affect the absorption maxima. However, crystal structures of these mutants are not available to analyze the cause-effect relationships.

One of the major limitations in using rhodopsin to study spectral tuning is that rhodopsin, as a membrane bound protein, is difficult to express, purify and crystallize. Even though different rhodopsin mutants have been produced for the purpose of spectral tuning, no crystal structures for these mutants are available to validate proposed hypotheses. A slight conformational change could have a

dramatic impact on the absorption profile of the chromophore. The overall conformation of the protein, as well as the side chain conformations, is critical for studying the protein-chromophore interactions' effect on the spectral characteristics of the chromophore. Therefore, various opsins might not be the best platform to study protein-chromophore interactions, even though they do provide a broad spectrum of absorptions with a single chromophore.

Although a number of investigations have targeted the wavelength regulation mechanism observed in color rhodopsins and bacteriorhodopsins, still no single conclusion could be drawn due to lack of crystal structures of color rhodopsins and the mutants of bacteriorhodopsin. Crystal structures are important in deciphering the subtle changes that lead to stereoelectronic alterations between the protein and the chromophore that could account for the wavelength tuning. Although crystal structures only provide one snapshot of the dynamic process of a protein's movement, they can provide valuable information about the dominant conformations of the flexible side-chains. Our approach for this problem has been to switch to another small cellular protein, which is easily expressed, structurally robust and most importantly amenable to both mutagenesis and crystallography. The protein would be engineered into a rhodopsin mimic that binds retinal as a protonated Schiff base (PSB).

Cellular Retinoic Acid Binding Protein II (CRABPII), which belongs to the family of lipid binding proteins, was chosen as a surrogate for rhodopsin. CRABPII is a small cellular protein, containing 137 amino acids. It forms a 10 β-

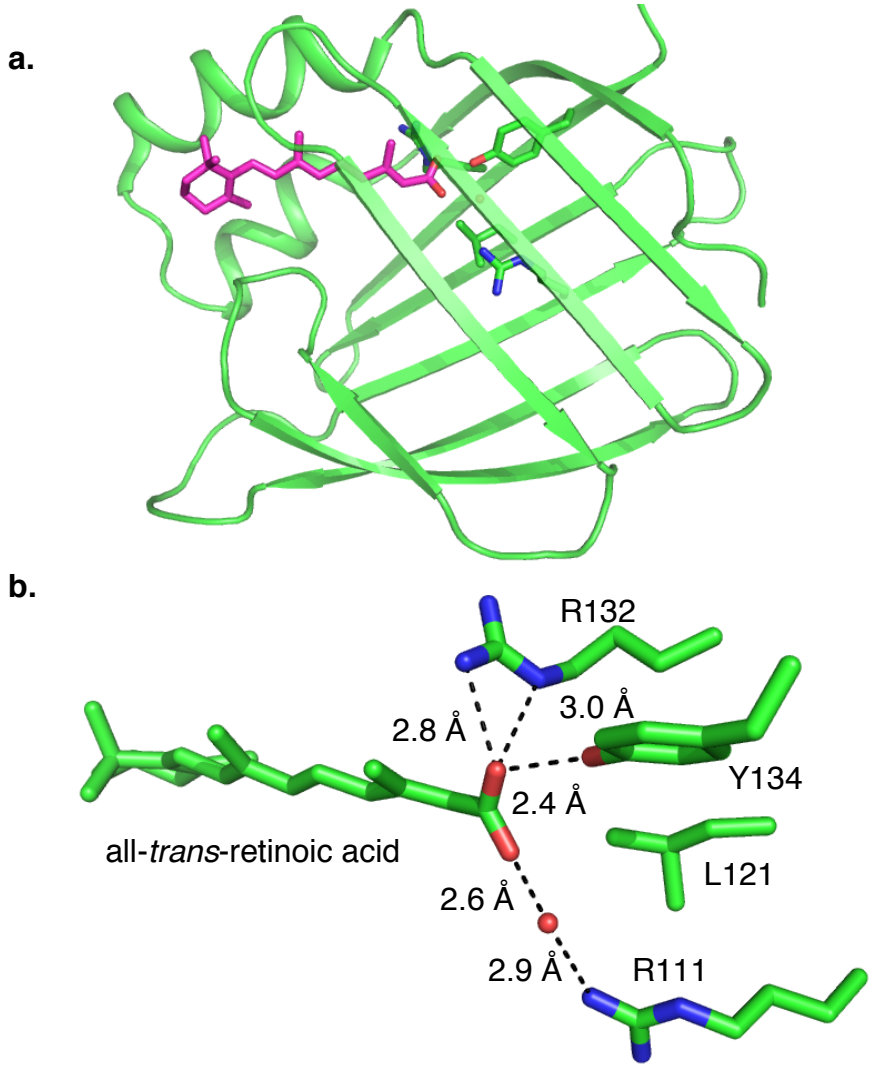


Figure II-1: Crystal structure of WT-CRABPII bound with retinoic acid. **a**. Cartoon of CRABPII bound with retinoic acid (PDB entry: 1CBS). **b**. Highlighted interactions of the chromophore carboxylic acid and CRABPII residues.

sheet barrel with a large binding pocket for its ligand, all-*trans*-retinoic acid (**Figure II-1a**).[5] There are two helices situated in the entry of the binding pocket.

Since 11-*cis*-retinal is extremely sensitive to light, all-*trans*-retinal is used instead to study the protein-chromophore interactions. The two isomers should share a similar mechanism intrinsically for spectral tuning as have been observed in microbial rhodopsins.

Although all-*trans*-retinal and all-*trans*-retinoic acid are structurally similar, CRABPII has a much lower affinity for all-*trans*-retinal, with a dissociation constant of 6000 nM compared to 1 nM for all-*trans*-retinoic acid.[6] The major interaction between retinoic acid and CRABPII is a salt bridge, facilitated by Arg132 directly and with Arg111 through a structured water mediated interaction as shown in **Figure II-1b**. Retinal, on the other hand, lacking the carboxylate moiety, does not bind well in the basic binding pocket.

To engineer CRABPII into a rhodopsin mimic, first a lysine residue was introduced at position 132. A Schiff base was formed as evident by the anticipated mass shift in MALDI-TOF, but no protonation was indicated, since there is no observed red shift to a peak above 420 nm, which is characteristic of retinal-PSB formation.

It was originally thought that a hydrophobic environment surrounding the lysine residue was important to lower its pK_a, in order to make it a better nucleophile to attack the retinal. It was realized later that hydrophobicity is not critical for Schiff base formation. What was important was the nucleophilic attack trajectory. In an elegant, crystallography based study, Bürgi and Dunitz determined that for a successful nucleophilic attack on a carbonyl, the amine must attack the aldehyde in a trajectory that bisects the aldehyde plane, with an optimal angle of 107°.[7] This specific angle for nucleophilic attack is due to a compromise between maximizing overlap of the lone pair of the amine with the π^*

orbital of the carbonyl group and minimizing repulsion with the π bonding electron cloud of the carbonyl group.

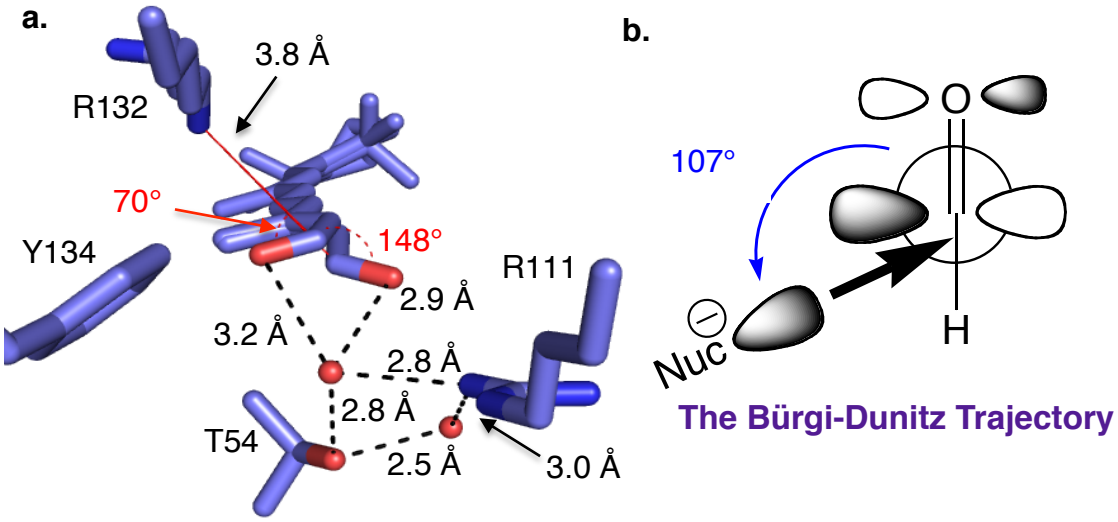


Figure II-2: **a**. Detailed hydrogen bonding interactions between R132K:Y134F–CRABPII and all-*trans*-retinal. **b**. Illustration of the Bürgi-Dunitz trajectory.

Double mutant R132K:Y134F does not form a Schiff base with all-*trans*-retinal, as evident from reductive amination studies. The crystal structure of R132K:Y134F with retinal further proved that Lys132 did not form a Schiff base with retinal. Comparison of the crystal structures of WT CRABPII and R132K:Y134F suggests that Arg111 and Thr54 are interacting with the aldehyde through a water-mediated hydrogen bonding network, and this interaction holds the aldehyde in an unfavorable position for the nucleophilic attack of lysine 132 (**Figure II-2a**). Consequently, removal of Arg111 and Thr54 interactions or restoration of Y134 hydrogen bonding interaction with the aldehyde might recover the reactivity of the aldehyde by allowing it to adopt a more favorable position for nucleophilic attack. This study suggested the importance of nucleophilic attack

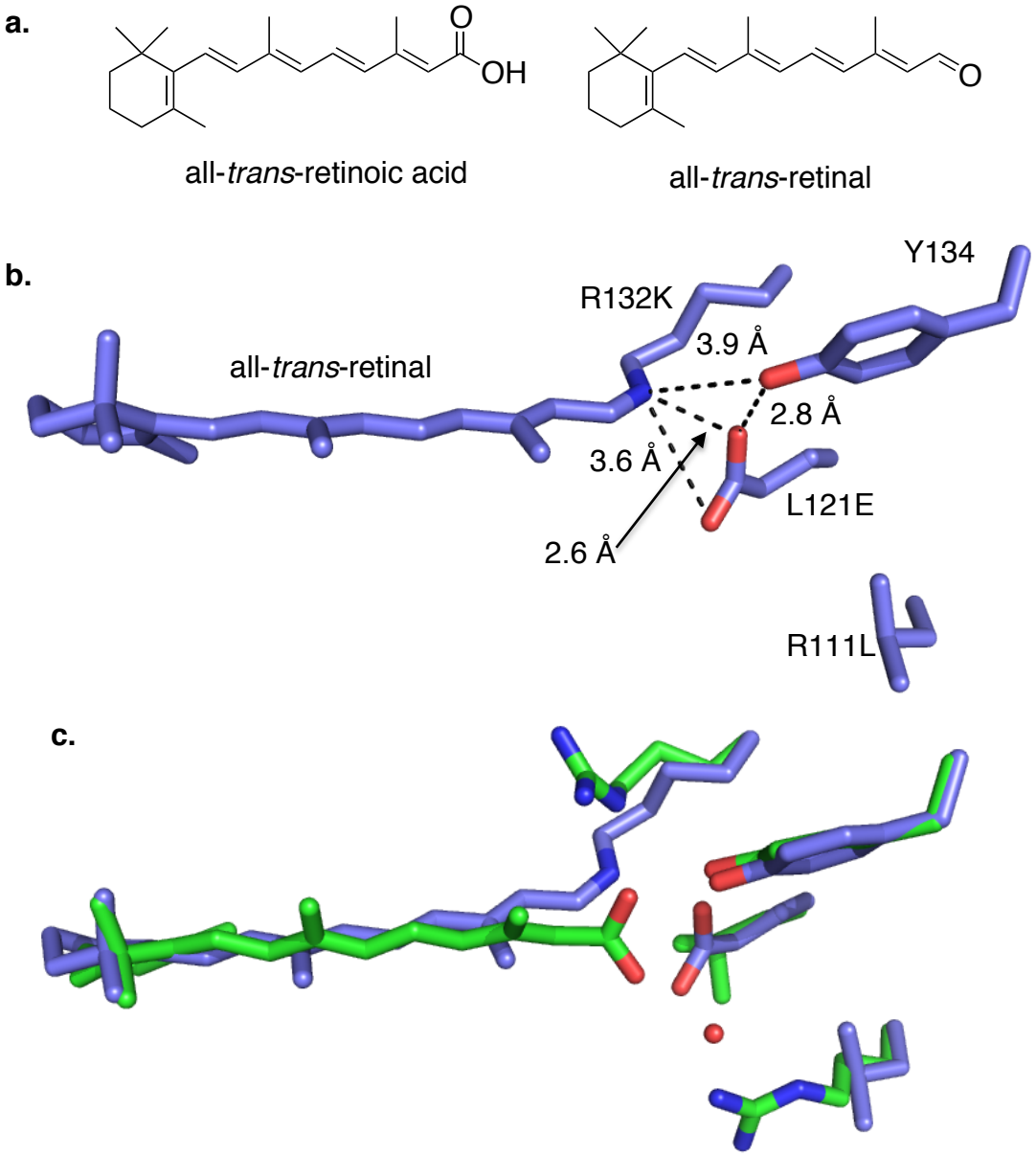


Figure II-3: **a.** Structures of all-*trans*-retinoic acid and all-*trans*-retinal. **b.** Crystal structure of R132K:R111L:L121E. **c.** Overlaid structures of WT-CRABPII-retinoic acid (green, 1CBS) and triple mutant R132K:R111L:L121E-retinal (cyan, 2G7B).

trajectory in the process of rhodopsin mimic engineering, and Bürgi-Dunitz trajectory should always be considered for future engineering.

R111L mutation was introduced to remove the observed water molecule, which holds the aldehyde in an unfavorable position for nucleophilic attack. However, still no obvious PSB was formed in double mutant R132K:R111L, due to the low pK_a of the retinal-PSB formed (<6.5). Learning from the rhodopsin systems, most of which have a counter anion close to the retinal-PSB to stabilize the protonated state of the Schiff base through charge compensation, a counteranion L121E was introduced. The retinal-PSB formed with triple mutant R132K:R111L:L121E binds retinal as shown in **Figure II-3** to form a stable PSB with a pK_a of 8.7. In this way, under physiological conditions, pH=7.3, the PSB is the major absorption at 449 nm. Control study with mutation L121Q verified the role of L121E as the counteranion of the PSB. Indeed R132K:R111L:L121Q has a much lower pK_a (<6.5), with an absorption maximum at ~370 nm, corresponding to unprotonated retinal-SB.[6, 8]

II.2 Proof of principle study: using C15 as the chromophore for wavelength tuning

After achieving a stable retinal protonated Schiff base through triple mutant of CRABPII, R132K:R111L:L121E, further mutagenesis studies on residues surrounding the chromophore were carried out to probe the spectral tuning mechanisms on this rhodopsin mimic. However, R132K:R111L:L121E-CRABPII based mutants did not respond to the polarity change in the binding

pocket, as indicated in **Table II-1.**[6, 9] Polarity switch at position 59 (**Figure II-4a**) from a positively charged to a negatively charged amino acid did not lead to the anticipated red shift. This was surprising, because removal of positive polarity, along with introduction of negative polarity in the ionone ring region should encourage positive charge delocalization of the iminium and result in a red shift.

Table II-1: CRABPII mutants with all-*trans*-retinal and all-*trans*-C15.

CRABPII mutant	λ_{max} with retinal	λ_{max} with C15
R132K-R111L-L121E-**R59**	449	n.d.[a]
R132K-R111L-L121E-**R59E**	450	424
R132K-R111L-L121E-**R59Q**	444	413
R132K-R111L-L121E-**R59L**	443	391
R132K-R111L-L121E-**R59W**	442	404

[a] because the pK_a of the PSB formed is too low to observe the absorption peak for PSB under physiological pH.

A closer examination of the crystal structure of R132K:R111L:L121E-CRABPII bound with retinal reveals that the binding pocket is open and the ionone ring of the chromophore is exposed to the aqueous environment. This is different in rhodopsins, which have the chromophore fully embedded inside their hydrophobic binding pockets. The water exposure of the retinal in R132K:R111L:L121E-CRABPII could possibly make the chromophore inert to changes in polarity near the ionone ring, due to high dielectric constant of aqueous system. To investigate this hypothesis, two different approaches were

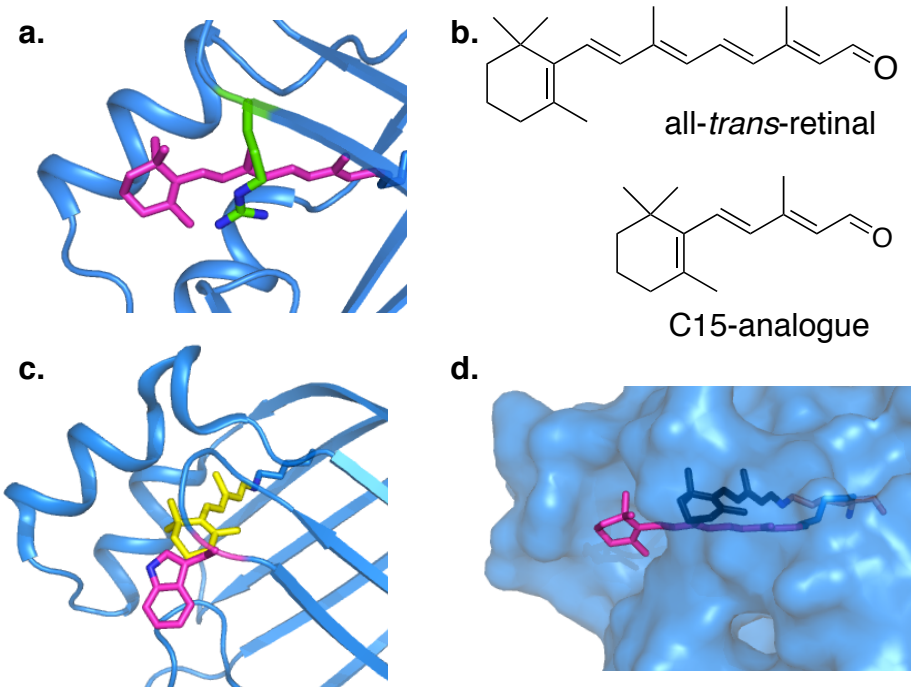


Figure II-4: Crystal structure of CRABPII mutant bound with C15. **a**. Crystal structure of R132K:R111L:L121E (KLE) bound with retinal, with Arg59 (green) and retinal (magenta) highlighted. **b**. Chemical structures of all-*trans*-retinal and all-*trans*-C15 analogue. **c**. Crystal structure of KLE-R59W bound to C15, with R59W (magenta) and C15 (yellow) highlighted. **d**. Overlaid structure of KLE-retinal and KLE-R59W-C15, showing that C15 is fully embedded within the protein binding pocket, while retinal (magenta) is exposed.

followed. One was to use a chromophore analogue that is shorter than retinal, so that it could be fully embedded inside the binding pocket. The other one is to engineer another protein, which can bind all-*trans*-retinal deep inside the binding pocket and fully sequester it from the aqueous solution, into a rhodopsin mimic.

The first approach was carried out by Dr. Lee from our lab using a shorter retinal analogue, C15 as shown in **Figure II-4a.**[9] The crystal structure of R132K:R111L:L121E:R59W bound to C15 was obtained, revealing that C15 is fully embedded inside the binding pocket.

Comparison of mutagenesis studies at position 59 for retinal and C15 is shown in **Table II-1**. For retinal, no red shift is observed for different amino acids replacement of Arg59. For the shorter C15 chromophore, the polarity change results in different degrees of spectral tuning. A negative polar residue, R59Q, placed close to the ionone ring results in a more red-shifted absorption compared to hydrophobic residue, R59L, while an acidic residue, R59E, leads to the most red shift. Tryptophan also results in a red shift, presumably because of its polarizable π electron cloud, which has been reported to cause red shift when it is placed near the ionone ring region.[10] Generally speaking, the absorption spectra of the mutants bound with C15 follow the general trend for wavelength regulation according to polarity change.

As a proof of principle study, the C15 analogue bound with different CRABPII mutants of Arg59 suggests that sequestration of the chromophore deep inside the binding pocket is critical for spectral tuning.

II.3 The 2nd generation rhodopsin mimic, based on CRBPII

In order to understand the intrinsic mechanisms responsible for the wide range of absorptions observed in rhodopsins, our approach is to use a small cellular protein, which is easily expressed and crystallized, and engineer it into a rhodopsin mimic that can bind all-*trans*-retinal as a protonated Schiff base.

CRABPII has been engineered into the first generation rhodopsin mimic through introduction of a lysine residue, which can have a favorable nucleophilic attack on the retinal through the appropriate Bürgi-Dunitz trajectory. The protonated Schiff base was stabilized through introduction of a counteranion that could form a salt bridge with the iminium. However, the ionone ring of retinal was found to be water exposed and this seemed to compromise the ability of the first generation rhodopsin mimic to be sensitive to polarity changes in the binding pocket. The polarity change induced by mutagenesis dissipates fast in the aqueous solution, which has a higher dielectric constant (~78) than that inside the hydrophobic protein binding pocket (from 2 to 10).[11] As a proof of principle study, shortening of the chromophore by two double bonds enabled the chromophore to be fully embedded and the chromophore started responding to polarity change in the expected manner.

The C15 study showed that having the chromophore fully embedded inside the binding pocket is important for spectral tuning. However, it does not represent what is happening in rhodopsin, as C15 is two double-bond shorter than the full-length retinal. Wavelength regulation studies on the full-length retinal are necessary for a direct comparison with the rhodopsin system.

We switched to another small cellular protein that had potential to bind the full-length retinal deeper inside the binding pocket and have the chromophore fully embedded. This way the polarity change might have a more dramatic effect on the absorptions of the chromophore, due to the lower dielectric constant of the

protein's interior. As shown in the following equation, the electric field exhibited by a point charge is inversely proportionate to the dielectric constant. The smaller the dielectric constant, the greater the electrostatic effect. Water has a dielectric constant of 78, while the dielectric constant inside a hydrophobic binding pocket was estimated to be ranging from 2 to 10.[11] Therefore in a hydrophobic environment, the electrostatic potential is more significant by at least 8 folds over the same distance.

$$E = \frac{Q}{4\pi\varepsilon * r^2}$$

(In the equation, Q is the charge for the particle, ε is the dielectric constant, r is the distance from the particle with charge Q to the E-field evaluation point.)

For this purpose, Cellular Retinol Binding Protein II (CRBPII) is a perfect target to be engineered into a rhodopsin mimic. WT CRBPII bound with retinol has been crystallized, showing that the chromophore binds deeper within the binding pocket. With the engineering strategies learned from the first generation rhodopsin mimic studies, CRBPII was our next target for engineering a rhodopsin protein mimic.

CRBPII binds both retinal and retinol with dissociation constants of 90 nM and 10 nM, respectively.[12] Human CRBPII is usually found in the small intestine, transporting the retinol/retinal across enterocytes to deliver retinol/retinal to the appropriate metabolic enzymes.[13, 14] CRBPII also belongs to the family of lipid

binding proteins,[15] it has 133 amino acids, much smaller than rhodopsin, which has 364 amino acids.

The sequence identity between CRBPII and CRABPII is only 35%, but CRBPII shares similar structural scaffolds with CRABPII. CRBPII also forms a 10 β-sheet barrel with two helices acting as the lid of the binding pocket, like CRABPII. However, the binding site in CRBPII is ~5 Å deeper than that in CRABPII as shown in **Figure II-5a**. I think this could be an evolutionary result to achieve different functions of these two proteins. CRABPII transfers retinoic acid to a gene regulation protein, retinoic acid receptor (RAR).[16, 17] The binding site of CRABPII is better to be shallow and open, to facilitate the transfer of retinoic

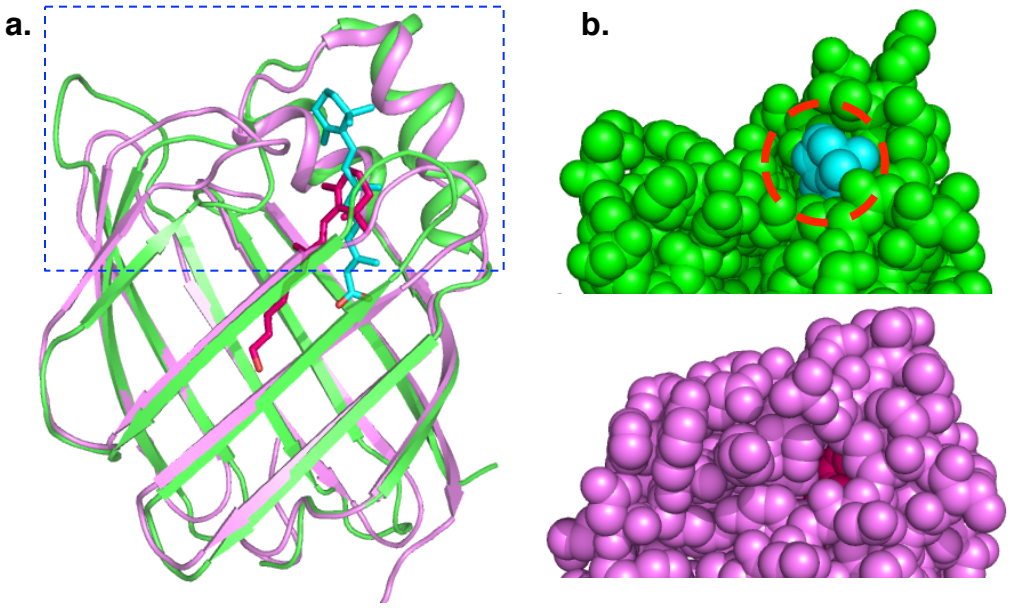


Figure II-5: Comparison of WT CRABPII and CRBPII. **a**. Overlaid crystal structures of WT CRABPII (green, 1CBS) bound with all-*trans*-retinoic acid (cyan) and WT CRBPII (violet, 2RCT) bound with all-*trans*-retinol (red). **b.** Space filling models for the crystal structures of WT CRABPII (top) and CRBPII (bottom) corresponding to the highlighted rectangular area in **a**. The red circular area highlights the retinoic acid ionone ring region, which is water exposed.

acid to RAR protein. While for CRBPII, it has to protect the retinal/retinol from oxidative degradation or isomerization, thus it is better to have the chromophore bound deeper inside a rigid binding pocket.

As a result, the retinal is fully embedded in the protein binding site of CRBPII, while in CRABPII, the ionone ring, highlighted within a red dashed circle (**Figure II-5b**), is exposed to the aqueous solution. With this more enclosed protein cavity, the dielectric constant in the binding pocket is lowered, and thus the electrostatic interactions between the protein and the chromophore will have more significantly effect on the absorptions of the chromophore.

II.4 Expression of WT CRBPII and characterization of WT CRBPII in *E. coli* system

WT CRBPII gene was cloned into both pETBlue-2 and pET17b vectors. Proteins expressed in pETBlue-2 vector have a 6-His-tag and the proteins are

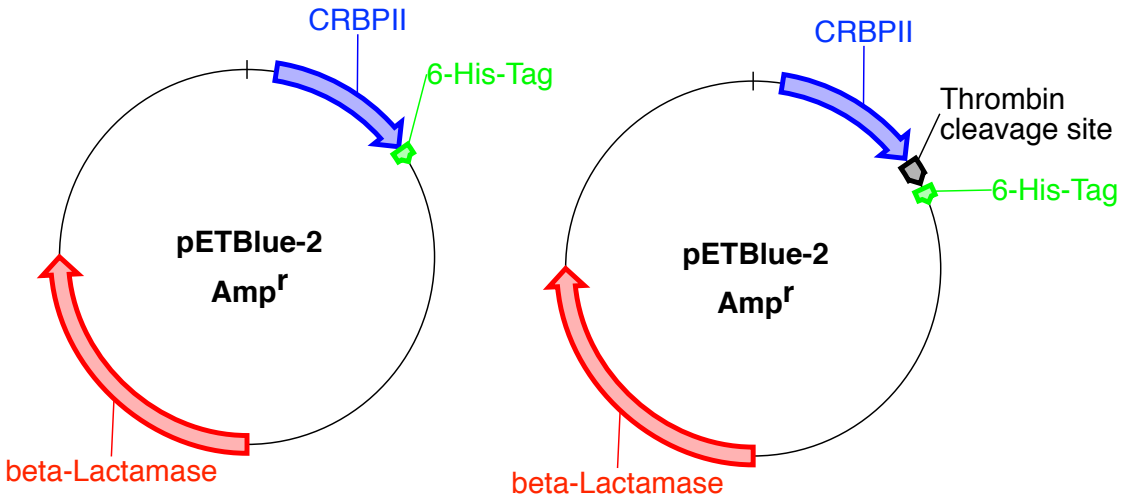


Figure II-6: Clone of CRBPII in pETBlue-2 vector with and without thrombin cleavage site introduced.

purified by Ni-NTA column. The 6-His-tag seemed to interfere with the crystallization of CRBPII protein. Introduction of a thrombin cleavage site between the CRBPII protein and the 6-His-tag could enable removal the 6-His-tag with thrombin digestion (**Figure II-6**). However, the whole purification process became too tedious and thrombin digestion was not efficient due to the short linkage between CRBPII and 6-His-tag.

Proteins expressed with the pET17b vector do not have affinity tag and the proteins are purified through ion exchange column. As CRBPII proteins have a pI of ~5.2, anion exchange column is applied at pH=8.0, so that the proteins are negatively charged and can stick to the column. Expression of WT CRBPII in pET-17b yields up to 30 mg/L of protein and the purification process is much easier. Most of the CRBPII proteins were expressed in pET17b without affinity tag.

Characterization of WT-CRBPII was carried out to make sure that the protein obtained was functional. As shown in **Figure II-7a,** the CD spectrum of WT-CRBPII shows a negative absorption peak at ~218 nm, characteristic of β-sheets.[18] This indicates the protein obtained has a correct folding, as β-sheet is the major secondary structure for CRBPII. No obvious absorption for α-helix is observed in the CD spectrum of WT-CRBPII. This could be due to the small proportion of the α-helices in CRBPII compared to the dominant β-sheets.

UV-vis titration of WT-CRBPII protein with all-*trans*-retinol shows the vibronic structure and a large red shift from 320 nm in ethanol solution to ~350

nm in CRBPII (**Figure II-7b**), characteristic of CRBPII as reported before.[19] The vibronic structure could be a result of sequestration of the chromophore from the aqueous solution, which usually has a line broadening effect and weakens the obvious vibronic structure detected in the UV-vis spectra. The red shift observed could be due to rigidification of rotation along the C6-C7 single bond. This leads to a smaller dihedral angle of 5.7° as observed in the crystal structure of WT-CRBPII bound with retinol, while the ionone ring is highly twisted for free retinol in solution.

All-*trans*-retinal bound with WT CRBPII also red shifts 10 nm to 390 nm, compared to the reported absorption at 395 nm when retinal is bound to CRBPII.[20] For most of the neutral molecules, when they are dissolved in non-polar solvent, a blue shift results due to destabilization of the excited state, which is usually more polar than the ground state. However, all-*trans*-retinal and all-*trans*-retinol are red shifted when bound in the hydrophobic binding pocket of CRBPII. This is not unique for CRBPII, 11-*cis*-retinal also red shifts dramatically to 420 nm when bound in the hydrophobic binding pocket of Cellular Retinaldehyde Binding Protein (CRALBP).

A few factors could contribute to this red shift. Sterically, a rigid binding pocket could possibly planarize the chromophore through appropriate protein-chromophore packing leading to a higher degree of conjugation. Crystal structure of retinal bound in WT-CRBPII shows that the ionone ring is around 45°

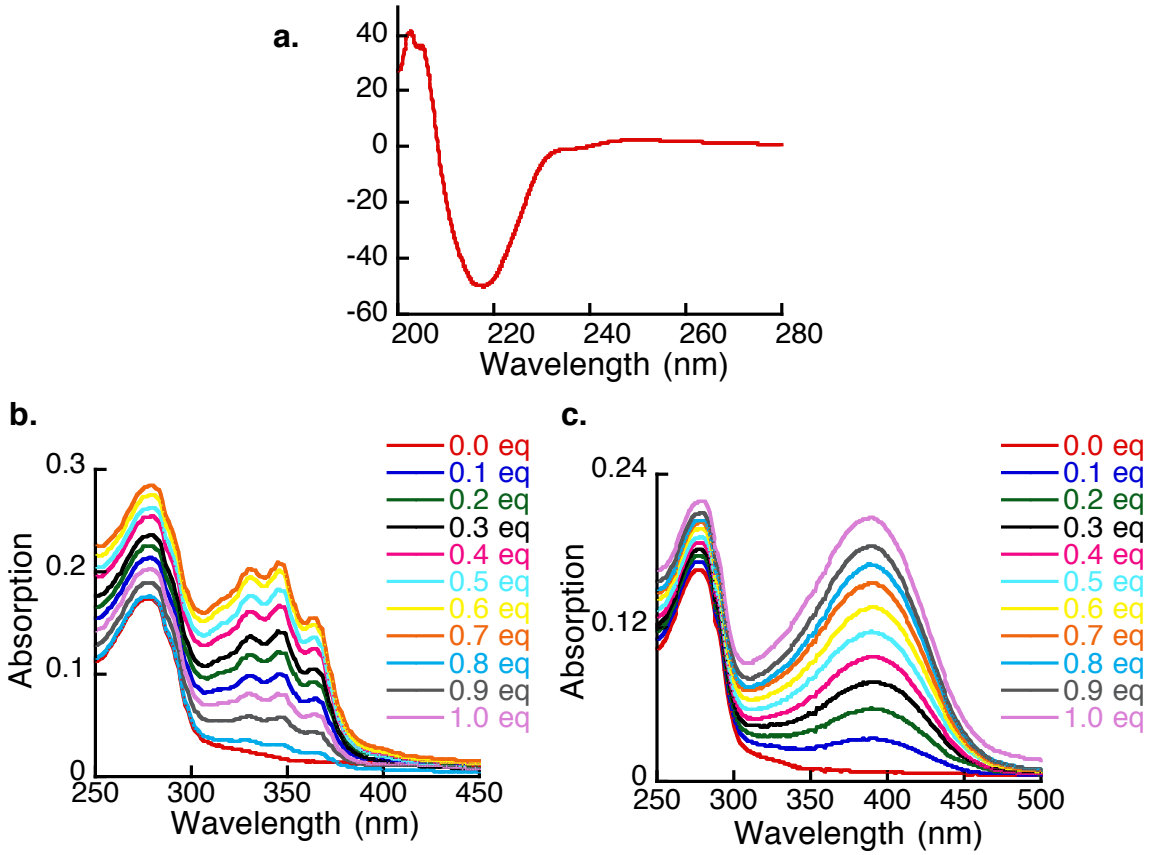


Figure II-7: Characterization of WT CRBPII. **a**. CD spectrum of WT-CRBPII. **b**. UV-vis spectra of WT-CRBPII titration with all-*trans*-retinol. **c**. UV-vis spectra of WT-CRBPII titration with all-*trans*-retinal.

twisted from the plane of the polyene. This is slightly planar than retinal in solution (~60°).[21] In this way, red shift could result even when there is no covalent bond formation between the chromophore and the protein.

In solution, either polar solvent or non-polar solvent, the bulk solvent molecules could always arrange their permanent dipoles or induced dipoles to stabilize the ground state of the chromophore to reach a minimum energy state (**Figure II-8a**). Upon excitation from the ground state to the excited state,

permanent dipoles in solvent can not reorient to stabilize the excited state within a short time frame (10^{-15} s), therefore a large energy gap will result.

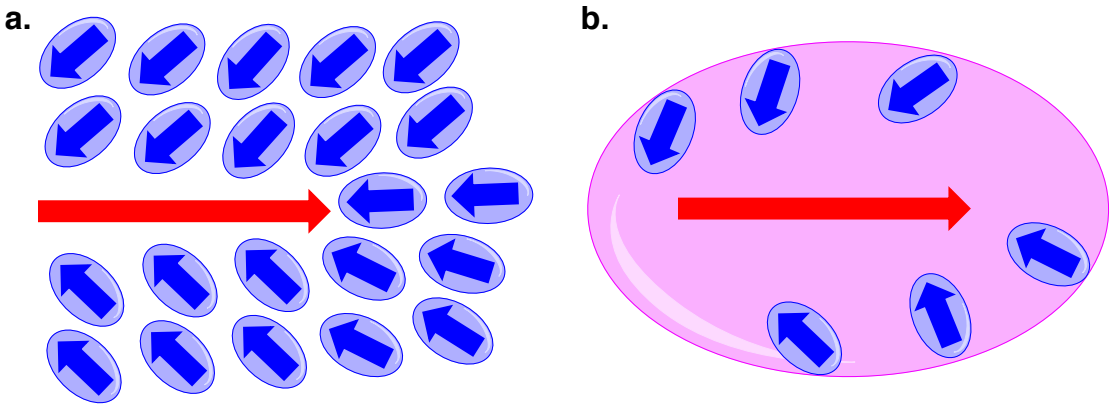


Figure II-8: Cartoon simulation of the dipole interactions in the solvent (a) and inside a protein binding pocket (b). Red arrows represent the dipoles of the ligand and the blue arrows represent the dipoles of the solvent or the residues inside the binding pocket. It shows that in solution, the solvent molecules could orient the permanent dipole or induced dipole to interact with the dipole of the ligand to achieve the minimum energy state. But in the case of protein binding pocket, due to restriction of the position of side chains in the binding pocket, the stabilization of the dipole of the ligand is limited and therefore can result in higher ground state energy.

The electrostatic interactions between the protein and the chromophore could contribute to the red shift observed. Inside the binding cavity of a protein, the dipoles of the side chains can orient in a way to stabilize the ground state of the chromophore to the largest extent possible. However, the orientations of the dipoles of the side chains are restricted by their spatial positions and allowed rotamers (**Figure II-8b**). With that said, protein environment has the advantages of positioning permanent dipoles or induced dipoles in an ordered and preorganized manner, to result in higher energy ground state than that in solution, leading to a red shift.

The organized dipoles inside the binding pocket can contribute to the opsin shift observed in the rhodopsins in a similar way as that of retinol/retinal in WT-CRBPII, by increasing the ground state energy. As protonated retinal-PSB is much more polar than neutral retinol/retinal, the protein environment could have a more significant effect on the absorption maximum of retinal-PSB.

II.5 Crystal structures of WT CRBPII-retinol and WT CRBPII-retinal

The high resolution crystal structure of WT CRBPII bound with retinol was available from the beginning.[22] Rafida Nossoni (Professor Jim Geiger's lab, MSU) was able to obtain the crystal structures of WT CRBPII bound with retinol and retinal to 1.2 Å resolution. These crystal structures greatly facilitated the engineering of CRBPII into a rhodopsin mimic.

WT-CRBPII-retinol has structured hydrogen-bonding water network inside the binding pocket. As shown in **Figure II-9**, Gln108 has a tight hydrogen bonding interaction with the alcohol functional group of retinol through a water molecule. The latter water molecule is hydrogen bonded with Lys40, which further interacts with T53 and a water molecule bound by two glutamine residues, Gln38 and Gln128. Besides, the retinol hydrogen bonds directly to the side chain of Thr51, which interacts with Thr53 indirectly through a water molecule.

Overlaying the crystal structures of WT CRBPII bound with retinol and retinal shows that the two chromophores are bound in a similar way (**Figure II-9**

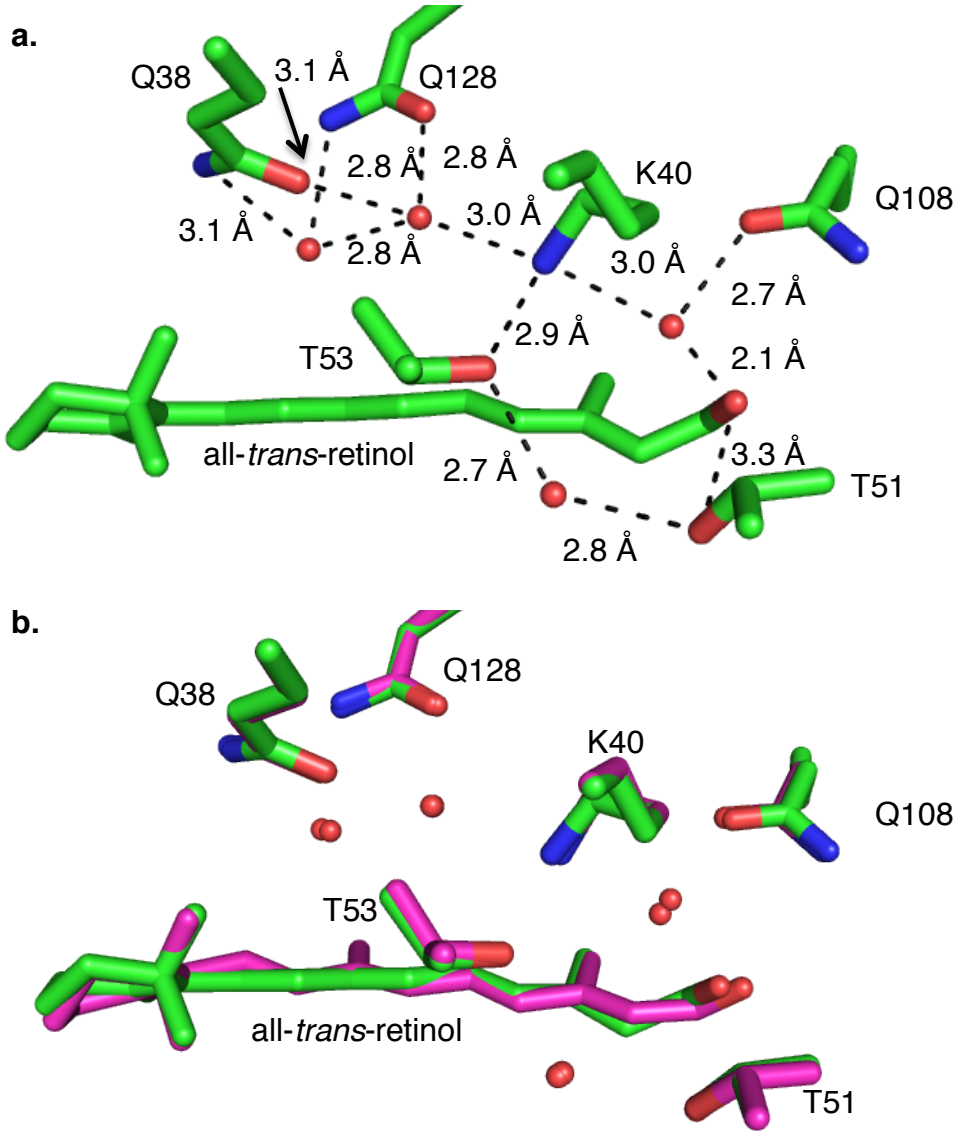


Figure II-9: Crystal structures of WT CRBPII bound with retinol and retinal. **a**. Crystal structures of WT CRBPII-retinol, showing the internal hydrogen bonding network among the side chain residues and retinol. **b**. Overlaid crystal structures of WT CRBPII-retinol (green) and WT CRBPII-retinal (magenta), showing that the two structures overlaid well with the side chains taking the same conformation due to similar water mediated hydrogen bonding network.

B). The positions of bound retinal and retinol do not change, with only a small difference for the aldehyde oxygen, mainly because the aldehyde is flat (sp^2 hybridized) while alcohol is not (sp^3 hybridized carbon character). All the

hydrogen bonding interactions are maintained in the crystal structure of CRBPII-retinal as that in CRBPII-retinol, indicating the importance of these hydrogen bonding interactions for retinal and retinol binding.

There is one lysine residue, Lys40, in close proximity to the aldehyde moiety in the crystal structure of WT CRBPII bound with retinal (**Figure II-9b**), but it does not form a Schiff base with the aldehyde. This could be explained by the unfavorable Bürgi-Dunitz trajectory for Lys40 to attack the aldehyde. Once again it shows the importance of the Bürgi-Dunitz trajectory for Schiff base formation in rhodopsin mimic engineering.

II.6 Introduction of a nucleophilic lysine residue

In order to engineer CRBPII into a rhodopsin mimic, a lysine residue, which can take a favorable Bürgi-Dunitz trajectory to attack the aldehyde, first needed to be introduced. After that, different strategies could be implemented to stabilize the protonated state of the SB, as the retinal-PSB is necessary to observe the red shift and to study wavelength regulation.

Upon studying the CRBPII-retinal crystal structure, three positions, 108, 106 and 51 (**Figure II-10a**), close to the aldehyde were chosen for in silico mutagenesis into lysine. All the possible rotamers of the three residues were considered in order to evaluate whether the putative lysine residue could adopt a favorable Bürgi-Dunitz trajectory to attack the aldehyde. The modeling result

showed that only Q108K has a conformation that could favorably attack the aldehyde (**Figure II-10a**).

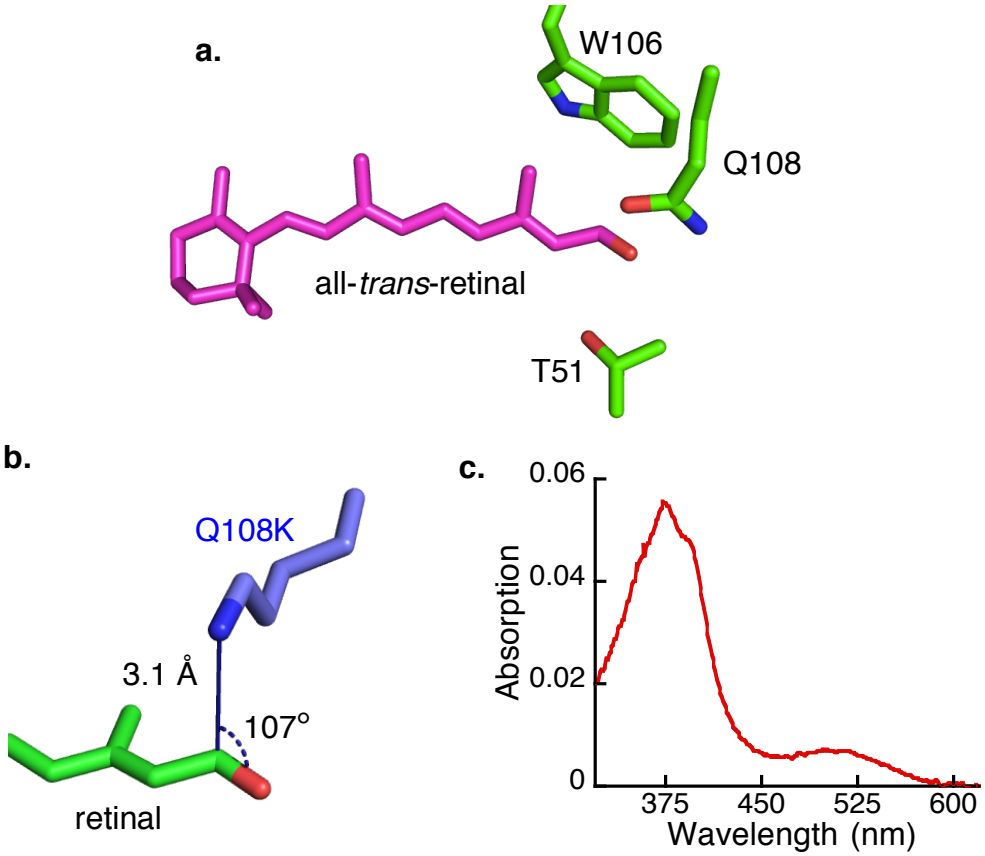


Figure II-10: Introduction of Q108K. **a**. Crystal structure of WT-CRBPII bound with retinal. **b**. Modeling of Q108K following the Bürgi-Dunitz trajectory to attack the aldehyde, based on the crystal structure of WT CRBPII-retinal. **c.** UV-vis spectrum of CRBPII single mutant Q108K bound with all-*trans*-retinal.

Nonetheless, all three mutations of WT-CRBPII, Q108K, W106K and T51K, were produced. Soluble protein was obtained only for the Q108K mutant, while the other two mutants, W106K and T51K, were found exclusively in the inclusion bodies during protein expression. W106 is highly conserved across the lipid binding protein family members and mutagenesis at this position disturbs the stability of the protein.[15] No functional protein of W106K could be obtained, as

the protein misfolded and aggregated in the form of insoluble inclusion bodies during protein expression. This indicates the importance of W106 in the correct folding of CRBPII. Often, aromatic residues are critical in the correct folding of protein due to π-π stacking, π-cation interactions and hydrophobic effect.[23]

Protein mutant T51K was not obtained either, due to difficulties encountered during protein expression. This could be due to the positive charge repulsion from Lys51, which could be placed close to Lys40. T51K was not pursued further partly because modeling had suggested that T51K would not be able to adopt the Bürgi-Dunitz trajectory to attack the retinaldehyde.

Q108K protein was obtained, with expression yield of ~1 mg/L, with IPTG induction at 32 °C. Most of the protein was found in inclusion bodies. The amount of soluble protein was sufficient to study the binding with all-*trans*-retinal. UV-vis spectrum of Q108K incubated with all-*trans*-retinal (**Figure II-10b**) shows that there is a small absorption peak at ~506 nm, while the major peak is at ~365 nm, typical of an un-protonated Schiff base of retinal. The absorption peak at 365 nm also exhibits vibronic structures, which is similar to WT-CRBPII bound with retinol, and indicates binding with the pocket.

The absorption peak at 506 nm was attributed to the retinal-PSB. Retinal absorbs at 380 nm in solution and ~390 nm when bound in CRBPII, and the un-protonated Schiff base usually peaks at ~365 nm. Only protonated Schiff base of retinal has been reported to absorb beyond 440 nm.

To confirm that the absorption peak at 506 nm is the protonated form of retinal-PSB and the peak at around 365 nm is the unprotonated Schiff base, acid titration was performed. If the peak at 506 nm increases and the peak at 365 nm decreases as the solution is acidified, it indicates that the peak at 506 nm is the protonated form. Unfortunately, acidification precipitated the protein slowly without causing an obvious increase of the absorption peak at 506 nm. This indicates the instability of single mutant Q108K under acidic conditions.

Later it was realized that Lys40, which is in close proximity to the PSB, could affect the stability of the protonated form of the single mutant Q108K due to charge repulsion. This also explains the low expression yield of Q108K-CRBPII and its instability under acidic conditions.

Lys40 also perturbs the pK_a of the retinal-PSB formed with single mutant Q108K-CRBPII. As the protonated form of lysine has a much higher pK_a (~9) than retinal-PSB (~6), Lys40 is more likely to be protonated than the retinal-PSB in the same environment. Due to charge repulsion, Lys40 could suppress the pK_a of retinal-PSB. That is probably why only a minor absorption peak is observed at 506 nm and a pK_a smaller than 6 is estimated for Q108K.

All the mutants in CRABPII (first generation rhodopsin mimic) absorb at around 450 nm when bound with all-*trans*-retinal, close to the absorption of retinal-PSB in aqueous solution, 440 nm. This indicates the protein environment had little impact on the chromophore. Absorption at 506 nm with Q108K-CRBPII

was promising as it showed that CRBPII has a more significant effect on the chromophore, presumably by embedding the chromophore deep inside the binding pocket.

However, the pK_a of Q108K is estimated to be lower than 6.0. This is not a good starting point for studying wavelength regulation of retinal-PSB. Optimization of the pK_a is necessary.

II.7 pK_a optimization for retinal-PSB

II.7.1 introduction of a counteranion, T51D

As mentioned above, another lysine residue in close proximity to the retinal-PSB likely lowers the pK_a of retinal-PSB due to charge repulsion. To increase the pK_a, two approaches could be followed. One is to introduce a counter-anion to stabilize the protonated state of the retinal-PSB through salt bridge formation and charge compensation. The other approach is to remove Lys40.

As seen in many rhodopsin systems, a counter anion is important for stabilizing the protonated state of the retinal-PSB. The specific orientation of the counter anion in relation to the PSB is critical for the stability of the PSB as well.[24, 25, 26, 27] Therefore, a counteranion could be introduced at an appropriate position to increase the pK_a of retinal-PSB. Modeling in Pymol

showed that introduction of an acidic residue at position 51, Asp51, had a good chance to have direct interactions with the retinal-PSB (**Figure II-11**), so double mutant Q108K:T51D-CRBPII was prepared.

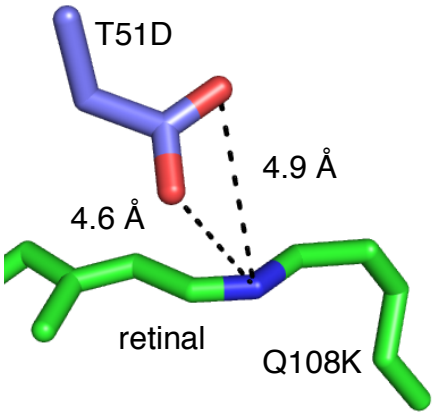


Figure II-11: Model structure of Q108K:T51D.

The expression of Q108K:T51D was not good. Induction of expression at 32 °C resulted in the formation of inclusion bodies exclusively. This is likely due to the incompatibility of having charged residues in the hydrophobic binding pocket. Introduction of T51D disturbs the folding of the protein. Refolding of the proteins from the isolated inclusion bodies was carried out first to obtain the soluble form of Q108K:T51D. Protein refolding yielded a good amount of active protein, which can bind retinal and form protonated Schiff base, with absorption maximum at ~474 nm. Later, expression conditions were optimized by lowering the induction temperature to 16 °C and ~10 mg/L of active soluble protein could be obtained, which was sufficient for the characterization of the protein and crystallographic studies.

As show in **Figure II-12**, introduction of an acidic residue T51D increases the pK_a of retinal PSB dramatically from < 6.0 to 9.2. Different from Q108K, which has a minor PSB absorption peak at 506 nm, double mutant Q108K:T51D has a single peak at 474 nm under physiological pH. It is ~3 pK_a units higher

than the pK_a of retinal-PSB in aqueous solution (~6.5). This study shows the importance of counteranion in stabilizing the protonated state of retinal-PSB.

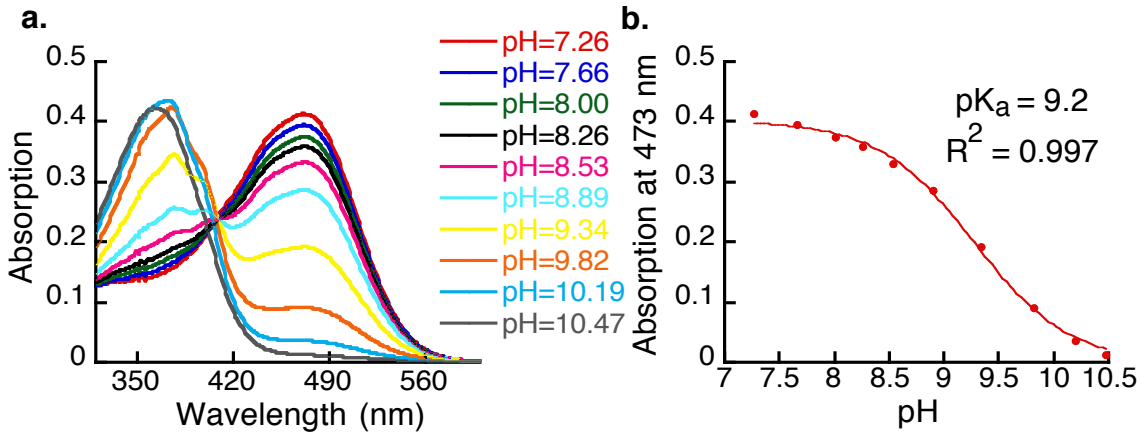


Figure II-12: Base titration of Q108K:T51D. **a**. UV-vis spectra of base titration of Q108K:T51D. The peak at 474 nm corresponds to retinal-PSB, while the absorption at around 370 nm corresponds to the unprotonated form, retinal-SB. As the pH of the solution increases, the absorption at 474 nm decreases, while the absorption at 370 nm increases. This shows the deprotonation of the PSB at higher pH. **b**. Absorption at 473 nm was plotted against the pH to obtain an apparent pK_a of 9.2.

Nonetheless, the pK_a of Q108K:T51D is not comparable to the pK_a of bovine rhodopsin, which is estimated to be above 16. As discussed earlier, the trajectory of the counteranion with respect to the imine is critical in determining the pK_a of ret-PSB. The counteranion, E113, in bovine rhodopsin has a perfect orientation to stabilize the PSB, with the overall dipole of the carboxylate pointing toward the iminium N-H. This could maximally balance the dipole of iminium and form a stable PSB.

Unfortunately, the crystal structure of Q108K:T51D bound with retinal could not be solved and we could not accurately dissect the exact interactions

between T51D and the retinal-PSB. Modeling of Q108K:T51D shows that due to the position of T51D and its allowed rotamers, there is no conformation of T51D that could place the carboxylate on the same plane as the iminium to maximally counteract the dipole of the iminium. Therefore, the trajectory of T51D is not optimal, but sufficient to charge compensate and form a salt bridge with the iminium to stabilize the retinal-PSB.

Besides the increase in pK_a, Q108K:T51D also blue shifts the absorption of retinal-PSB 32 nm from 506 nm to 474 nm as expected. Previous studies suggest that removal of counter-anion results in a red shift, and introduction of counter-anion leads to a blue shift.[24, 28, 29] This is because the counteranion can stabilize the positive charge of the iminium and reduce the delocalization of positive charge along the polyene. Introduction of a counteranion, T51D, agrees with the proposed hypothesis.

However, double mutant Q108K:T51D was not highly stable. Inclusion bodies were formed during protein expression and moderate expression level could only be achieved by lowering the induction temperature to 16 °C. Besides, Q108K:T51D has slow kinetics. It takes at least 2 hours of incubation at room temperature for Q108K:T51D to be completely bound with retinal and form PSB. The slower kinetics is probably due to possible salt bridge formation between T51D and Q108K before the nucleophilic attack (**Figure II-13**). Through this interaction, Q108K is held in a position not suitable for nucleophilic attack at retinaldehyde. In addition, introduction of an acidic residue in the binding pocket

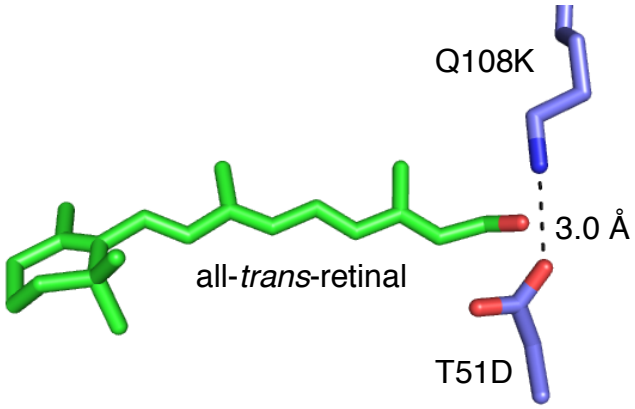


Figure II-13: Model structure of double mutant Q108K:T51D, showing one of the possible conformations of Q108K and T51D, forming salt bridge that could potentially slow down the imine formation.

can likely protonate Q108K and reduce the amount of the reactive nucleophilic lysine residue, which is neutral. It was observed later that lowering the pH of the buffer also slows down the formation of retinal-PSB with CRBPII mutants, possibly due to protonation of the lysine residue.

Based on the latter observations, double mutant Q108K:T51D was not the best platform for wavelength regulation. Another approach was pursued to increase the pK_a of retinal-PSB, with better stability and faster kinetics.

II.7.2 pK_a restoration of retinal-PSB through removal of K40

As mentioned earlier, there is another lysine residue inside the binding pocket in close proximity to the PSB, which likely disturbs the stability of the protein and surpresses the pK_a of retinal-PSB due to charge repulsion. Sheves and coworkers have shown that the pK_a of retinal-PSB is depressed as a result

of placing positive charge nearby in model compound studies.[27] To address this issue, different mutations of Lys40 were prepared to improve the stability and pK_a of CRBPII mutants.

CRBPII binding pocket is relatively hydrophobic, therefore a hydrophobic amino acid was considered first to replace Lys40. In order to maintain similar volume as the lysine residue and minimize differences in packing interactions in the binding pocket, leucine became the first choice to replace Lys40.

Double mutant Q108K:K40L was prepared, exhibiting a much better stability than Q108K and Q108K:T51D. Q108K:K40L can be expressed in soluble form when the expression is induced at 25 °C and expression level up to 100 mg/L can be achieved, even higher than that of WT CRBPII protein (~30 mg/L). This does suggest that Lys40 was disturbing the stability of the apo protein when Lys108 was introduced as a result of charge repulsion of the two residues in close proximity.

UV-vis base titration of Q108K:K40L shows that removal of Lys40 increases the pK_a of retinal-PSB drastically to 8.3 compared to Q108K (<6.5) (**Figure II-14**). Acidification of the protein solution does convert the absorption peak at ~370 nm to the absorption peak at 508 nm, suggesting the peak at ~370 nm corresponds to the unprotonated Schiff base and the peak at 508 nm the PSB. This also shows that the PSB is accessible by solvent molecules, as the PSB is titrable under acid-base conditions.

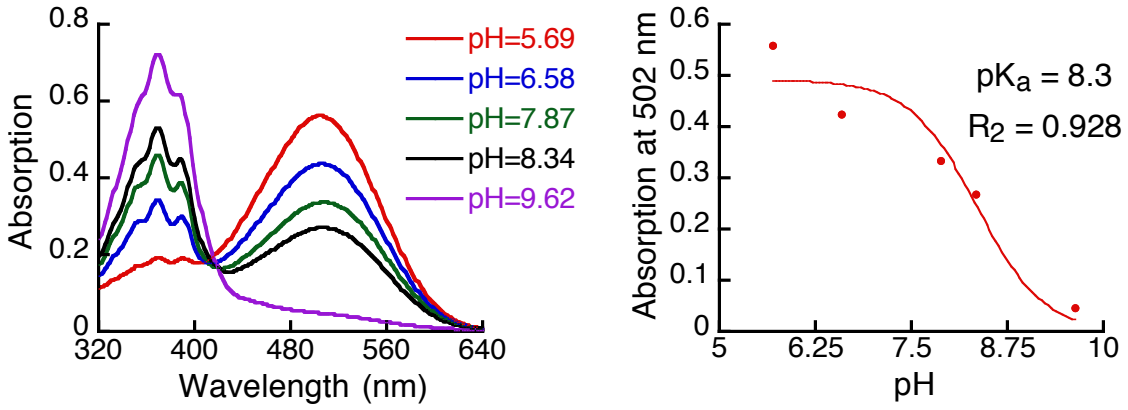


Figure II-14: UV-vis spectra of Q108K:K40L-retinal base titration.

Q108K:K40L does not shift the absorption maximum compared to Q108K. This was unexpected, as removal of a positively charged residue close to the Schiff base region should stabilize the positive charge localized on the iminium, resulting in less conjugation and thus a blue shift. The similar absorption maxima of Q108K:K40L to Q108K could suggest a proton transfer mechanism between the K40 and the PSB.

As shown in **Figure II-15,** the two different charged states are in equilibrium with each other. The equilibrium should be shifted to the K40 protonated and imine unprotonated since the pK_a of lysine residue is ~3 pK_a

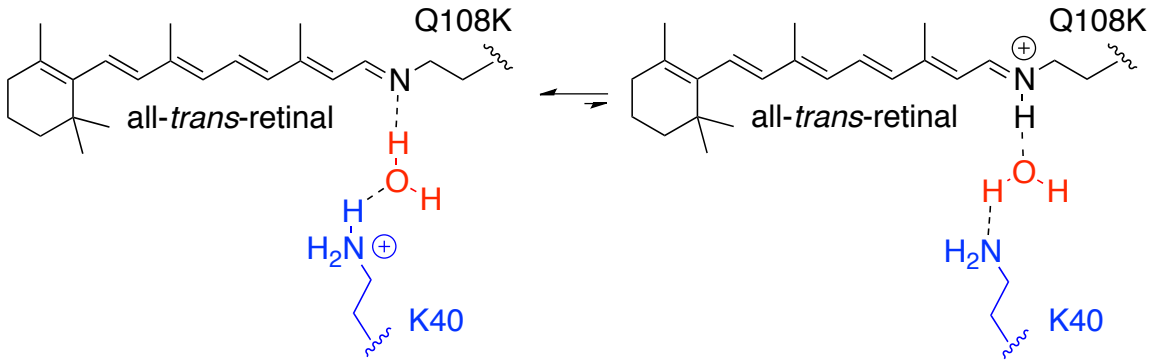


Figure II-15: Proposed mechanism for proton transfer in single mutant Q108K.

units higher than that of protonated Schiff base. This could explain the major absorption peak at ~365 nm in single mutant Q108K, corresponding to the unprotonated Schiff base. The minor absorption peak at 506 nm for single mutant Q108K corersponds to the right hand side of the equilibrium, with neutral K40 and the protonated Schiff base. PSB absorption is observed in UV-vis spectrum of single mutant Q108K, only when a neutral lysine residue is present in the binding pocket. In this way, the electronic environment with either a neutral lysine residue K40 or L40 are similar. As such it is possible that the absorption maximum for Q108K:K40L does not change much from that of Q108K, as observed.

Crystal structure of Q108K:K40L bound with retinal was solved (Rafida Nossoni, Professor Jim Geiger's lab, MSU). The electron density clearly showed the formation of the Schiff base between retinal and Q108K. As expected, the

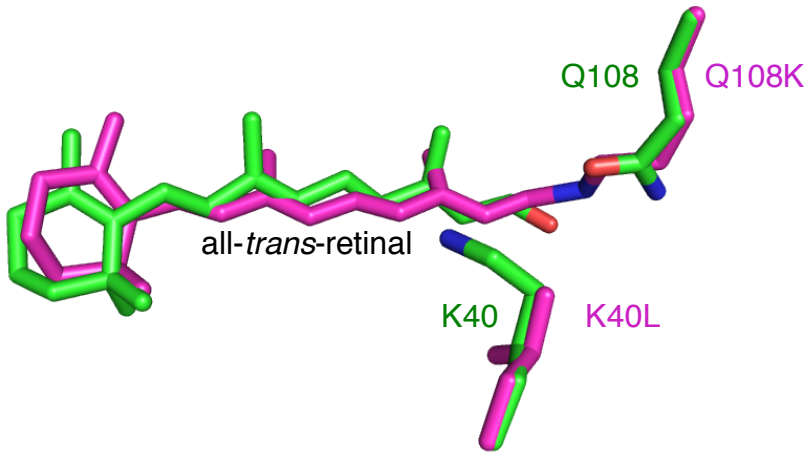


Figure II-16: Overlaid structure of WT-CRBPII-retinal (green) and Q108K:K40L-retinal (magenta).

chromophore was fully embedded inside the binding pocket and surrounded by side chains of the protein as shown in **Figure II-17**. This development sets the

stage for a number of different mutagenesis to study the cause-effect relationships in wavelength regulation.

Overlaying the crystal structures of WT-CRBPII and Q108K:K40L bound with retinal shows that the retinal rotates slightly in the case of Q108K:K40L from WT-CRBPII and the end of the chromophore is pulled up by Q108K as a result of Schiff base formation, as shown in **Figure II-16**. As expected, the position of the chromophore relative to the protein residues does not change much, because the binding pocket of CRBPII is rigid. The chromophore adopts a 6-s-*trans* conformation similar to that in the WT protein. The 6-s-*cis*, with the ionone ring 68° out of plane to avoid the steric clash of C5-methyl group and C8-hydrogen, was calculated to be 0.6 Kcal/mol more stable than the 6-s-*trans* conformation.[29] The preference for 6-s-*trans* in CRBPII is a result of packing between the chromophore and the protein, which will be elaborated later in detail.

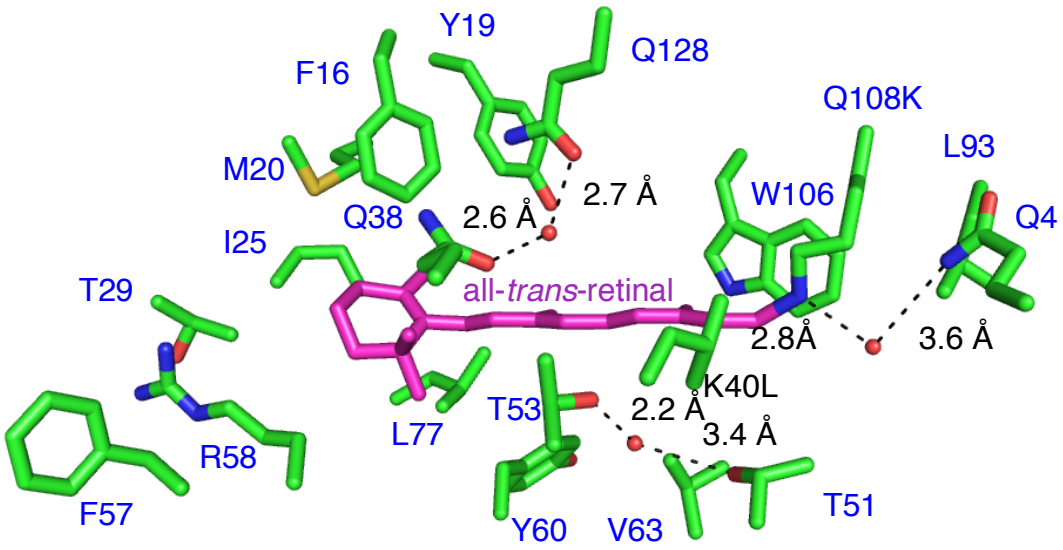


Figure II-17: Crystal structure of Q108K:K40L bound with all-*trans*-retinal, with the surrounding residues and water molecules highlighted.

With the crystal structure of Q108K:K40L bound with retinal, three factors can be considered to explain the 68 nm protein shift observed with Q108K:K40L, compared to retinal-PSB in solution.

Firstly, the chromophore is fully embedded inside the hydrophobic binding pocket of Q108K:K40L. Consequently, the low dielectric constant of the binding pocket could make electrostatic interactions between the surrounding residues inside the protein pocket and the chromophore to have a more significant impact. There are a few polar residues that line along the polyene, such as the two water molecules hydrogen bonded to Gln38 and Gln128 (**Figure II-17**). These polar residues could promote charge delocalization of the iminium charge and lead to red shift observed in Q108K:K40L. Mutagenesis studies can be carried out to check the effects of these polar residues.

Secondly, no obvious counteranion is found in the binding pocket of Q108K:K40L, while for retinal-PSB in buffer solution or first generation rhodopsin mimic, there is an anion or a glutamate residue that is the counter-anion for the PSB. It has been shown previously that introduction of a negatively charged residue could greatly stabilize the positive charge on iminium and lead to less delocalization of the positive charge.[30] Introduction of a negatively charged residue in the Schiff base region stabilizes the ground state of PSB more and at the same time destabilize the excited state of the PSB, due to the negative charge transfer toward the ionone ring of the chromophore in the excited state. This scenario would lead to an overall blue shift of absorption. We have also

shown that placing a negatively charged residue near the PSB leads to a dramatic blue shift (Q108K:T51D), with absorption maximum at 474 nm.

Thirdly, the 6-s-*trans* conformation of the retinal-PSB could also contribute to the red shift of Q108K:K40L compared to the free retinal-PSB in solution and also the first generation rhodopsin mimic as well. The retinal-PSB bound with Q108K:K40L adopts a 6-s-*trans* conformation, with the ionone ring twisted less than 30° out of the plane of the polyene. In solution, a mixture of 6-s-*trans* and 6-s-*cis* conformation is found, with 6-s-*cis* being slightly dominant, which is 68° twisted.[21a, 31] The crystal structure of the first generation of rhodopsin mimic R132K:R111L:L121E shows that the chromophore also adopts a 6-s-*cis* conformation and the ionone ring is ~40° out of plane of the polyene. 6-s-*trans* conformation found in Q108K:K40L-CRBPII results in a higher degree of planarization and thus red shifts. This agrees well with the prediction of Garavelli and coworkers, where calculations suggest an 80 nm red shift for transition from 6-s-*cis* to 6-s-*trans* conformation.[29]

Due to combination of the three factors discussed above, Q108K:K40L red shifts to 508 nm. Q108K:K40L has a pK_a of 8.3, ~2 pK_a units higher than that in solution, although no negatively charged amino acid is introduced to act as a counter anion for the retinal-PSB in Q108K:K40L. From analysis of crystal structure, there is no apparent solute anion in direct interaction with the PSB either. This is different from most of the natural vertebrate rhodopsins and

microbial rhodopsins, which have negatively charged amino acids such as aspartate or glutamate as counteranion. Invertebrate rhodopsins, like squid rhodopsin, use tyrosine as the counteranion and have a pK_a of ~10.[32, 33] The mechanism for CRBPII to stabilize the protonated state of the PSB is unique and remains to be fully understood.

In the crystal structure of Q108K:K40L, the imine adopts a *cis* conformation, although *trans*-imine is more stable than *cis*-imine, due to less steric hindrance, as observed in most rhodopsin crystal structures. As shown in **Figure II-18**, only *cis*-imine can form hydrogen bonding interactions indirectly with Gln4 through a water molecule. The electron density for the water molecule is not clear in the case of double mutant Q108K:K40L, which shows up in some crystal structures obtained but not all. In the crystal structures of other CRBPII mutants when Gln4 is present, the electron density for the water molecule is clear and does exist. This interaction can contribute to the stability of protonated Schiff base, as well as the *cis*-imine.

Besides the latter hydrogen bonding interaction, Trp106, 3.9 Å away from the iminium nitrogen and 3.9 Å away from C15, the carbon directly attached to imine nitrogen, can also contribute to the stability of retinal-PSB. W106 faces C15 and the possible position of iminium hydrogen with its π electron cloud. This conformation of tryptophan could stabilize the protonated state of PSB through π -cation interaction. A combination of these two interactions and other protein-chromophore interactions could increase the pK_a to 8.3.

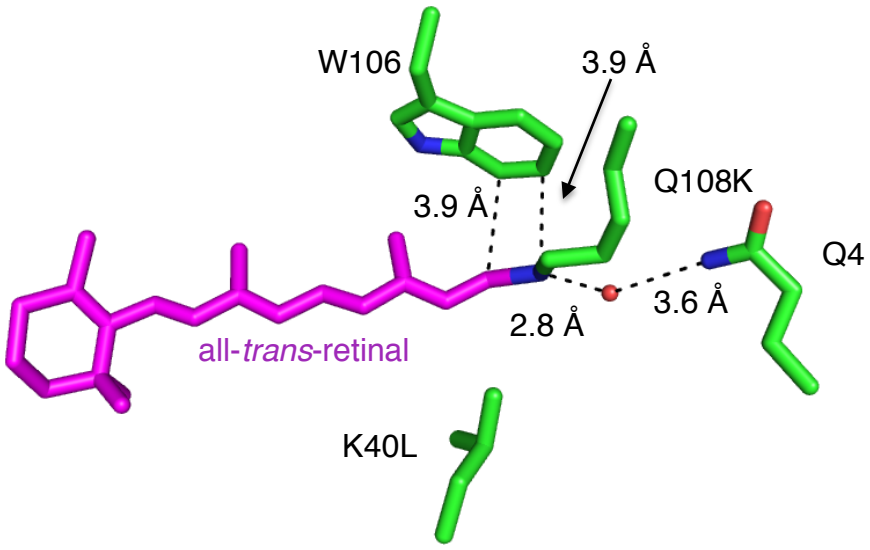


Figure II-18: Highlighted water-mediated hydrogen bonding interactions that stabilize the PSB in Q108K:K40L.

In summary, Q108K:K40L is stable, has high protein expression yield and a pK_a of ~8. It also has a good binding affinity for retinal, with a dissociation constant of 29±5 nM and faster Schiff base formation kinetics compared to Q108K:T51D. It forms PSB with retinal completely within 30 minutes, as compared to Q108K:T51D, which requires ~2 h to form the PSB. With all these characteristics, Q108K:K40L was deemed a good platform to study wavelength regulation.

II.8 Other mutations of K40

Since the importance of removing K40 was confirmed, other amino acids were also tried at position 40 to see their effects on wavelength regulation and the pK_a of the retinal-PSB.

Acidic residues, K40E and K40D were introduced, but the protein became unstable and no soluble protein could be obtained. In CRBPII, introduction of acidic residues inside the binding pocket usually disturbed the folding of the protein and resulted in inclusion bodies. Neutral polar residues such as asparagine and serine were tried. Double mutants Q108K:K40N and Q108K:K40S were successfully expressed in soluble form.

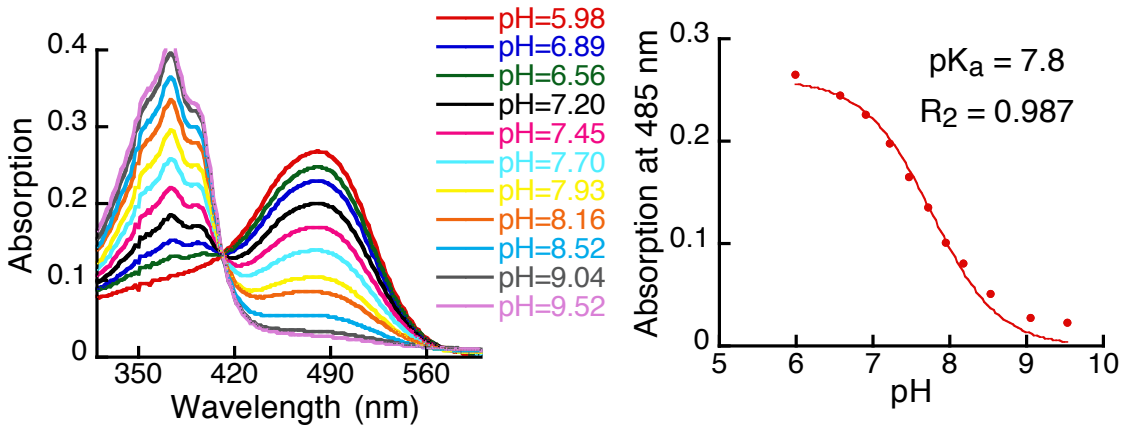


Figure II-19: UV-vis spectra of base titration of Q108K:K40N.

As expected, introduction of polar residues at position 40, which is close to the PSB, blue shifts 28 nm relative to Q108K:K40L and yields an absorption of 480 nm for Q108K:K40N. Although Asn is neutral, its negative polarity could still have a dramatic effect on the electrostatic projection on of the chromophore.

K40N introduces more negative polarity close to the Schiff base region, stabilizing the positive charge localized on iminium, resulting in a blue shift. The pK_a of the retinal-PSB formed with Q108K:K40N is 7.8, close to that of Q108K:K40L.

Q108K:K40S also blue shifts 26 nm relative to Q108K:K40L to 482 nm, similar to Q108K:K40N. Introduction of hydroxyl group containing polar residues in the Schiff base region could project more negative polarity in the PSB region and decrease the positive charge delocalization, leading to a blue shift. Ser is considered to be more polar than asparagine. However, these two mutations, K40S and K40N lead to similar degree of blue shift. This could be because serine is one carbon shorter than asparagine and thus further away from PSB compared to asparagine, which reduces the electric field projected on the chromophore, even though it is more polar. As a result, the electric field exerted on the chromophore from serine could be similar to asparagine and have similar effect in stabilizing the positive charge localized on the iminium.

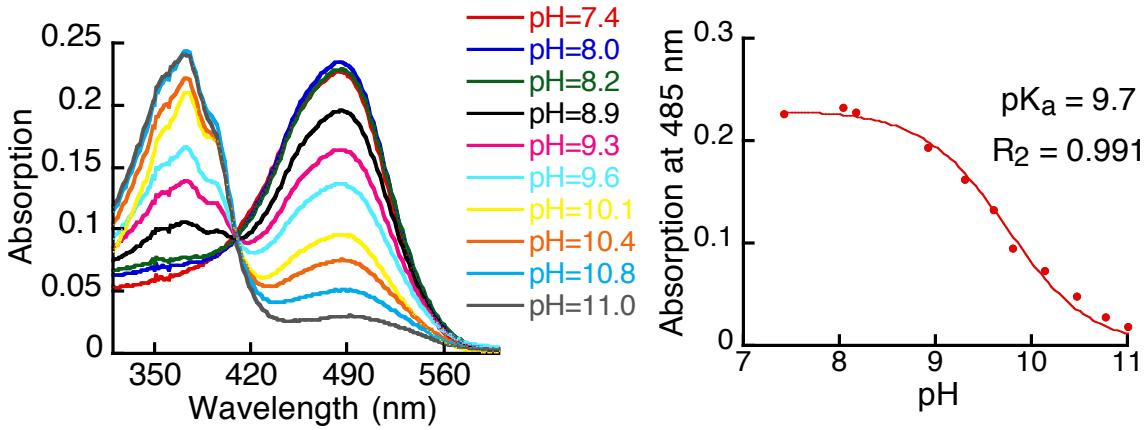


Figure II-20: Base titration of Q108K:K40S.

Q108K:K40S has a higher pK_a than both Q108K:K40N and Q108K:K40L (**Figure II-20**). Unfortunately, the crystal structure of Q108K:K40S could not be resolved to help analyze the factors that contribute to its increase in pK_a. It is possible that the decrease of volume by mutation K40S creates an empty space in the binding pocket, which allows a water molecule to be situated in that space and form hydrogen bonding interactions with K40S (**Figure II-21**). That same water molecule can also hydrogen bond with retinal-PSB to stabilize the PSB.

The crystal structure of Q108K:K40L shows that the imine of retinal-PSB adopts a *cis* conformation, with the lone pair of nitrogen rotating away from K40S. *Cis*-imine conformation can not interact with K40S directly or indirectly in any manner. To achieve the water-mediated interaction as shown in **Figure II-21**, the imine has to adopt the *trans* conformation to have a tight interaction with K40S through the water-mediated network. This means that the water-mediated interaction with Gln4, which stabilizes the *cis*-imine of retinal-PSB in Q108K:K40L, has to be interrupted. Hydroxyl group is considered to be a better

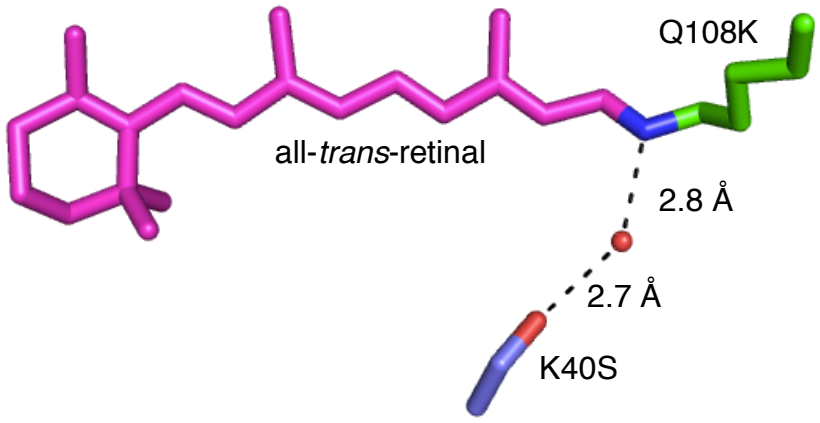


Figure II-21: Model structure of Q108K:K40S.

hydrogen bonding acceptor than amide, therefore K40S hydrogen bonding interaction could be more favorable than the Gln4 interaction and it could better stabilize the PSB protonation state through a much tighter hydrogen bonding interaction and increase its pK_a.

There could be other reasons that contribute to the increased pK_a of Q108K:K40S compared to Q108K:K40L. Previous studies have shown that increasing the polarity of the protein environment can also promote the charged state of a molecule.[34] Therefore, placing a more polar residue (K40S) and introduction of a putative water molecule in the PSB region can also increase pK_a of the retinal-PSB.

Up to now, I have shown that CRBPII has been engineered into a protein that can bind retinal as a protonated Schiff base through introduction of a lysine residue that can take a favorable Bürji-Duniz trajectory. Introduction of counteranion and negative polar residues in the Schiff base region leads to blue shift, as it can stabilize the positive charge localized on the iminium, and decrease positive charge delocalization. Introduction of a counteranion (T51D) or replacing the positively charged K40 with a hydrophobic residue (K40L) and negative polar residue (K40S) could both increase the pK_a of the retinal-PSB.

Materials and methods

UV-vis spectra were recorded using a Cary 300 Bio WinUV, Varian spectrophotometer. Fluorescence was recorded with a Fluorolog-3 (Instruments S. A., Inc.) fluorometer. All-*trans*-retinal was purchased from TRC and was used as received.

The second batch of retinal purchased from TRC was contaminated and retinal was degraded within a day once dissolved in ethanol even at -78 °C, indicated by an increase in the absorption peak at ~330 nm. Flash chromatography through silica gel was performed to purify the retinal before use.

Plasmid construction

CRBPII gene was cloned into pETBlue-2 vector through restriction sites NcoI and XhoI, and also cloned into pET17b vector through restriction sites NdeI and XhoI, from CRBPII in pET17b inserted between EcoRI and XhoI.

Cloning CRBPII **gene into pETBlue vector**

PCR was performed to mutate the stop codon to the codon of glycine, along with introduction of the NcoI cutting site (The stop codon in CRBPII needs to be mutated in order to express the six histidine residues in pETBlue vector).

Double digestion was carried out by incubating the reaction solutions in **Table II-2** at 37 °C for 3 h. Agarose gel (0.85% Agarose) DNA electrophoresis was performed for both reactions separately to purify the digested gene and vector, followed by extraction of DNA from the corresponding DNA band in agarose gel, using *QIAquick gel extraction kit*.

Table II-2: Double digestion of pETBlue-2 vector and CRBPII (between NcoI sites and XhoI)

	DNA	XhoI restriction enzyme	NcoI restriction enzyme	Buffer (10x)	Water (dd)	Total volume
pETBlue-2	10 µg	30 units	30 units	5 µl	/	50 µl
CRBPII/p17b	10 µg	30 units	30 units	5 µl	/	50 µl

Concentrations of the double digested gene and vector were estimated by DNA electrophoresis, roughly comparing the brightness of the DNA of interest with the standard double digested vector and gene with a gradient of concentrations (or Nanodrop).

(Estimation of the molecular weight of DNA = 134 x 3bp x 330 = 1.3×10^5

Molecular weight of vector = (3653-124) x 330 = 1.2×10^6)

Table II-3: Ligation reactions for 200 ng vector scale

	Amount of Vector	Amount of Gene	Molar ratio of gene:vector	Buffer (5x)	T4 DNA ligase	Water (dd)	Total volume
A	200 ng	100 ng	5:1	4 µl	2 µl	/	20 µl
B	200 ng	200 ng	10:1	4 µl	2 µl	/	20 µl
C	200 ng	300 ng	15:1	4 µl	2 µl	/	20 µl

Ligation reactions (**Table II-3**) were incubated at 16 °C for 12 h or at 25 °C for 2-3 h. Ligation reactions were heat inactivated at 60 °C for 2 min and then

transformed to XL1-Blue competent cells (5 µL ligation solution + 50 µL XL1-Blue competent cells) by heat shock at 42 °C. Five clones were picked and inoculated overnight in 10 mL LB solution with final concentrations of Ampicillin at 100mg/L and Tetracyclin at 12.5 mg/L. Plasmid was purified using Qiagen DNA Miniprep Kit and double digestion of the purified DNA with XhoI (NEB) and NcoI (NEB) was performed at 37 °C for 2 h, followed by DNA electrophoresis. The plasmids with a digested band with the size of CRBPII gene were sent for sequencing to verify the correct clones.

Similar procedures were performed for ligation of CRBPII into pET17b between the NdeI and XhoI sites. It is important to make sure that both the start codon and stop codon of the gene is in frame with the start codon of the plasmid.

Mutagenesis

All the mutations were made using ***Stratagene's QuickChange Site-Directed Mutagenesis Protocol.***

Table II-4: PCR Reaction solution

DNA template	Primer forward	Primer reverse	Turbo pfu DNA polymerase	dNTP(10 mM of each)	Buffer (10x)	Water (dd)	Total volume
70 ng	20 pmol	20 pmol	1 unit	1 µL	5 µL	/	50 µL

Table II-4 continued

PCR temperature control cycles

1 x	95 °C	3 min
20 x	95 °C	30 sec
	(T_m-4) °C	1 min
	72 °C	3.5 min
1x	72 °C	10 min
1x	25 °C	5 min

Note: T_m stands for the melting temperature of the primer, which is calculated using the following website.

http://www.promega.com/techserv/tools/biomath/calc11.htm

For some of the mutations, repeated primer sequences were inserted into the DNA following the primer region. If this happens, annealing temperature can be raised to (T_m-2) °C or (T_m-3) °C and the elongation time of the third step at 72 °C can be shortened to 3 min and 10 sec.

WT-CRBPII gene sequence

ATGACGAGGGACCAGAATGGAACCTGGGAGATGGAGAGTAATGAAAACTTT
GAGGGCTACATGAAGGCCCTGGATATTGATTTTGCCACCCGCAAGATTGCA
GTACGTCTCACTCAGACGCTGGTTATTGATCAAGATGGTGATAACTTCAAGA
CAAAAACCACTAGCACATTCCGCAACTGGGATGTGGATTTCACTGTTGGAGT
AGAGTTTGACGAGTACACAAAGAGCCTGGATAACCGGCATGTTAAGGCACT
GGTCACCTGGGAAGGTGATGTCCTTGTGTGTGTGCAAAAGGGGGAGAAGGA
GAACCGCGGCTGGAAGAAGTGGATTGAGGGGGACAAGCTGTACCTGGAGC

TGACCTGTGGTGACCAGGTGTGCCGTCAAGTGTTCAAAAAGAAG**TGA**

WT-CRBPII amino acid sequence

T R D Q N G T W E M E S N E N F E G Y M K A L D I D F A T R K I A V R L T Q

T L V I D Q D G D N F K T K T T S T F R N W D V D F T V G V E F D E Y T K S

L D N R H V K A L V T W E G D V L V C V Q K G E K E N R G W K K W I E G D

K L Y L E L T C G D Q V C R Q V F K K K

Primers used for mutagenesis:

Stop codon to Gly

Forward: 5'-GTTCAAAAAGAAG**GGA**CTCGAGCAG-3'

Reverse: 5'-CTGCTCGAG**TCC**CTTCTTTTTGAAC-3'

NcoI site modification

Forward: 5'-GGATCCGAAT**CCATGG**CGAGGGACCAG-3'

Reverse: 5'-CTGGTCCCTCG**CCATGG**ATTCGGATCC-3'

NdeI site modification

Forward: 5'-GTGTGCTGGAA**CATATG**ACGAGGAC-3'

Reverse: 5'-GTCCTCGT**CATATG**TTCCAGCACAC-3'

Q108K

Forward: 5'-CCGCGGCTGGAAG**AAG**TGGATTGAGGGGG-3'

Reverse: 5'-CCCCCTCAATCCA**CTT**CTTCCAGCCGCGG-3'

K40L

Forward: 5'-CTCACTCAGACG**CTG**GTTATTGATCAAGATGG-3'

Reverse: 5'-CCATCTTGATCAATAAC**CAG**CGTCTGAGTGAG-3'

K40S

Forward: 5’-CTCACTCAGACG**TCG**GTTATTGATCAAGATGG-3’

Reverse: 5’-CCATCTTGATCAATAAC**CGA**CGTCTGAGTGAG-3’

T51D

Forward: 5’-GGTGATAACTTCAAG**GAT**AAAACCACTAGCAC-3’

Reverse: 5’-GTGCTAGTGGTTTT**ATC**CTTGAAGTTATCACC-3’

T51V

Forward: 5’-GGTGATAACTTCAAG**GTA**AAAACCACTAGCAC-3’

Reverse: 5’-GTGCTAGTGGTTTT**TAC**CTTGAAGTTATCACC-3’

T53C

Forward: 5’-CAAGACAAAA**TGC**ACTAGCACATTCCG-3’

Reverse: 5’-CGGAATGTGCTAGT**GCA**TTTTGTCTTG-3’

T51V:T53C

Forward: 5’-CAAG**GTA**AAA**TGC**ACTAGCACATTCCG -3’

Reverse: 5’-CGGAATGTGCTAGT**GCA**TTT**TAC**CTTG -3’

Y60W

Forward: 5’-CACATTCCGCAAC**TGG**GATGTGGATTTCAC-3’

Reverse: 5’-GTGAAATCCACATC**CCA**GTTGCGGAATGTG-3’

A33W

Forward: 5’-CGCAAGATT**TGG**GTACGTCTCAC-3’

Reverse: 5’-GTGAGACGTAC**CCA**AATCTTGCG-3’

R58W

Forward: 5'-CTAGCACATTC**TGG**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**CCA**GAATGTGCTAG-3'

R58F

Forward: 5'-CTAGCACATTC**TTC**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**GAA**GAATGTGCTAG-3'

R58Y

Forward: 5'-CTAGCACATTC**TAC**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**GTA**GAATGTGCTAG-3'

T29L

Forward: 5'-GATTTTGCC**CTG**CGCAAGATTGC-3'

Reverse: 5'-GCAATCTTGCG**CAG**GGCAAAATC-3'

Y19W

Forward: 5'-CTTTGAGGGC**TGG**ATGAAGGC-3'

Reverse: 5'-GCCTTCAT**CCA**GCCCTCAAAG-3'

Q4W

Forward: 5'-GACGAGGGAC**TGG**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**CCA**GTCCCTCGTC-3'

Q4F

Forward: 5'-GACGAGGGAC**TTC**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**GAA**GTCCCTCGTC-3'

Q4R

Forward: 5'-GACGAGGGAC**AGG**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**CCT**GTCCCTCGTC-3'

Q4K

Forward: 5'-GACGAGGGAC**AAG**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**CTT**GTCCCTCGTC-3'

Q4H

Forward: 5'-GACGAGGGAC**CAC**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**GTG**GTCCCTCGTC-3'

XL1-Blue competent cells preparation

XL1-Blue *E.coli* cells were streaked on an agar plate with tetracycline (12.5 mg/L) and incubate at 37 °C overnight, using a sterile wood stick.

One colony was picked from the plate and innoculated in sterile LB solution (10 mL), with tetracycline (12.5 mg/L), in a 37 °C shaker at 220 RPM overnight. The overnight culture (1 mL) was transferred to sterile LB solution (200 mL) with tetracycline (12.5 mg/L) and it was kept shaking at 220 RPM at 37 °C for about 2 h until the OD_{600nm} reached a value between 0.4 and 0.8. At the same time, a 500 mL centrifuge bottle was sterilized with 20% bleach for 2 h and rinsed with sterile water 6 to 7 times before use.

The cells were harvested by centrifugation at 3500 RPM for 5 min at 4 °C. The cells were resuspended with sterile sodium chloride solution (0.9%, 100 mL). The cells were harvested by centrifugation at 3000 RPM for 4 min at 4 °C.

From this stage on, two protocols were followed.

A: The cell pellet was resuspended with calcium chloride solution (100 mM, 15% Glycerol v/v, 10 mL) on ice and was incubated on ice for at least half an hour. The cells were fractioned in sterile eppendorf tubes in portions of 100 µL and flash frozen with liquid nitrogen.

B: The cell pellet was resuspended with calcium chloride solution (100 mM, 50 mL) and incubated on ice for 30 min. The cells were spun down by centrifugation again at 3000 RPM for 4 min and then resuspended with calcium chloride solution (100 mM, 15% Glycerol v/v, 10 mL) on ice. The cells were fractioned in sterile eppendorf tubes in portions of 100 µL and flash frozen with liquid nitrogen.

Note: Protocol A has one less step, incubation with calcium chloride solution (100 mM, 50 mL), and therefore the competency was lower than that obtained from protocol B. XL1-Blue competent cells are used for transformation of the PCR product and need to be highly competent. This is because the DNA from PCR reaction is nicked and does not go through the cell membrane as easily as circular DNA during heat shock. Therefore, it is better to follow protocol B to prepare XLBlue-2 competent cells. Preparation of BL21 competent cells follows the same method as XLBlue cells, except that BL21 cells is resistant to chloramphenicol.

Note that the cells become fragile after addition of calcium chloride. Cells should be handled with gentleness on ice after treatment with calcium chloride.

Heat shock transformation

The DpnI digested PCR solution (5 µL) was added to competent cells (50 or 100 µl). The cells were incubated on ice for 10 to 30 min, followed by heat shock at 42 °C for 50 sec. The cells were put back on ice immediately.

Sterile LB solution (500 µL) was added to the cells and shaken at 37 °C for 1 h. (This step is optional. Only when the competency of the competent cells is low or the PCR reaction is not successful, such as a weak band is observed in the DNA gel, LB solution (500 µL) would be added to the cells and incubated, after heat shock. Alternatively, right after heat shock and incubation on ice for 2 min, the cells are plated directly on agar plates.)

The cells were harvested by centrifugation at 5000 RPM for 1 min. The cells were resuspended in LB (50 µL) and then plated on agar plate with the appropriate antibiotics.

Protein expression and purification in pET-17b system

CRBPII DNA was transformed into BL21 (DE3) pLysS competent cells. A single colony was picked from the transformation plate and innoculated in 10 mL of LB with a final concentration of ampicillin at 100 mg/L and chloramphenicol at 27 mg/L at 37 °C overnight. Next 2 mL of the overnight culture was transferred to 1 L of LB media with the same concentration of ampicillin and chloramphenicol and the media was incubated at 37 °C until OD_{600} reached 0.4-0.8. Isopropyl-1-thio-D-galactopyranoside (IPTG, Gold Biotechnology) was added to a final concentration of 1 mM to induce protein expression, and the cell culture were

shaken at RT or 16 °C overnight.

Cells were harvested by centrifugation at 5000 RPM for 10 min in Beckmann J2-21M/E centrifuge and resuspended with 50 mL Tris binding buffer (10 mM Tris, pH=8.0), followed by 3 min of sonication to cause cell lysis. DNase (20 µL, Roche, recombinant, 25 unit/µL) was added to the lysed solution and kept on ice for half an hour. The lysed cell solution was spun down for 15 min at 8000 RPM. The supernatant was applied to a fast Q ion exchange column, which was pre-equilibrated (washed with stripping buffer, 40 mL of 2 M NaCl, and equilibrated with 100 mL of Tris binding buffer). Protein was washed with Tris binding buffer and eluted with the elution buffer (10 mM Tris, 100 mM Sodium Chloride, pH=8.0). The eluted protein solution was desalted with Amicon filter membrane (cutoff 10 kDa) and then applied to a source Q Fast Protein Liquid Chromatography (FPLC). The protein was subject to a gradient of sodium chloride solution from 0 mM to 1 M at pH=8.0 (25 mM Tris and Tris•HCl), the protein of interest was eluted at ~40 mM NaCl.

Protein purification in pETBlue-2 system

Ni-NTA column was washed with 20 mL stripping buffer (100 mM EDTA, 500 mM NaCl, 20 mM Tris, pH=8.0), recharged with 20 mL Ni solution (50 mM $NiSO_4$), followed by two times of washing with 20 mL binding buffer (pH=8.0, 10 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4). The column was ready to be used.

The cells were harvested by centrifugation at 5000 RPM for 10 min in a

Beckmann J2-21M/E centrifuge and resuspended with 25 mL binding buffer (pH = 8.0, 10 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4), followed by sonication (3 time, 1 min per pulse, pulse=60%, power 60%). DNase (20 μL, Roche, recombinant, 25 unit/μL) was added to the lysed solution and incubated on ice for half 1 h. The lysed solution was spun down for 20 min at 8,000 RPM. The supernatant was applied to the pre-equilibrated Ni-NTA column, washed with washing buffer twice (pH=8.0, 30 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4, 20 mL), and eluted with 20 mL elution buffer (pH=8.0, 250 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4).

The eluent from the Ni-NTA column was concentrated down to less than 3 mL and then applied to a pre-equilibrated sephadex size exclusion column (G-15) and phosphate buffer (20 mM NaH_2PO_4, 150 mM NaCl, pH=7.3) was applied to elute the protein. Fractions were collected in 150 drops / tube. UV-vis spectra or SDS-PAGE was used to detect the tubes containing expected protein other than imidazole. The protein usually comes out in tubes 5-9.

Alternatively, filters that have a mass cutoff of 10,000 Da could be used to get rid of imidazole.

Thrombin digestion and purification

For 6-His-tag protein with thrombin cleavage site (Leu-Val-Pro-Arg-Gly-Ser) introduced, thrombin digestion was performed after purification of 6-His-tag protein from sephadex column.

First the sephadex eluent that contained the desired protein was concentrated down to around 10 mL and thrombin was added and incubated at 4 °C for 4 h. The digested protein was applied to a Ni-NTA column, which was pre-equilibrated with PBS buffer. The flow through was collected. The cleaved short peptide with 6-His-tag and the uncut protein will bind to the Ni-NTA column and only the cut protein will be eluted.

Inclusion bodies purification

Solutions for inclusion bodies purification

Pellet washing solution:	50 mM Tris, 2 M urea, 500 mM NaCl and 2% Triton X-100, pH=8.0
Solubilizing buffer:	70 mM NaH_2PO_4, 6 M guanidine HCl, 14 mM β-mercaptoethanol, 80 mM NaCl, pH=7.8
Refolding buffer:	1mM EDTA, 2 mM Tris•HCl, 0.2 mM DTT, pH=9.0
Denaturing binding buffer:	8 M urea, 20 mM sodium phosphate, 500 mM NaCl, pH=8.0
Denaturing washing buffer:	8M urea, 20mM sodium phosphate, 500mM NaCl, pH=6.0
Denaturing eluting buffer:	8 M urea, 20 mM sodium phosphate, 500 mM NaCl, pH=4

Cells were harvested by centrifugation at 5000 RPM for 10 min, resuspended with 70 mL Tris buffer (10 mM Tris, pH=8.0), followed by sonication

(1 min x 4). Lysed solution was spun down at 5000 RPM for 10 min, washed twice with pellet washing solution, and then dissolved in 20 mL of solubilizing buffer. The solubilized pellet was centrifuged at 8000 RPM for 20 min and the supernatant was applied to 3 mL of denatured Ni-NTA column. The column was washed with 2 column volumes of denaturing washing buffer and eluted with 15 mL of denaturing eluting buffer.

Eluted solution was added to 20 fold refolding buffer at 4 °C, while mechanically stirring vigorously. After that, the solution was incubated at 4 °C overnight. Refolded protein was concentrated down and applied to ion exchanged column (fast Q colum) to get rid of urea.

Extinction coefficient determination of CRBPII proteins

Gill and von Hippel method is used to determine protein extinction coefficient.[35]

For denatured protein: $\varepsilon_d = 5690 \times Try + 1280 \times Tyr + 120 \times Cys$

In WT CRBPII, there are 4 Try, 4 Tyr and 2 Cys.

So the extinction coefficient for WT CRBPII is:

$\varepsilon_d = 5690 \times 4 + 1280 \times 4 + 120 \times 2 = 28{,}120$

For different mutants, ε_d was determined according to the latter protocol.

The absorptions of native protein in PBS and denatured protein in 6 M guanidine hydrogen chloride salt at 280 nm were measured at the same protein concentration.

The extinction coefficient for the native protein is derived as:

$\varepsilon_n = \varepsilon_d \times (A_{280}$ native$) \div (A_{280}$ denatured).

Fluorescence titration

The dissociation constant was determined by fluorescence titration as previously described in our lab.[36] Briefly, retinal was added to 0.5 μM of protein solution in different aliquots and the fluorescence of tryptophan was recorded at 352 nm when excited at 283 nm. Titration was stopped when the fluorescence remained constant. The data was corrected by subtracting the quenching effect from free ligand, as shown below after completion of both the protein titration and the blank titration.

1. The value of α for every point on the curve is determined as follows:

$$\alpha = \frac{F_{max} - F}{F_{max} - F_0}$$

F_{max} = fluorescence upon saturation

F_o = initial fluorescence

F = observed fluorescence

a = fraction of free binding sites

2. The free ligand concentration, R, was determined as

$$R = R_o - nP_o(1\text{-}a)$$

R_o = ligand concentration

n = number of binding sites / protein, assume n=1

P_o = protein concentration

3. The fluorescence contribution of the free ligand, F_R, was to be deduced from the blank (*N*-acetyltryptophanamide) titration.
4. The fluorescence contribution of the free ligand was subtracted from the actual readings and the corrected data for fluorescence was plotted vs. the ligand concentration.

$(F\text{-}F_R)$ vs. R_o

The plot is then fitted into the following function:

$$(F - F_R) = 1 + \frac{(F_{max} - 1) * [(P_0 + R_0 + K_d) - \sqrt{(P_0 + R_0 + K_d)^2 - 4 * P_0 * R_0}]}{2 * P_0}$$

UV-vis base titration of CRBPII mutants bound with all-*trans*-retinal

Protein solution (1 mL) with concentrations ranging from 5 μM to 40 μM was prepared in UV cuvets. Retinal (0.5 equiv) was added to the protein solution and incubated for half an hour to 4 h until the PSB formation reached its maximum. The protein solution was acidified using citric acid (1 M, ~10 μL) to pH ~5, in order to fully convert the Schiff base absorption peak at ~360 nm to the protonated Schiff base peak. Sodium hydroxide (1 M) was added in small portions to the protein solution to increase the pH gradually and the corresponding UV-vis spectra were recorded at each pH.

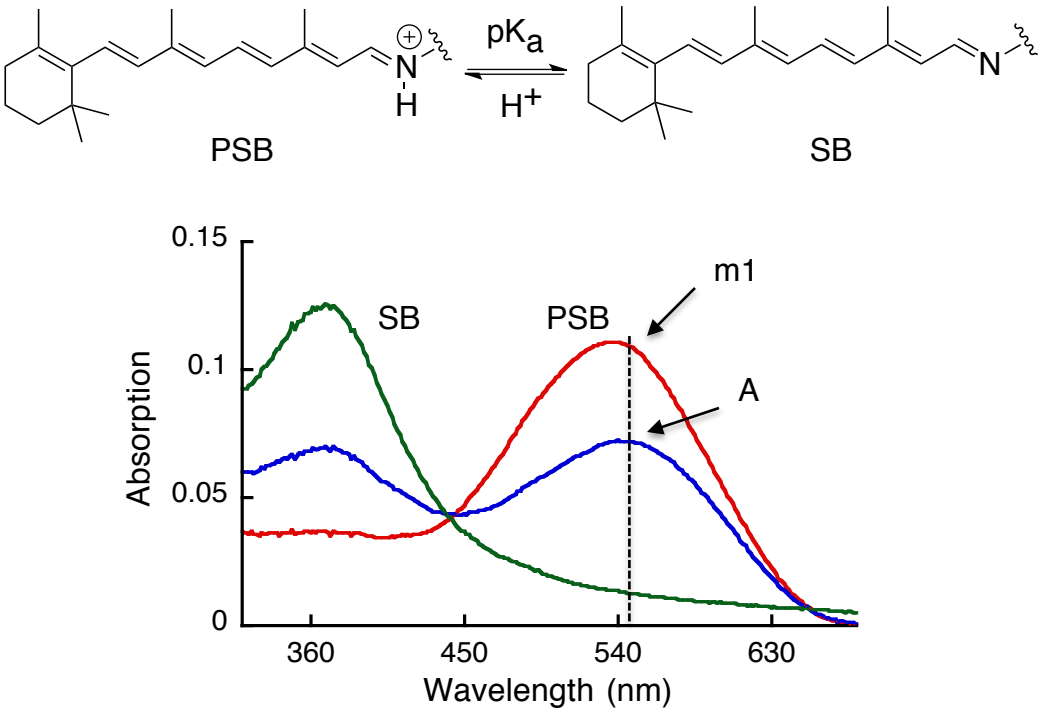


Figure II-22: Illustration of base titration of retinal-PSB.

The absorptions at selected wavelengths near the λ_{max} were plotted against the pH and fitted into the following function to obtain the pK_a value,

$$A = \frac{m1}{1 + 10^{(pH - pK_a)}}$$

The derivatization is as follows:

Since

$$pK_a = pH + \log \frac{[PSB]}{[SB]}$$

We can get

$$\frac{[PSB]}{[SB]} = 10^{(pKa - pH)}$$

Since in all cases of CRBPII mutants, the PSB formed are much more red shifted than SB or retinal, the PSB absorptions near their λ_{max} do not overlap with the absorptions of SB. Therefore, absorption A at a wavelength close to the PSB λ_{max} or beyond is only due to PSB. The correlation of A with the concentration of PSB, [PSB], is

$$A = [PSB] * \varepsilon$$

ε stands for the extinction coefficient of PSB

Assuming that when all of the Schiff base (SB) is protonated to be PSB, it has absorption of **m1** at a selected wavelength close to the PSB λ_{max} as shown in **Figure II-22**, then at a specific pH value, the ratio of the PSB versus SB can be correlated to the absorption of PSB at that specific wavelength, A, as described by the following equation:

$$\frac{[PSB]}{[SB]} = \frac{A/\varepsilon}{(m1 - A)/\varepsilon} = \frac{A}{(m1 - A)}$$

Combination of the latter two equations leads to the following:

$$A = \frac{m1}{1 + 10^{(pH - pK_a)}}$$

References

References

1. Wald, G., Mechanism of human color vision. *Am. J. Ophthalmol.* **1965,** *60* (6), 1132-&.

2. Wald, G., Receptors of human color vision. *Science* **1964,** *145* (363), 1007-&.

3. Brown, P. K.; Wald, G., Visual pigments in single rods + cones of human retina - direct measurements reveal mechanisms of human night + color vision. *Science* **1964,** *144* (361), 45-&.

4. Kim, S. Y.; Waschuk, S. A.; Brown, L. S.; Jung, K. H., Screening and characterization of proteorhodopsin color-tuning mutations in Escherichia coli with endogenous retinal synthesis. *Bba-Bioenergetics* **2008,** *1777* (6), 504-513.

5. (a) 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. *J. Mol. Biol.* **2006,** *363* (3), 687-701; (b) 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), 1241-1258.

6. 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. *J. Am. Chem. Soc.* **2007,** *129* (19), 6140-6148.

7. Burgi, H. B.; Dunitz, J. D., From crystal statics to chemical-dynamics. *Accounts Chem. Res.* **1983,** *16* (5), 153-161.

8. (a) Crist, R. M.; Vasileiou, C.; Rabago-Smith, M.; Geiger, J. H.; Borhan, B., Engineering a rhodopsin protein mimic. *J. Am. Chem. Soc.* **2006,** *128* (14), 4522-4523; (b) Vasileiou, C.; Wang, W. J.; Jia, X. F.; 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* **2009,** *77* (4), 812-822.

9. Lee, K. S. S., PhD. Dissertation, Michigan State University, East Lansing, 2011.

10. Beppu, Y.; Kakitani, T., Theoretical-study of color control mechanism in retinal proteins .1. role of the tryptophan residue, tyrosine residue and water molecule. *Photochem. Photobiol.* **1994,** *59* (6), 660-669.

11. (a) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W., Modeling electrostatic effects in proteins. *Bba-Proteins Proteom.* **2006,** *1764* (11), 1647-1676; (b) Oberoi, H.; Trikha, J.; Yuan, X. L.; Allewell, N. M., Identification and analysis of long-range electrostatic effects in proteins by computer modeling: Aspartate transcarbamylase. *Proteins* **1996,** *25* (3), 300-314.

12. Ong, D. E.; Chytil, F., Cellular retinol-binding protein from rat-liver - purification and characterization. *J. Biolog. Chem.* **1978,** *253* (3), 828-832.

13. Wolf, G., The intracellular vitamin-a-binding proteins - an overview of their functions. *Nutr. Rev.* **1991,** *49* (1), 1-12.

14. Li, E.; Norris, A. W., Structure/function of cytoplasmic vitamin A-binding proteins. *Annu. Rev. Nutr.* **1996,** *16*, 205-234.

15. Gunasekaran, K.; Hagler, A. T.; Gierasch, L. M., Sequence and structural analysis of cellular retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions. *Proteins* **2004,** *54* (2), 179-194.

16. Wolf, G., Cellular retinoic acid-binding protein II: A coactivator of the transactivation by the retinoic acid receptor complex RAR.RXR. *Nutr. Rev.* **2000,** *58* (5), 151-153.

17. Durand, B.; Saunders, M.; Leroy, P.; Leid, M.; Chambon, P., All-trans and 9-cis retinoic acid induction of crabpii transcription is mediated by rar-rxr heterodimers bound to dr1 and dr2 repeated motifs. *Cell* **1992,** *71* (1), 73-85.

18. Dodero, V. I.; Quirolo, Z. B.; Sequeira, M. A., Biomolecular studies by circular dichroism. *Front Biosci-Landmrk* **2011,** *16*, 61-73.

19. Li, E.; Yang, N. C. C.; Ong, D. E.; Gordon, J. I., Characterization of rat cellular retinol-binding protein-ii expressed in escherichia-coli. *J. Biolog. Chem.* **1987,** *262* (28), 13773-13779.

20. Macdonald, P. N.; Ong, D. E., Binding specificities of cellular retinol-binding protein and cellular retinol-binding protein, Type-Ii. *J. Biolog. Chem.* **1987,** *262* (22), 10550-10556.

21. (a) Hamanaka, T.; Kakudo, M.; Ashida, T.; Mitsui, T., Crystal-structure of all-trans retinal. *Acta. Crystall. B-Stru.* **1972,** *B 28* (Jan15), 214-&; (b)

Gilardi, R.; Karle, I. L.; Karle, J.; Sperling, W., Crystal structure of visual chromophores, 11-cis and all-trans retinal. *Nature* **1971,** *232* (5307), 187-&.

22. (a) Winter, N. S.; Bratt, J. M.; Banaszak, L. J., Crystal-structures of holo and apo-cellular retinol-binding protein-II. *J. Mol. Biol.* **1993,** *230* (4), 1247-1259; (b) Tarter, M.; Capaldi, S.; Carrizo, M. E.; Ambrosi, E.; Perduca, M.; Monaco, H. L., Crystal structure of human cellular retinol-binding protein II to 1.2 angstrom resolution. *Proteins* **2008,** *70* (4), 1626-1630.

23. (a) Salonen, L. M.; Ellermann, M.; Diederich, F., Aromatic rings in chemical and biological recognition: energetics and structures. *Angew. Chem. Int. Ed. Engl.* **2011,** *50* (21), 4808-4842; (b) Kong, F. R.; King, J., Contributions of aromatic pairs to the folding and stability of long-lived human gamma D-crystallin. *Protein Science* **2011,** *20* (3), 513-528; (c) Wu, L.; McElheny, D.; Huang, R.; Keiderling, T. A., Role of tryptophan-tryptophan interactions in trpzip beta-hairpin formation, structure, and stability. *Biochemistry* **2009,** *48* (43), 10362-10371.

24. 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. *P. Natl. Acad. Sci. USA* **1991,** *88* (8), 3079-3083.

25. Steinberg, G.; Ottolenghi, M.; Sheves, M., Pk(a) of the protonated schiff-base of bovine rhodopsin - a study with artificial pigments. *Biophys. J.* **1993,** *64* (5), 1499-1502.

26. 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. *J. Am. Chem. Soc.* **1993,** *115* (9), 3772-3773.

27. Baasov, T.; Sheves, M., Alteration of pka of the bacteriorhodopsin protonated schiff-base - a study with model compounds. *Biochemistry* **1986,** *25* (18), 5249-5258.

28. Livnah, N.; Sheves, M., Model compounds can mimic spectroscopic properties of bovine rhodopsin. *J. Am. Chem. Soc.* **1993,** *115* (1), 351-353.

29. 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. *Angew. Chem. Int. Ed. Engl.* **2010,** *49* (10), 1790-1793.

30. Sekharan, S.; Sugihara, M.; Buss, V., Origin of spectral tuning in rhodopsin - It is not the binding pocket. *Angew. Chem. Int. Ed. Engl.* **2007,** *46* (1-2), 269-271.

31. Childs, R. F.; Shaw, G. S.; Wasylishen, R. E., Solid-state studies of some retinal iminium salts and related-compounds - evidence for a 6-s-trans conformation. *J. Am. Chem. Soc.* **1987,** *109* (18), 5362-5366.

32. (a) Shimamura, T.; Hiraki, K.; Takahashi, N.; Hori, T.; Ago, H.; Masuda, K.; Takio, K.; Ishiguro, M.; Miyano, M., Crystal structure of squid rhodopsin with intracellularly extended cytoplasmic region. *J. Biol. Chem.* **2008,** *283* (26), 17753-17756; (b) Murakami, M.; Kouyama, T., Crystal structure of squid rhodopsin. *Nature* **2008,** *453* (7193), 363-U33.

33. Ota, T.; Furutani, Y.; Terakita, A.; Shichida, Y.; Kandori, H., Structural changes in the Schiff base region of squid rhodopsin upon photoisomerization studied by low-temperature FTIR spectroscopy. *Biochemistry* **2006,** *45* (9), 2845-2851.

34. Scheiner, S.; Duan, X. F., EFfect of intermolecular orientation upon proton-transfer within a polarizable medium. *Biophys. J.* **1991,** *60* (4), 874-883.

35. Gill, S. C.; Vonhippel, P. H., Calculation of protein extinction coefficients from amino-acid sequence data. *Anal. Biochem.* **1989,** *182* (2), 319-326.

36. Vasileiou, C., PhD. Dissertation, Michigan State University, East Lansing, 2005.

Chapter III. Mechanistic studies of wavelength tuning and pK_a regulation in CRBPII

III.1 General strategies for inducing red shift in rhodopsin mimic system

The retinal protonated Schiff base (Rt-PSB) absorbs in a wide range, from 420 nm to 570 nm, when bound in different opsins.[1] It is the most radically regulated protein-chromophore system in nature and has attracted a lot of interest.

Green (530 nm) and red (560 nm) rhodopsins red shift a lot compared to the absorption of retinal-PSB in aqueous solution (440 nm). The red shift observed in rhodopsins is termed the opsin shift.[2] Different hypotheses have been proposed to account for the opsin shift; but a clear picture is still not available that describes the cause of the opsin shift. This is partly because the crystal structures of color rhodopsins are not available, as rhodopsins are difficult to express, purify and crystallize, since they are membrane bound proteins. In addition, rhodopsin mutants for the purpose of spectral tuning have not been crystallized to study the cause-effect relationship of the mutagenesis on wavelength regulation. Although the crystal structures of different microbial rhodopsins are available, it is still not clear what leads to the 60 nm absorption

difference between sensory rhodopsin II (500 nm) and bacteriorhodopsin (560 nm).[3]

In addition, rhodopsins are naturally evolved to fulfill their function through isomerization of the chromophore.[4] They are all sensitive to light and isomerization takes place femto-seconds after light irradiation. Therefore, careful handling of these rhodopsins and complete shielding from light is necessary to obtain the dark state absorptions of the chromophore.

With CRBPII as a rhodopsin mimic, the drawbacks mentioned above for rhodopsins can be overcome. We have engineered CRBPII into a rhodopsin mimic that can bind all-*trans*-retinal as a protonated Schiff base with a pK_a of 8.3, which allows the PSB absorption to be clearly visible at physiological pH. Mutagenesis studies can be carried out to study the relationship between the protein environment and the absorption profiles of the chromophore. CRBPII mutants can be easily crystallized, providing a good platform to test different hypothesis for wavelength regulation.

The rationale for our wavelength regulation studies is based on two electronic characteristics of retinal-PSBs. One is that the positive charge on the iminium is more delocalized the ionone ring upon excitation from the ground state to the excited state.[5] If this is true, then a pre-organized electronic environment can have significant impact on the energy profile of the chromophore in the ground state and the excited state, via electrostatic perturbations. Increasing

negative polarity in the ionone ring region, along with reducing negative polarity in the PSB region can better stabilize the excited state than the ground state of the retinal-PSB, resulting in a decreased energy gap and red shift.

Besides the electric field projected by permanent dipoles, it has been suggested that introduction of polarizable residues such as Cys and Trp can also cause red shift.[6] However, this hypothesis has never been demonstrated in a protein system. With our robust rhodopsin mimic, we could test this hypothesis.

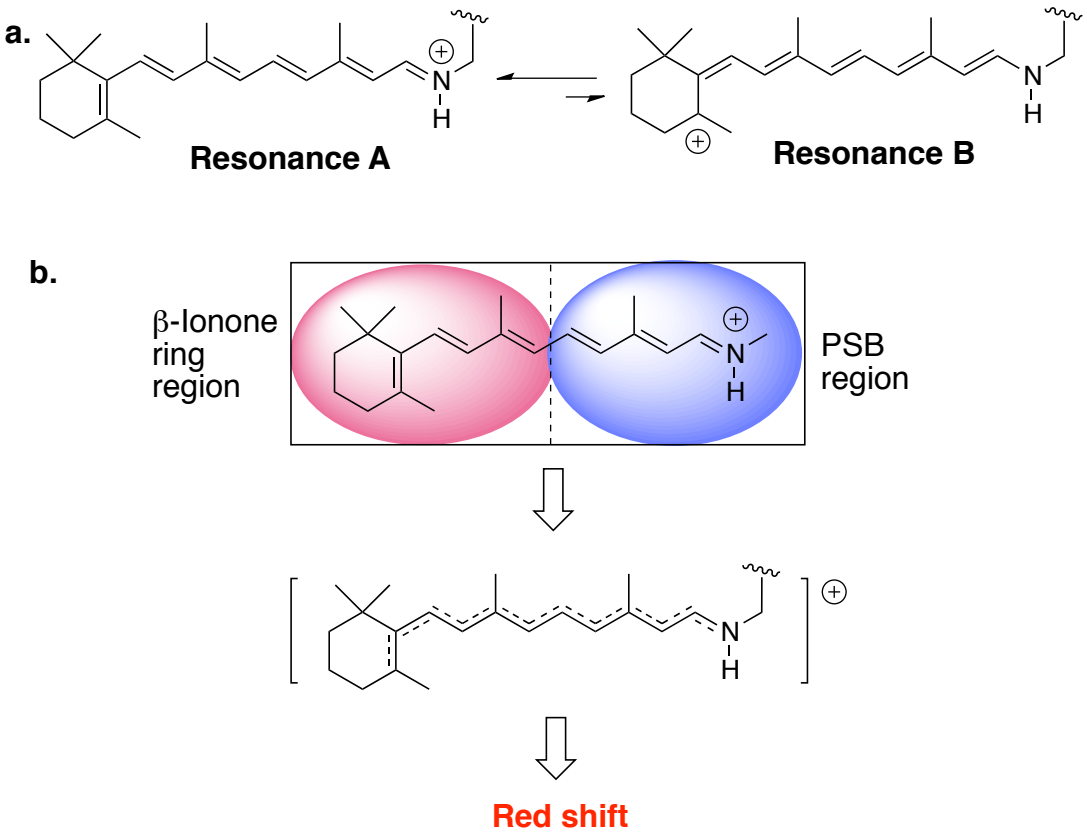


Figure III-1: General strategies for causing red shift. **a.** Resonance structures of retinal-PSB, showing the different localization of positive charge either on the PSB nitrogen or ionone ring. **b.** A simple recipe for causing red shift through electrostatic interactions with the protein by decreasing the negative polarity in the PSB region and increasing negative polarity in the β-ionone ring region. Blue color: positive polarity. Red color: negative polarity

The second hypothesis is that increased positive charge delocalization can lead to red shift (**Figure III-1a**). In order to stabilize the resonance structures that have delocalized positive charge, one can increase the negative polarity in the ionone ring region and decrease negative polarity in the PSB region (**Figure III-1b**), forming favorable electrostatic interactions with the delocalized positive charge.

Therefore, mutagenesis studies were carried out to decrease the negative dipoles in the PSB region and increase the negative polarity in the ionone ring region to test whether a red shift could be induced. Electrostatic calculations were carried out, based on the crystal structures of CRBPII mutants, to show the overall electrostatic potential projected on the Van der Waals surface of the chromophore from the protein environment, as a readout of the effect of electrostatic interactions on the absorption profiles of the chromophore.

III.2 Red-shift is induced by decreasing the negative polarity near the protonated Schiff base region

III.2.1 Mutations of T51

To induce red shift, negative dipoles in the PSB region are removed, to encourage the positive charge on the iminium to travel along the polyene towards the ionone ring. At the same time, the positive charge on the iminium can not be destabilized too much, as that might decrease the pK_a of the PSB to a point, where the pK_a drops below that of physiological pH.

Negatively polar residues can be removed and replaced with hydrophobic residues or even positively charged residues to decrease the negative polarity projected on the Schiff base region. CRBPII has a hydrophobic binding pocket, with only a few polar residues in the Schiff base region, including T51, Q4 and W106 that are within 5 Å from the iminium (**Figure III-2**). Mutagenesis studies on these three positions were carried out.

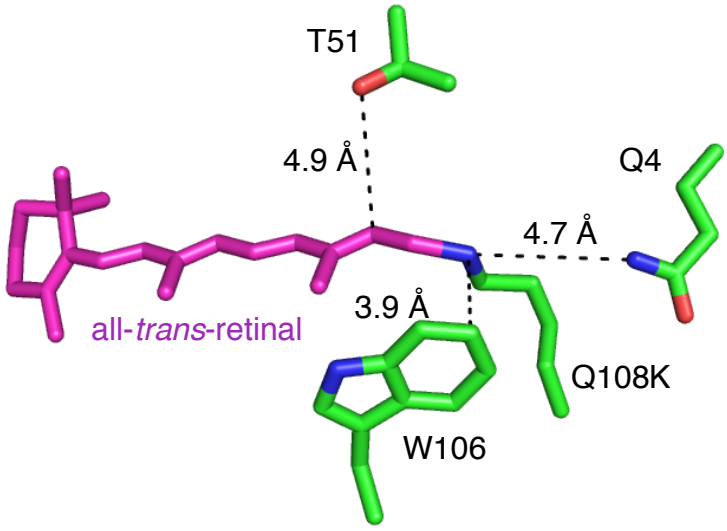


Figure III-2: Crystal structure of Q108K:K40L bound with all-*trans*-retinal. T51, Q4 and W106 are highlighted.

A hydrophobic residue was used to replace the negative polarity from the hydroxyl containing residue T51. In order to maintain the same volume of threonine, T51V was introduced. In this way, only electronic factors would be considered for the difference in absorption caused by mutant T51V, assuming little significant conformational change. To our great satisfaction Q108K:K40L: T51V-CRBPII triple mutant led to a 25 nm red shift, although the closest distance from Thr51 to the polyene is 4.9 Å and there is no hydrogen bonding interaction

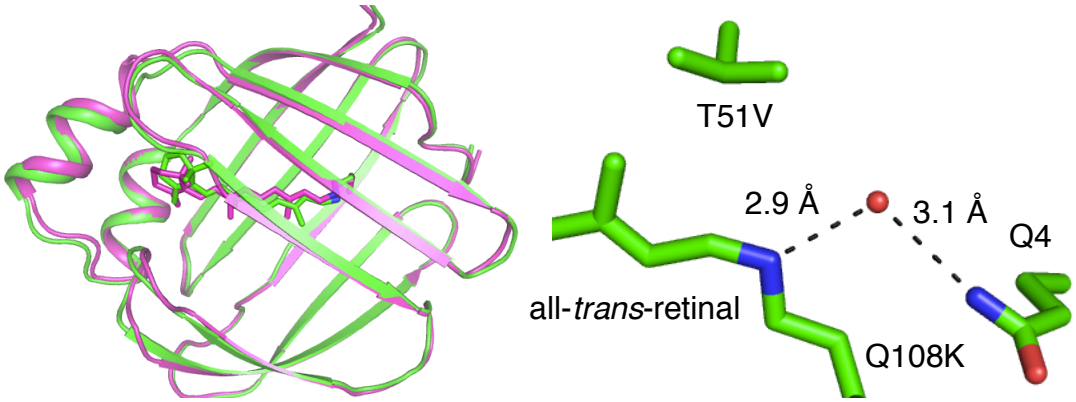


Figure III-3: The overlay of the crystal structures of Q108K:K40L:T51V (green) and Q108K:K40L (magenta) and the binding site of Q108K:K40L:T51V.

between the PSB and Thr51. The red shift agrees with the hypothesis that decreasing the negative polarity in the Schiff base region destabilizes the positive charge localized on the iminium, leading to a larger degree of positive charge delocalization along the polyene and results in a red shift.

Removal of the counteranion in bovine rhodopsin via mutation of E113A led to a 28 nm red shift,[7] close to the red shift caused by T51V in CRBPII. Considering a negatively charged glutamate is more polar than threonine and it has direct interactions with the iminium, removal of the counteranion E113 in bovine rhodopsin should have a more drastic effect in red shifting the absorption maximum than removal of threonine in CRBPII. The similar degree of red shift obtained might suggest that the CRBPII binding pocket is more sensitive to polarity change.

The crystal structure of Q108K:K40L:T51V was determined (Rafida Nossoni, Jim Geiger's Lab, MSU). As shown in **Figure III-3**, The imine adopts a

cis conformation and hydrogen bonds to a water molecule, which is bound to Gln4, like that in Q108K:K40L. The T51V mutation does not disturb this interaction. The overlay of the crystal structures of Q108K:K40L and Q108K:K40L:T51V shows that no significant structural change results from the T51V mutation. This indicates that the red shift is simply the result of decreasing the negative polarity in the PSB region.

Interestingly, removal of negatively polar residue T51 did not result in a decrease of the pK_a. This indicates that T51 does not directly contribute to the stability of the retinal-PSB, as evident from the crystal structure of Q108K:K40L, where T51 does not have any direct or indirect hydrogen bonding interactions with the PSB. The red shift caused by the T51V mutation is simply a result of the polarity change.

Another polar residue, Asn, was used to replace T51 to probe its effect on wavelength tuning. Interestingly, Q108K:K40L:T51N leads to a slight blue shift to 496 nm (compared to 508 nm with Q108K:K40L). Asparagine is less polar, but

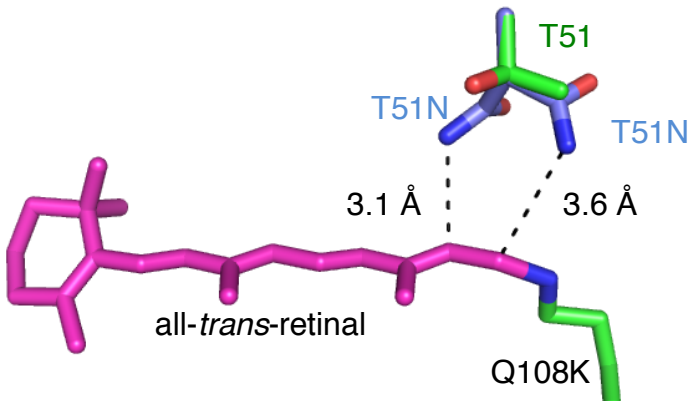


Figure III-4: Model structure of Q108K:K40L:T51N overlaid with crystal structure of Q108K:K40L:T51, with two possible rotamers of T51N highlighted.

one carbon shorter than threonine. This could position the amide group closer to the PSB than threonine as shown in **Figure III-4**. Distance is an important factor in the strength of an electric field.

$$E = \frac{Q}{4\pi\varepsilon * r^2}$$

(In the equation above, Q is the charge for each point of the residue, ε is the dielectric constant, r is the distance from the particle with charge Q to the E-field evaluation point.)

As indicated in the equation above, the electric field is inversely proportional to r^2. A small decrease in distance could result in a big increase of the electric field. Two possible conformers of asparagine (**Figure III-4**) could both move the amide group closer to the polyene as compared to threonine. Therefore, although asparagine might not be as polar as threonine, it could still project a larger amount of negative electrostatic potential on the PSB region of the chromophore.

If this is true, mutant Q108K:K40L:T51D should result in an even larger blue shift, as T51D is more polar than asparagine when deprotonated. Unfortunately, Q108K:K40L:T51D was only found in inclusion bodies during protein purification and protein refolding was not successful for this mutant.

The fact that Q108K:T51D can be expressed as a soluble protein and bind all-*trans*-retinal as a PSB, but not Q108K:K40L:T51D, indicates that K40 plays an important role in stabilizing T51D in the hydrophobic environment possibly

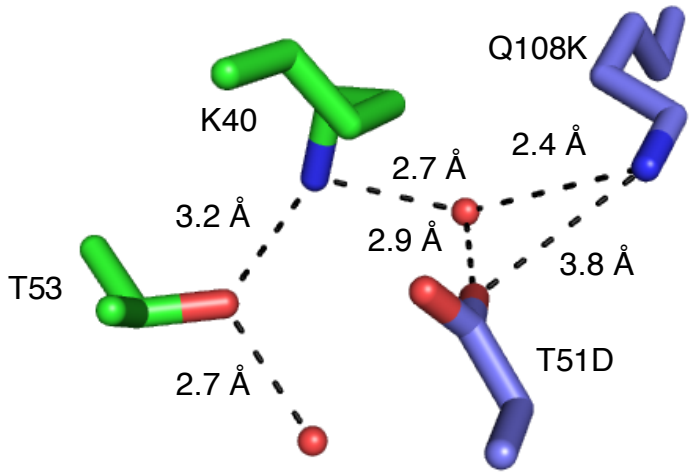


Figure III-5: Model structure of Q108K:T51D. Two water molecules are present in WT-CRBPII. Mutagenesis of Q108K and T51D were modeled in Pymol based on the crystal structure of WT CRBPII.

through hydrogen bonding interactions. As shown in **Figure III-5**, the conformation for Q108K and T51D were modeled in such a way that enables them to enjoy hydrogen bonding. Replacement of K40 with leucine disrupts this putative structured hydrogen bonding interaction and affects the correct folding of the protein.

Nonetheless, mutant Q108K:K40L:T51D with two extra mutations R58W: Y19W could be expressed as a soluble, functional protein. Both mutations R58W and Y19W are far away from K40L and T51D, therefore their direct interactions do not contribute to the stability of the protein. Tryptophans have been reported to be important in protein folding, due to better hydrophobic packing, π-π stacking and π-cation interactions.[8] Introduction of these two tryptophan residues could have greatly stabilized the protein, overcoming the destabilizing effect of placement of Asp51 in the hydrophobic cavity with K40L .

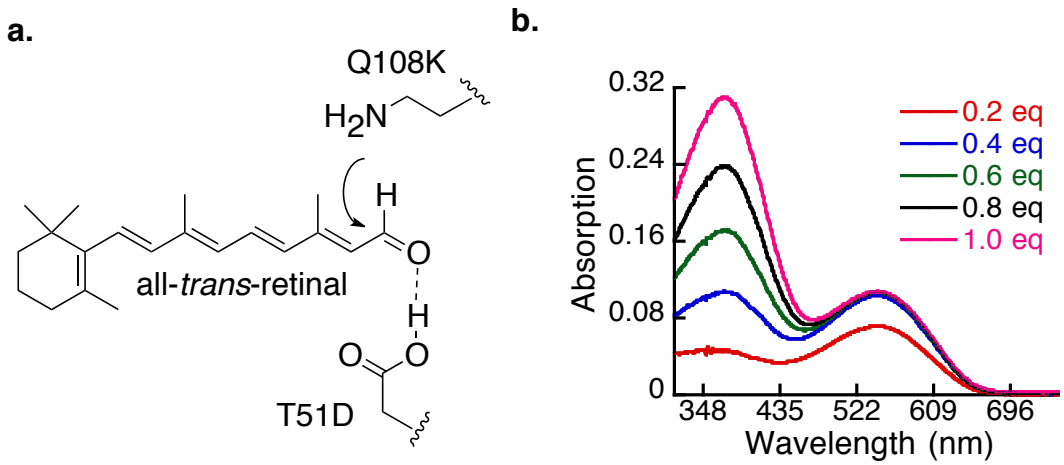


Figure III-6: Characterization of Q108K:K40L:R59W:Y19W:T51D. **a.** Proposed mechanism for activation of retinaldehyde by introduction of T51D in penta-mutant Q108K:K40L:R59W:Y19W:T51D. **b.** UV-vis spectra of retinal titration of penta-mutant Q108K:K40L:R59W:Y19W:T51D.

Q108K:K40L:R58W:Y19W:T51D results in a 7 nm red shift to 545 nm as compared to 538 nm with Q108K:K40L:R58W:Y19W. Introduction of negative polarity in the PSB region could stabilize the positive charge on the iminium and decrease the degree of positive charge delocalization, leading to a blue shift. The red shift obtained in mutant Q108K:K40L:R58W:Y19W:T51D is unexpected.

The Q108K:K40L:R58W:Y19W:T51D crystal structure could not be solved, to help explain this counter-intuitive result. Is the conformation of T51D different, rotating away from the PSB? The pK_a of this pentamutant is estimated to be 7.7, which is 1.4 pK_a units lower than the pK_a of Q108K:K40L:R58W:Y19W (9.1). This decrease in pK_a as a result of the T51D mutation indicates that T51D might be rotating away from the PSB or maybe there is an overall conformational change of the protein that leads to this red shift and decrease in pK_a.

Different from Q108K:T51D, which has slow PSB formation kinetics probably due to the salt bridge formed between T51D and Q108K, Q108K:K40L:R58W:R19W:T51D has fast PSB formation kinetics. The PSB is formed within 1 min after retinal is added. This fast kinetics suggests that T51D in the case of the the pentamutant does not form a salt bridge with Q108K, but instead might form a hydrogen bond with the retinaldehyde carbonyl group to activate the carbonyl group for electrophilic attack (**Figure III-6a**).

The potentially different roles of T51D in double mutant Q108K:T51D and penta-mutant Q108K:K40L:R58W:R19W:T51D is not conclusive, as both of these two mutants were not crystallized. More effort should be exerted to understand the different function of T51D.

The UV-vis retinal titration spectra of Q108K:K40L:R58W:R19W:T51D indicates that only a small portion of the penta-mutant is active, while the majority of the protein does not bind retinal as a PSB. As shown in **Figure III-6b,** incubation with 0.2 equiv of retinal generates an absorption peak at 545 nm attributed to the PSB. At 0.4 equiv, there is still a slight increase in the absorption for the PSB, but the peak at ~380 nm starts increasing, which signifies unbound retinal. After that, no increase in the PSB absorption is observed. This indicates that the active protein accounts for only 20% to 40% of the population. This phenomenon is not common for CRBPII mutants; it is observed only in a few CRBPII mutants when acidic residues are introduced. Once again this shows

that introduction of acidic residues in the hydrophobic cavity disturbs the stability of the protein.

In summary, we have shown that simply by changing the polarity of the amino acid at position 51 in the PSB region, we can regulate the wavelength in an expected manner (**Figure III-7**). The residue that projects the most negative electric field potential on the PSB region of the retinal-PSB is the most blue shifted. As it is more hydrophobic in the PSB region of Q108K:K40L-CRBPII, the chromophore seems to be more sensitive to polarity changes in the PSB region than the color rhodopsins.

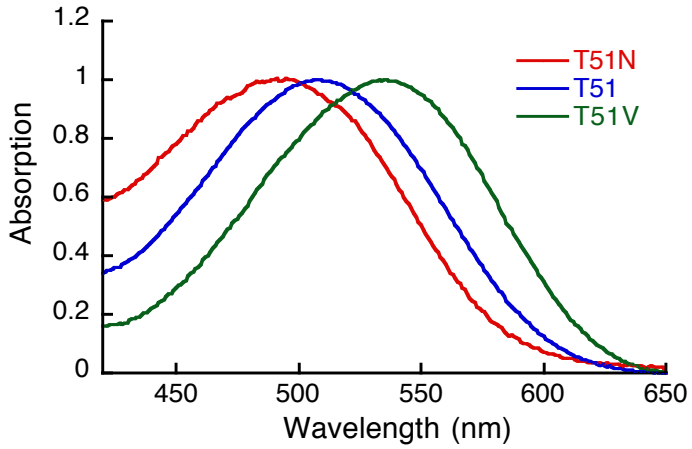


Figure III-7: UV-vis spectra of different T51 mutants (Q108K:K40L:T51N, Q108K:K40L:T51 and Q108K:K40L:T51V).

At the same time, the same residue, T51D, could have different functions depending on its context and interaction partners. Crystal structures are indeed critical in this case to provide the conformational changes that lead to these differences.

III.2.2 Red shift is induced by mutation Q4W

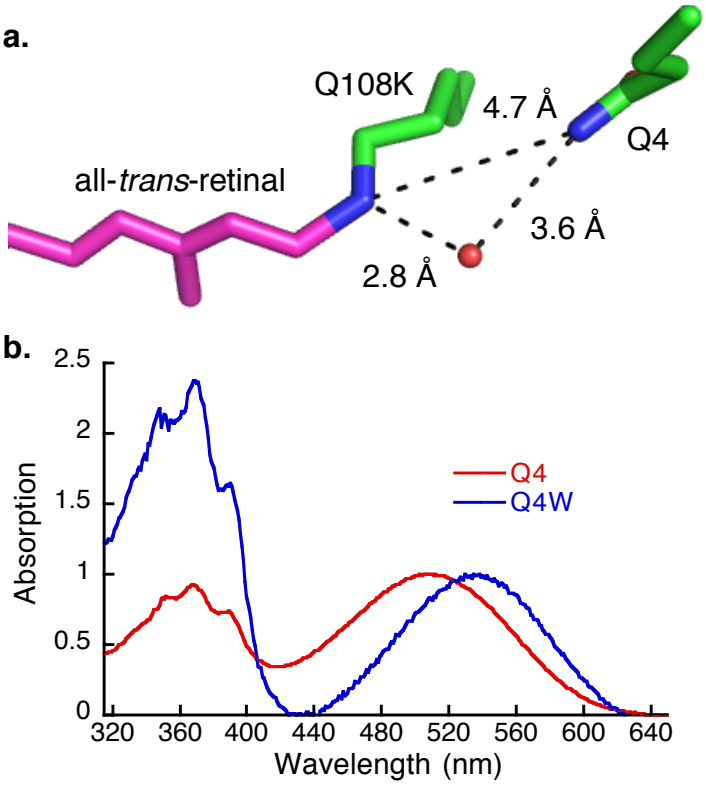


Figure III-8: Characterization of Q4W. **a**. Crystal structure of Q108K:K40L with Q4 highlighted. **b**. UV-vis spectra of Q108K:K40L:Q4 and Q108K:K40L:Q4W.

Encouraged by the large red shift caused by the T51V mutation, it seemed reasonable to further remove negative polarity in the PSB region to red shift the protein. The water-mediated hydrogen bonding interactions between Gln4 and the retinal-PSB revealed in CRBPII mutant structures, it seemed quite straightforward to alter position 4, in order to destabilize the positive charge on iminium for red shift. However, the Q108K:K40L:Q4W mutant would not crystallize. Based on the WT-CRBPII crystal structure, Gln4 seemed to be more than 5 Å away from the putative position of the iminium. In addition, the amide is

not as polar as hydroxyl containing amino acids or acidic amino acids. Therefore, we did not expect a large red shift for mutations of Gln4.

Mutagenesis at position 4 was carried out, since it is the only other obvious polar residue in the PSB region. The Q108K:K40L:Q4W-triple mutant gives a 25 nm red shift to 533 nm (**Figure III-8b**), which frankly was unexpected. However, once the crystal structure of the holo-CRBPII mutant was resolved, it revealed a water-mediated interaction between Gln4 and the iminium (**Figure III-8a**). Considering this the large red shift observed with mutations of Gln4 makes better sense.

As a water-mediated hydrogen bonding interaction can stabilize the protonated state of the PSB, removal of this interaction decreased the pK_a value of Q108K:K40L:Q4W (**Figure III-9**) by almost 2 units to 6.3 compared to Q108K:K40L (8.3). The detailed analysis of Gln4 will be discussed later with more systematic mutagenesis studies of the position, based on the hepta-mutant Q108K:K40L:T51V:T53C:R58W:T29L:Y19W and the use of crystal structures.

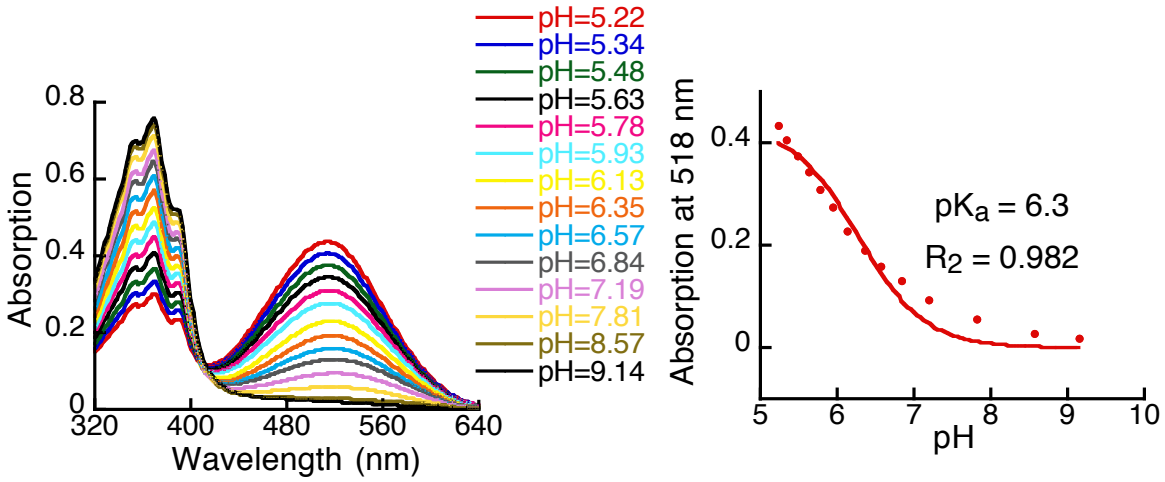


Figure III-9: UV-vis Base titration of Q108K:K40L:Q4W.

III.3 Mutagenesis studies in the middle of the polyene

III.3.1 A red shift results from the T53C mutation

As shown in **Figure III-10**, T53 is situated in the middle of the polyene, interacting with T51 via a water-mediated hydrogen bonding network. Previously, it has been reported that a polarizable environment could lead to a red shift of the retinal-PSB, probably by stabilizing the excited state of the retinal-PSB.[9, 10, 11, 12] This is because the positive charge of the retinal-PSB is transferred from the iminium end toward the ionone ring during electronic excitation.[5] Induced dipoles could be generated within the time frame of electronic excitation (10^{-15} sec) to interact favorably with the positive charge in the ionone ring region in the excited state and lower the energy gap.[12]

The effect of a polarizable environment in wavelength regulation has been tested with model compounds in polarizable solvents.[9, 10, 11, 12] Computational studies on halorhodopsin also shows that a polarizable environment is important in inducing a red shift.[13] However, the effect of polarizable residues has never been tested in a protein-chromophore system experimentally. This is because no systematic mutagenesis studies with the support of crystal structures have been done to dissect the contribution of the polarizable environment.

Engineered CRBPII mutants provide an optimal platform to study cause-effect relationships of different factors in spectral tuning through mutagenesis studies, with easily crystallized protein. Therefore, polarizable amino acids such as tryptophan and cysteine were introduced in CRBPII to probe its effect on wavelength tuning.

To see the effect of polarizability in spectral tuning, a cysteine was introduced at position 53. This is a structurally conservative mutation, as opposed to using tryptophan, which would be too bulky for this position. UV-vis spectrum of Q108K:K40L:T53C shows that T53C induces a small red shift to 513 nm compared with Q108K:K40L (508 nm).

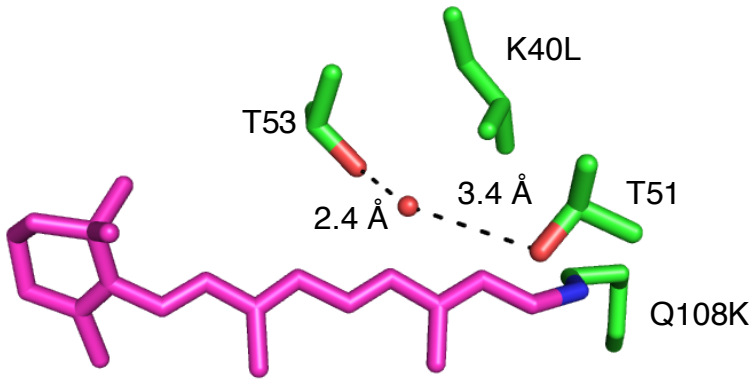


Figure III-10: Crystal structure of Q108K:K40L-retinal, with T53 and its hydrogen bonding interactions highlighted.

To show that the red shift caused by T53C is not simply an effect of removal of the negative polarity from the hydroxyl group of T53 or disruption of the water mediated hydrogen bonding network (**Figure III-10**), a control study with mutant Q108K:K40L:T53V was carried out. Satisfyingly, Q108K:K40L:T53V does not lead to a red shift as seen with T53C, instead it results in a slight blue

shift when T53V is introduced to Q108K:K40L. T53V has been introduced in other mutants of CRBPII as well, it does not lead to red shift in any cases that is comparable to T53C (**Table III-1**).

Table III-1: Comparison of mutations at Thr53 position

CRBPII mutant	λ_{max} (nm)	pK_a
Q108K:K40L:**T53**	508	8.3
Q108K:K40L:**T53C**	513	7.3
Q108K:K40L:**T53V**	503	8.3
Q108K:K40L:R58W:Y19W:T51V:**T53**	577	9.2
Q108K:K40L:R58W:Y19W:T51V:**T53C**	591	8.4
Q108K:K40L:R58W:Y19W:T51V:**T53V**	577	9.6
Q108K:K40L:R58W:Y19W:T51V:Q4F:**T53**	597	n.d.
Q108K:K40L:T29L:R58W:Y19W:Q4W:T51V:**T53C**	613	7.7
Q108K:K40L:T29L:R58W:Y19W:Q4W:T51V:**T53V**	600	8.1

This shows that the red shift caused by T53C is not only a result of polarity change, but due to some other factors, likely the introduction of polarizability from T53C. T53V does not cause a dramatic blue shift or red shift, which indicates that Thr53 has almost equal effects in stabilizing the ground state and excited state of the chromophore. Mutation of T53 might disrupt the hydrogen bonding interactions with T51, as shown in **Figure III-10**. In the case of Q108K:K40L:T53V, removal of T53 might push the water molecule to form a tighter hydrogen bond with T51, leading to a 5 nm blue shift, because the water

molecules moves closer to the PSB, increasing the negative polarity in the PSB slightly. However, T53V causes either no change or only a 3 nm red shift in absorption, as the water molecule mediating T51 and T53 is not stabilized anymore. T53C decreases the pK_a by 1 unit, while T53V increases the pK_a by less than half a unit.

Furthermore, to show that the red shift caused by T53C is not a result of the reduction of the size of Thr53, Q108K:K40L:T53S was made. As expected, no obvious change in absorption maximum is observed by reducing the size of Thr53 to serine. As shown in **Table III-2**, T53C leads to a larger red shift as compared to T53S when introduced in either Q108K:K40L or Q108K:K40L:R58F-mutants. This further proves the unique role of T53C in interacting with the chromophore, perhaps because of the increased polarizability of Cys53, which can interact favorably with the excited state retinal-PSB.

Table III-2: Table of different T53C and T53S mutations

CRBPII mutant	λ_{max} (nm)	pK_a
Q108K:K40L:**T53**	508	8.3
Q108K:K40L:**T53S**	509	8.2
Q108K:K40L:**T53C**	513	7.3
Q108K:K40L:R58F:**T53**	523	8.6
Q108K:K40L:R58F:**T53S**	528	8.6
Q108K:K40L:R58F:**T53C**	537	8.5

Q108K:K40L:T53N was also prepared. It leads to an 8 nm blue shift to 500 nm, indicating that with a longer carbon chain, asparagine could possibly extend to a region closer to the PSB. That will stabilize the positive charge in the PSB region and decrease the positive charge delocalization, leading to a blue shift.

Table III-3: Summary of different Thr53 mutations on Q108K:K40L.

CRBPII mutant	λ_{max} (nm)	pK_a
Q108K:K40L	508	8.3
Q108K:K40L:T53C	513	7.3
Q108K:K40L:T53V	503	8.3
Q108K:K40L:T53S	509	8.2
Q108K:K40L:T53N	500	7.4

Therefore in summary, only T53C leads to a consistent red shift (**Table III-3**). For the first time, the unique role of cysteine in stabilizing the excited state of the chromophore is shown. The crystal structure of Q108K:K40L:T53C bound with retinal was determined to support this idea (Rafida Nossoni, Jim Geiger's lab, MSU). An overlay of the crystal structures of Q108K:K40L and Q108K:K40L:T53C shows that T53C adopts the same rotamer as Thr53 in the crystal structure of Q108K:K40L bound with retinal. The water molecule coordinating to both Thr51 and Thr53 is maintained when T53C is introduced. The clear density of the T53C in the crystal structure of Q108K:K40L:T53C (**Figure III-11b**) rules out the possibility of rotation of T53C away from the PSB,

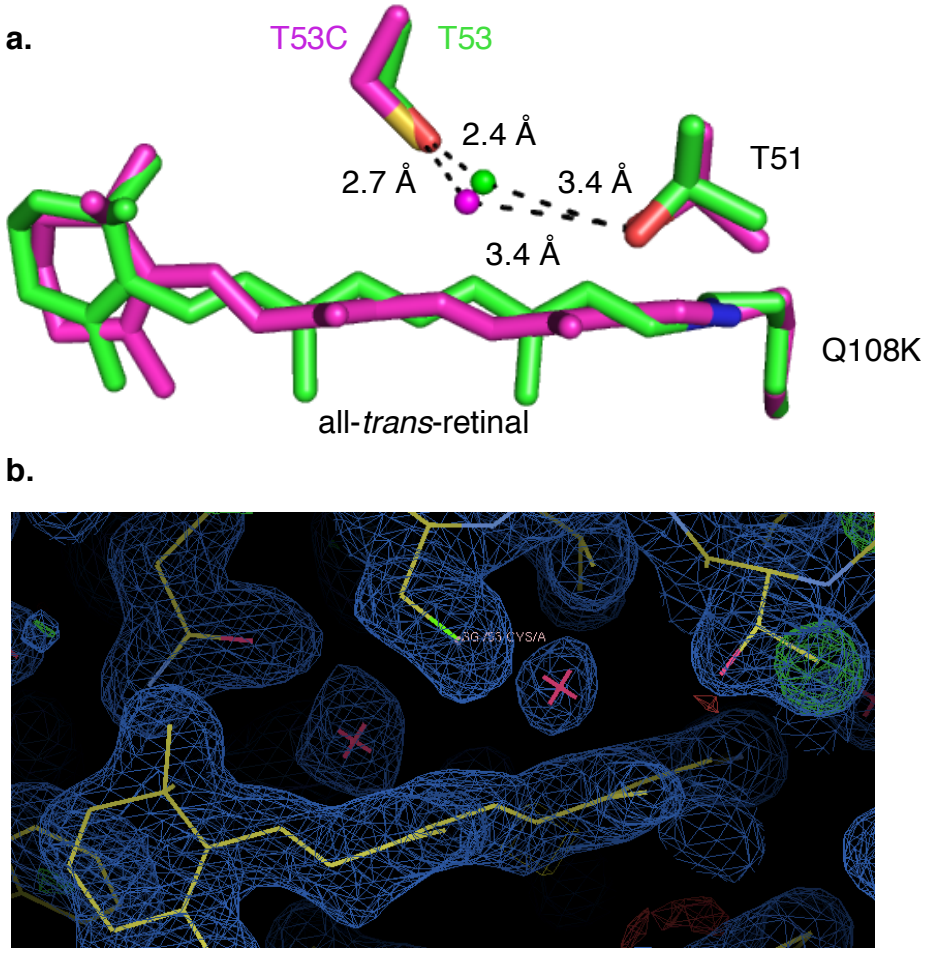


Figure III-11: Crystal structure of Q108K:K40L:T53C. **a**. Overlaid structures of Q108K:K40L and Q108K:K40L:T53C. **b**. Electron density of crystal structure of Q108K:K40L:T53C, showing only one rotamer for T53C.

at least not in the case of Q108K:K40L:T53C. This suggests that in the case of Q108K:K40L:T53C, the red shift is because of the polarizability of cysteine.

However, it could also be argued that the hydrogen bonding between T53C and the water molecule is weaker from the slightly longer distance between T53C and the water molecule, since oxygen is a better hydrogen bonding acceptor than sulfur. Because of this slight weakening of hydrogen bonding interaction, T53C could rotate away towards the ionone ring. If the sulfur group of T53C points toward the ionone ring, it would be able to stabilize the conjugated

positive charge further along the polyene and thus lead to more red shift. This rotamer of T53C, which points towards the ionone ring, is observed in crystal structures of other CRBPII mutants when T51V is present, which disrupts the water-mediated hydrogen bonding interactions with T53C. However, the density in the crystal structure of Q108K:K40L:T53C clearly shows one conformer for T53C, with the sulfhydryl group pointing towards the PSB as seen in **Figure III-11b**.

As shown in **Figure 11a**, the polyene of the retinal in the crystal structure of Q108K:K40L:T53C rotates ~90° from that of Q108K:K40L. The chromophore looks highly twisted at the C6-C7 single bond, leading to a lesser degree of conjugation and blue shift. However, T53C causes a minor red shift.

It is not clear why the chromophore adopts a highly twisted conformation in Q108K:K40L:T53C but still results in a red shift. It is possible that the polarizability of T53C actually leads to a red shift larger than 5 nm, but compromised by the blue shift resulting from the twisted chromophore.

III.3.2 The red shift induced by mutation Y19W

It is exciting to see that placing a polarizable residue, T53C, in the middle of the polyene could induce a red shift, through interaction of the induced dipole with the excited state of the chromophore. Similar strategies were carried out to introduce more polarizable residues in the middle of the polyene. One approach is to introduce tryptophan residues in positions near the middle of the polyene. In

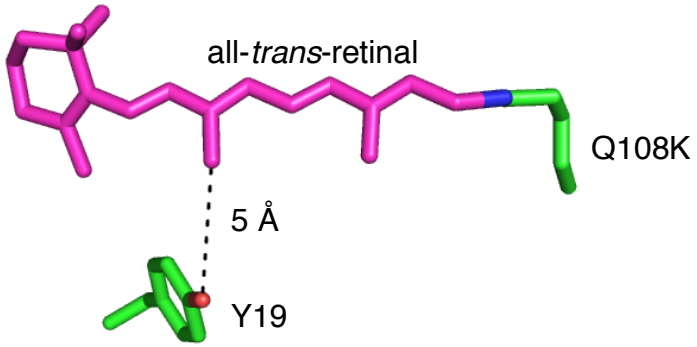


Figure III-12: Crystal structure of Q108K:K40L showing the position of Tyr19.

order not to introduce too great of a steric clash in the binding pocket by introduction of tryptophan residues, targeted positions should have residues such as tyrosine or phenylalanine, which are structurally comparable to tryptophan.

The first suitable candidate was Tyr19. Tyr19 is 5 Å away from the C9-methyl group and is placed in a position roughly in the middle, but closer toward the ionone ring (**Figure III-12**). Since it is far away from the retinal and tyrosine is a large aromatic residue, mutations to smaller residues will make the distance to the chromophore even larger and create an empty space in the binding pocket, which might disturb the protein folding. Therefore tryptophan became the top choice to replace Tyr19.

As **Table III-4** shows introduction of Y19W to the Q108K:K40L double mutant red shifts 5 nm to 513 nm. Similarly when Y19W is introduced to the Q108K:K40L:R58W triple mutant, a 19 nm red shift is observed. For comparison, Y19F results in only 5 nm red shift when introduced to Q108K:K40L:R58W. This indicates that removal of the polarity from Y19 can

contribute slightly to the red shift, however, the majority of the red shift is induced by placing the polarizable tryptophan at position 19.

Table III-4: Mutants of Y19, based on Q108K:K40L and Q108K:K40L:R58W.

CRBPII mutant	λ_{max} (nm)	pK_a
Q108K:K40L:**Y19**	508	8.3
Q108K:K40L:**Y19W**	513	8.9
Q108K:K40L:R58W:**Y19**	519	8.7
Q108K:K40L:R58W:**Y19W**	538	9.1
Q108K:K40L:R58W:**Y19F**	524	9.2

We originally thought that the red shift is simply a result of the polarizability of Y19W, but the crystal structures of CRBPII mutants containing Y19W forced a double interpretation. The crystal structures of CRBPII mutants containing Y19 and Y19W (**Figure III-13,** refined by Rafida Nossoni, Jim Geiger's lab, MSU) show that the two residues adopt the same rotamer and the overall structures are similar. However, the ionone ring of the chromophore is translated by about 1.4 Å when Y19W is introduced as a result of steric clash (2.3 Å) between Y19W and the 5-methyl group on the ionone ring. This movement will position the ionone ring of the chromophore in a different electronic environment. After the translational movement, the ionone ring is closer to the more electron rich side of the binding pocket, where Q38 and Q128 are situated, with two water molecules tightly bound through hydrogen bonding interactions, as shown in **Figure III-13**.

The idea of an electronic interaction between Y19W and the delocalized positive charge of the chromophore was further rejected after a closer look at the crystal structures. The reason is that with the conformation of Y19W almost perpendicular to the polyene, no optimal electronic interactions could be expected. A parallel π p stacking might be expected to contribute more to cause red shift, but is not possible for any rotamer of Y19W. The effect of Y19W will be further explored later along with other mutations.

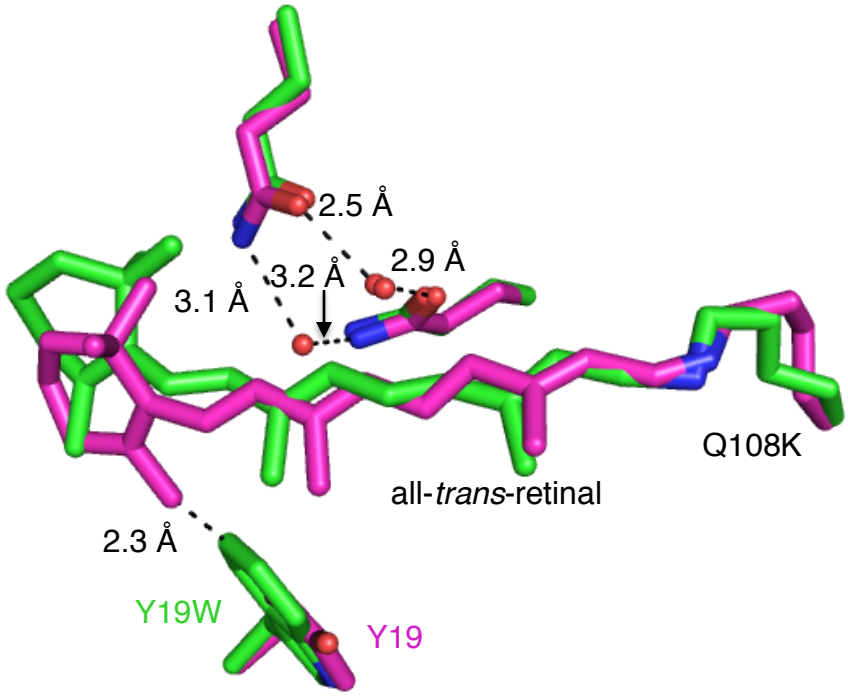


Figure III-13: The crystal structures of Q108K:K40L:T51V:R58F (magenta) and Q108K:K40L:T51V:R58W:Y19W (green) are overlaid.

III.3.3 Red shift and blue shift can result from Y60W mutation due to different protein conformations

Similar to Y19, residue Y60 is also situated in the middle of the polyene, around 4 Å away from the polyene, however with the hydroxyl group pointing away from the chromophore (**Figure III-14**). A tryptophan residue was introduced at this position with the intention of increasing the electronic interaction between residue 60 and the delocalized positive charge of the polyene system. Modeling suggested Y60W could possibly adopt a conformation that would position the tryptophan to be parallel to the polyene, leading to a favorable interaction to longer absorption wavelenth, although it would still be more than 4 Å away.

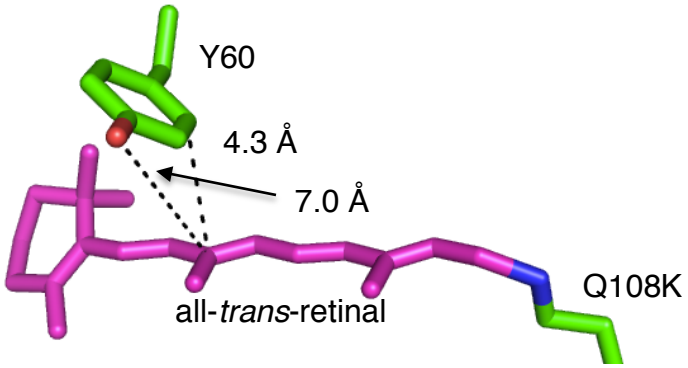


Figure III-14: Crystal structure of Q108K:K40L bound with all-*trans*-retinal, with Tyr60 highlighted.

To this end the triple mutant Q108K:K40L:Y60W was generated. When expression was induced with IPTG at 32 °C, the UV-vis spectra of Q108K:K40L:Y60W incubated with retinal shows a 5 nm red shift in absorption maximum to 513 nm. However, when the induction temperature was lowered to

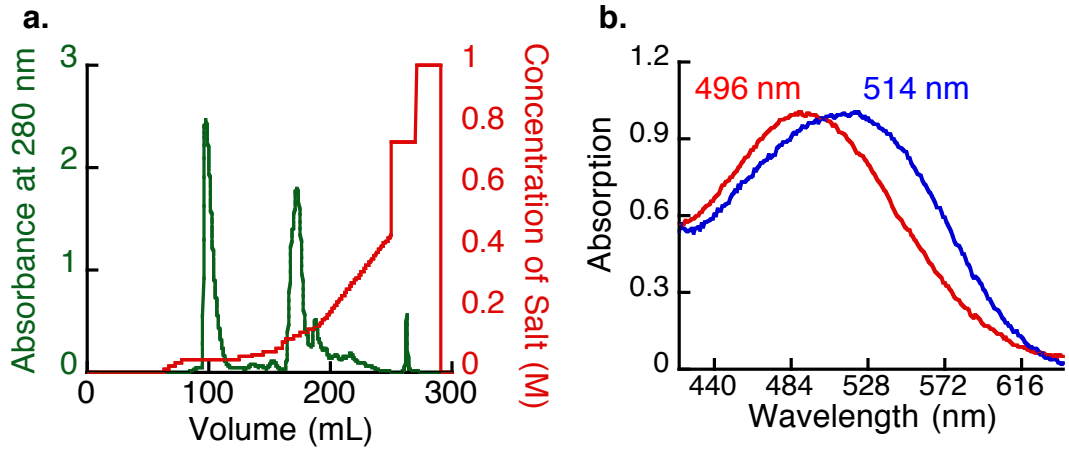


Figure III-15: FPLC trace of Q108K:K40L:Y60W and UV-vis characterization **a.** Diagram of FPLC fraction collection of CRBPII mutant Q108K:K40L:Y60W, monitored at a UV absorption of 280 nm, attributed to absorptions of tryptophan residues from the protein. **b.** UV-vis spectral overlay of 40 mM salt elution (red) and 150 mM salt elution (blue) of Q108K:K40L:Y60W incubated with all-*trans*-retinal.

16 °C to improve the expression yield, incubation of the protein with retinal resulted in a blue shift to 496 nm.

This was confusing at the beginning. It was realized later that when the protein is expressed at 25 °C, two different populations of the protein are produced, exhibiting different pI values. The two fractions of protein are eluted at two different salt concentrations, 40 mM and 150 mM NaCl, at pH=8.0, from the source Q ion exchange column (**Figure III-15a**). Incubation of the protein eluted with 40 mM salt shows absorption at 496 nm, while the protein eluted with 150 mM salt shows absorption at 514 nm (**Figure III-15b**). Routinely, CRBPII mutants usually elute with 40 mM salt, indicating that the fraction collected at 150 mM is different from the normal eluent. A few experiments were carried out to understand what gives rise to the differences between the two fractions.

First, we had to make sure that the different absorption maxima of retinal-PSB formed with the two fractions of Q108K:K40L:Y60W is not a result of different salt concentrations. The concentrated proteins were diluted with 100 mM NaCl PBS solution to make the final concentration of NaCl to be ~100 mM for both fractions. No changes in absorption for the two fractions were observed. This indicated that the absorption of the retinal-PSB formed is salt concentration independent.

Second, to understand whether these two fractions are due to different aggregations of the protein, size-exclusion chromatography was used. It was already known that the normal elution fraction at 40 mM salt concentration is a monomer. The 40 mM and 150 mM elution fractions were subjected to size exclusion chromatography, and both fractions were identified as monomer with molecular weights of around 16 kDa. This excluded different aggregation states to be the reason for the two proteins eluting at the two different salt concentrations. It does indicate that the differences between the proteins are likely due to different conformational states.

Third, to check whether the two different elution fractions are interchangeable or not, the two elution fractions were subjected to the source Q anion exchange column again respectively. If these two fractions are interchangeable with each other, there should be an equilibrium state where the ratio of one versus the other is a constant. However, the 40 mM protein eluent still eluted at 40 mM salt concentration, and the 150 mM protein eluent still eluted

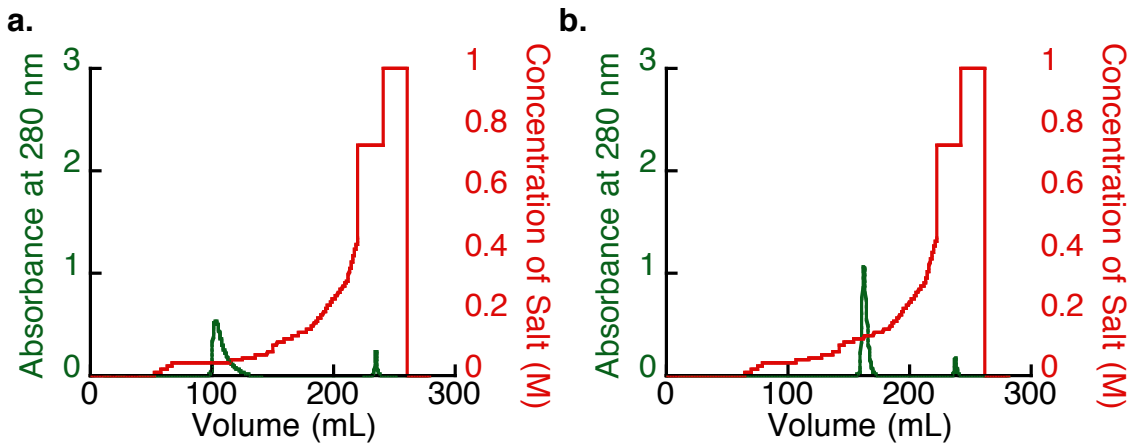


Figure III-16: Reinjection of 40 mM (a) and 150 mM (b) salt elution of Q108K:K40L:Y60W, monitored at UV absorption of 280 nm.

at 150 mM salt concentration (**Figure III-16**). From these results, it can be concluded that these two different elution fractions do not interchange with each other once the native protein is folded.

Fourth, since the two elution fractions are not interchangeable after the proteins are correctly folded, the question remains whether it is possible to interchange the two different elution fractions through denaturing of the protein first and then refolding of the protein. If this is possible, it will show that the two elution fractions are simply a result of different conformations of the proteins.

The two fractions of protein were first denatured with 8 M urea solution and then refolded by fast dilution into Tris Buffer (pH=9.0). Refolding was allowed to occur through incubation at 4 °C overnight. The refolded protein was concentrated and subjected to source Q ion exchange chromatogrphy.

It was found that after the protein eluted at 150 mM was denatured and refolded, it had converted into a mixture of protein that eluted at 40 mM and 150

mM salt concentration, with the 40 mM elution fraction being the major fraction (**Figure III-17**). The UV-vis of the protein eluted at 40 mM salt concentration of the refolded protein showed the same spectrum as the native 40 mM elution protein, indicating that the 40 mM elution fraction from the refolded 150 mM fraction protein behaves the same as the native 40 mM elution fraction.

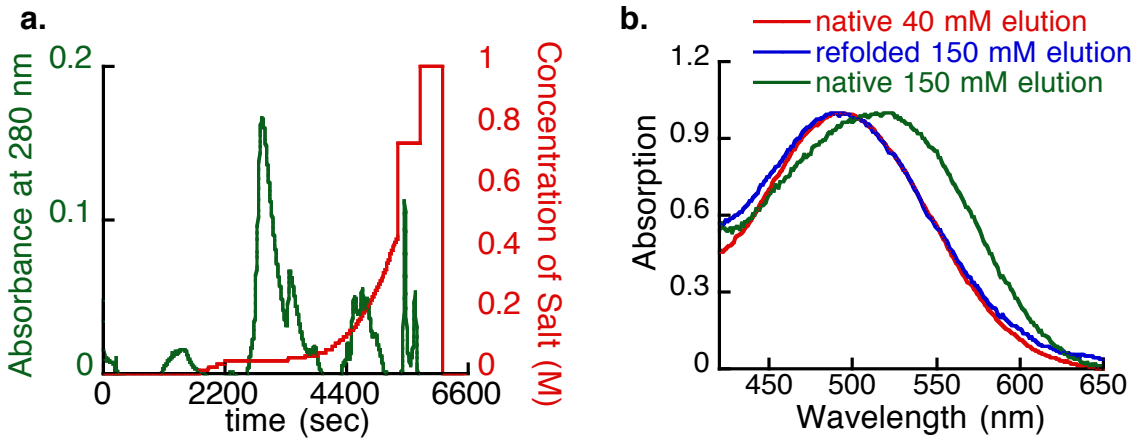


Figure III-17: Characterization of refolded 150 mM elution. **a.** Source Q chromatograph of refolded 150 mM elution of Q108K:K40L:Y60W, monitored at 280 nm. b. UV-vis comparison of native 40 mM elution and 40 mM elution of refolded 150 mM elution.

These experiments show that the two conformations of protein are formed during the protein folding step and are not interchangeable. The 40 mM elution fraction is probably the thermodynamically more stable one. That is why the denatured protein can be converted more favorably to the 40 mM elution fraction when it is allowed to be refolded slowly at 4 °C. This also explains that when the protein expression is induced at lower temperature, 16 °C, the 40 mM elution fraction is the major form, while at higher induction temperature, 32 °C, 150 mM elution fraction is the major form. At 16 °C, the protein is allowed to fold slowly

into a thermodynamically stable form. On the other hand, at 32 °C, higher temperature accelerates protein expression and folding rate, leading to the production of the 150 mM elution fraction as the major product.

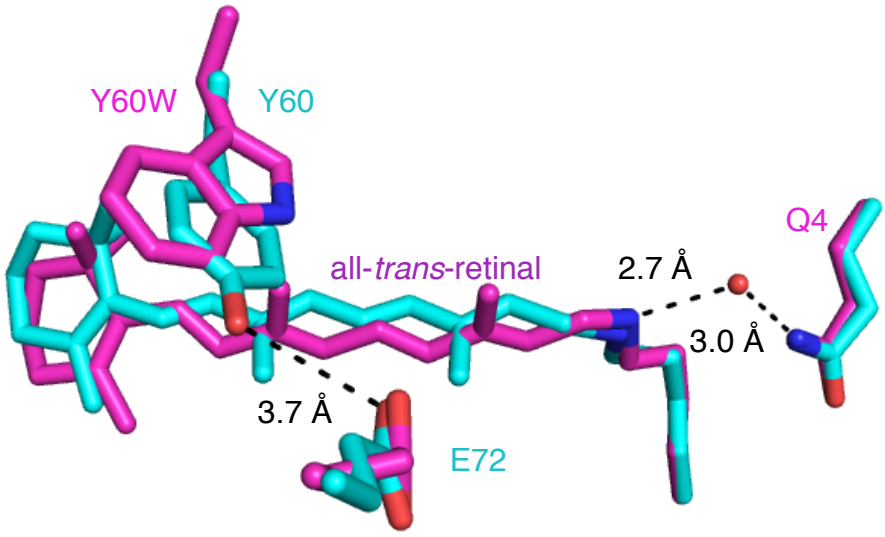


Figure III-18: Crystal structure of Q108K:K40L:Y60W 40 mM elution (magenta) overlaid with Q108K:K40L (cyan) bound with all-*trans*-retinal.

Crystal structures of the two fractions can help uncover the different conformations that lead to the different behaviors of the two protein fractions of Q108K:K40L:Y60W. The crystal structure of protein eluted at 40 mM salt concentration of Q108K:K40L:Y60W protein bound with retinal was obtained to high resolution (Rafida Nossoni, Jim Geiger's lab, MSU). However, the crystal structure of holo-Q108K:K40L:Y60W 150 mM elution fraction did not exhibit any density for the chromophore. From the available structural data, we could speculate the reasons for the differences of these two protein elution fractions.

Y60 is placed on a β-sheet where the gap between the two sheets is most open compared to all the other gaps. Y60 also hydrogen bonds with E72, which

is sitting in the neighboring β-sheet, either directly or through a water mediated interaction. The blue shift caused by mutation of Y60W for the 40 mM elution fraction seems to be a result of the conformational change of the chromophore.

The backbone of the crystal structures of Q108K:K40L and Q108K:K40L: Y60W (40 mM elution) overlay very well, with an RMS of 0.301 Å for all 133 residues. However, the plane of the polyene in the Q108K:K40L:Y60W-retinal rotates almost 90° from that in Q108K:K40L, while the ionone ring of the two chromophores overlay nearly perfectly (**Figure III-18**). As a result, the retinal in Q108K:K40L:Y60W is highly twisted, with a 6-s-*cis* conformation while that in Q108K:K40L adopts a 6-s-*trans* conformation. The highly twisted nature of the conformation of Q108K:K40L:Y60W (40 mM elution) could result in a large decrease in conjugation for the polyene system, leading to a blue shift.

It is likely that the binding pocket of Q108K:K40L:Y60W (40 mM protein eluent) allows enough freedom for the polyene to rotate and adopt a 6-s-*cis* conformation, as the 6-s-*cis* conformation is 0.6 Kcal/mol more stable than 6-s-*trans*.[14] The ionone ring of the chromophore is rigidified by the protein environment, with the rotation restricted, while the polyene is more flexible for CRBPII mutants without Q4, R58 and Y19 mutations. A highly twisted chromophore is observed in mutants Q108K:K40L:T51V and Q108K:K40L:T53C as well.

Unlike Q108K:K40L:Y60W, which produced protein fractions that eluted at 150 mM salt concentration, Q108K:K40L:Y60F and Q108K:K40L:Y60H yields

proteins that eluted at 40 mM salt concentration as the major fraction, similar to most of the CRBPII mutants. A blue shift results for both Q108K:K40L:Y60F and Q108K:K40L:Y60H to 494 nm. This shows that the phenomenon is unique for Q108K:K40L:Y60W, which might indicate the unique role of tryptophan in wavelength tuning. A large effort is still underway to understand how the 150 mM elution fraction results in a red shift.

III.3.4 Red shift is induced by placing polar residues at position 119

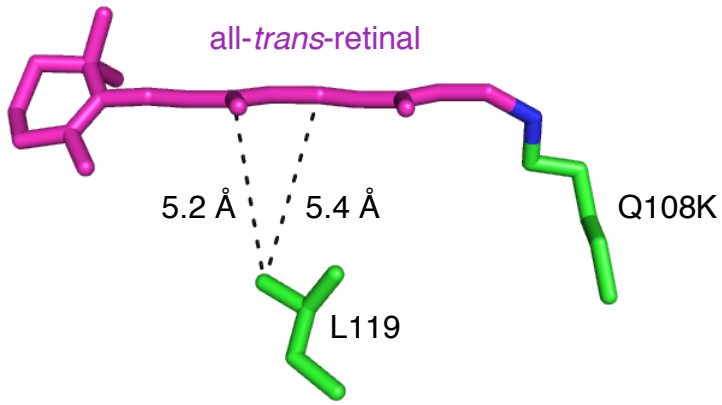


Figure III-19: Crystal structure of Q108K:K40L, with L119 highlighted.

The last position studied in the middle of the polyene is Leu119. Leu119 is on a β-sheet, pointing toward the middle of the polyene and is more than 5 Å from the conjugated polyene (**Figure III-19**). It was found that the chromophore is sensitive to a polarity change in CRBPII, even for residues 5 Å away from the polyene, due to the hydrophobicity of the binding pocket. It is interesting to note that mutations at position 119 could also generate some degree of disturbance

on the electronic profile of the chromophore, although it is far away from the polyene.

Starting from a neutral residue, a number of polar residues were substituted at this position to probe the effect of polarity change projected on the middle of the polyene on wavelength regulation. As **Table III-5** shows, the general trend is clear that when negative polar residues are introduced, a red shift is observed. This agrees with the hypothesis that increasing the negative polarity in the middle of the polyene can increase the charge delocalization through favorable electronic interactions with the delocalized positive charge. Distance plays an important role when comparing the red shift caused by introduction of L119Q with that of L119N. L119Q results in a 14 nm red shift, while L119N results in only 5 nm red shift, since it has a shorter side chain as thus is further away from the polyene (**Figure III-20**).

Table III-5: Mutagenesis studies at position 119.

CRBPII mutant	λ_{max} (nm)	Protein shift (nm)
Q108K: K40L: L119	508	0
Q108K: K40L: L119C	510	2
Q108K: K40L: L119T	516	8
Q108K: K40L: L119N	513	5
Q108K: K40L: L119Q	522	14
Q108K: K40L: L119F	509	1
Q108K: K40L: L119Y	518	10
Q108K: K40L: L119D	n.d.	n.d.
Q108K: K40L: L119E	n.d.	n.d.

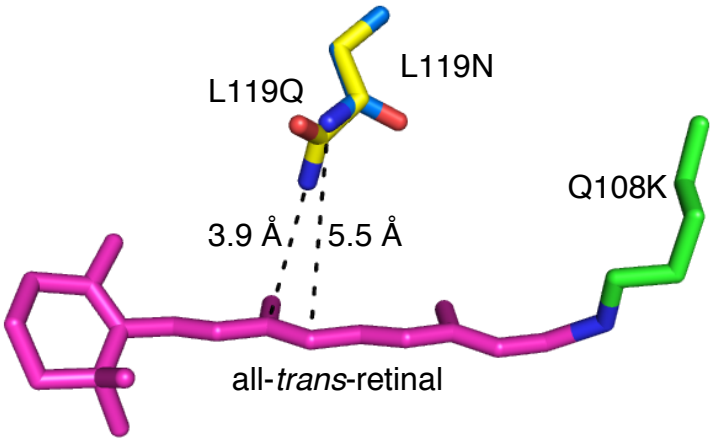


Figure III-20: Modeled structures of Q108K:K40L:L119Q and Q108K:K40L:L119N.

L119Y results in a 10 nm red shift, due to the negative polarity introduced from the hydroxyl group. The crystal structure of Q108K:K40L:L119Y was not resolved, but mutation of L119Y was modeled in Pymol based on the crystal structure of Q108K:K40L bound with retinal. Two possible rotamers were modeled for L119Y, with one of them pointing the hydroxyl group toward the ionone ring region of the chromophore, but 6.6 Å away, and the other one pointing toward the middle of the polyene, 2.9 Å away **(Figure III-21)**. Either

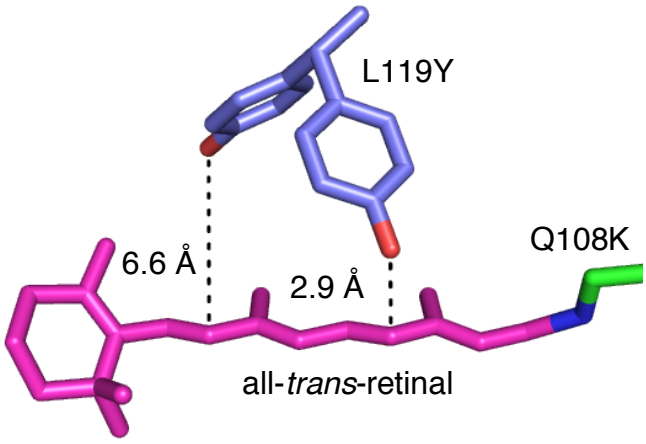


Figure III-21: Model structure of Q108K:K40L:L119 with two possible rotamers.

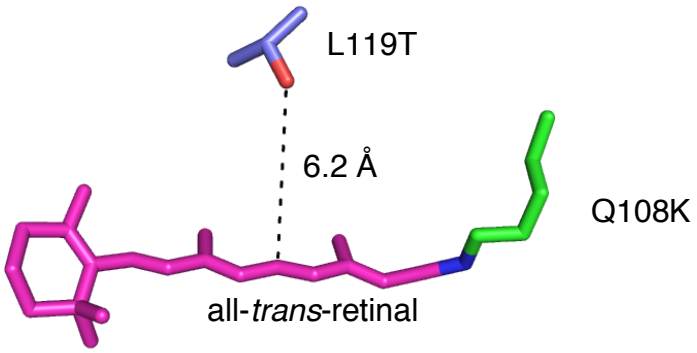


Figure III-22: Model structure of Q108K:K40L:L119T.

conformation could increase the positive charge delocalization along the polyene. In contrast to L119Y, L119F does not result in an obvious red shift, as phenylalanine is nonpolar. The comparison of L119Y and L119F supports the idea that the red shift of L119Y is a result of polarity from the hydroxyl group.

L119T results in an 8 nm red shift, due to the negative polarity from the hydroxyl group (**Figure III-22**) and L119C only leads to a 2 nm red shift (**Figure III-23**). Attempts were also made to introduce the more polar aspartic acid and glutamic acid at position 119, however, no soluble proteins were obtained for characterization.

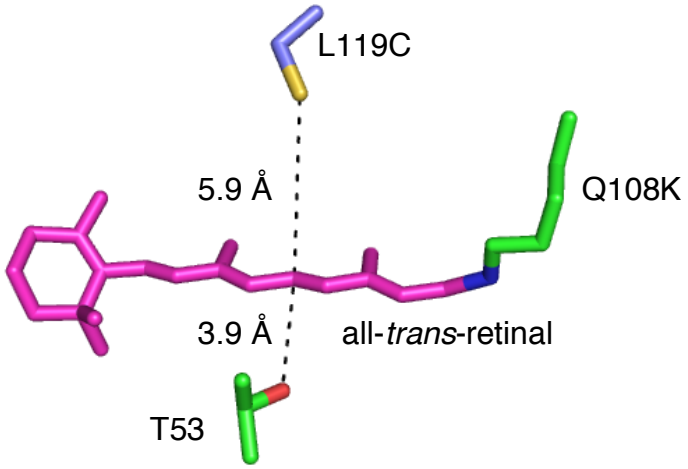


Figure III-23: Model structure of Q108K:K40L:L119C.

Mutagenesis studies at position 119 further demonstrate that the electronic interactions between the surrounding residues and the polyene system play an important role in the absorption profiles of the retinal-PSB. This is due to the perturbation of the different electronic state of the chromophore in the ground state and excited state.

Unexpectedly, L119 mutations slow down protonated Schiff base formation. It takes at least 2 hours for the PSB formation to be complete for L119 mutants, compared to the maturation time of Q108K:K40L that require half an hour. The reason for this slower kinetics is not clear yet, as no crystal structures are available for L119 mutants and it is hard to judge how a residue in the middle of the binding pocket can affect the Schiff base formation.

Table III-6: UV-vis data for mutants containing L119Q

CRBPII mutant	λ_{max} (nm)	Protein shift by L119Q (nm)
Q108K:K40L:L119	508	0
Q108K:K40L:L119Q	522	14
Q108K:K40L:T51C:L119	513	0
Q108K:K40L:T51C:L119Q	522	9
Q108K:K40L:T51V:L119	533	0
Q108K:K40L:T51V:L119Q	543	10
Q108K:K40L:T51V:T53C:L119	539	0
Q108K:K40L:T51V:T53C:L119Q	547	8
Q108K:K40L:T51V:T53C:R58W:L119	585	0
Q108K:K40L:T51V:T53C:R58W:L119Q	556	-29

The other intriguing fact about L119 mutations is that they do not lead to red shift in all cases. As shown in **Table III-6**, L119 results in red shift before R58W is introduced and results in blue shift when introduced to the pentamutant Q108K:K40L:T51V:T53C:R58W (585 nm).

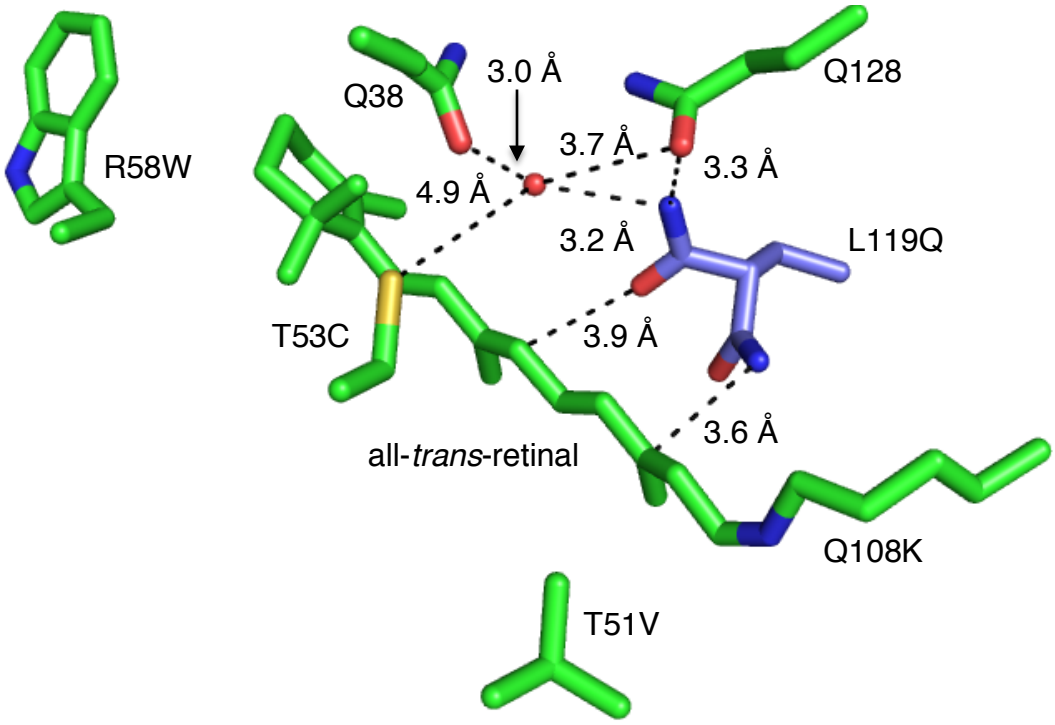


Figure III-24: Model structure of Q108K:K40L:T51V:T53C:R58W:L119Q with L119Q showing two possible conformers.

It is interesting that the effect of L119Q depends on the protein environment it is in. This might be because L119Q can take two different conformations (**Figure III-24**), positioning the amide group either more toward the ionone ring or the Schiff base. These two different conformations will result in different effects. If the amide group is pointing toward the ionone ring, red shift can result, as more negative polarity is projected toward the ionone ring, encouraging the positive charge delocalization. The other conformation, which

points toward the PSB region, could have the opposite effect, decreasing positive charge delocalization and leading to a blue shift.

The blue shift of the L119Q mutation in the presence of R58W could also be due to the bulkiness of Gln119, which pushes the chromophore towards R58W and changes the conformation of the chromophore in the ionone ring region in the presence of R58W. It requires more investigation to probe the different roles of L119Q.

III.4 Probing the effects of residues close to the ionone ring region on wavelength tuning

III.4.1 Mutations of F16 and A33

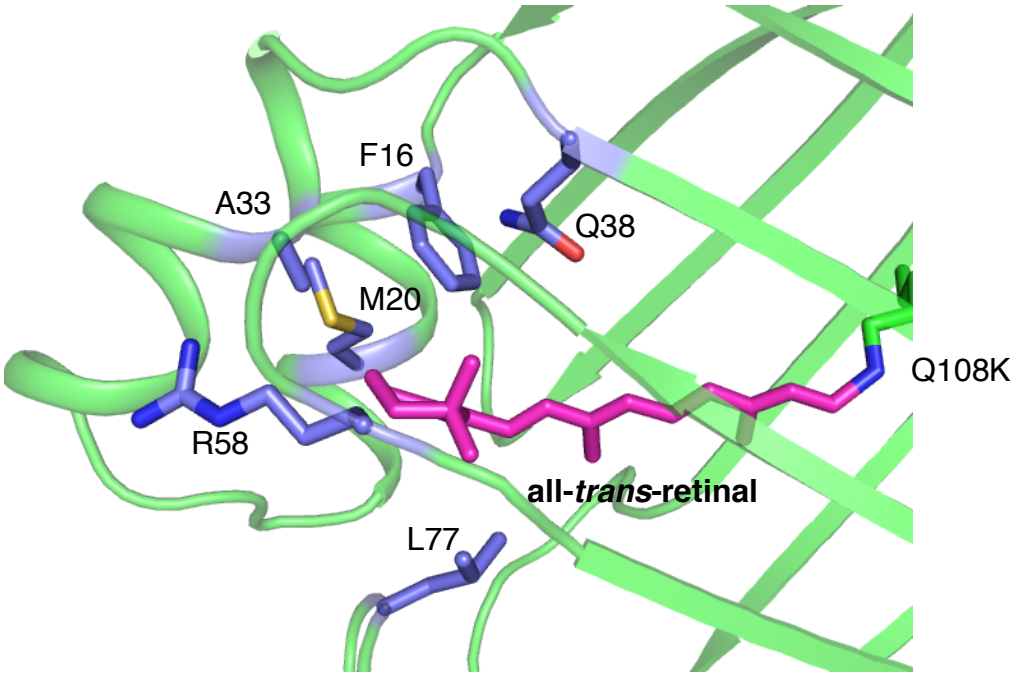
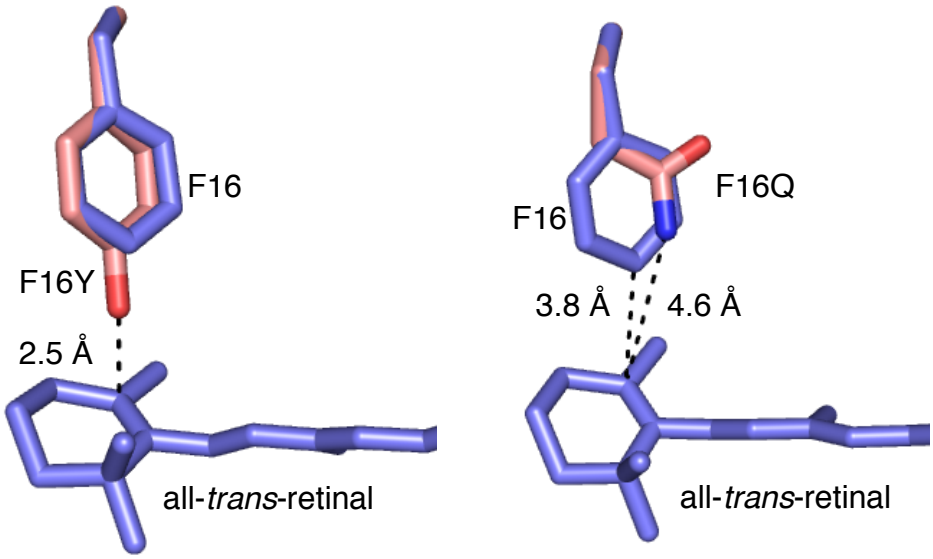


Figure III-25: Crystal structure of Q108K:K40L bound with all-*trans*-retinal, with surrounding residues in the β-ionone ring region highlighted.

I have shown that mutations carried out in the PSB region and the middle region of the chromophore all resulted in some degree of regulation of the absorption maxima of the retinal-PSB formed with CRBPII mutants, mostly due to electronic interactions of the surrounding residues with the chromophore. It is interesting to see whether mutations in the ionone region will have similar effects. However, factors other than electronic interactions should be considered for mutants made in the ionone ring region. This is because the packing between the chromophore and the protein residues in the ionone ring region can directly dictate the conformation of the chromophore. As it has been suggested a number of times in the literature, rotation along C6-C7 dictates to a large degree the absorption maximum of retinal-PSB.[14, 15, 16, 17]

There are quite a few positions in the ionone ring region in close proximity to the retinal, such as Leu77, Phe16, Ala33, Met20, T29L and Arg58 (**Figure III-25**). Mutations at each of these positions were carried out to change the electronics of the side chain to observe the cause-effect relationship. For some of these positions, the results obtained were quite unexpected. For example, when polar residues were introduced to replace nonpolar residues in some of these positions such as F16Y and A33S, a red shift was expected because negative polar residues could increase the negative polarity projected on the ionone region to favorably interact with the delocalized positive charge, however a blue shift was observed. This indicates that some conformational changes might have been responsible for the observed and unexpected results.

It is highly likely that mutations of Phe16 lead to conformational change of the chromophore in the ionone ring region, as Phe16 seems to play a significant role in restricting the rotation of the ionone ring through tight packing with the



Q108K:K40L:F16Y
λ_{max} = 486

Q108K:K40L:F16Q
λ_{max} = 486

Figure III-26: Model structure of F16 mutations. **a**. Model structure of Q108K:K40L:F16Y based on the crystal structure of Q108K:K40L, showing that F16Y is crashing in to the C5-methyl group of retinal. **b**. Model structure of Q108K:K40L:F16Q based on the crystal structure of Q108K:K40L, showing one possible rotamer of F16Q .

chromophore. The Q108K:K40L:F16Y-triple mutant could possibly disturb the planarity of the ionone ring by clashing into carbon 5 of the retinal as shown in the modeled structure of Q108K:K40L:F16Y based on the crystal structure of Q108K:K40L (**Figure III-26**). Disturbance of planarity would result in less conjugation, therefore leading to a blue shift, which is observed at 486 nm.

With that in mind, mutant Q108K:K40L:F16Q was made to place a small polar residue, glutamine, at position 16 (**Figure III-26b**). Although glutamine

probably does not sterically interact with the chromophore, due to its smaller size, it would allow free rotation of the ionone ring. A twisted 6-s-*cis* conformation is more stable than the flat 6-s-*trans* conformation,[17] this could also result in rotation of the ionone ring to a more stable conformation with the ionone ring rotating out of the plane of the polyene that leads to the observed 20 nm blue shift. However, at this stage, the crystal structures of F16Y or F16Q are not available to verify this hypothesis.

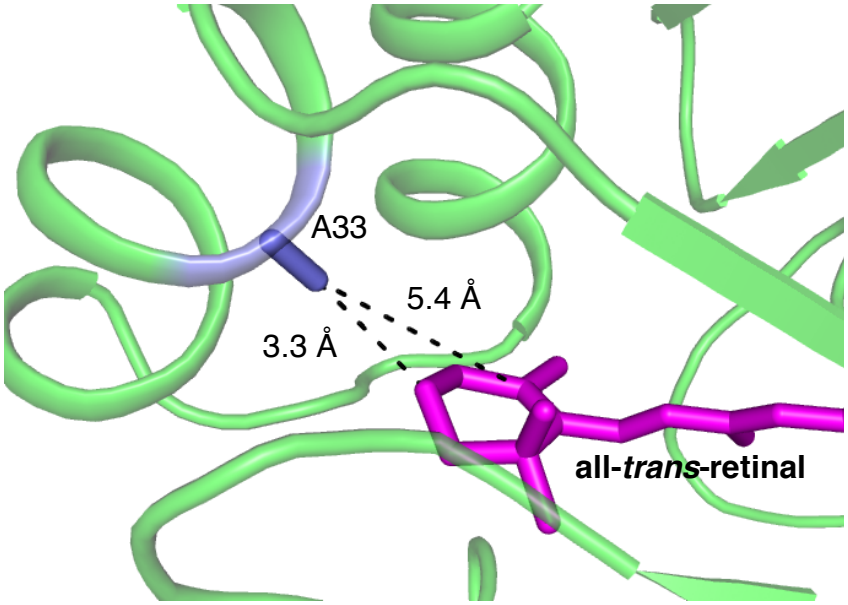


Figure III-27: Crystal structure of Q108K:K40L with A33 highlighted.

For position Ala33, Q108K:K40L:A33S was prepared intending to increase the negative polarity in the ionone region (**Figure III-27**), while maintaining a similar size to alanine. Increased negative potential in the ionone ring region could interact favorably with the delocalized positive charge and lead to a red

shift. Unexpectedly, Q108K:K40L:A33S triple mutant blue shifts to 502 nm as compared to Q108K:K40L (508 nm).

III.4.2 Mutations of L77

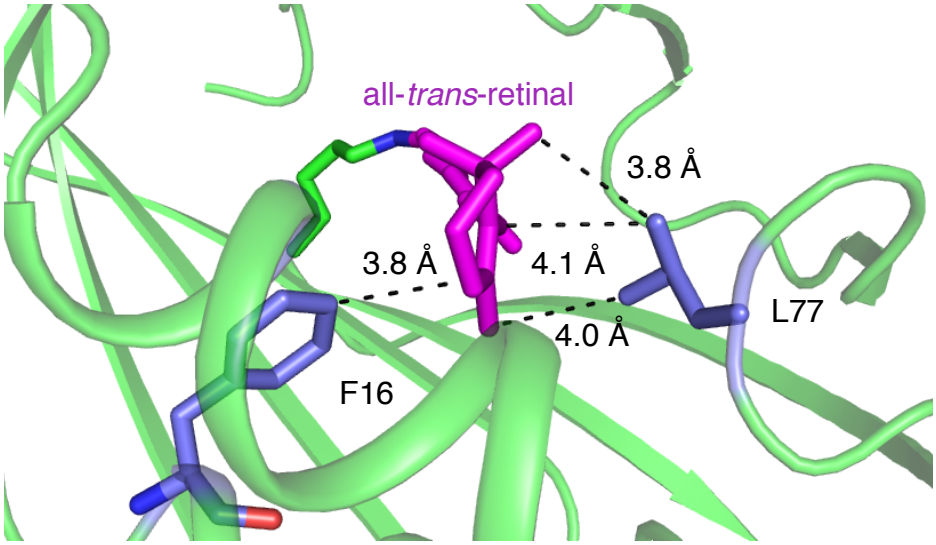


Figure III-28: Crystal structure of Q108K:K40L with L77 and F16 highlighted.

Similar to Phe16, Leu77 also has a tight packing interaction with the ionone ring (**Figure III-28**). It is believed that Leu77 prevents the retinal from adopting a 6-s-*cis* conformation. Q108K:K40L:L77E was made to increase the negative polarity in the ionone ring region. Q108K:K40L:L77E does not lead to a red shift but a slight blue shift to 504 nm. Since L77E is on the loop, switching from a hydrophobic residue to a hydrophilic residue such as glutamic acid could move the conformation of the loop so that L77E can flip out of the binding pocket to get more solvated in the aqueous solution.

Surprisingly, introduction of even neutral polar residues, L77Q, L77M and L77C, all lead to blue shifts (**Table III-7**). Only when large aromatic residues

such as phenylalanine and tryptophan, are introduced, no change is observed in absorption maxima. Supposedly they can also have a tight packing with the chromophore and rigidify the rotation of the ionone ring. This indicates that the chromophore is sensitive to mutations at position 77 conformationally, but inert to polarity changes of the residue.

Table III-7: Mutagenesis studies of position 77.

CRBPII mutant	λ_{max} (nm)	Protein shift (nm)
Q108K: K40L: L77	508	0
Q108K: K40L: L77E	504	-4
Q108K: K40L: L77C	501	-7
Q108K: K40L: L77M	500	-8
Q108K: K40L: L77Q	500	-8
Q108K: K40L: L77F	508	0
Q108K: K40L: L77W	509	1

The polarity changes made in the ionone ring region do not seem to contribute to red shift. It is possible that the local dielectric constant in the ionone ring region is higher than that deep inside the binding pocket, due to the relative openness in this region of the binding pocket (**Figure III-29**). This makes the chromophore inert to polarity changes in the ionone ring region.

At the same time, as the conformation of the ionone ring is sensitive to the surrounding residues and can change its relative position versus the polyene, generating a more twisted chromophore along the C6-C7 single bond could decrease the degree of conjugation and lead to blue shift. This has been shown

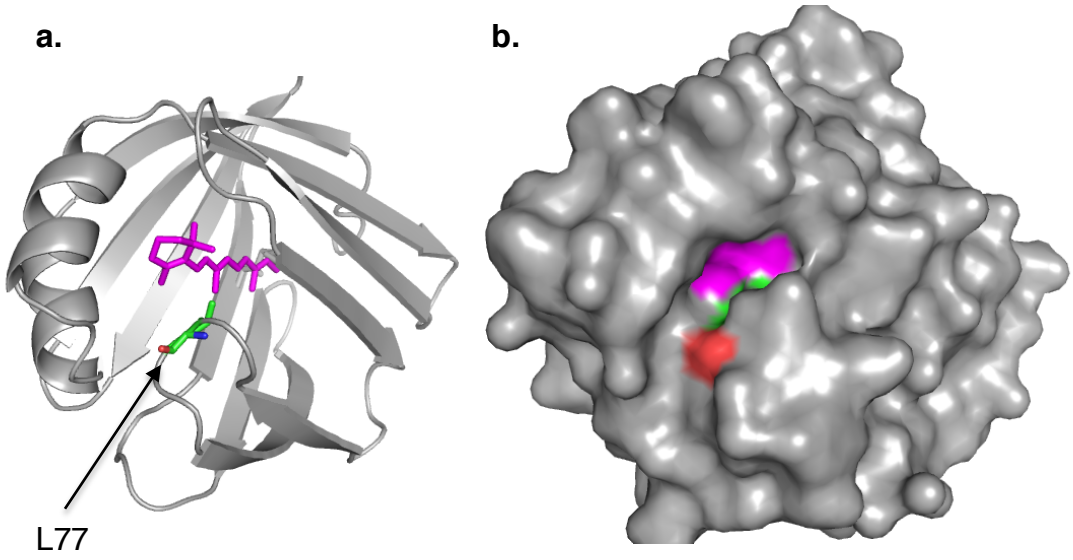


Figure III-29: Relative openness in the β-ionone ring region. (a) Cartoon and (b) surface of the crystal structure of Q108K:K40L with the ligand retinal (magenta) and residue L77 (C-green, O-red, N-blue) highlighted, showing that there is a slight openness as shown in (b), with the chromophore and L77 slightly exposed to water.

before with mutagenesis studies in bacteriorhodpsin, where mutation of Met118, which has tight packing with the chromophore, results in a dramatic blue shift.[18]

Table III-8: Absorption data of mutants with L77T.

CRBPII mutant	λ_{max} (nm)	Protein shift by L77 mutant (nm)
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4H:L77	585	0
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4H:L77T	563	-22
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4W:L77	613	0
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4W:L77T	593	-20
Q108K:K40L:T51V:R58Y:Y19W:L77	565	0
Q108K:K40L:T51V:R58Y:Y19W:L77T	555	-10

As shown in **Table 8**, placing a neutral polar residue, L77T, in different CRBPII mutants all lead to a blue shift. This result is counter-intuitive, and most probably suggests that conformational change of the chromophore is the major cause for the blue shift observed.

Crystal structure of Q108K:K40L:R58Y:T51V:Y19W:L77T was solved (Camille Watson, Babak Borhan's lab, MSU). Overlay of the crystal structures of Q108K:K40L:R58Y:T51V:Y19W:L77T and Q108K:K40L:R58Y:T51V:Y19W gives an RMS of 0.252 Å for all the 133 amino acids, indicating that introduction of L77T barely changed the conformation of the overall protein scaffold.

However, as shown in **Figure III-30**, introduction of L77T, which is smaller than L77 opens up space for the chromophore to move down toward L77T for tighter packing. As a result, the ionone ring of the chromophore translates

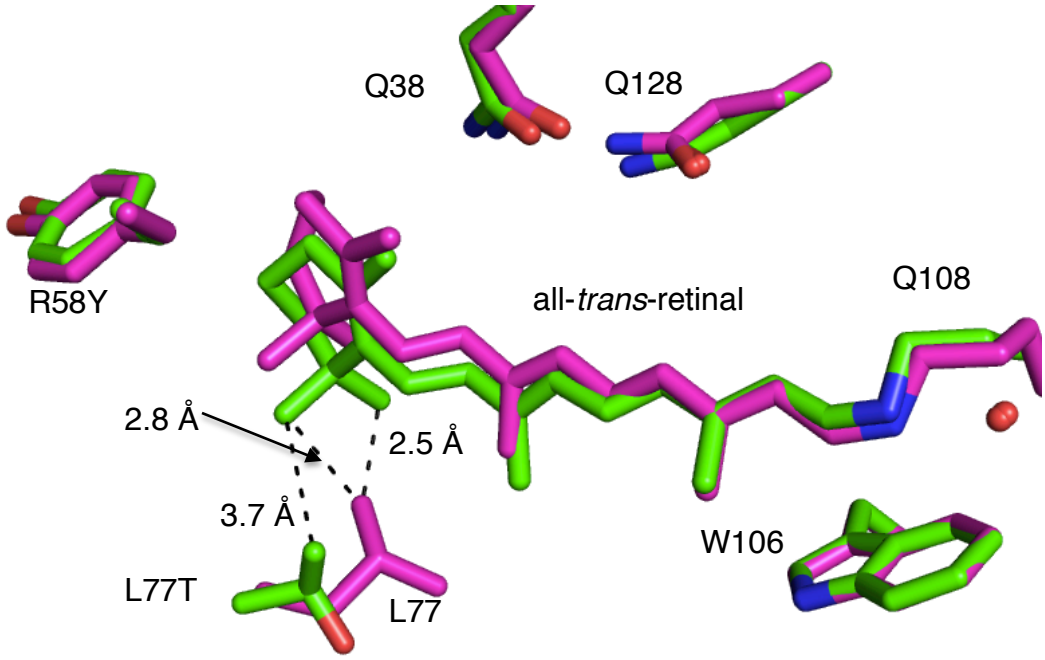


Figure III-30: Overlaid crystal structures of CRBPII mutants with and without L77T. Q108K:K40L:R58Y:T51V:Y19W (magenta) and Q108K:K40L:R58Y:T51V:Y19W:L77T (green).

downward, away from the more polar side of the pocket, where Q38 and Q128 reside. This translation might lead to the reduction of negative polarity projected on the ionone ring and lead to less stabilization of the excited state of the chromophore and thus the observed blue shift.

Interestingly, L77T adopts a conformation, with the hydroxyl group pointing away from the chromophore, which might allow better solvation of the hydroxyl group. Therefore, L77T is not in an optimal position to interact with the delocalized positive charge of the retinal-PSB.

The L77T studies illustrate the importance of crystal structures in understanding the wavelength regulation. Unexpected conformational change of the chromophore or protein might lead to unanticipated effects.

III.4.3 introduction of polar residues at positions 20 and 29

It was found that the chromophore is not as sensitive to polarity change in the ionone ring region as in the middle of the polyene and PSB region. This could be explained by two reasons, one of them being that the conformation of

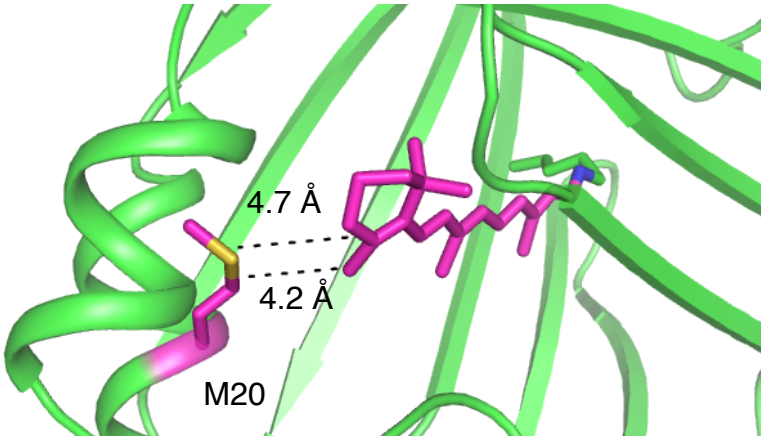


Figure III-31: Graphic showing the position of M20 in the crystal structure of Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4H.

the ionone ring is sensitive to the surrounding residues. The energy difference between 6-s-*cis* and 6-s-*trans* was calculated to be only 0.6 Kcal/mol,[14] a slight change in the side chain might lead to different hydrophobic packing and changes in the conformation of the chromophore. Therefore the net effect of polarity change in the ionone ring region is complicated by potential conformational changes, as well as translational motion.

Table III-9: Mutants of Met20 in different templates

CRBPII mutant	λ_{max} (nm)	Protein shift by M20 mutant (nm)
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4F:M20	613	0
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4F:M20Q	612	-1
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4F:M20E	618	5
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4W:M20	613	0
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4W:M20T	609	-4
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4W:M20C	605	-8
Q108K:K40L:T51V:T53C:R58W:M20	585	0
Q108K:K40L:T51V:T53C:R58W:M20W	569	-16

The second reason is that the binding cavity has a slight opening to the surface and consequently the local dielectric constant in the ionone ring region can be higher than that deeper inside the binding pocket. As reported previously,

the local dielectric constant tends to decrease as it goes deeper into the hydrophobic binding pocket.[19] Therefore, polarity change in the ionone ring region dissipates faster and requires a larger polarity change to induce a similar red shift as that deep inside the binding pocket.

Met20 is on one of the α-helices, pointing toward the ionone ring of the retinal-PSB (**Figure III-31**). A few Met20 mutants were prepared as shown in **Table III-9**. Interestingly, neutral polar residue M20Q has almost no effect, while M20E results in a 5 nm red shift, due to its more drastic polarity change through introduction of a negatively charged amino acid in the ionone ring region. M20T and M20C both lead to a slight blue shift, probably due to conformational change of the chromophore due to reduction of size, which allows more space for free rotation of the ionone ring. Similarly, M20W also leads to a large blue shift, as M20W probably clashes into the ionone ring and forces it to twist along the single bond C6-C7. Generally speaking, it was found that mutagenesis in the ionone

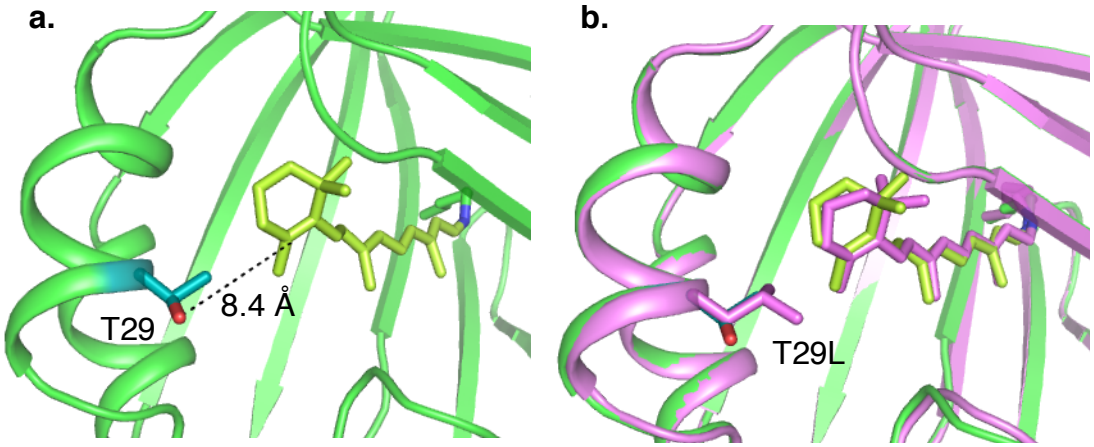


Figure III-32: Mutation of T29. **a.** Crystal structure of Q108K:K40L:T51V: R58W:Y19W with T29 highlighted. **b.** Overlaid crystal structure of Q108K: K40L:T51V:R58W:Y19W (green) and Q108K:K40L:T51V:T53C:R58W: Y19W:T29L (pink), with T29 and T29L highlighted.

ring region is more complicated as it might also involve conformational change of the chromophore.

Table III-10: Mutants of Thr29.

CRBPII mutant	λ_{max} (nm)	Protein shift by T29 mutant (nm)
Q108K:K40L:T51V:T53C:R58W:Y19W:Q4W:**T29**	613	0
Q108K:K40L:T51V:T53C:R58W:Y19W:Q4W:**T29L**	613	0
Q108K:K40L:T51V:T53C:R58W:Y19W:Q4W:**T29D**	616	3
Q108K:K40L:T51V:T53C:R58W:Y19W:Q4W:**T29E**	616	3
Q108K:K40L:T51V:T53C:R58W:Y19W:Q4W:**T29W**	615	2
Q108K:K40L:**T29**	508	0
Q108K:K40L:**T29E**	496	-12

T29 is also a residue that resides on the α-helices. The hydroxyl group is pointing outside of the binding pocket due to better solvation of the polar hydroxyl group. As a result the hydroxyl group is 8.4 Å away from the double bond of ionone ring (**Figure III-32a**). A hydroxyl group so far away from the chromophore, which is also slightly water exposed, can not have significant electrostatic interactions with the chromophore. Replacement of T29 with T29L does not cause any change in the absorption maximum of the chromophore, but it seems to stabilize the protein better due to hydrophobic packing. Therefore, most of the later series of CRBPII mutants contain mutation T29L. As shown in

Figure III-32b, replacement of T29 by leucine does not cause any conformational change of the helix either.

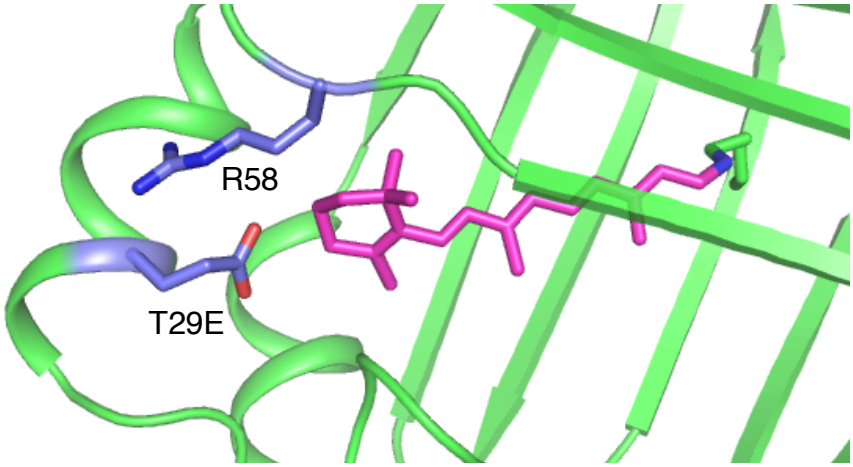


Figure III-33: Model structure of Q108K:K40L:T29E based on the crystal structure of Q108K:K40L.

Polar residues such as T29D and T29E were also examined at this position. Surprisingly, both mutations led to a 3 nm red shift in the more red shifted series of CRBPII mutants. The minor red shift indicates that indeed a dramatic polarity change in this region can only result in a minor protein shift. It is interesting to see that only with R58W mutation in place, a red shift results, while for Q108K:K40L, introduction of T29E yields a blue shift. That might be a result of salt bridge interaction between T29E and R58 as shown in **Figure III-33**. Another reason might also be that R58W covers up the pocket more effectively and therefore, creates a more hydrophobic, more shielded binding pocket. The polarity of T29E can have a stronger effect, as will be discussed next.

III.4.4 Red shift as a result of placing aromatic residues at position 58

Arg58 resides on a loop, which partially covers the binding pocket of CRBPII. The original intention of mutation Arg58 was to remove the positive polarity of Arg58 in the ionone region in order to promote positive charge delocalization toward the ionone ring. As **Figure III-34** shows, the positively charged guanidine part of Arg58 is more than 7.8 Å away from the double bond carbon of the ionone ring and is water exposed. As such a large red shift is not expected by removal of Arg58. Nonetheless, a few mutations to remove the positive charge of R58 were carried out.

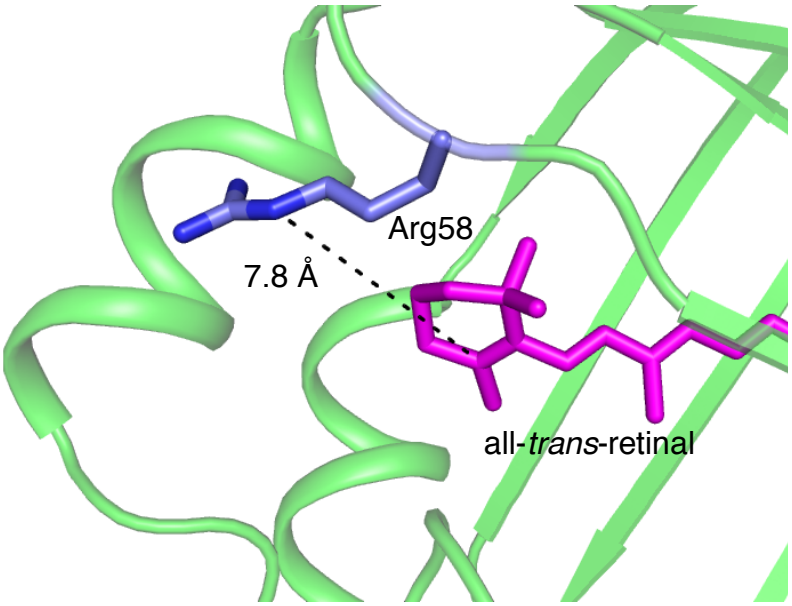


Figure III-34: Crystal structure of Q108K:K40L with Arg58 highlighted.

Negatively charged residues were introduced, considering that switching of the polarity in this position should have the most significant change on the overall electrostatic potential projected on the retinal. However, both mutants, Q108K:K40L:R58E and Q108K:K40L:R58D, do not lead to red shift, but a slight

blue shift. Two factors could account for the fact that no red shift is observed by introduction of negatively charged residues at position 58.

First, Arg58 is far away from the conjugated system of the chromophore, it does not project much positive electrostatic field on the chromophore, since electric field projected on the chromophore is inversely proportional to r^2. Besides, many calculations have shown that for charged residues that are water exposed, the effect is buffered by water and ions in the solution to a large extent and most of these residues could even be considered as neutral.[13]

Secondly, the ionone ring is close to the surface of the binding pocket and relatively exposed to the aqueous environment compared to the residues deep inside the binding pocket. Therefore, the electric field projected from Arg58 mutations dissipates fast in this region and as a result does not contribute much

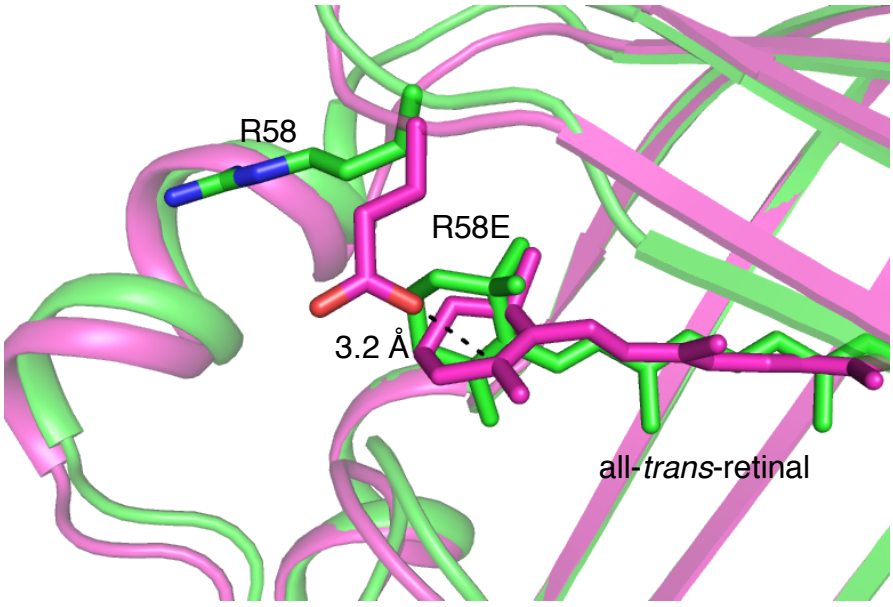


Figure III-35: Crystal structure overlay of Q108K:K40L (magenta) and Q108K:K40L:R58E (green), with position 58 highlighted.

to the overall electrostatic projection on the chromophore.

As the crystal structure of Q108K:K40L:R58E bound with retinal (**Figure III-35,** Rafida Nossoni, Jim Geiger's lab, MSU) shows, R58E is 3.2 Å away form the double bond of the ionone ring. It is in a good position to influence the electric field projected on the ionone ring region of the chromophore, but R58E results in a blue shift instead. This indicated that the dielectric constant for this region is high and it greatly reduces the electronic effects of the polar residue R58E.

However, there could be other reasons that R58E causes blue shift. The overlaid crystal structures of Q108K:K40L:R58E and Q108K:K40L bound with retinal show that the chromophore in Q108K:K40L:R58E rotates almost 90° away from its original position in Q108K:K40L (**Figure III-35**), which could place the chromophore in a position with slightly different electrostatic environment and lead to the observed blue shift. It is not clear how the R58E mutation leads to the rotation of the whole chromophore.

Besides acidic residues, small neutral residues were also placed at position 58, like Q108K:K40L:R58L, Q108K:K40L:R58A, Q108K:K40L:R58Q, all of which result in similar degree of blue shift. As even R58E and R58D do not result in red shift, it is not surprising to find that simple removal of the positive charge does not result in red shift either.

However, when large aromatic residues, such as R58W, R58F and R58Y are introduced, different degrees of red shifts are observed (**Table III-11**). The

effect of position 58 on wavelength regulation is intriguing. Mutation of Arg58 into either small neutral residues or negatively charged residues all lead to blue shift, and red shift results only when large aromatic residues such as Phe, Tyr and Try were utilized (**Table III-11**).

Table III-11: Mutagenesis studies of position Arg58.

CRBPII Mutant	λ_{max} (nm)	pK_a	K_d (nM)
Q108K:K40L:**R58**	508	8.3	29±5
Q108K:K40L:**R58D**	500	8.6	26±6
Q108K:K40L:**R58E**	500	8.8	24±5
Q108K:K40L:**R58L**	500	8.1	16±7
Q108K:K40L:**R58A**	499	8.1	20±6
Q108K:K40L:**R58Q**	499	8.4	43±3
Q108K:K40L:**R58W**	519	8.7	43±4
Q108K:K40L:**R58F**	524	8.6	27±6
Q108K:K40L:**R58Y**	535	9.5	10±7

Arg58 in CRBPII corresponds to Arg59 in the first generation rhodopsin mimic based on CRABPII, according to the sequence alignment and crystal structure comparison. In CRABPII, full length retinal did not respond to a polarity change at position 59, but a shorter chromophore, C15 analogue, did respond in an expected manner. This was so because C15 could be fully embedded inside the binding pocket. However, in CRBPII, the role of position 58 in wavelength regulation is different from that in the CRABPII-C15 system. The red shift

induced by placing aromatic residues at position 58 is not simply a result of polarity change at position 58. The conclusion is drawn because mutations R58D, R58E, R58L, R58Q and R58A all lead to blue shift, although positive polarity from Arg58 is removed.

The red shift induced by introduction of aromatic residues at position 58 in CRBPII mutants is believed to be a result of more effective shielding of the binding pocket and thus the chromophore. Different CRBPII mutants with mutation R58W were crystallized. As shown in **Figure III-36**, although R58W adopts different rotamers in the crystal structures of different CRBPII mutants, most of the rotamers of R58W cover the binding pocket more effectively as compared to mutants that have Arg in position 58.

The space filling representation of Q108K:K40L:R58 shows that the blue color of the ionone ring is partially exposed. As a result, the chromophore is not sensitive to polarity changes in the ionone ring region. The blue shift caused by mutation of R58 into smaller residues such as leucine and Gln might be that Leu and Gln can not cover the pocket as well as Arg, due to their smaller sizes.

Three out of four rotamers of R58W could well seal the hole in the ionone ring region. We believe that with the R58W mutation in place, the chromophore can be better sequestered from the aqueous solution and thus polarity changes lead to a larger impact on the electrostatic potential projected on the chromophore.

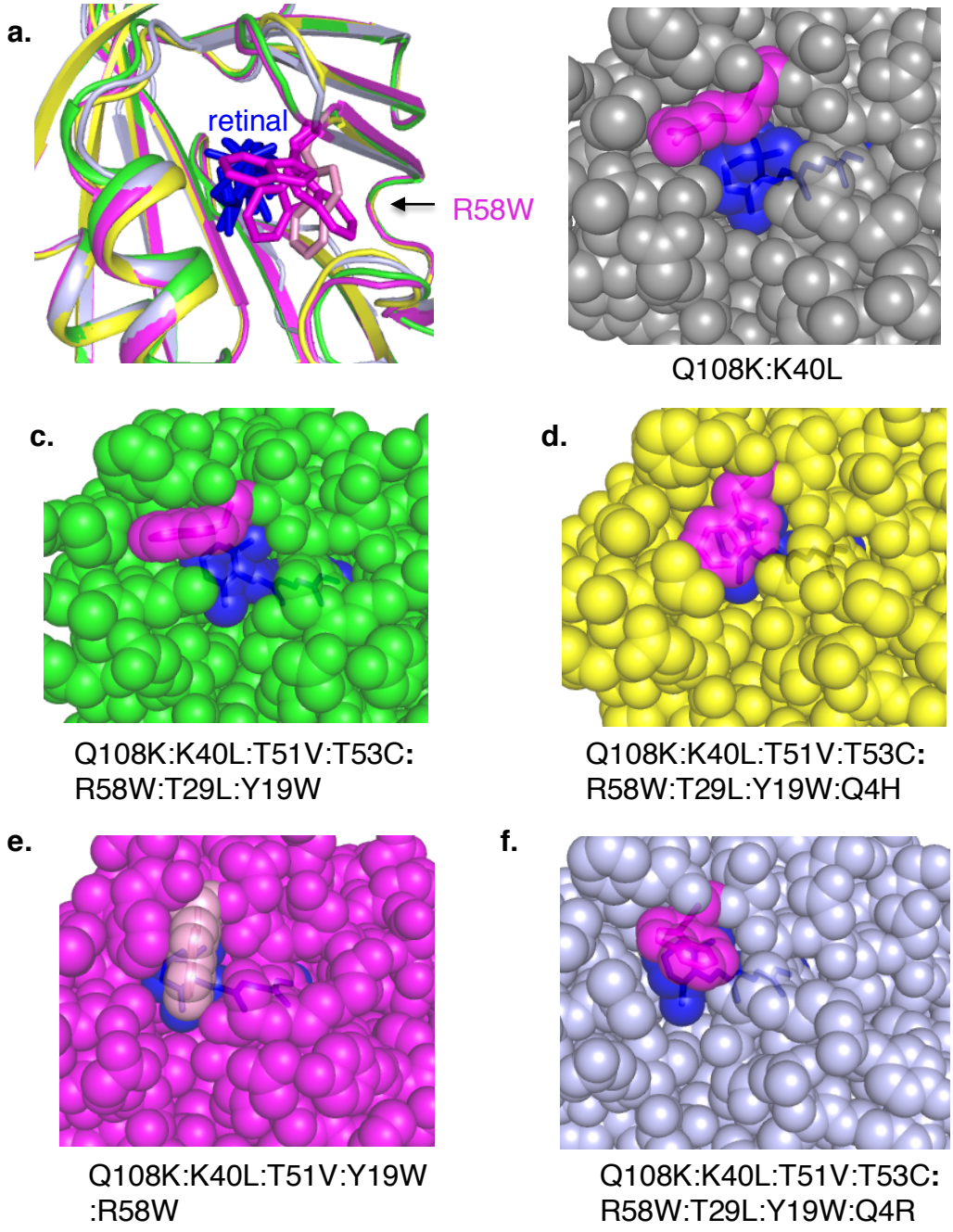


Figure III-36: Different R58W rotamers for different CRBPII mutants. **a.** Overlaid crystal structures of CRBPII mutants with R58W mutation, showing different rotamers of R58W in different mutants, Q108K:K40L:T51V:T53C**:** R58W:T29L:Y19W (green), Q108K:K40L:T51V:Y19W:R58W (magenta), Q108K:K40L:T51V:T53C**:**R58W:T29L:Y19W:Q4H (yellow), Q108K:K40L:T51V: T53C**:**R58W:T29L:Y19W:Q4R (light blue). **b-f.** Space filling model of Q108K: K40L and different CRBPII mutants corresponding to the structures in **a** with retinal (blue) and 58 (magenta) highlighted.

III.4.5 Enhanced red shift in the presence of R58W for T51V, T53C and Y19W

R58 resides in the entry of the binding pocket and it acts as a lid for the cavity. Introduction of large aromatic residues at position 58 can better seal the binding pocket and sequester the chromophore from the bulk aqueous solution outside the pocket. This way, the dielectric constant inside the binding pocket is decreased. A more hydrophobic environment with lower dielectric constant can enable the same polarity change to have a more significant impact on the chromophore inside the pocket. Thus the same mutation made inside the binding pocket should result in a larger protein shift in the presence of R58W.

Table III-12: Comparison of protein shift caused by the same mutant with and without R58W.

Q108K:K40L (KL), λ_{max} = 508 nm			
CRBPII Mutant	λ_{max} (nm)	$\lambda_{mutant} - \lambda_{KL}$ (nm)	v_{KL}-v_{mutant} (cm^{-1})
KL-T51V	533	25	923
KL-T53C	513	5	192
KL-Y19W	513	5	192
KL-Q4W	533	25	923
Q108K:K40L:R58W (KLW), λ_{max} = 519 nm			
CRBPII Mutant	λ_{max} (nm)	$\lambda_{mutant} - \lambda_{KLW}$ (nm)	v_{KLW}-v_{mutant} (cm^{-1})
KLW-T51V	565	46	1569
KLW-T53C	540	21	749
KLW-Y19W	538	19	680
KLW-Q4W	550	31	1086

As shown in **Table III-12**, comparison of the protein shifts for mutants that belong to either the Q108K:K40L or Q108K:K40L:R58W series indicates that the same mutation made on Q108K:K40L:R58W family lead to a greater red shift than that for Q108K:K40L family of mutant. For example, addition of one more mutation, T51V, to Q108K:K40L results in 923 cm^{-1} protein shift, while in the presence of R58W, the same mutation T51V leads to 1569 cm^{-1} red shift. The same phenomenon is observed with mutants that contain Y19W and T53C. This increase of sensitivity could be attributed to the increased hydrophobicity of the binding pocket in the presence of R58W, by encapsulating the chromophore more effectively from the aqueous solution.

The enhanced effect of mutations T51V, T53C and Y19W is not due to direct interactions with R58W. As shown in **Figure III-37**, R58W is far away from mutations T51V, T53C and Y19W, but still has a large impact on the three

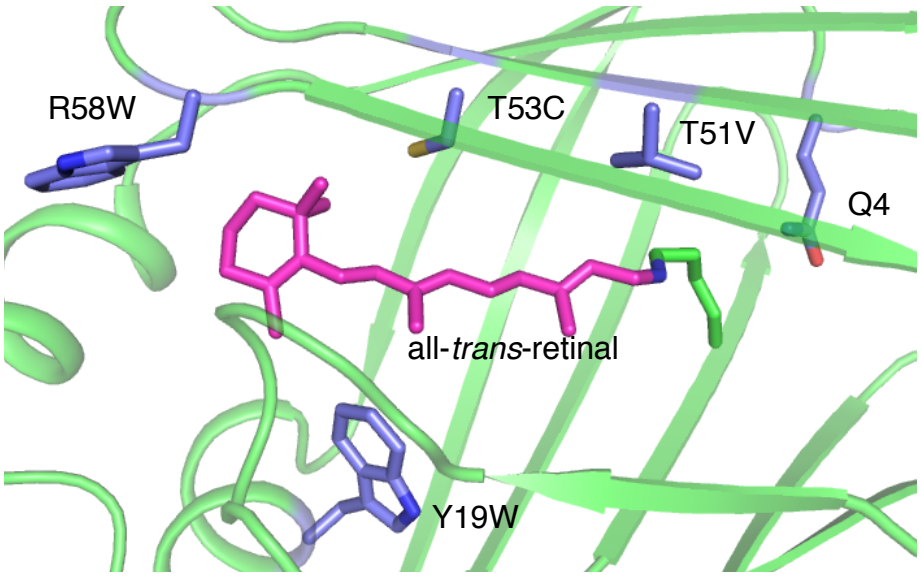


Figure III-37: Crystal structure of CRBPII mutant Q108K:K40L: R58W:T51V:T53C:Y19W:T29L.

mutations, which are spatially spread out in the binding cavity.

An interesting trend for the enhancement of red shift is observed. T53C and Y19W, which are closer toward the mouth of the binding pocket, have a ~4 fold increase in red shift, while for T51V, only a ~2 fold increase of protein shift is observed when R58W is present. It seems that the enhanced effect decreases as the residues are more embedded inside the binding pocket. It argues that covering up the binding pocket more as a result of the R58W mutation changes the dielectric environment close to the opening of the binding pocket more, and lowers the dielectric constant in that region to a larger extent. Therefore, polarity change by mutation of residues in the middle of the polyene are enhanced to a greater extent. T51V is situated deeper inside the binding pocket, thus the dielectric environment is not changed as much compared to T53C and Y19W, through introduction of R58W. For Q4W mutation, as it is situated close to the other end of the binding pocket and very far away from R58W, supposedly the dielectric environment does not vary much with introduction of R58W. Therefore, smaller increases in red shift (1.1 fold) is observed with introduction of R58W for Q4W.

Comparison of the Q108K:K40L:R58W, Q108K:K40L:R58F and Q108K:K40L:R58Y series of mutants shows that in the presence of R58W, the chromophore is most sensitive to polarity change, compared to Q108K:K40L:R58F and Q108K:K40L:R58Y (**Table III-13**). However, introduction of R58Y to Q108K:K40L leads to the largest amount of red shift compared to the

other R58 mutants. Q108K:K40L:R58Y absorbs at 535 nm compared to 519 nm with Q108K:K40L:R58W. And surprisingly, R58Y increases the pK_a of the retinal-PSB by one to two units, although it is far away from the PSB. It is interesting to note that R58Y results in a larger red shift compared to R58F and R58W when introduced to Q108K:K40L, but least sensitive to induce polarity change with T51V, T53C and Y19W.

Table III-13: Comparison of Q108K:K40L:R58, Q108K:K40L:R58W, Q108K:K40L: R58F, Q108K:K40L:R58Y series of mutants.

CRBPII Mutant	λ_{max} (nm)	$\lambda_{mutant} - \lambda_{ref}$ (nm)	$\nu_{ref}-\nu_{mutant}$ (cm^{-1})
Q108K:K40L (KL), λ_{max} = 508 nm			
KL-T51V	533	25	923
KL-T53C	513	5	192
KL-Y19W	513	5	192
Q108K:K40L:R58W (KLW), λ_{max} = 519 nm			
KLW-T51V	565	46	1569
KLW-T53C	540	21	749
KLW-Y19W	538	19	680
Q108K:K40L:R58F (KLF), λ_{max} = 524 nm			
KLF-T51V	561	37	1259
KLF-T53C	537	13	462
KLF-Y19W	537	13	462
Q108K:K40L:R58Y (KLY), λ_{max} = 535 nm			
KLY-T51V	563	28	929
KLY-T53C	540	5	173
KLY-Y19W	n.d.	n.d.	n.d.

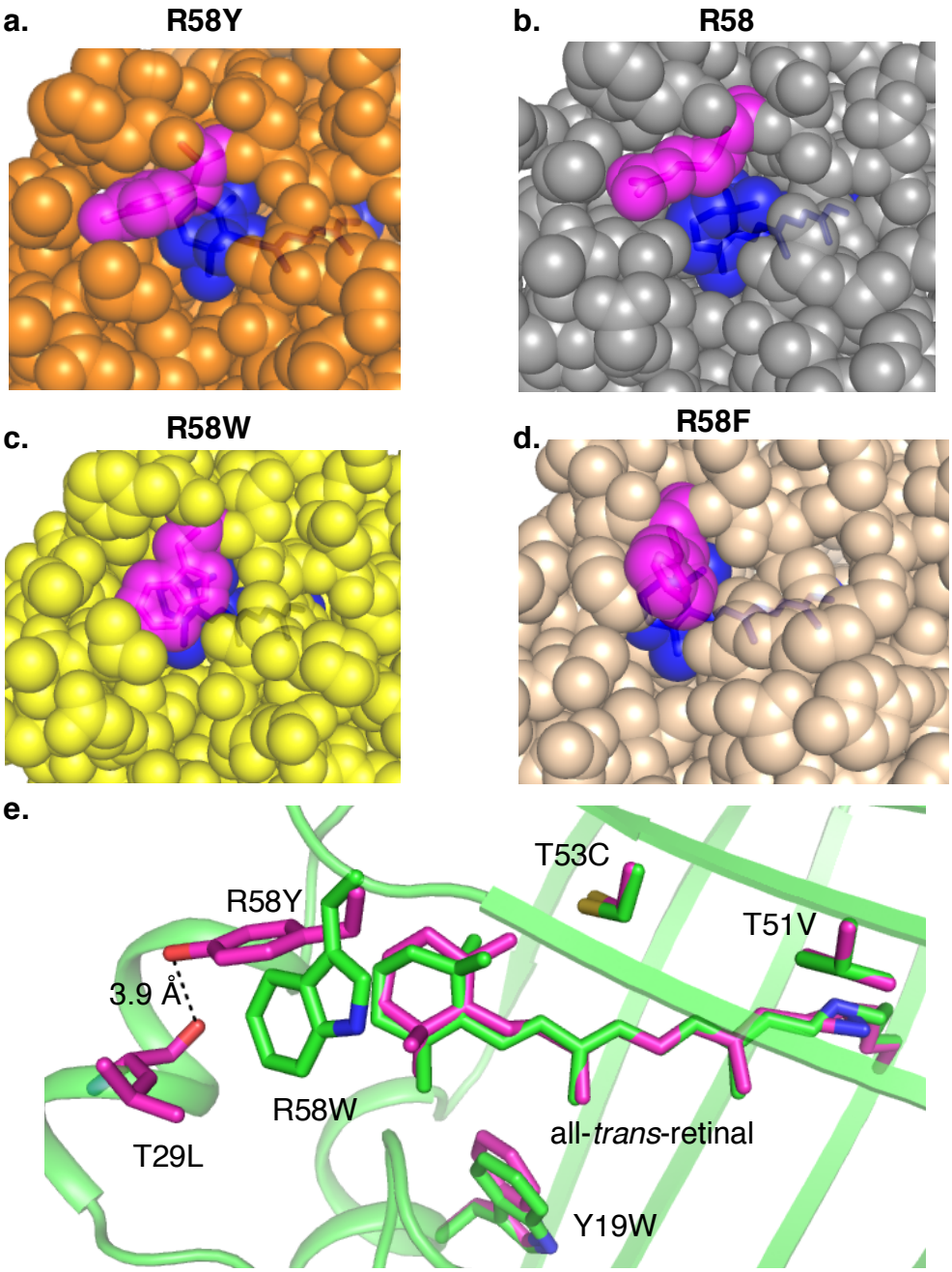


Figure III-38: Comparison of different R58 mutants. (a) Q108K:K40L:T51V:T53C:**R58Y**:T29L:Y19W:Q4H, (b) Q108K:K40L:**R58**, (c) Q108K:K40L:T51V:T53C:**R58W**:T29L:Y19W:Q4H and (d) Q108K:K40L:**R58F** with retinal in blue and 58 residue in magenta. (e) Overlaid crystal structures of CRBPII mutants Q108K:K40L:R58W:T51V:T53C:Y19W:T29L:Q4H (green) and Q108K:K40L:R58Y:T51V:T53C:Y19W:T29L:Q4H (magenta).

Q108K:K40L:R58Y was initially made to induce a red shift, considering that R58Y can increase the negative polarity projected on the ionone ring, and also could cover the binding pocket well. According to the model of Q108K:K40L:R58Y in Pymol, R58Y could adopt a conformation which projects the hydroxyl group toward the double bond of the ionone ring to interact the delocalized positive charge. However, all the crystal structures of CRBPII mutants containing R58Y show only one conformation, unlike R58F and R58W, which exhibit multiple conformations in different crystal structures. R58Y points away from the retinal as shown in **Figure III-38e**, presumably due to better solvation of the hydroxyl group of R58Y. Consequently, R58Y adopts similar conformation as R58.

As shown in **Figure III-38**, comparison of the space filling models of crystal structures of CRBPII mutants containing R58Y with R58, R58F, R58W shows that R58Y similar to R58, leaves a hole accessible to the binding pocket. Therefore, mutations of T51V and T53C contribute as much red shift in the presence of R58Y as that of R58 (**Table III-13**). Different from R58Y, R58F could also take multiple conformations similar to R58W. The space filling models of the crystal structure of Q108K:K40L:R58F shows that R58F can cover up the hole, but not as well as R58W, because it is smaller. Therefore, R58F could enhance the sensitivity of the chromophore to polarity change, but not as efficiently as R58W, as shown in **Table III-13**.

Although Q108K:K40L:R58W has the lowest absorption among the three aromatic mutations (R58W, R58F and R58Y), the observed enhancement in red shift for mutants in the Q108K:K40L:R58W family are the largest. Overlaid crystal structures of mutants containing R58Y and R58W (**Figure III-38e**) shows that the backbone of the protein and most of the side chains barely change between these two mutants. Even the chromophores adopt the same trajectory for the two mutants. The most significant difference lies in position 58. In the presence of R58W, the pocket is more sequestered from the aqueous environment than R58Y. Another difference between the two crystal structures of R58W and R58Y is that the retinal in the crystal structure of R58Y adopts the 6-s-*cis* conformation, while in the case of R58W, retinal adopts the 6-s-*trans* conformation. It is not clear yet how R58Y affects the conformation of retinal with regard to the rotation along C6-C7.

R58Y containing mutants are usually 1 pK_a unit higher than R58W containing mutants. The increase of pK_a by introduction of R58Y is not due to an allosteric effect, as overlaid crystal structures of Q108K:K40L:R58W:T51V:T53C:Y19W:T29L:Q4H and Q108K:K40L:R58Y:T51V:T53C:Y19W:T29L:Q4H show that the residues close to the PSB region do not change position. It is usually thought that only counteranion and residues close to the PSB region can affect the stability of the PSB, either through hydrogen bonding or electrostatic interactions with the iminium. To the best of our knowledge, it has never been

reported before that a mutation in the ionone ring region could alter the pK_a of retinal-PSB.

This increase of pK_a could be due to the increase of negative polarity in the ionone region, contributed from the polarized water molecules in the ionone ring region in R58Y containing mutants. Stabilization of the resonance structures of retinal-PSB with delocalized positive charge along the polyene can drive the equilibrium toward the protonated state of PSB, and therefore increase the pK_a as shown in **Figure III-39**.

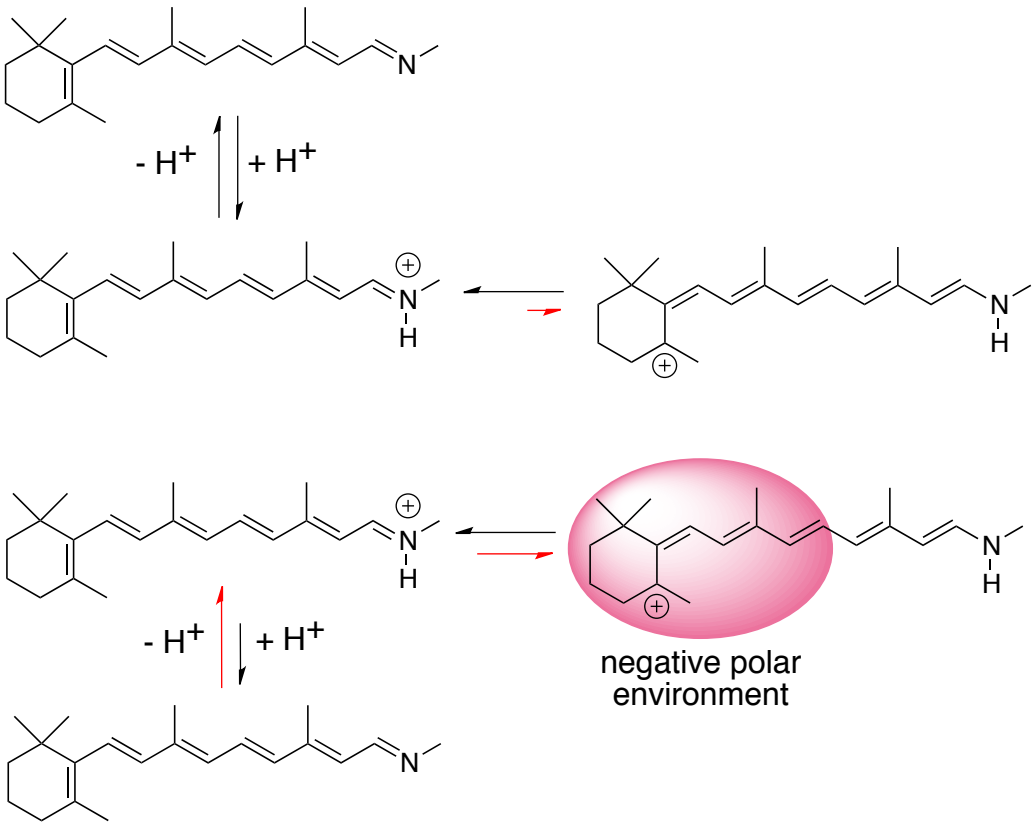


Figure III-39: Proposed mechanism for the increase of pK_a with introduction of R58Y, through stabilization of the positive charge in the β-ionone ring region.

III.5 Additive effects observed for mutations T51V, T53C, Q4W and Y19W

III.5.1 Additive effects of red shift caused by T51V, T53C and Q4W

Table III-14: Additive effects of mutants T51V, T53C and Q4W in the absence and presence of R58W.

Q108K:K40L (KL), λ_{max} = 508 nm				
CRBPII Mutant	λ_{max} (nm)	λ_{mutant} - λ_{KL} (nm)	ν_{KL}-ν_{mutant} (cm^{-1})	Additive effect (cm^{-1})
KL-T51V	533	25	923	1115
KL-T53C	513	5	192	
KL-T51V-T53C	539	31	1132	1132
KL-T51V	533	25	923	
KL-T53C	513	5	192	2038
KL-Q4W	533	25	923	
KL-T51V-T53C-Q4W	574	66	2263	2263
Q108K:K40L:R58W (KLW), λ_{max} = 519 nm				
CRBPII Mutant	λ_{max} (nm)	λ_{mutant} - λ_{KLW} (nm)	ν_{KLW}-ν_{mutant} (cm^{-1})	Additive effect (cm^{-1})
KLW-T51V	565	46	1569	2318
KLW-T53C	540	21	749	
KLW-T51V-T53C	585	66	2173	2173
KLW-T51V	565	46	1569	
KLW-T53C	540	21	749	3404
KLW-Q4W	550	31	1086	
KLW-T51V-T53C-Q4W	613	94	2955	2955

According to our hypothesis, the overall electrostatic potential projected on the Van der Waals surface of the chromophore is the major factor in determining the absorption profiles of the retinal-PSB in CRBPII. The overall electrostatic potential is a sum of all the electric field from all points. The polarity change caused by each point mutation should be additive if the two points are not interfering with each other. Therefore, the red shift caused by each point mutation should be additive. The synergistic additive effect had been illustrated before in color rhodopsin mutants as well.[20]

It is interesting to find that the effects of some of these individual mutants are indeed additive. For example, as shown in **Table III-14**, the protein shift caused by T51V and T53C is 923 cm^{-1} and 192 cm^{-1}, respectively, when introduced to the Q108K:K40L-double mutant. The protein shift of Q108K:K40L: T51V:T53C-tetra mutant is 1132 cm^{-1}, which is similar to the theoretical value of simply adding up the red shift caused by T51V (923 cm^{-1}) and T53C (192 cm^{-1}) individually. Similarly, the protein shift caused by T51V and T53C is 1569 cm^{-1} and 749 cm^{-1}, respectively, when introduced to Q108K:K40L:R58W. The protein shift induced by mutations of T51V and T53C together in addition to Q108K:K40L:R58W is 2174 cm^{-1}, also close to the theoretical value of simply adding up 1569 cm^{-1} and 749 cm^{-1}.

Likewise, introduction of mutations T51V, T53C and Q4W at the same time to Q108K:K40L:R58W results in a 2959 cm^{-1} protein shift, comparable to the value of 3404 cm^{-1} by addition of the individual protein shifts from each mutation. In this way, a dramatically red shifted mutant, Q108K:K40L:R58W: T51V:T53C:Q4W, with absorption maximum of 613 nm, is achieved. The absorption of this mutant surpasses all the reported rhodopsin absorption maxima and even reaches the value of retinal-PSB in vacuum, which was considered to be the ceiling for retinal-PSB. Supposedly, the positive charge of retinal-PSB in vacuum is the least stabilized, due to complete separation from the counteranion. Therefore it should have the largest degree of charge delocalization and red shift. However, we have shown that through appropriate protein-chromophore interactions, it is possible to surpass this limit. We have also shown that it is not the counteranion that is dictating the absorption maxima, but the binding pocket.

The red shift of the Q108K:K40L:R58W-triple mutant is not strictly additive with the red shift of Q108K:K40L:T51V or Q108K:K40L:T53C. The protein shift induced by introduction of R58W:T51V or R58W:T53C to Q108K:K40L is much larger than simply adding up the red shift caused by each individual mutant. This agrees with our hypothesis that the effect of R58W is not electrostatic in nature. It leads to an increased hydrophobicity of the pocket and renders the chromophore more sensitive to changes in the polarity of its environment.

III.5.2 Partially additive effects of red shift caused by Y19W

Y19W results in a 5 nm red shift when added to Q108K:K40L, which increases to 19 nm in the presence of R58W. The red shift of Y19W is considered to be a result of either introduction of electron-rich polarizable tryptophan residue, which could encourage the delocalization of the positive charge along the polyene, or changing of the chromophore's trajectory slightly by pushing it toward the more electron negative side of the pocket, where Q38 and Q128 reside (**Figure III-13**).

Table III-15: Additive effect of Y19W with either T51V or T53C.

Q108K:K40L (KL), λ_{max} = 508 nm				
CRBPII Mutant	λ_{max} (nm)	$\lambda_{mutant} - \lambda_{KL}$ (nm)	$\nu_{KL}-\nu_{mutant}$ (cm^{-1})	Additive effect (cm^{-1})
KL-T51V	533	25	923	1115
KL-Y19W	513	5	192	
KL-T51V-Y19W	537	31	1132	1132
Q108K:K40L:R58W (KLW), λ_{max} = 519 nm				
CRBPII Mutant	λ_{max} (nm)	$\lambda_{mutant} - \lambda_{KLW}$ (nm)	$\nu_{KLW}-\nu_{mutant}$ (cm^{-1})	Additive effect (cm^{-1})
KLW-T51V	565	46	1569	2249
KLW-Y19W	538	19	680	
KLW-T51V-Y19W	577	58	1936	1936
KLW-Y19W	538	19	680	1429
KLW-T53C	540	21	749	
KLW-T53C-Y19W	556	37	1282	1282

When Y19W is introduced together with T51V or T53C, a synergistic red shift is observed as shown in **Table III-15**, both in the presence and absence of R58W. This means that the red shift caused by these three residues does not interfere with each other.

However, Y19W does not contribute to the red shift when Q4 mutants are present (**Table III-16**). This indicates that one of the changes induced by the Q4W mutation is the same as that induced by Y19W, although the two positions are ~15 Å away from each other. The explanation for this will be discussed later after clarifying the role and function of Q4.

Table III-16: No additive effect for Q4W and Y19W.

CRBPII mutant	λ_{max} (nm)
Q108K:K40L:Q4W	533
Q108K:K40L:Q4W:Y19W	533
Q108K:K40L:R58W:Q4W	550
Q108K:K40L:R58W:Q4W:Y19W	553

III.6 Detailed studies on Gln4

III.6.1 Red shift and a slight drop in pK_a as a result of the removal of Gln4

The crystal structures of CRBPII mutants show that Gln4 stabilizes the protonated state of the PSB through hydrogen-bonding interactions mediated by a water molecule (**Figure III-40**). Removal of Gln4 by Q4W mutation results in a

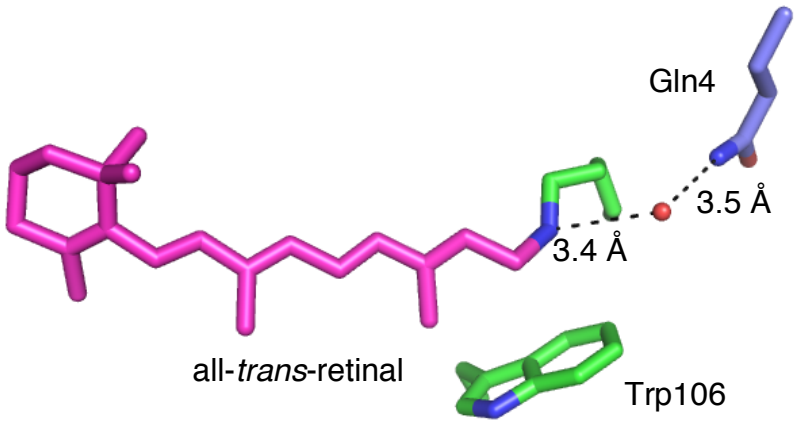


Figure III-40: Crystal structure of Q108K:K40L:T51V:T53C:R58W:T29L:Y19W bound with retinal, with Q4 and W106 highlighted.

dramatic red shift, possibly through disruption of the water mediated hydrogen bonding network. The importance of Gln4 for the absorptions of the retinal-PSB and the pK_a of the PSB is further investigated here.

To fully explore the role of Gln4, different amino acids were introduced to hepta-mutant Q108K:K40L:R58W:T51V:T53C:T29L:Y19W, which has an absorption maximum of 591 nm when bound to retinal. Crystal structures of Gln4 mutants aid in the study to reveal cause-effect relationships of Gln4 mutants, together with the absorptions of the retinal-PSB.

As shown in **Table III-17**, replacement of Gln4 with hydrophobic amino acids, such as Ala, Leu, Phe and Trp, in Q108K:K40L:T51V:T53C:R58W:T29L:Y19W-hepta mutant, all lead to the same extent of red shift to ~613 nm. This is likely due to the disruption of the hydrogen bonding interactions of Gln4 with the network of water molecules, which as a result destabilizes the interaction of the water molecule with the PSB. Less stabilization of the positive charge on

the iminium encourages further delocalization of the positive charge along the polyene, leading to a red shift.

Table III-17: Gln4 mutagenesis studies.

CRBPII Mutant	λ_{max} (nm)	$\lambda_{mut} - \lambda_{Q4}$ (nm)	pK_a	K_d (nM)
KLVCWLW-Q4	591	0	8.2	38±10
KLVCWLW-Q4F	613	22	7.5	58±12
KLVCWLW-Q4W	613	22	7.7	103±10
KLVCWLW-Q4L	613	22	7.9	57±8
KLVCWLW-Q4A	613	22	7.0	150±12
KLVCWLW-Q4N	612	21	7.2	65±12
KLVCWLW-Q4T	610	19	7.8	63±8
KLVCWLW-Q4D	613	22	n.d.[a]	165±9
KLVCWLW-Q4K	616	25	7.2	12±8
KLVCWLW-Q4R	622	31	6.5	183±11
KLVCWLW-Q4E	590	-1	n.d.[a]	162±20
KLVCWLW-Q4H	585	-6	7.9	18±5

KLVCWLW is abbreviated for hepta-mutant Q108K:K40L:T51V:T53C:R58W:T29L:Y19W. ([a] CRBPII mutants containing Q4D and Q4E were not stable and pK_a could not be determined, as the protein denatured upon acidification.)

Since we hypothesize that Gln4 stabilizes the PSB through the water-mediated hydrogen bonding interaction, shortening the carbon chain of Q4 by one carbon to Q4N or Q4D can disrupt the hydrogen bonding interactions. Indeed, this leads to the same extent of red shift as mutations of Gln4 into

hydrophobic residues in Q108K:K40L:T51V:T53C:R58W:T29L:Y19W-hepta mutant (22 nm, **Table III-17**). Since there is no difference in the protein shift caused by Q4N and Q4D, it suggests that Asp4 remains neutral in the hydrophobic binding pocket. The same phenomenon is observed with Glu4.

Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:**Q4E** has the same absorption maximum as Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:**Q4**, because Q4E can possibly maintain the same kind of hydrogen bonding interactions with the water molecule to stabilize the protonated Schiff base. Also, Q4E does not cause any blue shift compared to Q4 when introduced in the Q108K:K40L:T51V:T53C:R58W:T29L:Y19W-hepta mutant. This indicates that Q4E probably stays neutral in the hydrophobic binding pocket and therefore has similar polarity to Gln4.

Mutations of Gln4 into basic residues, Lys4 and Arg4, in Q108K:K40L: T51V:T53C:R58W:T29L:Y19W-hepta mutant result in more red shift as compared to the rest of the Gln4 mutants (**Table III-17**). This is because introduction of basic residues not only disrupts the water-mediated hydrogen bonding interactions, but also destabilizes the positive charge on the iminium through charge repulsion.

In this way, the most red-shifted mutant, Q108K:K40L:T51V:T53C:R58W: T29L:Y19W:Q4R, absorbing at 622 nm is obtained, with the lowest pK_a of 6.5 in the series of Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4 mutants. Crystal structures of Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4 and Q108K:K40L:

T51V:T53C:R58W:T29L:Y19W:Q4R have been resolved (Rafida Nossoni, Jim Geiger's lab, MSU) and allow comparison to unravel differences caused by Q4R mutation.

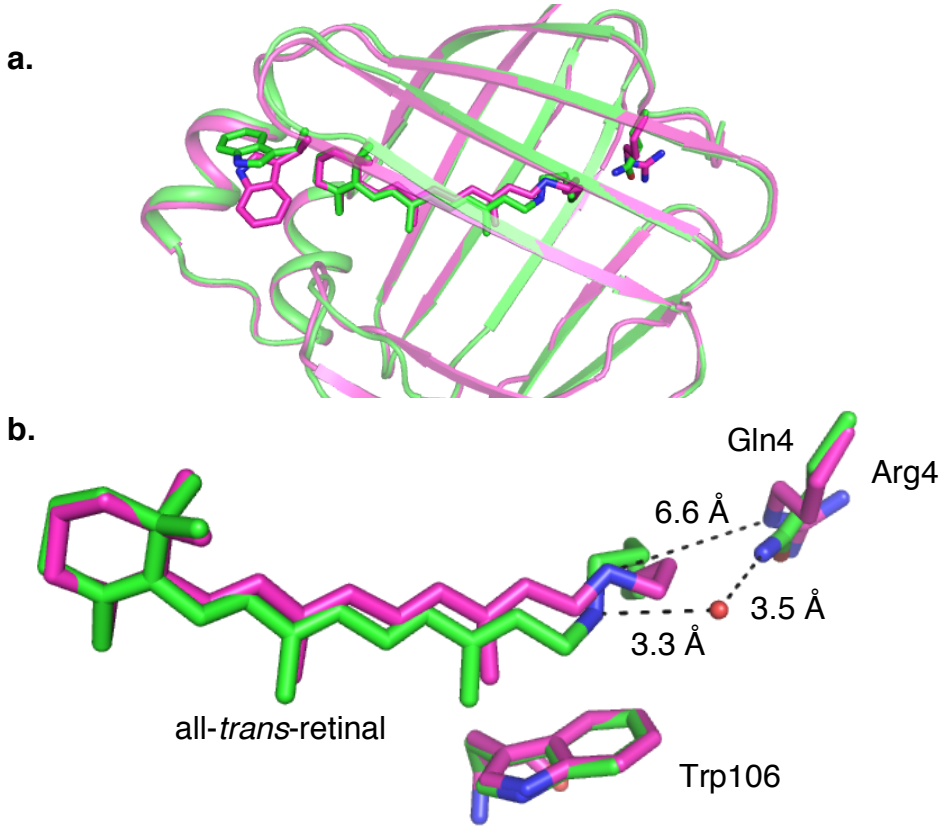


Figure III-41: Comparison of Gln4 and Arg4 mutants. (a) Cartoon and (b) zoom in view of crystal structure overlay of Q108K:K40L:T51V:T53C: R58W:T29L:Y19W (green) and Q108K:K40L:T51V:T53C:R58W:T29L:Y19W: Q4R (magenta) bound with retinal.

As shown in **Figure III-41**, the overlaid crystal structures of Gln4 and Arg4 show that the backbone of the two proteins and the side chains do not move significantly, with RMS = 0.415 Å for all the 133 residues. However, a significant conformational change of the imine is clearly evident. The imine changes its conformation from the *cis* to *trans*-isomer. Replacement of Q4 with arginine leads to the disruption of the water molecule mediated hydrogen bonding

interaction and as a result a highly ordered water molecule is not observed for the Q4R containing mutant.

Trans-imine is more stable than *cis*-imine by one to two kcal/mol. However, *cis*-imine can be stabilized by the hydrogen bond interactions with Gln4, via a water molecule. Moreover, the *cis*-imine can project the iminium hydrogen to the π electron cloud of Trp106, which could also contribute to the stability of *cis*-imine through π-cation interaction. These two interactions compensate for the steric hindrance of *cis*-imine and make *cis*-imine to be the more stable isomer in the presence of Gln4. In short, the water mediated hydrogen bonding interaction locks the conformation of the imine as *cis*. This is why in most of the crystal structures of CRBPII mutants containing Gln4, *cis*-imine is observed, along with the water molecule mediating Gln4 and the *cis*-imine. Removal of the water molecule by Gln4 mutations destabilizes the *cis*-imine conformation of the PSB and a *trans*-imine is observed instead, leading to a decreased pK_a, due to less stabilization.

Disturbance of the hydrogen-bonding network leads to a decrease of pK_a by ~1 unit in Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4 series of mutants as shown in **Table III-17**. This shows that Gln4 does play a role in stabilizing the pK_a of the PSB, however, not to a large extent. This can be explained by the weak hydrogen bonding interactions between the water molecule and Gln4. The density for the water molecule is found in most of the cases where the Gln4

residue is maintained, however, the distance between the water molecule and Gln4 is around 3.5 Å, indicating a weak hydrogen bonding interaction. Removal of this hydrogen bond does not abolish the protonated state of the retinal-PSB. This indicates that the retinal-PSB is stabilized by other interactions as well, such as the π-cation interaction between Trp106 and other polar residues along the polyene, through stabilizing the resonance structures of retinal-PSB.

A pK_a change is observed for Gln4 mutants after the PSB is formed. After retinal is added to the protein solution, the PSB and SB are formed slowly and PSB formation reaches a maximum after some time. Then the absorption for PSB starts decreasing slowly and at the same time SB absorption starts increasing. Acidification of the solution can convert the SB absorption to PSB absorption. This indicates that the PSB is converted to SB due to the pK_a change, as the pH of the solution remains the same.

The change of pK_a indicates conformational change of the protein or the chromophore. One hypothesis for this phenomenon is that the *cis*-imine is the kinetic product of retinal-PSB, which has a higher pK_a, while the *trans*-imine is the thermodynamic product with a lower pK_a for Gln4 mutants, due to perturbation of the water mediated hydrogen bonding interaction. As the isomerization from *cis* to *trans* is slower than the formation of PSB, the kinetic product of *cis*-imine forms first with higher pK_a, but gradually isomerizes to *trans*-

imine with lower pK_a and reaches equilibrium. (It was observed that change in pK_a accelerates at higher temperature, with light exposure, or in the presence of high concentration of imidazole.)

Mutagenesis studies of Gln4 shows the importance of the water-mediated hydrogen bonding interaction with the PSB to stabilize the positive charge on the iminium. Removal of this stabilization results in positive charge delocalization along the polyene, leading to a red shift. Although the latter hydrogen bonding interactions are weak, removal of Gln4 still has a significant impact on the wavelength regulation.

Now let's revisit the reason why Q4W cancels the contribution of Y19W in wavelength regulation, as discussed before in section **III.5.2**. After learning about the role of Gln4 mutations in wavelength regulation, this might suggest that both mutations are leading to the same conformational change of the chromophore. The ionone ring of retinal is translated by ~1.4 Å away from Y19W when Y19W is introduced, to avoid steric clash into Y19W. As a result, the ionone ring is situated in a position with more negative electron density from Q38 and Q128. Since Gln4 mutation changes the conformation of imine from *cis* to *trans*, this could also induce a small amount of translation of the chromophore in the same way as Y19W. Therefore, in the presence of the Gln4 mutation, the red shift caused by Y19W is canceled out. This might be why the red shift caused by Y19W and Q4W are not additive. However, the crystal structures of Gln4 mutants without Y19W are not available to support this hypothesis.

III.6.2 Blue shift is induced by the Q4H mutation.

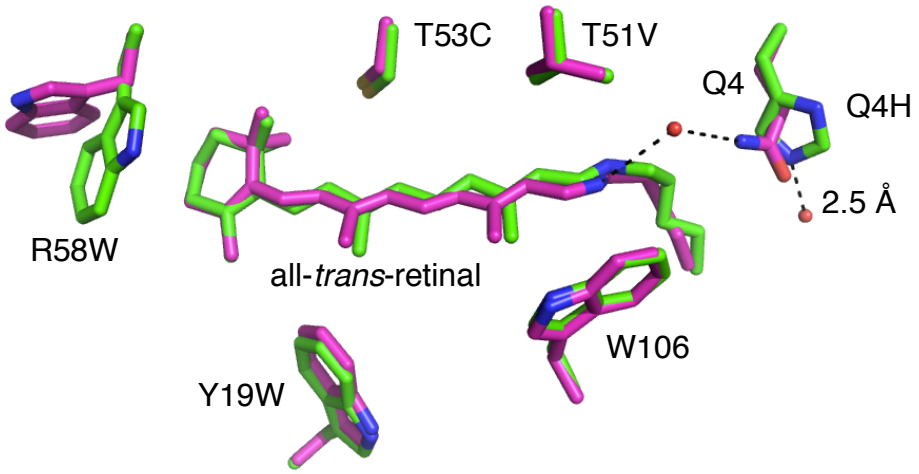


Figure III-42: Overlaid crystal structures of Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4 (magenta) and Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4H (green).

The only Q4 mutation in the Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4 series that leads to a blue shift is Q4H, from 591 nm to 585 nm. Histidine could provide the side chain to maintain the same kind of hydrogen bonding interactions with the water molecule network as seen with Q4. Unexpectedly, the crystal structure of mutant Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4H reveals that Q4H actually points away from the Schiff base. As a result, it can not form hydrogen bonding interactions with the protonated Schiff base via a water molecule.

III.7 Toward the most red-shifted CRBPII mutant by addition of A33W

As shown before, introduction of R58W enhances the red shift caused by mutations T51V, T53C and Y19W dramatically through better covering up the

binding pocket and making the chromophore more sensitive to polarity change. Attempts at trying to close the binding pocket even more efficiently through introduction of a large aromatic residue in the mouth of the pocket were carried out through mutations of F57W and I25F. However, these two mutations do not result in further red shift; presumably F57W flips out of the pocket in the presence of R58W due to steric clash and does not have an effect in spectral tuning (**Table III-18**). I25F was made in order to compliment R58W as it is situated next to R58W. But the space filling model of the crystal structure shows that I25 is already buried by the protein residues, therefore it can not cover up the pocket any more.

Table III-18: Summary data for F57W and I25F mutants.

CRBPII Mutant	λ_{max} (nm)	Protein Shift of mutation
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4L:**F57**	613	0
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4L:**F57W**	609	-4
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:**I25**	591	0
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:**I25F**	589	-2

Ala33, the final amino acid to be probed for this purpose, resides on one of the two α-helices, which cover up the binding pocket. Modeling of A33W in Pymol shows that A33W clashes with either R58W or the chromophore, due to the significant steric demand of tryptophan compared to alanine (**Figure III-43b**). Therefore, this mutation was not considered until a similar mutation in the

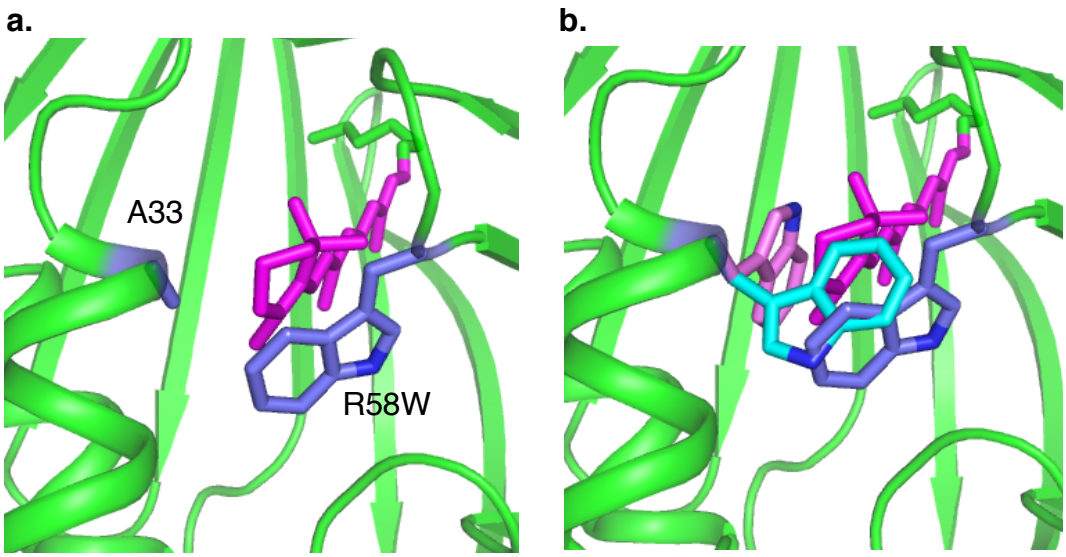


Figure III-43: Modeling of A33W. **a.** Cartoon of crystal structure of Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4H, with residues A33 and R58W highlighted. **b.** In silico mutageneis of A33W showing that it clashes into either R58W or retinal.

reengineered CRABPII, with a deeper chromophore binding pocket (Tanya Berbasova's work, Babak Borhan's lab, MSU), was introduced and exhibited a dramatic red shift.

Introduction of A33W in the most red shifted mutant known, Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4R, results in an even further red shifted mutant from 622 nm to 644 nm, the most red shifted retinal-PSB so far obtained. In order to understand the role of A33W in causing this red shift, different mutations of A33W were prepared. It was found that A33W does not result in a red shift in all the cases.

As shown in **Table III-19**, A33W results in different degrees of blue shift and red shift depending on which CRBPII mutants it is added to. A general trend can be found for determining whether red shift or blue shift results. The

observation that A33W results in red shift when it is introduced to Q108K:K40L:T51V:T53C:R58W:T29L:Y19W and Q108K:K40L:T51V:T53C:R58W:T29L:Q4F, but blue shift to Q108K:K40L:T51V:T53C:R58W:T29L, indicates that the presence of Y19W or Q4F is important for A33W to cause a red shift. Interestingly, this agrees with the previous result that Q4 mutation leads to the same effect as Y19W, since Q4W mutation cancels out the red shift caused by Y19W, possibly due to the same geometry change of the chromophore as discussed before.

Table III-19: Protein shift caused by A33W.

CRBPII mutants	λ_{max} (nm) with **A33**	λ_{max} (nm) with **A33W**	Protein shift of A33W (nm)
Q108K:K40L (KL)	508	498	-10
KL:Q4W	533	513	-20
KL:T51V:T53C	539	533	-6
KL:T51V:Y19W	537	522	-15
KL:T51V:T53C:R58W:T29L	585	566	-19
KL:R58F:Y19W	537	543	6
KL:T51V:T53C:R58W:T29L:Y19W	591	606	15
KL:T51V:T53C:R58W:T29L:Q4F	613	629	16
KL:T51V:T53C:R58W:T29L:Y19W:Q4F	613	636	23
KL:T51V:T53C:R58W:T29L:Y19W:Q4R	622	644	22

R58 mutation to aromatic residues is also necessary for A33W to induce red shift. This is evident from the comparison of KL:Q4W, KL:T51V:Y19W with

the last five entries of **Table III-19**. Without R58W (R58F), even if Q4W or Y19W are present, A33W does not cause red shift. It requires the complementation of A33W and R58W to better seal the pocket.

To understand the role of A33W in regulation of the absorption maxima for retinal-PSB, different A33 mutations were prepared. As shown in the last 6 entries of **Table III-20**, the largest amount of red shift results from tryptophan, followed by tyrosine and phenylalanine. This supports our hypothesis that A33W could most effectively close the cavity in complementation with R58W, and consequently result in a larger red shift.

Electrostatic interactions also play a role in the A33 mutagenesis studies as expected. A33E is 10 nm more red shifted than A33L, because the negative polarity from A33E could encourage the positive charge delocalization, leading to red shift, as compared to A33L. A33Y results in 4 nm more red shift than A33F, possibly due to the hydroxyl group of tyrosine, which could project negative polarity on the ionone ring region, inducing more positive charge delocalization.

A number of R58 mutants were prepared in the presence of the A33W mutation to ascertain whether R58W is necessary to maintain the observed red shifts. As shown in the first seven entries of **Table III-20**, large aromatic amino acids R58W, R58F and R58Y result in more red shift than the rest of the amino acids. R58W and A33W together generate the most red shifted mutant in Table III-20, indicating A33W and R58W work complimentarily to cover up the cavity.

Table III-20: Different combination of A33X and R58X mutants.

CRBPII Mutants	λ_{max} (nm)	pK_a
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:**R58W**:A33W	636	7.8
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:**R58F**:A33W	616	7.7
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:**R58Y**:A33W	600	8.1
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:**R58L**:A33W	583	6.8
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:**R58H**:A33W	573	6.1
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:**R58E**:A33W	586	7.0
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:**R58A**:A33W	576-590	6.7
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:R58W:**A33W**	636	7.8
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:R58W:**A33F**	616	7.6
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:R58W:**A33Y**	620	7.9
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:R58W:**A33H**	614	7.7
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:R58W:**A33L**	605	7.4
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:R58W:**A33E**	615	7.6
Q108K:K40L:T51V:T53C:T29L:Y19W:Q4F:R58W:**A33**	613	7.6

As shown in **Table III-20**, changes in position 58 results in more dramatic changes of absorption maxima, for example changing from R58W to R58H leads to a 63 nm blue shift. This shows the significant role of R58W in regulating the wavelength of PSB in CRBPII mutants, not through change in polarity of position 58, but the dielectric environment of the protein.

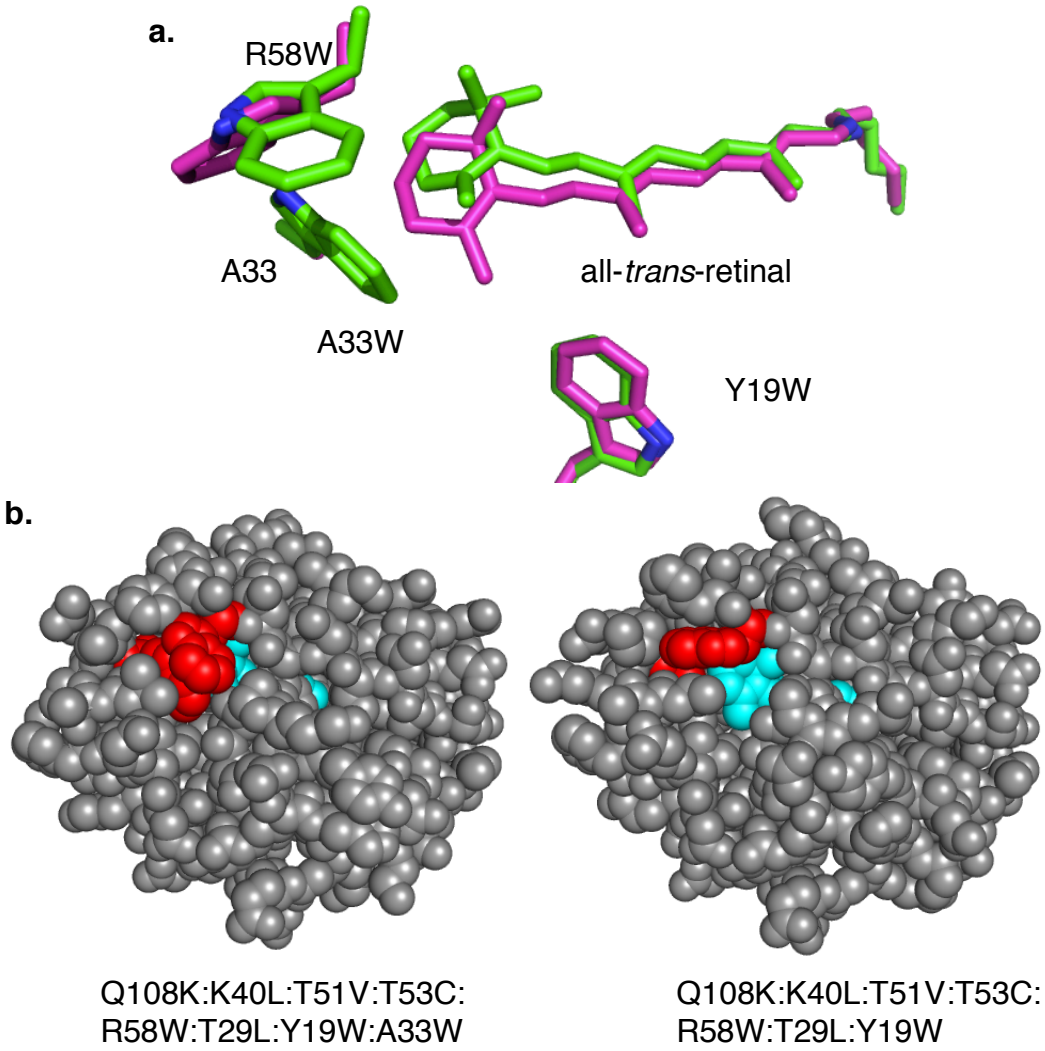


Figure III-44: Crystal structures of A33W. **a**. Crystal structure overlay of Q108K:K40L:T51V:T53C:R58W:T29L:Y19W (green) and Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:A33W (magenta) bound with retinal. **b**. Space filling models of Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:A33W and Q108K:K40L:T51V:T53C:R58W:T29L:Y19W bound with retinal (retinal shown in cyan color, R58W in red and A33W or A33 in red).

Fortunately, the crystal structure of Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:A33W was solved (Rafida Nossoni, Jim Geiger's lab, MSU) to support our hypothesis about the role of A33W. Overlaying the crystal structures of Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:A33W and Q108K:K40L:T51V:T53C:R58W:T29L:Y19W reveals that the overall structure of the protein does not

change with the introduction of A33W, not even the helix where A33W resides. RMS of the two structures is 0.398 Å.

As shown in **Figure III-44a**, introduction of the bulky residue tryptophan residue at position 33 locks the conformation of R58W (flipped out), in order to avoid steric clash with A33W. At the same time, the chromophore is translated slightly away from A33W due to the steric demand of A33W. With the introduction of A33W, the binding cavity is more embedded. As demonstrated in **Figure III-44b**, the chromophore is fully covered in the crystal structure of Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:A33W, while in Q108K:K40L: T51V:T53C:R58W:T29L:Y19W a small opening is observed, as a result of R58W rotameric possibilities. A33W mutagenesis studies demonstrates the importance of fully sequestrating the chromophore, and increasing the chromophore's sensitivity to polarity.

III.8 Dissecting the role of Q38 and Q128

Ordered water molecules are found in the hydrophobic binding pocket of CRBPII. As structured water molecules could project negative polarity on the chromophore and change its electronic characteristics, they could play an important role in the absorption and the pK_a of the chromophore.

A water network corresponding to Q38 and Q128 is found in most of the CRBPII mutants if the two residues are maintained. A detailed hydrogen bonding network is illustrated in **Figure III-45,** based on the crystal structure of

Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4H, as it is at 1.1 Å resolution and can display the density and location of water molecules accurately.

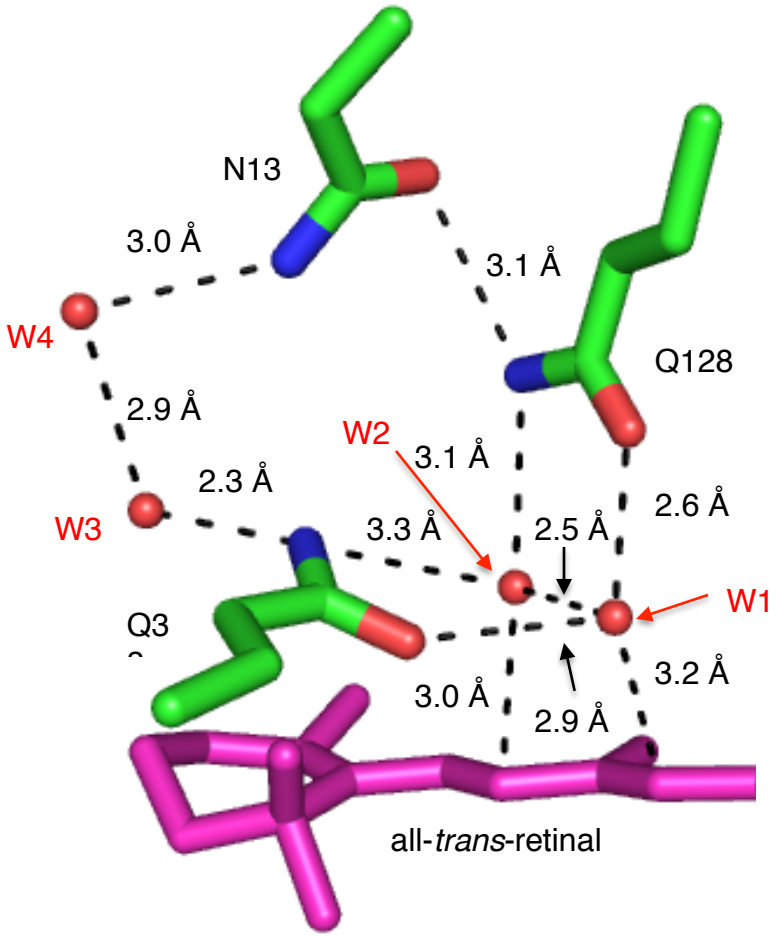


Figure III-45: Detailed hydrogen bonding network surrounding Q38, Q128 and N13 in the crystal structure of Q108K:K40L:T51V:T53C:R58W: T29L:Y19W:Q4H.

As shown in **Figure III-45**, two water molecules, W1 and W2, close to the middle of the polyene, are bound by two glutamine residues, Q38 and Q128. Q38 is close to the ionone ring and interacts with Q128 through one or two water molecules. These two glutamine residues are important for the red shift of the retinal-PSB formed in CRBPII. As shown in **Table III-21**, whenever hydrophobic

mutations are introduced at positions 38 and Q128, a large blue shift results, which is likely due to the perturbation of the hydrogen bonding network.

To show that the two water molecules are important for the red shift observed in most of CRBPII mutants, electrostatic calculations were performed with and without the two water molecules, which are hydrogen bonded to Q38 and Q128. As shown in **Figure III-46,** for the calculation with two water

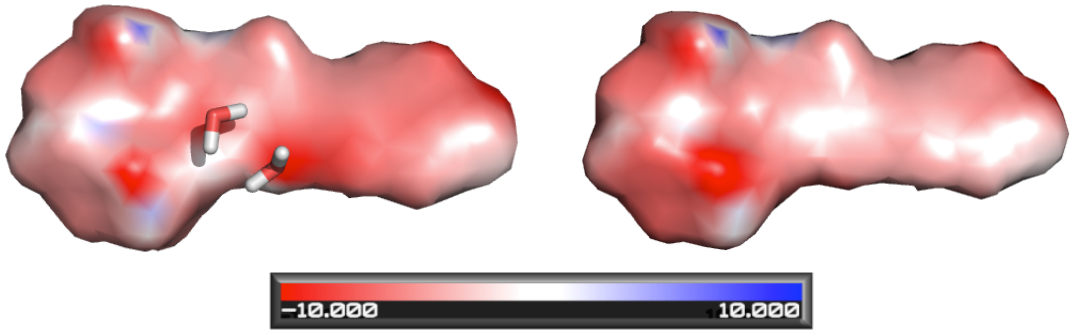


Figure III-46: Comparison of electrostatic calculations with and without water molecules. Electrostatic potential projected on the Van der Waals surface of the chromophore based on crystal structure of Q108K:K40L:T51V:T53C: R58W:T29L:Y19W:Q4R, with two water molecules (left) and without two water molecules (right). The unit for the scale is kT/e. Left end of the chromophore is the β-ionone ring region.

molecules, the middle of the polyene exhibits more negative electrostatic potential than the one without the two water molecules. This is because the two highly polarized water molecules are close to the middle of the polyene and project negative electrostatic potential on the chromophore, thus affecting the overall electrostatic potential projected on the chromophore. Due to the increase of negative potential in the ionone ring region, the migration of the positive

charge localized on the PSB along the polyene towards the ionone ring towards the ionone ring is favored, leading to the observed red shift.

Table III-21: Table of Q38 and Q128 mutants

CRBPII mutants	λ_{max} (nm)	pK_a	K_d (nm)
Q108K:K40L:R58W:T51V:T53C:T29L	585	7.9	29±4
Q108K:K40L:R58W:T51V:T53C:T29L:**Q128L**	532	5.8	32±11
Q108K:K40L:R58W:T51V:T53C:T29L:**Q38W**	538	7.8	15±7
Q108K:K40L:R58W:T51V:T53C:T29L:**Q38M**	513	7.5	19±5
Q108K:K40L:R58W:T51V:T53C:T29L:**Q128L:Q38M**	504	7.2	16±4
Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4F	613	7.5	58±12
Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4F:**Q38W**	577	5.9	n.d.
Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W:**Q38E**	590	6.0	196±19
Q108K:K40L:R58W:T51V:T53C:Q128E	~555	<6.5	n.d.
Q108K:K40L:R58W:T51V:T53C:Q128K	553	n.d.	n.d.

n.d. not determined.

Whenever hydrophobic residues are introduced at position 38 or 128, the tight hydrogen bonding interaction between the two water molecules, Q38 and Q128, will be disturbed. Some Q38 and Q128 mutations decrease the pK_a dramatically. As we have discussed earlier, stabilization of the resonance structure of retinal-PSB with delocalized positive charge could also stabilize the protonated state of imine, and lead to an increase in pK_a. Mutations of Q38 or Q128 support this hypothesis. Mutant Q108K:K40L:R58W:T51V:T53C:T29L:

Q128L and Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4F:**Q38W** decreased the pK_a by ~2 units as shown in **Table III-21**.

III.9 Overall electrostatic potential projected on the chromophore of a few mutants with crystal structures refined

Through combinations of different mutations, a full spectrum of CRBPII mutants were obtained, as shown in **Figure III-47**. This is the first time a chromophore has been regulated over such a wide range based on only one protein, CRBPII. The most red shifted mutant obtained surpasses the most red shifted retinal-PSB reported so far, and even surpasses the value of retinal-PSB in vacuum, which was considered to be the maximum for retinal-PSB. We have shown that through proper protein-chromophore interactions, it is possible to change a chromophore's absorption profiles, especially for chromphores that are highly polarizable.

With crystal structures for some of these mutants in hand, we were able to perform electrostatic calculations to evaluate the effect of the overall electrostatic potential projected on the chromophore on the absorption maxima. To illustrate the importance of dielectric environment on the electrostatic potential projected on the chromophore, two different dielectric constants were applied for the calculation of Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4R. As shown in **Figure III-48**, the chromophore exhibits much more intense negative potential on the middle and ionone ring region of the chromophore at lower dielectric constant

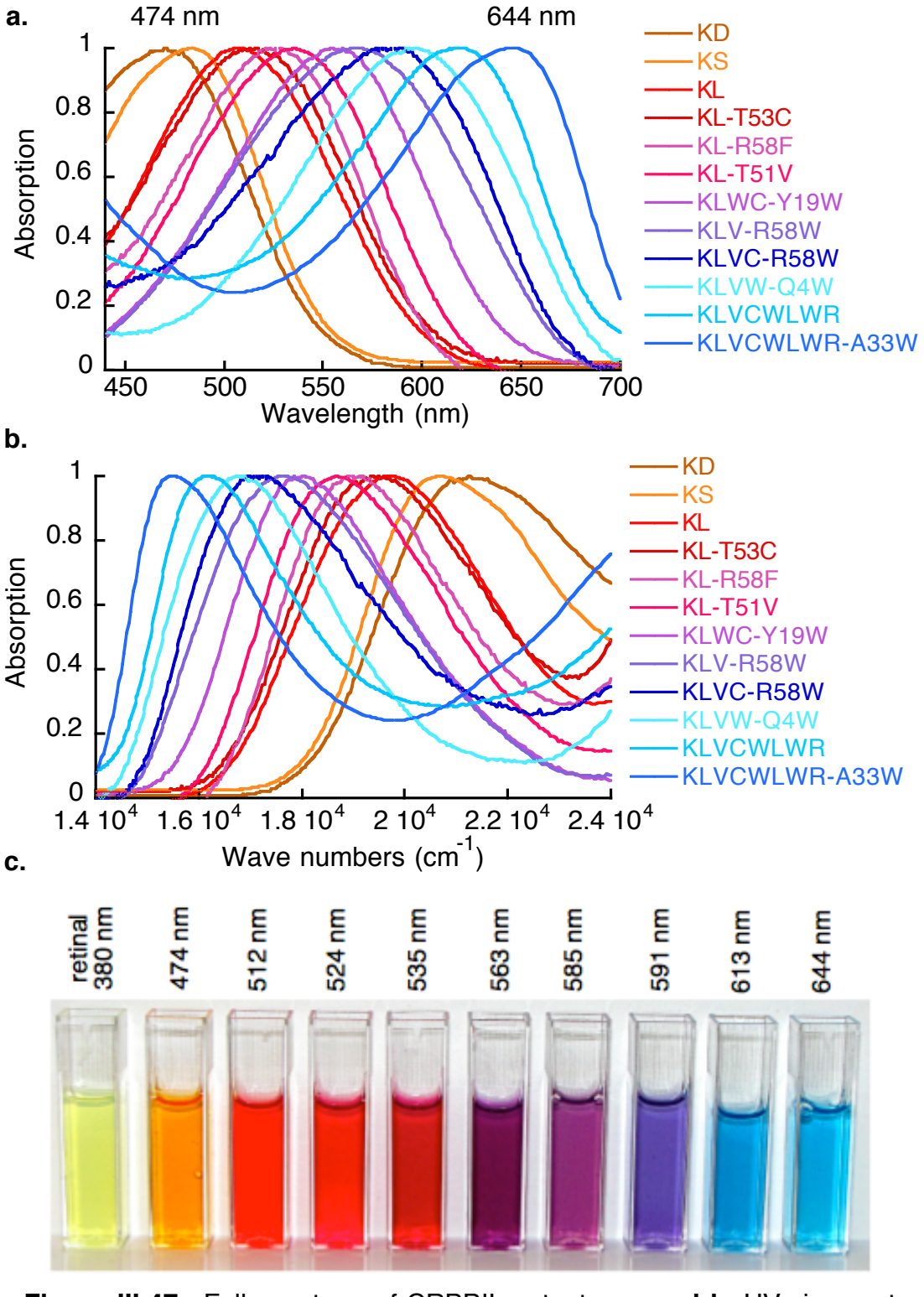


Figure III-47. Full spectrum of CRBPII mutants. **a and b.** UV-vis spectra of different CRBPII mutants bound with all-*trans*-retinal. **c.** Protein solution of different CRBPII mutants incubated with all-*trans*-retinal.

(3, left panel). This is why a fully enclosed binding pocket is important for the electrostatic interactions to be more effective. The R58W mutation can cover the pocket more and possibly lower the dielectric constant inside the binding pocket.

The electrostatic calculations for CRBPII mutants with absorptions ranging from 508 nm to 622 nm were performed. As shown in **Figure III-49,** it is clear that the more red shifted CRBPII mutants exhibit more negative potential in the ionone ring region and less negative potential in the Schiff base region.

However, the electrostatic projection is generated on a low-level calculation. Dynamic movement of the protein is not taken into account and the charges of amino acids inside the binding pocket are assigned according to the pK_a of each residue at pH=7, without considering the depression of pK_a inside the hydrophobic binding pocket. The binding pocket was considered to have the same dielectric constant, which is usually not the case. In the calculation, a dielectric constant of 3 was assigned for mutants with R58W and without R58W, unless specified.

Moveover, crystal structures do not show the position of hydrogens of

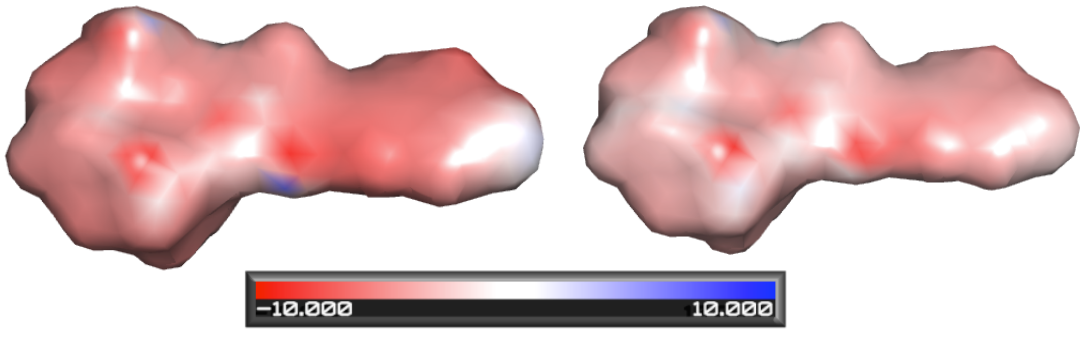


Figure III-48: Comparison of electrostatic calculations of Q108K:K40L:R58W:T51V :T53C:T29L:Y19W:Q4R with dielectric constant of 3 (left) and 6 (right) applied. The unit for the scale is kT/e.

water molecules or hydroxyl groups, however, these hydrogens are critical in electrostatic interactions as they are polarized and exert a significant amount of positive charge. Different positions could lead to different projections of positive potential on the chromophore.

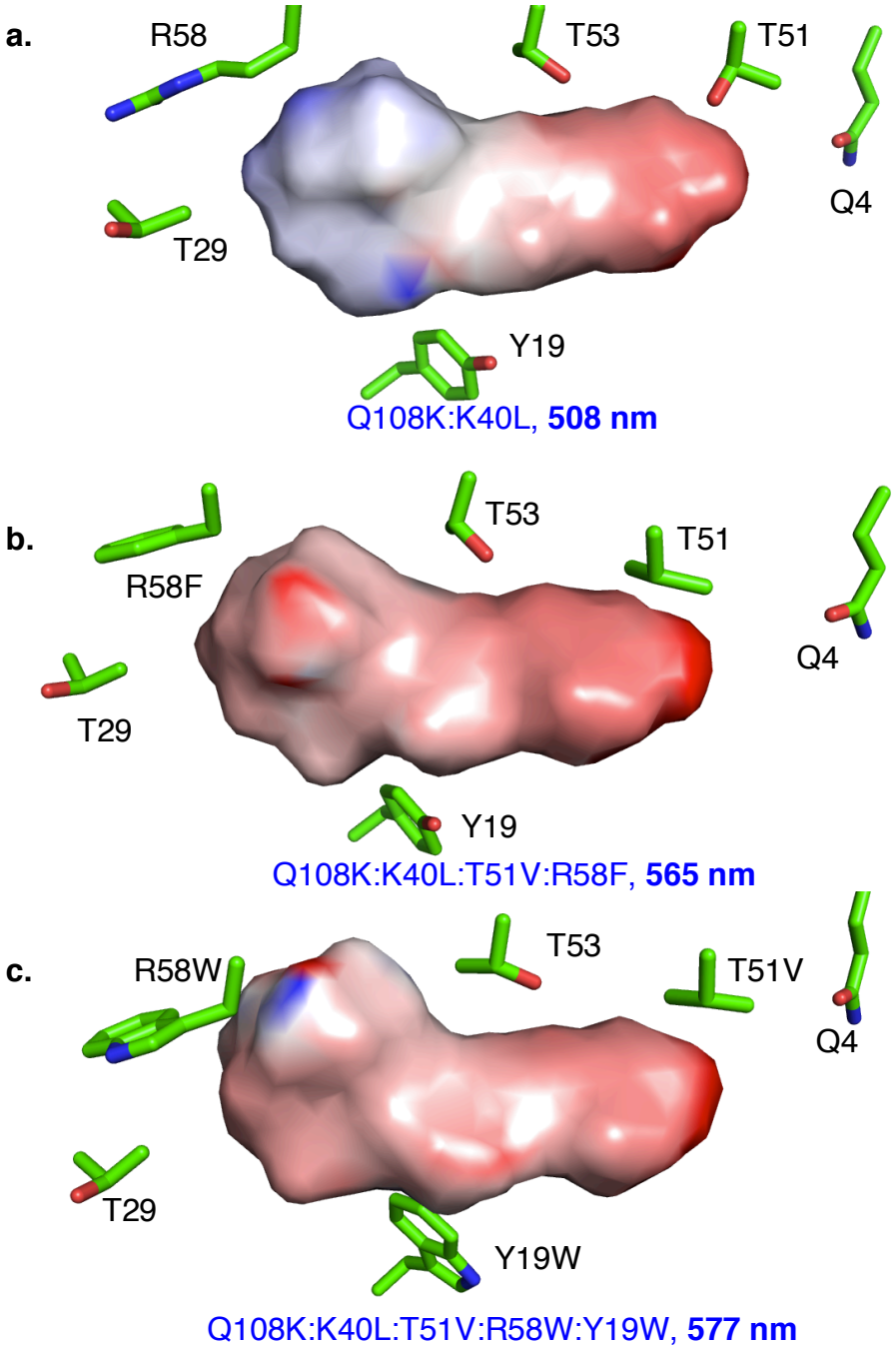
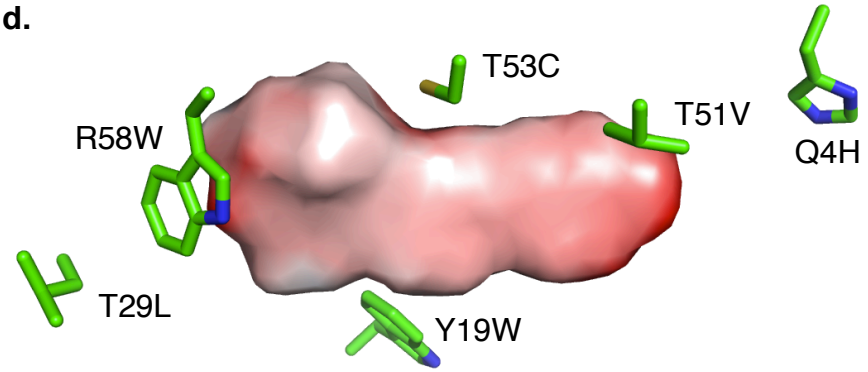


Figure III-49: Electrostatic calculations of CRBPII mutants with dielectric constant of 3 applied.

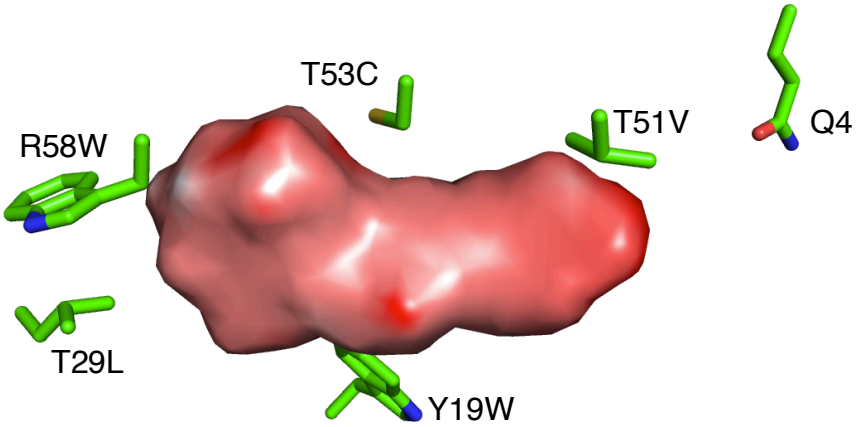
Figure III-49 continued

d.



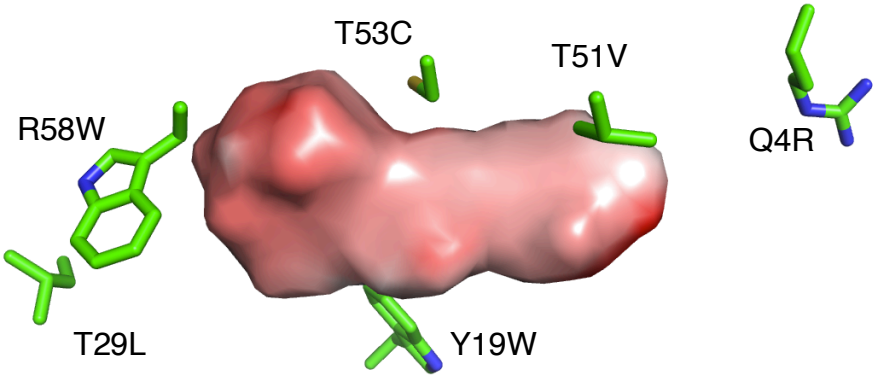
Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4H, **585 nm**

e.



Q108K:K40L:T51V:T53C:R58W:T29L:Y19W, **590 nm**

f.



Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4R, **622 nm**

III.10 Conclusions and outlook

We have engineered a small cellular protein into a rhodopsin mimic, which can bind all-*trans*-retinal as a protonated Schiff base. We have shown that through appropriate protein-chromophore interactions, the absorptions of the chromophore can be regulated to a large extent, from 474 nm to 644 nm, mainly by changing the electronic interactions between the protein and the chromophore.

Increased negative electrostatic potential on the ionone ring and decreased negative electrostatic potential in the PSB region can encourage the positive charge delocalization and lead to red shift. Hydrophobicity of the binding pocket and good insulation of the chromophore from the bulk aqueous environment is extremely important, since dissipation of the electric field is inversely proportional to the dielectric constant of the environment. The dielectric constant of water is ~78, while that deep inside a hydrophobic binding pocket is estimated to range from 2 to 10. This could cause more than an 8 fold difference in the strength of the electrostatic potential projected on the surface of the chromophore, depending on how well the chromophore is sequestered from the aqueous solution. We have observed a single mutation R58W, which could enhance the sensitivity of the chromophore toward the change in polarity of the binding pocket dramatically.

Besides electrostatic interactions, the conformation of the chromophore due to different protein-chromophore packing can also contribute to the overall

conjugation of the system. This requires more investigation. Although quite a few high resolution crystal structures of CRBPII mutants bound with retinal were obtained, the densities of the chromophore for some of the mutants are not clear enough to assign the conformation of the chromophore accurately, especially for the shorter series of mutants (possibly due to more flexibility without Y19W, Q4H or Q4R mutation).

Although it seems pretty clear that electrostatic interactions are playing a major role in tuning the wavelength of retinal-PSB formed with CRBPII mutants, the electrostatic calculation applied is not rigorous enough to quantify the absorptions, leading to a number of unanswered questions regarding the importance of water molecules in the binding pocket. However, with this radically regulated chromophore and the availability of crystal structures, a new platform is available for biophysicists to test different hypotheses for wavelength regulation of retinal-PSB and find a suitable method to quantify retinal-PSB wavelength regulation in different protein environments.

As we have shown that we were able to regulate the wavelength of retinal-PSB through appropriate protein-chromophore interactions, it would be interesting to go back to rhodopsin systems and change rhodopsin's absorption, enabling rhodopsins to respond to different light regimes.

III.11 Summary of all the mutants

Table III-22: Summary of CRBPII mutants complexed with all-*trans*-retinal.

Entry	CRBPII mutant	λ_{max} (nm)	K_d (nM)	ε_{280nm} /1000	pK_a
0	WT	390	23±10	28.9	N.A.
1	Q108K	506	48±4	27.9	<6
2	Q108K:K40L	508	29±5	27.8	8.3
3	Q108K:T51D	474	23±6	28.5	9.2
4	Q108K:K40S	482	28±12	27.0	9.7
5	Q108K:K40N	480	2.9±6.1	28.3	7.6
6	Q108K:K40L:T51V	533	19±7	27.8	8.3
7	Q108K:K40L:T51N	496	n.d.	27.2	7.9
8	Q108K:K40L:T51C	497	14±8	27.2	6.8
9	Q108K:K40L:T53C	513	54±7	28.0	7.5
10	Q108K:K40L:T53V	503	1.3±3.5	27.9	8.3
11	Q108K:K40L:T53S	508	6.9±7.7	27.2	8.2
12	Q108K:K40L:T53N	500	1.5±5.1	27.8	7.8
13	Q108K:K40L:R58W	519	45±3	33.3	8.7

Table III-22 continued

14	Q108K:K40L:R58F	523	36±5	28.1	8.7
15	Q108K:K40L:R58Y	535	29±12	28.8	9.5
16	Q108K:K40L:R58L	500	22±6	26.8	8.5
17	Q108K:K40L:R58E	500	25±4	27.0	8.9
18	Q108K:K40L:R58D	500	28±5	26.7	8.9
19	Q108K:K40L:R58A	499	26±4	27.0	8.6
20	Q108K:K40L:R58Q	499	43±3	28.0	8.4
21	Q108K:K40L:R58W:T51V	565	63±4	32.2	8.4
22	Q108K:K40L:R58W:T53C	540	4.4±10	31.7	8.3
23	Q108K:K40L:R58W:T51V:T53C	585	58±8	33.0	7.4
24	Q108K:K40L:R58W:Y19W	538	54±16	39.5	9.1
25	Q108K:K40L:R58W:Q4W	550	42±6	37.7	7.5
26	Q108K:K40L:R58W:T51V:Q4W	595	92±9	38.5	7.2
27	Q108K:K40L:R58W:T51V:T53C:Y19W	590	55±5	38.1	8.4
28	Q108K:K40L:R58W:T53C:Q4W	575	6±5	37.0	7.1
29	Q108K:K40L:R58W:T51V:Y19W	577	86±6	37.6	9.6
30	Q108K:K40L:R58W:Y19W:Q4W	553	26±6	44.9	8.7

Table III-22 continued

31	Q108K:K40L:T51V:T53C	539	30±14	26.6	8.4
32	Q108K:K40L:Y19W	513	2±4	32.4	8.9
33	Q108K:K40L:T51V:Y19W	537	12±3	32.5	9.3
34	Q108K:K40L:T51V:T53C:Y19W	538	7±4	31.5	8.9
35	Q108K:K40L:Q4W	533	1.6±2.5	33.5	6.2
36	Q108K:K40L:Q4W:T51V	565	58±10	33.7	7.2
37	Q108K:K40L:Q4W:T53C	542	0.03±0.7	33.7	5.4
38	Q108K:K40L:Q4W:Y19W	533	<1	37.2	7.9
39	Q108K:K40L:Q4W:T51V:T53C	574	11±5	33.8	6.9
40	Q108L:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W	404	<1	43.9	N.A.
41	Q108K:K40L:R58W:T51V:T53V:Y19W	577	1.2±2.9	37.9	9.7
42	Q108K:K40L:R58W:T51D:Y19W	545	36±11	35.2	7.7
43	Q108K:K40L:R58W:T53C:Y19W	556	60±6	36.6	8.6
44	Q108K:K40L:R58W:Y19F	526	50±4	30.8	9.2
45	Q108K:K40L:R58L:T51V:T53C	531	22±3	27.5	7.3
46	Q108K:K40L:R58E:T51V:T53C	517	18±5	27.4	8.3
47	Q108K:K40L:R58F:L119Q	523	315±15	26.7	8.8

Table III-22 continued

48	Q108K:K40L:R58F:T51V	561	16±4	27.2	8.7
49	Q108K:K40L:R58F:T53C	537	0.4±2	27.0	8.5
50	Q108K:K40L:R58F:T53S	528	3±2	27.4	8.6
51	Q108K:K40L:R58F:Y19W	537	13±2	31.1	9.4
52	Q108K:K40L:R58F:T51V:T53C	571	0.9±3	26.7	7.8
53	Q108K:K40L:R58Y:T51V	563	42±4	28.8	10.0
54	Q108K:K40L:R58Y:T53C	540	19±4	29.3	9.1
55	Q108K:K40L:R58Y:T51V:T53C	576	65±8	28.7	8.4
56	Q108K:K40L:R58Y:T51V:T53C:T29L:Y19W:Q4W	593	127±16	40.4	8.7
57	Q108K:K40L:R58Y:T51V:Y19W	565	49±5	32.8	10.2
58	Q108K:K40L:R58W:T51V:T53C:T29L	586	30±3	32.6	7.9
59	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W	591	40±9	35.7	8.2
60	Q108K:K40L:R58W:T51V:T53C:Q4W	612	80±4	39.3	7.4
61	Q108K:K40L:R58W:T51V:T53C:T29L:Q4W	613	34±4	38.3	7.3
62	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4F	613	69±10	38.3	7.6
63	Q108K:K40L:R58W:T51V:T53C:T29:Y19W:Q4A	613	164±11	39.4	7.1
64	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W	613	108±10	44.1	7.7

Table III-22 continued

65	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4L	614	65±7	39.1	7.9
66	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4R	622	194±11	40.5	6.7
67	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4H	585	19±4	37.6	7.8
68	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4K	618	15±7	38.3	7.3
69	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4E	590	171±18	39.0	N.D.
70	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4D	612	141±13	41.7	N.D.
71	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4N	610	68±11	41.4	7.3
72	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4T	608	68±7	38.9	7.8
73	Q108K:K40L:R58W:T51V:T53C:T29E:Y19W:Q4W	616	28±5	43.0	7.7
74	Q108K:K40L:R58W:T51V:T53C:T29D:Y19W:Q4W	616	38±5	43.8	7.3
75	Q108K:K40L:R58F:T51V:T53C:T29L:Y19W:Q4H	574	0.2±2	30.6	8.0
76	Q108K:K40L:R58Y:T51V:T53C:T29L:Y19W:Q4H	575	0.3±1.7	33.8	8.7
77	Q108K:K40L:A33W	498	63±5	32.2	9.0
78	Q108K:K40L:A33W:Q4W	527	67±6	39.5	7.4
79	Q108K:K40L:A33W:T51V:Y19W	522	13±3	36.4	8.9
80	Q108K:K40L:A33W:T51V:T53C	533	8.9±6.7	32.6	7.7
81	Q108K:K40L:A33W:T51V:T53C:Y19W	527	58±10	37.2	8.1

Table III-22 continued

82	Q108K:K40L:A33W:T51V:T53C:R58W:T29L	566	32±3	39.0	8.6
83	Q108K:K40L:A33W:T51V:T53C:R58W:T29L:Q4F	629	106±5	39.3	7.3
84	Q108K:K40L:A33W:T51V:T53C:R58W:T29L:Y19W:Q4F	637	60±4	42.9	7.8
85	Q108K:K40L:A33W:T51V:T53C:R58W:T29L:Y19W:Q4R	644	42±6	42.4	6.9
86	Q108K:K40L:A33W:R58F:Y19W	543	78±6	37.0	9.8
87	(Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4F:A33Y	615-620	82±5	39.1	7.7
88	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4F:A33L	605-613	65±6	37.7	7.4
89	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4F:A33H	614	95±7	37.4	7.7
90	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4F:A33E	615	77±5	37.8	7.6
91	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4F:A33F	616	65±7	37.2	7.6
92	Q108K:K40L:R58Y:T51V:T53C:T29L:Y19W:Q4F:A33W	600	15±6	38.4	8.1
93	Q108K:K40L:R58F:T51V:T53C:T29L:Y19W:Q4F:A33W	616	88±5	38.0	7.7
94	Q108K:K40L:R58L:T51V:T53C:T29L:Y19W:Q4F:A33W	583	22±5	38.3	6.8
95	Q108K:K40L:R58A:T51V:T53C:T29L:Y19W:Q4F:A33W	576-590	3.4±4.6	37.7	6.7
96	Q108K:K40L:R58E:T51V:T53C:T29L:Y19W:Q4F:A33W	586	2.1±6.2	38.2	7.0
97	Q108K:K40L:R58H:T51V:T53C:T29L:Y19W:Q4F:A33W	573	2.5±7.9	39.9	6.1

Table III-22 continued

98	Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4F:A33W:M20W	627	168±9	49.2	7.7
99	Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4F:A33W:M20E	635	194±22	43.6	7.8
100	Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4F:M20E	618	205±20	37.1	7.6
101	Q108K:K40L:T51V:T53C:R58W:T29D:Y19W:Q4W:M20E	618	197±9	43.8	7.7
102	Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4F:M20Q	612	135±8	37.7	7.4
103	Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4F:I25Q	614	164±14	38.2	7.8
104	Q108K:K40L:R58W:T51V:T53V:T29L:Y19W:Q4W	600	81±8	43.5	8.1
105	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W:Y60T	608	53±16	43.5	7.2
106	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W:Y60Q	608	16±15	43.3	7.3
107	Q108K:K40L:R58Y:T51V:Y19W:L77T	555	13±1	33.8	9.1
108	Q108K:K40L:R58W:T51V:T53C:T29L:Q128L	532	32±11	33.0	6.0
109	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4F:Q38W	577	n.d.	43.9	6.1
110	Q108K:K40L:R58W:T51V:T53C:T29L:Q128L:Q38M	504	16±4	33.5	7.6
111	Q108K:K40L:R58W:T51V:T53C:T29L:Q38M	513	19±5	32.8	7.5
112	Q108K:K40L:R58W:T51V:T53C:T29L:Q38W	538	15±7	34.0	8.0
113	a. Q108K:K40L:Y60W (40 mM elution)	496	5.5±6.4	32.0	8.6
	b. Q108K:K40L:Y60W (150 mM elution)	514	82±7	32.0	9.1

Table III-22 continued

114	Q108K:K40L:Y60F	494	21±15	26.5	8.2
115	Q108K:K40L:Y60H	494-502	11±13	26.1	8.0
116	Q108K:K40L:L119Q	522	56±15	27.6	9.1
117	Q108K:K40L:L119N	513	15±8	28.1	
118	Q108K:K40L:L119T	516	50±10	27.1	
119	Q108K:K40L:L119Y	518	46±7	28.8	8.6
120	Q108K:K40L:L119F	509	47±13	26.9	8.7
121	Q108K:K40L:L119C	510	19±5	27.1	8.3
122	Q108K:K40L:L119Q:T53C	522		28.0	
123	Q108K:K40L:L119Q:T51V	543	220±11	27.7	9.4
124	Q108K:K40L:L119Q:T51V:T53C	547		27.8	9.0
125	Q108K:K40L:F16Y	486	33±5	28.7	8.6
126	Q108K:K40L:F16Q	486		27.8	
127	Q108K:K40L:F16W	490	5.1±9.6	33.8	8.4
128	Q108K:K40L:L77W	509	65±3	32.5	8.4
129	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W:L77T	595	33±2	45.6	
130	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4H:L77T	563	40±6	38.1	7.3

Table III-22 continued

131	Q108K:K40L:L77F	508		27.9	
132	Q108K:K40L:L77E	504	87±17	27.9	
133	Q108K:K40L:L77C	501	7.5±16	29.1	
134	Q108K:K40L:A33S	502	35±3	28.8	8.6
135	Q108K:K40L:S76L	484	2.2±8.3	27.5	9.0
136	Q108K:K40L:R104L	489		29.2	
137	Q108K:K40L:E72L	507	6.6±1.2	27.7	7.5
138	Q108K:K40L:L77W:R58W	505	87±7	39.6	8.6
139	Q108K:K40L:L77W:Y60W	483	2.7±1.1	37.5	9.0
140	Q108K:K40L:R58W:T51V:T53C:L119Q	556	17±8	33.3	8.4
141	Q108K:K40L:R58W:T51V:T53C:T29L:Q4W:Q38E	590	196±19	39.0	6.0
142	Q108K:K40R:R58W:T51V:T53C:T29L:Y19W:Q4W	550	50±3	42.7	
143	Q108K:K40L:R58W:T51V:T53C:T29L:Q4F	613	74±6	32.8	6.8
144	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W:M20C	609	108±5	45.9	7.6
145	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W:M20T	605	130±6	44.0	7.6
146	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W:F16W	588	44±20	51.5	6.1
147	Q108K:K40L:R58W:T51V:T53C:T29L:Q4W:C95A	610	37±8	40.6	7.5

Table III-22 continued

148	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W:Q44W	609	84±6	51.1	7.5
149	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4L:F57W	609	129±11	49.6	7.5
150	Q108K:K40L:R58W:T51V:T53C:Y19W:Q4L:T29W	616	49±6	44.0	7.9
151	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W:Y60T:E72W	603	21±13	48.0	6.4
152	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:Q4W:Y60Q:E72W	603	20±15	48.5	7.2
153	Q108K:K40L:R58W:T51V:T53C:T29L:Q4F:Y60W:E72W	553	97±40	50.0	< 4
154	Q108K:K40L:R58W:T51V:T53C:T29L:Y19W:I25F	589	5.5±6.1	38.4	8.3
155	Q108K:K40L:R58W:T51V:T53C:Q128E	555		34.0	
156	Q108K:K40L:R58W:T51V:T53C:Q128K	553		32.4	
157	Q108K:K40L:A33W:T51V:T53C:R58W:T29L:Y19W	606	5.9±6.1	43.8	8.7
158	Q108K:K40L:R58F:T51V:T53S	577	0.6±3.2	26.7	9.0
159	Q108K:K40L:R58W:Q4W:A33W	549	87±11	46.2	7.0
160	Q108K:K40L:R58W:T51V:T53C:Y19W:Q4R	622	70±6	38.0	6.7
161	Q108K:K40L:R58W:T51V:T53C:Y19W:Q4R:A33W	644	11±13	44.1	6.5
162	Q108K:K40L:R58W:T51V:T53C:Y19W:Q4W	613	65±8	44.5	7.6
163	Q108K:K40L:R58W:T51V:T53M	544	86±7	32.6	8.6
164	Q108K:K40L:R58W:T51V:T53N	540	11±6	32.4	8.6

Table III-22 continued

165	Q108K:K40L:R58W:T51V:T53S	596	5.6±5.8	32.7	9.2
166	Q108K:K40L:R58W:T51V:T53S:Y19W	600	7.6±2.8	37.1	9.7
167	Q108K:K40L:R58W:T51V:T53S:Y19W:Q4W	622	45±9	45.1	8.4
168	Q108K:K40L:R58W:T51V:T53S:Y19W:T29L	600	1.8±5.3	38.6	9.7
169	Q108K:K40L:R58W:T51V:T53S:Y19W:T29L:Q4F	622	73±7	38.1	8.1
170	Q108K:K40L:R58W:T51V:T53S:Y19W:T29L:Q4F:A33W	646	42±5	43.1	8.4
171	Q108K:K40L:R58W:T51V:T53S:Y19W:T29L:Q4H	594	8.9±4.3	38.2	8.4
172	Q108K:K40L:R58W:T51V:T53S:Y19W:T29L:Q4R	626	14±10	38.8	7.0
173	Q108K:K40L:R58W:T51V:T53S:Y19W:T29L:Q4R:A33W	642	3±10	43.0	7.1
174	Q108K:K40L:R58W:T51V:Y19W:Q4W	593	219±26	44.6	8.0
175	Q108K:K40L:R58W:T53M	558,482	121±22	32.1	8.7
176	Q108K:K40L:R58W:T53S	554	2.8±4	32.6	8.7
177	Q108K:K40L:R58W:T53S:Q4W	578	1.3±4.8	39.0	7.6
178	Q108K:K40L:R58W:T53S:Y19W	565	2.5±5.5	37.4	9.1
179	Q108K:K40L:R58W:T53V	541	1.5±4.4	32.5	9.0
180	Q108K:K40L:R58W:T53W	~485	351±37	38.0	9.3
181	Q108K:K40L:T29L	500	2.7±6.3	26.9	8.3

Table III-22 continued

182	Q108K:K40L:T51V:T53C:Y19W:T29L	533	32±7	31.8	7.9
183	Q108K:K40L:T51V:T53S	534	3.4±3.	27.2	8.7
184	Q108K:K40L:T53M	498	28±10	27.1	8.0
185	Q108K:K40L:T53W	491	280±18	33.0	8.3

Materials and methods

Experiments for mutagenesis, protein expression and characterizations are described in Chapter II.

Molecular modeling and electrostatic calculation

Energy minimization and molecular dynamics were performed with Discover module in InsightII® obtained from Accelrys.[21] The structures for computational studies are obtained from the crystal structures available from the protein database. The conformer of the mutated residues is chosen based on the lowest energy conformation and the double bond is added on the corresponding positions on bound retinal or C_{15} aldehyde before subjecting it to computational study. The number of energy minimization steps is set to 5000 at 298 K for 1 femtosecond per step using CVFF forcefield in InsightII before molecular dynamic calculations. After the structure was minimized, molecular dynamic calculation was performed. The molecular dynamic was equilibrated for 5000 steps for 1 femtosecond using CVFF forcefield, starting at 0 °K. The obtained structures were subject to the subsequent rounds of calculation, where the temperature of each round was increased by 50 °K until 298 °K was reached.

Electrostatic potential calculation was performed using standard procedures described for APBS package.[22] The dielectric constant for the protein was set at 3 while the dielectric constant for water was set for 78 before

the start of calculations. The charge and radius of atoms on each residue within the protein was added by a web-base program pdb2pqr (http://www.poissonboltzmann.org/pdb2pqr/d/web-servers) using Amber94 forcefield before subjecting it for electrostatic potential calculation.[23]

All the protein figures were generated by PyMol Molecular Graphics Systems, version 1.2 educational, copyright by DeLano Scientific LLC.

Primers used for mutagenesis

K40R

Forward: 5'-CACTCAGACG**AGG**GTTATTGATCAAG-3'

Reverse: 5'-CTTGATCAATAAC**CCT**CGTCTGAGTG-3'

T51D

Forward: 5'-GGTGATAACTTCAAG**GAT**AAAACCACTAGCAC-3'

Reverse: 5'-GTGCTAGTGGTTTT**ATC**CTTGAAGTTATCACC-3'

T51V

Forward: 5'-GGTGATAACTTCAAG**GTA**AAAACCACTAGCAC-3'

Reverse: 5'-GTGCTAGTGGTTTT**TAC**CTTGAAGTTATCACC-3'

T51C

Forward: 5'-GGTGATAACTTCAAG**TGT**AAAACCACTAGCAC-3'

Reverse: 5'-GTGCTAGTGGTTTT**ACA**CTTGAAGTTATCACC-3'

T51N

Forward: 5'-GGTGATAACTTCAAG**AAC**AAAACCACTAGCAC-3'

Reverse: 5'-GTGCTAGTGGTTTT**GTT**CTTGAAGTTATCACC-3'

T53N

Forward: 5’-CTTCAAGACAAAA**AAC**ACTAGCACATTCCG-3’

Reverse: 5’-CGGAATGTGCTAGT**GTT**TTTTGTCTTGAAG-3’

T53D

Forward: 5’-CTTCAAGACAAAA**GAT**ACTAGCACATTCCG-3’

Reverse: 5’-CGGAATGTGCTAGT**ATC**TTTTGTCTTGAAG-3’

T53V

Forward: 5’-CTTCAAGACAAAA**GTC**ACTAGCACATTCCG-3’

Reverse: 5’-CGGAATGTGCTAGT**GAC**TTTTGTCTTGAAG-3’

T53C

Forward: 5’-CAAGACAAAATGCACTAGCACATTCCG-3’

Reverse: 5’-CGGAATGTGCTAGTGCATTTTGTCTTG-3’

T53S

Forward: 5’-CAAGACAAAAAGCACTAGCACATTCCG-3’

Reverse: 5’-CGGAATGTGCTAGTGCTTTTTGTCTTG-3’

Y60W

Forward: 5’-CACATTCCGCAAC**TGG**GATGTGGATTTCAC-3’

Reverse: 5’-GTGAAATCCACATC**CCA**GTTGCGGAATGTG-3’

Y60D

Forward: 5’-CACATTCCGCAAC**GAT**GATGTGGATTTCAC-3’

Reverse: 5’-GTGAAATCCACATC**ATC**GTTGCGGAATGTG-3’

Y60F

Forward: 5’-CACATTCCGCAAC**TTT**GATGTGGATTTCAC-3’

Reverse: 5’-GTGAAATCCACATC**AAA**GTTGCGGAATGTG-3’

Y60H

Forward: 5’-CACATTCCGCAAC**CAT**GATGTGGATTTCAC-3’

Reverse: 5’-GTGAAATCCACATC**ATG**GTTGCGGAATGTG-3’

L77E

Forward: 5’-GTACACAAAGAGC**GAG**GATAACCGG-3’

Reverse: 5’-CCGGTTAATCCT**CGC**TCTTTGTGTAC-3’

L77M

Forward: 5’-GTACACAAAGAGC**ATG**GATAACCG-5’

Reverse: 5’-CGGTTATC**CAT**GCTCTTTGTGTAC-3’

L77N

Forward: 5’-GTACACAAAGAGC**AAC**GATAACCGGCATG-3’

Reverse: 5’-CATGCCGGTTATC**GTT**GCTCTTTGTGTAC-3’

L77Q

Forward: 5’-CACAAAGAGC**CAG**GATAACCGGC-3’

Reverse: 5’-GCCGGTTATC**CTG**GCTCTTTGTG-3’

L77C

Forward: 5’-GTACACAAAGAGC**TGC**GATAACCGGCATG-3’

Reverse: 5’-CATGCCGGTTATC**GCA**GCTCTTTGTGTAC-3’

L77F

Forward: 5’-GTACACAAAGAGC**TTC**GATAACCGGCATG-3’

Reverse: 5’-CATGCCGGTTATC**GAA**GCTCTTTGTGTAC-3’

L77W

Forward: 5’-GTACACAAAGAGC**TGG**GATAACCGGCATG-3’

Reverse: 5’-CATGCCGGTTATC**CCA**GCTCTTTGTGTAC-3’

L77S

Forward: 5’-CACAAAGAGC**TCA**GATAACCGGC-3’

Reverse: 5’-GCCGGTTATC**TGA**GCTCTTTGTG-3’

L77T

Forward: 5’-CACAAAGAGC**ACC**GATAACCGGC-3’

Reverse: 5’-GCCGGTTATC**GGT**GCTCTTTGTG-3’

L77G

Forward: 5’-CACAAAGAGC**GGA**GATAACCGGC-3’

Reverse: 5’-GCCGGTTATC**TCC**GCTCTTTGTG-3’

L77A

Forward: 5’-CACAAAGAGC**GCA**GATAACCGGC-3’

Reverse: 5’-GCCGGTTATC**TGC**GCTCTTTGTG-3’

L119Q

Forward: 5’-GCTGTACCTGGAG**CAG**ACCTGTGGTGAC-3’

Reverse: 5’-GTCACCACAGGT**CTG**CTCCAGGTACAGC-3’

L119C

Forward: 5’-GCTGTACCTGGAG**TGT**ACCTGTGGTGAC-3’

Reverse: 5’-GTCACCACAGGT**ACA**CTCCAGGTACAGC-3’

L119E

Forward: 5’-GCTGTACCTGGAG**GAG**ACCTGTGGTGAC-3’

Reverse: 5’-GTCACCACAGGT**CTC**CTCCAGGTACAGC-3’

L119K

Forward: 5’-GCTGTACCTGGAG**AAG**ACCTGTGGTGAC-3’

Reverse: 5’-GTCACCACAGGT**CTT**CTCCAGGTACAGC-3’

L119T

Forward: 5’-GCTGTACCTGGAG**ACA**ACCTGTGGTGAC-3’

Reverse: 5’-GTCACCACAGGT**TGT**CTCCAGGTACAGC-3’

L119D

Forward: 5’-GCTGTACCTGGAG**GAT**ACCTGTGGTGAC-3’

Reverse: 5’-GTCACCACAGGT**ATC**CTCCAGGTACAGC-3’

L119Y

Forward: 5’-GCTGTACCTGGAG**TAC**ACCTGTGGTGAC-3’

Reverse: 5’-GTCACCACAGGT**GTA**CTCCAGGTACAGC-3’

L119F

Forward: 5’-GCTGTACCTGGAG**TTC**ACCTGTGGTGAC-3’

Reverse: 5’-GTCACCACAGGT**GAA**CTCCAGGTACAGC-3’

K40L:Q38E

Forward: 5’-CGTCTCACT**GAG**ACG**CTG**GTTATTG-3’

Reverse: 5’-CAATAAC**CAG**CGT**CTC**AGTGAGACG-3’

K40L:I42D

Forward: 5’-CAGACG**CTG**GTT**GAT**GATCAAGATGG-3’

Reverse: 5’-CCATCTTGATC**ATC**AAC**CAG**CGTCTG-3’

Q108K:W106I

Forward: 5’-GAACCGCGGC**ATC**AAG**AAG**TGG-3’

Reverse: 5’-CCA**CTT**CTT**GAT**GCCGCGGTTC-3’

Q108K:W106F

Forward: 5’-GAACCGCGGC**TTC**AAGAAGTGG-3’

Reverse: 5’-CCACTTCTT**GAA**GCCGCGGTTC-3’

Q108K:W106H

Forward: 5’-GAACCGCGGC**CCA**AAG**AAG**TGG-3’

Reverse: 5’-CCA**CTT**CTT**TGG**GCCGCGGTTC-3’

Q108K:W106E

Forward: 5’-GAACCGCGGC**GAG**AAG**AAG**TGGATTG-3’

Reverse: 5’-CAATCCA**CTT**CTT**CTC**GCCGCGGTTC-3’

K40E:I42T

Forward: 5’-CTCAGACG**GAG**GTT**ACT**GATCAAGATGG-3’

Reverse: 5’-CCATCTTGATC**AGT**AAC**CTC**CGTCTGAG-3’

K40L:I42E

Forward: 5’-CAGACG**CTG**GTT**GAG**GATCAAGATGG-3’

Reverse: 5’-CCATC**TTG**ATC**CTC**AACCAGCGTCTG-3’

F16Y

Forward: 5’-GAGTAATGAAAAC**TAT**GAGGGCTACATG-3’

Reverse: 5’-CATGTAGCCCTC**ATA**GTTTTCATTACTC-3’

F16Q

Forward: 5’-GAGTAATGAAAAC**CAG**GAGGGCTACATG-3’

Reverse: 5’-CATGTAGCCCTC**CTG**GTTTTCATTACTC-3’

F16W

Forward: 5’-GAGTAATGAAAAC**TGG**GAGGGCTACATG-3’

Reverse: 5’-CATGTAGCCCTC**CCA**GTTTTCATTACTC-3’

F16W:Y19W

Forward: 5’-GAGTAATGAAAAC**TGG**GAGGGC**TGG**ATG-3’

Reverse: 5’-CAT**CCA**GCCCTC**CCA**GTTTTCATTACTC-3’

E72L

Forward: 5’-GTA GAG TTT GAC **CTG** TAC ACA AAG AGC-3’

Reverse: 5’-GCT CTT TGT GTA **CAG** GTC AAA CTC TAC-3’

E72A

Forward: 5’-GTA GAG TTT GAC **GCG** TAC ACA AAG AGC-3’

Reverse: 5’-GCT CTT TGT GTA **CGC** GTC AAA CTC TAC-3’

E72F

Forward: 5’-GTA GAG TTT GAC **TTC** TAC ACA AAG AGC-3’

Reverse: 5’-GCT CTT TGT GTA **GAA** GTC AAA CTC TAC-3’

E72W

Forward: 5’-GTA GAG TTT GAC **TGG** TAC ACA AAG AGC-3’

Reverse: 5’-GCT CTT TGT GTA **CCA** GTC AAA CTC TAC-3’

R104L

Forward: 5’-GAAGGAGAAC**CTC**GGCTGGAAGAAG-3’

Reverse: 5’-CTTCTTCCAGCC**GAG**GTTCTCCTTC-3’

R104Y

Forward: 5’-GAAGGAGAAC**TAC**GGCTGGAAGAAG-3’

Reverse: 5’-CTTCTTCCAGCC**GTA**GTTCTCCTTC-3’

A33S

Forward: 5’-CCGCAAGATT**AGC**GTACGTCTCAC-3’

Reverse: 5’-GTGAGACGTAC**GCT**AATCTTGCGG-3’

A33W

Forward: 5’-CGCAAGATT**TGG**GTACGTCTCAC-3’

Reverse: 5’-GTGAGACGTAC**CCA**AATCTTGCG-3’

A33Y

Forward: 5’-CGCAAGATT**TAT**GTACGTCTCAC-3’

Reverse: 5’-GTGAGACGTAC**CTC**AATCTTGCG-3’

A33E

Forward: 5’-CGCAAGATT**GAG**GTACGTCTCAC-3’

Reverse: 5’-GTGAGACGTAC**CTC**AATCTTGCG-3’

A33L

Forward: 5’-CGCAAGATT**CTG**GTACGTCTCAC-3’

Reverse: 5’-GTGAGACGTAC**CAG**AATCTTGCG-3’

A33H

Forward: 5’-CGCAAGATT**CAC**GTACGTCTCAC-3’

Reverse: 5’-GTGAGACGTAC**GTG**AATCTTGCG-3’

A33F

Forward: 5’-CGCAAGATT**TTC**GTACGTCTCAC-3’

Reverse: 5’-GTGAGACGTAC**GAA**AATCTTGCG-3’

F130Y

Forward: 5’-GCCGTCAAGTG**TAC**AAAAAGAAGTTGG-3’

Reverse: 5’-CCAACTTCTTTTT**GTA**CACTTGACGGC-3’

K41L:I43D

Forward: 5’-CAGACG**CTG**GTT**GAT**GATCAAGATGG-3’

Reverse: 5’-CCATCTTGATC**ATC**AAC**CAG**CGTCTG -3’

T51V:T53C

Forward: 5’-CAAG**GTA**AAA**TGC**ACTAGCACATTCCG -3’

Reverse: 5’-CGGAATGTGCTAGT**GCA**TTT**TAC**CTTG -3’

T51V:T53S

Forward: 5’-CAAG**GTA**AAA**AGC**ACTAGCACATTC-3’

Reverse: 5’-GAATGTGCTAGT**GCT**TTT**TAC**CTTG-3’

T51V:T53N

Forward: 5’-CAAG**GTA**AAA**AAC**ACTAGCACATTC-3’

Reverse: 5’-GAATGTGCTAGT**GTT**TTT**TAC**CTTG-3’

T51V:T53W

Forward: 5’-CAAG**GTA**AAA**TGG**ACTAGCACATTC-3’

Reverse: 5’-GAATGTGCTAGT**CCA**TTT**TAC**CTTG-3’

T51V:T53M

Forward: 5’-CAAG**GTA**AAA**ATG**ACTAGCACATTC-3’

Reverse: 5’-GAATGTGCTAGT**CAT**TTT**TAC**CTTG-3’

T53S (2nd)

Forward: 5’-CAAGACAAAA**AGC**ACTAGCACATTC-3’

Reverse: 5’-GAATGTGCTAGT**GCT**TTTTGTCTTG-3’

L119Q: L117N

Forward: 5'-CAAGCTGTAC**AAC**GAG**CAG**ACC-3'

Reverse: 5'-GGT**CTG**CTC**GTT**GTACAGCTTG-3'

L119Q:L117Q

Forward: 5'-CAAGCTGTAC**CAG**GAG**CAG**ACC-3'

Reverse: 5'-GGT**CTG**CTC**CTG**GTACAGCTTG-3'

L117Q

Forward: 5'-CAAGCTGTAC**CAG**GAGCTGACC-3'

Reverse: 5'-GGTCAGCTC**CTG**GTACAGCTTG-3'

L117E

Forward: 5'-CAAGCTGTAC**GAG**GAGCTGACC-3'

Reverse: 5'-GGTCAGCTC**CTC**GTACAGCTTG-3'

L117T

Forward: 5'-CAAGCTGTAC**ACA**GAGCTGACC-3'

Reverse: 5'-GGTCAGCTC**TGT**GTACAGCTTG-3'

L115E

Forward: 5'-GGGGACAAG**GAG**TACCTGGAGC-3'

Reverse: 5'-GCTCCAGGTA**CTC**CTTGTCCCC-3'

L115Q

Forward: 5'-GGGGACAAG**CAG**TACCTGGAGC-3'

Reverse: 5'-GCTCCAGGTA**CTG**CTTGTCCCC-3'

Q128L

Forward: 5'-CAGGTGTGCCGT**CTG**GTGTTCAAAAAG-3'

Reverse: 5'-CTTTTTGAACAC**CAG**ACGGCACACCTG-3'

Q128K

Forward: 5'-CAGGTGTGCCGT**AAG**GTGTTCAAAAAG-3'

Reverse: 5'-CTTTTTGAACAC**CTT**ACGGCACACCTG-3'

Q128R

Forward: 5'-CAGGTGTGCCGT**AGG**GTGTTCAAAAAG-3'

Reverse: 5'-CTTTTTGAACAC**CCT**ACGGCACACCTG-3'

Q128E

Forward: 5'-CAGGTGTGCCGT**GAG**GTGTTCAAAAAG-3'

Reverse: 5'-CTTTTTGAACAC**CTC**ACGGCACACCTG-3'

S76A

Forward: 5'-GAGTACACAAAG**GCC**CTGGATAACCGG-3'

Reverse: 5'-CCGGTTATCCAG**GGC**CTTTGTGTACTC-3'

S76L

Forward: 5'-GTACACAAAG**CTG**CTGGATAACCG-3'

Reverse: 5'-CGGTTATCCAG**CAG**CTTTGTGTAC-3'

R58W

Forward: 5'-CTAGCACATTC**TGG**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**CCA**GAATGTGCTAG-3'

R58E

Forward: 5'-CTAGCACATTC**GAG**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**CTC**GAATGTGCTAG-3'

R58L

Forward: 5'-CTAGCACATTC**CTG**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**CAG**GAATGTGCTAG-3'

R58F

Forward: 5'-CTAGCACATTC**TTC**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**GAA**GAATGTGCTAG-3'

R58D

Forward: 5'-CTAGCACATTC**GAC**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**GTC**GAATGTGCTAG-3'

R58Y

Forward: 5'-CTAGCACATTC**TAC**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**GTA**GAATGTGCTAG-3'

R58Q

Forward: 5'-CTAGCACATTC**CAG**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**CTG**GAATGTGCTAG-3'

R58H

Forward: 5'-CTAGCACATTC**CAC**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**GTG**GAATGTGCTAG-3'

T29L

Forward: 5'-GATTTTGCC**CTG**CGCAAGATTGC-3'

Reverse: 5'-GCAATCTTGCG**CAG**GGCAAAATC-3'

R58A

Forward: 5'-CTAGCACATTC**GCG**AACTATGATGTG-3'

Reverse: 5'-CACATCATAGTT**CGC**GAATGTGCTAG-3'

Y19W

Forward: 5'-CTTTGAGGGC**TGG**ATGAAGGC-3'

Reverse: 5'-GCCTTCAT**CCA**GCCCTCAAAG-3'

Y19F

Forward: 5'-CTTTGAGGGC**TTC**ATGAAGGC-3'

Reverse: 5'-GCCTTCAT**GAA**GCCCTCAAAG-3'

Y19A

Forward: 5'-CTTTGAGGGC**GCG**ATGAAGGC-3'

Reverse: 5'-GCCTTCAT**CGC**GCCCTCAAAG-3'

Y19Q

Forward: 5'-CTTTGAGGGC**CAG**ATGAAGGC-3'

Reverse: 5'-GCCTTCAT**CTG**GCCCTCAAAG-3'

Q4F

Forward: 5'-GACGAGGGAC**TTC**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**GAA**GTCCCTCGTC-3'

Q4L

Forward: 5'-GACGAGGGAC**CTG**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**CAG**GTCCCTCGTC-3'

Q4A

Forward: 5'-GACGAGGGAC**GCG**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**CGC**GTCCCTCGTC-3'

Q4W

Forward: 5'-GACGAGGGAC**TGG**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**CCA**GTCCCTCGTC-3'

Q4R

Forward: 5'-GACGAGGGAC**AGG**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**CCT**GTCCCTCGTC-3'

Q4K

Forward: 5'-GACGAGGGAC**AAG**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**CTT**GTCCCTCGTC-3'

Q4H

Forward: 5'-GACGAGGGAC**CAC**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**GTG**GTCCCTCGT C-3'

Q4E

Forward: 5'-GACGAGGGAC**GAG**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**CTC**GTCCCTCGTC-3'

Q4D

Forward: 5'-GACGAGGGAC**GAC**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**GTC**GTCCCTCGTC-3'

Q4N

Forward: 5'-GACGAGGGAC**AAC**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**GTT**GTCCCTCGTC-3'

Q4T

Forward: 5'-GACGAGGGAC**ACA**AATGGAACC-3'

Reverse: 5'-GGTTCCATT**TGT**GTCCCTCGTC-3'

Q4W:R2W

Forward: 5'-CAGATGACG**TGG**GAC**TGG**AATGG-3'

Reverse: 5'-CCATT**CCA**GTC**CCA**CGTCAGATG-3'

T51V:T53V

Forward: 5’-CAAG**GTA**AAA**GTC**ACTAGCACATTC -3’

Reverse: 5’-GAATGTGCTAGT**GAC**TTT**TAC**CTTG -3’

L119W

Forward: 5’-GTACCTGGAG**TGG**ACCTGTGGTG-3’

Reverse: 5’-CACCACAGGT**CCA**CTCCAGGTAC-3’

R58W:Y60W

Forward: 5'-GCACATTC**TGG**AAC**TGG**GATGTGGATTTC-3'

Reverse: 5'-GAAATCCACATC**CCA**GTT**CCA**GAATGTGC-3'

R58W:Y60T

Forward: 5'-GCACATTC**TGG**AAC**ACA**GATGTGGATTTC-3'

Reverse: 5'-GAAATCCACATC**TGT**GTT**CCA**GAATGTGC-3'

R58W:Y60Q

Forward: 5'-GCACATTC**TGG**AAC**CAG**GATGTGGATTTC-3'

Reverse: 5'-GAAATCCACATC**CTG**GTT**CCA**GAATGTGC-3'

Q97W

Forward: 5'-GTGTGTGTG**TGG**AAGGGGGAG-3'

Reverse: 5'-CTCCCCCTT**CCA**CACACACAC-3'

Q97W:C95A

Forward: 5'-GTCCTTGTG**GCT**GTG**TGG**AAGGGGGAG-3'

Reverse: 5'-CTCCCCCTT**CCA**CAC**AGC**CACAAGGAC-3'

Q97F:C95A

Forward: 5'-GTCCTTGTG**GCT**GTG**TTC**AAG GGGGAG-3'

Reverse: 5'-CTCCCCCTT**CCA**CAC**GAA**CACAAGGAC-3'

C95A

Forward: 5'-GATGTCCTTGTG**GCT**GTG**CAA**AAGGG-3'

Reverse: 5'-CCCTT**TTG**CAC**GAA**CACAAGGACATC-3'

K40L:Q38M

Forward: 5’-CGTCTCACT**ATG**ACG**CTG**GTTATTG-3’

Reverse: 5’-CAATAAC**CAG**CGT**CAT**AGTGAGACG-3’

K40L:Q38L

Forward: 5’-CGTCTCACT**CTG**ACG**CTG**GTTATTG-3’

Reverse: 5’-CAATAAC**CAG**CGT**CAG**AGTGAGACG-3’

K40L:Q38S

Forward: 5’-CGTCTCACT**TCG**ACG**CTG**GTTATTG-3’

Reverse: 5’-CAATAAC**CAG**CGT**CGA**AGTGAGACG-3’

K40L:Q38C

Forward: 5’-CGTCTCACT**TGC**ACG**CTG**GTTATTG-3’

Reverse: 5’-CAATAAC**CAG**CGT**GCA**AGTGAGACG-3’

K40L:Q38W

Forward: 5’-CGTCTCACT**TGG**ACG**CTG**GTTATTG-3’

Reverse: 5’-CAATAAC**CAG**CGT**CCA**AGTGAGACG-3’

R58W:F57W

Forward: 5'-CACTAGCACA**TGGTGG**AACTATGATG-3'

Reverse: 5'-CATCATAGTT**CCACCA**TGTGCTAGTG-3'

T29W

Forward: 5'-GATTTTGCC**TGG**CGCAAGATTGC-3'

Reverse: 5'-GCAATCTTGCG**CCA**GGCAAAATC-3'

Q44W:K40L

Forward: 5'-**CTG**GTTATTGAT**TGG**GATGGTGATAAC-3'

Reverse: 5'-GTTATCACCATC**CCA**ATCAATAAC**CAG**-3'

K40:I42T

Forward: 5’-GACGAAGGTT**ACT**GATCAAGATGG-3’

Reverse: 5’-CCATCTTGATC**AGT**AACCTTCGTC-3’

K40:I42D

Forward: 5’-GACGAAGGTT**GAT**GATCAAGATGG-3’

Reverse: 5’-CCATCTTGATC**ATC**AACCTTCGTC-3’

F64Y

Forward: 5’-GATGTGGAT**TAT**ACTGTTGGAGTAG-3’

Reverse: 5’-CTACTCCAACAGT**ATA**ATCCACATC-3’

K75D

Forward: 5’-CGAGTACACA**GAT**AGCCTGGATAACC-3’

Reverse: 5’-GGTTATCCAGGCT**ATC**TGTGTACTCG-3’

R35D

Forward: 5’-GATTGCAGTA**GAT**CTCACTCAGACG-3’

Reverse: 5’-CGTCTGAGTGAG**ATC**TACTGCAATC-3’

R58W:N59D

Forward: 5'-GCACATTC**TGGGAT**TATGATGTG-3'

Reverse: 5'-CACATCATA**ATCCCA**GAATGTGC-3'

T29D

Forward: 5'-GATTTTGCC**GAT**CGCAAGATTGC-3'

Reverse: 5'-GCAATCTTGCG**ATC**GGCAAAATC-3'

T29E

Forward: 5'-GATTTTGCC**GAG**CGCAAGATTGC-3'

Reverse: 5'-GCAATCTTGCG**CTC**GGCAAAATC-3'

R30D

Forward: 5’-GATTTTGCCACC**GAT**AAGATTGCAG-3’

Reverse: 5’-CTGCAATCTT**ATC**GTTGGCAAAATC-3’

M20T:Y19W

Forward: 5'-CTTTGAGGGC**TGGACC**AAGGCCCTG-3'

Reverse: 5'-CAGGGCCTT**GGTCCA**GCCCTCAAAG-3'

M20C:Y19W

Forward: 5'-CTTTGAGGGC**TGGTGC**AAGGCCCTG-3'

Reverse: 5'-CAGGGCCTT**GCACCA**GCCCTCAAAG-3'

M20D:Y19W

Forward: 5'-CTTTGAGGGC**TGGGAT**AAGGCCCTG-3'

Reverse: 5'-CAGGGCCTT**ATCCCA**GCCCTCAAAG-3'

M20E:Y19W

Forward: 5'-CTTTGAGGGC**TGGGAG**AAGGCCCTG-3'

Reverse: 5'-CAGGGCCTT**CTCCCA**GCCCTCAAAG-3'

M20W

Forward: 5'-CTTTGAGGGC**TACTGG**AAGGCCCTG-3'

Reverse: 5'-CAGGGCCTT**CCAGTA**GCCCTCAAAG-3'

M20W:Y19W

Forward: 5'-CTTTGAGGGC**TGGTGG**AAGGCCCTG-3'

Reverse: 5'-CAGGGCCTT**CCACCA**GCCCTCAAAG-3'

M20Q:Y19W

Forward: 5'-CTTTGAGGGC**TGGCAG**AAGGCCCTG-3'

Reverse: 5'-CAGGGCCTT**CTGCCA**GCCCTCAAAG-3'

L93Q

Forward: 5'-GAA GGT GAT GTC **CAG** GTG TGT GTGC-3'

Reverse: 5'-GCACACACAC**CTG**GACATCACCTTC-3'

I42Q

Forward: 5’-CAGACGCTGGTT**CAG**GATCAAGATGG-3’

Reverse: 5’-CC ATC TTG ATC**CTG**AACCAGCGTCTG-3’

F49Q

Forward: 5’-GATGGTGATAAC**CAG**AAGACAAAAACCAC-3’

Reverse: 5’-GTGGTTTTTGTCTT**CTG**GTTATCACCATC-3’

R58W:V62R

Forward: 5’-C**TGG**AACTATGAT**CGC**GATTTCACTG-3’

Reverse: 5’-CAGTGAAATC**GCG**ATCATAGTTCCAG-3’

References

References

1. Brown, P. K.; Wald, G., Visual pigments in single rods + cones of human retina - direct measurements reveal mechanisms of human night + color vision. *Science* **1964,** *144* (361), 45-&.

2. Motto, M. G.; Sheves, M.; Tsujimoto, K.; Baloghnair, V.; Nakanishi, K., Opsin Shifts in Bovine Rhodopsin and Bacteriorhodopsin - Comparison of 2 External Point-Charge Models. *J. Am. Chem. Soc.* **1980,** *102* (27), 7947-7949.

3. (a) Shimono, K.; Ikeura, Y.; Sudo, Y.; Iwamoto, M.; Kamo, N., Environment around the chromophore in pharaonis phoborhodopsin: mutation analysis of the retinal binding site. *BBA Biomembranes* **2001,** *1515* (2), 92-100; (b) Shimono, K.; Iwamoto, M.; Sumi, M.; Kamo, N., Effects of three characteristic amino acid residues of Pharaonis phoborhodopsin on the absorption maximum. *Photochem. Photobiol.* **2000,** *72* (1), 141-145.

4. 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 Biochem. Sci.* **2009,** *34* (11), 540-552.

5. Mathies, R.; Stryer, L., Retinal has a highly dipolar vertically excited singlet-state - implications for vision. *P. Natl. Acad. Sci. USA* **1976,** *73* (7), 2169-2173.

6. Soederhjelm, P.; Husberg, C.; Strambi, A.; Olivucci, M.; Ryde, U., Protein influence on electronic spectra modeled by multipoles and polarizabilities. *J. Chem. Theory Comput.* **2009,** *5* (3), 649-658.

7. 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. *P. Natl. Acad. Sci. USA* **1991,** *88* (8), 3079-3083.

8. Kong, F. R.; King, J., Contributions of aromatic pairs to the folding and stability of long-lived human gamma D-crystallin. *Protein Sci.* **2011,** *20* (3), 513-528.

9. Baasov, T.; Sheves, M., Alteration of pka of the bacteriorhodopsin protonated schiff-base - a study with model compounds. *Biochemistry* **1986,** *25* (18), 5249-5258.

10. Rosenber.B; Krigas, T. M., Spectral shifts in retinal schiff base complexes. *Photochem. Photobiol.* **1967,** *6* (10), 769-&.

11. Baasov, T.; Sheves, M., Model compounds for the study of spectroscopic properties of visual pigments and bacteriorhodopsin. *J. Am. Chem. Soc.* **1985,** *107* (25), 7524-7533.

12. Irving, C. S.; Byers, G. W.; Leermake.Pa, Spectroscopic model for visual pigments - influence of microenvironmental polarizability. *Biochemistry* **1970,** *9* (4), 858-&.

13. Sakurai, M.; Sakata, K.; Saito, S.; Nakajima, S.; Inoue, Y., Decisive role of electronic polarization of the protein environment in determining the absorption maximum of halorhodopsin. *J. Am. Chem. Soc.* **2003,** *125* (10), 3108-3112.

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. *Angew. Chem. Int. Ed. Engl.* **2010,** *49* (10), 1790-1793.

15. Wada, A.; Tsutsumi, M.; Inatomi, Y.; Imai, H.; Shichida, Y.; Ito, M., Retinoids and related compounds. Part 26. Synthesis of (11 Z)-8,18-propano- and methano-retinals and conformational study of the rhodopsin chromophore. *J. Chem. Soc. Perk. T. 1* **2001,** (19), 2430-2439.

16. Trabanino, R. J.; Vaidehi, N.; Goddard, W. A., III, Exploring the molecular mechanism for color distinction in humans. *J. Phys. Chem. B* **2006,** *110* (34), 17230-17239.

17. Cembran, A.; Gonzalez-Luque, R.; Altoe, P.; Merchan, M.; Bernardi, F.; Olivucci, M.; Garavelli, M., Structure, spectroscopy, and spectral tuning of the gas-phase retinal chromophore: The beta-ionone "handle" and alkyl group effect. *J. Phys. Chem. A* **2005,** *109* (29), 6597-6605.

18. (a) Hackett, N. R.; Stern, L. J.; Chao, B. H.; Kronis, K. A.; Khorana, H. G., Structure-function studies on bacteriorhodopsin .5. effects of amino-acid substitutions in the putative helix-F. *J. Biol. Chem.* **1987,** *262* (19), 9277-9284; (b) Mogi, T.; Marti, T.; Khorana, H. G., Structure-function studies on bacteriorhodopsin .9. substitutions of tryptophan residues affect protein-retinal interactions in bacteriorhodopsin. *J. Biol. Chem.* **1989,** *264* (24), 14197-14201; (c) Greenhalgh, D. A.; Farrens, D. L.; Subramaniam, S.; Khorana, H. G., Hydrophobic amino-acids in the retinal-binding pocket of bacteriorhodopsin. *J. Biol. Chem.* **1993,** *268* (27), 20305-20311.

19. Wisz, M. S.; Hellinga, H. W., An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants. *Proteins* **2003,** *51* (3), 360-377.

20. Neitz, M.; Neitz, J.; Jacobs, G. H., Spectral tuning of pigments underlying red-green color-vision. *Science* **1991,** *252* (5008), 971-974.

21. *DISCOVER User Guide*. MSI: San Diego, 1995.

22. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A., Electrostatics of nanosystems: Application to microtubules and the ribosome. *P. Natl. Acad. Sci. USA* **2001,** *98* (18), 10037-10041.

23. (a) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A., PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007,** *35*, W522-W525; (b) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004,** *32*, W665-W667.

Chapter IV

Developing CRBPII derivatives into fluorescent and chromophoric tags

IV.1 Introduction of GFP and its derivatives

Cellular biology has developed substantially since the introduction of fluorescent protein tags. Fusion of fluorescent proteins with the protein of interest allows visualization of the targeted protein. Fluorescent proteins have been widely applied to study the expression level, localization, trafficking, and molecular functions of proteins *in vivo* and in tissue culture.[1]

Green Fluorescent Protein (GFP) is the first and most used fluorescent protein for imaging in live cells without fixation. Its importance has been realized and three people, Osamu Shimomura, Martin Chalfie and Roger Tsien have contributed the most to the development of GFP, and were awarded Nobel Prize in 2008.

GFP was first found in *Aequorea victoria* by Osamu Shimomura when he was a student in the laboratory of Professor Frank Johnson at Princeton University in 1960s.[2] The chemical structure of the chromophore formed in GFP was proposed and verified to be 4-(*p*-hydroxybenzylidene)imidazolidin-5-one (HBI) in 1979.[3] Douglas Prasher sequenced and cloned the GFP gene in 1992, but without the success of autofluorescence when expressing the GFP in *E. coli*,

which was later realized to be due to inhibition by a few extra amino acids in the N-terminus.[4] Martin Chalfie got the gene from Douglas Prasher and successfully expressed GFP that could auto-catalytically generate the fluorophore by itself.[5] The auto-fluorescence ability of GFP makes them suitable as fluorescent tags. Tsien's group worked on GFP in parallel and generated much brighter monomeric GFP, as well as different colors to enrich the fluorescent protein palette.

The GFP protein has 238 amino acids. It forms an 11-β-sheet barrel with an α-helix going through the center of the barrel (**Figure IV-1a**).[6] The fluorophore is formed from residues 65, 66 and 67, which belong to the α-helix. The chemical structure of the chromophore formed in GFP and the proposed mechanism for maturation of GFP is illustrated in **Figure IV-b**. The formation of the GFP chromophore starts from cyclization of the amide backbone, followed by dehydration to form the imidazole ring. It is further oxidized to generate the fluorophore in WT-GFP, and hydrogen peroxide as a byproduct, which impedes high level expression of GFP in live cells.[7]

WT-GFP is not bright due to the high pK_a of Tyr66. The majority of the fluorophore is protonated in WT-GFP; only a small portion of the fluorophore is deprotonated, which turns out to be the fluorescent form that emits with a quantum efficiency of 65%.[8] A single mutation made by Tsien's group at position S65 to Ala, Leu, Cys or Thr disrupts the hydrogen bonding network that stabilizes

the protonated state of Tyr66.[9] This decreases the pK_a of Tyr66 and makes the deprotonated fluorophore exclusively. In this way, the brightness of GFP was increased dramatically. Mutation S65T also accelerates the maturation of the fluorophore by four-fold, which greatly facilitates the application of GFP as a fluorescent protein tag.

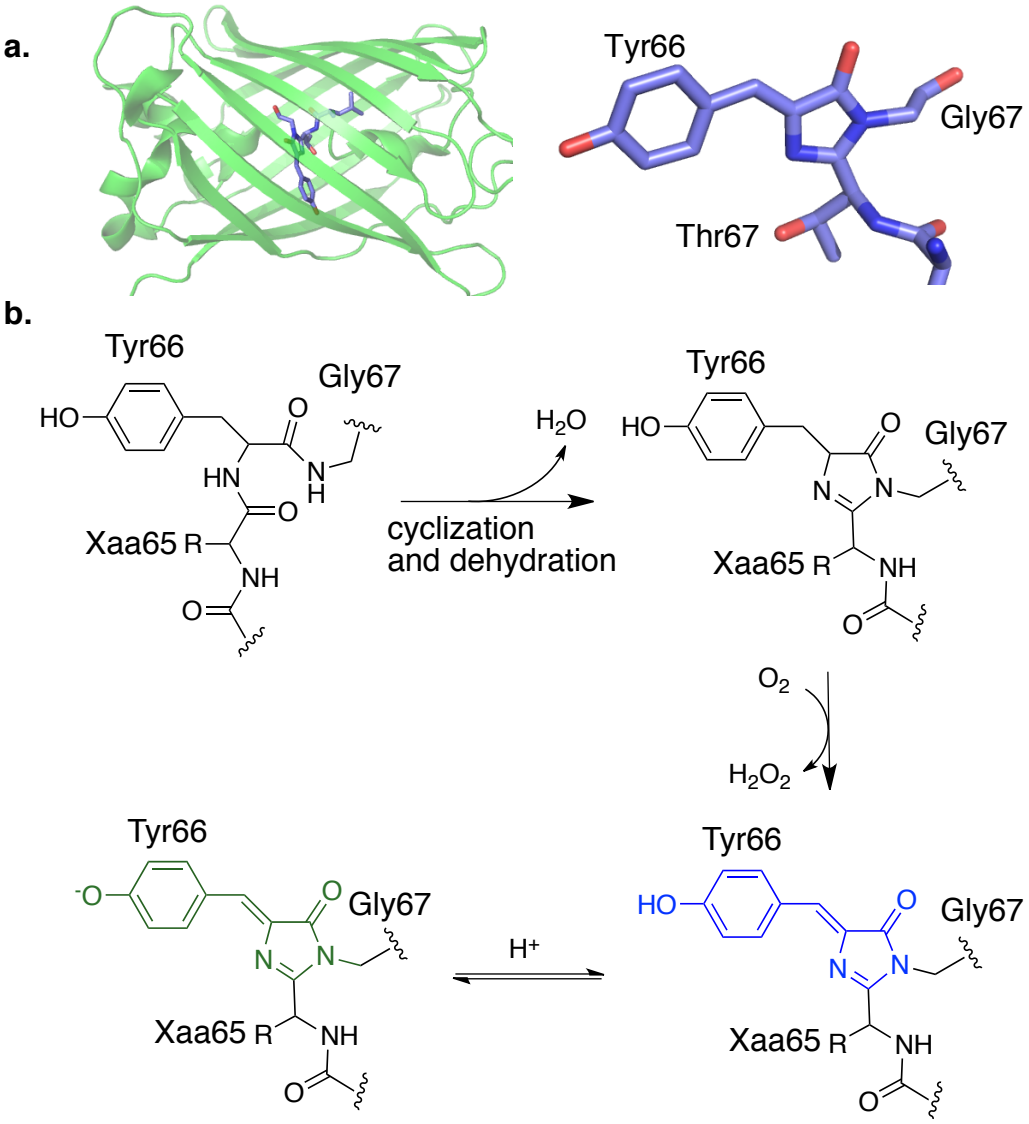


Figure IV-1: Illustration of GFP protein structure and mechanism of chromophore formation. **a.** Crystal structure (left) and chromophore (right) of EGFP (PDB entry: 2Y0G). **b.** Molecular mechanism for the chromophore maturation of green fluorescent proteins.

The spectral characteristics of GFP can be varied by either changing the amino acid at Tyr66 or by changing the binding pocket of the GFP fluorophore. Blue, cyan and orange fluorescent proteins were made in the late 90s.[10] However, genetic modification of *Aequorea victoria* GFP (avGFP) to red fluorescent protein was not achieved until 2008.[11]

Development of red fluorescent proteins was initiated from a new source of red fluorescent protein, called DsRed, which was found in coral.[12] It shares a similar structural scaffold with that of avGFP, although they have only 23% sequence homology.[12] The chromophore of DsRed has a more conjugated system, as a result of one more oxidation step as shown in **Figure IV-2**. Mutations at position 65 generate different variants of DsRed with fluorescence at different wavelengths.

Different from avGFP, DsRed forms a tetramer naturally and its half time of maturation at 37 °C is around 10 hours. A total number of 33 mutations were made to speed up the maturation, disrupt the interactions in the interface of the tetramer, and still maintain the fluorescence.[13] Starting from the reengineered monomeric red fluorescent protein, different derivatives have been generated to obtain fluorescent proteins with different spectral characteristics, photo-bleaching stabilities and maturation rates.[14] Various excitation and emission spectra of fluorescent proteins available have made possible the visualization of multiple

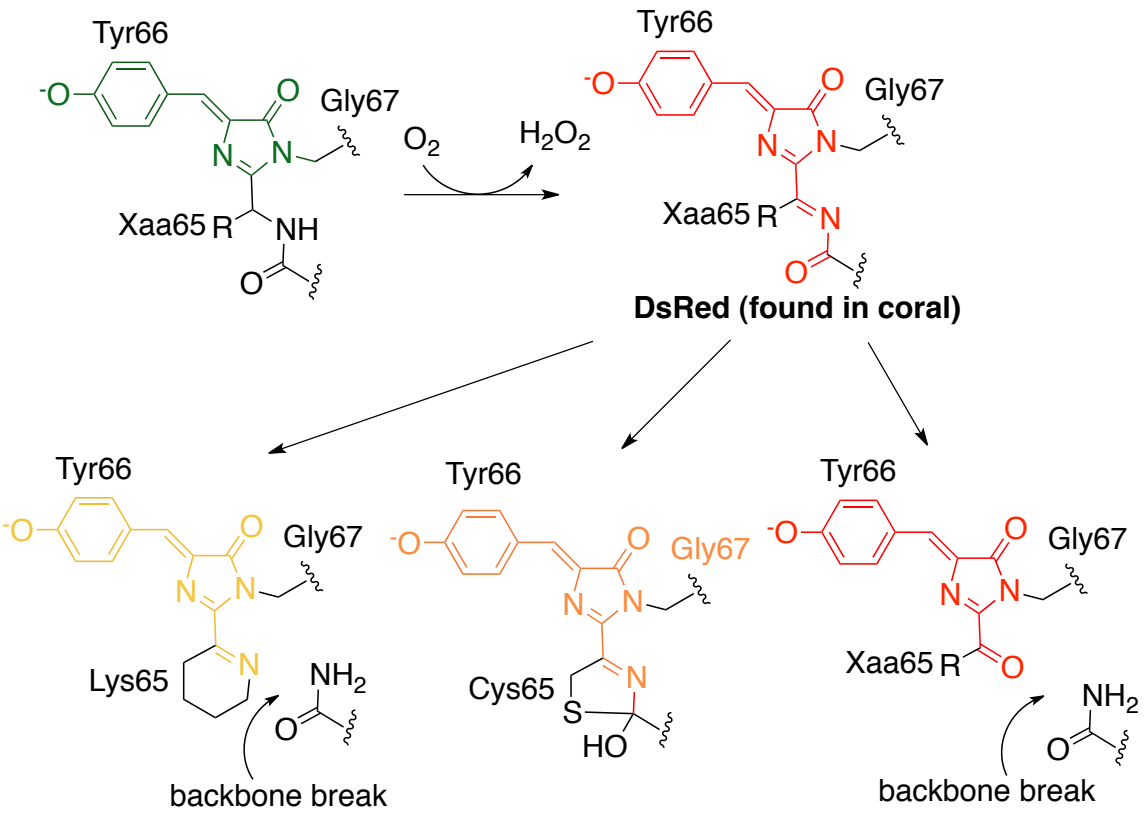


Figure IV-2: Molecular mechanism for the formation of the fluorophore and molecular structure of different variants developed from DsRed.

protein or multiple events at the same time through tagging with different fluorescent proteins.[15]

Fluorescent protein tags have been used extensively to check the expression level and localization of proteins of interest through fusing with the fluorescent protein.[1a] Fluorescent proteins have been used to study the mobility of the targeted protein by photobleaching the fluorescent protein in one location first and observing the recover of fluorescence due to protein movement.[16] Fluorescent proteins have also been used as a signal readout for gene

transcription activation by inserting the gene of fluorescent protein downstream of the promoter region.[17] Once the promoter is activated, fluorescent protein will be expressed and thus fluorescence will be observed. Fluroescent proteins can be used to study protein-protein interactions or conformational change of the peptides and protein during some biological events, through FRET or split fluorescent proteins.[18]

Besides the above mentioned applications, recently developed photoactivatable fluorescent proteins have been used for super-resolution fluorescent microscopic studies.[19, 20] Photoactivation of only a few fluorescent molecules at a time allows the visualization of the fluorescence emission from these individual fluorescent molecules. In this way, the fluorescence noise from neighboring fluorescent molecules can be minimized. It is possible to increase the precision in space and increase the resolution to tens of nanometers, while the resolution of a traditional microscope is limited to hundreds of nanometers. Super resolution fluorescence microscopic studies will enable visualization of single protein molecules inside the cell, which can greatly facilitate the single molecule study of protein dynamics, interactions and function.

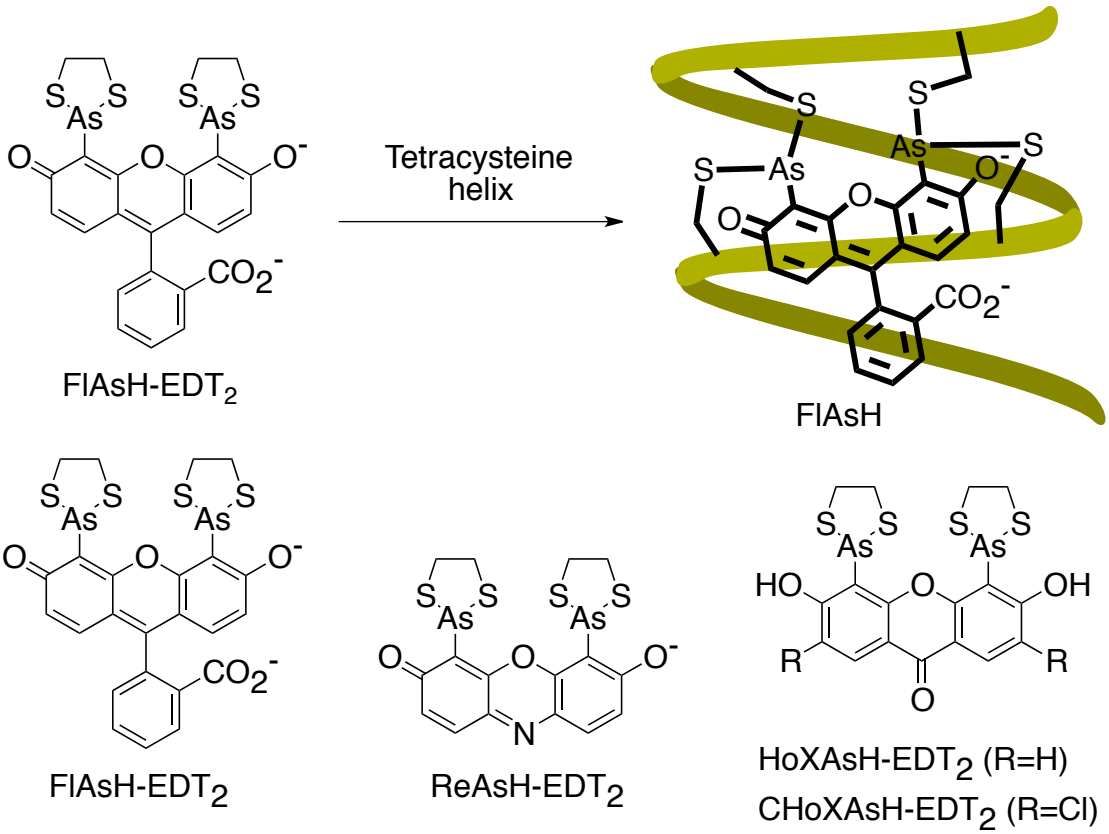
Both reversible and irreversible photoactivatable fluorescent proteins have been developed. For reversible photoactivatable fluorescent proteins, fluorescence could be switched on and off many times, or fluorescence could be switched forward and backward from one wavelength to another wavelength with

a pulse of light at specific wavelengths. This is usually due to the *cis-trans* isomerization of the double bond, which changes the protein-chromophore interactions, leading to changes in spectral characteristics.[21] Irreversible photoactivation usually involves chemical structural changes of the chromophore, like bond cleavage of the chromophore from the backbone of the protein.[22]

IV.2 Fluorescent protein tags other than GFP

IV.2.1 FlAsH tag

FlAsH, which is abbreviated for Fluorescein Arsenical Helix Binder, was developed by Tsien's group. The idea is to use a short peptide, which can form an alpha-helix and contains four cysteines at the i, i + 1, i + 4, and i + 5 positions, as a peptide tag to chelate a bis-arsenic-fluorescein molecule as shown in **Figure IV-3.**[23] The fluorescence of fluorescein is enhanced a thousand fold, upon binding to the tetra-cysteine motif in FlAsH. This is believed to be due to decreased non-radiative relaxation as a result of restricted rotation when bound to a rigid peptide. However, arsenic compounds are usually toxic. A tight chelating reagent 1,2-ethanedithiol (EDT) has to be added to prevent cellular cysteine chelating to arsenic, and thus minimize the toxic effects of arsenic compounds in live cells. The *in vitro* association and dissociation rate constants of FlAsH were determined to be in the range of 10^{5} $M^{-1}s^{-1}$ and 10^{-6} $M^{-1}s^{-1}$, which corresponds to a dissociation constant of 10^{-11} M. Such tight binding



	FlAsH	ReAsH	CHoXAsH
λ_{Ex} (nm)	500	593	380
λ_{Em} (nm)	529	608	430

Figure IV-3: Illustration of FlAsH mechanism and three different colors of FlAsH system.

allows addition of arsenic compounds at micromolar concentration and removal of them after minutes to minimize the toxicity.

Multi-color probes that cover the entire visible light spectrum, have been designed by addition of different organoarsenic compounds as shown in **Figure IV-3**. Since the tetracysteine helix can not distinguish among different organoarsenic compounds, multiple tags of FlAsH or ReAsH can not be used at the same time in the same cell.

IV.2.2 SNAP tag

The SNAP tag originated from the human DNA repair protein, O^6-alkylguanine-DNA alkyltransferase (hAGT). hAGT is a suicide enzyme. It irreversibly transfers the alkyl group to one of its cysteine groups in the active site from the alkylated guanine, in order to fulfill its role as a DNA repair protein (**Figure IV-4a)**.[24] hAGT is specific for O^6-benzyl-guanine, but substitutions in the *para*-position of the benzyl group do not seem to interfere with the activity of the alkyl transferase.[25] Due to this characteristic, Johnsson's group covalently labeled hAGT protein with any molecule by attaching it to the benzyl group (**Figure IV-4b**).[26]

Optimization of hAGT protein as a tag starts with truncation of 30 amino acids in the C-terminus of the protein, which does not contribute to the activity of the enzyme. As a result the hAGT-tag has 177 amino acids. Mutant G160W was introduced to increase its activity towards the unnatural substrate, BG (**Figure IV-4b**). Initially, the hAGT tag suffered from background fluorescence problems due to the activity of endogenous hAGT protein. This problem was solved by designing an inhibitor, CG (**Figure IV-4c**), which is specific for WT-hAGT protein. At the same time a hAGT mutant, ^{M}AGT, whose activity towards BG is not affected by CG, was engineered through directed evolution.[27]

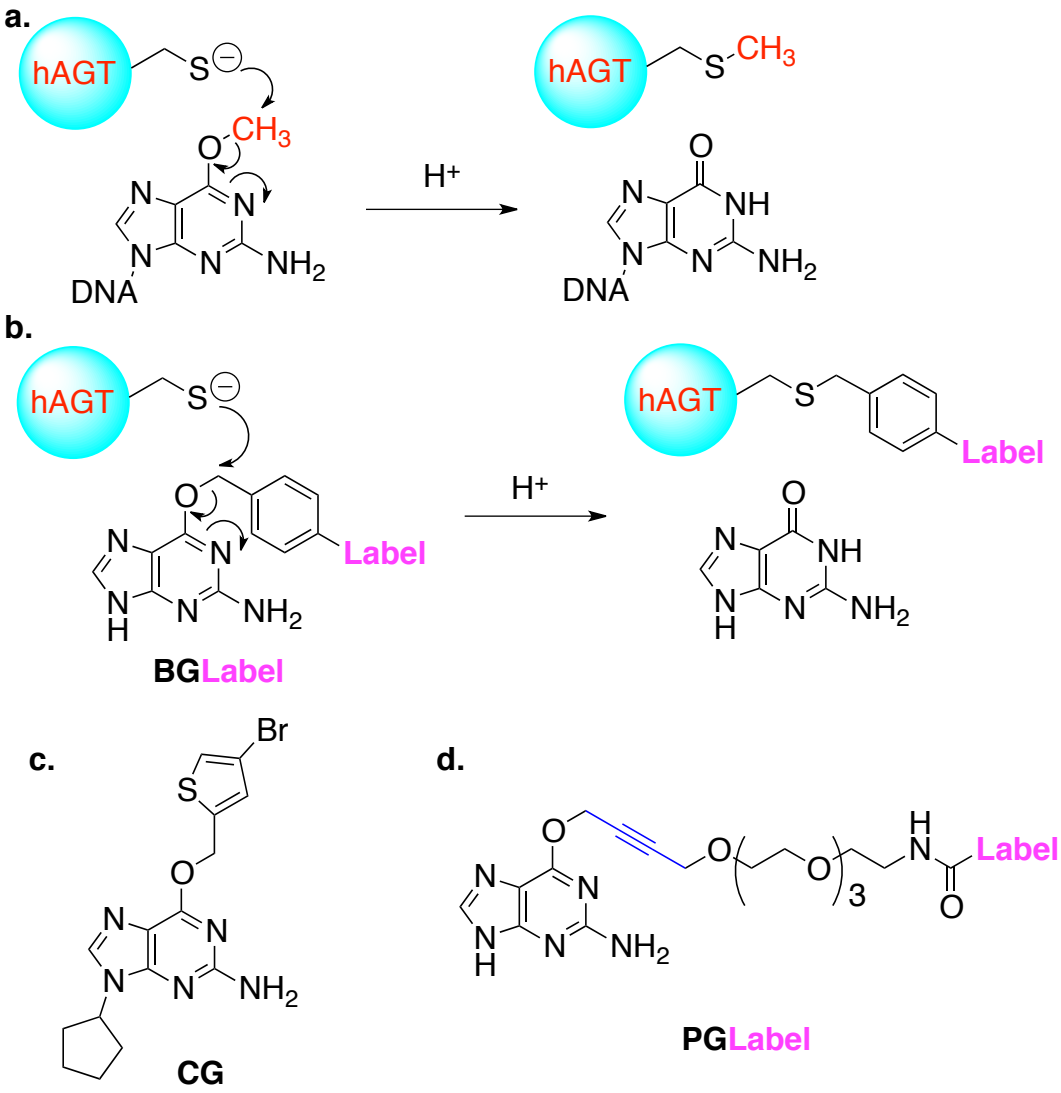


Figure IV-4: Illustration of SNAP protein tag mechanism. **a.** The mechanism of the DNA repair enzyme hAGT. **b.** General ways of labeling hAGT with a labeled benzylguanine. **c.** Molecular structure of WT-hAGT inhibitor CG. **d.** Molecular structure of labeled propargylguanine.

The application of SNAP tag is broad. Tagging different fluorophores to benzyl-guanine (BG) would enable different spectral characteristics. Thus it can easily overcome the problem of low brightness for red fluorescent protein and it can easily reach to the near-IR spectrum by attaching more conjugated compounds. By applying cell-impermeable fluorophores and cell-permeable

fluorophores, it is possible to distinguish cell surface proteins from intracellular proteins.

SNAP technique is not selective with different labels. This limits the usage of multiple tags at one time as well. The challenge was overcome by developing different mutants of AGT that specifically target a different set of substrates. Johnsson's group thus developed another derivative of ^{L}AGT, which is specific for O^6-propargylguanine (**Fiugre IV-4d**).[28] This has enabled application of multiple tags at the same time.

IV.2.3 Modified ligases for fluorescent tag

Specific labeling of proteins using ligases has also been developed for imaging live cells. Ting's group has evolved a few sets of orthogonal ligases with a short target peptide and modified substrates, which enabled direct or indirect visualization of the target peptide.[29] As shown in **Figure IV-5a**, different lipoic acid analogues with labels attached could be used as the substrate for modified lipoic acid ligase, which catalyzes the amide bond formation between the lipoic acid analogues and a lysine residue of an orthogonal peptide (less than 15 amino acids), as illustrated in **Figure IV-5b**. Due to the tight binding pocket of lipoic acid ligase, mutation W37A was found to be necessary for accommodating the labeled substrates for the ligase.

Labeling with an azide group or alkyne group allows further modification through click chemistry, which has a broad scope of possible labeling with any

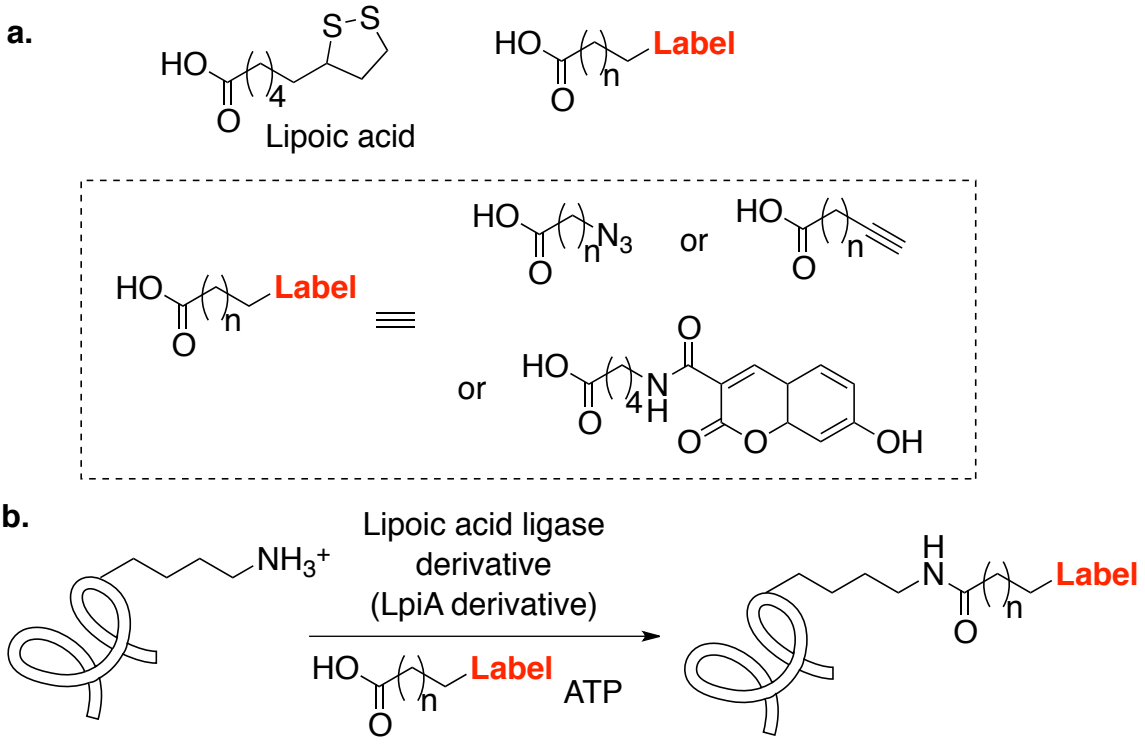


Figure IV-5: Demonstration of using lipoic acid ligase for fluorescent protein tag development. **a**. Chemical structure of lipoic acid and its analogues with different labeling groups. **b.** Strategies to tag different label to a short peptide, which is the substrate of a lipoic acid ligase derivative.

functional groups or fluorophores.[29d] However, this approach requires two steps of washing, which increases the background signal. Direct labeling with compact fluorophores, such as coumarin, makes this method much more convenient for *in vivo* fluorescent labeling.[29h, 29j]

Using biotin ligase to ligate biotin to a specific target peptide substrate allows coupling with streptavidin coated quantum dots.[29c] Quantum dots have attracted a lot of attention as a fluorescent probe due to their much brighter fluorescence and photostability compared to normal fluorophore molecules,

although quantum dots have cell-permeability problems due to their relatively large size.

What's more, different fluorescent probes based on fluorescence activation upon binding to a tight binding pocket have also been reported. Verkhusha's group has developed a near-IR fluorescent protein which has high affinity phytochrome and turns on the fluorescence of phytochrome upon binding, with excitation at 690 nm and emission at 713 nm.[30] It has long maturation time, with $t_{1/2}$ of around 3 hours and about one fifth the brightness of EGFP. This is a breakthrough for near-IR fluorescent protein development, without introduction of exogenous chromophores. But the application is limited to hosts with endogenous phytochrome, which include plants and some strains of bacteria.

IV.3 Engineering of CRBPII into a fluorescent protein tag

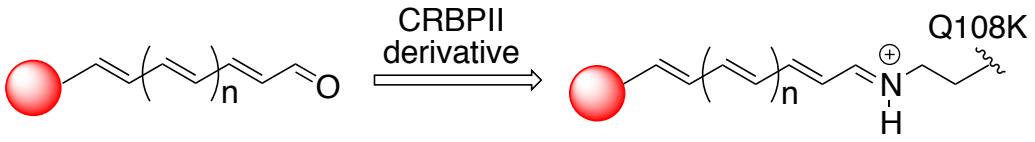
Red and near-IR fluorescent proteins are highly desired. This is because the longer the wavelength, the less diffraction will result and the deeper the light can penetrate into tissues. Red and near-IR fluorescence are also out of the spectral window where molecules inside cells absorb or emit. Thus using red and near-IR fluorescent probes can get around the background fluorescence inside cells. However, there are few choices of bright red and near-IR fluorescent protein tags available with good photostability. Therefore there is an urgent need to develop brighter and more photostable fluorescent protein tags

with red and near-IR emissions to complement the existing fluorescent protein palette and for deep tissue studies.

CRBPII has been engineered into a rhodopsin mimic that binds all-*trans*-retinal as a Protonated Schiff Base (PSB) and can regulate the wavelength of the bound chromophore greater than 170 nm. The high expression level of CRBPII derivatives in *E. coli* cells, up to 200 mg/L LB, and the stability of the protein-chromophore complex make them suitable to be developed into fluorescent tags, if appropriate fluorophores can be applied. Furthermore, if the wide range of wavelength regulation observed with retinal-PSB could be reproduced with other fluorophores, different CRBPII mutants could be applied at the same time to afford multi-color fluorescent protein probes by addition of just one fluorophore.

It is interesting to note that the Schiff base of retinal red shifts dramatically (more than 60 nm) upon protonation. This characteristic makes the conjugated polyene aldehyde a good target to be used as a fluorophore-precursor for CRBPII derivatives. This is because the red shift can help reduce the background fluorescence introduced from excessive amounts of polyene-aldehyde added, by excitation at a wavelength away from the absorption of polyene-aldehyde but optimal for the PSB formed. Additionally, the CRBPII binding site is long and rigid, thus polyene systems could better fit into the binding pocket with high affinity.

As shown in **Figure IV-6**, besides the polyene-aldehyde moiety, the general structure of fluorophores should also include an electron-donating group



represents a head group with electron-donating group;

n = 0, 1, 2, 3.

Figure IV-6: General structure of polyene aldehyde to be applied to CRBPII derivatives for the development of fluorescent protein tags.

in conjugation with the polyene system. This is because most of the fluorescent molecules have an electron push-pull system. This increases the degree of conjugation and distributes the double bonds more evenly, thus increasing the double bond character of the single bonds and rigidifying the single bond rotation, increasing quantum efficiency. In the protein-fluorophore complex formed, the electron deficient positively charged PSB is the electron pull end, while the electron-donating group is the electron push end.

Sequestering of the fluorophore in a hydrophobic binding pocket can decrease interactions with solvent molecules, suppressing the non-radiative relaxation through heat transfer with the solvent molecules. In this way, the quantum efficiency can be enhanced. Furthermore, rigidifying the binding pocket by introduction of large side chains can tighten the protein-fluorophore packing and restrict chromophore rotation and vibrational freedom, thus increasing quantum efficiency.

IV.4 Spectroscopic characterization of CRBPII mutants with different chromophores.

IV.4.1 Fluorescence of retinal-PSB formed in CRBPII mutants

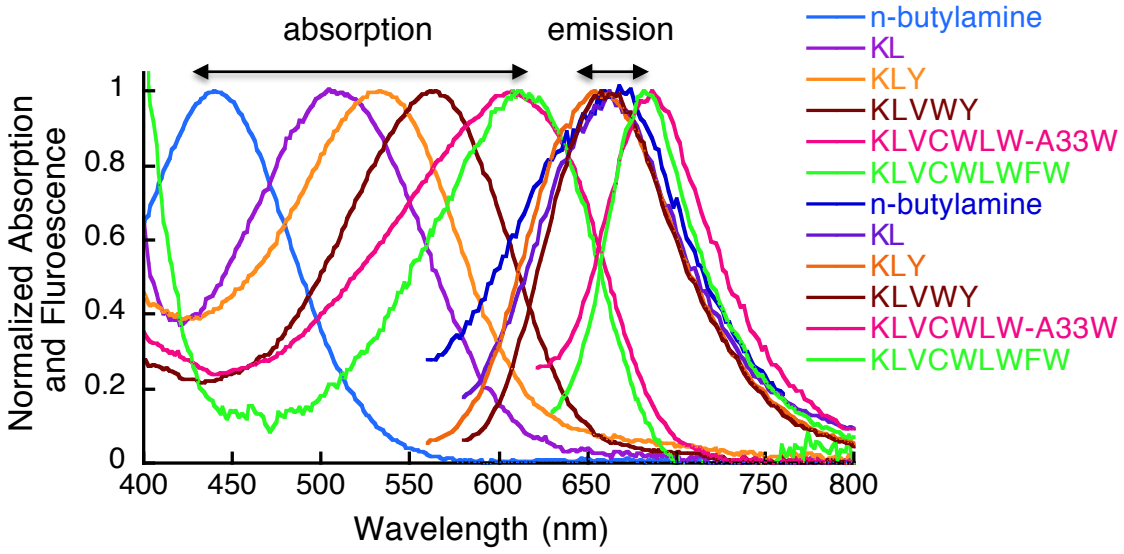


Figure IV-7: Normalized UV-vis and fluorescence spectra of retinal-PSB with n-butylamine and different CRBPII mutants.

The absorption wavelength of retinal-PSB bound to CRBPII mutants can be widely regulated depending on protein-chromophore interactions, including both electronic and steric factors. We were also interested in whether retinal-PSB formed with CRBPII derivatives can be developed into fluorescent tags.

Retinal and retinal-PSB are reported to exhibit low quantum efficiency, around 0.02% in solution. However, recently a rhodopsin based fluorescent probe was reported to study the electronic potential in *E. coli* membranes.[31] The quantum efficiency of the retinal-PSB based fluorescent rhodopsin was reported to be 0.2%, ten times that of the free chromophore in solution. This is possibly due to restricted rotation inside the rhodopsin binding pocket. Due to the large

Stokes shift with the excitation maximum at ~560 nm and emission maximum at 710 nm, the fluorescence was detectable using a sensitive detector (a self built advanced microscope was applied) due to the high signal/noise contrast.

Table VI-1: Fluorescent characterizations of retinal-PSB.

Retinal PSB	Excitation (nm)	Emission (nm)	QY (%)
n-BuNH_2	440	660	0.033
KL	506	614/660	0.029
KLY (KL:**R58Y**)	533	654	0.12
KLVY (KLV:**R58Y**)	561	660	0.14
KLVWY (KLVW:**R58Y**)	563	656	0.14
KLVY-A33W (KL:**R58Y**:A33W)	566	662	0.13
KLVF (KLV:**R58F**)	561	664	0.12
KLVCWLW-A33W	605	684	0.079
KLVCFLWF-A33W	610	676	0.18

Note: Abbreviations for CRBPII mutants. KL (Q108K:K40L), KLY (Q108K:K40L:R58Y), KLVY (Q108K:K40L: T51V: R58Y), KLVWY (Q108K:K40L:T51V:Y19W:R58Y), KLVY-A33W (Q108K: K40L:T51V:R58Y:A33W), KLVF (Q108K:K40L:T51V:R58F), KLVCWLW-A33W (Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:A33W), KLVCFLWF-A33W (Q108K:K40L:T51V:T53C:R58F:T29L:Y19W:Q4F:A33W).

Encouraged by the latter study, fluorescence spectra and quantum efficiencies of a few retinal-bound CRBPII mutants were measured, in order to evaluate whether retinal-CRBPII complexes could be used as fluorescent probes. Surprisingly, although the absorption of the retinal-PSB could be tuned over a wide range of wavelength using different CRBPII mutants, the fluorescence emission maxima do not vary much (**Figure IV-7**).

Similar phenomena have been observed with photoactive yellow fluorescent protein,[32] in which the excitation maxima varied from 441 nm to 478 nm with different mutants, while the emission maxima varied only from 500 nm to 511 nm. The authors of the study suggest that different mutants perturb the width of the excited state surface, without disturbing the S1 energy level. A wider width of the excited state surface will increase the Franck-Condon factor for transitions to higher vibrational energy levels, therefore leading to blue shift and broadening the absorption peak, and vice versa. As the overall S1 energy level does not change, the fluorescence emission does not vary much.

In the case of CRBPII-retinal, the absorption maxima were regulated over an even more dramatic spectral range, from 440 nm to 610 nm, while the emission varied only from 660 nm to 680 nm. The shapes of the absorption spectra do not vary much, indicating that the mechanism in CRBPII-retinal is different from photoactive yellow fluorescent protein.

Table IV-1 lists the quantum efficiency and spectral characteristics of retinal PSB formed with n-butylamine and different CRBPII mutants. There is no discernable correlation between the position of the absorption maximum and the quantum efficiency of the mutant (extinction coefficients were not determined).

Surprisingly, double mutant Q108K:K40L (KL) has similar fluorescence quantum efficiency as the free retinal-PSB in buffer. This is unexpected, as sequestration of the chromophore inside the rigid binding cavity should restrict the non-radiative relaxation processes such as isomerization and heat transfer

with the solvent. Mutations R58Y and R58F increase the quantum efficiency by about 3 fold, while other mutations like T51V and Y19W do not seem to make a difference. It is not clear how R58Y and R58F contribute to the increase of quantum efficiency observed.

The best quantum efficiency obtained, 0.18%, is similar to the quantum efficiency (0.2%) of the rhodopsin, which is used as a fluorescent probe in a recent report.[31] It was hoped that as the absorption is red shifted dramatically compared to free retinal-PSB or retinal, it might also lead to high signal-noise ratio, using excitation wavelength away from 440 nm. With longer exposure time, fluorescent imaging might be possible.

CRBPII protein was expressed in U2-OS cell lines using a CMV promoted vector expressing the fusion protein, EGFP-KLVF(CRBPII)-RB (EGFP: Enhanced Green Fluorescent Protein; KLVF: Q108K:K40L:T51V:R58F-CRBPII; RB: RetinoBlastoma protein). After two days of transfection and observation of nucleus localized green fluorescence, which indicated sufficient amount of fusion protein expression, retinal solution (0.01 M in ethanol, 10 μM, 30 μM and 60 μM final concentration) was added to the cell. A large amount of retinal was added to increase the concentration of retinal-PSB formed with CRBPII mutant Q108K:K40L:T51V:R58F (KLVF). However, retinal at such concentration was found to be toxic to U2-OS cells, as the cell morphology changed and began dying after a few hours. It is well known that vitamin A is important for cell

growth, but an excessive amount is detrimental. Therefore, it is not too surprising that with high concentrations of retinal, cells died.

The toxicity issue can be solved by lowering the concentration of retinal to sub μM levels, however, it would make imaging more difficult due to the low fluorescence signal. Generally speaking, retinal is not suitable as a fluorophore precursor for our system, if comparative fluorescence to GFP is desired.

IV.4.2 Characterization of azulene polyene bound to different CRBPII mutants

Besides retinal, a few other chromophores were also characterized with different CRBPII mutants (**Figure IV-8a**) (For synthesis, please refer to Dr. Lee's thesis[33]). Similar to retinal, the other three chromophores also contain the long polyene-aldehyde moiety for binding to the protein. The azulene polyene analogue (Azu) has a less electron-donating head group compared to merocyanine analogues, which have an electron-rich nitrogen to form a push-pull system.

Figure IV-8b illustrates that one of the dominant resonance structures of Azu, which generates an aromatic seven-member ring, exhibits a negative charge on the carbon that is in direct conjugation to the polyene system. The negative charge could be delocalized towards the positively charged iminium end through resonance structures, once the PSB is formed.

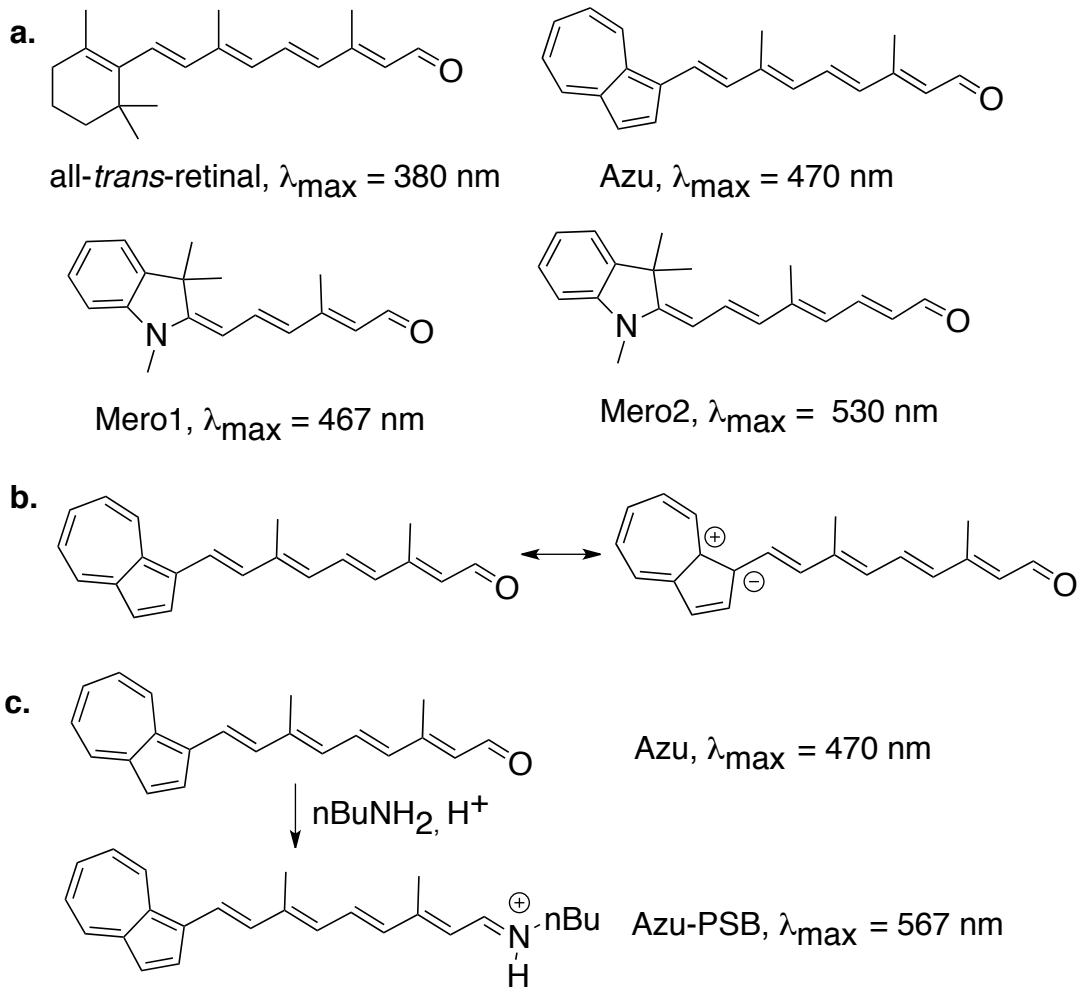


Figure IV-8: **a.** Chemical structures of different chromophores. **b.** Resonance structure of Azu head group. **c**. Formation of protonated Schiff base of Azu with n-butylamine.

The azulene head group by itself is fluorescent with quantum efficiency of ~5% in hexane,[34] due to its rigid structural scaffold. Similar compounds to Azu (referring to **Figiure IV-8**) with the same number of double bonds have been synthesized bound to bacteriorhodopsin (BR) and characterized. The BR-complex absorbs at 750 nm.[35] Thus the Azu chromophore could be an interesting target as a near-IR fluorescent tag if the quantum efficiency is high.

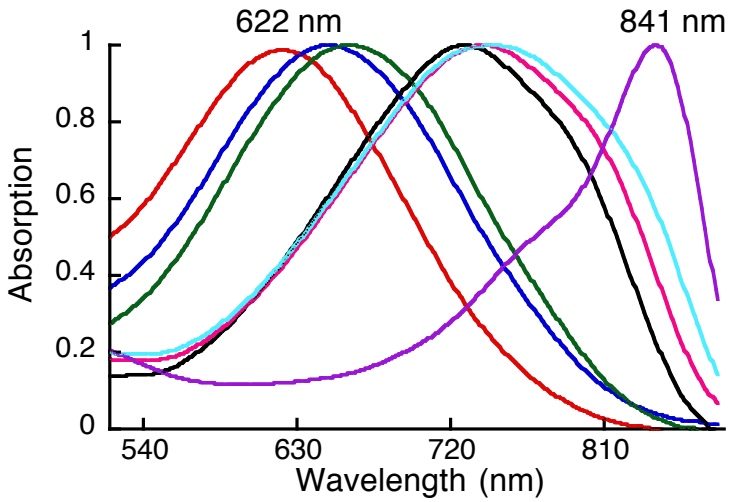


Figure IV-9: UV-vis spectra of azulene polyene aldehyde (Azu) with CRBPII mutants, corresponding to CRBPII mutants in **Table IV-2**.

Azu red shifts ~100 nm upon formation of PSB with n-butylamine as shown in **Figure IV-8c**. This red shift is crucial to avoid the background fluorescence resulting from unbound aldehyde. Double mutant Q108K:K40L red shifts the absorption spectrum further from 567 nm to 622 nm. A large protein shift is desired, so that even if there is non-specific binding of the aldehyde to amine groups inside the cells, excitation at a longer wavelength can circumvent the fluorescence from non-specific binding. Furthermore, incubation of Azu with different CRBPII mutants results in a wide range of absorption spectra, from 622 nm to 841 nm (**Figure IV-9**). The most red shifted mutant is 91 nm more red shifted than a similar chromophore bound with bacteriorhodopsin.[35]

As expected, a similar trend for wavelength regulation is observed for retinal and Azu as shown in **Table IV-2**. Comparison of the protein shift of mutations introduced into Q108K:K40L (KL) for retinal and azulene-polyene indicates that mutations in CRBPII have similar effects on retinal and Azu. This

further supports our hypothesis that it is the electrostatic interactions between the protein and the chromophore that perturb the energy gap from the ground state to the excited state of the chromophore. If conformational change plays the major role in spectral turning, it is hard to imagine how the same mutants could result in similar degree of conformational change, which lead to similar protein shift for two different chromophores.

Table IV-2: Summary of CRBPII mutants with retinal and Azu.

CRBPII Mutant	λ_{max} (nm)		protein shift compared to KL (cm^{-1})	
	Azu	retinal	Azu	retinal
KL	622	508	0	0
KLY	651	535	716	993
KLCY	663	540	994	1166
KLVCY	732	576	2415	2323
KLVCF	741	571	2581	2171
KLVCW	747	585	2690	2591
KLVCWLWF	841	613	4186	3371

Note: abbreviations for CRBPII mutants. KL, Q108K:K40L; KLY, Q108K:K40L:R58Y; KLCY, Q108K:K40L:R58Y:T53C; KLVCY, Q108K:K40L: R58Y:T53C:T51V; KLVCF, Q108K:K40L:R58F:T53C:T51V; KLVCW, Q108K: K40L:R58W:T53C:T51V; KLVCWLWF, Q108K:K40L:R58W:T53C:T51V:T29L: Y19W:Q4F.

Disappointingly, strong fluorescence was not observed for the Azu bound with CRBPII mutants. Different derivatized azulene head groups by themselves have quantum yields up to 20%, while the quantum yield of the Azu bound to

CRBPII mutants is estimated to be lower than 0.2% (worse than retinal). This might indicate that the flexible polyene could be leading to non-radiative decay.

Overall, Azu compounds behave similarly to retinal with respect to wavelength regulation, rate of formation and quantum efficiency of fluorescence. At this stage, they are not suitable to be used as chromophores for a fluorescent tag.

IV.4.3 Characterization of Mero1 with CRBPII mutants

Cyanine dyes are known for their fluorescent properties,[36] contributed partially from the push-pull system. Once Mero1 forms an iminium with the protein, it becomes a cyanine dye-like chromophore. Therefore, Mero1 is highly likely to be fluorescent when bound to CRBPII mutants.

The aldehyde form of Mero1 absorbs at 467 nm in ethanol and it is red-shifted 109 nm upon formation of PSB with n-butylamine in PBS buffer solution (**Figure IV-10**). Mero1 by itself has a high extinction coefficient, 86,000 M^{-1} cm^{-1}, which is important for its brightness (brightness = $\varepsilon \times \Phi$). Once Mero1 forms the iminium, the absorption peak becomes much sharper and the extinction coeffient increases by almost two fold compared to the aldehyde form (**Figure IV-10c**). This is attributed to the increased conjugation due to better stabilization of the resonance structures as shown in **Figure IV-10b** when Mero1 forms a protonated Schiff base.

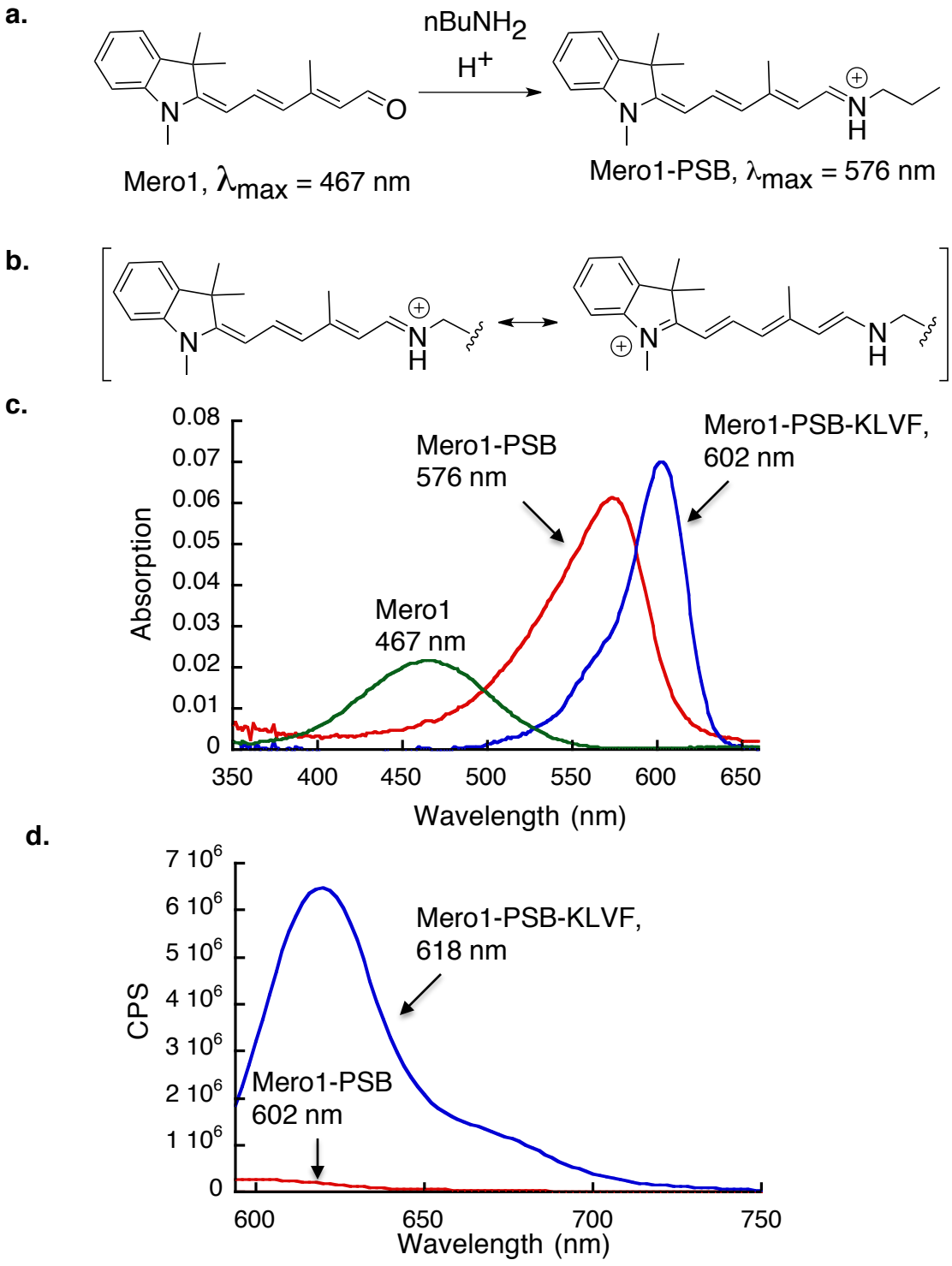


Figure IV-10: Characterization of Mero1. **a.** Formation of Mero1-PSB; **b.** Resonance structure of Mero1-PSB; **c.** UV-vis spectra of Mero1, Mero1-PSB and Mero1-PSB formed with CRBPII mutant KLVF (Q108K:K40L:T51V:R58F); **d.** Fluorescence spectra of Mero1.

Two dominant resonance structures of Mero1-PSB are illustrated in **Figure IV-10b**. The positive charge is delocalized towards the two nitrogen ends. As a result, the π bonds are more delocalized over the polyene carbons, which restricts the rotation along the single bonds and thus limits the number of vibrational energy levels and makes the peak sharper. Restriction of rotation along the polyene system also increses the population of the planar form due to less single/double bond alternation, thus increasing the absorption of the planar form. Mero1 with CRBPII mutant KLVF generates a similar amount of increase in extinction coefficient and a small red shift (**Figure IV-10c**).

Incubation of different CRBPII mutants with Mero1 did not result in wavelength regulation, which was observed with retinal and azulene analogue. It is disappointing but not unexpected that Mero1-PSB has its positive charge fully delocalized along the polyene. The mechanism for CRBPII mutants to cause dramatic red shift compared to free retinal-PSB in solution is by promotion of the positive charge delocalization from the iminium end towards the ionone ring end through electrostatic interactions. However, in the case of Mero1-PSB, the positive charge is already delocalized on the two nitrogens on both ends. Therefore, the polarity changes in the protein cavity of existing CRBPII mutants do not exert a significant impact on the absorption profiles of Mero1-PSB.

Background fluorescence introduced through non-specific binding of Mero1 with amine-containing functional groups, such as lysine residues or terminal amine groups could pose a problem. This is because the absorption of

Mero1 bound with CRBPII mutants is only 20 to 30 nm more red shifted compared to Mero1-PSB. Background fluorescence can not be avoided by exciting at a much longer wavelength than 576 nm, the absorption of the Mero1-PSB formed non-specifically. With the available confocal laser resources at MSU, the most optimal excitation wavelength is at 594 nm. It is further away from 576 nm, but still close to the excitation maximum of Mero1-PSB formed with CRBPII mutants. Therefore, *in vitro* fluorescence assays described below were excited at 594 nm.

Table IV-3: Characterization of Mero1-PSB.

	Mero1-$nBuNH_2$	KL-Mero1	KLV-Mero1	KLVF-Mero1	mRaspb-erry
λ_{max} (nm)	576	600	599	603	598
Emission (nm)	602	618	616	618	625
ε_{594nm} (M^{-1} cm^{-1})	140,000	260,000	240,000	230,000	86,000
QY	1.2%	8.4%	15%	18%	15%
[a]B/1000	1.7	22	36	41	13
B relative to mRaspberry	0.13	1.7	2.8	3.1	1.0
K_d (nM)	/	35 ± 21	632 ± 306	66 ± 32	/

Note: abbriviations of CRBPII mutants. KL, Q108K:K40L; KLV, Q108K:K40L:T51V; KLVF, Q108K:K40L:T51V:R58F. [a]B (Brightness) = ε_{594nm} x QY.

Satisfyingly, it was found that Mero1 bound to CRBPII mutants exhibited much higher quantum yield than the Mero1-PSB formed with n-butylamine in solution. The summary of a few CRBPII mutants bound with Mero1 is tabulated in **Table IV-3**. When Mero1 is bound to double mutant Q108K:K40L (KL), the quantum efficiency increases by seven fold. This is probably a result of chromophore sequestration from the bulk solvent, reducing the non-radiative relaxation through heat transfer in the binding pocket.

It is interesting to note that introduction of T51V increases the quantum efficiency by almost two fold to 15%. The crystal structures of Q108K:K40L (KL) and Q108K:K40L:T51V (KLV) bound with Mero1 were determined (Camille Watson, Babak Borhan's lab, MSU) in order to compare the different protein-chromophore interactions that might lead to the increased quantum efficiency. There are two molecules in the asymmetric units of both crystal structures, each of which is termed chain A and B. The chromophore densities in chain B for both crystal structures are more occupied, therefore comparisons were made for chain B from both crystal structures.

As shown in Figure **IV-11a**, the protein backbone structures of the two CRBPII mutants, Q108K:K40L (KL) and Q108K:K40L:T51V (KLV) bound with Mero1, overlay pretty well except at position 51. Mutation of a polar residue Thr51 into a hydrophobic residue, Val51, causes a slight inward movement of Val51 towards the chromophore. This possibly leads to tighter packing of the protein with the chromophore.

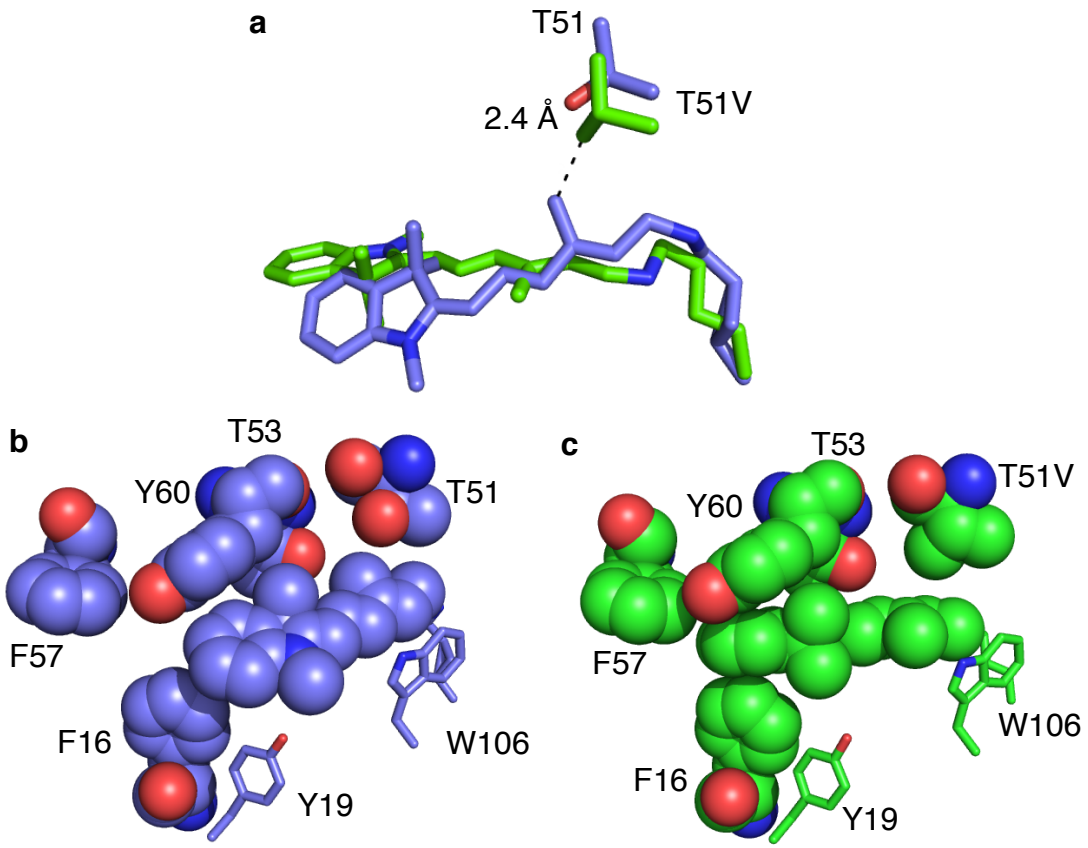


Figure IV-11: Comparison of Q108K:K40L (KL) and Q108K:K40L:T51V (KLV) bound with Mero1. **a.** Overlaid crystal structures of KL and KLV bound with Mero1. Crystal structures of KL (b) and KLV (c) with space filling models shown for the surrounding residues.

As shown in **Figure IV-11a**, the inward movement of Val51 drives Mero1 to change its geometry to avoid steric clash. Consequently, the chromophore is sandwiched between residues Tyr60 and Phe16 as shown in **Figure IV-11c**, as compared to the relatively open cavity in double mutant Q108K:K40L as shown in **Figure IV-11b**. Such tight packing as a result of the T51V mutation greatly rigidifies the chromophore, resulting in less non-radiative decay and thus higher quantum efficiency.

CRBPII mutant Q108K:K40L:T51V:R58F (KLVF) bound with Mero1 yields the highest quantum yield. The brightness of KLVF-Mero1 is 24 fold higher than that of free Mero1-PSB in solution (**Table IV-3**), due to a combination of higher quantum yield and higher extinction coefficient at 594 nm. Comparison of KLVF-Mero1 with one of the bright red fluorescent protein with similar excitation and emission, mRasperry, shows that our fluorescent protein tag is more than three fold brighter than mRasperry. This provides an alternative to red fluorescent protein with higher brightness.

With enhanced brightness and high binding affinity of Q108K:K40L:T51V:R58F (KLVF) with Mero1, it might be possible to avoid background fluorescence due to the large contrast in brightness with low concentrations of Mero1 needed.

Besides brightness, faster kinetics is also required for CRBPII to be used as a fluorescent protein tag. Shown in **Figure IV-12** and **Figure IV-13** is the UV-vis kinetic studies of Mero1-PSB formation with three CRBPII mutants Q108K:K40L (KL), Q108K:K40L:T51V (KLV) and Q108K:K40L:T51V:R58F (KLVF) at 37 °C. KL has a much faster PSB formation kinetics ($t_{1/2}$=4.2 min) than KLV ($t_{1/2}$=26 min) and KLVF ($t_{1/2}$=28 min), indicating mutation T51V slows down the formation of PSB with Mero1.

The slow PSB formation kinetics can be due to either the binding or the PSB bond formation. The UV-vis spectra of CRBPII mutants incubated with Mero1 indicate that binding of Mero1 happens right away. A red shift results immediately after Mero1 is added to Q108K:K40L:T51V (KLV, 501 nm) and Q108

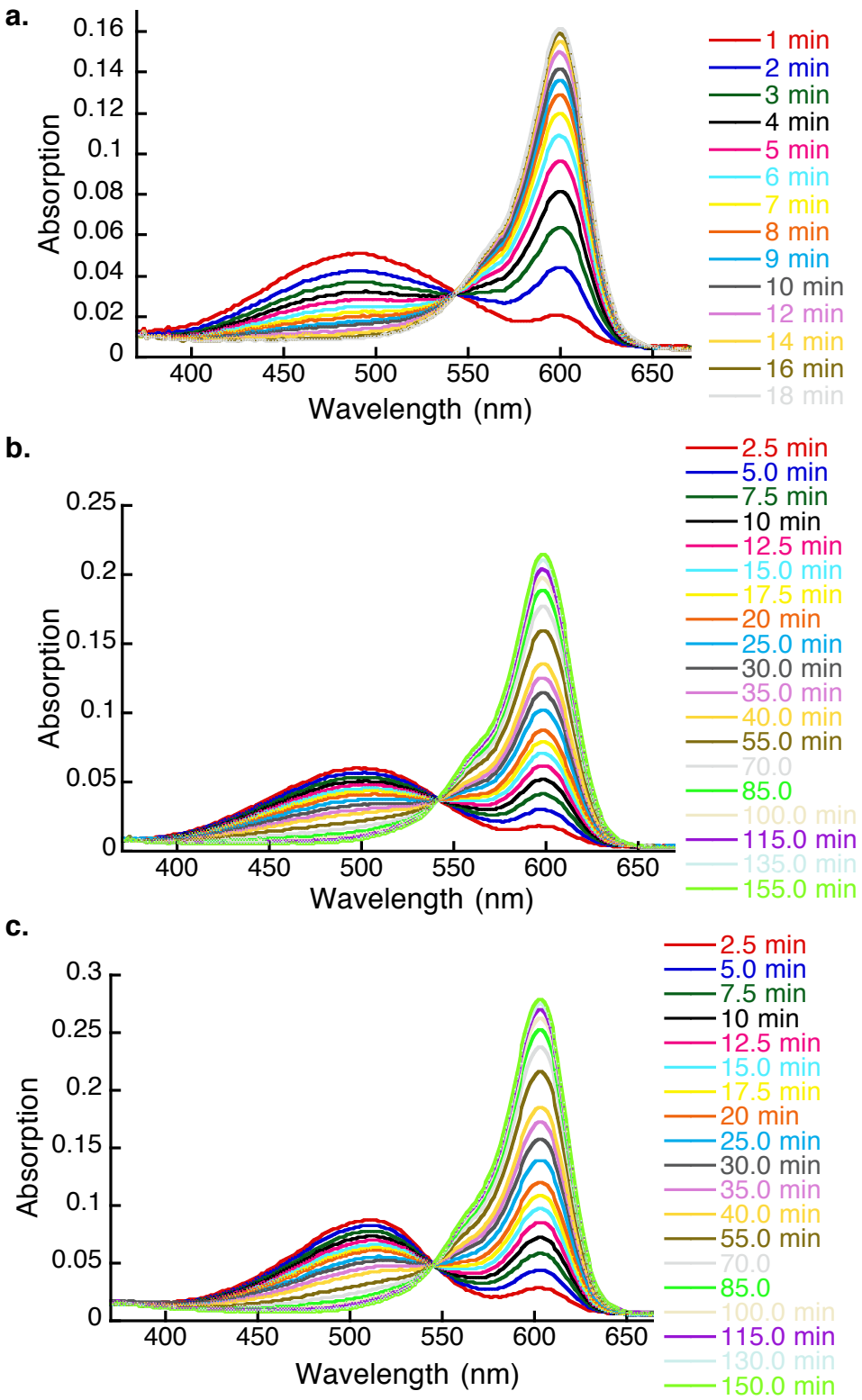


Figure IV-12: UV-vis studies of Mero1-PSB formation with **a.** Q108K:K40L; **b.** Q108K:K40L:T51V; **c.** Q108K:K40L:T51V:R58F.

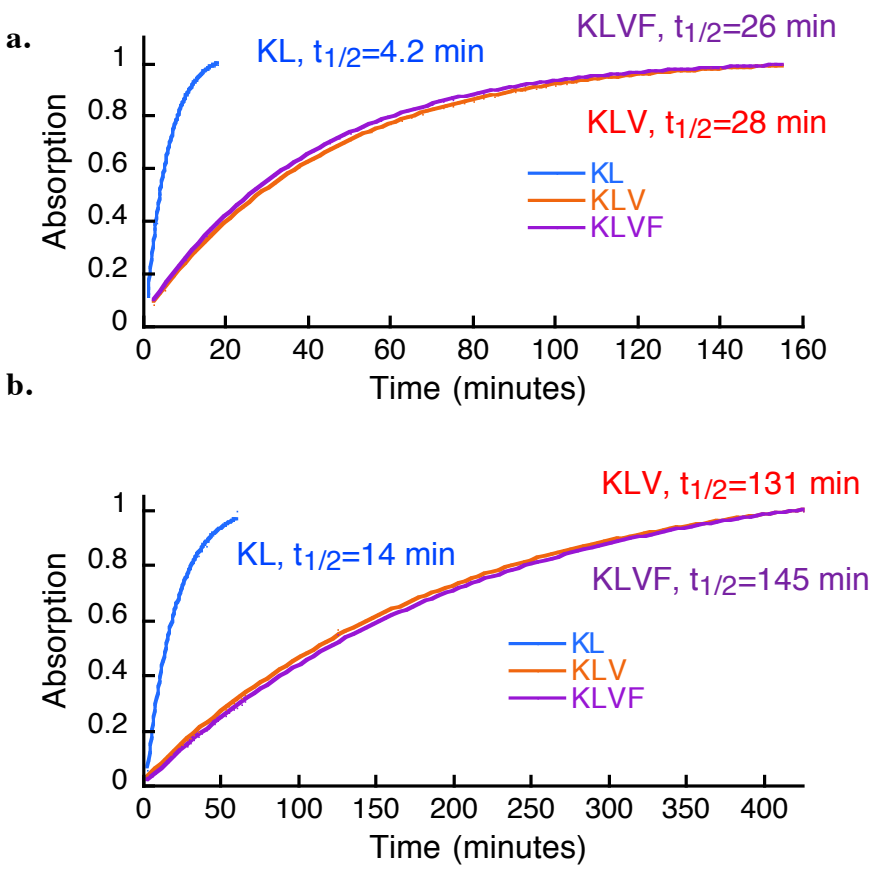


Figure IV-13: Kinetics of Mero1-PSB formation with KL (Q108K:K40L); KLV (Q108K:K40L:T51V); and KLVF (Q108K:K40L:T51V:R58F) at 37 °C (a) and RT (b).

K:K40L:T51V (KLVF, 513 nm) as shown in **Figure IV-12b** and **c**, as compared to free Mero1 in PBS buffer (489 nm). Only when Mero1 is captured in the binding pocket of CRBPII mutants will it exhibit a different spectrum from Mero1 in PBS buffer. As the rate of Mero1 binding is fast, the rate limiting step must be the PSB bond formation.

It is interesting to note that the maturation half life of Q108K:K40L:T51V (KLV) and Q108K:K40L:T51V:R58F (KLVF) is around six times longer than Q108K:K40L (KL) at 37 °C (**Figure IV-13**a). This demonstrates the important

role of T51 in facilitating PSB formation. It is likely that T51 assists the formation of PSB by hydrogen bonding to Mero1 aldehyde, activating the aldehyde or orienting the aldehyde in a better position for the lysine to attack.

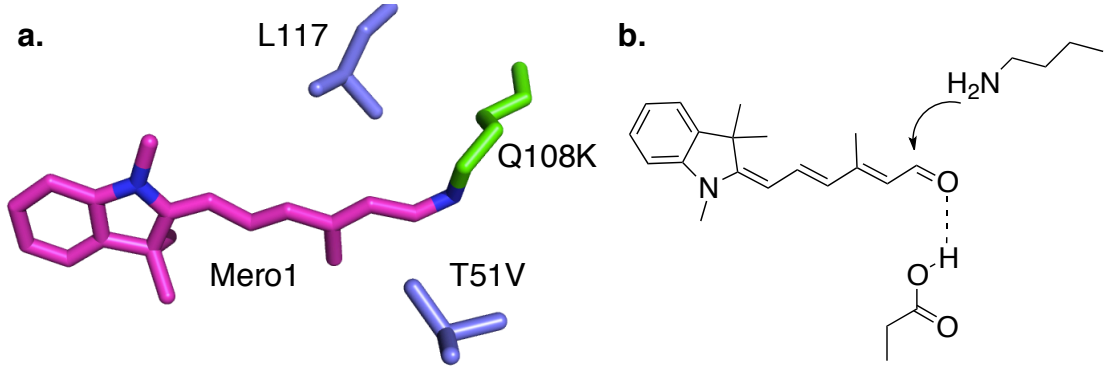


Figure IV-14: Proposed mechanism for Mero1 aldehyde activation by introduction acidic residues at positions 117 and 51. **a.** Crystal structure of Q108K:K40L:T51V, with T51V and L117 highlighted; **b**. Proposed mechanism for activation of aldehyde for electrophilic attack through introduction of acidic residues in position 51 and 117.

The role of T51 in PSB formation is further supported by mutants Q108K:K40L:**T51D**:R58W:Y19W and Q108K:K40L:T51V:T53C:R58W:T29L:Q4F:A33W:**L117E** discussed later. Acidic amino acid, T51D and L117E, are introduced in each of the latter two mutants, at a position close to the putative aldehyde binding site (**Figure IV-14a**). Both mutations facilitate the PSB formation considerablly and consequently the PSB can be formed completely within minutes at room temperature as shown in **Figure IV-15**.

The proposed mechanism is illustrated in **Figure IV-14b.** Acidic residues in these two positions could donate the acidic hydrogen to the aldehyde through hydrogen bonding to activate the aldehyde. That same hydrogen bonding could also help the aldehyde orient at an optimal position for nucleophilic lysine to attack.

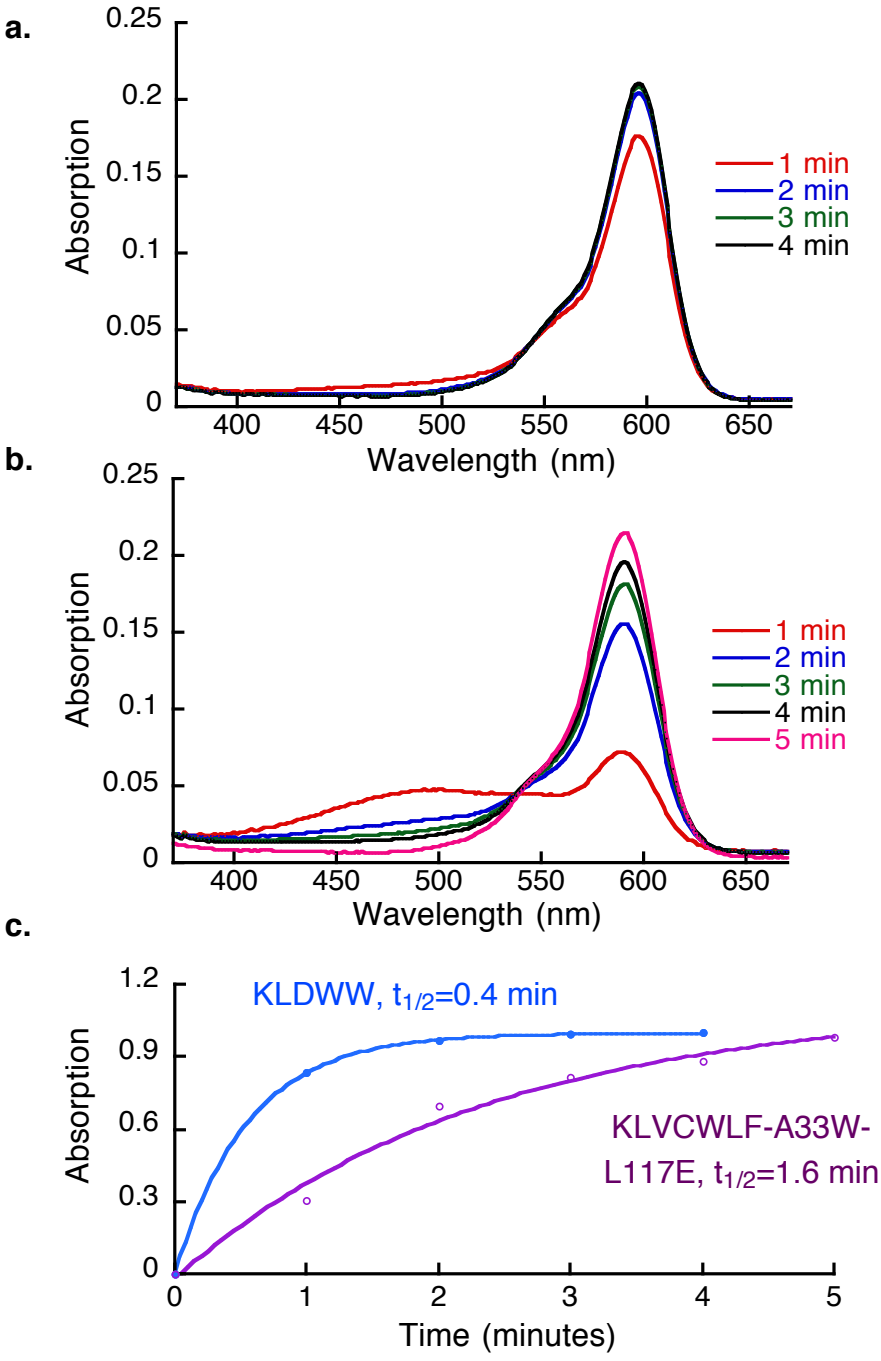


Figure IV-15: Kinetics of Mero1-PSB formation with (a) KLDWW (Q108K: K40L:T51D:R58W:Y19W) and (b) KLVCWLF-A33W-L117E (Q108K:K40L: T51V:T53C:R58W:T29L:Q4F:A33W:L117E) at RT. (c) Comparison of KLDWW and KLVCWLF-A33W-L117E maturation kinetics.

Q108K:K40L:**T51D**:R58W:Y19W and Q108K:K40L:T51V:T53C:R58W:T29L:Q4F:A33W:**L117E** show over 100 fold faster PSB formation kinetics than Q108K:K40L:T51V:R58F (KLVF). Also, the quantum efficiency for the two were determined to be 16% and 18%, respectively, which is similar to KLVF (18%). However, the expression levels of Q108K:K40L:**T51D**:R58W:Y19W (~10 mg/L) and Q108K:K40L:T51V:T53C:R58W:T29L:Q4F:A33W:**L117E** (~40 mg/L) are not as good as Q108K:K40L:T51V:R58F (KLVF, ~100 mg/L) in *E. coli*. at 16 °C. As eukaryotic cells have different protein expression and folding systems from prokaryotic cells, these two mutants might still be able to be expressed in sufficient amounts in eukaryotic cells at 37 °C for imaging.

Another interesting observation is that PSB formation is highly temperature dependent as shown in **Figure IV-13**. Comparison of maturation half time for PSB formation at RT and 37 °C for different CRBPII mutants shows that it takes three to four fold more time for the maturation to happen at RT. This could be explained by the Arrhenius equation, considering the protein structure is stable at both temperatures and the reaction is going through the same mechanism and energy barrier.

$$k = Ae^{\frac{-E}{RT}}$$

(A refers to frequency factor, k refers to rate constant, E refers to activation energy, R refers to the gas constant, T refers to Kelvin temperature)

According to the Arrhenius equation, if the activation energy remains the same, higher temperature will lead to faster rate constant. Therefore, at 37 °C

the maturation time is faster. In the following *in vivo* assay, the cells will be incubated at 37 °C after Mero1 is added for faster PSB formation kinetics.

IV.5 Fluorescent microscopic assay based on *E. coli* cells

With the bright CRBPII mutant Q108K:K40L:T51V:R58F (KLVF)-Mero1 in hand, *in vivo* assays in prokaryotic cells were first carried out to investigate the possibility of using KLVF-CRBPII as a fluorescent protein tag.

E. coli cells expressing Q108K:K40L:T51V:R58F (KLVF) and Q108K:K40L (KL) were subjected to wide-field fluorescent microscopic studies along with the control studies of *E. coli* cells without any transfection and *E. coli* cells expressing WT-CRBPII, which can not form PSB with Mero1. Four hours after induction of protein expression at 26 °C, Mero1 was added to a final concentration of 3 μM and then incubated at 37 °C with vigorous shaking. The cells were monitored at different time points in a wide-field fluorescent microscope. Since the available filter for the wide-field fluorescent microscope in our lab provides only blue and green light excitation, green light was used to excite the processed cells. However, the wavelength of green light is ~540 nm, where Q108K:K40L: T51V:R58F (KLVF)-Mero1 has low absorption. This is not the optimal set up for fluorescent microscopic studies.

As shown in **Figure IV-16**, it is obvious that both control cell lines, *E. coli* without transfection vector and *E. coli* with WT-CRBPII are barely fluorescent compared to *E. coli* cell lines expressing Q108K:K40L (KL) and

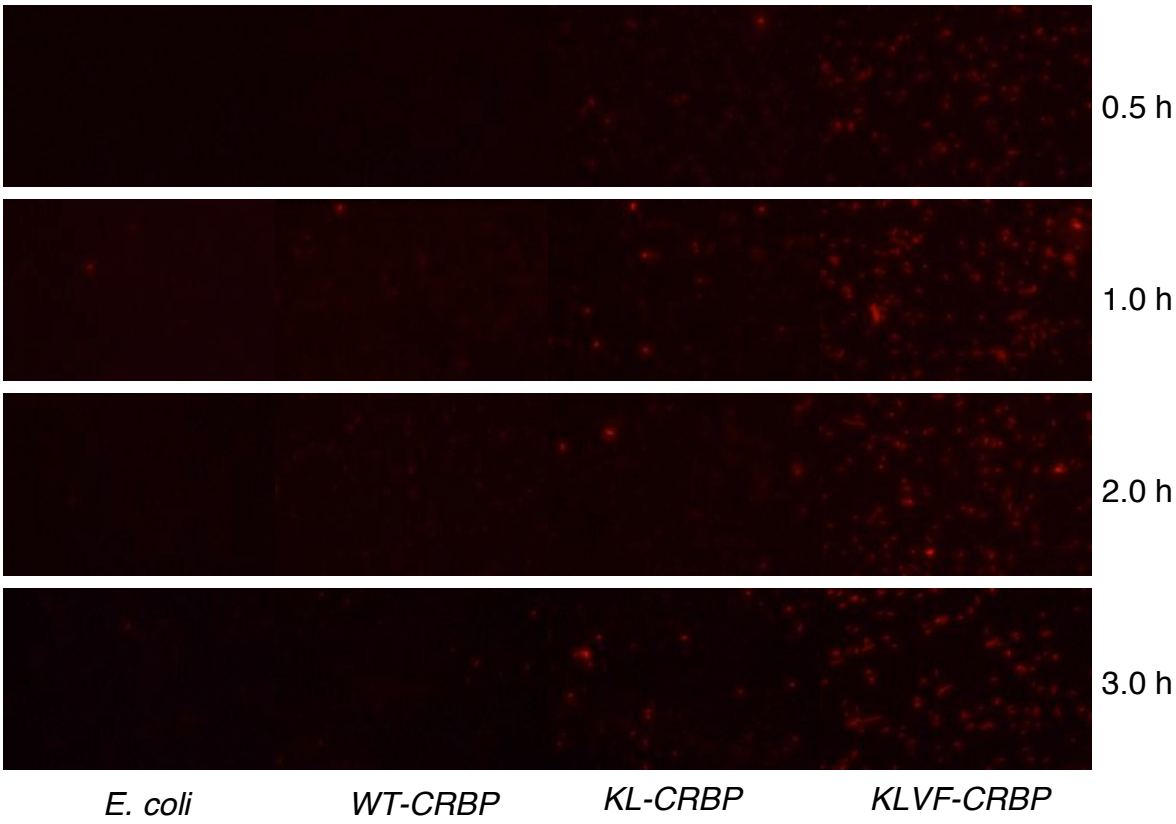


Figure IV-16: Wide-field fluorescent microscopic pictures of *E. coli* cells with and without different CRBPII construct at different time points. Green laser light was used for excitation and a red light filter was used for fluorescence emission cut off.

Q108K:K40L:T51V:R58F (KLVF). In *E. coli* expressing WT-CRBPII, a few dim specs are observed, but they are not comparable to *E. coli* expressing Q108K:K40L (KL) and Q108K:K40L:T51V:R58F (KLVF). Comparison of KLVF with KL shows that KLVF exhibits a brighter fluorescence, due to its higher quantum efficiency.

Although the excitation wavelength in the widefield microscope is not optimal for KLVF-Mero1, intense bright fluorescence is observed, which is partially because of the high protein expression level in *E. coli* (up to 100 mg/L

LB). If we assume the protein expression level is 50 mg/L LB and the cell density is ~10^9/mL, then each *E. coli* cell has the following amount of protein:

$$50\ \mathrm{mg}/(1000\ mL * 10^9\ /mL) = 5 * 10^{-11}\mathrm{mg}$$

For an average-sized *E. coli* cell (2 μm long and 0.5 μm in diameter), its volume is ~0.4 μm^3, therefore the local concentration of the CRBPII mutant protein inside the *E. coli* cell can be estimated to be:

$$\frac{(5 * 10^{-11}\mathrm{mg} * \frac{10^{-3}g}{mg})/16{,}000Da}{0.4\ \mu m^3 * 10^{-15}\frac{L}{\mu m^3}} = 8\ mM$$

With such high local concentration of CRBPII mutant protein, Q108K:K40L:T51V:R58F (KLVF), inside the cells and the high affinity toward Mero1, the formation of PSB with KLVF is presumably faster than non-specific binding. Obvious fluorescence could be observed 30 min after Mero1 was added for *E. coli* cells expressing KLVF and continues to become brighter after 1 h, remaining constant afterwards. This agrees well with the maturation half time of KLVF-Mero1.

This preliminary study demonstrates that Mero1 can penetrate the membrane easily and the high binding affinity of Q108K:K40L:T51V:R58F (KLVF) towards Mero1 along with much more enhanced fluorescence compared to that of nonspecific binding can minimize background fluorescence to a negligible extent in prokaryotic systems. Due to the high expression level of KLVF and much higher brightness of KLVF complexed with Mero1, no washing

was needed in *E. coli* studies. Encouraged by this result, we proceeded to studies in eukaryotic cells.

IV.6 In vivo imaging of KLVF-Mero1 in a mammalian cell line

To prove that Q108K:K40L:T51V:R58F (KLVF)-Mero1 can be used as a general fluorescent protein tag, fluorescent assays in mammalian cell lines were carried out. For better visualization of the red fluorescence from KLVF-Mero1, the protein was to be localized in a targeted compartment to distinguish the localized fluorescence from background fluorescence. In order to achieve this localization, the protein was inserted into a construct of pEGFP-RB in between EGFP and RB (RetinoBlastoma) protein. EGFP helps visualization of the expression of KLVF. RB protein is a tumor suppressor protein that plays an important role in cell cycle, suppressing cell division.[37] It is localized inside the nucleus, is a large protein, with 928 amino acids. Fusion of KLVF with RB protein will direct the localization of the protein to the nucleus. At the same time, construct PEGFP-KLVF was also made by inserting a stop code after the KLVF gene, so that expression terminates before the RB protein. With this construct, the protein will be scattered throughout the whole cells.

Human osteosarcoma cells (U2-OS) were transiently transfected with fused EGFP-CRBPII (KLVF)-RB construct, which is under the control of the CMV promoter. Forty eight hours after transfection, Mero1 (1 mM, in ethanol) was

added to a final concentration of 0.25 μM. Surprisingly, minutes after Mero1 was added, bright red fluorescence was observed in the cytosol of all the cells under wide-field fluorescent microscope with green light excitation. Since EGFP was coexpressed, green fluorescence localization indicates which cells are successfully transfected. The intense red fluorescence lighting up right away is not due to KLVF-Mero1, but non-specific binding with amine-containing molecules in the cytosolic solution.

This is different from that in *E. coli* cells, where background fluorescence from nonspecific binding is not obvious compared to the fluorescence from KLVF-Mero1. Two factors could be accounting for the obvious background fluorescence generated in mammalian cells. One is that there are specific compartments in eukaryotic cells with basic or acidic environments, but not in *E. coli*. These environments can catalyze the formation of Mero1 with amine containing groups. The other factor is that the excellent protein expression level in *E. coli* results in much higher local concentration of CRBPII mutant protein to compete with the non-specific binding. As a result, the relative concentration of Mero1-PSB formed with CRBPII mutant is higher than the non-specific binding.

Also, the green light excitation is not optimal for KLVF-Mero1 in order to avoid the background fluorescence. As shown in **Figure IV-17**, within the window of green light excitation, Mero1-PSB from non-specific binding has a much higher absorption extinction coefficient than that of KLVF-Mero1. Therefore, although KLVF-Mero1 has higher quantum efficiency, the low

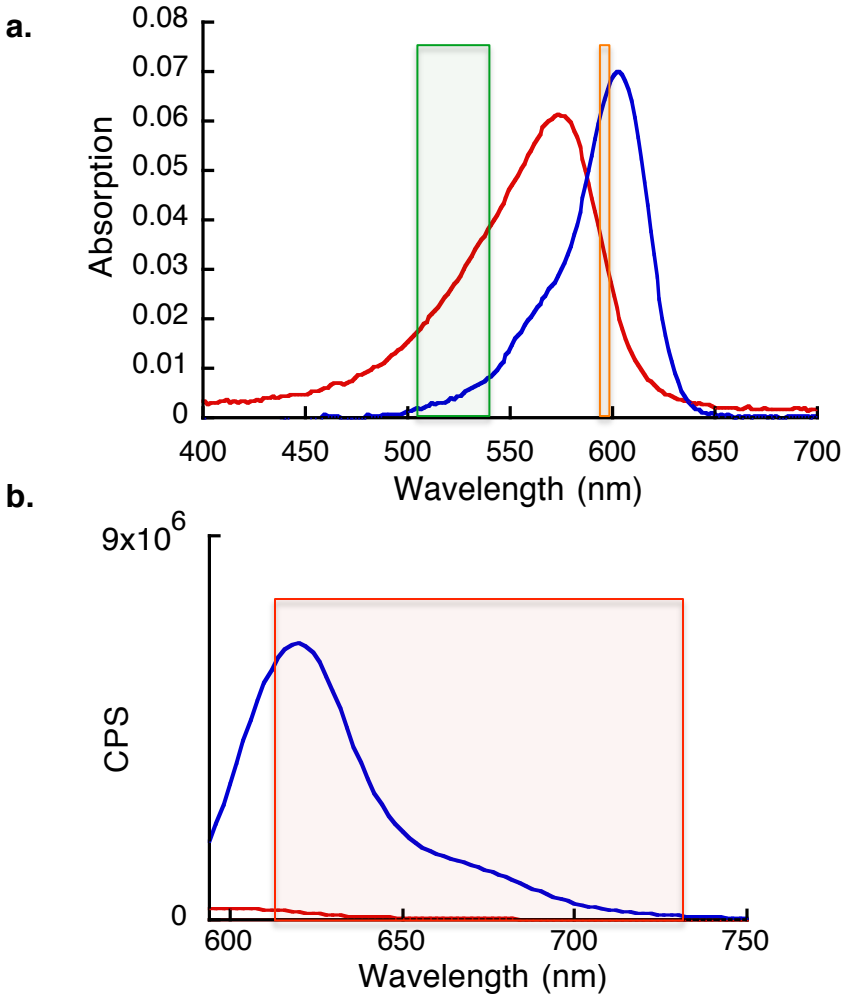


Figure IV-17: Illustration of optimal excitation and emission filter. **a.** UV-vis spectrum of Mero1-PSB formed with n-butylamine (red) and CRBPII mutant KLVF (Q108K:K40L:T51V:R58F, blue) and showing roughly two different excitation light. Green box indicated the light source with green light filter used in our wide-field microscope and orange box indicated confocal laser light at 594 nm. **b.** Emission spectra of Mero1-PSB with n-butylamine (red) and KLVF (blue). The red box indicated roughly the emission cut off filter.

absorption extinction coefficient in the green light region makes the background fluorescence significant, if a large amount of non-specific binding is present.

With a confocal laser microscope, excitation laser light can be set at 594 nm, which favors KLVF-Mero1 as shown in **Figure IV-17**. Besides, emission

filter that cut off light shorter than 615 nm can be used to favor KLVF-Mero1 and minimize the background fluorescence as well.

In order to minimize the background fluorescence from non-specific binding, sub-micromolar concentration of Mero1 dye was added to the cell culture. The dye was washed away after two to four hours of incubation, to ensure complete formation of KLVF-Mero1-PSB and prevention of further non-

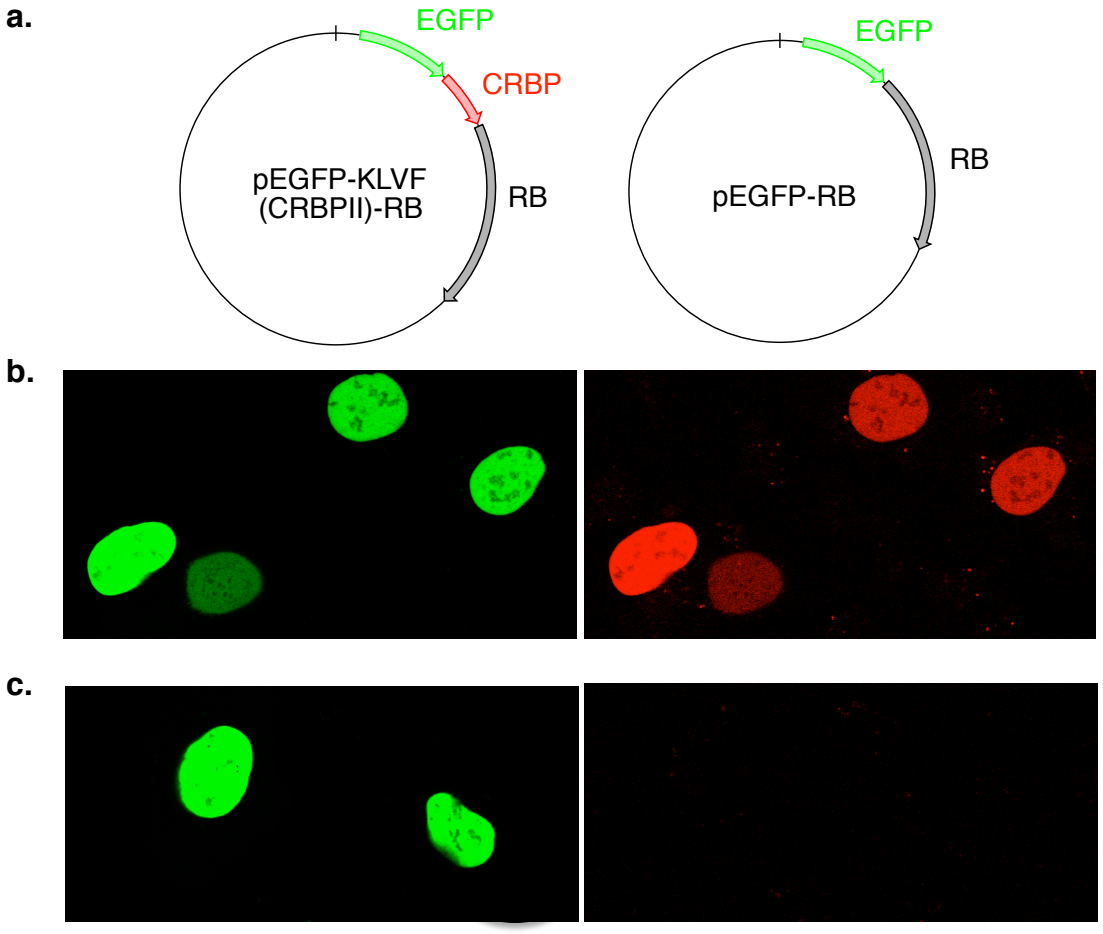


Figure VI-18. *In vivo* fluorescent microscopic studies. **a.** Constructs of pEGFP-KLVF(CRBPII)-RB and pEGFP-RB. Confocal microscopic pictures of U2-OS cells transfected with construct pEGFP-KLVF(CRBPII)-RB **(b)** and the control pEGFP-RB **(c)**. EGFP was excited at 488 nm and KLVF-Mero1 was excited at 594 nm. Red fluorescence was observed only when CRBPII was included in the construct with merocyanine aldehyde added. Negligible background fluorescence was observed in the control study.

specific binding. It is interesting to note that further incubation of the cells at 37 °C diminished the background fluorescence slowly, which was also observed under wide-field fluorescent microscope. After overnight incubation, most of the background fluorescence had cleared; consequently the localized red fluorescence from KLVF-Mero1 was clearly evident in the nucleus. Confocal microscopic pictures were scanned 12 h after washing away the chromophore.

As shown in **Figure VI-18**, irradiation at 488 nm, which excites EGFP, lights up green fluorescence localized in the nucleus, while irradiation at 594 nm, the optimal excitation wavelength for KLVF-Mero1 complex, lights up red fluorescence. Co-localization of the green and red fluorescence clearly shows that KLVF-Mero1 could be used as a red fluorescent protein tag. As a control, when the cells were transfected with EGFP-RB fused construct without KLVF-CRBPII, only green fluorescence was observed. It is exciting to note that the background fluorescence was barely visible at this stage.

A photobleaching assay was done to compare the photostability of KLVF-Mero1 fluorescent complex with EGFP. Cells were transfected with EGFP-KLVF without any localization peptides, thus the fluorescence will be seen all over the cells. Cells were irradiated with 488 nm Argon laser light and 594 nm Helium-Neon laser light simultaneously and microscopic pictures were taken every 30 sec. Green fluorescence and red fluorescence from one cell were plotted against time as shown in **Figure IV-19**. EGFP exhibits stronger fluorescence and better signal/noise contrast versus KLVF-Mero1. However, with signal/noise contrast of

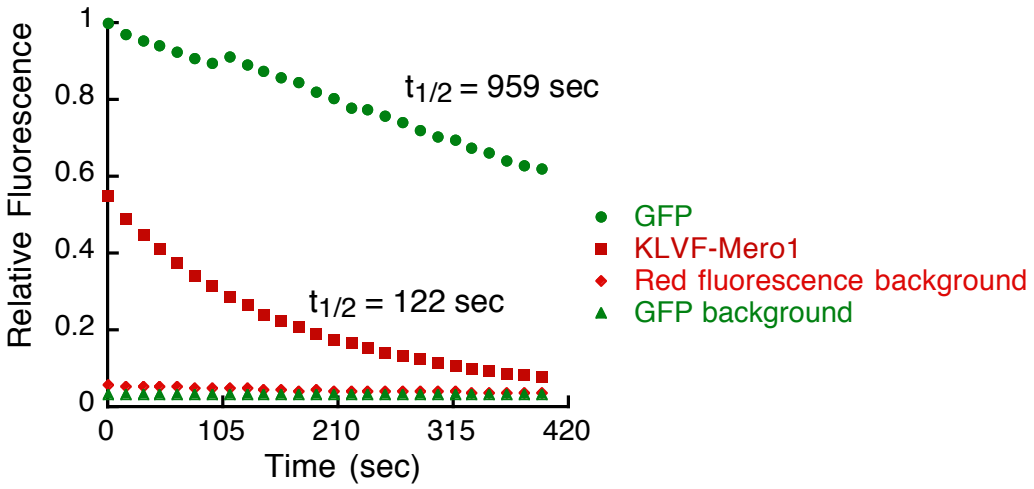


Figure IV-19: Photobleaching studies of U2-OS cells transfected with EGFP-KLVF.

~7 for red fluorescence, the localization of the fluorescent protein was clearly visible.

Although KLVF-Mero1 has a slightly higher brightness compared to EGFP from *in vitro* experimental data, *in vivo* imaging in live cells showed the opposite. This is probably due to two reasons. One of them is that the chromophore is washed away 12 h before microscopic pictures are taken. During the 12 h time window, new EGFP protein is expressed, making the concentration of EGFP higher than that of KLVF-Mero1. The second reason is that the emission filter used for KLVF-Mero1 is a 615 nm longpass filter, which cuts off a part of the emission spectrum of KLVF-Mero1. Therefore, although theoretically KLVF-Mero1 is brighter than GFP, in the in vivo experiment KLVF-Mero1 shows lower brightness than EGFP. However, most of the red fluorescent proteins, which have similar spectral characteristics as KLVF-Mero1, have much lower brightness compared to EGFP, thus, highlighting the advantage of KLVF-Mero1.

A better way to avoid the background fluorescence should be saught soon so that fluorescence images can be scanned shortly after the chromophore is added. Use of a CRBPII mutant that has faster kinetics, but with comparable quantum yield might minimize the background fluorescence by removing the excessive amount of Mero1 right after it is added. Another way is to switch to another chromophore, whose wavelength can be tuned to be far away from the absorption of non-specific binding.

In summary, the fluorescent microscopic pictures taken after the background fluorrescence diminished shows that KLVF-Mero1 could be used as a fluorescent protein tag to image the protein of interest *in vivo*. Due to the versatility of the binding pocket, different retinal analogues could be accomodated in the binding pocket and a broad palette of colors could be achieved using different chromophores. With derivatized merocyanine compounds or a more conjugated chromophore such as Mero2 (**Figure IV-8**), a near-IR fluorescent protein tag can be achieved. Moreover, directed evolution (Fluorescence Activated Cell Sorting) can be applied to modify the CRBPII mutants to achieve fluorescent proteins with desired properties.

IV.7 Using CRBPII derivatives as a chromophoric tag for protein expression and purification

As a result of the wide spectrum of colors obtained with retinal-CRBPII derivatives and its photostability under visible light irradiation, we are pursuing

the development of these proteins as chromophoric tags as well for better visualization of protein expression and protein purification.

After expression of different CRBPII mutants in *E. coli*, retinal was added to the cell culture and incubated at RT overnight with vigorous shaking at 250 RPM. The cells became colorful depending on which mutant was expressed, as shown in **Figure IV-20**. This could be used as an indication of the protein expression level in *E. coli* cells. CRBPII can be tagged to a protein of interest for tracking the protein during purification.

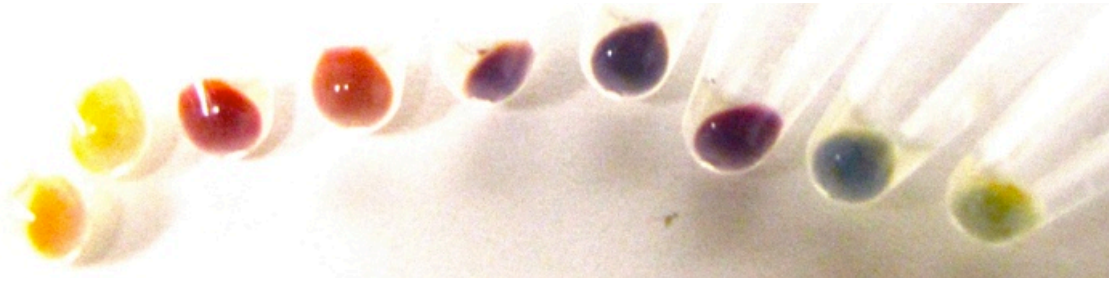


Figure IV-20*: E. coli* cells expressing different CRBPII mutants incubated with retinal overnight and then spun down. From left to right, the mutants are Q108K:K40L:F16W (490 nm), Q108K:T51D (474 nm), Q108K:K40L:R58Y (535 nm), Q108K:K40L:T53C:R58Y (540 nm), Q108K:K40L:T51V:R58Y (563 nm), Q108K:K40L:T51V:T53C:R58Y (573 nm), Q108K:K40L:T51V:Y19W: R58Y (567 nm), Q108K:K40L:T51V:T53C:R58Y:T29L:Y19W:Q4W (593 nm), Q108K:K40L:T51V:T53C: R58W:T29L:Y19W:Q4L (613 nm).

Materials and methods

Cloning of Q108K:K40L:T51V:R58F (KLVF)-CRBPII into pEGFP-C2-RB vector

Q108K:K40L:T51V:R58F (KLVF)-CRBPII gene was amplified out of the KLVF-CRBPII-pET17b vector using the following two primers: 5'-CTCGAGCATGACGAGGGACCAGAATGGAACC-3' and 5'-AAGCTTGATCTC TTCTTTTTGAACACTTGACGGCACAC-3'. The amplified gene and the vector of pEGFP-C2-RB, which has RB (retinoblastoma protein) inserted in a single restriction cutting site BamHI, were digested with restriction enzyme XhoI and HindIII for 5 h at 37 °C. The digested DNA was purified by agarose gel DNA electrophoresis, followed by Qiagen Gel Extraction Kit. Ligation was set up at RT overnight using T4 ligase following the protocol (Invitrogen).[38]

Determination of extinction coefficient of Mero1 PSB formed with n-butylamine.

Mero1 aldehyde (390 μM, 2 μL) was incubated with n-butylamine (2 μL) in ethanol (50 μL) for 0.5 h until Schiff base was completely formed. The Schiff base was acidified with 5 μL 50% HCl to form the protonated Schiff base. UV-vis spectrum of the prepared solution was measured in a final volume of 1 mL PBS buffer. The extinction coefficient at 594 nm can be derived from the following equation:

$$Extinction\ coefficient\ at\ 594\ nm = \frac{Absorption\ at\ 594\ nm}{Concentration\ of\ merocyanine}$$

Determination of extinction coefficient of Mero1 PSB formed with CRBPII mutants

Mero1 aldehyde (390 μM, 2 μL) was incubated with CRBPII protein solution in PBS buffer (~10 μM, 998 μL) for more than 8 h at RT until all of Mero1 was bound as PSB. UV-vis spectrum of the prepared solution was measured. The extinction coefficient at 594 nm can be derived from the following equation:

$$Extinction\ coefficient\ at\ 594\ nm = \frac{Absorption\ at\ 594\ nm}{Concentration\ of\ merocyanine}$$

Determination of quantum efficiency of Mero1 PSB formed with n-butylamine and CRBPII mutants

Quantum efficiency for Mero1-PSB formed with n-butylamine and CRBPII mutants were determined by comparing to rhodomine-6G, which has a quantum efficiency of 95% when excited at 480 nm.[39]

Three samples of Rhodamine-6G solution with absorptions at 480 nm ranging from 0.01 to 0.1 were excited at 480 nm, fluorescence emission spectra were collected from 500 nm to 750 nm. The total photons emitted, which are the integration of the emission spectrum, were plotted against the absorption at 480 nm together with point (0, 0). A linear function was fitted to the points to get a slope for Rhodamine-6G.

For Mero1-PSB, two or three samples of the Mero1-PSB solution were excited at 594 nm, fluorescence emission was measured starting from 594 nm to 750 nm. Similarly, the total photons of the emission were plotted against the

absorption at 594 nm together with point (0,0). A linear function was fitted to the points to get a slope for Mero1-PSB.

The quantum efficiency can be derived from the following equation:

$$Quantum\ efficiency = \frac{Slope\ of\ Mero1\_PSB}{Slope\ of\ rhodamine\_6G} * 95\%$$

Determination of dissociation constant for CRBPII mutants with Mero1

The stock protein solution (~0.5 μM, 40 mL) was prepared in PBS buffer. The protein solution was distributed in portions of 3 mL into the silylated vials. Mero1 was added into each of the prepared protein solutions to a final concentration of 0.2 equiv, 0.4 equiv, 0.6 equiv, 0.8 equiv, 1.0 equiv, 1.2 equiv, 1.6 equiv, 2.0 equiv and 3.0 equiv of the protein solution (as the amount of Mero1 added was up to 10 μL to a 3 mL of protein solution, the concentration of the protein could be considered to be the same for all the prepared solutions). The solutions were incubated for 8 hours in dark at RT. Fluorescence emission spectra for each of the prepared solution were measured with excitation at 594 nm. Fluorescence emission intensity at 630 nm were plotted against the adjusted concentration of Mero1. All the points were fitted to the following equation:

$$Fluo = F * [\frac{(P + x + K_d) - \sqrt{(P + x + K_d)^2 - 4 * P * x}}{2 * P}]$$

$$\text{Mero1-PSB} \overset{K_d}{\rightleftharpoons} \text{Mero1 + CRBPII mutant}$$

Total concentration of CRBPII mutant = *P*;

Total concentration of Mero1 added = *x* (this is the variable);

Assume the concentration of Mero1-PSB formed in equilibrium is *A*,

Then the concentration of free CRBPII mutant is *(P – A)*;

The concentration of free merocyanine is *(x – A);*

Since $K_d = \frac{[CRBPII][Merocyanine]}{[PSB]}$

Then $K_d = \frac{(P-A)(x-A)}{A}$

Thus $A = \frac{(P+x+K_d)-\sqrt{(P+x+K_d)^2-4*P*x}}{2}$

The amount of complex formed, A, could be converted to the fluorescence obtained at different concentrations of Mero1. When the CRBPII mutant is saturated with Mero1, the amount of fluorescence should be the maximum amount of fluorescence possible for CRBPII mutant with concentration of *P*. And if we assume the fluorescence at saturation is *F*, then the amount of complex formed should be:

$$A = \frac{Fluorescence}{F} * P$$

Combining the last two equations, the folowing equation relates fluorescence to the amount of merocyanine added: s

$$Fluo = F * [\frac{(P + x + Kd) - \sqrt{(P + x + Kd)^2 - 4 * P * x}}{2 * P}]$$

CRBPII mutant KLVF is used as an example. The table below shows the concentration of Mero1 added and the corresponding fluorescence emission of Mero1-PSB formed at 630 nm.

Table IV-4: Fluorescence titration of KLVF-CRBPII (4.20×10^{-7} M) with Mero1.

Conc. of Mero1, x (M)	Fluo. at 630 nm
0.0000	0.0000
8.40×10^{-8}	2.22×10^{7}
1.68×10^{-7}	4.08×10^{7}
2.52×10^{-7}	4.99×10^{7}
3.36×10^{-7}	5.93×10^{7}
4.20×10^{-7}	7.81×10^{7}
5.04×10^{-7}	9.62×10^{7}
6.72×10^{-7}	1.10×10^{8}
8.40×10^{-7}	1.05×10^{8}
1.05×10^{-6}	1.10×10^{8}

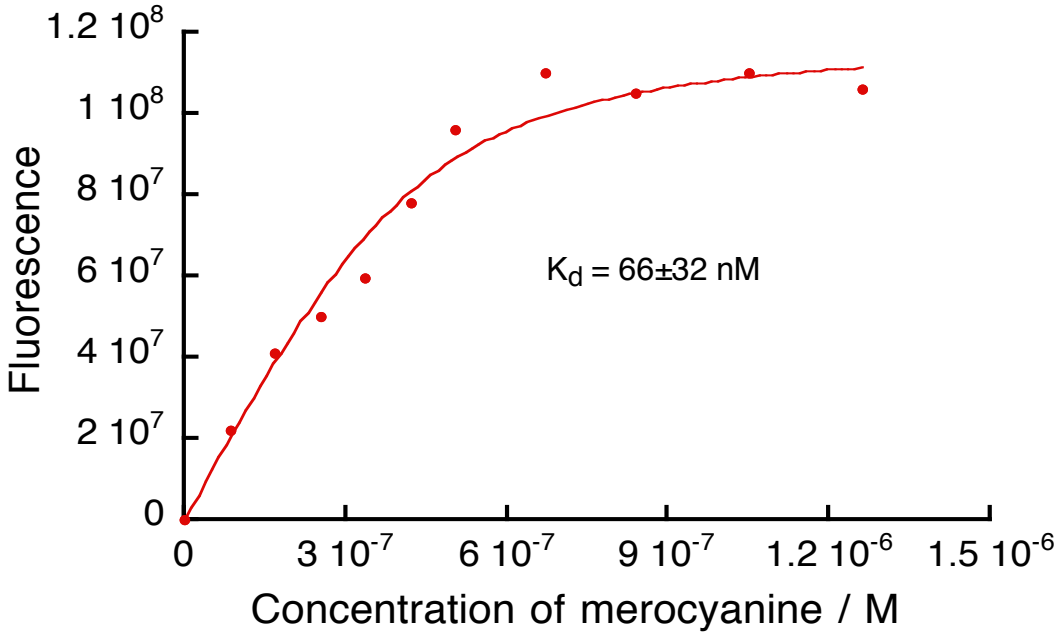


Figure IV-21: Fluorescence titration of KLVF with Mero1.

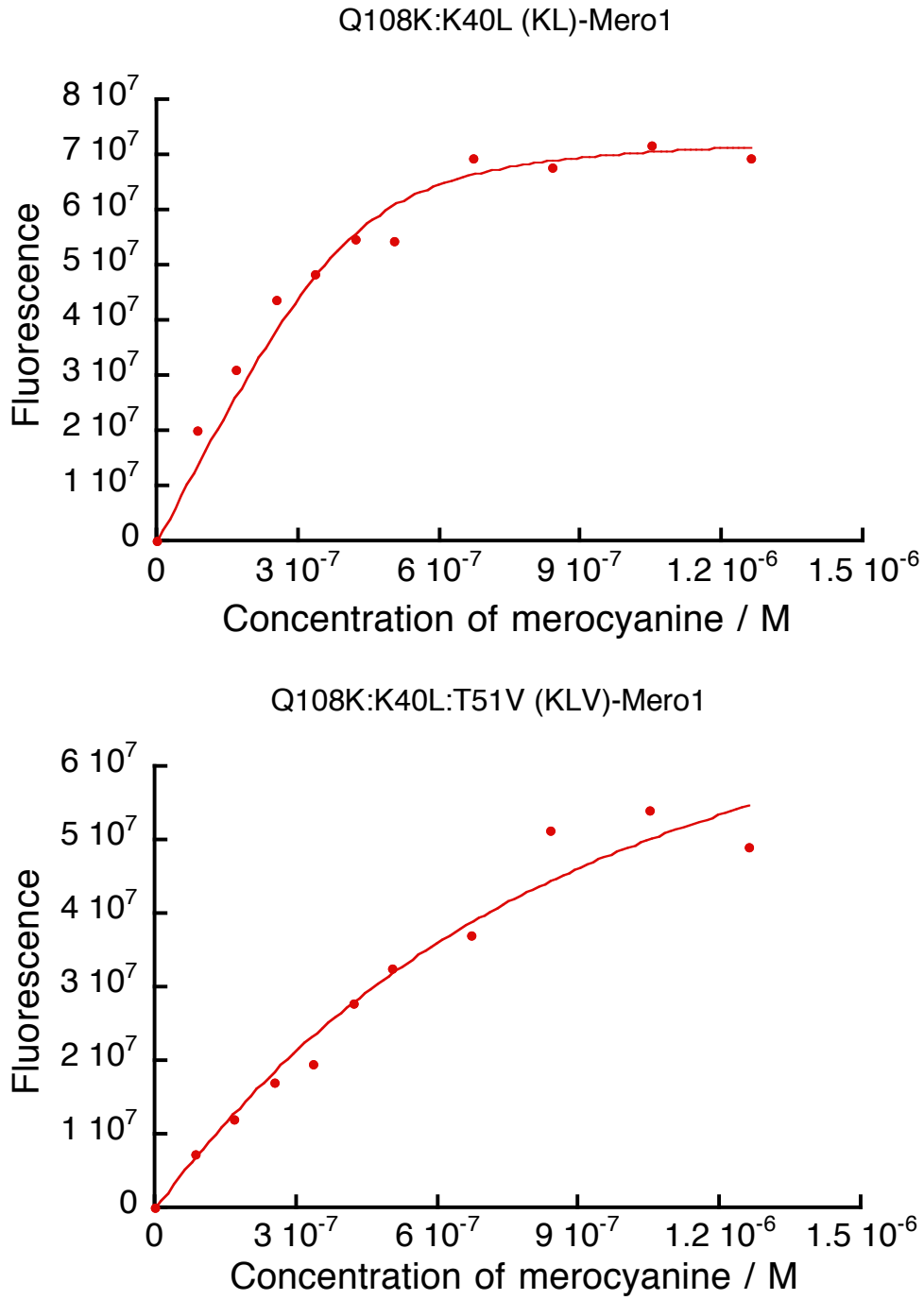


Figure IV-22: Fluorescence titrations of KL and KLV with Mero1.

Similarly, dissociation constant was obtained for Q108K:K40L (KL) and Q108K:K40L:T51V (KLV) to be 74±44 nM and 632±306 nM, respectively.

Fluorophore maturation kinetics

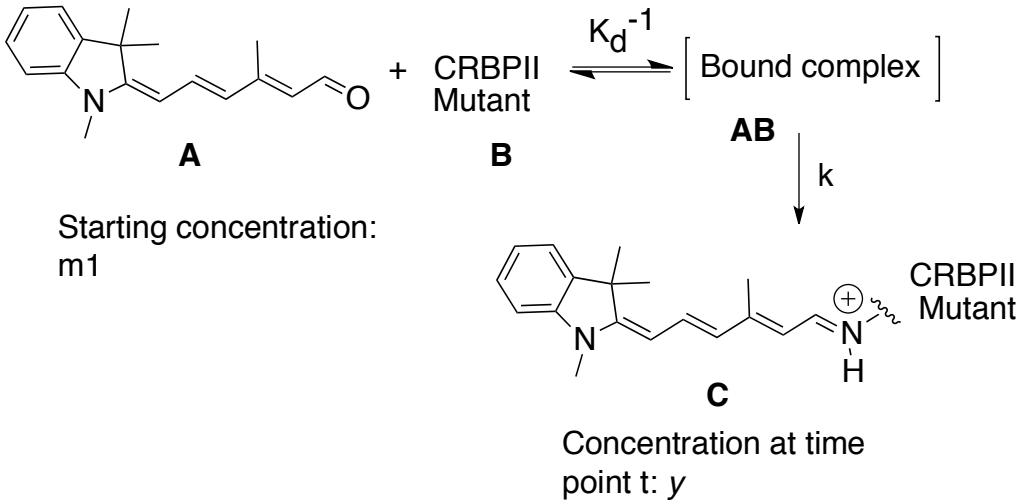


Figure IV-23: Scheme of Mero1-PSB formation with CRBPII mutant.

UV-vis spectrometer was used to record the conversion of merocyanine aldehyde to merocyanine PSB at 37 °C and 20 °C. The ratio of merocyanine to protein was roughly 1:10, with merocyanine final concentration around 1 μM. UV-vis spectra were recorded every minute for fast maturation protein or every two and half minutes for slow maturation protein. Absorptions at 602 nm, which corresponds to the absorption of merocyanine-PSB, were plotted against time and fitted to the following derived function:

$$y = m1 - e^{-kt+m2}$$

A red shift is observed right after Mero1 was added to the protein solution (the λ_{max} for Mero1 aldehyde in PBS buffer is ~490 nm), indicating Mero1 was sequestered by the protein immediately. Thus the binding step is fast and the rate-limiting step is the formation of the protonated Schiff base.

Since $K_d = \frac{[A]*[B]}{[AB]}$

Then

$$[AB] = \frac{[A] * [B]}{K_d}$$

If [B] >>K_d then [AB] >> [A], we can assume that all of the starting material A formed the reactive intermediate [AB] immediately. Assuming the original concentration of Mero1 is m1, and the formation of the PSB at the time point is y, then the concentration of [AB] at any time t is: *(m1-y).* Because the rate of PSB formation only depends on the concentration of the protein-chromophore complex, [AB], the rate of PSB formation at time point t is:

$$\frac{dy}{dt} = k[AB] = k * (m1 - y)$$

so

$$\frac{dy}{(m1 - y)} = k * dt$$

Integration of two sides:

$$\int_0^y \frac{dy}{(m1 - y)} = \int_0^t k * dt$$

leads to:

$$-\ln(m1 - y) = kt + m2$$

where *m3* is a constant, thus,

$$y = m1 - e^{-kt+m2}$$

Live cell imaging in *E. coli*

BL21 cells, BL21 cells carrying WT-CRBPII-pET17b expression vector and BL21 cells carrying KL-CRBPII-pET17b, KLVF-CRBPII-pET17b expression vectors were innoculated overnight in 10 mL LB media. The overnight culture (500 µL) was transferred to LB media (10 mL) with Tet added for BL21 cells and Amp/Tet added for BL21 cells carrying CRBPII gene. The cells were kept shaking at 37 °C for 2 h before IPTG was added at a final concentration of 1 mM. The cells were kept shaking at 26 °C for 4 h before being transferred to eppendorf tubes (1.5 mL) in 500 µL portions. Mero1 was added to the cells at a final concentrations of 1 µM, 3 µM and 6 µM. The cells were vigorously shaken at 220 RPM at 37 °C and the fluorescence image were taken at 0.5 h, 1 h, 2 h, and 3 h time points.

A small portion of the cell culture mixed with Mero1 (50 µL) was spun down at 5000 RPM for 2 min. The cell pellet was resuspended gently in PBS buffer (50 µL). The resuspended cells (1 µL) was pipetted on a microscope glass slide and covered up with a cover glass. Olympus fluorescent microscope with 10x objective lens and 40x magnification, with 100 W Mercury lamp source was used. The cells were excited with green light and the microscopic images were taken with filters that pass red and blue light.

Live cell imaging in human osteosarcoma cells

Human osteosarcoma cells (U2-OS) were cultured in a 4-well microscopic chamber slide (Thermo-Scientific) with Dulbecco's modified Eagle medium

without phenol red (DMEM, Sigma), supplemented with 10% v/v heat inactivated fetal bovine serum (FBS, USA Scientific) and 10 mM of PSG (penicillin, streptomycin, and glutamine). The cells were maintained at 37 °C and under 5% CO_2. When the cells reach 50% to 70% mobility on the well, they were transfected with pEGFP-CRBP-RB and pEGFP-RB using NanoJuice Transfection Reagents and Kits (Novagen). Merocyanine solution (1 mM in ethanol) was added to each well to a final concentration of 0.25 μM after 48 h of transfection. The cells were incubated at 37 °C under 5% CO_2 for 2 h. The cells were then washed with 5% DMSO PBS solution 5 times and the media was changed back to DMEM media without phenol red. The cells were incubated at 37 °C under 5% CO_2 for at least 12 h before confocal microscopic images were taken.

Microscopic pictures were taken on a Zeiss 510mete Confocal Laser Scanning Microscope with a 63x oil objective (NA 1.40). GFP was excited with a 488 nm Argon laser line and emission was detected from 505 to 530 nm. KLVF-Mero1 was excited with a 594 nm Helium-Neon laser line and emission was detected with a 615 nm longpass filter.

Photobleaching study

The transfected cells with EGFP-KLVF-Mero1 was subject to photobleaching study. EGFP was irradiated with a 488 nm 10 mW Argon laser light with 20% transmission, fluorescence was detected from 505 nm to 530 nm. Simultaneously KLVF-Mero1 was irradiated with a 594 nm 2 mW Helium-Neon laser at 100% transmission, fluorescence was detected with 615 nm longpass

filter. Microscopic pictures were taken every 30 sec. Fluorescence from one whole cell was quantified for all the time points and plotted against time. Fluorescence from a small region outside the fluorescent cell was quantified and also plotted against time.

Table IV-4: Summary of different CRBPII mutants bound with Mero1.

Entry	CRBPII Mutant	QY (%)	ε_{594nm}/ 100,000	B^a /1000	Relative B^a vs. Mero1-$BuNH_2$
1	n-$BuNH_2$-Mero1	1.20	1.43	1.7	1.0
2	Q108K	15	n.d.	n.d.	n.d.
3	Q108K:K40L (KL)	8.40	2.59	22	12
4	Q108K:K40L:T51V (KLV)	15	2.40	36	21
5	Q108K:K40L:R58Y	10	2.60	26	15
6	Q108K:K40L:T51V:R58W:Y19W	18	n.d.	n.d.	n.d.
7	Q108K:K40L:T51V:T53V:R58W:Y19W	16	n.d.	n.d.	n.d.
8	Q108K:K40L:T51V:T53C	15	n.d.	n.d.	n.d.
9	Q108K:K40L:T51V:F16W	8.70	n.d.	n.d.	n.d.
10	Q108K:K40L:T51V:R58W:Q4W	10.2	n.d.	n.d.	n.d.
11	Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4F:A33W	8	n.d.	n.d.	n.d.
12	Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:A33W	12	n.d.	n.d.	n.d.
13	Q108K:K40L:T51V:R58W:Y19W:T53V:A33W	13	n.d.	n.d.	n.d.
14	Q108K:K40L:T51V:R58W:Y19W:T53V:A33W:Q4F	6.9	n.d.	n.d.	n.d.

Table IV-4 continued

15	Q108K:K40L:T51V:R58Y:Y19W:T53V:A33W	16.6	n.d.	n.d.	n.d.
16	Q108K:K40L:T51V:R58Y:Y19W:T53V:Q4F:A33W	6.2	n.d.	n.d.	n.d.
17	Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:Q4R:A33W	5.0	n.d.	n.d.	n.d.
18	Q108K:K40L:T51V:T53C:R58W:T29L:Q4F:A33W	11	n.d.	n.d.	n.d.
19	Q108K:K40L:T51V:Y19W:A33W	19	n.d.	n.d.	n.d.
20	Q108K:K40L:R58W:L77W	5.8	n.d.	n.d.	n.d.
21	Q108K:K40L:L77W:Y60W	12	n.d.	n.d.	n.d.
22	Q108K:K40L:T51V:T53C:R58W:T29L:Q4F:A33W:L117E	18	2.88	52	30
23	Q108K:K40L:R58F	7.5	n.d.	n.d.	n.d.
24	Q108K:K40R	7.6	n.d.	n.d.	n.d.
25	Q108K:K40L:T51V:R58F:L117E	16	2.34	37	22
26	Q108K:K40L:T51V:R58Y:A33W	20	2.29	46	27
27	Q108K:K40L:T51V:A33W	15	n.d.	n.d.	n.d.
28	Q108K:K40L:Y60W	5.0	n.d.	n.d.	n.d.
29	Q108K:K40L:R58W	9.3	2.49	23	13
30	Q108K:K40L:T51V:L119Q:Y60W	8.0	n.d.	n.d.	n.d.

Table IV-4 continued

31	Q108K:K40L:T51D:R58W:Y19W	15.5	2.89	45	26
32	Q108K:K40L:T51V:R58W:A33W:L117E	14	2.22	31	18
33	Q108K:K40L:T51V:R58Y:A33W:L117E	17.4	n.d.	n.d.	n.d.
34	Q108K:K40L:T51V:R58F (KLVF)	18	2.26	41	24
35	Q108K:K40L:T51V:R58Y	17.6	n.d.	n.d.	n.d.
36	Q108K:K40L:T51V:R58W	18.8	n.d.	n.d.	n.d.

Note: All the data refered to excitation at 594 nm.

B^a stands for brightness. $B^a = QY \times \varepsilon_{594nm}$.

(n.d. not determined)

References

References

1. (a) Giepmans, B. N. G.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y., Review - The fluorescent toolbox for assessing protein location and function. *Science* **2006,** *312* (5771), 217-224; (b) Cho, W. H.; Stahelin, R. V., Membrane-protein interactions in cell signaling and membrane trafficking. *Annu. Rev. Bioph. Biom.* **2005,** *34*, 119-151.

2. Zimmer, M., GFP: from jellyfish to the Nobel prize and beyond. *Chem. Soc. Rev.* **2009,** *38* (10), 2823-2832.

3. Cody, C. W.; Prasher, D. C.; Westler, W. M.; Prendergast, F. G.; Ward, W. W., Chemical-structure of the hexapeptide chromophore of the aequorea green-fluorescent protein. *Biochemistry* **1993,** *32* (5), 1212-1218.

4. 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), 229-233.

5. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C., Green fluorescent protein as a marker for gene-expression. *Science* **1994,** *263* (5148), 802-805.

6. Ormo, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J., Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **1996,** *273* (5280), 1392-1395.

7. Jackson, S. E.; D Craggs, T.; Huang, J. R., Understanding the folding of GFP using biophysical techniques. *Expert. Rev. Proteomic* **2006,** *3* (5), 545-559.

8. Jung, G.; Zumbusch, A., Improving autofluorescent proteins: Comparative studies of the effective brightness of green fluorescent protein (GFP) mutants. *Microsc. Res. Techniq.* **2006,** *69* (3), 175-185.

9. Heim, R.; Cubitt, A. B.; Tsien, R. Y., Improved green fluorescence. *Nature* **1995,** *373* (6516), 663-664.

10. (a) Heim, R.; Prasher, D. C.; Tsien, R. Y., Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *P. Natl. Acad. Sci. USA* **1994,** *91* (26), 12501-12504; (b) Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y., Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* **1995,** *20* (11), 448-455; (c) Hein, R.; Tsien, R. Y., Engineering green fluorescent protein for improved brightness,

longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.* **1996,** *6* (2), 178-182.

11. Mishin, A. S.; Subach, F. V.; Yampolsky, I. V.; King, W.; Lukyanov, K. A.; Verkhusha, V. V., The first mutant of the Aequorea victoria green fluorescent protein that forms a red chromophore. *Biochemistry* **2008,** *47* (16), 4666-4673.

12. Yarbrough, D.; Wachter, R. M.; Kallio, K.; Matz, M. V.; Remington, S. J., Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-angstrom resolution. *P. Natl. Acad. Sci. USA* **2001,** *98* (2), 462-467.

13. Bevis, B. J.; Glick, B. S., Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). *Nat. Biotechnol.* **2002,** *20* (1), 83-87.

14. 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. *Nat. Biotechnol.* **2004,** *22* (12), 1567-1572.

15. Chudakov, D. M.; Matz, M. V.; Lukyanov, S.; Lukyanov, K. A., Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiol. Rev.* **2010,** *90* (3), 1103-1163.

16. (a) Rayan, G.; Guet, J. E.; Taulier, N.; Pincet, F.; Urbach, W., Recent Applications of Fluorescence Recovery after Photobleaching (FRAP) to Membrane Bio-Macromolecules. *Sensors-Basel* **2010,** *10* (6), 5927-5948; (b) McNally, J. G., Quantitative FRAP in analysis of molecular binding dynamics in vivo. *Method Cell Biol.* **2008,** *85*, 329-351.

17. Anderson, M. T.; Tjioe, I. M.; Lorincz, M. C.; Parks, D. R.; Herzenberg, L. A.; Nolan, G. P.; Herzenberg, L. A., Simultaneous fluorescence-activated cell sorter analysis of two distinct transcriptional elements within a single cell using engineered green fluorescent proteins. *P. Natl. Acad. Sci. USA* **1996,** *93* (16), 8508-8511.

18. (a) Susperreguy, S.; Prendes, L. P.; Desbats, M. A.; Charo, N. L.; Brown, K.; MacDougald, O. A.; Kerppola, T.; Schwartz, J.; Piwien-Pilipuk, G., Visualization by BiFC of different C/EBP beta dimers and their interaction with HP1 alpha reveals a differential subnuclear distribution of complexes in living cells. *Exp. Cell. Res.* **2011,** *317* (6), 706-723; (b) Kerppola, T. K., Biomolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annu. Rev. Biophys.* **2008,** *37*, 465-487; (c) Kerppola, T. K., Design and implementation of bimolecular fluorescence complementation (BiFC) assays

for the visualization of protein interactions in living cells. *Nat. Protoc.* **2006,** *1* (3), 1278-1286; (d) Hu, C. D.; Chinenov, Y.; Ranez-Carrozzi, V. R.; Grinberg, A. V.; Kerppola, T. K., gelFRET and BiFC: Visualization of protein interactions in vitro and in living cells. *Biophysical Journal* **2002,** *82* (1), 2a-2a; (e) Hoffmann, C.; Gaietta, G.; Bunemann, M.; Adams, S. R.; Oberdorff-Maass, S.; Behr, B.; Vilardaga, J. P.; Tsien, R. Y.; Eisman, M. H.; Lohse, M. J., A FlAsH-based FRET approach to determine G protein - coupled receptor activation in living cells. *Nat. Methods* **2005,** *2* (3), 171-176.

19. Flors, C., Photoswitching of monomeric and dimeric DNA-intercalating cyanine dyes for super-resolution microscopy applications. *Photoch. Photobio. Sci.* **2010,** *9* (5), 643-648.

20. (a) Schermelleh, L.; Heintzmann, R.; Leonhardt, H., A guide to super-resolution fluorescence microscopy. *J Cell Biol* **2010,** *190* (2), 165-175; (b) Huang, B., Super-resolution optical microscopy: multiple choices. *Curr Opin Chem Biol* **2010,** *14* (1), 10-14; (c) Huang, B.; Bates, M.; Zhuang, X. W., Super-Resolution Fluorescence Microscopy. *Annu. Rev. Biochem.* **2009,** *78*, 993-1016.

21. (a) Faro, A. R.; Carpentier, P.; Jonasson, G.; Pompidor, G.; Arcizett, D.; Demachy, I.; Bourgeois, D., Low-Temperature Chromophore Isomerization Reveals the Photoswitching Mechanism of the Fluorescent Protein Padron. *J. Am. Chem. Soc.* **2011,** *133* (41), 16362-16365; (b) Subach, F. V.; Patterson, G. H.; Renz, M.; Lippincott-Schwartz, J.; Verkhusha, V. V., Bright Monomeric Photoactivatable Red Fluorescent Protein for Two-Color Super-Resolution sptPALM of Live Cells. *J. Am. Chem. Soc.* **2010,** *132* (18), 6481-6491; (c) Nienhaus, K.; Nar, H.; Heilker, R.; Wiedenmann, J.; Nienhaus, G. U., Trans-cis isomerization is responsible for the red-shifted fluorescence in variants of the red fluorescent protein eqFP611. *J. Am. Chem. Soc.* **2008,** *130* (38), 12578.

22. (a) Henderson, J. N.; Gepshtein, R.; Heenan, J. R.; Kallio, K.; Huppert, D.; Remington, S. J., Structure and Mechanism of the Photoactivatable Green Fluorescent Protein. *J. Am. Chem. Soc.* **2009,** *131* (12), 4176-+; (b) Subach, F. V.; Malashkevich, V. N.; Zencheck, W. D.; Xiao, H.; Filonov, G. S.; Almo, S. C.; Verkhusha, V. V., Photoactivation mechanism of PAmCherry based on crystal structures of the protein in the dark and fluorescent states. *P. Natl. Acad. Sci. USA* **2009,** *106* (50), 21097-21102; (c) Brakemann, T.; Stiel, A. C.; Weber, G.; Andresen, M.; Testa, I.; Grotjohann, T.; Leutenegger, M.; Plessmann, U.; Urlaub, H.; Eggeling, C.; Wahl, M. C.; Hell, S. W.; Jakobs, S., A reversibly photoswitchable GFP-like protein with fluorescence excitation decoupled from switching. *Nat. Biotechnol.* **2011,** *29* (10), 942-U132.

23. Griffin, B. A.; Adams, S. R.; Tsien, R. Y., Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998,** *281* (5374), 269-272.

24. (a) Morimoto, K.; Dolan, M. E.; Scicchitano, D.; Pegg, A. E., Repair of O-6-Propylguanine and O-6-Butylguanine in DNA by O-6-Alkylguanine-DNA Alkyltransferases from Rat-Liver and E-Coli. *Carcinogenesis* **1985,** *6* (7), 1027-1031; (b) Pegg, A. E., Repair of O-6-alkylguanine by alkyltransferases. *Mutat. Res-Rev. Mutat.* **2000,** *462* (2-3), 83-100; (c) Pegg, A. E.; Dolan, M. E.; Scicchitano, D.; Morimoto, K., Studies of the Repair of O-6-Alkylguanine and O-4-Alkylthymine in DNA by Alkyltransferases from Mammalian-Cells and Bacteria. *Environ. Health Persp.* **1985,** *62* (Oct), 109-114; (d) Spratt, T. E.; Meyer, A. S., Sequence specificity of repair of O-6-methyl guanine by O-6-alkylguanine-DNA alkyltransferases due to differences in the rate of methyl transfer not differences in binding. *Abstr. Pap. Am. Chem. S.* **2003,** *226*, U322-U322.

25. Daniels, D. S.; Woo, T. T.; Luu, K. X.; Noll, D. M.; Clarke, N. D.; Pegg, A. E.; Tainer, J. A., DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol.* **2004,** *11* (8), 714-720.

26. (a) George, N.; Pick, H.; Vogel, H.; Johnsson, N.; Johnsson, K., Specific labeling of cell surface proteins with chemically diverse compounds. *J. Am. Chem. Soc.* **2004,** *126* (29), 8896-8897; (b) Juillerat, A.; Gronemeyer, T.; Keppler, A.; Gendreizig, S.; Pick, H.; Vogel, H.; Johnsson, K., Directed evolution of O-6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. *Chem. Biol.* **2003,** *10* (4), 313-317; (c) Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K., A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **2003,** *21* (1), 86-89; (d) Keppler, A.; Pick, H.; Arrivoli, C.; Vogel, H.; Johnsson, K., Labeling of fusion proteins with synthetic fluorophores in live cells. *P. Natl. Acad. Sci. USA* **2004,** *101* (27), 9955-9959.

27. Juillerat, A.; Heinis, C.; Sielaff, I.; Barnikow, J.; Jaccard, H.; Kunz, B.; Terskikh, A.; Johnsson, K., Engineering substrate specificity of O-6-alkylguanine-DNA alkyltransferase for specific protein labeling in living cells. *Chembiochem* **2005,** *6* (7), 1263-1269.

28. Heinis, C.; Schmitt, S.; Kindermann, M.; Godin, G.; Johnsson, K., Evolving the substrate specificity of O-6-alkylguanine-DNA alkyltransferase through loop insertion for applications in molecular imaging. *ACS Chem. Biol.* **2006,** *1* (9), 575-584.

29. (a) Chen, I.; Choi, Y. A.; Ting, A. Y., Phage display evolution of a peptide substrate for yeast biotin ligase and application to two-color quantum dot

labeling of cell surface proteins. *J. Am. Chem. Soc.* **2007,** *129* (20), 6619-6625; (b) Fernandez-Suarez, M.; Chen, T. S.; Ting, A. Y., Protein-protein interaction detection in vitro and in cells by proximity biotinylation. *J. Am. Chem. Soc.* **2008,** *130* (29), 9251-+; (c) Howarth, M.; Liu, W. H.; Puthenveetil, S.; Zheng, Y.; Marshall, L. F.; Schmidt, M. M.; Wittrup, K. D.; Bawendi, M. G.; Ting, A. Y., Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nat Methods* **2008,** *5* (5), 397-399; (d) Baruah, H.; Puthenveetil, S.; Choi, Y. A.; Shah, S.; Ting, A. Y., An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photo-cross-linking. *Angew. Chem. Int. Ed.* **2008,** *47* (37), 7018-7021; (e) Howarth, M.; Ting, A. Y., Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat. Protoc.* **2008,** *3* (3), 534-545; (f) Fernandez-Suarez, M.; Baruah, H.; Martinez-Hernandez, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y., Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* **2007,** *25* (12), 1483-1487; (g) Puthenveetil, S.; Liu, D. S.; White, K. A.; Thompson, S.; Ting, A. Y., Yeast Display Evolution of a Kinetically Efficient 13-Amino Acid Substrate for Lipoic Acid Ligase. *J. Am. Chem. Soc.* **2009,** *131* (45), 16430-16438; (h) Cohen, J. D.; Thompson, S.; Ting, A. Y., Structure-Guided Engineering of a Pacific Blue Fluorophore Ligase for Specific Protein Imaging in Living Cells. *Biochemistry* **2011,** *50* (38), 8221-8225; (i) Jin, X.; Uttamapinant, C.; Ting, A. Y., Synthesis of 7-Aminocoumarin by Buchwald-Hartwig Cross Coupling for Specific Protein Labeling in Living Cells. *Chembiochem* **2011,** *12* (1), 65-70; (j) 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. *P. Natl. Acad. Sci. USA* **2010,** *107* (24), 10914-10919.

30. Filonov, G. S.; Piatkevich, K. D.; Ting, L.-M.; Zhang, J.; Kim, K.; Verkhusha, V. V., Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nat. Biotechnol.* **2011,** *29* (8), 757-U133.

31. 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), 345-348.

32. Philip, A. F.; Nome, R. A.; Papadantonakis, G. A.; Scherer, N. F.; Hoff, W. D., Spectral tuning in photoactive yellow protein by modulation of the shape of the excited state energy surface. *P. Natl. Acad. Sci. USA* **2010,** *107* (13), 5821-5826.

33. Lee, K. S. S. Michigan State University, East Lansing, 2011.

34. Liu, R. S. H.; Asato, A. E., Tuning the color and excited state properties of the azulenic chromophore: NIR absorbing pigments and materials. *J. Photoch. Photobio. C* **2003,** *4* (3), 179-194.

35. Singh, A. K.; Hota, P. K., Development of bacteriorhodopsin analogues and studies of charge separated excited states in the photoprocesses of linear polyenes. *Photochem. Photobiol.* **2007,** *83* (1), 50-62.

36. Sameiro, M.; Goncalves, T., Fluorescent Labeling of Biomolecules with Organic Probes. *Chem. Rev.* **2009,** *109* (1), 190-212.

37. (a) Bartek, J.; Bartkova, J.; Lukas, J., The retinoblastoma protein pathway in cell cycle control and cancer. *Exp. Cell Res.* **1997,** *237* (1), 1-6; (b) Cobrinik, D.; Dowdy, S. F.; Hinds, P. W.; Mittnacht, S.; Weinberg, R. A., The Retinoblastoma Protein and the Regulation of Cell Cycling. *Trends Biochem. Sci.* **1992,** *17* (8), 312-315; (c) Ewen, M. E., The Cell-Cycle and the Retinoblastoma Protein Family. *Cancer Metast. Rev.* **1994,** *13* (1), 45-66; (d) Riley, D. J.; Lee, E. Y. H. P.; Lee, W. H., The Retinoblastoma Protein - More Than a Tumor-Suppressor. *Annu. Rev. Cell Biol.* **1994,** *10*, 1-29; (e) Rubin, E.; Tamrakar, S.; Ludlow, J. W., Protein phosphatase type 1, the product of the retinoblastoma susceptibility gene, and cell cycle control. *Pediatr. Pathol. Mol. M* **1999,** *18* (4-5), 395-415; (f) Weinberg, R. A., The Retinoblastoma Protein and Cell-Cycle Control. *Cell* **1995,** *81* (3), 323-330; (g) Wiman, K. G., The Retinoblastoma Gene - Role in Cell-Cycle Control and Cell-Differentiation. *Faseb J.* **1993,** *7* (10), 841-845.

38. Graham, D. E.; Xu, H. M.; White, R. H., Identification of coenzyme M biosynthetic phosphosulfolactate synthase - A new family of sulfonate-biosynthesizing enzymes. *J. Biol. Chem.* **2002,** *277* (16), 13421-13429.

39. Kubin, R. F.; Fletcher, A. N., Fluorescence Quantum Yields of Some Rhodamine Dyes. *J. Lumin.* **1982,** *27* (4), 455-462.

Chapter V

Design of a photo-switchable protein tag for affinity purification of protein of interest

V.1 Introduction of azo-compounds

V.1.1 Photophysical properties of azo-compounds

Photochromism is a phenomenon describing the color change upon light irradiation. It usually involves chemical interconversion of two forms of the molecule; these two forms have different absorptions. It was first discovered in 1867, when Fritzsche found an orange colored tetracene solution bleached in the daytime and recovered at night.[4]

To illustrate the chemical structural changes upon light excitation, a few types of representative photochromic organic compounds are shown in **Figure V-1**. As shown, most of the interconversion of the two different states of the compounds involves bond formation or bond cleavage to generate a more conjugated system, with longer absorption wavelength for the compounds on the right.

However, azobenzene type photochromic compounds change their photophysical properties by undergoing bond isomerization and result in a large conformational change. Due to this characteristic of azobenzene, it has been widely used in chemical biological systems and photoswitchable materials.

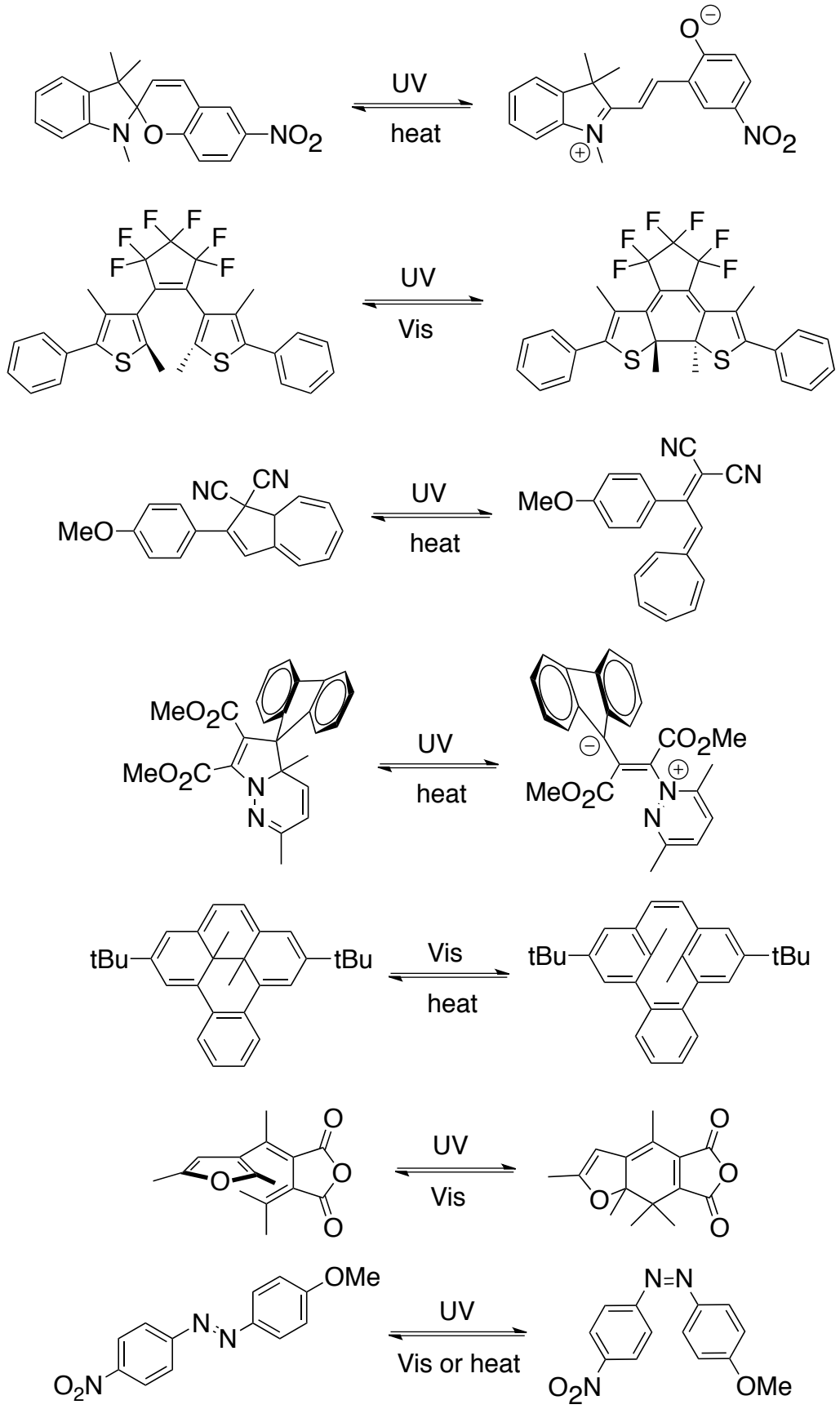


Figure V-1: A few examples of photochromic interconversions.[1]

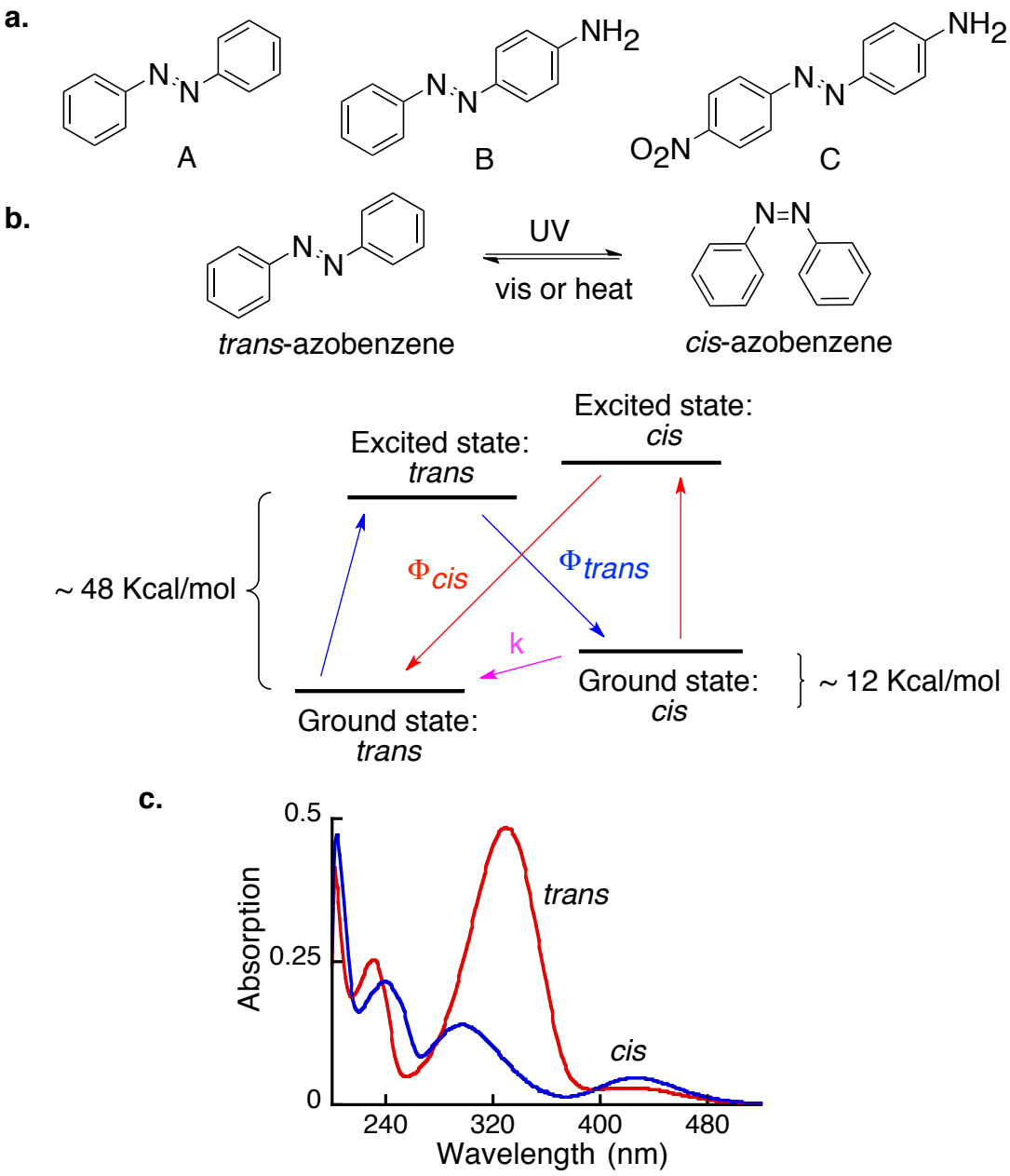
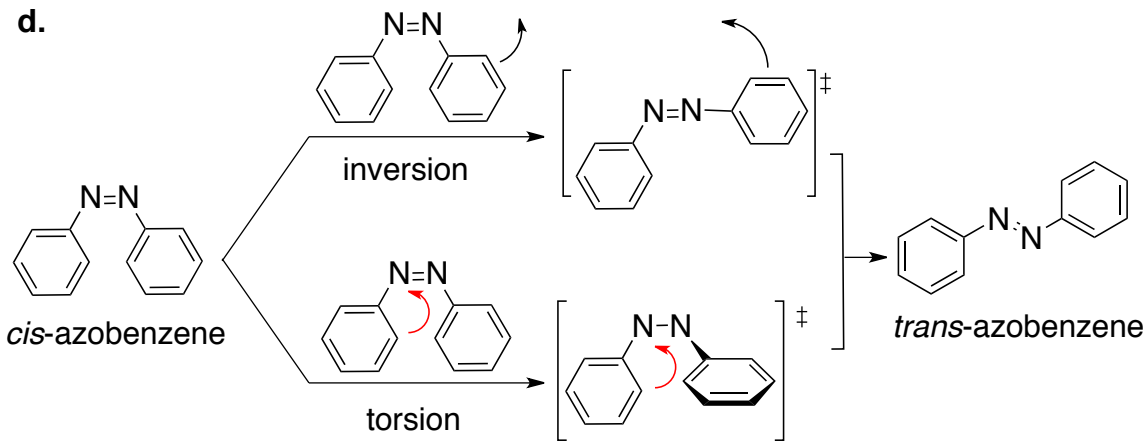


Figure V-2: Physical properties of azobenzene. **a.** Chemical structures of three spectroscopic classes of azobenzene compounds. **b.** Isomerization of azobenzene and the energy diagram for the isomerization.[2] **c.** UV-vis spectra for *trans*-azobenzene and *cis*-azobenzene. **d**. Two proposed mechanisms for isomerization.

Figure V-2 continued.



There are generally speaking three spectroscopic types of azobenzene compounds as shown in **Figure V-2a:** A. azobenzenes (without polar substitutions), B. aminoazobenzenes (with an electron-donating group substituted in the *para* position) and C. pseudostilbenes (with an electron donating group and electron-withdrawing group substituted in *para* positions of the two benzene rings).[2] Because of their different electronic polarization properties, their spectra are slightly different. Usually azobenzene compounds exhibit two absorption peaks, one with maximum at around 360 nm, the other one at around 420 nm (**Figure V-2c**). Pseudostilbenes have the most polarized system and exhibit the most intense absorption in the blue visible light spectrum, followed by aminoazobenzene and azobenzene.

Type A *trans*-azobenzene compounds exhibit a major absorption peak in the UV region (~360 nm, due to vertical electronic excitation from π to π*) and a small absorption peak in the visible blue light region (~420 nm, due to electronic

excitation from n to π*) (**Figure V-2c**).[5] Upon isomerization to *cis*, the absorption at 360 nm decreases while the absorption peak in the blue visible spectrum part increases. In order to avoid the steric repulsion of the *cis*-orientation, the two benzene rings are twisted out of the plane of the diazo, leading to the decreased UV absorption, as a result of the decrease in π orbital conjugation. However, this also leads to the increase of the orbital overlap from n to π* and therefore the blue visible light absorption increases slightly.

Isomerization quantum efficiency was found to be dependent on the excitation wavelength. In a hydrophobic solvent, such as hexane, the quantum efficiency of the *trans to cis* photoconversion is 0.40 and 0.12, when excited at ~360 nm (π-π*) and ~440 nm (n-π*), respectivel. Similarly, for *cis to trans* photoconversion, the quantum efficiency is 0.53 and 0.25, excited at 290 nm (π-π*) and 430 nm (n-π*), respectively.[6]

Because of the slightly separated absorptions for *trans* and *cis* isomers in the UV region, 360 nm versus 290 nm, upon UV light irradiation at ~360 nm, where *trans*-azobenzene has an high absorption extinction coefficient, *trans*-azobenzene is isomerized to *cis*-azobenzene efficiently (**Figure V-2b**). Under blue visible light irradiation, where *cis*-azobenzene has more absorption than *trans*-azobenzene, *cis*-azobenzene will isomerize back to *trans* more efficiently.

At the same time, *trans*-azobenzene is around 12 Kcal/mol more stable than *cis*-azobenzene, as a result of release of the steric strain originating from the two phenyl ring in the *cis*-orientation, the *cis*-azobenzene will thermally relax

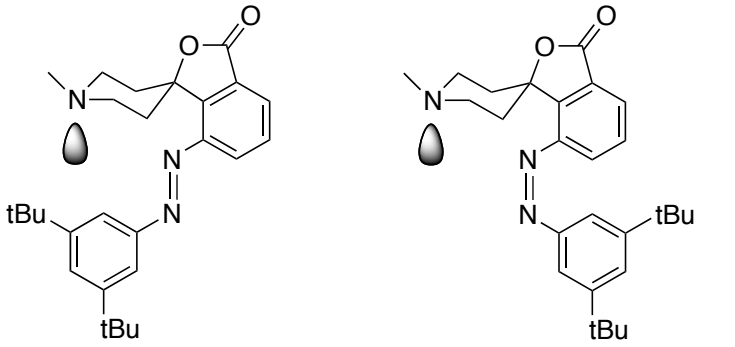
back to *trans*-azobenzene. Due to the overlapping of absorptions between *trans* and *cis*-azobenzene, upon light irradiation, these three processes, *trans* isomerization to *cis*, *cis* isomerization to *trans* and thermal relaxation can happen at the same time. The photostationary state depends on the competition among the three processes. Usually for unsubstituted azobenzene compounds, at the photostationary state, the percentage of *cis*-azobenzene can reach up to 95% when excited at ~360 nm.

Two possible isomerization mechanisms are possible for azobenzene compounds. As illustrated in **Figure V-2d**, azobenzene can undergo CNN inversion or torsion around the N=N bond to isomerize from *cis* to *trans* or *trans* to *cis.* Rau proposed that when excited at the π-π* absorption band, azobenzene would go through the N=N torsion mechanism, while when excited at n-π* absorption band, azobenzene would go through CNN inversion mechanism.[7] Understanding the correct mechanism of isomerization could help design the correct steric demand for photoisomerization of azobenzene. However, this is still under debate.[6]

Cis-azobenzene compounds usually have short lifetimes, on the order of hours, minutes and seconds for type **A, B** and **C** compounds, respectively (**Figure V-2a)**. Introduction of electron donating groups or strong electron-withdrawing groups greatly shorten the lifetime of the *cis*-isomer. This is likely due to weakening of the N=N double bond through polarization, which lowers the energy barrier for thermal relaxation from *cis* back to the thermally more stable

trans-isomer.[2] Depending on the application of azobenzene, the types of azobenzene compounds with suitable *cis*-isomer lifetime should be considered.

V.1.2 Applications of azobenzene compounds as a photoswitch



inaccessible nucleophilic/basic site **neutral**

accessible nucleophilic/basic site **basic**

Figure V-3: Photoswitching the Brønsted basicity of amine group using azobenzene.

Due to the large conformational change of azobenzene compounds upon isomerization and its photostability, they have been widely applied in different areas, for example organic chemistry, material science, and especially in chemical biological systems, to obtain photoswitchable properties.

In the field of organic chemistry, azobenzene has been used to switch the basicity of an amino group. As shown in **Figure V-3,**[8] with *trans*-azobenzene the amino group is blocked for reactivity. Photoswitching to *cis*-azobenzene moves the sterically bulky di-*tert*-butyl substituted benzene group, opening up access for the lone pair of amino group to act as a nucleophilic or Brønsted base. In this

way, a photoswitchable Brønsted base is achieved and its basicity in reaction solely depends on which light irradiation is used.

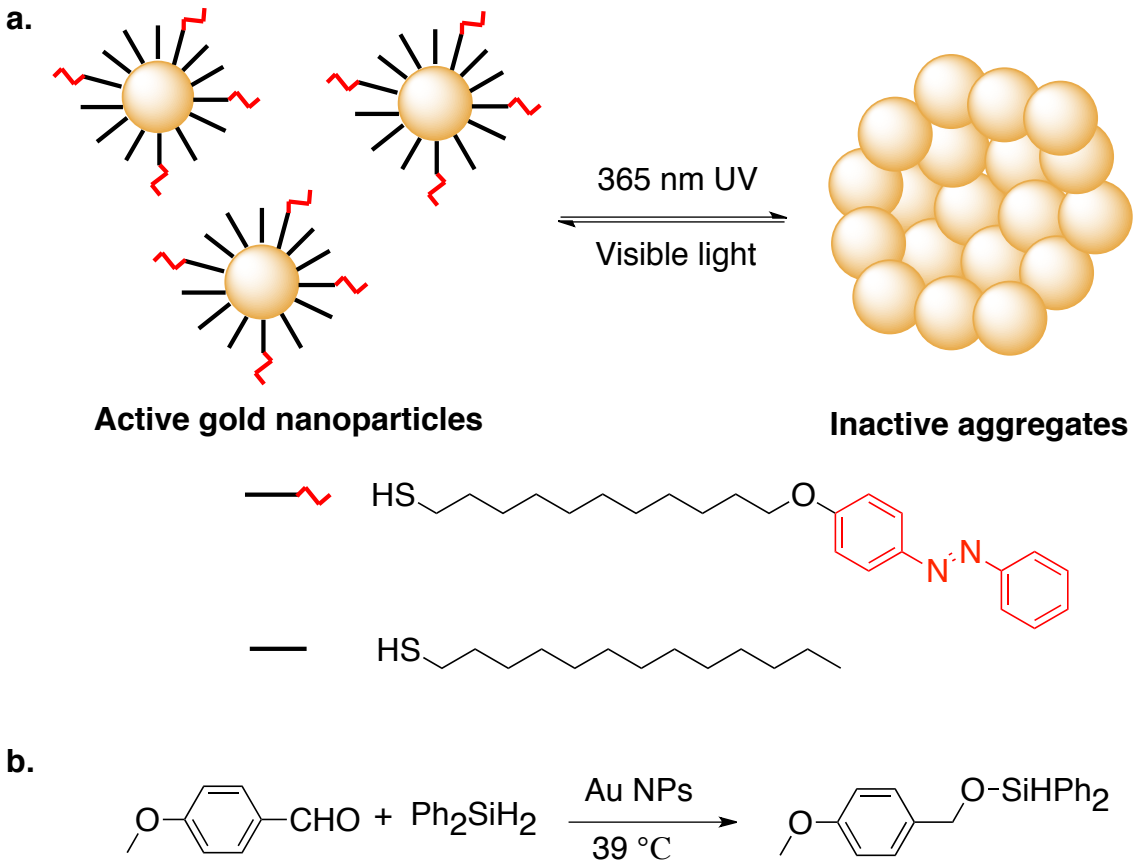


Figure V-4: Photoswitchable catalysis for hydrosilylation of *p*-methoxy-benzaldehyde with gold nanoparticles[3] (reaction was carried out in dry toluene).

Cis-azobenzene was found to have a higher dipole moment than *trans*-azobenzene; therefore *cis*-azobenzene is more soluble in polar solvents, while *trans*-azobenzene is more soluble in non-polar solvents. Utilizing this property of azobenzene compounds, Grzybowski designed gold nanoparticles with photo-controlled catalytic activity through chelating of azobenzene to gold nanoparticles.[3] Under visible light, *trans*-azobenzene is dominant and therefore

can be easily solvated in a nonpolar solvent, such as toluene. Upon UV light irradiation, *trans*-azobenzene is isomerized to *cis*-azobenzene and the increased polarity of *cis*-azobenzene promotes the aggregation of nanoparticles in the nonpolar solvent, like reverse micelles. The hydrosilylation of methoxybenzaldehyde was monitored and indeed UV light switched off the catalytic activity of gold nanoparticle and visible light switched on the catalytic activity.

Besides activation of catalytic activity, azobenzene has also been used as a reversible light induced chirality switch by Haberhauer and his coworkers, through fusing azobenzene to rigid prochiral compounds.[9] And interestingly, disulfated-azobenzene has been used in crystallography to obtain different crystal forms because of different packing interactions of *trans* and *cis*-azobenzene with the host molecule.[10]

In the field of chemical biology, azobenzene is most widely used for switching the binding affinity of two interacting partners. The binding of azobenzene incorporated molecules with their target, such as an enzyme, DNA, RNA or different surfaces, could be controlled by applying different light. In some other cases, azobenzene is incorporated into the backbone and thus through isomerization from the *trans*-isomer to *cis*-isomer, the distance between the two molecules linked to the azobenzene can change dramatically from around 10 Å to 5 Å. Therefore, large conformational changes can result.

Westmark and coworkers designed photoswitchable transition-state-analogue inhibitors of cysteine and serine proteases.[11] A five fold increase of cysteine protease, papain, enzymatic activity was obtained with irradiation of light from 330 to 370 nm. At the photostationary state, the *cis*-azobenzene isomer inhibitor was estimated to be 83% and did not inhibit the cysteine protease. Irradiation with visible light switched the azobenzene to the *trans*-isomer (**Figure V-5a**), which has a higher affinity for papain to inhibit the active site and therefore lowers its enzymatic activity. It was shown that the regulation system could be cycled a few times without obvious change in its activity.

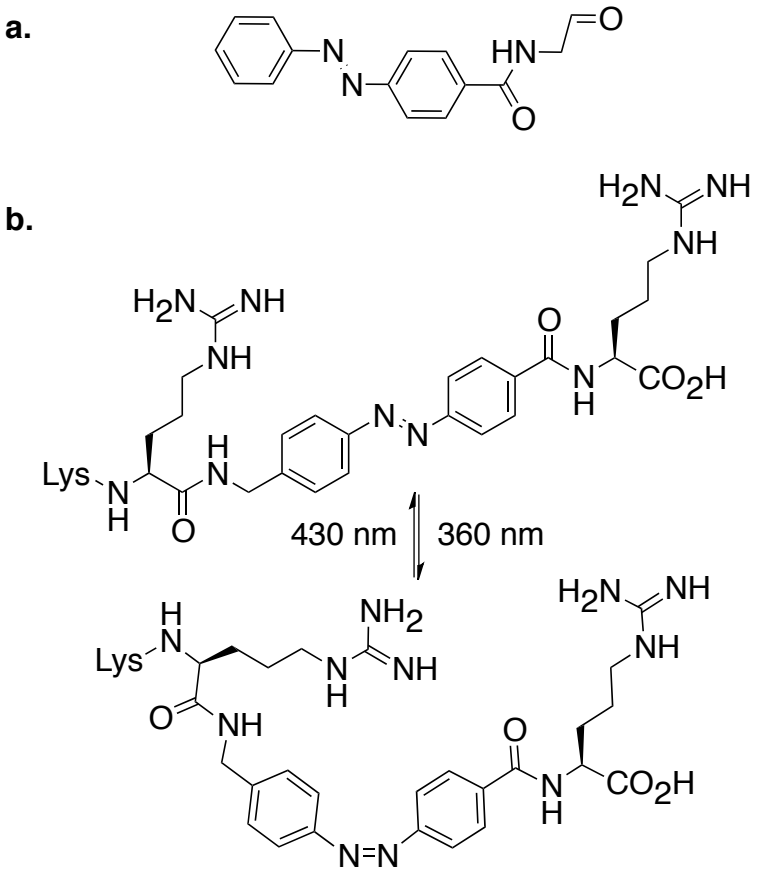


Figure V-5: a. The inhibitor of papain protease. **b.** Peptide KRAzR and its isomerization, with the *trans*-isomer binding better to the RNA aptamer.

Azobenzene has been incorporated into peptide backbones to synthesize peptidomimetics. The large conformational change induces large secondary structural change of the peptides.[12] Hayashi and coworkers were able to design an RNA aptamer, which has photoregulated affinity towards a photoresponsive peptide, KRAzR.[13] The RNA aptamer was obtained through in vitro selection from 70 random nucleotide sequence pool of RNA. As shown in **Figure V-5b**, the *trans*-azobenzene conformation will possibly render a better hydrogen bonding interaction of the two arginine residues with the RNA aptamer.

The activity of RNase H also could be regulated through photoresponsive sense DNA.[14] RNase H only digests RNA when it is hybridized with antisense DNA. The availability of antisense DNA could be tuned through an azobenzene incorporated DNA double strand. As shown in **Figure V-6**, an unnatural base made with azobenzene was incorporated into the sense DNA strand. When azobenzene is *trans*, it is flat with maximum π-π stacking, while upon UV

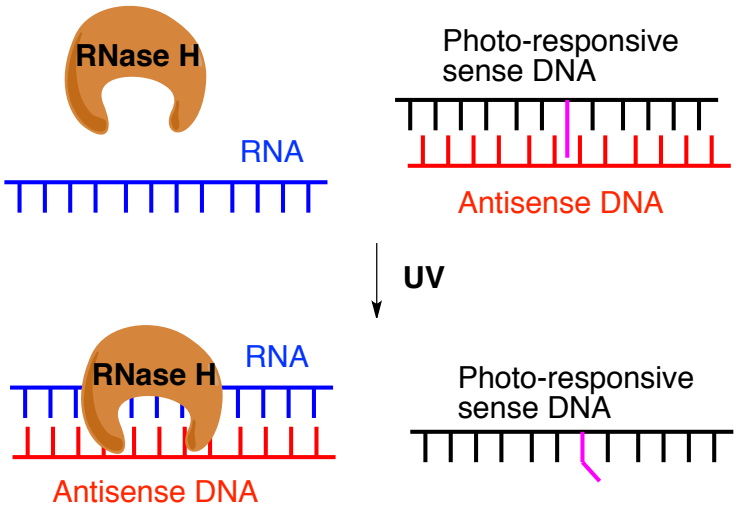


Figure V-6: Diagram for photoregulation of RNase H activity.

irradiation, *trans*-azobenzene isomerizes to *cis*-azobenzene. The *cis*-isomer is twisted and it will destabilize the DNA duplex. The DNA double strand will break and make the antisense DNA available to hybridize with RNA, ready for RNase H digestion. Indeed under UV light irradiation, the RNA is hydrolyzed by RNase H two to four fold more than in the dark state.

Incorporation of azobenzene could make interactions photoswitchable; this has been applied to obtain photoswitchable interactions on surfaces as well. RGD peptides are known to interact with integrins in the extracellular matrix of cells. As shown in **Figure V-7,** fusing RGD peptides to azobenzene will make it possible to switch the accessibility of the RGD peptide.[15] When azobenzene is in the *trans* conformation, the RGD peptide is more accessible and higher cell adhesion is observed, while under UV light when azobenzene isomerizes to the

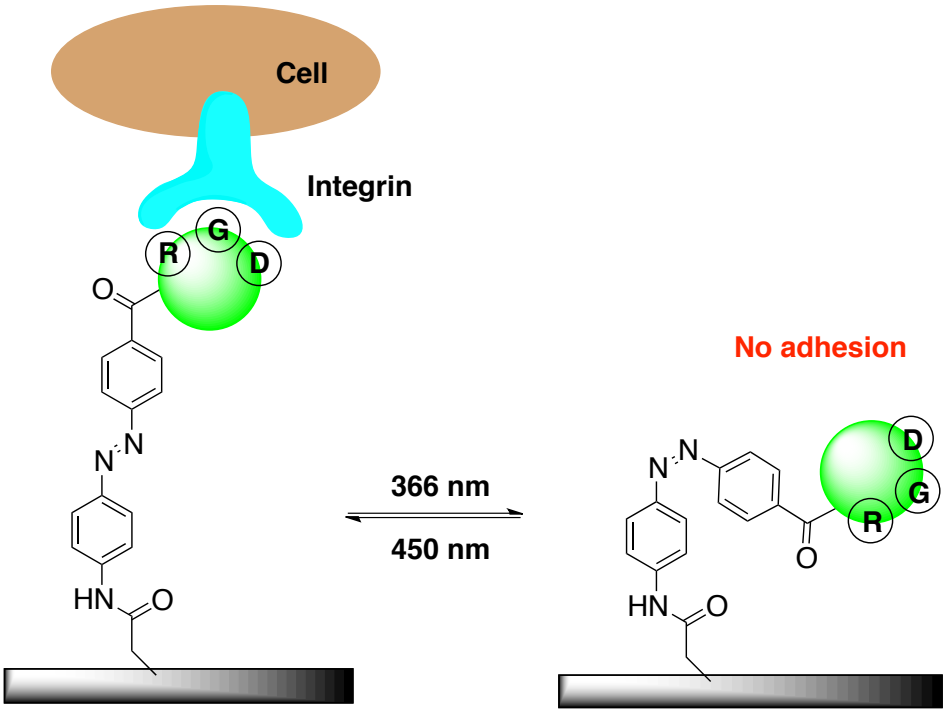


Figure V-7: Diagram for photoregulation of surface adhesion.

cis-conformation, the RGD peptide is not as accessible and therefore, lower cell adhesion is observed.

Azobenzene has also been used for photo-regulated allosteric control of an glutamate receptor.[16] This glutamate receptor is known to regulate glutamate-mediated ion channels. Upon binding of glutamate, the conformation of the glutamate receptor changes and opens the ion channel to release sodium and calcium cations and uptake potassium cations. Based on this, Volgraf and coworkers designed a ligand as shown in **Figure V-8**.[16] The succinimide moiety is covalently ligated to a cysteine residue on the glutamate receptor protein. With *cis*-azobenzene tethered, the glutamate can bind to the glutamate receptor

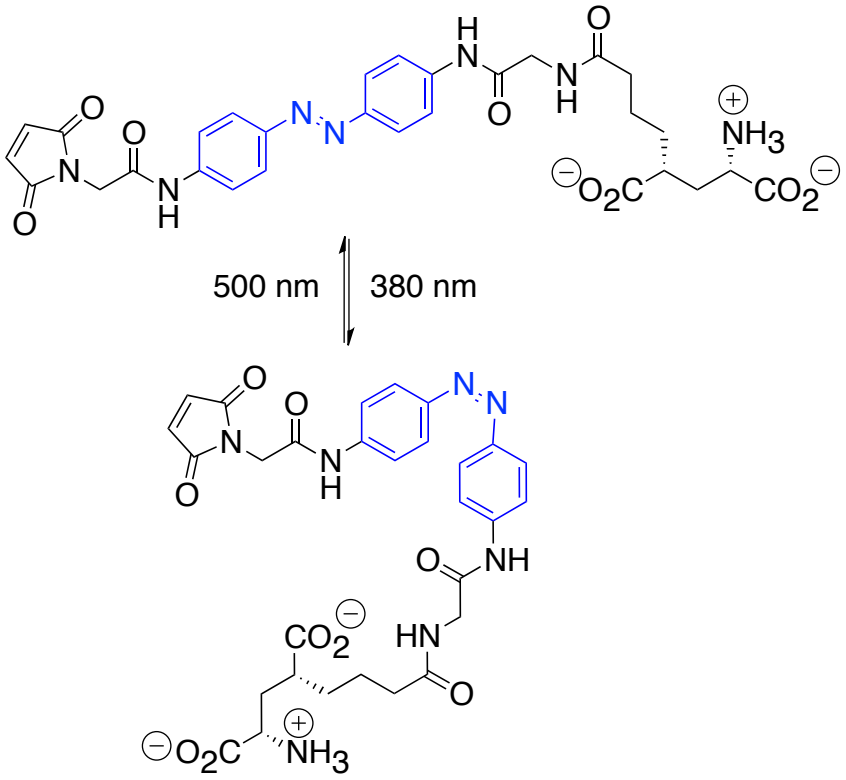


Figure V-8: Chemical structures of the molecule used for photoregulation of glutamate receptor.

protein and induce the channel opening, while the *trans*-azobenzene would project the glutamate in a position notable to approach the binding site. Therefore, the ion channel is gated by light driven isomerization of azobenzene.

There are many more interesting examples of azobenzene in use as a photoswtich. Azobenzene has greatly broadened the scope of photo regulation in *in vitro* studies. However, as azobenzene incorporated compounds are usually big and specific ligation to a target inside the cell is a challenge, *in vivo* studies using azobenzene are still lacking.

V.2 To develop a photoswitchable protein tag for protein purification

V.2.1 General scheme for photoswitchable protein tag

As has been shown, azobenzene has been widely used as a photoswitch; our goal is to develop a photoswitchable protein tag with azobenzene-like compounds. A protein tag that responds to different light irradiation could be applied for many purposes; for example: using photo-regulated protein-protein interactions for yeast two hybrid system, spatio-rearrangement of a protein due to different protein interaction partners, and photoswitchable affinity protein purification.

We are interested in finding a method to enable practical photoswitchable protein affinity purification. The general scheme is described in **Figure V-9**. A protein tag, which has specific binding affinity for the *trans*-isomer, together with the protein of interest will be bound to *trans*-azobenzene, which is immobilized on

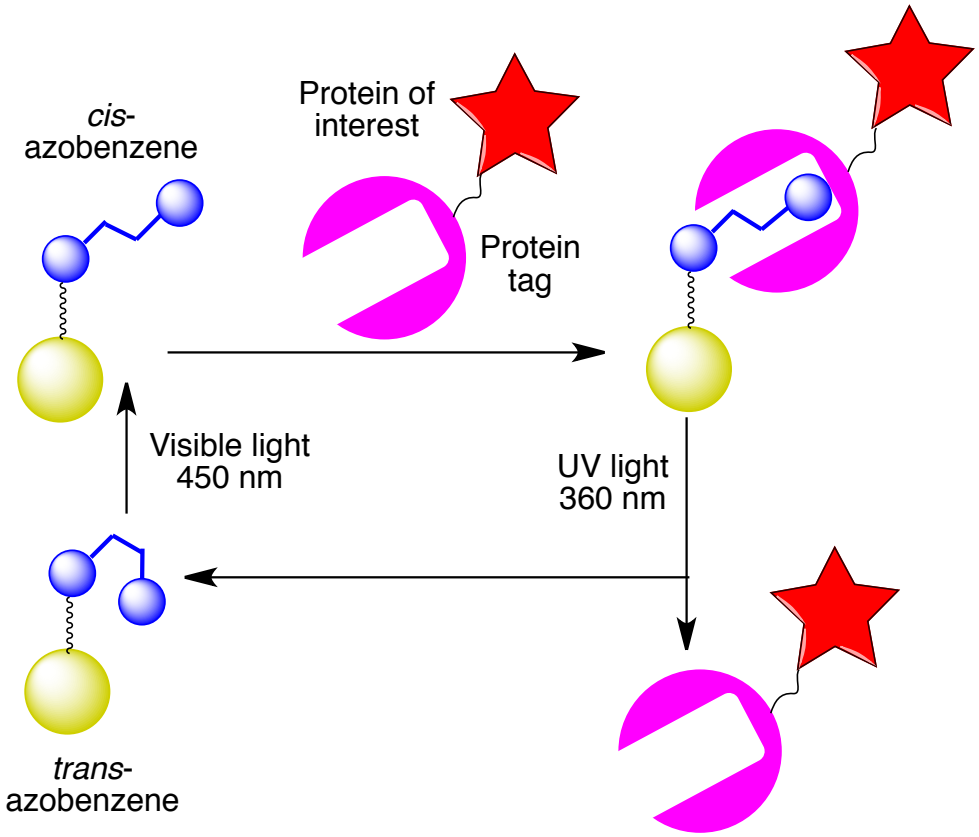


Figure V-9: General scheme for photoswitchable affinity protein purification by tagging the protein of interest with photoswitchable protein tag.

a solid surface. Upon light irradiation with UV light, *trans*-azobenzene isomerizes to *cis*-azobenzene. As a result, the protein tag loses its binding affinity for *cis*-azobenzene and dissociates from the beads, along with the protein of interest. The beads with *cis*-azobenzene could be recycled either through thermal relaxation or visible light irradiation to regenerate the *trans*-azobenzene.

In this purification method, light is used as the eluent to release the protein, without introducing a high concentration of eluent molecules as in most affinity protein purifications. The high concentration of eluent could interfere with the protein activity or crystallization in subsequent experiments. In normal affinity protein purification, after the elution of protein, usually another purification step is

necessary to get rid of the eluting agent in the eluted protein solution, either through dialysis or size exclusion chromotography. With a photoswitchable affinity protein tag, the eluted protein can be directly subjected to further characterization.

V.2.2 Previous studies of photoswitchable protein binding interactions with azobenzene

As described earlier, photoswitchable protein affinity purification exhibits advantages over the traditional affinity protein purification method. The challenge lies in developing a peptide or protein tag that could show affinity specifically for one isomer of azobenzene or its derivatives. Prior to this, some work has been reported in trying to find an antibody, a protein tag or a short peptide in order to achieve selective photoswitchable interactions with azobenzene.

Harada and coworkers developed a monoclonal antibody against an azobenzene containing peptide hapten as shown in **Figure V-10**.[17] Fluorescence quenching using 100% *trans*-azobenzene peptide and 82% *cis*-

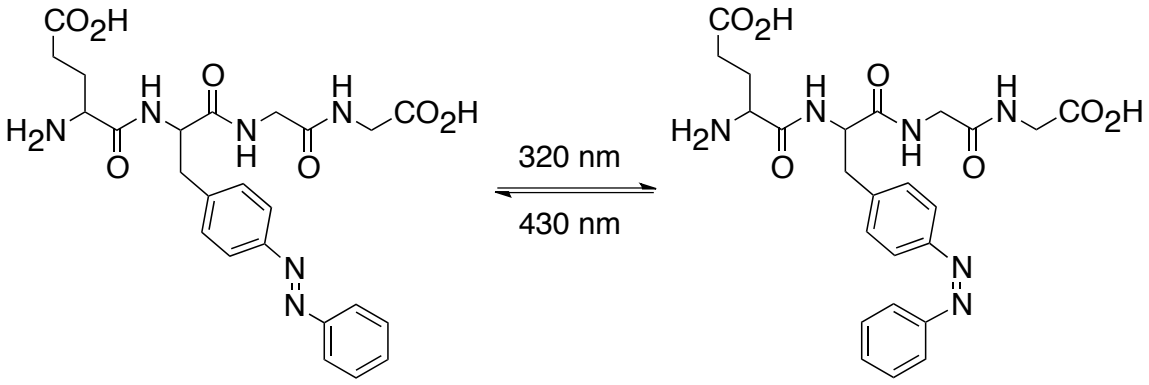


Figure V-10: Photoswitchable azobenzene containing hapten peptide.

azobenzene peptide showed that the antibody has higher affinity towards *trans*-azobenzene peptide. HPLC was used to monitor the photoswitchable binding event of the antibody and azobenzene-containing peptide hapten. It was interesting to find that the concentration of free hapten increased by six fold upon UV light irradiation due to low affinity of *cis*-azobenzene peptide towards the antibody.

To prove that the isomerization happened inside the binding pocket,[18] high concentrations of antibody were used to ensure the majority of the *trans*-azobenzene hapten was bound inside the antibody. Picosecond pulsed laser was applied to isomerize the *trans*-azobenzene hapten to *cis*. Dissociation of azobenzene-hapten and re-association of the hapten with the antibody is not possible in the picosecond time scale. Therefore, if isomerization still takes place, it is the bound azobenzene hapten inside the antibody cavity. In this way, it was confirmed that the antibody has a large cavity, in which the isomerization could take place.

However, antibody expression as a protein tag fused with another protein is not practical. Furthermore, the peptide is vulnerable to degradation with proteases in the crude protein solution and might have non-specific binding interactions with cell lysate. Its use is not practical for protein purification.

Recently, Pearson and coworkers were able to add a photoswitchable azobenzene moiety into the phenylalanine based trifluoromethylketone inhibitor of α-chymotrypsin and immobilize the molecule on a gold surface to obtain

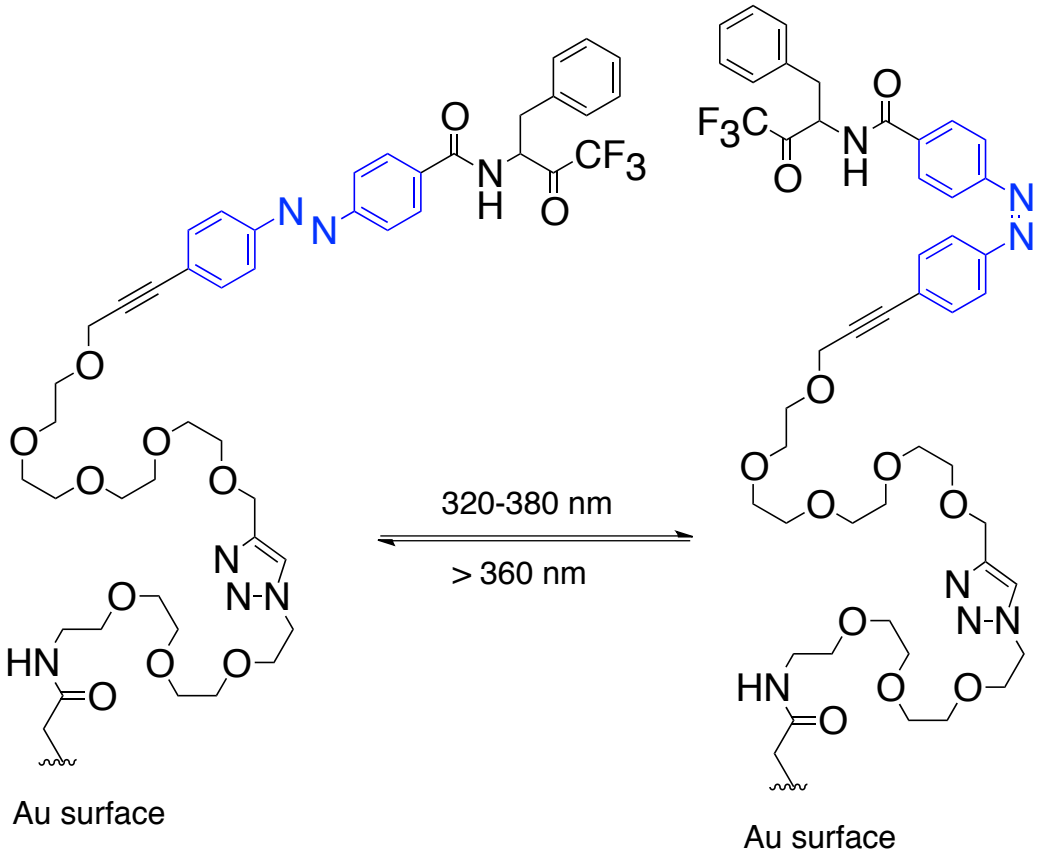


Figure V-11: Photoreversible isomerization of α-chymotrypsin inhibitor immobilized on gold surface.

reversible photoregulation of binding toward α-chymotrypsin (**Figure V-11**).[19, 20]

The binding affinity toward *cis*-azobenzene incorporated inhibitor is 5.3 fold higher than the *trans*-azobenzene inhibitor. This strategy could be further extended to other proteases. However, for each specific interaction partner, a specific design needs to be developed and tuned to obtain the highest contrast in binding to *cis* vs *trans*-azobenzene moiety. The precondition for this to work is that part of the azobenzene moiety should be involved in binding. A more general photoswitchable protein tag with a larger difference in binding affinity should be sought.

More recently, Chen and coworkers developed a 7-mer peptide with higher affinity toward *cis*-azobenzene through phage display as shown in **Figure V-12**. The largest ratio of apparent binding constants toward azobenzene under UV light versus visible light is 3.2.

This simplified system inspired us to develop a protein or larger peptide tag for exclusive affinity toward one isomer of azobenzene derivatives for protein purification purpose. Considering that a 7-mer peptide is small, but can achieve a three fold binding affinity preference for one isomer over the other, it could be possible to gain better selectivity with a larger peptide, such as a 24-mer or 48-

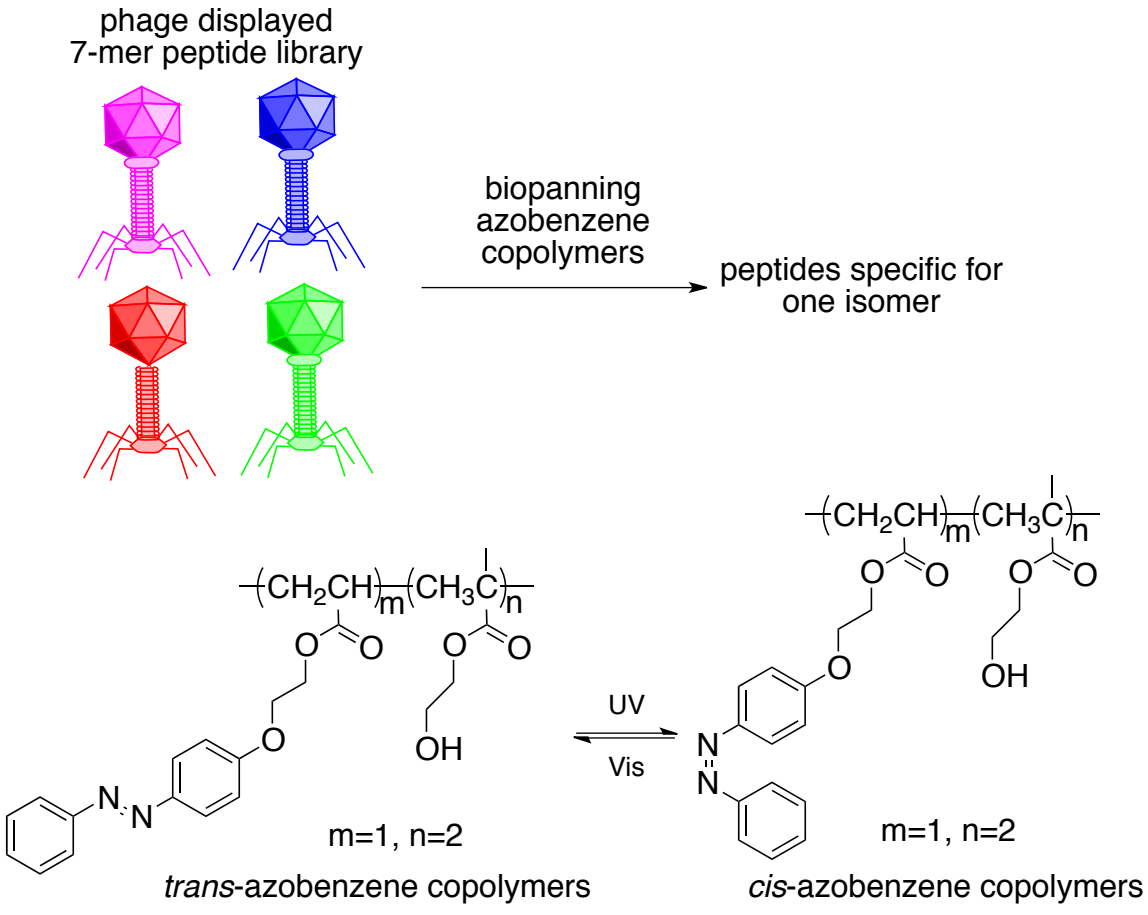


Figure V-12: Phage display 7-mer peptides against *cis*-azobenzene copolymers.

mer. Phage display will be applied for the directed evolution of a peptide or protein to bind *trans*-azobenzene derivatives. With this powerful and efficient protein engineering tool, it is likely to select a larger peptide or protein that has high and specific affinity for one isomer of azobenzene.

V.3 Brief introduction of phage display and comparison with other display methods

Protein engineering to obtain desired properties or enzymatic activities is a broad and interesting area. Semi-rational engineering and *in-silico* protein engineering could greatly assist the process, but only when the crystal structure and binding mode or reaction mechanism are already known. Directed evolution methods have proved to be a powerful method for selecting proteins with desired properties, because the selection could be from a pool of up to 10^{12} diversified

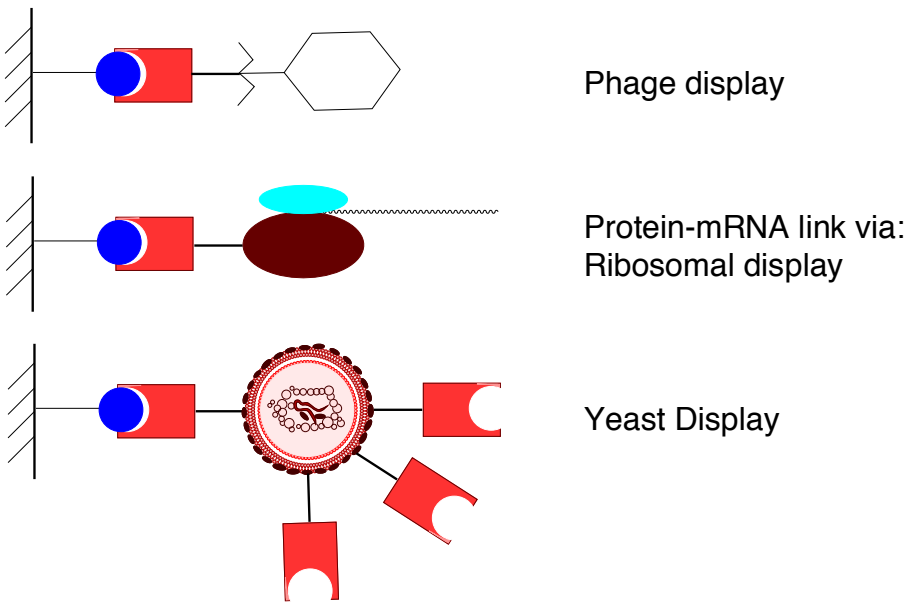


Figure V-13: Illustration of three kinds of major display methods.

peptides or protein, depending on the directed evolution method utilized.

Generally speaking, there are three major display methods for directed evolution, phage display,[21] yeast display,[22] and ribosomal display.[23] As the name suggests, phage display is to display peptides or proteins on the surface of phages, yeast display on the surface of yeast, while ribosomal display is display of proteins and peptides on the mRNA-protein complex (**Figure V-13**). For all of these displaying methods, it is possible to correlate the displayed peptide to its DNA sequence, which is extremely relevant to identify the selected peptide. In addition, with the protein or peptides exposed, direct selection against some desired molecules could be performed as shown in **Figure V-13**.

Each display method has its advantages and disadvantages. Hoogenboom has presented a good comparison of the three different display methods, as shown in **Table V-1.**[24] The library of yeast display is the smallest and it requires cell-sorting. With no access to cell sorting instrument in our lab, we turned to the other two selection methods. As phage display is technically easier than ribosomal display and the library generated is comparable to ribosome display, phage display was preferred and chosen for the directed evolution of the photoswitchable protein tag against azobenzene compounds.

Phage display has proven to be a powerful technique for drug discovery and protein engineering.[21, 25] By cloning the gene of interest into the phagemid, the protein corresponding to that gene can be displayed on the surface of the

phage. This makes it possible for direct selection of the desired peptides from a large library of various displayed peptides through affinity binding. The number of peptides in the peptide library can be up to hundreds of millions through random mutagenesis in the targeted gene or by using a short random peptide sequence.

Table V-1: Comparison of the three major display methods.

Name	Phage	Ribosome	Yeast Cell
Valency of display	Monovalent; multivalent	Monovalent	Multivalent
Typical maximum library size	10^{10} to 10^{11}	10^{12} to 10^{13}	10^{7}
Selection scope	Versatile	Limited	Cell Sorting
Main application	Affinity maturation; Stability increase	Affinity maturation; Stability increase	Affinity maturation; Stability increase; Expression increase
Main strength	Technically robust; Easy to use; automated	Intrinsic mutagenesis; Fastest of all systems; Amenable to automation	Fast in combination with random mutagenesis; direct screening for kinetics with cells
Main weakness	Introduction of diversity by cloning is slow; large libraries difficult to make; not truly monovalent	Limited selection scope; technically sensitive	Sorting expertise and equipment needed; transformation efficiency

The general protocol for phage display is shown in **Figure V-14**. A large library of mutants of the targeted gene can be generated through random mutagenesis, gene shuffling or site-targeted random mutagenesis.

After cloning of the genes into the phagemid, expression of different peptides on the surface of the bacteriophage could be achieved through

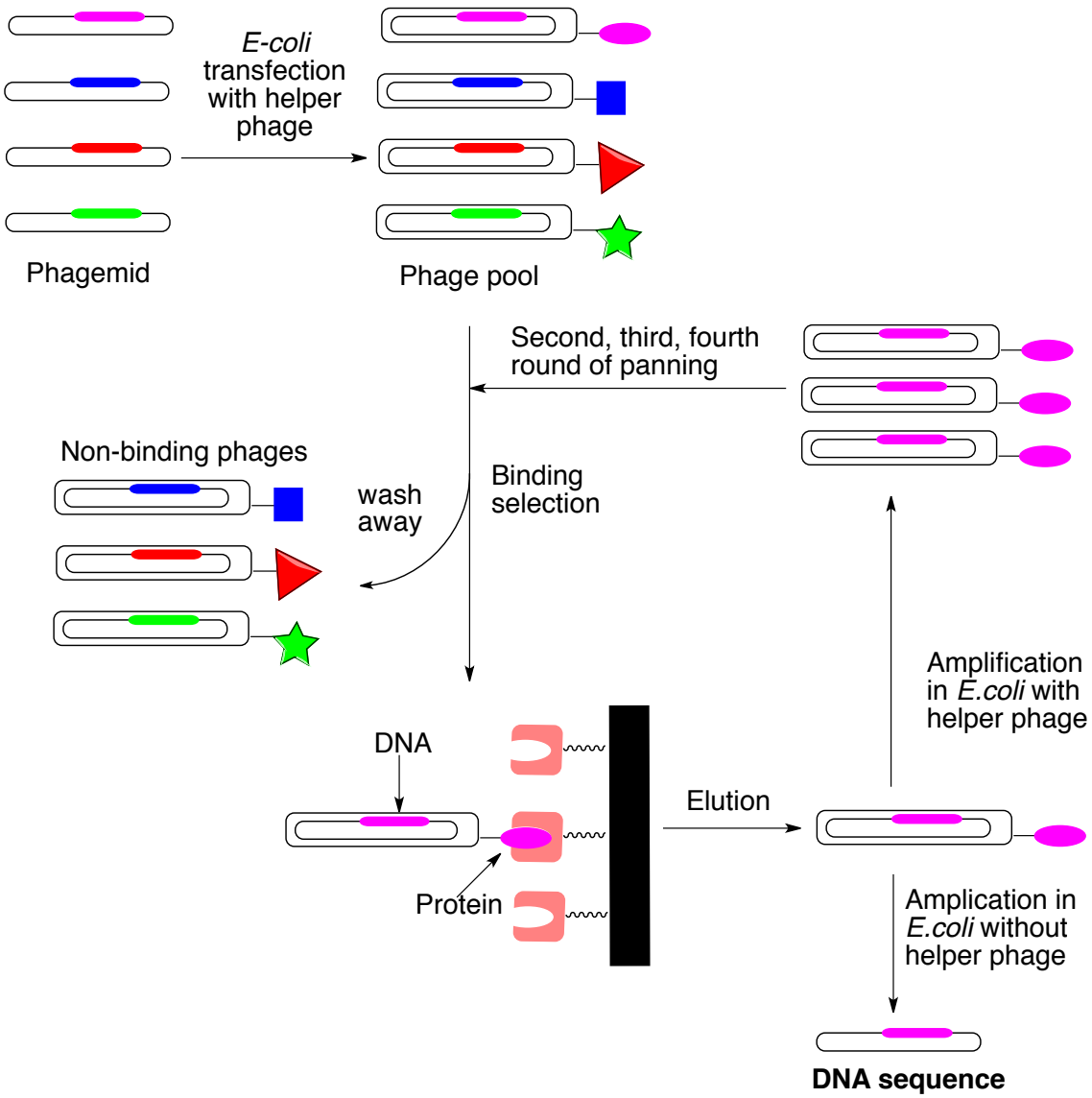


Figure V-14. General protocol for phage display. Different color coded genes represent different sequences, corresponding to the displayed protein of the same color on the surface of the phage.

transformation of the phagemid into a bacterial host, usually *E. coli* with pili, followed by infection with helper phage. The phage pools are then subjected to binding selection, also called affinity panning.

Only those phages displaying peptides that have high binding affinity toward the immobilized molecules, will be captured on the solid phase, while the unbound phages can be washed away. The bound phages are eluted and amplified through infection in host *E. coli*. bacteria again and subject to a few more rounds of panning processes, until the desired peptides are enriched by millions of fold. DNA sequences corresponding to the bound peptides are determined to assign the amino acid composition of the peptides and evaluate the important residues contributing to the binding affinity. Usually a second library of phage peptides could be generated based on the selected peptide in the first phage library, and subjected to affinity panning again with more stringent selection conditions to obtain higher binding affinity peptides.

V.4 Design of phage library based on WT CRABPII for selective binding affinity of *trans*-azobenzene derivatives

Phage display, as a powerful tool, has been applied widely to select peptides and proteins with high affinity binding towards the desired targets. Skerra's group has shown that human lipocalin 2 protein can be redesigned to bind a molecule rather different from its native ligand, through randomizing the residues in the binding cavity utilizing phage display.[26] This is encouraging for

us and similar strategies will be applied to find tight binders against *trans*-azobenzene derivatives.

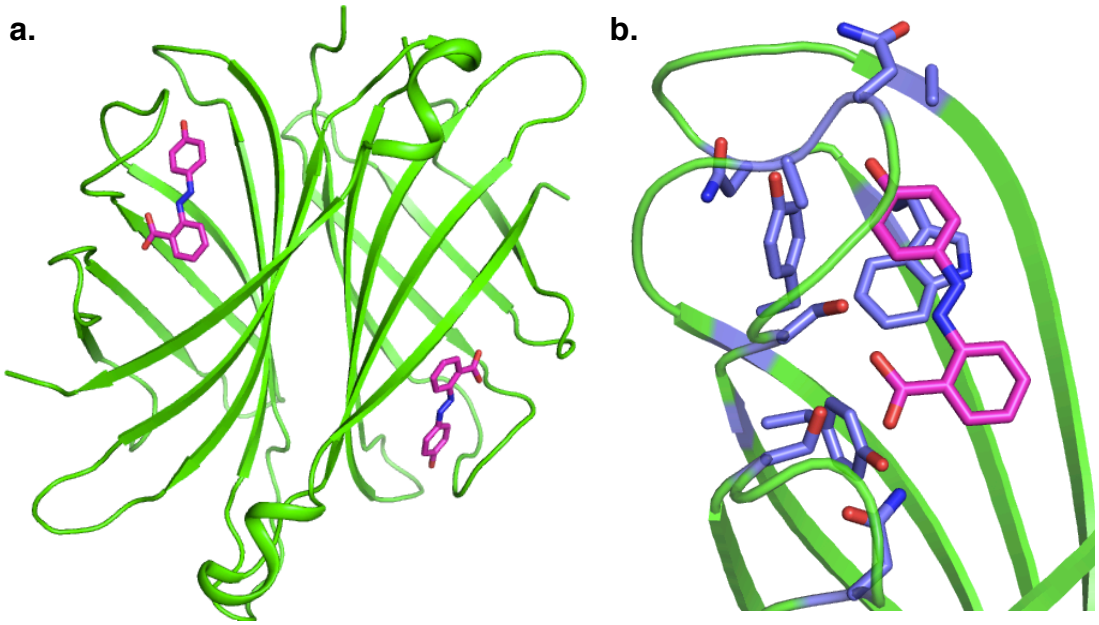


Figure V-15: Cartoon (a) and zoom in (b) of the crystal structure of streptavidin bound with 2-(4'-hydroxyphenylazo)benzoic acid (magenta). PDB entry: 1SRE.

An appropriate protein platform to start with is critical. Looking through the literature, the only protein found that has been reported to have μM affinity toward azobenzene derivatives is streptavidin.[27, 28] Crystal structures revealed that hydrogen bonding and π-π stacking interactions are important for the binding of streptavidin against 2-(4'-hydroxyphenylazo)benzoic acid (**Figure V-15**).

However, streptavidin has femtomolar binding affinity toward its native ligand, biotin, which exists in most of the cells. Biotin could inhibit the binding for azobenzene in strepatavidin, as the two molecules are bound in the same cavity. Besides, streptavidin naturally is a tetramer. Although there were strategies to perturb the tetramer interface interactions and generate a monomer streptavidin,

its affinity for its native ligand biotin drops by 10^7 fold.[29] The affinity of monomer streptavidin against 2-(4'-hydroxyphenylazo)benzoic acid is not reported, but as 2-(4'-hydroxyphenylazo)benzoic acid was bound in the same binding site as biotin, it is likely that the affinity possibly decreased dramatically as well. In conclusion, streptavidin is not a good template to start with, for designing into a protein with better affinity toward one isomer of azobenzene compounds.

CRABPII has a large binding cavity. Our work illustrated previously that CRABPII can accommodate different sized molecules. It was also observed that 11-*cis*-retinal could isomerize inside the CRABPII binding cavity, which makes it likely for the similar sterically-demanding isomerization of azobenzene to take place.[30] This is critical for designing a photoswitchable protein tag, because only

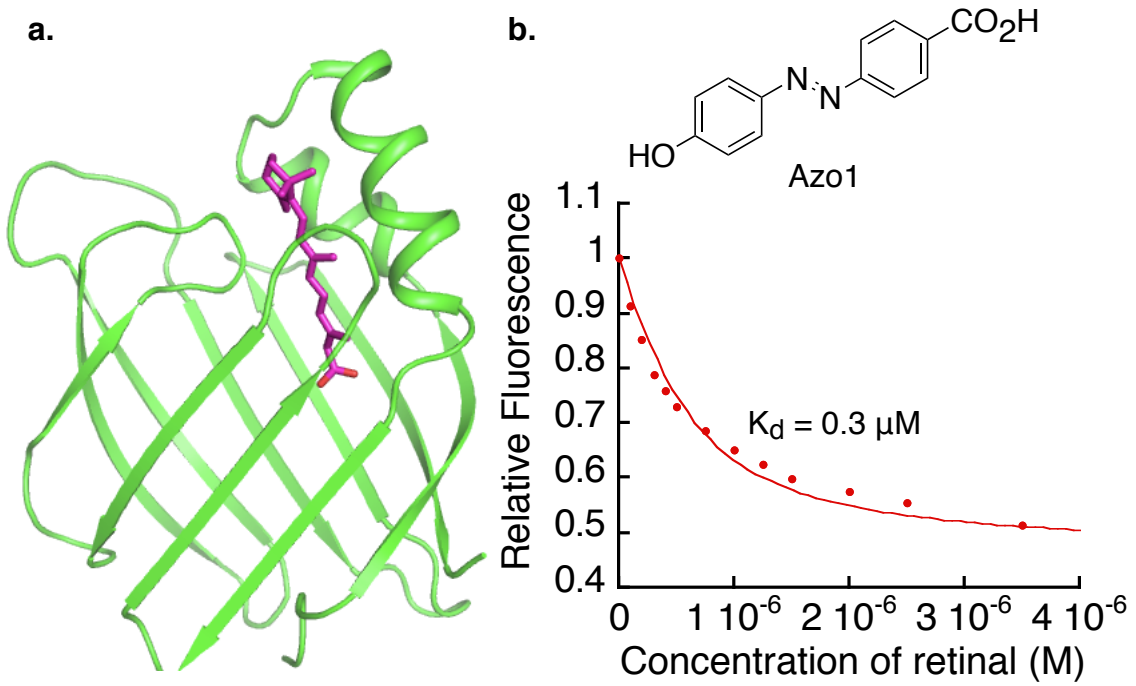


Figure V-16: a. Cartoon for the crystal structure of WT-CRABPII with retinoic acid bound. **b.** Chemical structure of Azo1 and fluorescence titration of WT-CRABPII with Azo1.

in this way will the elution of the protein tag upon UV light irradiation be efficient.

Besides, the binding pocket of CRABPII is open as shown in **Figure V-16a**, which allows the azobenzene molecules to be attached via a linkage that protrudes outside of the binding pocket and is immobilized on a solid surface. Therefore, CRABPII was chosen as the starting point for affinity maturation through randomizing important binding site residues by phage display.

Due to easy access, Azo1 (**Figure V-16b**) was tried for initial screening with WT CRABPII, showing a 0.3 μM dissociation constant. This was exciting and motivated us to move further with CRABPII for phage display.

Looking through the binding site of CRABPII, there are quite a few residues that could have potential direct interactions with the azobenzene derivatives. However, in order to limit the size of the phage library to 10^9 to 10^{10}, which is dictated by the transformation efficiency, only 8 residues in the binding site were randomized, as shown in **Figure V-17**. Even with 8 residues, the

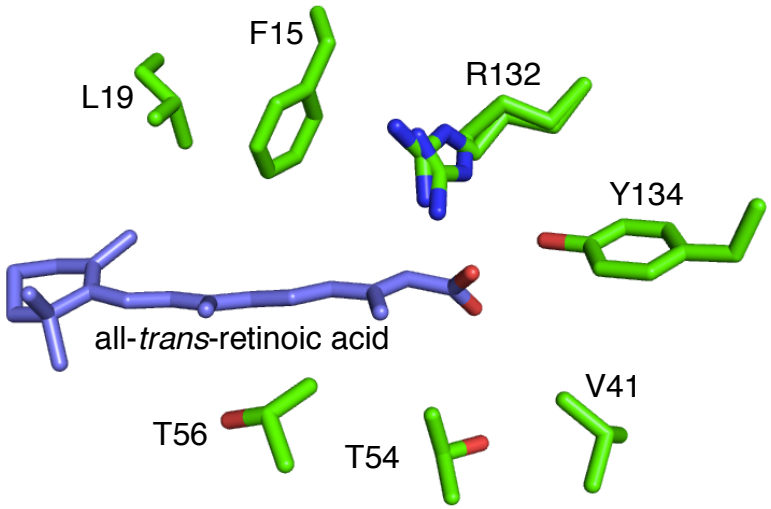


Figure V-17: Highlighted residues in the binding site of CRABPII.

variety of the phage library could be up to 20^8, as each position has the possibility to be any of the 20 amino acids. This number, 2.56×10^{10}, reaches the limit of the size of the phage library. These 8 positions were chosen for the first phage library generation, not only because they are the hot spots for possible interactions with azobenzene, but also limited by the way we are generating the libraries as shown in **Figure V-18**.

The basic strategy to generate the phage library is illustrated in **Figure V-18**. Two sets of primers, P1, P2 and P3, P4 with degenerate codons for the selected eight amino acids are used to generate two fragments P12 and P34, through the polymerase chain elongation reactions. The positions of interest are situated at least 9 bases from each end of the primers for better binding of the primers to the DNA template. The length of primers is kept shorter than 80 bases (the quality of a shorter primer is better and PCR reaction works better). P12 and P34 should overlap by at least 12 bases, so that annealing between the complimentary strands for P12 and P34 is tighter and the chances for elongation is higher during PCR.

The fragments P12 and P34 were combined and SfiI restriction cleavage site was introduced by primers P5, P6, with an SfiI site placed at least 6 bases away from the terminal, as SfiI restriction enzyme is more efficient when the cleavage site is away from the end. The fragment P56 obtained was subject to double digestion with SfiI restriction enzyme at 50 °C and then ligated with the SfiI digested phagemid vector pComb-3X overnight at RT. The degenerated

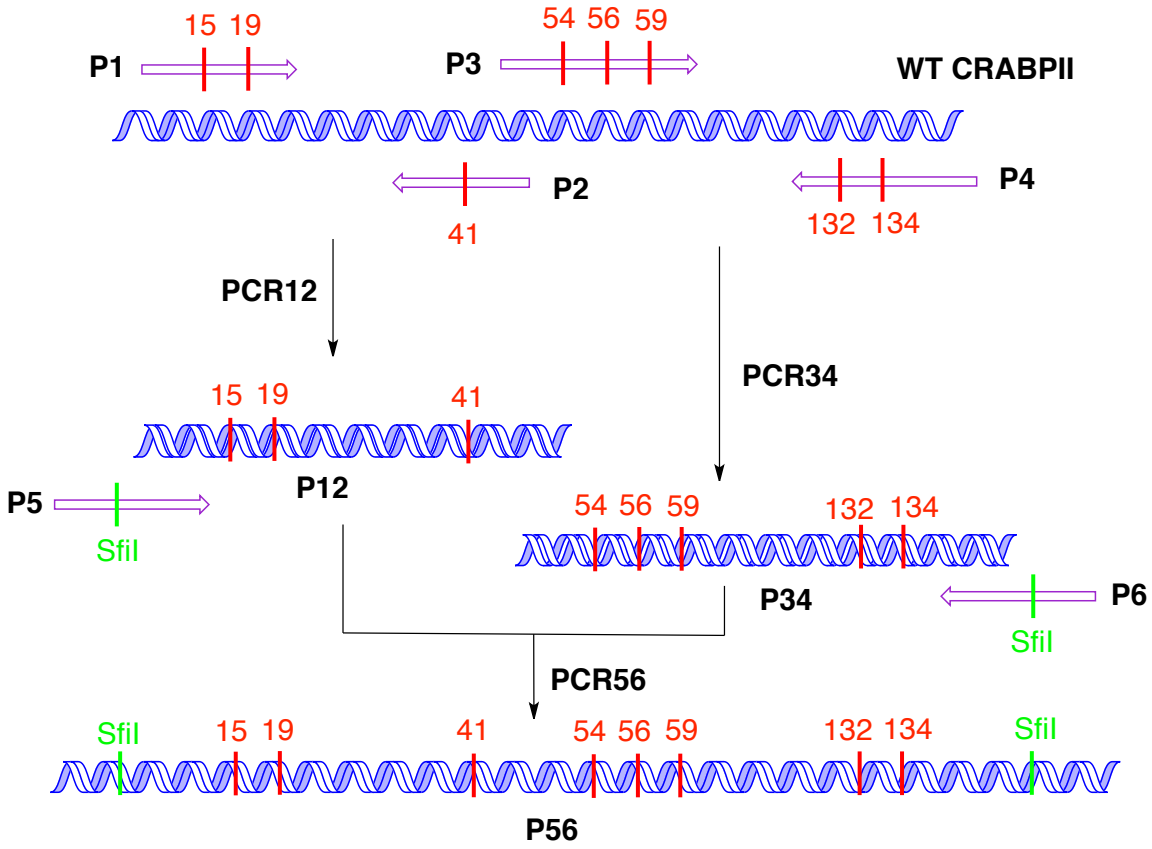


Figure V-18: General strategy to generate the phage library based on WT-CRABP gene. Two sets of primers were used to introduce degenerated sites in positions 15, 19, 41, 54, 56, 59, 132 and 134. SfiI restriction cleavage sites were introduced with primers P5 and P6.

CRABPII gene was inserted after OmpA leader peptide sequence and before Gene III of phage coating protein III, as shown in **Figure V-19a**.

Gene III encodes the phage coating protein pIII, as shown in **Figure V-19b**, which is involved in the infectious process of phage, through interaction of pIII with the pili of bacterial host. When CRABPII is coexpressed with pIII, the fusion protein can be packaged on the surface of the phage.

Remember that only when Gene III is expressed after the inserted CRABPII gene will the protein be displayed on the surface of phage. Therefore, the gene has to be in frame with Gene III during cloning. And also note that there is an Amber stop code between the inserted CRABPII gene and Gene III (**Figure V-19a**), therefore a strain of host *E. coli* bacteria with the amber suppressor gene should be used, such as XL1-Blue strain and ER2738 strain bacteria.

The ligated product was transformed into ER2738 competent cells and a phage library with variant CRABPII mutants displayed on the surface of the

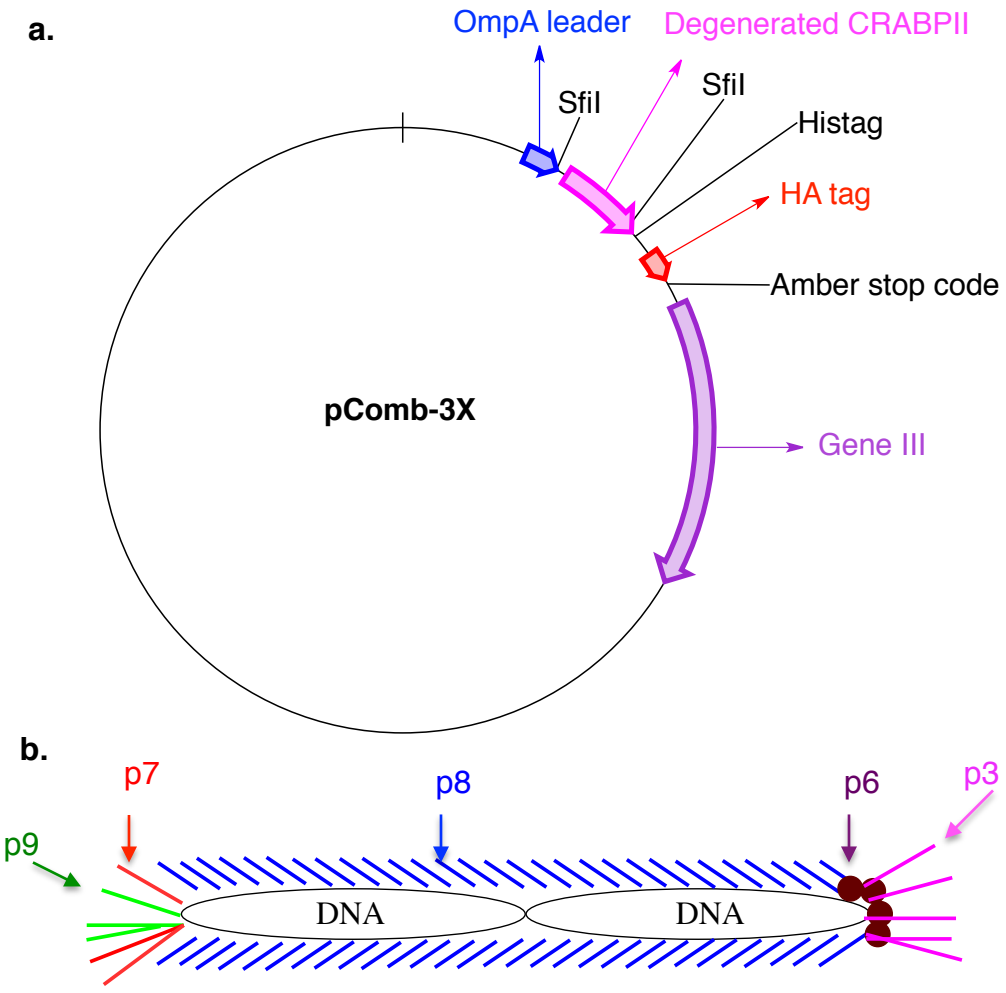


Figure V-19: Illustration of phagemid and phage. **a.** Map of pComb-3X with CRABPII inserted; **b.** Illustration of M13 phage, with the coating proteins highlighted .

phage was generated by addition of helper phage. Helper phage is necessary, as it provides all the necessary capsid proteins and packaging enzymes to help generate the phage. Low level expression of fused CRABPII mutant-phage coating protein III from phagemid makes it more likely for one phage to have only one copy of phagemid protein incorporated, in order not to lower the infection ability of the phage too much. For detailed procedures, one can follow the prototocl in *Phage Display: A Laboratory Manual*.[31]

V.5 Synthesis of azobenzene compounds and characterization

Cis-azobenzene is more polar and therefore more soluble in aqueous solution than *trans*-azobenzene, thus *cis*-azobenzene is more accessible under phage display selection processes. In order to eliminate the discrimination of selection favoring *cis*-azobenzene due to the poorer solubility of *trans*-azobenzene, a carboxylic acid moiety or hydroxyl groups are introduced to increase the solubility of azobenzene in aqueous solution. This could also minimize the degree of aggregation of the azobenzene molecules.

In order to study the photoisomerization properties of azobenzene compounds, Azo1 was synthesized first, due to its simplicity, following the Azo dye synthesis method. As shown in **Figure V-20a**, first the diazonium salt is formed by reacting *p*-amino-benzoic acid with sodium nitrite under acidic conditions, followed by nucleophilic attack from phenolate to form the final Azo1.

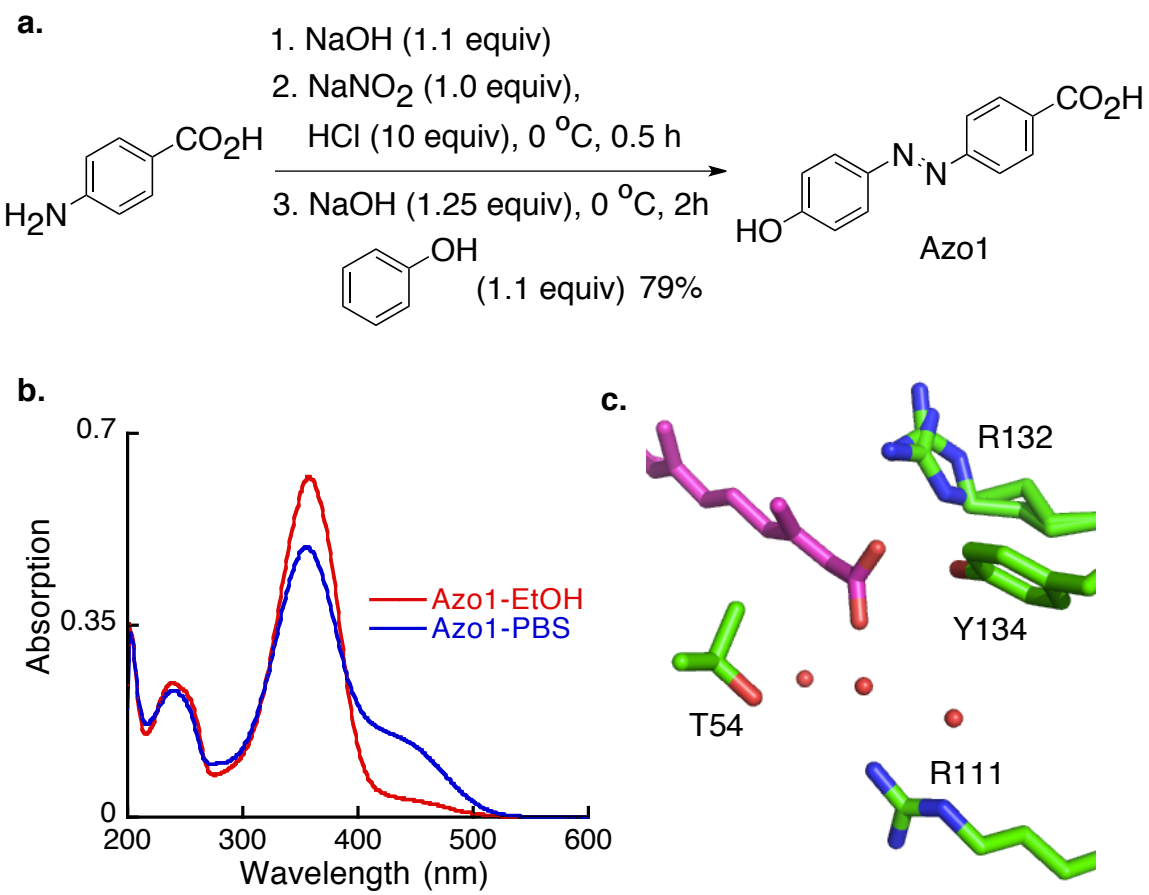


Figure V-20: **a.** Scheme of Azo1 synthesis. **b.** UV-vis characterization of Azo1. **c.** Interactions of WT-CRABPII with retinoic acid.

UV-vis absorption of Azo1 in PBS buffer is different from that in ethanol (**Figure V-20b**), probably because in PBS buffer, the hydroxyl group is partially deprotonated, which increases its electron-donating characteristics and polarizes Azo1 more with its push-pull system. UV irradiation of Azo1 did not show difference in absorption spectra, which is due to the short lifetime of *cis*-Azo1. As has been discussed previously, the electron donating groups and electron withdrawing groups generate highly polarized azobenzene compounds. This decreases the energy barrier for thermal isomerization, which greatly shortens the lifetime for the thermodynamically less stable *cis*-isomer. The time gap for

this experiment after Azo1 is irradiated is minimally 30 sec, presumably longer than the lifetime of *cis*-Azo1. Therefore the *cis*-isomer could not be observed.

Fluorescence titration of Azo1 with WT CRABPII shows a K_d of 0.3 μM, which is encouraging. The native ligand for WT-CRABPII is retinoic acid, which also has a carboxylic acid moiety. The anion binding hole in the CRABPII pocket, constituted with two arginine residues, R132 and R111 (**Figure V-20c**), probably can have favorable interactions with the carboxylic acid from Azo1. Therefore, the carboxylic acid moiety has to be maintained.

In order to lengthen the lifetime of azobenzene compounds, the electronic effects of electron donating hydroxyl groups and electron withdrawing carboxylic acid groups have to be attenuated. Azo2 was prepared to completely remove the push-pull system in Azo1, but still maintain the carboxylic acid.

A similar scheme, like the synthesis of Azo1, with diazonium salt as the intermediate, was applied to synthesize Azo2 (**Figure V-21a**), but was not successful. This is because benzene is not as good a nucleophile as phenolate to attack the diazonium salt and form the product, Azo2. Instead, the synthesis of Azo2 follows the scheme as shown in **Figure V-21b**. First, nitrosobenzene was generated by reduction of nitrobenzene to phenylhydroxylamine, followed by oxidation with sodium perchromate. Then *p*-aminophenyl-acetic acid coupled with nitrosobenzne to yield Azo2. This is a more general scheme for synthesizing azobenzene compounds.

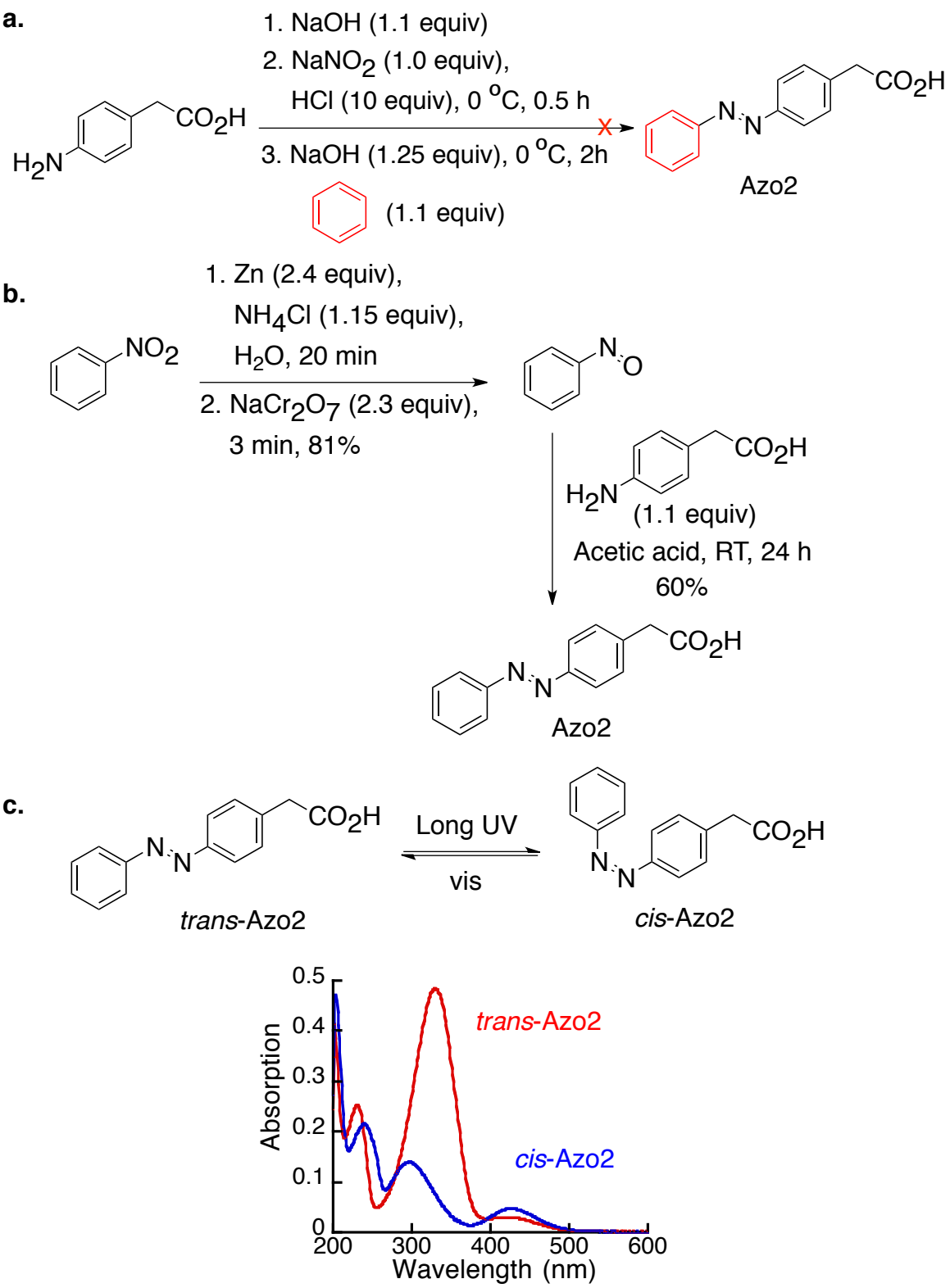


Figure V-21: (a) The failed and (b) successful synthesis of Azo2, and (c) UV-vis characterization of *trans*-Azo2 and *cis*-Azo2.

UV-vis and NMR studies showed that Azo2 isomerizes readily under long UV light to *cis*-Azo2, as shown in **Figure V-21c**. This agrees with the longer lifetime of *cis*-Azo2. However, the affinity of Azo2 for WT CRABPII drops dramatically, which is probably due to the unfavorable conformation of carboxylic acid and its interactions with R132 and R111 in the binding pocket.

Azo3 was also prepared following the scheme in **Figure V-22a**, with *p*-amino-benzoic acid coupled with nitrosobenzene. The hydroxyl group of Azo1 is removed to partially abolish the push-pull system in Azo3. Different from Azo2, the carboxylic acid is directly attached to the benzene ring, in order to restore the binding with WT-CRABPII. Azo3 is expected to isomerize and also bind to WT CRABPII. Interestingly, the isomerization characteristics of Azo3 are similar to Azo2, but the binding affinity of Azo3 for WT CRABPII, 570±26 nM, is close to Azo1. This further supports the importance of attaching the carboxylic acid directly to the benzene ring.

As shown in the right panel of **Figure V-22b**, addition of Azo3 quenches the fluorescence of tryptophan, due to the short distance of Azo3 from the tryptophans in WT CRABP, when it is bound in the binding cavity. Irradiation with UV light isomerizes Azo3 from *trans* to *cis,* resulting in less fluorescence quenching. Irradiation with visible light recovers the fluorescence quenching to the original level. This photoswitching of fluorescence quenching levels could be cycled several times, showing the reversible isomerization process of Azo3.

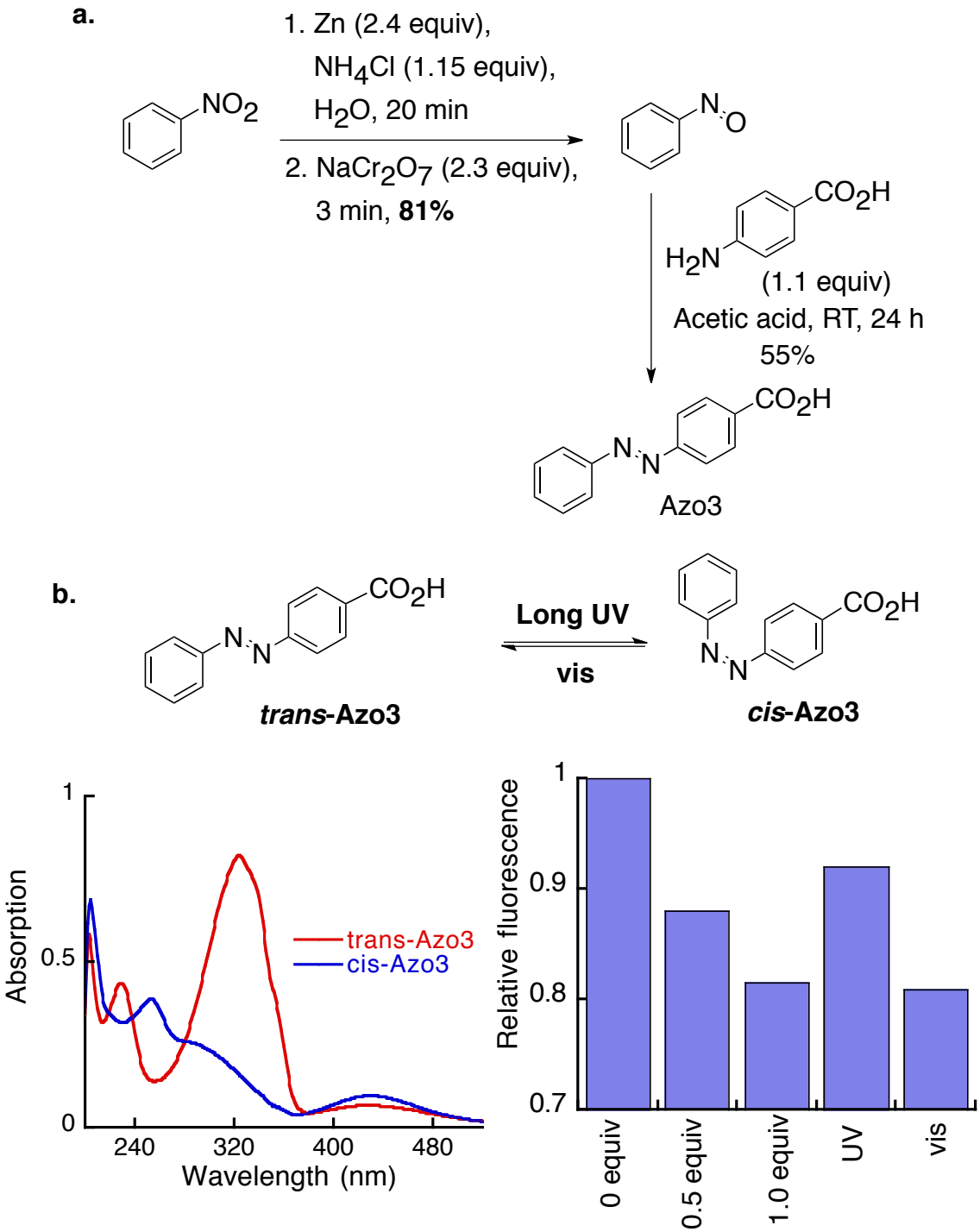


Figure V-22: Synthesis and characterization of Azo3. **a.** Scheme of the Azo3 synthesis. **b.** UV-vis spectra of isomerization of Azo3 from *trans* to *cis* with long UV light (left) and tryptophan fluorescence quenching experiment using Azo3 (right).

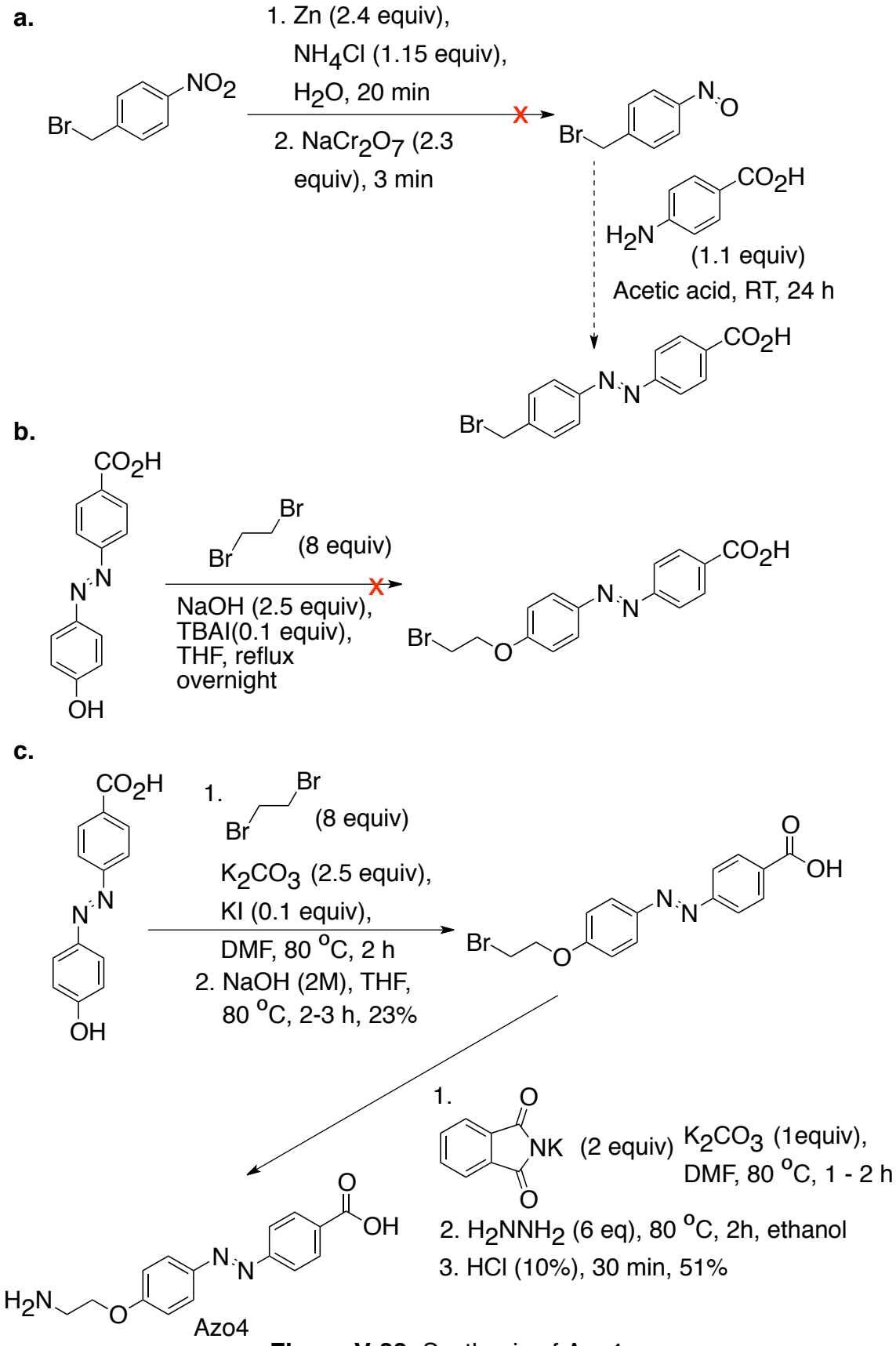


Figure V-23: Synthesis of Azo4.

The change of fluorescence quenching level upon isomerization could be because of less binding affinity of *cis*-Azo3 for WT-CRABP. It could also be due to less absorption of *cis*-Azo3 at 350 nm, the emission maximum of tryptophan, resulting in less quenching. If it is the former case, it indicates that WT-CRABPII preferably favors *trans*-azobenzene.

Encouraged by the latter result, an azobenzene compound with carboxylic acid directly attached to the benzene, like Azo3, but with an amine group tethered on the other end, facilitating immobilization on solid surfaces, will be synthesized. As shown in **Figure V-23a**, 4-bromomethylnitrobenzene was used to obtain 4-bromomethylnitrosobenzene, following similar procedures as discussed previously. 4-Bromomethylazobenzene could be obtained and converted to an amine through Gabriel amine synthesis. However, the first step of the reaction did not work. This is because 4-bromomethylnitrobenzene is solid, while nitrobenzene is liquid. The difference makes the reduction of 1-bromomethylnitrobenzene with zinc in aqueous solution extremely difficult. This route was not pursued further.

Instead Azo1 was used as a starting material and alkylated with 1,2-dibromoethane to form the corresponding ether. The product could be converted to an amine by Gabriel amine synthesis. However, the S_N2 reaction in THF, as shown in **Figure V-23b,** is sluggish and mostly starting material was recovered.

DMF was suggested to be a solvent more suitable for S_N2 reaction and the reaction was carried out in DMF as shown in **Figure V-23c**. The reaction

proceeded much faster, but both the carboxylate and phenol group reacted with 1,2-dibromoethane. Therefore, a large excess of 1,2-dibromoethane was added to ensure full conversion in the first step. The crude product was subjected to basic conditions, in order to hydrolyze the ester group. The bromo-compound was converted to the amine, yielding the final product, Azo4.

In Azo4, the oxygen forms an ether linkage, which greatly reduces its electron donating characteristics and increases the lifetime of *cis*-isomer of Azo4. The amine group introduced will be a handle to immobilize the molecule on the solid surface for phage display and also for future photoswitchable protein purification.

Azo4 can readily undergo isomerization under irradiation with long UV light or visible light as shown in **Figure V-24**. An NMR study revealed that the photostationary state of Azo4 under long UV light is a mixture of 92% *cis*-Azo4 and 8% *trans*-Azo4, while the photostationary state of Azo4 under visible light or room light is a mixture of 75% *trans*-Azo4 and 25% of *cis*-Azo4. 100% of *trans*-Azo4 could be achieved by thermal relaxation at 60°C in less than an hour (NMR studies). Fluorescence quenching experiments with *trans*-Azo4 and *cis*-Azo4 **(Figure V-24)** shows that WT-CRABPII binds *trans*-Azo4 with about three fold higher affinity. This is a good starting point to achieve higher binding affinity of WT-CRABPII for *trans*-Azo4.

Azo1, Azo2, and Azo3 are all soluble in polar solvents such as methanol, ethanol, acetone and DMSO. However, Azo4 is not soluble in most solvents,

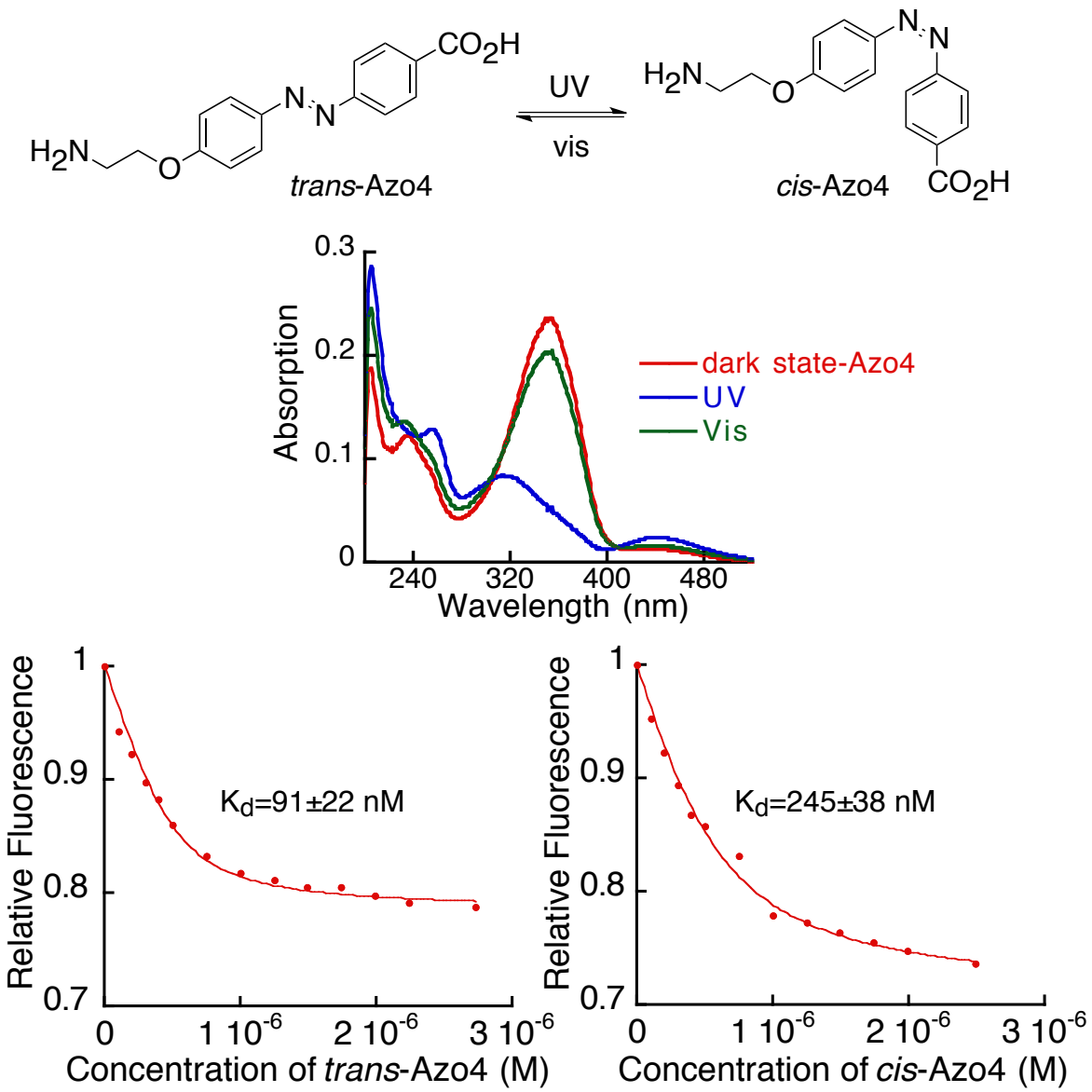


Figure V-24: Reversible isomerization of Azo4 and fluorescence quenching titration.

polar or nonpolar. Its solubility in polar solvent, like acetone and methanol increases when acidified or basified. This is unexpected, as Azo4 is likely in its zwitterionic form.

It is not clear why suppression of one of the charges, either carboxylate or protonated amine, could lead to higher solubility in polar solvent. It is possible that the compound forms an aggregate easily through intermolecular salt bridge

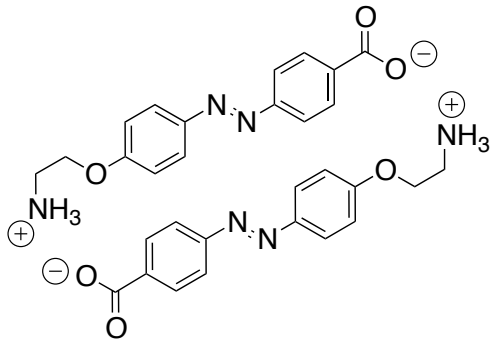


Figure V-25: Illustration of the possible aggregates formed by Azo4.

and π-π stacking (**Figure V-25**), which makes it highly stable and hard to solubilize. Neutralization of one of the charges would break the two salt bridges and make the aggregates less stable and thus soluble in polar solvent, which could solubilize the mono-charged Azo4.

For phage selection, Azo4 is to be immobilized on a solid surface. NHS (N-hydroxysuccinimide) activated magnetic beads, which react with primary amine, was first tried. However, Azo4 did not seem to react with the NHS ester of the magnetic beads. The only possible way to quantify how much Azo4 reacts with the magnetic beads is either through subtraction of the total amount of Azo4 added by the amount unreacted measured by UV-vis absorption, or through hydrolysis of the magnetics beads to quantify how much Azo4 is hydrolyzed. The first method was applied and it was found that Azo4 did not react at all. As magnetic beads are red in color, it is hard to tell whether Azo4 was coupled to the magnetic beads by color change. These magnetic bead are expensive as well and not practical for future purification of protein. Therefore, we switched to NHS-activated agarose beads for immobilization of Azo4 molecules.

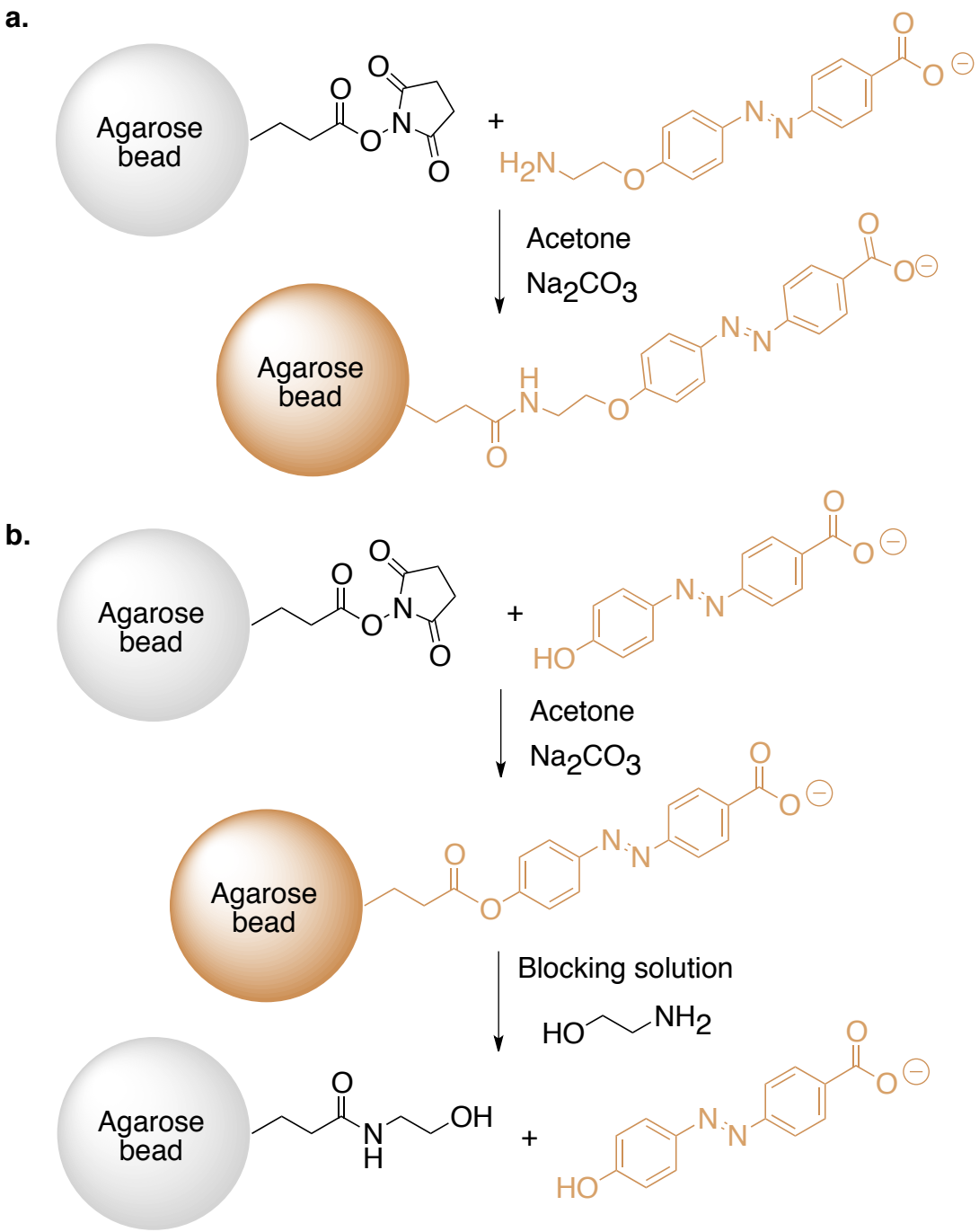


Figure V-26: Reaction scheme for NHS-activated agarose beads reaction with Azo4 and Azo1.

Since NHS-activated Agar beads are preserved in dry acetone, the coupling reaction was set up in dry acetone as well with a few chips of sodium

carbonate added to increase the solubility of Azo4 (**Figure V-26a**). Reaction was shaken vigorously at RT for 2 hours before it was quenched with amino-ethanol solution and washed with water and ethanol a couple of times to collect all the unreacted Azo4 for measuring the coupling efficiency. The coupling efficiency of Azo4 was found to be ~2 μmol/mL of resin, close to the reported value by the company. As agar beads are white and Azo4 is yellow, coupling of Azo4 to the agar beads can be visualized through the color change as well.

Attempts to couple Azo1 to the NHS-activated agar beads through ester bond formation with the phenolic hydroxyl group failed, as the ester bond formed is unstable and was cleaved during the blocking stage with amino-ethanol (**Figure V-26a**). It is observed that the yellow color added to agar beads was cleaved upon amino-ethanol addition, supposedly the amino-ethanol reacted with the Azo1 ester.

With Azo4 immobilized on a solid surface and the phage library generated, the stage is ready for biopanning to select phages that display CRABPII mutants that bind to *trans*-Azo4.

V.6 Biopanning

For selection of phages from the phage libraries, the general procedure follows the protocol from Phage Display: A Laboratory Manual.[31] The general scheme is illustrated in **Figure V-27**. First, the phage library is incubated with the functionalized and sterilized agarose beads. For initial screening, optimal

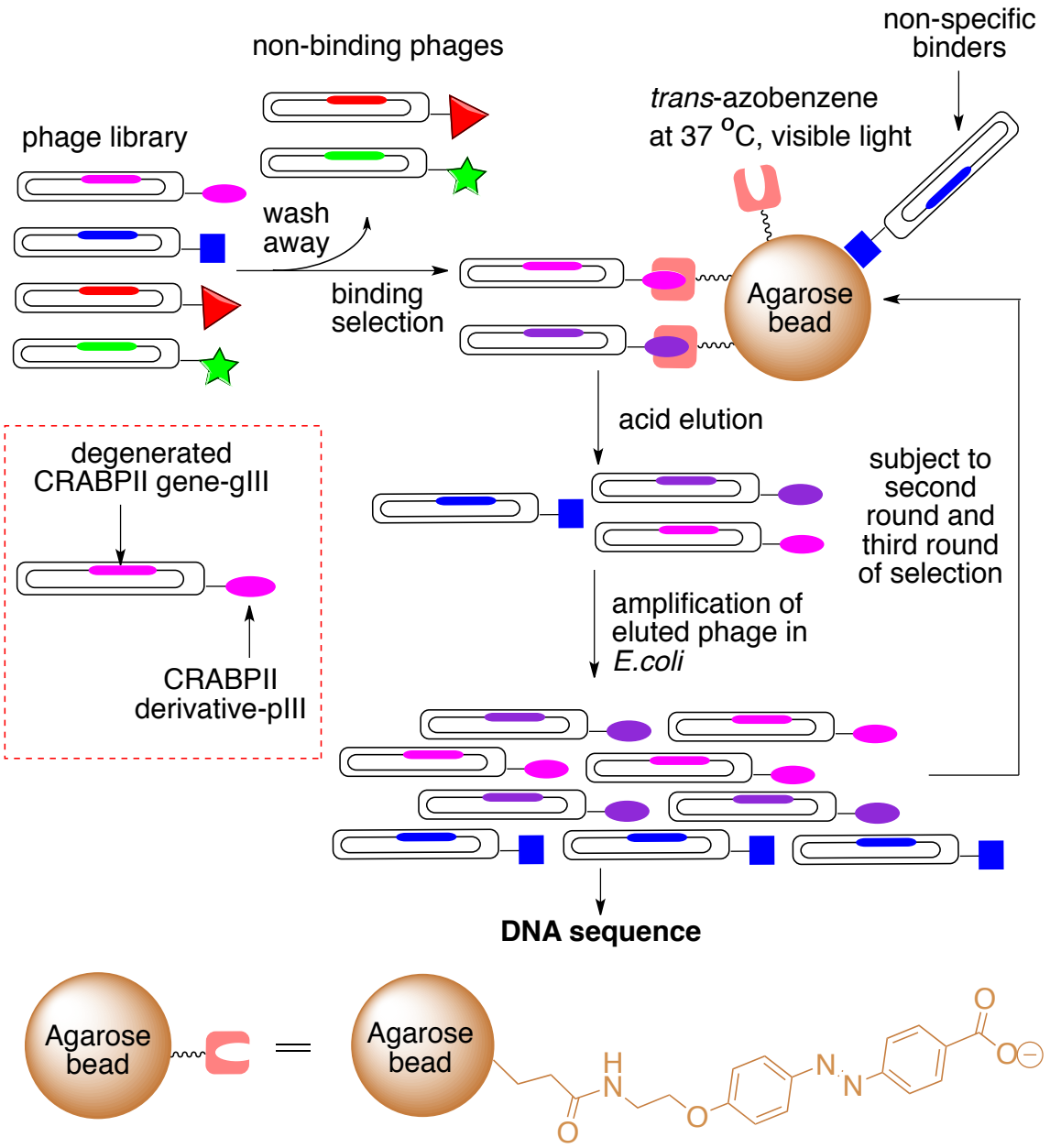


Figure V-27: General scheme for biopanning.

condition (pH~8, no competing reagent added) was used to look for binders with both high affinity and low affinity. BSA was added to reduce non-specific binding.

Nonspecific acid elution was used to find some binders first. After identifying some good binders, specific elution method would be used, such as addition of *trans*-Azo4 molecule to compete for binding with phage to elute out

the phage that binds to immobilized *trans*-Azo4. A negative selection would entail subjecting the eluted phage back to agarose beads with Azo-4 already irradiated with UV light (*cis*-form). The unbound phage will be collected to ensure that the negatively selected phage in this step does not bind to *cis*-Azo4.

The amount of acid eluted phage in the first round was determined through titering. Simply put, the eluted phage (in various dilution) was allowed to infect bacteria and was plated on agarose plates with ampicillin. Only the bacteria that pick up the phage and have phagemid will be resistant to ampicillin and can survive. The count of the colonies will tell how many phages were eluted in total. Usually in the first round of panning, the number of phage in 1 μL of eluted phage solution ranges from 10 to 1000.

The eluted phage from the first round was amplified by infection of the *E. coli* together with helper phage. Usually the same phage could be amplified by up to 1000 copies. Therefore, after subjecting the amplified phage back to the second round of panning, it will enrich the specific binders, but not non-specific binders. Titering of the second round eluted phage should get a number 10 to 1000 times larger than the first round. Similarly the third round eluted phage will also be 10 to 1000 times larger than the second round. If this trend is obtained, that means likely some specific binders are found. After 3 to 4 rounds, 10 to 20 colonies could be picked from the titering of the eluted phage and the DNA is extracted for sequencing. The pattern of mutations could be obtained to find out which mutations frequently show up in the selected phage.

I have tried to generate the library twice and pan it twice. The first library generated was not good, due to technical problems (PCR and ligation). The randomness in the library is limited. Phages that have CRABPII in frame with gIII in that library was estimated to be ~10% of the whole library, based on the sequencing results of 8 clones. Out of the 8 clones, four of them do not have CRABPII inserted. Of the four clones with CRABPII gene, one of them has base insertions and deletions. As a result, the whole gene is out of frame and the corresponding protein and pIII protein could not be expressed and displayed on the surface of phage. Two of them have stop-codon replacing the original amino acid, which terminates the expression before pIII. Only one clone is in frame, without stop codon, and can express CRABPII together with pIII protein.

The second phage library generated was better in terms of CRABPII gene being in frame of gIII. However, ligation did not work well, as the amount of degenerated CRABPII gene from PCR was not enough. As a result, after the electroporation, roughly only 10^8 varieties of phages were obtained, as identified by titering. For both libraries, no enrichment was observed during the affinity panning. The number of phage obtained after each round stayed the same, which implied that they are non-specific binders.

The efficiency of PCR was greatly improved by addition of 10% DMSO in the PCR reaction solution. This shows that the low efficiency of PCR is probably due to the aggregation of the primers, as the primers used are up to 70 bases,

which can adopt different conformational states. DMSO could disrupt those interactions and free the primers to bind to DNA for initiation of the PCR reaction.

At this point, no positive hit has been obtained. This could be due to many reasons. Although Azo4 is coupled to agarose beads, whether the solubility of Azo4 is good enough and whether the linker is long enough to allow the Azo4 molecule to be accessible by the phage is not know. The use of polyethyleneglycol linker could increase Azo4 solubility in water and also make the AZO4 more accessible.

The other problem might lie in the design of the library. The variety of the library generated might not be large enough or maybe the positions selected are not the best positions to affect the folding and interaction of the protein with Azo4. If this is the case, different sites could be tried.

Usually, a positive control is used in phage display to make sure the selection method and the procedure of library generation and phage amplification are OK. However, in our case, it is hard to immobilize retinoic acid on solid surface or on beads to act as a positive control.

Materials and methods

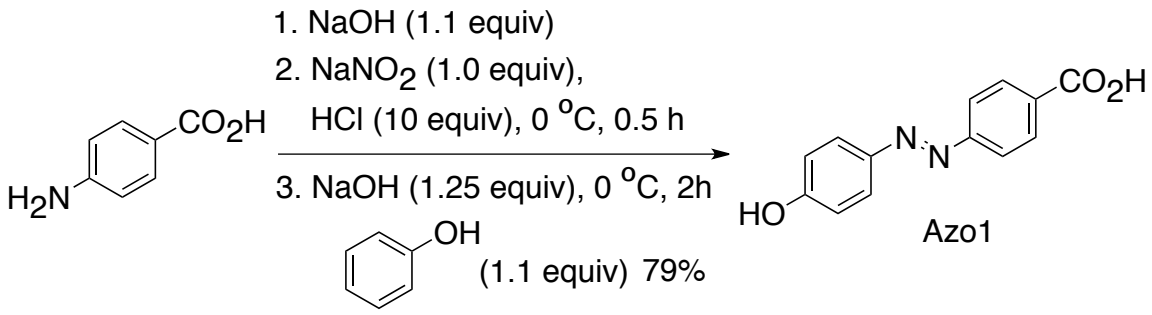


Figure V-28: Synthesis of Azo1.

Sodium hydroxide (75mmol, 3 g, 1.1 equiv) was dissolved in water (150 mL) and *p*-aminobenzoic acid (68 mmol, 9.32 g, 1 equiv) was added into this alkaline solution. The solution was cooled to 0 °C and sodium nitrite (68 mmol, 4.68 g, 1 equiv) was added. The mixture was stirred in ice bath until all the solid was dissolved. This solution was transferred slowly into concentrated chloric acid (20 mL, 720 mmol 10.6 equiv) with ice (20 g) in ice bath. The solution was stirred for additional 30 min at 0 °C.

Phenol (75 mmol, 7.04 g, 1.1 equiv) and sodium hydroxide (85 mmol, 3.4 g, 1.25 equiv) were mixed in water (50 mL) and cooled down in ice bath. The dissolved phenoxide solution was pipetted dropwise to the latter solution, while stirring at 0 °C. The slurry was stirred for 1 h at 0 °C. The product was collected by filtration and recrystallized in ethanol as orange crystals in 79% yield. ^{1}H NMR (500 MHz, CD_3OD): δ 6.96 (d, *J*=9 Hz, 2H), 7.89 (d, *J*=9 Hz, 2H), 7.92 (d,

J=9 Hz, 2H), 8.19 (d, *J*= 8.5 Hz, 2H) ppm. ^{13}C NMR (500 MHz, CD_3OD) δ 169.5, 163.2, 157.1, 147.9, 133.2, 132.1, 126.7, 123.5, 117.1 ppm.[32]

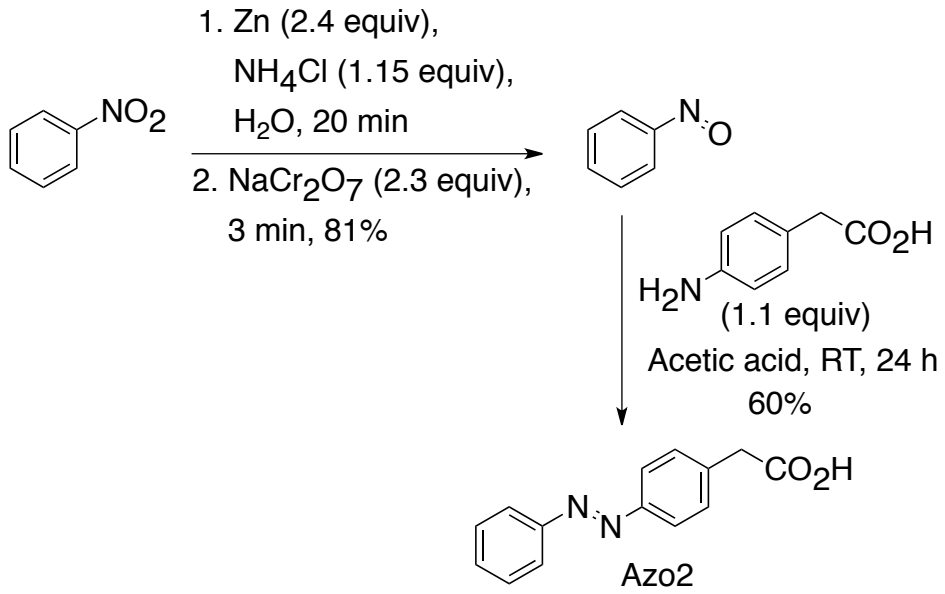


Figure V-29: Synthesis of Azo2.

Zinc dust (58 mmol, 3.8 g, 2.4 equiv) was added to a stirred suspension of nitrobenzene (24.4 mmol, 3.0 g, 1.0 equiv) in water (50 mL) containing ammonium chloride (28 mmol, 1.5 g, 1.15 equiv) at RT. Stirring was continued for 20 min after addition was complete. The slurry solution was filtered and washed with hot water (~90 °C, 60 mL). Crushed ice (20 g) was added to the aqueous solution, followed by concentrated sulfuric acid (7.5 mL) and more crushed ice (10 g) at 0 °C. To the rapidly stirred acid solution was added rapidly an ice-cold solution of sodium dichromate dehydrate (17 g, 2.3 equiv) in water (7.5 mL). After 3 min the product was collected by filtration and washed twice

with water as straw-colored solid. Crude nitrosobenzene (~2.1 g) was obtained in 81% yield, mp 67 °C.[33]

To a hot solution of *p*-aminophenyl-acetic acid (0.5 g, 3.6 mmol, 1.1 equiv) in glacial acetic acid (3 mL) was added nitrosobenzene (0.356 g, 3.3 mmol, 1 equiv). The reaction was allowed to stand for more than 24 h. The orange precipitate was filtered and washed with acetic acid (20%) and water and recrystallized with 95 % ethanol (8 mL). Azo2 (0.475 g, 2.0 mmol) was obtained in 60% yield. ^{1}H NMR (500 MHz, CD_3OD): δ 3.76(s, 2H), 7.51(m, 5H), 7.91(m, 4H) ppm. ^{13}C NMR (500 MHz, CD_3OD) δ 175.3, 154.3, 153.2, 139.9, 132.5, 131.7, 130.6, 124.2, 124.1, 42.0 ppm.[34]

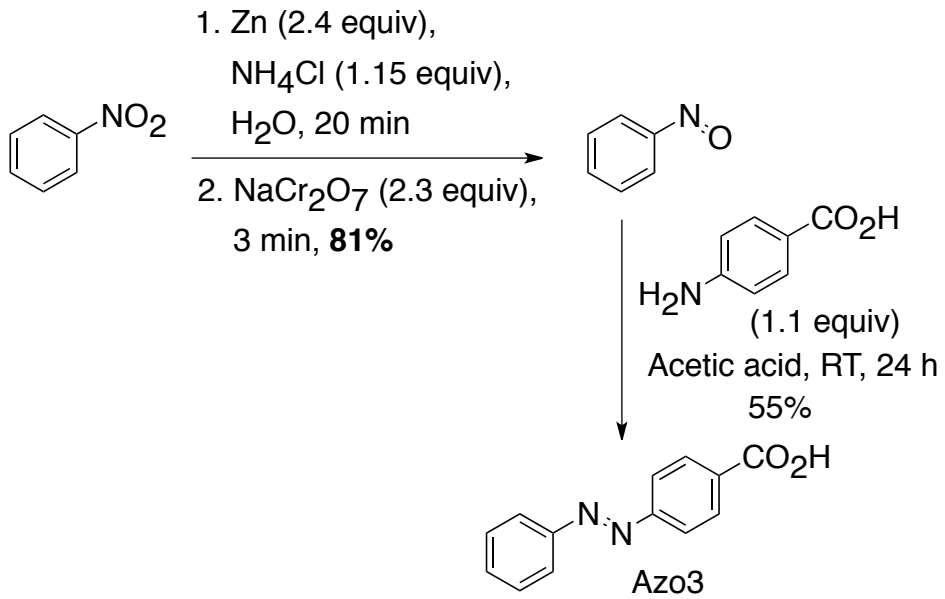


Figure V-30: Synthesis of Azo3

To a hot solution of *p*-aminobenzoic acid (0.48 g, 3.5 mmol) in glacial

acetic acid (3 mL) was added nitrosobenzene (0.35 g, 3.2 mmol). The reaction was allowed to stand for more than 24 h. The orange precipitate was filtered and washed with acetic acid (20%) and water and recrystallized with 95% ethanol (8 mL). Azo3 (0.39 g, 1.7 mmol) was obtained in 55% yield.

^{1}H NMR (500 MHz, CD_3OD) δ 7.60 (m, 3H), 7.99 (dd, *J*=8.5 Hz, *J*=2.0 Hz, 2H), 8.0 (d, *J*=8.4 Hz, 2H), 8.23 (d, *J*=9 Hz, 2H) ppm. ^{13}C NMR (500 MHz, CD_3OD): δ 169.9, 156.8, 154.4, 135.1, 133.4, 132.3, 130.8, 124.6, 124.0 ppm.

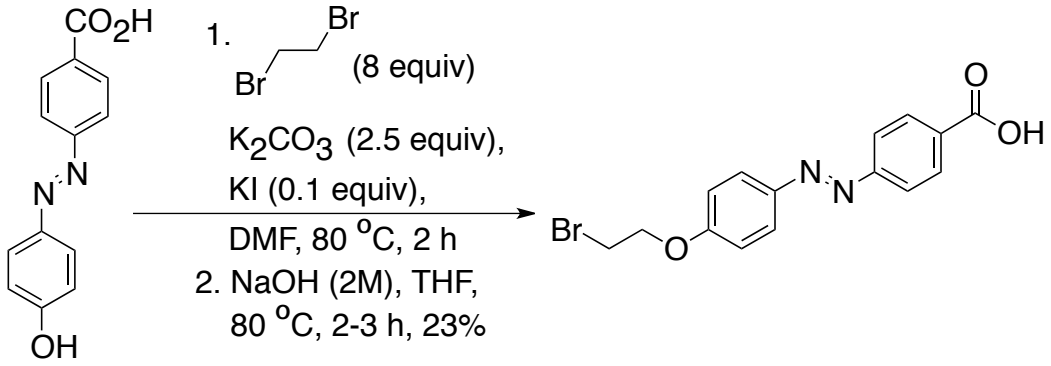


Figure V-31: Synthesis of 4-((4-(2-bromoethoxy)phenyl)diazenyl)benzoic acid

Azo1 (12 mmol, 3 g, 1 equiv) was dissolved in DMF (70 mL). Potassium carbonate (29 mmol, 4.08 g, 2.5 equiv), sodium iodide (1.2 mmol, 0.176 g, 0.1 equiv) and 1,2-dibromoethane (96 mmol, 17.8 g, 8 equiv) were added to the solution at RT. The mixture was stirred at 80 °C for 2 h. Water (50 mL) was added to quench the reaction. The product was extracted with dichloromethane (20 mL) twice and the organic layer was combined and washed with water (10 mL) 8 times, followed by brine solution once and dried with sodium sulfate. The organic solvent was removed by rotavapping, to get a mixture of mono and di-

alkylated products. A flash silica gel column was run with 50% dichloromethane and hexane solvent system to roughly separate the dialkylated product.

The dialkylated product was dissolved in THF (30 mL) and sodium hydroxide solution (30 mL, 2 M). The solution was kept stirring at 80 °C for 3 h till all the ester was hydrolyzed. The reaction was quenched with concentrated chloric acid till pH=3. The solution was extracted with ethyl acetate (15 mL) three times. The organic layer were combined and washed with water (10 mL) and brime solution (20 mL), dried with sodium sulfate. The solvent was removed by rotavap to get an orange solid with an overall yield of 23%. ^{1}H NMR (500 MHz, CD_3OD) δ 3.79 (t, *J*=5.5 Hz, 2H), 4.45 (t, *J*=5.5 Hz, 2H), 7.14 (d, *J*=9 Hz, 2H), 7.87 (d, *J*=8.5 Hz, 2H), 7.97 (d, *J*=9.5 Hz, 2H), 8.11 (d, *J*=9.0 Hz, 2H).

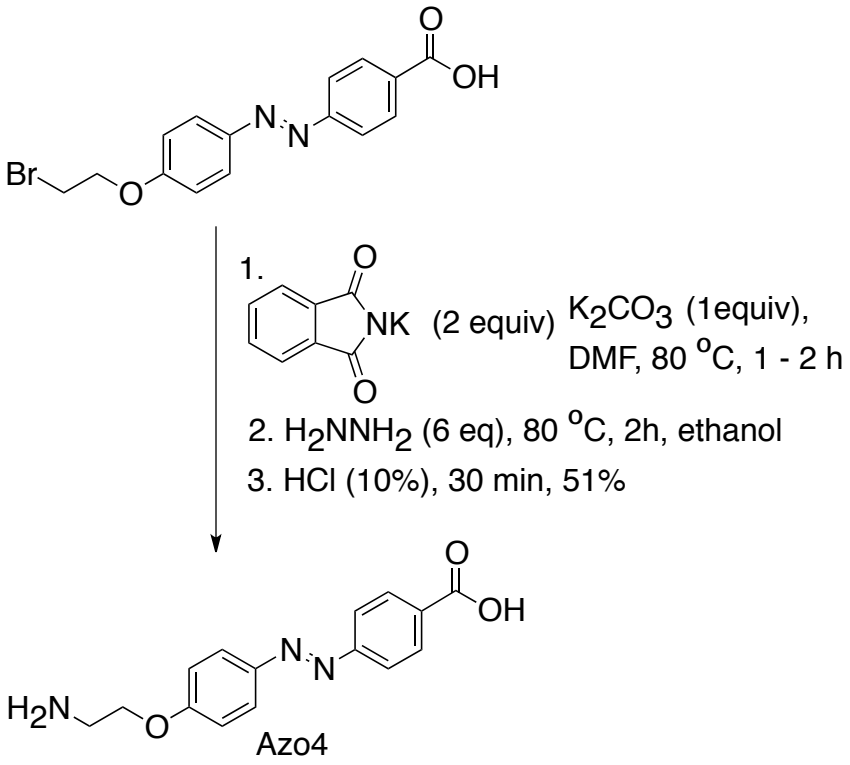


Figure V-32: Synthesis towards Azo4

4-((4-(2-bromoethoxy)phenyl)diazenyl)benzoic acid (1.45 mmol, 0.5 g, 1 equiv) was dissolved in DMF (30 mL). Potassium carbonate (1.45 mmol, 0.2 g, 1 equiv) and potassium phthalimide (2.9 mmol, 0.53 g, 2 equiv) were added and stirred at 80 °C for 2 h. The reaction was quenched with 10% chloric acid to make the final pH=4. The product was extracted with ethyl acetate (20 mL) twice and the organic layer was washed with water (15 mL) eight times, followed by brime solution (10 mL) and dried with sodium sulfate.

Ethanol (20 mL) was added to the solid obtained. Hydrazine monohydrate (800 µL, 11 mmol) was added to the hot slurry. The reaction was kept stirring at 80 °C for about 1 h until a lot of precipitate formed. The reaction was quenched with 10% chloric acid until pH=7. The precipitate was filtered and washed with a large amount of ethyl acetate.

The orange solid obtained was washed with hot ethanol and then recrystallized with small amount of sodium hydroxide solution and ethanol. ^{1}H NMR (500 MHz, CD_3OD): δ 3.07 (t, *J*=5.0 Hz, 2H), 4.15 (t, *J*=5.0 Hz, 2H), 7.15 (d, *J*=9.5 Hz, 2H), 7.86 (d, *J*=9.0 Hz, 2H), 7.96 (d, *J*=9.0 Hz, 2H), 8.11 (d, *J*=8.5 Hz, 2H). ^{13}C NMR (500 MHz, CD_3OD): δ 174.9, 163.5, 155.4, 148.7, 141.3, 131.4, 126.2, 123.1, 116.2, 71.3, 42.1 ppm.

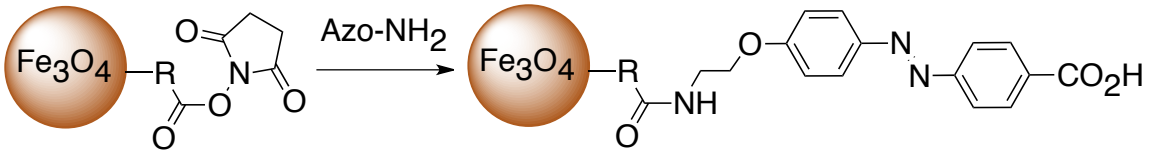


Figure V-33: Immobilization of Azo4 on magnetic beads.

The sodium salt of Azo4 (7 mg, 23 µmol, 10 equiv) and potassium carbonate (3.2 mg, 23 µmol, 10 equiv) was dissolved in DMF (800 µL) and dry acetone (200 µL). This solution was added to the magnetic beads (9 mg, 2.3 µmol, 1 equiv) in a 1.7 mL Epperdorf tube. The reaction was shaken vigorously for 1 h at RT. Water (200 µL) was added to the reaction and it was shaken overnight at RT. The magnetic beads were spun down and washed with ethanol and water. The beads were incubated with blocking buffer (3-hydroxylpropylamine, 2 M, sodium hydroxide, 20%, 1 mL). Then the beads were washed 4 times with PBS buffer.

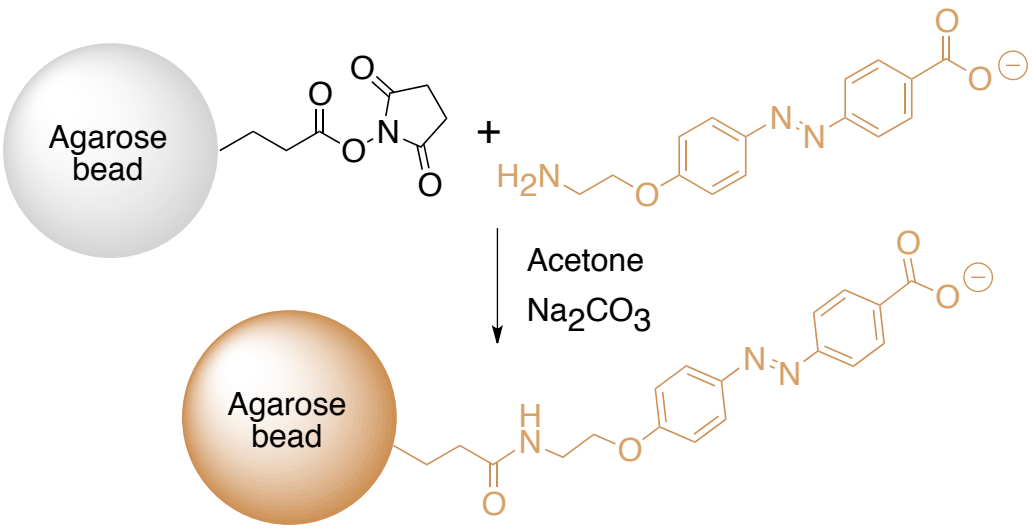


Figure V-34: Coupling of Azo4 to the agarose beads.

Azo4 (0.24 mg, 0.84 µmol) was dissolved in dry acetone (300 µL) with sodium carbonate (1 mg). The solution (20 µL) at this stage was saved in an eppendorf tube for measuring of Azo4 coupling efficiency later). NHS-activated agarose beads in dry acetone (400 µL, 50% v/v) was added to the Azo4 solution and kept shaking vigorously at RT for 2 h.

Ethanol (500 µL) and water (200 µL) were added to the reaction and it was kept shaking for another 5 min. The agarose beads were spun down and the supernatant was gently pipetted out into a 15 mL orange tube. Blocking buffer (1 M aminopropanol solution, pH=9.0, 1 mL) was added to the beads and it was kept shaking for another 15 min. The beads were spun down and the supernatant were pipetted out and combined with the previous supernatant solution. The beads were washed several times with ethanol and water and combined with the latter supernant solution. The unbound Azo4 was measnured by UV-vis spectrometry. The amount of Azo4 coupled to the beads could be obtained through subtraction of the added Azo4 by the unreacted Azo4.

UV study of isomerization of different Azo compounds

As all the Azo compounds are recrystallized in hot ethanol, therefore *trans*-azobenzene is predominant in the dark state.

Azo compounds were dissolved in ethanol and absorption spectrum were recorded from 600 nm to 200 nm. The UV cuvette was irradiated with hand held long UV light for 1 min and absorption spectrum was recorded again. For Azo4, after UV irradiation, bright projector light was applied to isomerize *cis*-Azo4 back to *trans*-Azo4 and UV-vis spectrum was measured again.

NMR studies of isomerization of Azo4

Azo1, Azo2, Azo3 were dissolved in deuterated methanol. Azo4 was dissolved in deuterated methanol as well, but with addition of sodium carbonate (1 mg) for better solubility. For isomerization from *trans* to *cis*-isomer, the NMR

tube was kept in the cold room and irradiated with hand-held long UV light overnight to reach photostationary state. As NMR tube blocks some of the UV light and much higher concentration solution was used in NMR studies than UV-vis spectral studies, much longer time was required to isomerize *trans* to *cis*-azobenzene than in UV cuvet.

After UV irradiation, the NMR tube was left at RT exposed to projector white light and NMR was taken to reach the photostationary state under visible light.

NMR tube was covered with alumina foil and placed in a 60 °C water bath for 1 h and NMR data was collected. In all of these NMRs, the integration of the aromatic hydrogens were used to determine the ratio of *trans-cis* isomer.

Phage library construction

WT CRABP DNA sequence

ATG CCA AAC TTC TCT GGC AAC TGG AAA ATC **ATC** CGA TCG GAA AAC TTC GAG GAA TTG CTC **AAA** GTG CTG GGG GTG AAT GTG ATG CTG AGG **AAG** ATT GCT GTG GCT GCA GCG TCC AAG CCA **GCA** GTG GAG ATC AAA CAG GAG GGA GAC ACT **TTC** TAC ATC AAA ACC TCC ACC ACC GTG CGC **ACC** ACA GAG ATT AAC TTC AAG GTT GGG GAG **GAG** TTT GAG GAG CAG ACT GTG GAT GGG AGG **CCC** TGT AAG AGC CTG GTG AAA TGG GAG AGT **GAG** AAT AAA ATG GTC TGT GAG CAG AAG CTC **CTG** AAG GGA GAG GGC CCC AAG ACC TCG TGG **ACC** AGA GAA CTG ACC AAC GAT GGG GAA

CTG **ATC** CTG ACC ATG ACG GCG GAT GAC GTT GTG **TGC** ACC AGG GTC TAC GTC CGA GAG TGA

WT CRABP amino acid sequence

M P N F S G N W K I I R S E N F E E L L K V L G V N V M L R K I A V A A A S K P A V E I K Q E G D T F Y I K T S T T V R T T E I N F K V G E E F E E Q T V D G R P C K S L V K W E S E N K M V C E Q K L L K G E G P K T S W T R E L T N D G E L I L T M T A D D V V C T R V Y V R E

Primers for library generation

Two sets of primers have been tried.

The first trial, which failed, because the gene was out of frame with the start codon of the vector and also primer P4 was too long in order to introduce degenerated codon for three amino acids. As a result, the PCR did not work well. SacI and SpeI restriction cleavage sites were introduced. These two restriction cleavage sites could show up in the degenerated sites inside the CRABPII gene, while SfiI is more unique and chances for these sites to show up inside the CRABPII gene is small. Therefore, more varieties could be obtained with SfiI restriction cleavage site.

The first set of primers

Primer **P1** covers 15, 19; primer **P2** covers 41;

Primer **P3** covers 54, 56, 59; primer **P4** covers 121, 132, 134.

Primers **P5, P6** introduce the restriction cleavage sites on both ends of the CRABP gene.

Symbol for degenerated bases:

M (AC); W (AT); Y (CT); V (ACG); D (AGT); N (AGTC); R (AG); S (CG); K (GT); H (ACT); B (CGT).

P1: 5'-C ACC TGG AAA ATC ATC CGA TCG GAA AAC NNK GAG GAA TTG NNK AAA GTG CTG GGG GTG AAT GTG-3'

P2 (reverse complimentary): 5'-GAA AGT GTC TCC CTC CTG TTT GAT CTC KNN TGC TGG CTT GGA CGC TGC AGC-3'

P3: 5'-C AAA CAG GAG GGA GAC ACT TTC TAC ATC AAA NNK TCC NNK ACC GTG NNK ACC ACA GAG ATT AAC TTC AAG G-3'

P4 (reverse complimentary): 5'-CT GCA GAA TTC TCA CTC TCG GAC KNN GAC KNN GGT GCA CAC AAC GTC ATC CGC CGT CAT GGT KNN GAT CAG TTC CCC-3'

P5 (SacI restriction site introduction)

5'-**GAG CTC** AAC TTC TCT GGC AAC TGG AAA ATC ATC CGA TCG-3'

P6 (Reverse complimentary, SpeI restriction site introduction)

5'-AC TAG TCT GCA GAA TTC TCC CTC TCG GAC-3'

NOTE: P6 made the gene out of frame, missing one base, should be

5'-ACT AGT ACT GCA GAA TTC TCC CTC TCG GAC-3'

Primers for trial 2

P1 covers 15, 19; **P2** covers 41; **P3** covers 54, 56, 59; **P4** covers 132, 134

P1: 5'-C AAC TGG AAA ATC ATC CGA TCG GAA AAC NNK GAG GAA TTG NNK AAA GTG CTG GGG GTG AAT GTG-3'

P2 (reverse complimentary): 5'-GAA AGT GTC TCC CTC CTG TTT GAT CTC KNN **TGC** TGG CTT GGA C-3'

P3: 5'- C AAA CAG GAG GGA GAC ACT TTC TAC ATC AAA NNK TCC NNK ACC GTG NNK ACC ACA GAG-3'

P4 (reverse complimentary): 5'-CT GCA GAA TTC TCC CTC TCG GAC KNN GAC KNN GGT GCA CAC AAC GTC-3'

P5-SfiI

5'-CAT GCC ATG ACT GTG GCC CAG GCG GCC AAC TTC TCT GGC AAC TGG AAA ATC ATC CGA TCG-3'

P6-SfiI (reverse complimentary)

5'-CAC AGT CAT GGC ATG CTG GCC GGC CTG GCC CCT GCA GAA TTC TCC CTC TCG GAC-6'

Eight PCR reactions were set up to amplify segment **P12** and segment **P34** applying the following protocols.

PCR for P12	**PCR for P34**
Water (dd)	Water (dd)
DMSO (5 µL)	DMSO (5 µL)
WT CRABPII DNA, 100 ng	WT CRABPII DNA 100 ng
Primer P1, 100 pmole	Primer P3, 100 pmole
Primer P2, 100 pmole	Primer P4, 100 pmole
Buffer (10x), 5 µL	Buffer (10x), 5 µL
dNTP (10mM), 2 µL	dNTP (10mM), 2 µL
Turbo polymerase, 1 µL	Turbo polymerase, 1 µL
Volume in total = 50 µL	Volume in total = 50 µL

PCR Thermal Control Cycles

1x	95 °C	3 min
30x	95 °C	30 sec
	55 °C	50 sec
	72 °C	50 sec
1x	72 °C	10 min
1x	25 °C	10 min

Note: 10% DMSO is found to be important for PCR reactions to work when the primer is long.

Agarose DNA gel (1%) electrophoresis was performed to isolate the PCR product. The corresponding band of PCR product was cut out. The DNA was extracted using QIAGEN Gel Exctraction Kit. The concentration of fragments P12 and P34 were roughly estimated by comparison of ethidinium bromide fluorescence with a series of standard DNA solution or NanoDrop.

PCR for P56

Water (dd)
DMSD (5 µL)
DNA fragment P12 (100 ng)
DNA fragment P34 (200 ng)
dNTP (10mM), 2µL
Buffer (10x), 5.5 µL
Turbo polymerase, 1µL

Volume in total = 50 µL

Primer P5, 100 pmole
Primer P6, 100 pmole

PCR Thermal Control Cycles

1x	95 °C	3 min
30x	95 °C	30 sec
	55 °C	50 sec
	72 °C	1 min
1x	72 °C	10 min
1x	25 °C	10 min

Eight PCR reactions were set up to combine fragment P12 and P34 and introduce SfiI cutting site on both ends using primers P5-SfiI and P6-SfiI. The PCR reaction solutions to ligate fragment P12 and P34 were set up by addition of all the above solutions except primers P5 and P6. After finishing the PCR thermal cycles, P5 and P6 were added to the eight PCR reaction tubes and 0.2 µL extra turbo polymerase was added and subjected to the same PCR cycles.

Agarose DNA gel (1 %) was run to isolate the PCR product. The DNA was extracted from the cut out band using QIAGEN Gel Exctraction Kit. Both the PCR product CRABP-P56 and pComb-3X were digested using SfiI restriction endonuclease enzyme in a 50 °C water bath for 5 h. Agarose DNA gel (1 %) was run again to isolate the double digested DNA, followed by DNA extraction using QIAGEN Gel Exctraction Kit. The concentration of the digested DNA was estimated using NanoDrop or UV spectrometer by determining the absorption at 260 nm.

Ligation of digested pComb-3X and CRABP-P56 at RT overnight.

Control Ligation	**Phage library Ligation**
Water(dd)	Water(dd)
pComb-3X SfiI cut, 100 µg	pComb-3X SfiI cut, 1.4 µg
CRABP-P56 SfiI cut, 0 µg	CRABP-P56 SfiI cut, 0.7 µg
Buffer(5x), 4 µL	Buffer(5x), 40 µL
T4 ligase 1 µL	T4 ligase 10 µL
Volume in total=20 µL	Volume in total=200 µL

The ligation product was heat inactivated at 60 °C for 2 min. Then the ligated DNA was purified by QIAGEN PCR purification kit. The DNA was eluted with two portions of ddWater (15 µL and 10 µL). DNA gel electrophoresis was performed to check whether there are any ligated product, which was supercoiled and should travel faster than the digested vector.

Constructruction of phage library through electroporation to transform the ligated DNA into cells

1. The purified ligation product (2 µL out of 25 µL) was subject to DNA gel electrophoresis to make sure the ligation worked.

2. The electroporation cuvette and ER2738 electrocompetent cells (purchased from Lucigen) were incubated on ice for 10 min.

3. Six reactions of competent cells (25 µL) and purified ligated DNA (2 µL) were premixed on ice.

4. The mixed cells were transferred to a 1 mm electroporation cuvet (Bio-Rad). The moisture was wiped off the cuvette before electroporation. Ecor1 electroporation set up was applied for electroporation using Bio-Rad Micropulser electroporator. Immediately after the pulse, recovery media (1000 µL), which comes with the purchased competent cells, was added to the cuvette. The cells were pipetted out and transferred to a 10 mL tube for each reaction and kept shaking at 37 °C for 1 h.

5. For tittering, the 1 mL cell culture from all the six reactions were combined first and 2 µL of the combined cells were diluted into 1000 µL LB

media to get 1/500th diluted solution 1. The diluted solution (2 µL) was further diluted into 1000 µL LB to get 1/250000th diluted solution 2. Solution 1 (5 µL) was plated and solution 2 (1 µL , 2 µL , 10 µL , 50 µL) were also plated on agarose plates with Ampicillin, Tetracycline and LB.

$$\text{Number of phages} = \frac{number\ of\ colonies}{dilution\ times * volume} * 6000\ \mu L$$

(Ten colonies were inoculated for miniprep DNA purification and sequencing to check the diversity of the library.)

6. The combined cell culture (6 mL) was added to SB media (500 mL) and ampicillin (50 mg/mL, 400 µL) was added. The culture was shaken at 37 °C for 1 h at 220 RPM. Another portion of ampicillin (50 mg/mL, 600 µL) was added and shaken for another hour until the OD600 reached 0.4 to 0.6.

7. The amplified helper phage (250 µL, titerring=1015) was added to the cell culture and incubated at 37 °C for half an hour. Then the cells were shaken for 1.5 h before kanamycin (50 mg/mL, 700 µL) was added.

8. The cells were shaken overnight at 37 °C at 250 RPM.

Purification of the phage library

1. The overnight cell culture was spun down at 4 °C at 8500 RPM for 30 min. The supernatant was transferred to two sterile 500 mL centrifuge bottles, followed by addition of 20% volume of the sterile PEG8000/NaCl solution (20% PEG, 2.5 M NaCl) and incubated no ice for 1 h. The phage particles were collected by centrifugation at 9000 RPM for 20 min at 4 °C.

2. The phage pellet was resuspended with PBS buffer (40 mL) and transferred into two 30 mL sterile centrifuge bottles. The PEG8000/NaCl solution (20% volume of the supernatant) was added and incubated on ice for 1 h. The phage was collected by centrifugation at 10,000 RPM for 25 min at 4 °C. The phage pellet was resuspended in PBS (5 mL, 0.02% sodium azide, and 1 pellet of Roche complete protease inhibitors).

3. The solution was centrifuged at 10,000 RPM at 4 °C for 10-20 min to spin down bacteria. The supernatant solution was passaged through a 0.4 μm sterile filter to further separate phage from bacteria. The solution was fractioned in sterile eppendorf tubes in portions of 500 μL and frozen in -80 °C freezer for future use.

1st Round of Panning

1. The functionalized agar beads were incubated with methanol (1 mL) overnight, then washed several times with sterile PBS buffer.

2. ER2738 cells was inoculated in LB solution (2 mL). The cell culture (500 μL) was later transferred into LB solution (10 mL) 1 h before phage elution and let shaken at 37 °C until OD600 reached 0.4-0.8.

3. The sterile agar beads were incubated with sterile BSA (600 μL, 2.5% in PBS) at 37 °C in the shaker for 1 h (37 °C is necessary to ensure the azobenzene is adopting *trans*-form).

4. The beads were spun down and the BSA solution was pipetted out.

5. The phage library (500 μL) was mixed with sterile BSA solution (100 μL, 5% in PBS) and added to the processed azo-beads. The slurry was shaken at 250 RPM at 37 °C for 1.5 h.

6. The beads were spun down and the unbound phage was pipetted out. The beads were washed seven times with PBST solution (0.05% Tween 20 in PBS, 1.2 mL). During each wash, the supernatant was pipetted out as much as possible without disturbing the azo-beads.

(Washing method was modified later to use a 0.45 μm centrifugal filter to get rid of the unbound phage. Short UV light was used to irradiate the centrifugal filter for 20 min to sterilize it before use. The beads together with the phage were transferred to the centrifugal filter and spun down to wash away the unbound phages. PBST (300 μL) was added to the beads and the filter was centrifuged to facilitate the solution to pass through. This procedure was repeated 6 to 7 times. Then the beads were transferred to a sterile eppendorf tubes with two portions of elution buffer (0.1 M Glycine, pH=2.0, 300 μL × 2). The beads were shaken for 15 min, then spun down. The supernatant was pipetted into an eppendorf tube with Tris base solution (2 M, 36 μL) for neutralization.)

7. The phage was eluted with Glycine buffer (0.1 M, 400 + 200 μl, pH=2.0). The solution along with beads shaken at 37 °C for 15 min. The beads were spun down and the supernatant was pipetted into an eppendorf tube and neutralized with Tris base (2 M, 36 μL).

8. Five eppendorf tubes of cell culture (1 mL) were spun down at 4000 RPM for 5 min at RT and resuspended in 50 µL of LB solution for tittering of the eluted phage.

9. For tittering, $1/10^{th}$, $1/100^{th}$, $1/1000^{th}$ dilution of the eluted phage were prepared. Different dilutions of the eluted phage (1 µL) were added to the resuspended cells and incubated at 37 °C for 0.5 h. The cells were plated in prewarmed Amp/Tet/LB agar plates.

Note: Affinity panning is a messy process, therefore a waste bucket with 20% bleach solution was prepared and all the washing solution and pipets could be sterilized in the bleach solution to reduce contamination of phage. And after each step, if possible the bench was bleached with 20% bleach solution to minimize contamination. Cells before addition of phage could be processed in a different room from the room where phage is processed to reduce chances of contamination.

Amplification of phage from 1^{st} round of panning

1. ER2738 cells were inoculated overnight.

2. The overnight cell culture (1 mL) was transferred to LB solution (10 mL) with tetracyclin (12.5 mg/L) and shaken at 37 °C for a couple of hours until the OD_{600} reached 0.4-0.8. The cells were spun down at 4,000 RPM at RT for 5 min and resuspended with LB solution (1 mL).

3. The eluted phage from first panning was incubated with 600 µL of the resuspended cells at 37 °C for 20 min. Then the cells were transferred to 20 mL

of SB media (30 g tryptone, 20 g yeast extract, 10 g MOPS (3-[N-morpholino]-propanesulfonic acid), pH=7.0) with ampicillin (final concentration=100 mg/L) and tetracyclin (final concentration=12.5 mg/L). The media was shaken at 280 RPM at 37 °C for ~6 h, until OD600 reached 0.4 (It was observed that first some of cells died because they are not resistant to ampicillin, but later the cell density increased as the population of ampicillin resistant cells grew). Helper phage (50 µL, titerring=10^{15}) was added to the cells and incubated for 0.5 h before kanamycin (final concentration=70 mg/L) was added. The cell culture was shaken overnight at 37 °C.

Purification of amplified 1st round panning eluted phage

1. The overnight cell culture was splitted into two 30 mL sterile centrifuge bottles and spun down at 10,000 RPM for 15 min to remove the cell pellets.

2. The supernatant was decanted to a sterile tube. PEG-NaCl solution (20% volume of the supernantant) was added to the supernatant and incubated on ice for 1 h. Some cloudiness should be seen. The phage particles were collected by centrifugation at 12,000 RPM for 20 min.

3. The phage pellet was resuspended with PBS buffer (1 mL).

4. The phage was incubated at 70 °C for 20 min to kill bacteria. At this stage, it was ready for the 2nd round of panning.

(It was later realized that incubation at 70 °C is only suitable for phages displaying short peptides, but not big proteins that will be denatured at 70 °C.)

5. The 2nd round of panning and 3rd round of panning followed the same protocol as the 1st round of panning. If the result is positive, after each round of panning, the count of eluted phage should increase by 10 to 1000 folds each round, as the selected phage is amplified and selected in each round, but for non-specific binding phage, the number of bound phage remains the same during each selection.

Amplification of helper phage

1. ER2738 cells were inoculated in LB media (3 mL) with tetracycline (final concentration=12.5 μg/mL) overnight.

2. The overnight culture (1 mL) was transferred into SB media (500 mL) with tetracycline (12.5 μg/mL) and shaken at 37 °C for 2 h. Then helper phage (Invitrogen, 1 mL, 10^{12} tittering) was added and incubated at 37 °C for 0.5 h first and shaken at 37 °C for 2 h before kanamycin (70 mg/L) was added. The cell culture was shaken at 37 °C overnight.

3. The overnight cell culture was spun down by centrifugation at 4 °C at 8,000 RPM for 30 min. The supernatant was transferred to two sterile centrifuge bottles and PEG8000/NaCl solution (20% volume of the supernatant) was added and the mixture was incubated on ice for 1 h. The phage particles were collected by centrifugation at 8,000 RPM for 50 min at 4 °C. The phage pellet was resuspended with PBS (40 mL) and transferred into two 30 mL centrifuge bottles. PEG8000/NaCl solution (20% volume of the supernatant) was added and

incubated on ice for 1 h. The phage was collected by centrifugation at 10,000 RPM for 25 min. The phage was resuspended with PBS (5 mL) containing 0.02% sodium azide and 1 pellet of Roche complete protease inhibitors. The phage solution was centrifuged at 8,000 RPM at 4 °C for 10 min and the supernatant was passaged through a 0.4 μm filter. The solution was fractioned in portions of 500 μL to epperdorf tubes and frozen in -80 °C freezer.

Determination of the titer of helper phage

The overnight ER2738 culture (300 μL) was transferred to SB media (3 mL) and shaken for 1 h at 37 °C. Helper phage with different dilution (10^{-9} and 10^{-12}, 1 μL) was added to 50 μL of the cells. The cells were incubated at RT for 20 min and then plated on agar plates with Kanamycin (70 mg/L).

References

References

1. Cusido, J.; Deniz, E.; Raymo, F. M., Fluorescent Switches Based on Photochromic Compounds. *Eur. J. Org. Chem.* **2009,** (13), 2031-2045.

2. Yager, K. G.; Barrett, C. J., Novel photo-switching using azobenzene functional materials. *J. Photoch. Photobio. A* **2006,** *182* (3), 250-261.

3. Wei, Y. H.; Han, S. B.; Kim, J.; Soh, S. L.; Grzybowski, B. A., Photoswitchable catalysis mediated by dynamic aggregation of nanoparticles. *J. Am. Chem. Soc.* **2010,** *132* (32), 11018-11020.

4. Russew, M. M.; Hecht, S., Photoswitches: from molecules to materials. *Adv. Mater.* **2010,** *22* (31), 3348-3360.

5. Kumar, G. S.; Neckers, D. C., Photochemistry of azobenzene-containing polymers. *Chem. Rev.* **1989,** *89* (8), 1915-1925.

6. Conti, I.; Garavelli, M.; Orlandi, G., The different photoisomerization efficiency of azobenzene in the lowest n pi* and,pi pi* singlets: The role of a phantom state. *J. Am. Chem. Soc.* **2008,** *130* (15), 5216-5230.

7. Rau, H.; Luddecke, E., On the rotation-inversion controversy on photo-isomerization of azobenzenes - experimental proof of inversion. *J. Am. Chem. Soc.* **1982,** *104* (6), 1616-1620.

8. Stoll, R. S.; Peters, M. V.; Kuhn, A.; Heiles, S.; Goddard, R.; Buhl, M.; Thiele, C. M.; Hecht, S., Photoswitchable catalysts: correlating structure and conformational dynamics with reactivity by a combined experimental and computational approach. *J. Am. Chem. Soc.* **2009,** *131* (1), 357-367.

9. Haberhauer, G.; Kallweit, C., A bridged azobenzene derivative as a reversible, light-induced chirality switch. *Angew. Chem. Int. Ed.* **2010,** *49* (13), 2418-2421.

10. Clever, G. H.; Tashiro, S.; Shionoya, M., Light-triggered crystallization of a molecular host-guest complex. *J. Am. Chem. Soc.* **2010,** *132* (29), 9973-9975.

11. Westmark, P. R.; Kelly, J. P.; Smith, B. D., Photoregulation of enzyme-activity - photochromic, transition-state-analog inhibitors of cysteine and serine proteases. *J. Am. Chem. Soc.* **1993,** *115* (9), 3416-3419.

12. Ruck-Braun, K.; Kempa, S.; Priewisch, B.; Richter, A.; Seedorff, S.; Wallach, L., Azobenzene-Based omega-Amino Acids and Related Building Blocks: Synthesis, Properties, and Application in Peptide Chemistry. *Synthesis-Stuttgart* **2009,** (24), 4256-4267.

13. Hayashi, G.; Hagihara, M.; Dohno, C.; Nakatani, K., Photoregulation of a peptide-RNA interaction on a gold surface. *J. Am. Chem. Soc.* **2007,** *129* (28), 8678.

14. Matsunaga, D.; Asanuma, H.; Komiyama, M., Photoregulation of RNA digestion by RNase H with azobenzene-tethered DNA. *J. Am. Chem. Soc.* **2004,** *126* (37), 11452-11453.

15. Auernheimer, J.; Dahmen, C.; Hersel, U.; Bausch, A.; Kessler, H., Photoswitched cell adhesion on surfaces with RGD peptides. *J. Am. Chem. Soc.* **2005,** *127* (46), 16107-16110.

16. Volgraf, M.; Gorostiza, P.; Numano, R.; Kramer, R. H.; Isacoff, E. Y.; Trauner, D., Allosteric control of an ionotropic glutamate receptor with an optical switch. *Nat. Chem. Biol.* **2006,** *2* (1), 47-52.

17. Harada, M.; Sisido, M.; Hirose, J.; Nakanishi, M., Photoreversible antigen-antibody reactions. *Febs Lett.* **1991,** *286* (1-2), 6-8.

18. Harada, M.; Sisido, M.; Hirose, J.; Nakanishi, M., Photocontrolled Uptake and Release of Photochromic Haptens by Monoclonal-Antibodies - Evidence of Photoisomerization inside the Hapten-Binding Site. *B Chem Soc Jpn* **1994,** *67* (5), 1380-1385.

19. Pearson, D.; Downard, A. J.; Muscroft-Taylor, A.; Abell, A. D., Reversible photoregulation of binding of alpha-chymotrypsin to a gold surface. *J. Am. Chem. Soc.* **2007,** *129* (48), 14862.

20. Pearson, D.; Abell, A. D., Structural optimization of photoswitch ligands for surface attachment of alpha-chymotrypsin and regulation of its surface binding. *Chem-Eur. J.* **2010,** *16* (23), 6983-6992.

21. Bratkovic, T., Progress in phage display: evolution of the technique and its applications. *Cell. Mol. Life Sci.* **2010,** *67* (5), 749-767.

22. (a) Gai, S. A.; Wittrup, K. D., Yeast surface display for protein engineering and characterization. *Curr. Opin. Struc. Biol.* **2007,** *17* (4), 467-473; (b) Boder, E. T.; Wittrup, K. D., Yeast surface display for directed evolution of protein expression, affinity, and stability. *Method Enzymol.* **2000,** *328*, 430-444.

23. (a) Lipovsek, D.; Pluckthun, A., In-vitro protein evolution by ribosome display and mRNA display. *J. Immunol. Methods* **2004,** *290* (1-2), 51-67; (b) Schaffitzel, C.; Hanes, J.; Jermutus, L.; Pluckthun, A., Ribosome display: an in vitro method for selection and evolution of antibodies from libraries. *J. Immunol. Methods* **1999,** *231* (1-2), 119-135.

24. Hoogenboom, H. R., Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* **2005,** *23* (9), 1105-1116.

25. (a) Molek, P.; Strukelj, B.; Bratkovic, T., Peptide Phage Display as a Tool for Drug Discovery: Targeting Membrane Receptors. *Molecules* **2011,** *16* (1), 857-887; (b) Pande, J.; Szewczyk, M. M.; Grover, A. K., Phage display: Concept, innovations, applications and future. *Biotechnol. Adv.* **2010,** *28* (6), 849-858.

26. Kim, H. J.; Eichinger, A.; Skerra, A., High-affinity recognition of lanthanide(iii) chelate complexes by a reprogrammed human lipocalin 2. *J. Am. Chem. Soc.* **2009,** *131* (10), 3565-3576.

27. Weber, P. C.; Pantoliano, M. W.; Simons, D. M.; Salemme, F. R., Structure-based design of synthetic azobenzene ligands for streptavidin. *J. Am. Chem. Soc.* **1994,** *116* (7), 2717-2724.

28. Farrera, J. A.; Canal, I.; Hidalgo-Fernandez, P.; Perez-Garcia, M. L.; Huertas, O.; Luque, E. J., Towards a tunable tautomeric switch in azobenzene biomimetics: Implications for the binding affinity of 2-(4 '-hydroxyphenylazo)benzoic acid to streptavidin. *Chem. Eur. J.* **2008,** *14* (7), 2277-2285.

29. Qureshi, M. H.; Yeung, J. C.; Wu, S. C.; Wong, S. L., Development and characterization of a series of soluble tetrameric and monomeric streptavidin muteins with differential biotin binding affinities. *J. Biol. Chem.* **2001,** *276* (49), 46422-46428.

30. Lee, K. S. S. Michigan State University, East Lansing, 2011.

31. Barbas, C. F., *Phage display : a laboratory manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

32. Oertel, U.; Mart, H.; Komber, H.; Bohme, F., Media effects on the optical properties of proton donor group containing azobenzene derivatives. *Opt. Mater.* **2009,** *32* (1), 54-61.

33. (a) Shine, H. J.; Zmuda, H.; Park, K. H.; Kwart, H.; Horgan, A. G.; Brechbiel, M., Benzidine Rearrangements .16. The Use of Heavy-Atom Kinetic Isotope Effects in Solving the Mechanism of the Acid-Catalyzed Rearrangement of Hydrazobenzene - the Concerted Pathway to Benzidine and the Non-Concerted Pathway to Diphenyline. *J. Am. Chem. Soc.* **1982,** *104* (9), 2501-2509; (b) Shine, H. J.; Zmuda, H.; Kwart, H.; Horgan, A. G.; Brechbiel, M., Benzidine Rearrangements .17. The Concerted Nature of the One-Proton Para-Semidine Rearrangement of 4-Methoxyhydrazobenzene. *J. Am. Chem. Soc.* **1982,** *104* (19), 5181-5184.

34. Iyi, N.; Kurashima, K.; Fujita, T., Orientation of an organic anion and second-staging structure in layered double-hydroxide intercalates. *Chem. Mater.* **2002,** *14* (2), 583-589.