FUNCTIONAL CONTROL OF SOLUBLE RHODOPSIN MIMICS USING HIGH RESOLUTION STRUCTURE-BASED DESIGN AND EVALUATION

By

Alireza Ghanbarpour

A DISSERTATION

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

Chemistry – Doctor of Philosophy

2019

ABSTRACT

FUNCTIONAL CONTROL OF SOLUBLE RHODOPSIN MIMICS USING HIGH RESOLUTION STRUCTURE-BASED DESIGN AND EVALUATION

By

Alireza Ghanbarpour

Visualizing the microenvironments of protein/small molecule interactions is the missing link in evaluating the structure-function relationship in many “redesigned” protein systems requiring small molecule binding.

The primary goal of this thesis is to manage protein/small molecule interactions to achieve new functions through rational protein engineering in a protein scaffold, which is “evolutionarily naive.” The snapshots of each engineering step are collected using high resolution protein crystallography, opening doors to the design strategies of future measures. Finally, the mechanism of the system is elucidated by connecting structural information and biochemical assays.

The protein scaffolds used in our study are hCRBPII and CRABPII, belonging to the iLBP protein family. By reengineering their binding pockets, we generated a rhodopsin mimic ligating with small molecules with aldehyde functionalities through protonated Schiff base formation. In Chapter I, we employ the aforementioned strategy to create a new model system based on reengineered CRABPII, mimicking the critical steps of microbial rhodopsin isomerization in a single crystal. Using atomic resolution crystal structures, different mechanisms of retinal/protein interactions with light are demonstrated. Specially, a new photoswitchable protein is identified that does not require chromophore isomerization or a conformational change.

In Chapter II, the effect of ligand binding on the conformational states of the domain-swapped dimer of hCRBPII is investigated. A new protein conformational switch is created through a designed disulfide bond that can be activated and adopt new conformations in response to retinal/fatty acid binding and/or reduction potential of the environment. A novel allosterically regulated zinc-binding site is engineered, whose binding affinity can be tuned by the conformational states of our protein. Additionally, using merocyanine, a synthetic fluorophore, a new "swap back" domain-swapped dimer is identified in hCRBPII at atomic resolution, leading to the largest conformational change in the protein. This demonstrates the power of our system to adopt new conformations with different small molecules. Through systematic mutational studies and high resolution crystal structures, the role of the hinge loop region in imposing new conformations/functions in the iLBP family is explored.

In Chapter III, the discovery of the domain-swapped trimer as an unprecedented fold for the iLBP family is mentioned. Through a designed disulfide bond and metal-binding site, we are able to favor trimer formation. The mechanism of each step is examined using crystal structures and binding and stability assays.

Finally, in Chapter IV, the mechanism of a new class of fluorescent protein tags using the hCRBPII rhodopsin mimic bound with synthetic fluorophores is inspected. By exploiting high resolution crystallography, the microenvironments of protein/ligand interactions is visualized in different fluorescent protein tags applications.

ACKNOWLEDGMENTS

I would like to thank my undergraduate and master's supervisors, Professor Hamid Reza Khavasi, Professor Mostafa Pouramini, from Shahid Beheshti University, and Dr. Mojtaba Bagherzadeh from Sharif University of Technology, who gave me the chance to begin my scientific journey as a researcher in their labs.

I am very thankful to my PhD supervisor, Professor James Geiger, who provided me with great knowledge and resources and helped me visualize protein structures from a unique perspective. Through our many conversations, he not only educated me on how to be a structural biologist, but also taught me what it means to approach science with humility. I am very grateful for all the support and invaluable advice I received from him in every step of my education.

As a first year PhD student, I received helpful guidance from the senior graduate students, Dr. Nosrati, Dr. Assar, and Dr. Yapici. In particular, Dr. Nosrati played an essential role in my training.

Many of my projects would have been hard to accomplish without the help of my talented collaborators, who gave me immense support from beginning to end. I would like to thank Dr. Nairat, who assisted me with laser experiments, Dr. Pinger, who assisted me with the radioactive zinc labelling assay, Dr. Santos, who generously helped me with the mutagenesis studies of many joined projects, Dr. Assar, who assisted me with the thermal stability assay, Professor Jin, who patiently provided support for the Isothermal calorimetry assay, and Mr. Adam Jenkins, who performed the time-resolved experiment on our system from Professor Larsen's lab at UC Davis.

I am very grateful to Dr. Chrysoula Vasileiou for providing invaluable ongoing guidance and feedback.

I would like to express my sincere gratitude to Dr. Borhan, who educated me in various aspects of my projects. He has had a great influence on me as a researcher by helping me to think critically about my science and develop outstanding ideas during my projects.

I would like to thank my other committee members, Professor Jian Hu, who generously gave me freedom to use the mosquito robot in his lab at any time, and Professor Daniel Jones, who kindly provided support and consultation during my PhD studies.

Lastly, I am very thankful for my family, especially my wife and my parents, who have always been supportive, loving, understanding, and encouraging during my education.

TABLE OF CONTENTS

CHAPTER IV: ELUCIDATING THE MECHANSIM OF HCRBPII-BASED FLUORESCENT PROTEIN TAGS USING HIGH-RESOLUTION CRYSTAL STRUCTURES.. 186

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

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

M	molar
μM	micromolar
nM	nanomolar
mM	millimolar
ε	extinction coefficient
h	hour
m	minute
sec	second
$t_{1/2}$	maturation half-life
φ	quantum yield
CRABPII	cellular retinoic acid binding protein II
hCRBPII	human cellular retinol binding protein II
EGFP	enhanced GFP
FABP	fatty acid binding protein
iLBP	intracellular lipid binding protein
HPLC	high pressure liquid chromatography
SB	Schiff base
PSB	protonated Schiff base
TF	transcription factor
EF	elongation factor
7TM	seven transmembrane helix
XFEL	x-ray free electron laser
TLC	thin layer chromatography

UV-vis	ultraviolet visible spectroscopy
FTIR	Fourier transform infrared spectroscopy
PCR	polymerase chain reaction
DS	domain-swapped
FRET	Förster resonance energy transfer
MBP	maltose-binding protein
BLA	β-lactamase
LOV	light-oxygen-voltage-sensing domain
BLUF	blue-light using FAD
FLARE	fast light and activity-regulated expression
TEV	tobacco etch virus
CD	circular dichroism
ITC	isothermal titration calorimetry
WT	wild type
hCRBPII-HIV	HIV-protease connected to the c-terminus of hCRBPII
DT	diphteria toxin
RNase A	ribonucleases
BS-RNase	bovine seminal ribonucleases
CV-N	Cyanovirin-N
PAGE	polyacrylamide gel electrophoresis
ESPT	excited-state proton transfer
KLK	Q108K:K40L:T51K
KLF	Q108K:K40L:T51F

A1	Q108K:R58L:T53A: K40D:Q38F:Q4F
KD	Q108K:T51D
M1	R111K:Y134F:T54V:R132Q:P39Y:R59Y
H1	Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L117D:Q38Y
AG1	Q108K:T51D:A28C:L36C:F57H
L1	Q108K:K40L:T53S:Y19W:R58W
L2	Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W
L3	Q108K:K40H:T53A:R58L:Q4F:Q38F

CHAPTER I: MIMICKING MICROBIAL RHODOPSIN ISOMERIZATION IN A SINGLE CRYSTAL

I-1. Rhodopsin and retinal isomerization

The conversion of light into chemical energy is one of the most essential processes in nature. Photoreceptor proteins are at the heart of many biological functions mandating the conversion of light energy into chemical energy. These proteins encompass all kingdoms of life; eukaryotes, bacteria, and archaea, and they are responsible for sensing, intracellular and intercellular signaling, as well as energy production in lower organisms. [1-2]

These proteins' ability to respond with light arises from a chromophore enclosed inside the binding cavity of proteins, such as bilin proteins bound to bilin, flavoproteins bound

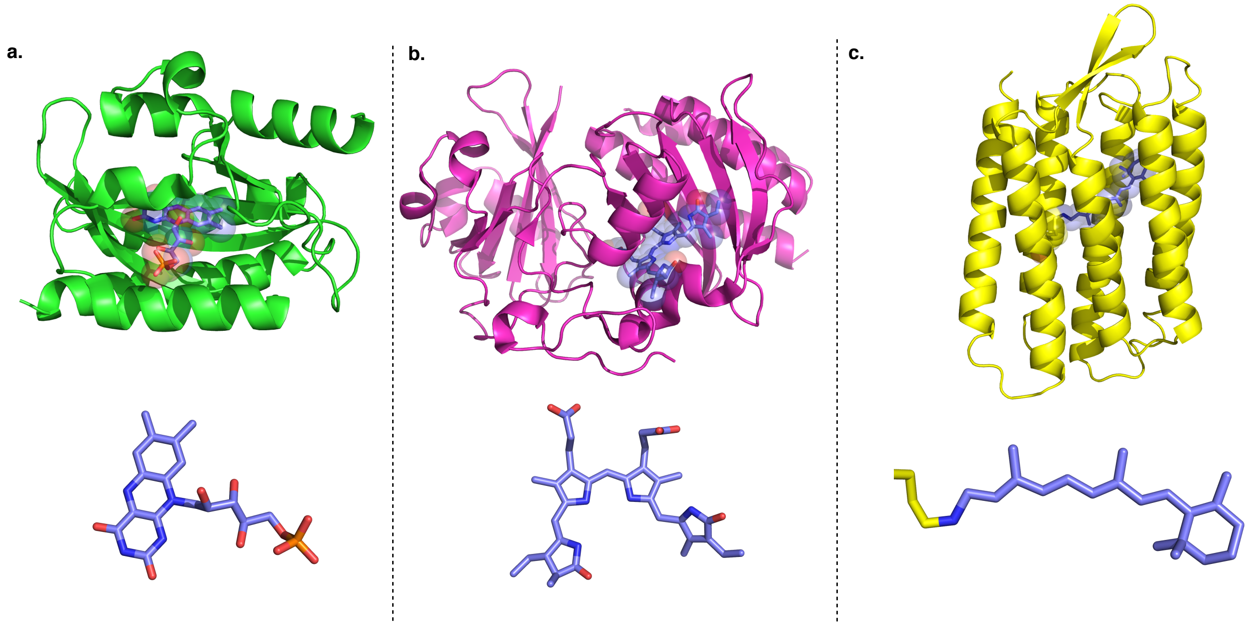


Figure I-1. a. The LOV domain bound with flavin chromophore. b. Phytochrome bound with a bilin chromophore. c. Rhodopsin that is covalently attached to the protein through Lsy216. The chromophores have been shown with stick representation (carbons in blue, nitrogen in dark blue, oxygen in red, and phosphorous in orange).

with flavin chromophore, and retinylidene proteins bound with retinal (Figure I-1). Rhodopsin is one of the earliest and most important classes of photoreceptors, employing retinal as a covalently-bound chromophore.[1-4] The rhodopsin family is a large and functionally diverse group of proteins that are found in all kingdoms of life. Their functional diversity comprises proton or ion pumps, [5] ion channels,[6] signaling proteins,[7] vision,[8] and sensing.[9] The apo-proteins, referred to as opsins, are seven transmembrane helix ("7TM") integral membrane proteins, and form an imine (Schiff Base) between retinal, or a closely-related analog, and a buried lysine residue to generate the pigment.[1,10] Vast literature that includes time-resolved spectroscopy, structural biology — most notably X-ray crystallography [11-14] — and biochemistry [15-16] describes the structures, spectroscopy and lifetimes of the various intermediates in the process and identifies the key protein residues essential for its function.[1, 17-19] There are two pivotal classes of rhodopsin proteins: microbial or type I rhodopsin, and animal, or type II rhodopsin. Structurally, these proteins share a seven transmembrane helix (7-TM) fold, but share little sequence similarity with one another. While microbial rhodopsins are employed in lower organisms for energy production and intracellular and intercellular signaling, animals employ a G-protein-coupled receptor (GPCR) family of rhodopsins mostly for signaling processes. Color vision, non-vision transduction, and maintenance of the circadian clock are some important examples of where these proteins are utilized.[1]

Animal rhodopsin mostly uses 11-*cis* retinal as a chromophore (Figure I-2). Upon absorption of a light photon, the chromophore undergoes selective isomerization to generate all-*trans* retinal, a thermodynamically more stable isomeric form. Then, at the end of the photocycle, after the protein adopts the active meta II state, which is essential for G-protein activation [20] and where the chromophore gets deprotonated, the all-*trans* retinal becomes hydrolyzed from the protein binding pocket. Through a catalytic cycle in which many proteins and enzymes cooperate together, the 11-*cis* retinal is regenerated and delivered to the opsin (apo protein) to restore the pigmented rhodopsin. [21-22] Although hydrolyzing the chromophore after isomerization is the hallmark of many rhodopsins' functions, some bistable rhodopsins have recently been identified that do not require the hydrolysis of the chromophore, and through irradiation with light, the chromophore becomes reset to the original state.[23]

However, in microbial rhodopsin, the isomerization ensues from all-*trans* to 13-*cis* retinal (Figure I-2a), in which a thermodynamically more stable isomeric form undergoes selective isomerization to produce a thermodynamically less stable isomer. Also, in contrast to most animal rhodopsins, the chromophore does not get hydrolyzed, but instead, through thermal adaptation, the chromophore and conformation of the proteins revert back to their original states.[24] A recent study showed that in the case of animal rhodopsin, the only role of protein in the isomerization of retinal is to preselect the active conformer of retinal, which is 11-*cis*-12-s *trans* versus 11-*cis* 12-s-*cis*. [25]

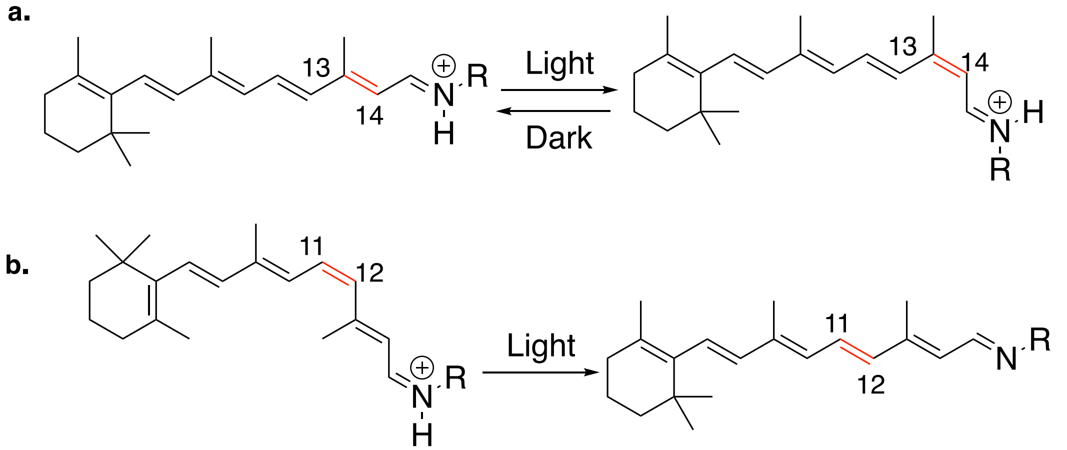
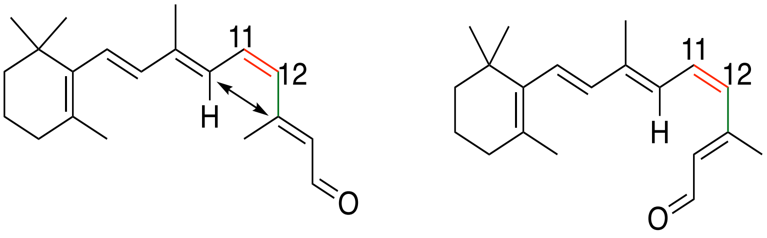


Figure I-2. Photoisomerization of the chromophore-bound protein in: a. microbial rhodopsin; b. animal rhodopsin.

According to this study, the isomerization of 11-*cis* to all-*trans* retinal happens both solution and the protein binding pocket with 100% selectivity. However, inside the proteins this isomerization ensues with a high quantum yield, since the binding pocket is well-engineered to bind with the active conformer, 11-*cis*-12-s *trans,* where the methyl group of C13 and hydrogen of C10 sterically clash with each other, causing the fast photoisomerization of retinal (Figure I-3).[25] In contrast, in microbial rhodopsin photoisomerization, the proteins need to perform three important tasks. First, they must overcome the energy barrier for isomerization. Second, they need to prevent the formation of other isomers, since there is more than one double bond that can potentially get isomerized. The irradiation study of all-*trans* retinal in solution leads to the distribution of different isomers highlighting the role of the protein in the selective isomerization of retinal.[26] And third, the proteins must reset the chromophore at the end of isomerization to the original state through a thermal process, where the energy of the absorbing photon is gone. These tasks demonstrate the nature and complexity of the biophysical processes that need to be performed in retinal isomerization by proteins in the microbial rhodopsin family.



11-*cis*-12-s-*trans* 11-*cis*-12-s-*cis*

Figure I-3. Two different possible conformations for 11-*cis* retinal in solution: 11-*cis*-12-s-*trans* and 11-*cis*-12-s-*cis*. 11-*cis*-12-*s*-*trans* conformations generate a steric between the hydrogen of C10 and methyl group of C13.

One of the long-lasting, remaining questions that has attracted many researchers in the past few decades is how microbial rhodopsin can selectively perform the isomerization of retinal around one double bond, while retinal photoisomerization in solution gives rise to the distribution of different isomers. [27-28]

To answer this challenging question, research groups have taken different approaches, such as the chemical modification of retinal [27, 29] or mutational studies of the protein.[17, 30] Although these approaches unravel different biophysical features and thermodynamic driving forces of rhodopsin isomerization, they have some inherent limitations: first, some of the key residues in the photocycle of rhodopsin are conserved residues, and mutations of them can lead to insoluble expressions of the protein; second, the chemical modification of retinal might lead to a completely different trajectory of the chromophore inside the binding cavity of the protein. Creating a new photoactive model system using a soluble protein, which is evolutionary naive, not only provides us with an easier system to study the isomerization pathway of retinal in comparison with a membrane protein, but it would also reveal a new insight into how evolution develops such photoactive proteins from scratch. In this chapter, the photocycle of microbial rhodopsin and the key residues in this event will be discussed first, and then we will describe our strategy to create a new model system that mimics all the important features of rhodopsin isomerization.

I-2. Microbial rhodopsin features

The microbial rhodopsin family encompasses more than 7,000 photoactive proteins in different organisms from different kingdoms of life.[31] This family offers an explicit example of nature's diverse evolution of a single protein scaffold, giving rise to functionally-distinct proteins ranging from ion/proton pumps, sensors, channels, signaling proteins, etc. [1-2, 8]

Although all the proteins in this family share a 7-TM helical scaffold and a conserved Lys that generates a covalent link to retinal through protonated Schiff base (PSB) formation, there are many variations in the residues, the orientation of loops, and the water molecules' positions in the protein structures. The binding pocket of retinal perhaps is the most conserved part among all the different proteins in this family.[1]

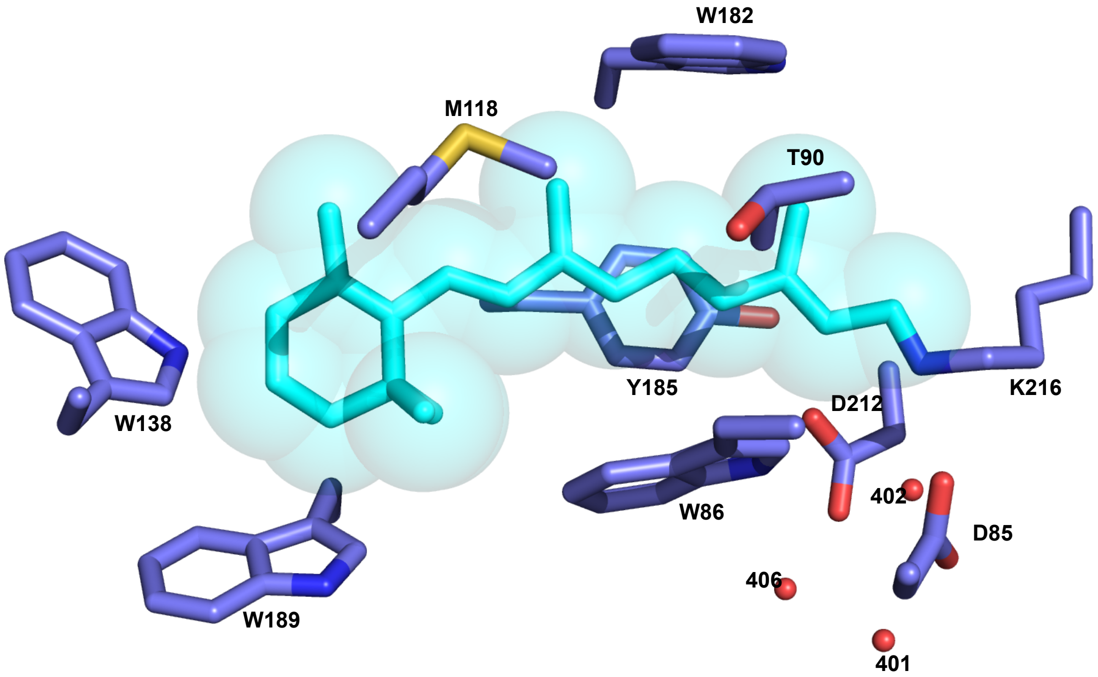


Figure I-4. Bacteriorhodopsin binding pocket contains aromatic residues such as W182, W138, W189, W86, and Y185, that sandwich the chromophore through aromatic interaction. D85 and D212 act as a counter ion for the iminium, stabilizing the positive charge. The water molecules shown in red mediate the hydrogen bond between the nitrogen of the iminium and D85 and D212. Retinal is shown in cyan, and the protein backbone carbon, nitrogen, and sulfur atoms are shown in purple-blue, dark blue, and dark yellow, respectively (PDB code:1C3W).

Bacteriorhodopsin, one of the most well-studied systems in the microbial rhodopsin family, highlights the role of aromatic residues, as shown in Figure I-4, for holding the chromophore in the specific trajectory needed for isomerization and a net conformational change in the protein that gives rise to its functional regulation. The presence of these aromatic residues is diffuse in the family of rhodopsin proteins.[1]

For instance, many spectroscopic studies, along with recent time-resolved X-ray Free Electron Laser (XFEL) experiments,[32] revealed that a conserved Trp182 acts as a mechanical transformer, leading to a net conformational change in the protein upon isomerization of retinal by tilting the α-helices in response to retinal photoisomerization. Other studies speculated the importance of Trp83 in directing the photoisomerization towards the C13-C14 double bond by providing a tuned steric interaction in the vicinity of the chromophore.[17,30] Furthermore, another conserved residue, Tyr185, forms a hydrogen bond with the nitrogen of iminium, which is important for stabilizing the PSB and holding the trajectory of the chromophore during the photocycle. Interestingly, this residue is conserved in animal rhodopsin as well.[33]

Since the chromophore is protonated inside the binding cavity, there are some acidic residues, such as Asp85 and Asp212, inside the binding cavity that act as counter ions to stabilize the positive charge on the nitrogen of the iminium (Figure I-4). Notably, these residues are conserved between all the proton pumps in this family. Additionally, mutational studies of these residues, specially Asp212, showed they are important not only for proton pumping activity, but also for preventing other isomerizations.[5, 34]

In addition to these residues, the ordered water molecules in the binding pocket, mostly residing between acidic residues and the chromophore, are crucial for adjusting the pK_a of the chromophore and surrounding residues and dictating the directionality of the ion/proton channels from the cytoplasmic side to the extracellular side and vice versa. [35] A very recent time-resolved x-ray experiment on bacteriorhodopsin demonstrated that the water molecules between the chromophore and acidic residues become disordered in less than 500 fs, prior to retinal isomerization, where they can provide space for the isomerization of the imine bond and C13-C14 double bond.[36]

I-2-1. Bacteriorhodopsin photocycle

The photocycle of the microbial rhodopsin family has been extensively studied using many different spectroscopic and biophysical techniques.[6, 24, 26, 37-38] In all the proteins in the microbial rhodopsin family, the photocycle begins with selective isomerization of the chromophore, and after the formation of several intermediates and conformational changes of the protein, the structure of the chromophore resets to the original state. The bacteriorhodopsin photocycle is one of the most well-studied

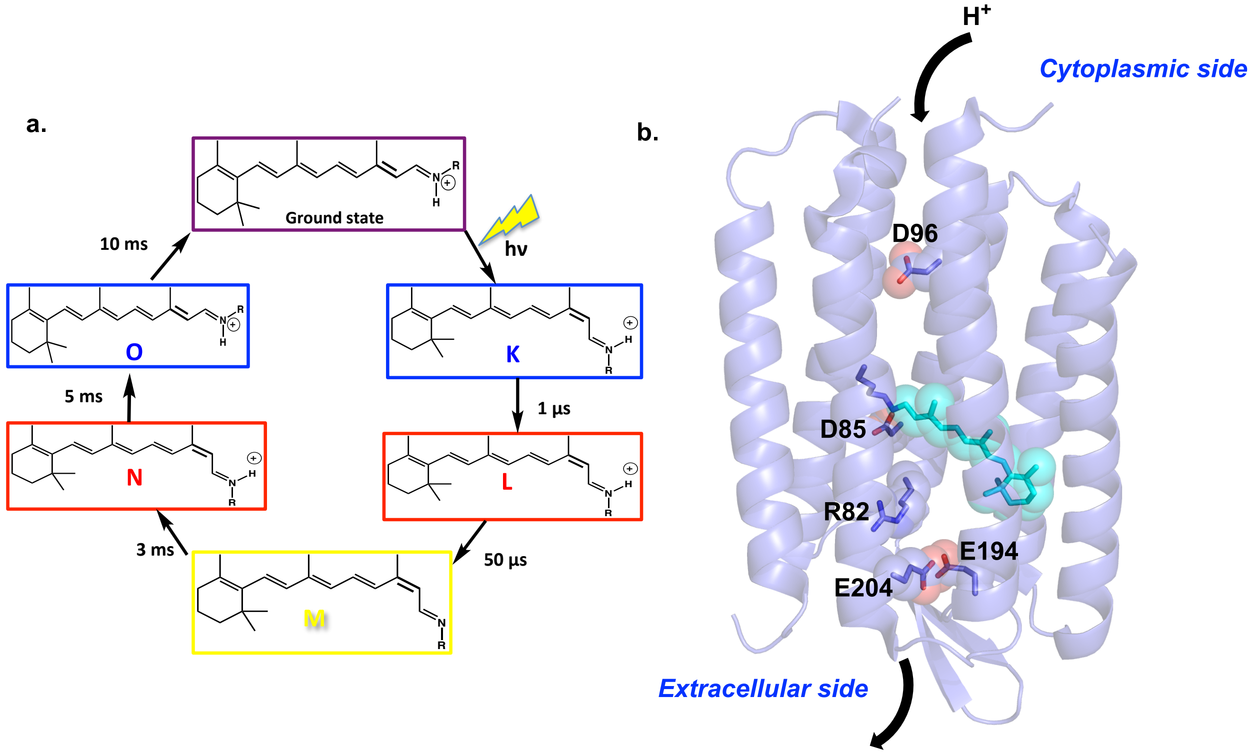


Figure I-5. a. Photocycle of bacteriorhodopsin, where different intermediates form during the photocycle. The lifetime of each is shown underneath the retinal structure. b. The residues that are important in proton-pumping activity, protonation and deprotonation of iminium, are shown as transparent purple spheres. Retinal is shown as cyan spheres. R82, E194, and E204 are referred to the proton releasing complex, facilitating the proton transfer to the bulk.

photocycles in this protein family. This protein acts as a unidirectional proton pump, in which it translocates the proton from the cytoplasmic side to the extracellular side. This activity, regulated by light, is crucial in halobacteria cells, where it leads to large gradient potential between the cells and the environment, resulting in energy production within the cells.[39] However, this process, like other photo-induced systems,[3] passes through a series of short-lived intermediate events, as shown in Figure I-5a, where the lifetime of each event has been tuned through many years of evolution to create one of the most efficient photosystems in nature.[40]

The photocycle of bacteriorhodopsin begins with the photoisomerization of the chromophore to generate 13-*cis*-15-*anti* from all-*trans* retinal less than 500 fs. The K intermediate that is formed initially (13-*cis* 15-anti) is red-shifted (590 nm vs. 570 nm, which is an original absorption of the retinal/protein complex) where the chromophore conformation is highly twisted. This intermediate possesses extra free energy for the net photoinduced proton pump activity in the protein.[41] After about a microsecond, the K intermediate relaxes to the blue-shifted L intermediate (550 nm), where the chromophore is planar and creates a hydrogen bond network with the neighboring residues. The K-to-L transitions are escorted by reorientations of the Schiff base pointing toward the intracellular side, producing a local conformational change in the protein.[38] Remarkably, the L intermediate is the incipient platform for proton-pumping activity, where the proton of the iminium switches to the primary acceptor group (D85) resulting in the formation of the significantly blue-shifted M intermediate (412 nm), which is not absorbing at the wavelength in which the irradiation occurs.[38] At this step, as a result of a change in the pK_a of the chromophore and subsequently Asp85, the Arg82 that originally maintains a salt bridge with Asp 85, swings its conformation away from the cytoplasmic site and toward the proton releasing complex (E194 and E204) in the extracellular site.[1, 42] The reconfiguration of Arg82 during the photocycle has been detected using recent XFEL time-resolved crystallography as well.[12] This local conformational change induces the proton release through the proton-releasing complex by altering the pK_a of E204 and E194 that now are deprotonated. After this step, the N intermediate (520 nm) is generated where the largest conformational change in the protein backbone ensues. This conformational change (tilting the α-helices) situates the chromophore in the cytoplasmic side, where the chromophore becomes reprotonated again through D96, Figure I-5b.[43] During the N-to-O transition (PSB λ_{max} = 640 nm), the chromophore reverts back to the original state and the D96 positioned in the cytoplasmic site becomes reprotonated again.[44] Finally, in the last step, the D85 delivers its proton to the proton-releasing complex and the protein conformation relapses back to the original state.[1] The mechanism of the unidirectionality of the proton pump in bacteriorhodopsin, which makes the ion/proton channel different from ion/proton pumps, has been controversial. However, one of the most popular mechanisms proposed is that through the photocycle, the proton accessibility of the imine varies between the extracellular site and cytoplasmic site that prevents the proton backflow.[45] Furthermore, the crystal structure of bacteriorhodopsin, along with time-resolved FTIR, depicts seven ordered water molecules in the extracellular half while only two molecules have been identified in the cytoplasmic side.[16, 46] This disconnection in the water-mediated hydrogen bond networks between the cytoplasmic and extracellular sites makes the dark inactive state and demands the light-induced conformational change to occur for a proton uptake from the cytoplasmic side and a fast proton release to the extracellular side.[1, 16]

It is important to mention that a wealth of cryo-kinetic data detected the structure of different intermediates formed in the photocycle by prolonging the lifetime of these intermediates through certain mutations, irradiation with a specific wavelength of light, or a change in pH. [38, 41, 43, 47] Additionally, recent X-ray free electron laser (XFEL) studies detected different stages of the photocycle with femtosecond time resolution. [12, 36]

I-3. Engineering hCRABPII to mimic microbial rhodopsin isomerization [48]

I-3-1. Choosing an initial protein template

We have had a long-standing interest in understanding the protein-chromophore interactions giving rise to wavelength regulation in rhodopsins,[49-59] using protein mimics to investigate the latter phenomena. We have now turned our attention to understanding the factors that govern the isomerization events by investigating the minimal requirements for bacteriorhodopsin in a model system to mimic the hallmarks of the natural system. Human Cellular Retinol Binding Protein II (hCRBPII) and Human Cellular Retinoic Acid Binding Protein II (hCRABPII), two small, soluble cytosolic proteins from the intracellular lipid binding protein family, were used as templates for protein reengineering.[49-59] To initiate our investigation, we required a platform that would support photoinduced isomerization of the bound chromophore. As described in previous studies by Nosrati *et al*, a number of hCRABPII variants exhibited photoinduced changes in the protonation state of a retinylidene-bound protein; namely, an initially-observed pigmented protein-chromophore complex reverts to a colorless state in time.[49] X-ray crystallography and UV-vis spectroscopy showed that imine bond isomerization from 15-*syn* to 15-*anti* is

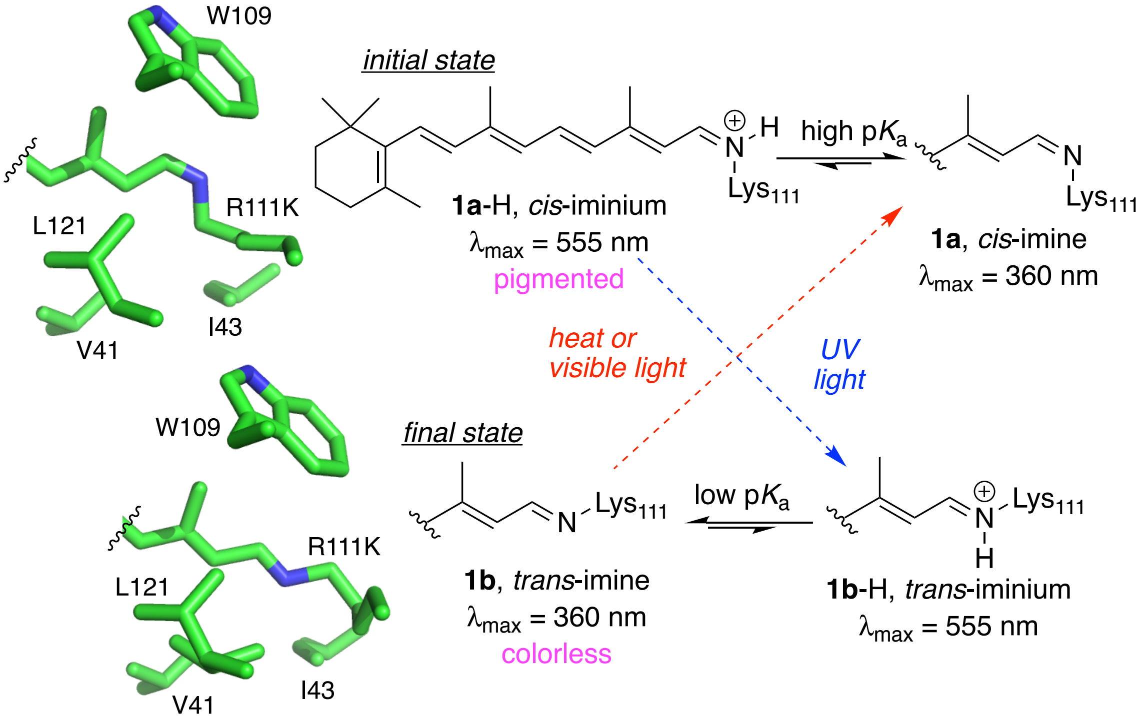


Figure I-6. Photocycle of hCRABPII hexamutant **M1** (R111K:R134F:T54V:R132Q:P39Y:R59Y) bound to retinal. The photoinduced isomerization of the imine functionality leads to changes in its pK_a and consequently its protonation state. The hydrophilic and hydrophobic environment of the imine nitrogen atom for the *cis* and *trans* geometry, respectively, is highlighted in the figure obtained from the crystal structure of each form (PDB IDs: 4YFP and 4YFQ).

responsible for this change. The initial kinetically-formed, 15-*cis* isomer orients the imine nitrogen in a polar environment that supports a high pK_a regime, and thus yields a protonated Schiff base (PSB) that is a colored pigment. Time-dependent 15-*cis* to 15-*trans* imine isomerization yields a thermodynamic product that places the nitrogen atom in a low pK_a regime, leading to the unprotonated Schiff base (SB) that is colorless (Figure I-6) (Figures I-6 to I-22 in this chapter along with the context were reproduced by the permission of Journal of American Chemical Society)[48]. Additionally, green and UV light irradiation interconverts between the PSB and SB, respectively, in solution and crystalline states.[48-49] We previously demonstrated that the pK_a difference between the two isomers is the driving force for 15-*syn* (high pK_a) to 15-*anti* (low pK_a) isomerization, where the imine nitrogen experiences two different environments in each isomeric form (Figure **I**-6).[49] Having shown the ability to photochemically isomerize the imine bond, the next goal was to design a protein capable of isomerizing the C13 double bond, reminiscent of the action of microbial rhodopsin. In general terms, this would be achieved by mimicking the placement of critical residues in bacteriorhodopsin — one of the most well-studied systems in the microbial rhodopsin family — in the binding pocket of our surrogate protein. Having high-resolution structures on hand, along with the high structural tolerance of these proteins to mutation, provides us the opportunity to explore their capabilities as an isomerase. The very first goal was to obtain a high resolution structure of protein-bound 13-*cis* retinal in our system. However, attempts to obtain the crystal structure of 13-*cis* bound protein using different variants of hCRABPII and hCRBPII were not fruitful (As shown in Table **I**-1). Finally the crystal structure of **M1** incubated with 13-*cis* showed an electron density in the binding cavity that was more similar to the all-*trans* bound form (Figure **I**-7a). The UV-vis spectra of **M1** with all-*trans* retinal and 13-*cis* revealed that **M1** gives little PSB with 13-*cis* at physiological pH, in contrast to all-*trans* (Figure **I**-7). However, the fine structure of the retinal SB in the UV-vis spectrum (at ~369 nm) proved that 13-*cis* forms a robust SB within the protein.[48]

Consistent with this result, the HPLC extraction of **M1** after incubation with 13-*cis* in the dark (~24 hours) revealed that almost half of the 13-*cis* was converted to all-*trans* at room temperature, (Figure **I**-7b), indicating slow thermal isomerization of 13-*cis* to all-

trans in the **M1** binding pocket. Additionally, light irradiation of the chromophore led to the

Table I-1. The initial efforts at obtaining a 13-*cis* bound crystal structure (Other than the first four variants of hCRBPII, the rest of the mutants belong to CRABPII).

Variants	Consequence
Q108K:K40L-(hCRBPII)	No crystal
Q108K:K40L-T5V-(hCRBPII)	No crystal
Q108K:K40L-T5V:R58F-(hCRBPII)	No crystal
Q108K:K40L-T5V:R58F:L117E-(hCRBPII)	No crystal
R111K:Y134F:T54V:R132Q:P39Q:R59Y:A32W:F3Q	No electron density for retinal
R111K:Y134F:T54V:R132Q:P39Q:R59Y:A32W:F3Q:L121W	No electron density for retinal
R111K:Y134F:T54V:R132Q:P39Q:R59Y:A32W:F3Q:L121H	No electron density for retinal
R111K:Y134F:T54V:R132Q:P39Y:R59Y:L121M	No crystal
R111K:Y134F:T54V:R132Q:P39Y:R59Y:L121Y	No crystal
R111K:Y134F:T54V:R132Q:P39Y:R59Y:L121W	No crystal
R111K:Y134F:T54V:R132Q:P39Y:R59Y:L121Q	No crystal
R111K:Y134F:T54V:R132Q:P39Y:R59Y:L121E	No crystal
R111K:Y134F:T54V:R132Q:P39Y:R59Y:L121D	No crystal
R111K:Y134F:T54V:R132Q:P39Y:R59Y:L121N	No crystal
R111K:Y134F:T54V:R132Q:P39N:R59Y:L121E	No crystal
R111K:Y134F:T54V:R132Q:P39E:R59Y:L121E	No crystal
R111K:Y134F:T54V:R132Q:P39F:R59Y:L121E	No crystal
R111K:Y134F:T54V:R132Q:P39Y:R59Y:L121E:A32W	No crystal
R111K:Y134F:T54V:R132Q:P39F:R59Y:L121E:I31W	No crystal
R111K:Y134F:T54V:R132Q:P39N:R59Y:M121E	No crystal
R111K:Y134F:T54V:R132Q:P39F:R59Y	electron density was similar to all-*trans*

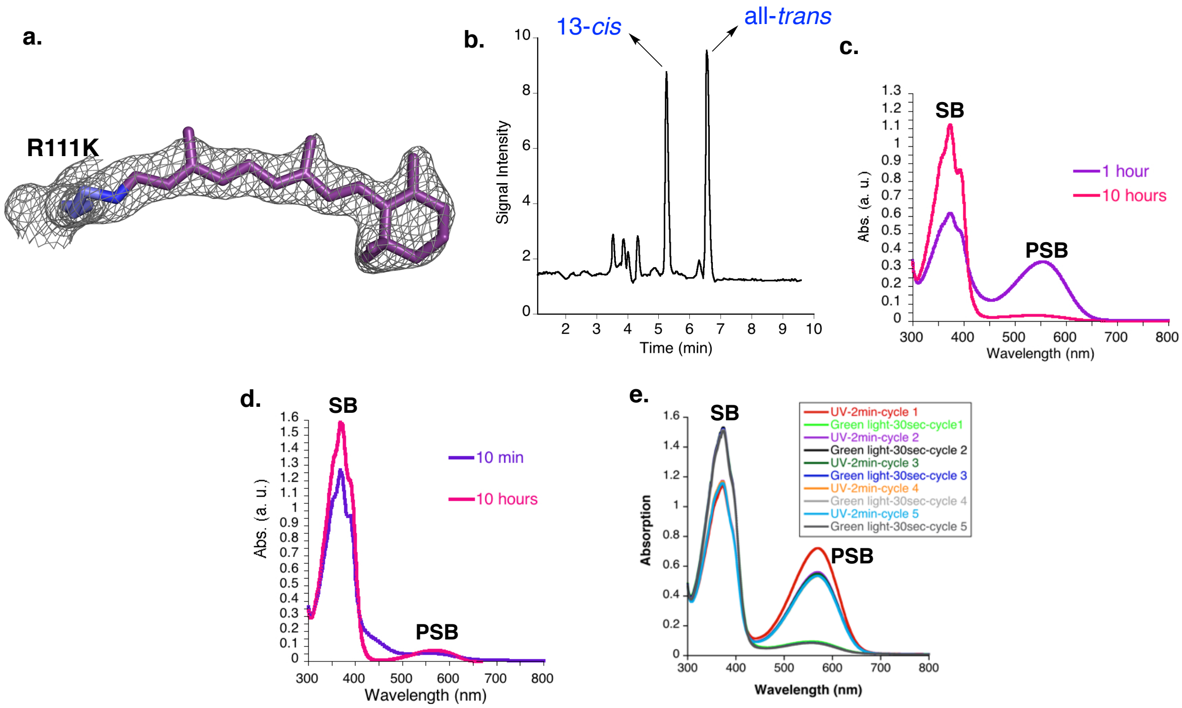


Figure I-7. a. Electron density of **M1** bound with 13-*cis* retinal showed electron density that is more similar to all-*trans* retinal. b. HPLC chromatogram of retinals extracted from 13-*cis* bound **M1** after 24 h of incubation in the dark demonstrated the dark isomerization of 13-*cis* to all-*trans* retinal. c. UV-vis spectrum of **M1**-L121E bound with all-*trans* retinal. d. UV-vis spectrum of **M1**-L121E bound with 13-*cis* retinal. e. Photo-irradiation of 13-*cis* retinal with UV (B.P. filter, 300-400 nm), green light (L.P. filter > 500 nm).

same imine bond isomerization (Figure I-7e).

Since we were able to show isomerization of the imine bond in crystals of R111K:Y134F:T54V:R132Q:P39Y:R59Y (**M1**),[49] we decided to employ this construct for further design.

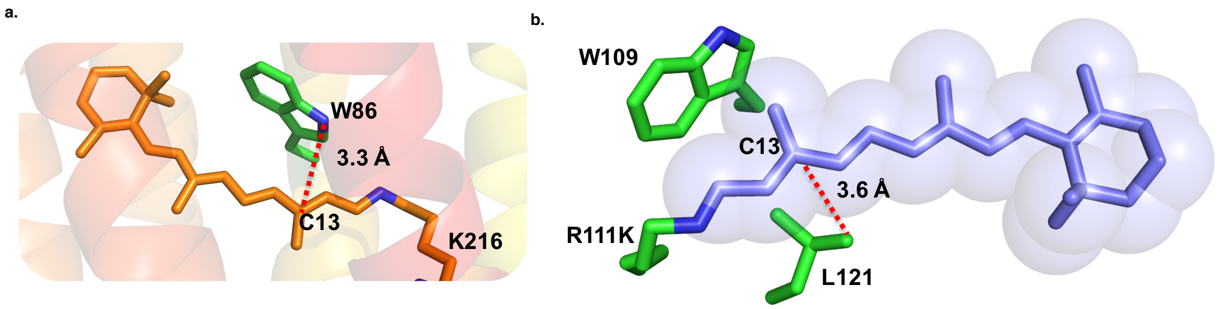


Figure I-8. a. Binding pocket of bacteriorhodopsin, where W86 is located in the vicinity of C13-C14. b. hCRABPII binding pocket. The L121 position is analogous to the W86 position in bacteriorhodopsin relative to retinal.

To mimic the microbial rhodopsin isomerization, it was first necessary to prevent the imine isomerization pathway described above. This was essential to keeping the imine protonated in all isomeric forms, and thus maintaining its bathochromicity for absorption of visible light to initiate isomerization of the C13 double bond. We feared that the imine bond isomerization could potentially lead to a conformation resulting in a low pK_a, and therefore, an unprotonated imine, which would siphon away a portion of the pigment without C13 isomerization. In essence, the unprotonated retinylidene would be an "off-

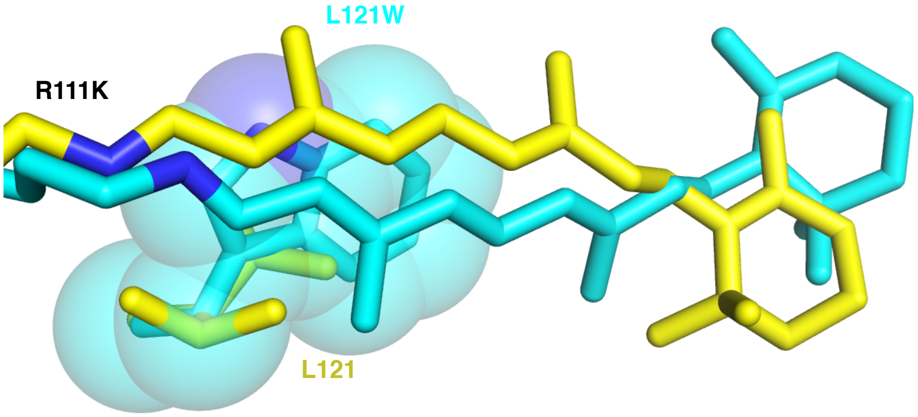


Figure I-9. Overlay of **M1** (yellow) and **M1**-L121W (cyan). Introducing Trp in the L121 position impinges on the bound retinal, which leads to the concomitant change in retinal trajectory.

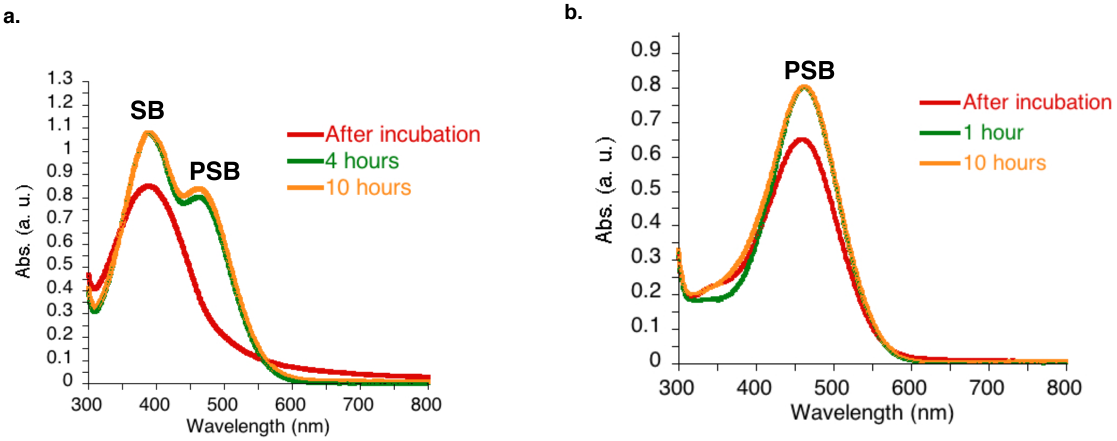


Figure I-10. UV-vis spectra of **M1** and **M1**-L121E upon incubation with: a. all-*trans* and b.13-*cis* retinal. The SB (Schiff Base, Imine) and PSB (Protonated Schiff Base, Iminium) peaks are highlighted.

cycle" well that removes the isomerizing species from the photocycle due to the lack of absorptivity at the wavelength of irradiation.

Furthermore, to perform a specific C13 isomerization, we needed to tune the steric interaction in the vicinity of the C13-C14 double bond to make this double bond distinguishable for the system in comparison with other double bonds. Using the all-*trans* retinal-bound structures as a guide, bulky residues were introduced in the vicinity of the C13-C14 double bond to introduce steric interactions. Inspired by bacteriorhodopsin, our first screening effort was centered on mutating L121 to aromatic residues to mimic the steric interactions of the natural protein's Trp 86.[1, 60] Leu121 is ~ 3.6 Å away from C13, making it a suitable candidate to fulfill this goal (Figure **I**-8). Therefore, L121W, L121Y, and L121F were incorporated in the **M1** construct. **M1**-L121F did not lead to soluble expression, but **M1**-L121W and **M1**-L121Y were solubly expressed and purified.[48]

The all-*trans* retinal-bound crystal structures of **M1**-L121W and L121Y were determined, but attempts at crystallizing the 13-*cis* bound variants were unsuccessful (Table **I**-1). The overlay of all-*trans* retinal-bound **M1**-L121W and **M1** showed that the chromophore had rotated about its C14-C15 and C6-C7 torsion angles such that the polyene methyl groups point in the opposite direction from those of all-*trans* retinal-bound **M1**, with a concomitant translation in the direction approximately perpendicular to the polyene axis (Figure **I**-9). **M1**-L121Y revealed some changes in retinal trajectory along the polyene but poor electron density for the β-ionone ring, indicating some disorder of the retinal inside the binding pocket. The failure in crystallizing 13-*cis*-bound proteins and the all-*trans* retinal chromophore disorder in these variants leads to the supposition that aromatic residues at position 121 may be too large to comfortably accommodate retinal. We envisioned the

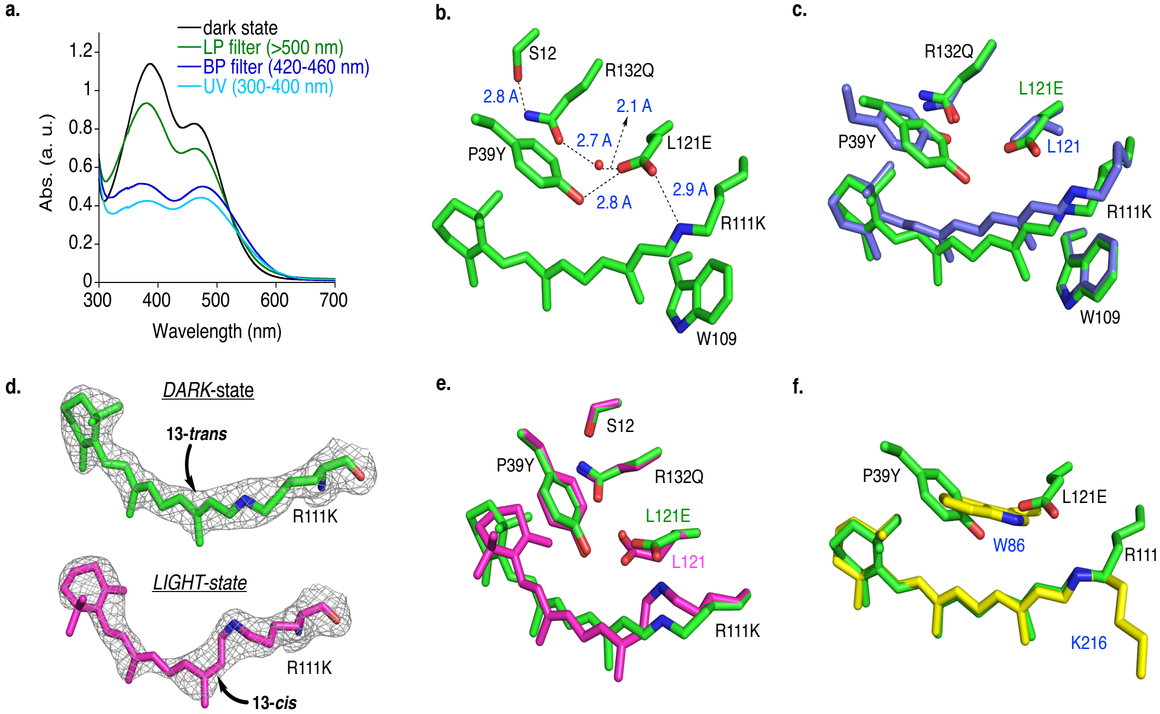


Figure I-11. a. UV-vis absorption spectrum of all-*trans*-retinal bound **M1**-L121E in the dark state (black) and after green light irradiation (long pass filter > 500 nm) (green), blue-light irradiation using band pass filter (440 nm ± 20 nm) (blue), and UV irradiation (UV band pass filter (300-400 nm)) (cyan). b. Hydrogen bonding network between the imine hydrogen of R111K and L121E, P39Y, H_2O-303, R132Q, and S12. c. Overlay of **M1** (carbon atoms in cyan) with all-*trans*-retinal bound **M1**-L121E (carbon atoms in green). d. Electron density (contoured at 1σ) of all-*trans*-retinylidene in **M1**-L121E in the dark state (green carbons) vs. 13-*cis*-*15*-*syn*-retinal imine (magenta carbons) generated after five-minute laser irradiation at ~400 nm. e. Overlay of all-*trans*-retinylidine-bound **M1**-L121E in the dark (green) with the 13-*cis*-15-*syn* imine (magenta) generated after five-minute laser irradiation at ~399 nm, showing the movement of Lys111, Glu121, and the rotation of the β-ionone ring in 13-*cis* upon isomerization. f. Overlay of all-*trans*-retinal bound **M1**-L121E imine (green) and all-*trans*-retinal bound bacteriorhodopsin (yellow carbons, PDB code: 1C3W), illustrating how the P39Y-L121E interaction mimics Trp86 in bacteriorhodopsin relative to retinal. All heteroatoms are colored by type, with N in blue and O in red.

placement of acidic residues could stabilize the protonated species to an extent that would prevent imine bond isomerization. The proximity of Leu121 to the SB nitrogen atom led to the introduction of the L121E mutation in **M1** yielding the R111K:Y134F:T54V:R132Q:P39Y:R59Y:L121E (**M1**-L121E) variant, Figure I-10, which successfully abolished the imine isomerization pathway. As shown in Figure I-11a, irradiation with either visible or UV light no longer switches the absorption between visible light-absorbing (high pK_a) and UV light-absorbing (low pK_a) forms, but instead leads to a decrease in the absorbance of both forms independent of the wavelength of light, presumably due to changes in the structure of the chromophore (*vida infra*).

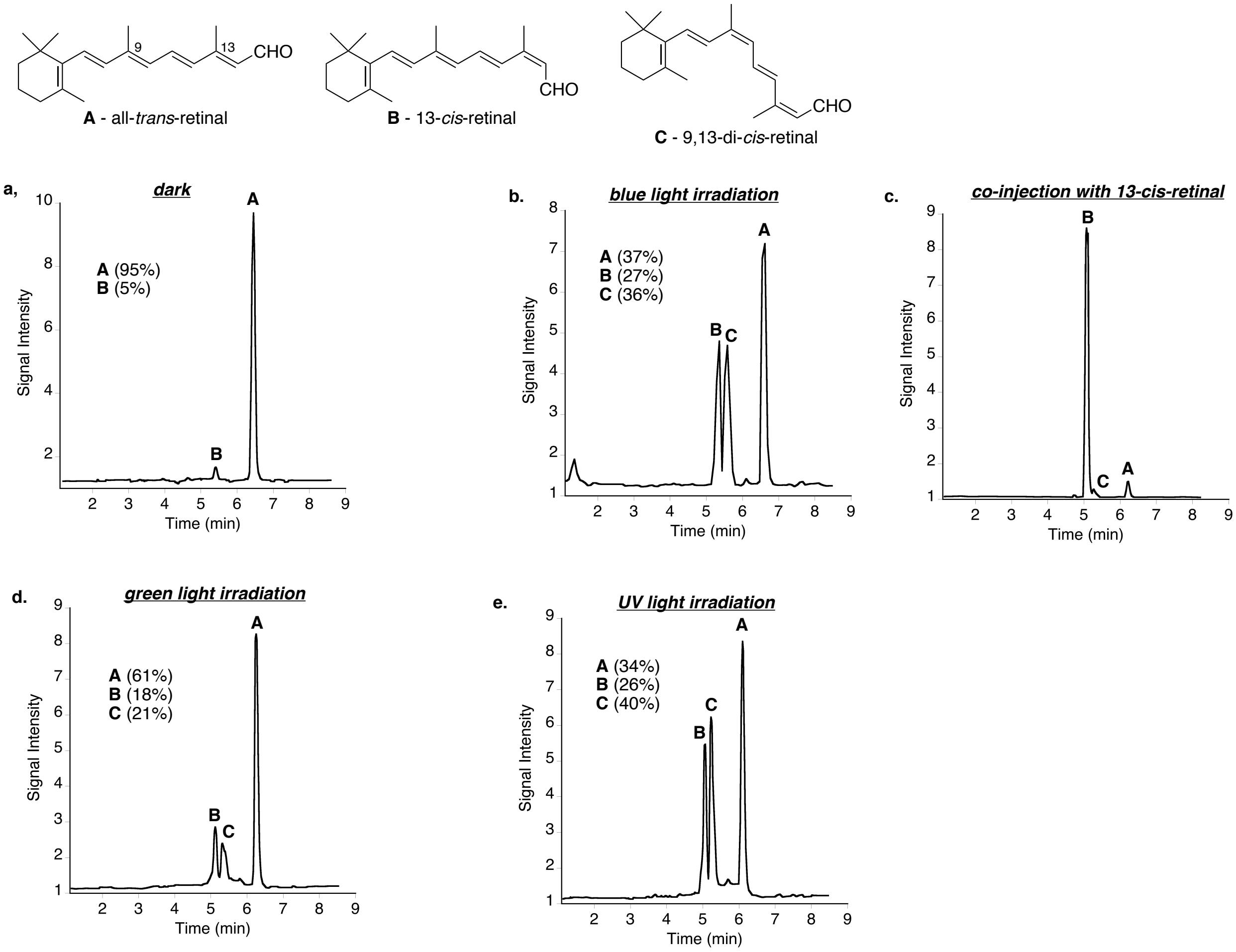


Figure I-12. HPLC chromatogram of retinal isomers extracted from all-*trans* retinal bound **M1**-L121E (recorded at 363 nm). a. Dark, after irradiation with mercury lamp- green light, blue light, UV, and coinjection with the pure 13-*cis* retinal (red).The ratios were corrected using the extinction coefficient of each isomer (all-*trans* 48,000 $M^{-1}cm^{-1}$, 13-*cis* 38,770 $M^{-1}cm^{-1}$, and 9,13-*dicis* retinal 32,380 $M^{-1}cm^{-1}$).

The structure of all-*trans* retinal-bound **M1**-L121E (Figure **I**-11b) was successfully determined and revealed a *trans*-imine, with the newly-installed Glu121 engaged as a salt bridge with the iminium (Figure I-11b). The overlay of **M1** and **M1**-L121E also reveals a serendipitous structural rearrangement (Figure **I**-11c). Tyr39, which hydrogen-bonds with Gln132 in **M1**, adopts a new conformation, presumably as a result of a stronger hydrogen bond with Glu121 (Figure **I**-11b). In this conformation, Tyr39 sterically impinges on the bound retinal, causing a translation of the polyene in the vertical axis, sandwiching the chromophore between Trp109, located on the opposite side of the chromophore, and Tyr39, while tilting the chromophore relative to the trajectory in **M1** (Figure I-11c).

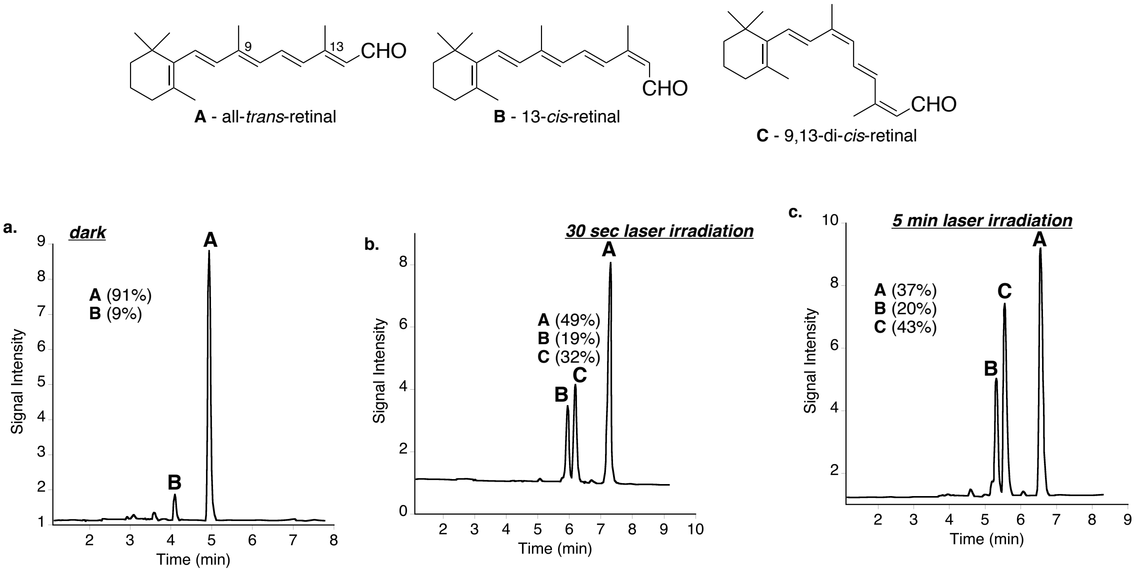


Figure I-13. HPLC chromatogram of **M1**-L121E bound with all-*trans* retinal before irradiation, after 30 sec laser irradiation, and after 5 min laser irradiation (recorded at 363 nm). The ratios were corrected using the extinction coefficient of each isomer. The ratios were corrected using the extinction coefficient of each isomer (all-*trans* 48,000 $M^{-1}cm^{-1}$, 13-*cis* 38,770 $M^{-1}cm^{-1}$, and 9,13-*dicis* retinal 32,380 $M^{-1}cm^{-1}$).

Figure I-11c shows that these changes occur in the polyene part of the chromophore, leaving the β-ionone ring in approximately the same position seen in **M1**. The direct salt bridge between Glu121 and the nitrogen atom of the *trans*-iminium gives a substantially blue-shifted spectrum (468 nm, compared to 568 nm of the *cis*-iminium of **M1**),[49] by localizing the positive charge on the polyene (Figures I-7c and I-10a).[59] Glu121 seems to not only shut off the imine bond isomerization pathway by substantially increasing the pK_a of the *trans*-iminium, but it also provides a steric interaction that orders the chromophore in the binding pocket by sandwiching the chromophore between Tyr39 and Trp109. Unfortunately, crystals of 13-*cis*-bound **M1**-L121E in the dark could not be obtained. Intrigued by the large drop in the absorption of both SB and PSB upon irradiation with light of any wavelength, we extracted the chromophore from the **M1**-L121E samples and analyzed them by HPLC (Figures I-12 and I-13). Two peaks are apparent in the HPLC traces for the light-irradiated **M1**-L121E samples: one that co-elutes with authentic 13-*cis* retinal, and another that was assigned as 9,13-di-*cis*-retinal, based on a ^{1}H-NMR aldehyde chemical shift comparison with the reported isomer (Figure I-14).[48, 61]

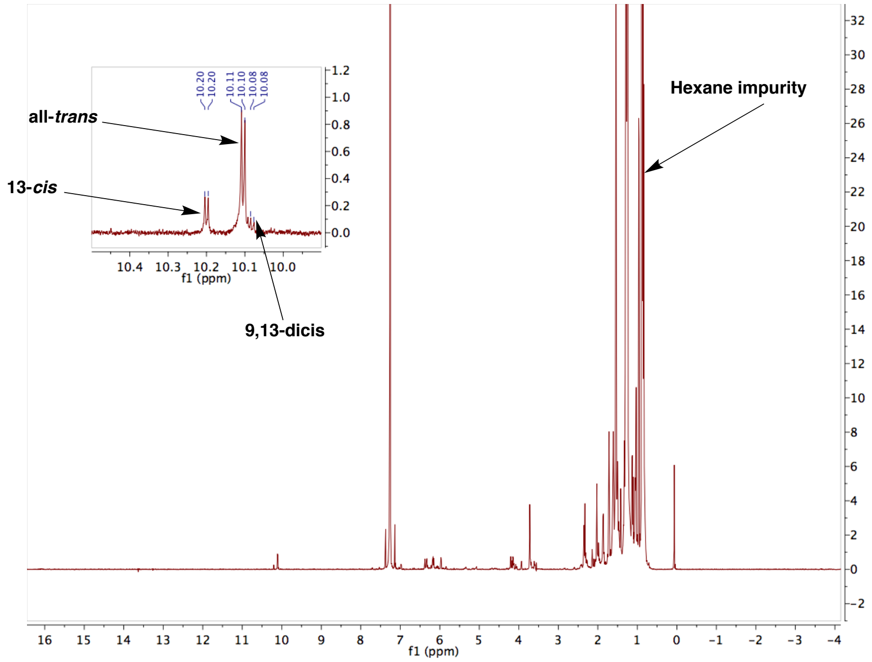


Figure I-14. ^{1}H-NMR spectrum of retinal extracted from the protein complex recorded in $CDCl_3$ solvent.

We surmised that the drop in absorption of our retinal-bound **M1**-L121E upon irradiation is due to both the generation of a new species with a lower extinction coefficient, along with some potential photobleaching. To gain further insight into the process, the irradiation of all-*trans* retinal-bound **M1**-L121E crystals with a 399 nm laser (Figure 11d) was performed. The crystals were frozen in liquid nitrogen after various time intervals of irradiation (30 seconds and 5 minutes), and their structures were determined. After five minutes and 30 seconds of laser irradiation, the electron density of the retinal shows no evidence of all-*trans* retinal and instead is consistent with its quantitative photo-conversion to 13-*cis* (Figures **I**-11d and **I**-15). As shown in Figure **I**-11e, the overlay of 13-*cis* generated after laser irradiation with all-*trans* shows both substantial β-ionone ring rotation and polyene movement. Furthermore, the conformational change of Glu121, which follows the movement of the imine nitrogen, and Lys108, is apparent. Interestingly, in all these structures, the position of Tyr39 remains fixed before and after irradiation. Tyr39 may act as a “steric pillar” that dictates the path of isomerization in a specific direction (Figure **I**-11e). The role of Tyr39 is reminiscent of Trp86 in bacteriorhodopsin, which sterically packs against the retinal and assists in directing the photoisomerization of the C13-C14 double bond [62]. As shown in the overlay of **M1**-L121E and bacteriorhodopsin (Figure **I**-11F), Glu121 and Tyr39 occupy a similar space relative to the retinal, as does Trp86 in bacteriorhodopsin. Once again, the fortuitous, yet critical, role of the L121E mutation is apparent, as it aids in orienting Tyr39 in a conformation that mimics Trp86 in bacteriorhodopsin, leading to specific 13-*cis* photoisomerization.[48]

I-3-2. Thermal stability of 13-*cis* retinal in the crystal

With photoisomerization to 13-*cis* confirmed, its thermal stability was investigated next. To find the shortest exposure time that can lead to 13-*cis* formation in the crystal, the crystal was irradiated with the indicated 399 nm laser (please see section I-7-10 for the details) for 30 seconds, and then the crystals were immediately flash-frozen in liquid N_2. The 13-*cis* generated after 30 seconds of laser irradiation depicted the same β-ionone ring rotation as observed for the five-minute laser-irradiated crystal (**M1**-L121E light state) (Figure I-15), while the overlay of this structure with the all-*trans* retinal-bound structure of **M1**-L121E (dark state) shows drastic differences in the retinal trajectory and β-ionone ring.

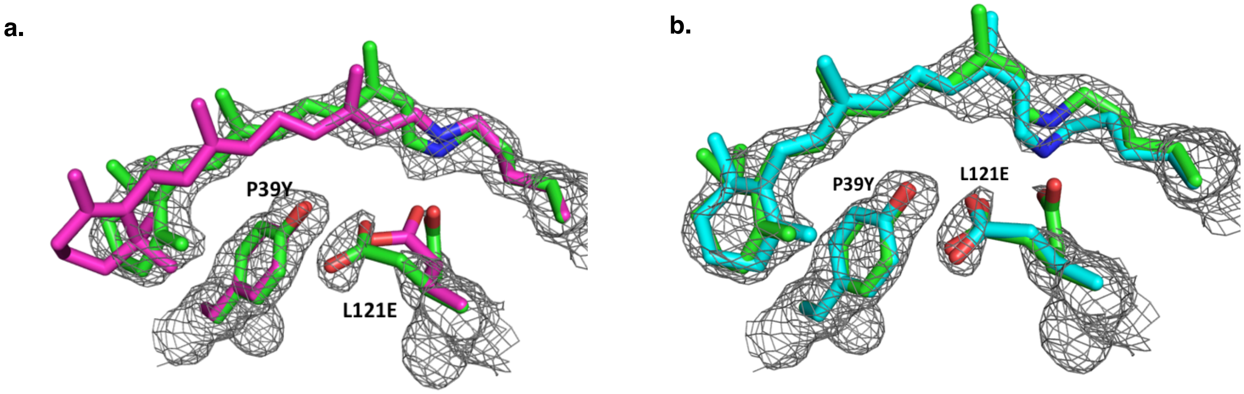


Figure I-15. a. Overlay of the dark state **M1**-L121E (pink) and after 30 sec laser irradiation (green-13-*cis*) revealing the changes in the polyene and ionone ring conformation of retinal after laser irradiation. Electron density belongs to the crystal irradiated for 30 sec (countered at 1σ). b. Overlay of 30 sec irradiation structure (green) with five min irradiation with laser (cyan) of **M1**-L121E, demonstrating the similar ionone ring and polyene conformation and verifying the formation of 13-*cis* after 30 sec laser irradiation. Electron density belongs to the crystal irradiated for 30 sec (countered at 1σ).

To this end, photochemically-generated 13-*cis*-bound **M1**-L121E crystals were left in the dark at room temperature for 10 and 25 minutes. We anticipated the thermal reversion of the 13-*cis* isomer to all-*trans*, since it appears to be the thermodynamically more stable isomer of retinal in **M1**-L121E. This would complete the photocycle that is

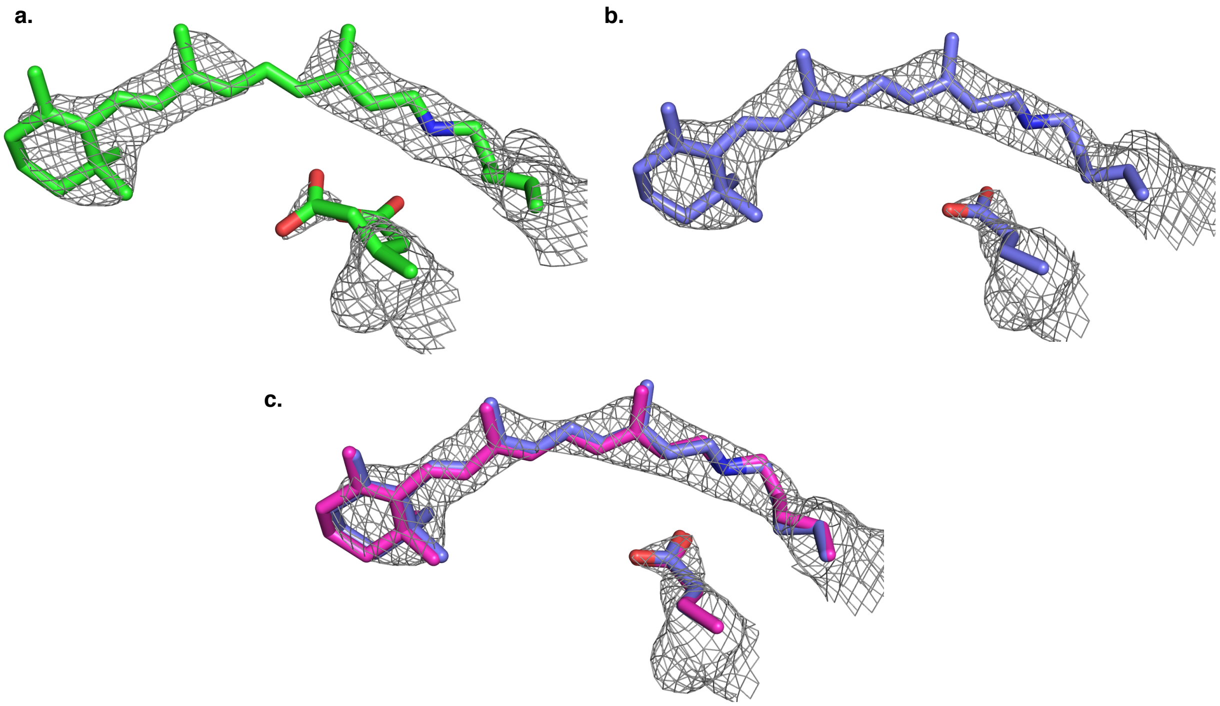


Figure I-16. a. All-*trans* retinal-bound **M1**-L121E crystal after 30 sec irradiation with laser (399 nm) and subsequent incubation in the dark for 10 minutes (electron density contoured at 1σ). This crystal structure indicates the regeneration of all-*trans* retinal through the thermal isomerization of 13-*cis*. Although the chromophore adopted the all-*trans* retinal conformation, L121E does not reset completely. b. Structure from an all-*trans* retinal-bound **M1**-L121E crystal after 30 sec irradiation with laser (399 nm) and subsequent incubation in the dark for 25 minutes (electron density contoured at 1σ), showing the regeneration of all-*trans* retinal and L121E adopted the same conformation as the initial dark state conformation. c. Overlay of structures from all-*trans* bound **M1**-L121E crystal (dark state, in pink) and **M1**-L121E irradiated with the laser (399nm) for 30 sec and subsequently incubated in the dark for 25 minutes (purple) verifies the similarity of the retinal trajectory in both structures, showing the thermal isomerization of the 13-*cis* created after 30 sec laser irradiation to produce all-*trans* retinal (RMSD=0.1). The electron density (contoured at 1σ) was calculated from the structure after 25 minutes incubation in the dark.

essentially identical to that of microbial rhodopsin, though on a significantly slower time scale. The structures of all the thermally-equilibrated crystals (after 10 and 25 minutes of equilibration) revealed electron density consistent with all-*trans*-bound **M1**-L121E, confirming the completion of a microbial rhodopsin photocycle (Figure I-16). The main characteristics of the structure, most notably the rotation and trajectory of the β-ionone ring, are clearly observed (Figure 1-16c). This is consistent with the fact that the extraction of irradiated **M1**-L121E shows the majority of the chromophore to be all-*trans* (Figures I-

12 and I-13). Thus, a complete photocycle, analogous to that of microbial rhodopsin, has been generated in an orthogonal protein and confirmed at atomic resolution.[48]

I-3-3. Atomic resolution crystal structure of 13-*cis* retinal in the dark

We were intrigued by the ionone ring rotation (Figure I-11e) accompanying the photoconversion to 13-*cis* and hypothesized that the stabilization of the rotated ionone ring conformation would also stabilize 13-*cis* in the binding pocket. To this end, Ala32, located at the mouth of the binding pocket, was mutated to aromatic amino acids to discourage rotation of the ring due to steric compression afforded by the larger amino acids. Gratifyingly, the structure of 13-*cis*-retinal **M1-**L121E:A32Y clearly shows 13-*cis* exclusively bound in the binding cavity (Figures I-17a and I-18a). Analysis of the high-resolution structure indicates that the 13-*cis* isomer is stabilized by the packing of several aromatic residues, including Tyr39, Tyr32, Tyr59, and Trp109, inside the binding pocket, mostly through "aromatic sandwiches" (Figure I-17a). Interestingly, this is similar to bacteriorhodopsin, where the chromophore finds itself well-packed in a sphere of aromatic residues, Figure I-17b. The overlay of this structure with the 13-*cis*-retinal **M1-**

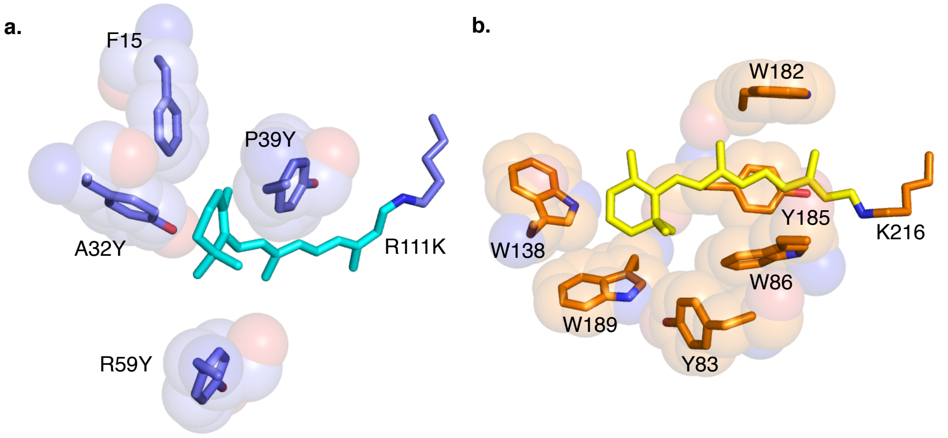


Figure I-17. a. Aromatic residues located in the binding pocket of 13-*cis* bound **M1**-L121E:A32Y; b. Bacteriorhodopsin binding pocket where the chromophore is sandwiched with aromatic residues.

L121E post-irradiation (13-*cis*-isomer) shows both the *β*-ionone and polyene moieties to be quite similar, while the overlay of this structure with the all-*trans* retinal-bound structures shows obvious structural variations (Figure I-18). These results demonstrate that by increasing the pK_a (using the L121E mutation) and altering the β-ionone ring packing (via the A32Y mutation), a 13-*cis* "unfriendly" binding pocket (**M1)** is converted into a 13-*cis*-specific protein, nestling a well-ordered 13-*cis* inside the binding pocket.[48]

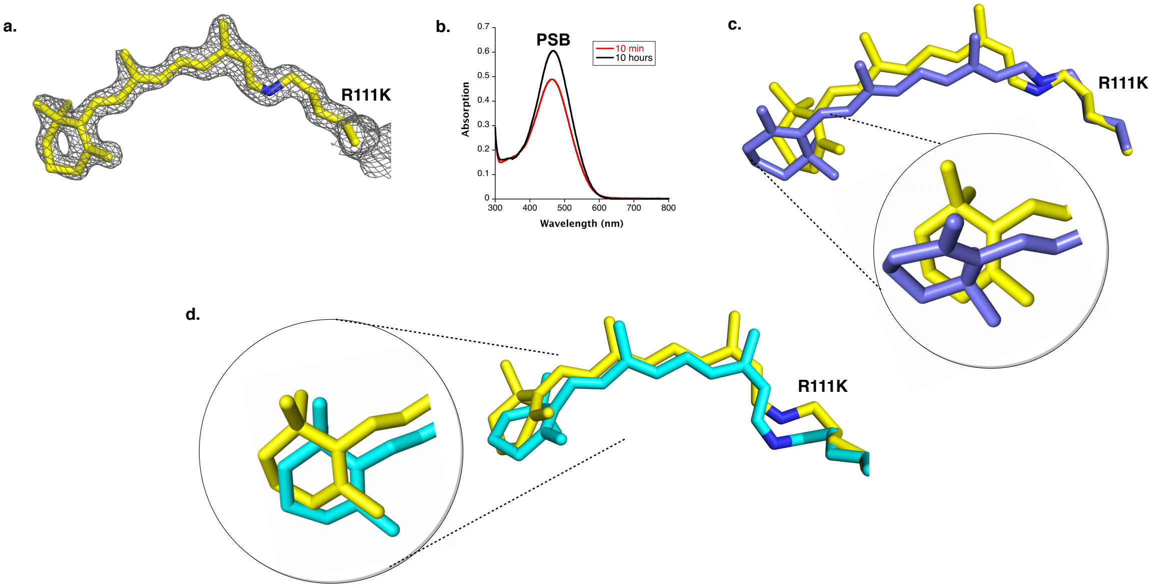


Figure I-18. a. 13-*cis* electron density in **M1**-L121E:A32Y variant countered at 1.5 σ. The atoms are colored by type, with the C atoms in yellow. b. UV-vis spectrum of **M1**-L121E:A32Y bound with the 13-*cis* retinal, depicting the stable PSB after 10 hours of incubation. c. Overlay of 13-*cis* bound **M1**:L121E:A32Y (yellow) with all-*trans* bound **M1**:L121E (purple) without irradiation of either. d. Overlay of 13-*cis* retinal bound **M1**-L121E-A32Y (yellow) with the structure from a crystal of all-*trans* bound **M1**-L121E that was irradiated with laser (5 min) and immediately frozen (cyan).

I-3-4. Mechanistic insight and prediction

Our interpretation of the results described above naturally leads to predictions regarding the mechanism in which both Glu121 and Tyr39 are essential for providing a 13-*cis* isomerizing system. To test this prediction, mutations were made to both Tyr39 and Glu121. As predicted, the P39Q mutation that led to **M1**-L121E:P39Q resulted in a protein complex incapable of isomerizing the C13 double bond. Instead, it resulted in yet another imine isomerizing system. The crystal structure of this mutant was obtained in the dark at 2.2 Å resolution. As shown in the overlay of **M1**:L121E P39Q (dark state) vs. **M1**-L121E (dark state), replacing Tyr39 with Gln causes no steric interaction in the vicinity of the chromophore around the C13-C14 double bond. Furthermore, the chromophore is not pushed to the retinal trajectory as seen for M1-L121E, a conformation that is

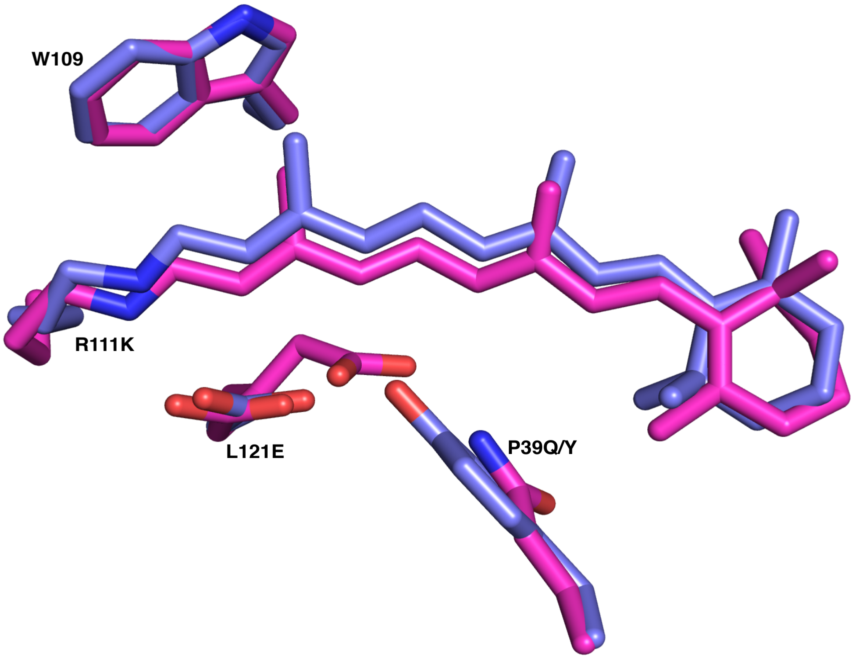


Figure I-19. Overlay of M1-L121E (purple-blue) with **M1**-L121E:P39Q (pink). The P39Q mutation does not increase the steric interaction in the vicinity of the C13-C14 double bond. Furthermore, in the absence of P39Y, the chromophore does not move to a trajectory seen for **M1**-L121E, highlighting the essential role of P39Y for defining the chromophore trajectory in **M1**-L121E and presumably directing the isomerization toward 13-*cis* isomer formation.

presumably necessary for retinal isomerization around a C13-C14 double bond (Figure **I**-19). Additionally, the UV spectrum of **M1**-P39Q:L121E, in contrast with **M1**-L121E, showed the PSB absorption (457 nm) red-shifted to a new PSB absorbing at 507 nm upon irradiation with UV (B.P 300-400 nm) (Figure **I**-20a). **I**nterestingly, irradiation with green light (L.P.>500 nm) retrieves the original PSB (457nm). The irradiation of the **M1**-P39Q:L121E crystal with UV showed a drastic change in color, presumably correlating with the shift observed in the UV-vis spectrum (Figure **I**-20-b). The crystal structure after UV irradiation with a handheld TLC UV lamp was determined at 2.6 Å resolution. The conformation of the polyene and β-ionone ring confirmed the presence of all-*trans* retinal after five minutes of UV irradiation (Figure **I**-20c). Furthermore, as shown in Figure **I**-20d, the overlay of the UV-irradiated structure with **M1**-L121E bound with 13-*cis*, generated after five minutes of laser irradiation, further confirmed that the product of **M1**:L121E P39Q irradiation is all-*trans*. (It is not correlated with the 13-*cis* retinal trajectory.) This demonstrates the importance of the steric interaction between Tyr39 and the chromophore in directing isomerization to the 13 bond.

To verify the role of the L121E mutation as an iminium-stabilizing species, we examined the photochemical characteristics of the structurally conservative **M1**-L121Q mutant. As shown in Figure I23a, irradiation of **M1**-L121Q in solution leads to behavior similar to mutants that undergo imine isomerization with two apparent pK_a's for each isomer. UV light irradiation (<400 nm) results in a loss of intensity of the 374 nm SB absorption, concomitant with a large increase in the PSB absorption at 522 nm. Conversely, irradiation with visible light (>500 nm) reduces the PSB absorption at 522 nm, and increases the 374 nm SB absorption. At first, this result was not unexpected, since the loss of the salt bridge afforded through the interaction of the iminium with Glu121 could reestablish the imine *cis*/*trans* isomerization observed before, leading to two states

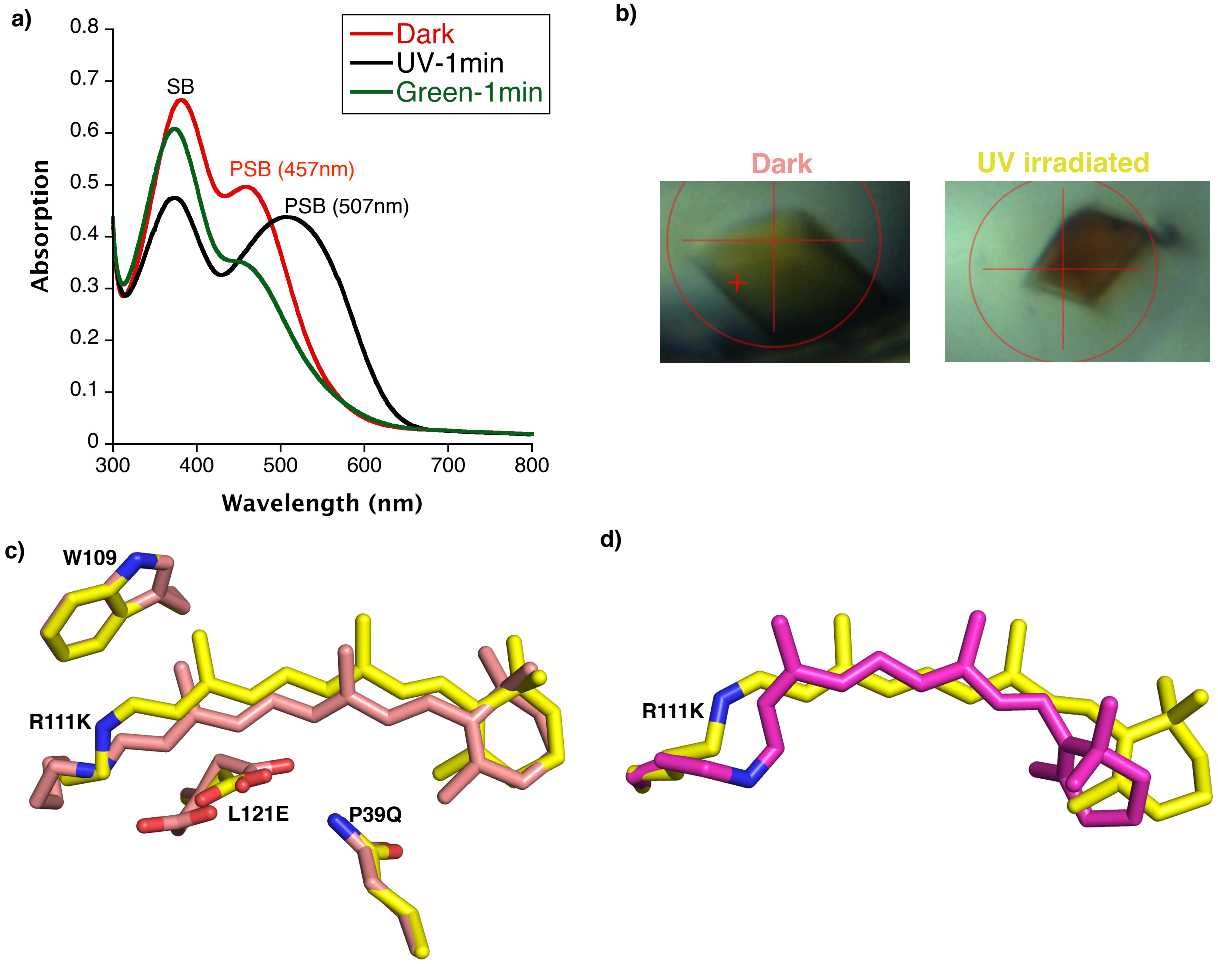


Figure I-20. a. UV-vis spectrum of **M1**-P39Q:L121E in the dark, after UV (B.P.300-400 nm) irradiation, and after subsequent green-light irradiation (L.P 500 nm). b. Drastic color change in the crystal upon irradiation with the UV. c. Overlay of dark state (pink) and UV-irradiated state (yellow) of **M1**-P39Q:L121E crystal structures, depicting the all-*trans* conformation for both states. d. Overlay of UV-irradiated crystal structure of **M1**-P39Q:L121E vs. the light state of **M1**-L121E.

with distinct pK_a's via iminium *cis*/*trans* isomerization. Nonetheless, crystal structure analysis reveals a surprising and unanticipated mechanism for the observed pK_a change: photo-irradiation leads to side-chain motion.[48]

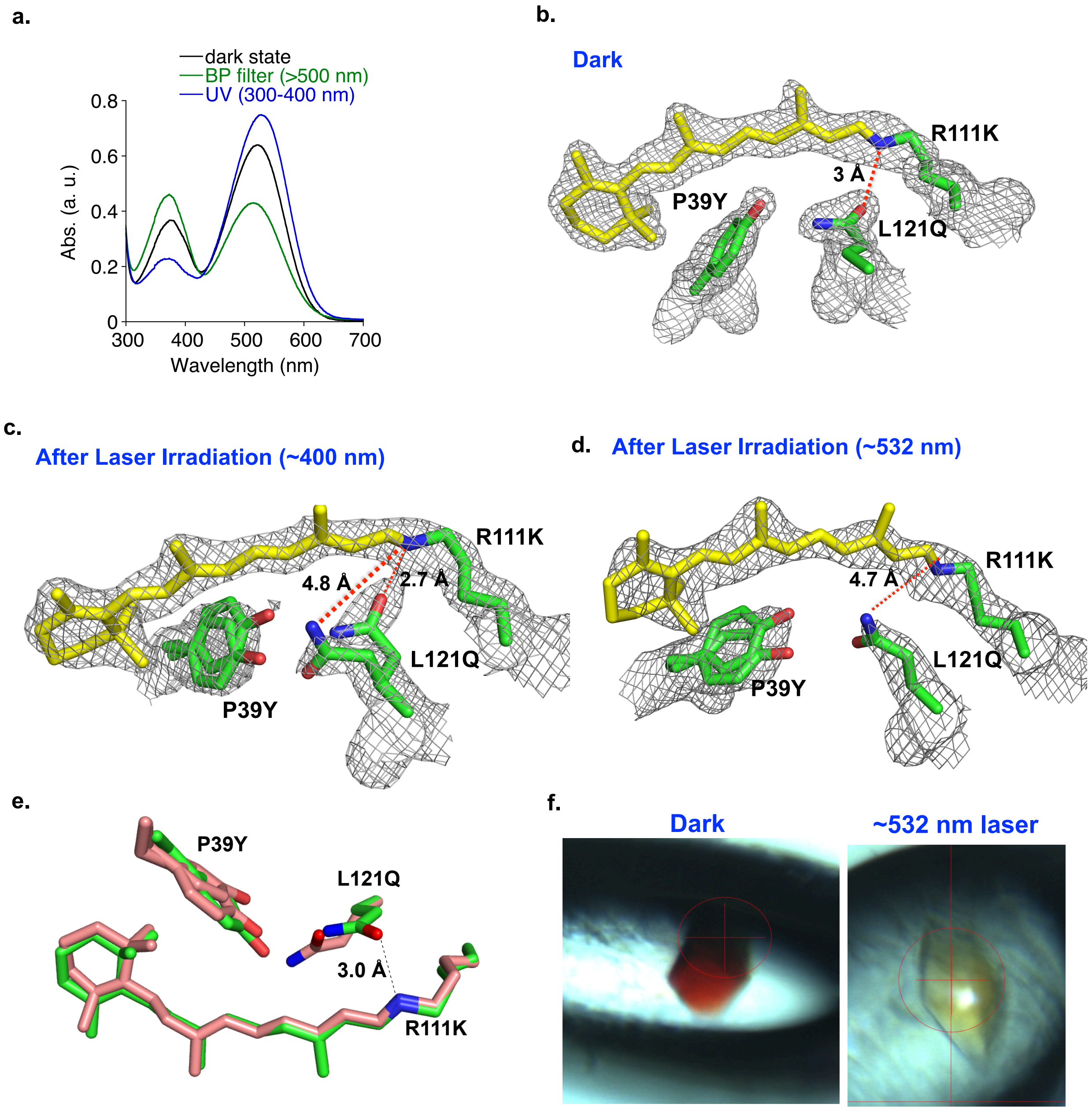


Figure I-21. a. UV-vis spectra of **M1**-L121Q before and after irradiation with green light, showing a high/low pK_a system where green-light irradiation leads to decreased PSB and increased SB absorption. b. Crystal structure of **M1**-L121Q obtained in the dark. L121Q has a direct hydrogen bond with R111K. c. Crystal structure of **M1**-L121Q after 5 min irradiation with 399 nm laser. L121Q and P39Y adopted two conformations but no isomerization occurred. This shows the importance of sidechain conformation stability in directing the photoisomerization product. L121Q in one conformation is far away from iminium (4.8 Å). d. After 10 min laser irradiation of crystals with 532 nm laser (close to PSB maximum absorption), the L121Q swings away from iminium, lowering the pK_a of the chromophore and leading to the SB formation. e. Overlay of the all-*trans*-bound **M1**-L121Q imine before (green carbons) and after (salmon carbons) laser irradiation at 528 nm, clearly showing the Gln121 movement. f. Drastic color change of crystals upon irradiation with 532 nm laser.

As illustrated in Figure I-21, the dark-state structure of **M1**-L121Q is almost identical to that of **M1**-L121E, with Gln121 making a direct hydrogen bond with the iminium nitrogen atom. However, irradiation with the 399 nm laser shows no evidence of chromophore isomerization or conformational change. Instead, Gln121 adopts two conformations, the original conformation seen in the dark, and a new position that is rotated away from the iminium, resulting in a naked imine nitrogen atom with no other interacting residues (Figure I-21c). Such an imine would be expected to have a substantially-suppressed pK_a relative to the dark state. Subsequent irradiation with a 532 nm laser leads to a complete conversion to the low pK_a form, where Gln121 makes no interaction with the imine (Figures I-21d and I-21e).

In contrast with the expected isomerization of the chromophore's double bond that would lead to a change in the environment of the imine, and thus its pK_a, the excited state of the **M1**-L121Q retinal complex instead leads to a change in the conformation of an interacting amino acid sidechain, effectively altering the environment. This is reminiscent, albeit on a much longer timescale, of the sub-picosecond X-ray laser time-resolved study of bacteriorhodopsin, which showed that the critical water molecule moves away from the chromophore in response to the change in the polarity of the excited state, before the chromophore itself displays substantial motion.[19] Such subtle changes in the structure of the surroundings upon irradiation would be difficult to identify in the absence of high-resolution structural interrogation of the irradiated systems. This may explain the complex hydrogen-bonding network holding the two key aspartates in place in bacteriorhodopsin.[5, 63] Perhaps this prevents the interacting residues from moving during photo excitation, resulting in a low pK_a form, which would sabotage the bacteriorhodopsin photocycle.[48]

I-3-5. Irradiation study of **M1**-L121E:A32Y

As cited above, we were able to confine the 13-*cis* retinal in the binding pocket of our protein by placing A32Y in the vicinity of the β-ionone ring, in the context of **M1**-L121E, giving rise to a well-ordered chromophore inside the binding cavity and well-diffracting crystals. The next question is what would happen to the **M1**-L121E-A32Y/13-*cis* retinal complex upon irradiation. To test this, a 399 nm laser was used as a source of irradiation.

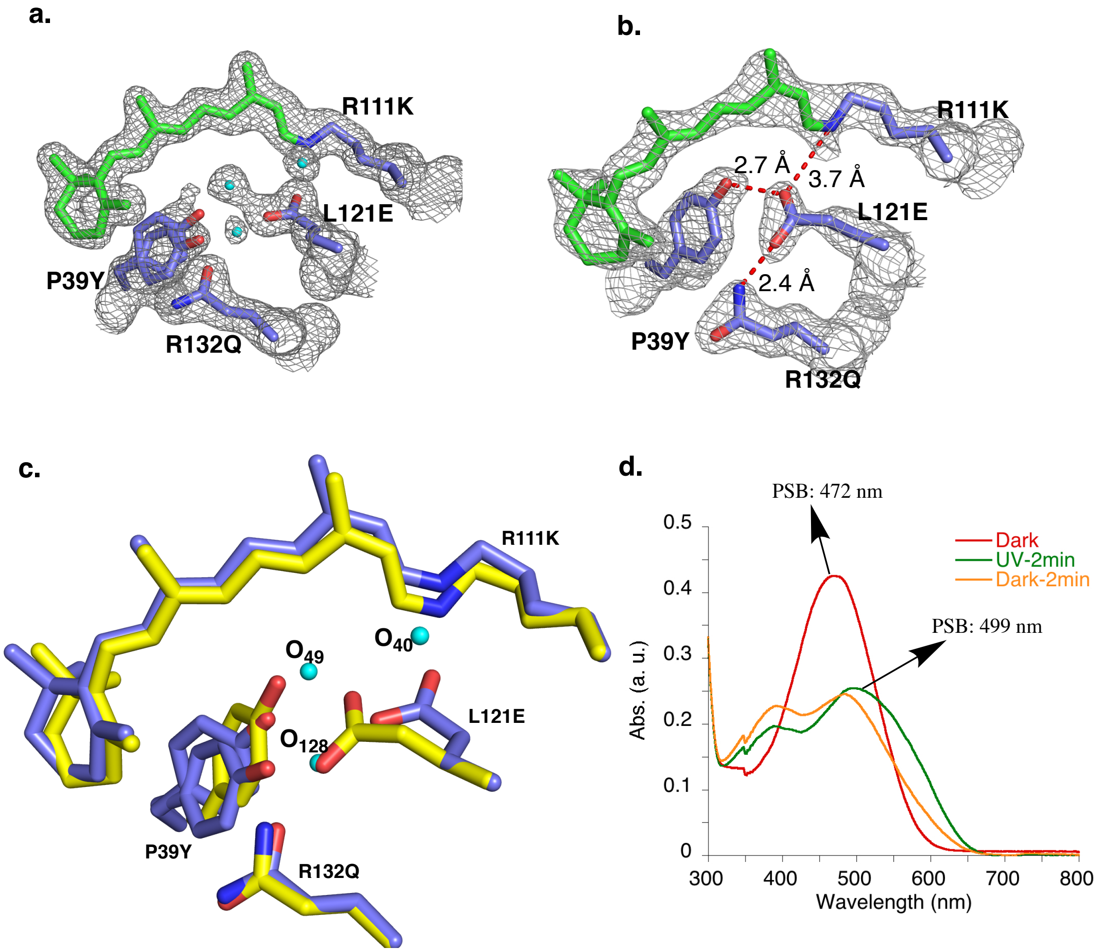


Figure I-22. a. Dark state of **M1**-L121E:A32Y bound with 13-*cis* retinal, showing well-ordered water molecules mediated between the retinal and protein (electron density is contoured to 1 σ). b. After 399 nm laser irradiation, no isomerization occurs; however, the Lys111 side chain and retinal polyene move vertically ~ 0.7 Å, which leads to Glu121 movement and disordering of the water molecules (electron density is contoured to 1 σ). c. Overlay of dark state (purple-blue) vs. light state demonstrated the subtle movement of the chromophore, Lys111 side chain, and Glu121, disordering the water molecules after irradiation.

As shown in Figure **I**-22, upon irradiation with the laser, the chromophore still holds the 13-*cis* retinal configuration; however, the position of Lys111, the retinal polyene, P39Y, and L121E change subtly, which disorders the ordered water molecules around the iminium and increases the distance between L121E and the nitrogen of the imine (from 2.8 to 3.7 Å). The movement of L121E and disordering of the water molecules around the polyene are expected to suppress the pK_a of the chromophore and shift the retinal absorption to an SB population. This observation is consistent with the UV irradiation of the **M1**-L121E:A32Y/retinal complex in solution, revealing that upon irradiation with UV, the SB population increases. Additionally, the UV-vis spectra depicts that upon irradiation with UV, a new, short-lived red-shifted PSB forms. (Due to the short lifetime of the red-shifted PSB, we were not able to completely detect the absorption wavelength by our method, but based on the color of the solution right after irradiation, it would be expected that the new PSB possesses a higher absorption wavelength than 499 nm.) As shown in Figure **I**-22c, increasing the distance between L121E and nitrogen of iminium (from 2.8 Å to 3.7 Å), along with dispelling the water molecules around the iminium, might lead to the new red-shifted PSB by allowing the delocalization of charge on the polyene. Overall, the UV-vis spectra and crystal structure together reveal another important photoswitching mechanism for the retinal/protein complex upon interaction with light. The irradiated chromophore can avoid the absorption of light through subtle movement in the binding pocket, which leads to the changes in the surrounding residues and water molecules, causing a low pK_a form of the chromophore and a bathochromic shift from the irradiated wavelength. Presumably, in this case, the binding pocket restricts any vibrational flexibility of the chromophore for isomerization where A32Y is perpendicular to the ionone ring, and in contrast to L121Q, the charged residue of L121E does not reorient without the impulsive motion of the Lys111 side chain. However, we cannot rule out another possible mechanism for this observation, in which the chromophore undergoes a fast photoisomerization and thermal reversion to the original state (representing the bacteriorhodopsin photocycle in reverse), which is not detectable within the time window of our method of study.

I-3-6. Other mutational studies on the M1 construct

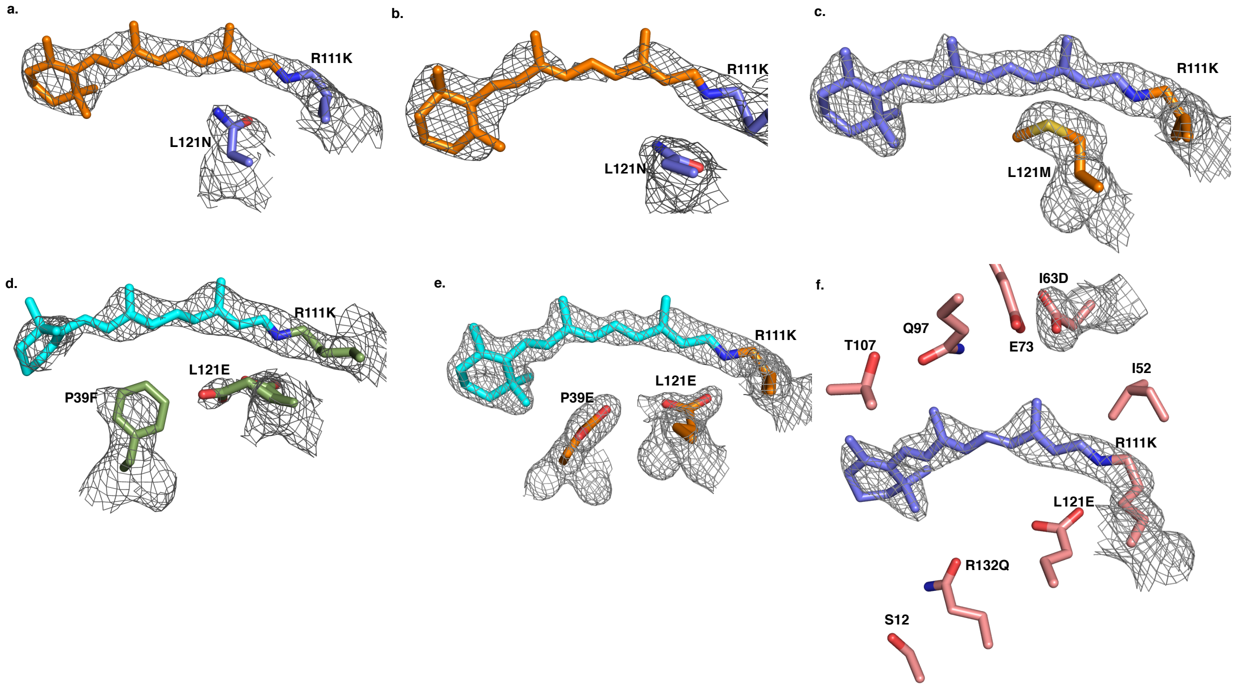


Figure I-23. a. Crystal structure of **M1**-L121N/all-*trans* retinal in the dark. b. Crystal structure of **M1**-L121N/all-*trans* retinal after 3 min laser irradiation with 399 nm laser. c. Crystal structure of **M1**-L121M /all-*trans* retinal in the dark (PDB code:6NNX). d. Crystal structure of **M1**-L121E:P39F/ all-*trans* retinal in the dark. e. Crystal structure of **M1**-L121E:P39E in the dark (PDB code:6NNY). f. Crystal structure of **M1**-L121E:I63D in the dark (PDB code: 6NOE). The presence of I63D can potentially generate a hydrogen bond network on the other side of the chromophore opposed to L121E. The electron density was contoured to 1 σ for the $2F_O$-F_C map.

Other than the mutational studies cited above, other variants using the **M1**-L121E construct were made, as shown in Table I-2.

The **M1**-L121N/retinal photocycle UV-vis spectra and the crystal structure were obtained, as shown in Figures I-23a and I-24a. The photoirradiation of **M1**-L121N/retinal depicted a two-pK_a system reminiscent of **M1**-L121Q . The crystal structure of the dark and laser-irradiated forms were obtained (Figures I-23A and I-23b), demonstrating that the chromophore does not undergo isomerization upon 399 nm laser irradiation. However, the exact position of L121N in the dark and light state is not as clear as L121Q in elucidating another photoinduced side-chain motion mechanism. **M1**-L121D was also

Table I-2. Other mutational studies on **M1** construct.

Variants	Crystal structure
M1-L121N	Available
M1-L121M	Available (PDB code:6NNX)
M1-L121D	Was not crystallized
M1-L121E:P39F	Available
M1-L121E:P39N	Available
M1-L121E:P39E	Available (PDB code:6NNY)
M1-L121E:F65Y	Was not solubly expressed
M1-L121E:I63D	Available (PDB code:6NOE)
M1-L121E:F3Q	Was not solubly expressed
M1-L121E:A33W:F3Q	Was not solubly expressed
M1-L121E:I31W	Was not solubly expressed
M1-L121E:A32W	Available
M1-L121E:T56Y	Was not solubly expressed
M1-L121E:I52E	Was not solubly expressed
M1-L121E:A23Y:I52E	Was not solubly expressed

created, which shows some photocycle behavior with a high photobleaching effect (Figure **I**-24b). The crystallization attempt for this variant was not successful. **M1**-121M/retinal was prepared and crystallized (Figure **I**-23c), and showed no tangible spectroscopic or structural variation upon light irradiation. The **M1**-L121E:P39F variant was made and crystallized in the presence of retinal. The crystal structure of this construct revealed a rotation of the β-ionone ring as a result of the P39F mutation, which possesses a conformation different from P39Y held by L121E. The chromophore possessed a low $\mathrm{p}K_\mathrm{a}$ without showing a photocycle (Figure **I**-23d). **M1**-L121E:P39E was successfully crystallized (Figure **I**-23e), and the spectroscopic data was measured. The UV-vis spectrum demonstrates the same feature as **M1**-L121E, where upon irradiation with different wavelengths, both the SB and PSB population decrease; however, the spectrum also shows some recovery of the SB population over time using UV-vis spectroscopy (Figure **I**-24e). The chromophore trajectory was very similar to **M1**-L121E, as shown in Figure **I**-23. The next mutant was **M1**-L121E:P39N, which depicted the photocycle of the protein/retinal complex upon irradiation with green and UV light, Figure **I**-24c. However, this mutant could not be crystallized. Some new variants were also made with the hope of stabilizing the 13-*cis*-15-*anti* configuration by imposing F3Q, I63D, F65Y,I52E, mutations to yield **M1**-L121E:F3Q, **M1**-L121E:A33W:F3Q, **M1**-L121E:I63D, **M1**-L121E:F65Y, **M1**-L121E:I52E, and **M1**-L121E:A32Y:I52E. Unfortunately, only **M1**-L121E:I63D was solubly expressed. The crystal structure of **M1**-L121E:I63D bound with all-*trans* was obtained (Figure **I**-23f); however, no interesting spectroscopic behavior or deviation from **M1**-L121E was observed. Finally, to elucidate the effect of aromatic residues in the vicinity of the β-ionone ring, **M1**-L121E:I31W, **M1**-L121E:A32W, and **M1**-

L121E:T56Y were created. Unfortunately, **M1**-L121E:I31W and **M1**-L121E:T56Y were not crystallized; however, **M1**-L121E:A32W was crystallized with retinal and showed no electron density for the chromophore, indicating a disordered ligand in the binding pocket.

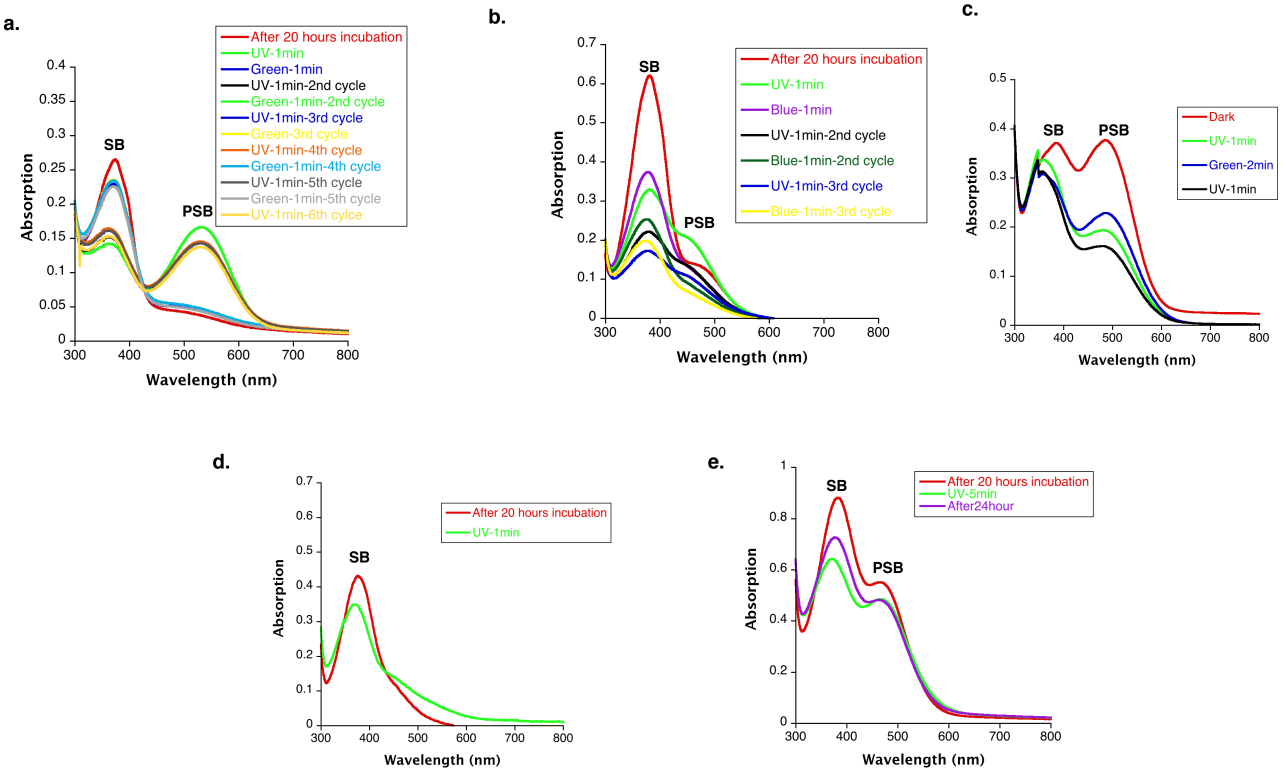


Figure I-24. UV-vis photocycle spectra of different all-*trans* retinal-bound variants: a. **M1**-L121N, b. **M1**-L121D, c. **M1**-L121E:P39N, d. M1-L121E:P39F, and e. **M1**-L121E:P39E. UV (B.P 300-400 nm) and green light irradiation (Y filter L.P>500 nm) were performed with a mercury lamp as the source of irradiation.

I-4. 13-*cis* bound hCRBPII crystal structure

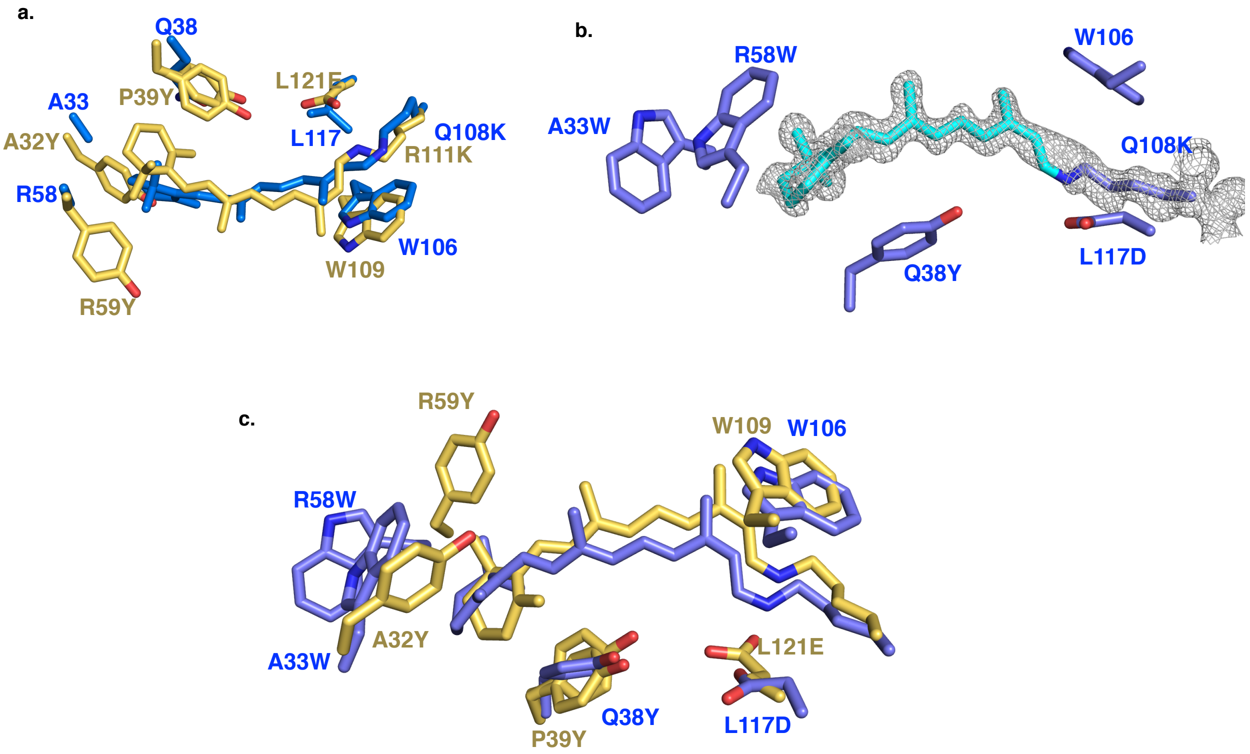


Figure I-25. Overlay of **M1**-L121E:A32Y-CRABPII (yellow) and Q108K:K40L-hCRBPII (blue) showing the equivalent positions between the two proteins in the retinal binding pocket b. Crystal structure of **H1** bound with 13-*cis* retinal. The electron density is contoured to 1σ. c. Overlay of **H1** (blue) and **M1**-L121E:A32Y bound with 13-*cis* retinal, highlighting the aromatic interactions holding the chromophore.

Having demonstrated the key elements of stabilizing 13-*cis* retinal in the protein binding pocket using the **M1**-L121E:A32Y variant of CRABPII, the next goal was to introduce these key residues in the hCRBPII binding pocket, which possesses almost 40% sequence similarity with CRABPII. The Q108K:K40L:T51V:T53S:R58W:Y19W:A33W:L117D:Q38Y (**H1**) variant was employed as our template, which contains several aromatic residues (R58W, Y19W, A33W, and Q38Y) with an acidic residue, L117D, in the vicinity of nitrogen of iminium. Notably, the 33, 38, 58, and 117 positions in CRBPII correlate with the 32, 39, 59, and 121 positions in CRABPII (Figure I-25a). To our delight, the crystal structure of 13-*cis* bound **H1** was obtained at 1.1 Å resolution, depicting a well-ordered ligand in the binding pocket (Figure

I-25b). Based on the overlay of **M1**-L121E:A32Y/13-*cis* and **H1**/13-*cis,* as anticipated, the **H1** construct creates aromatic sandwiches surrounding 13-*cis* retinal, leading to a well-ordered ligand in the binding pocket (Figure I-25c). This is in contrast to the all-*trans* bound **H1** crystal structure that led to a disordered chromophore. This once again demonstrates the critical role of the aromatic residues in stabilizing the 13-*cis* in the rhodopsin mimic.

I-5. Time-resolved x-ray crystallography experiment with rhodopsin mimic

As mentioned above, using CRABPII as an engineering template, we observed four distinct processes that occur upon irradiation of the retinal chromophore in the binding cavity of the protein: imine bond isomerization (**M1**), 13-*cis* specific isomerization (**M1**-L121E), photoinduced side-chain motion without a retinal conformational change (**M1**-L121Q), and retinylidene movement without isomerization, leading to the side-chain

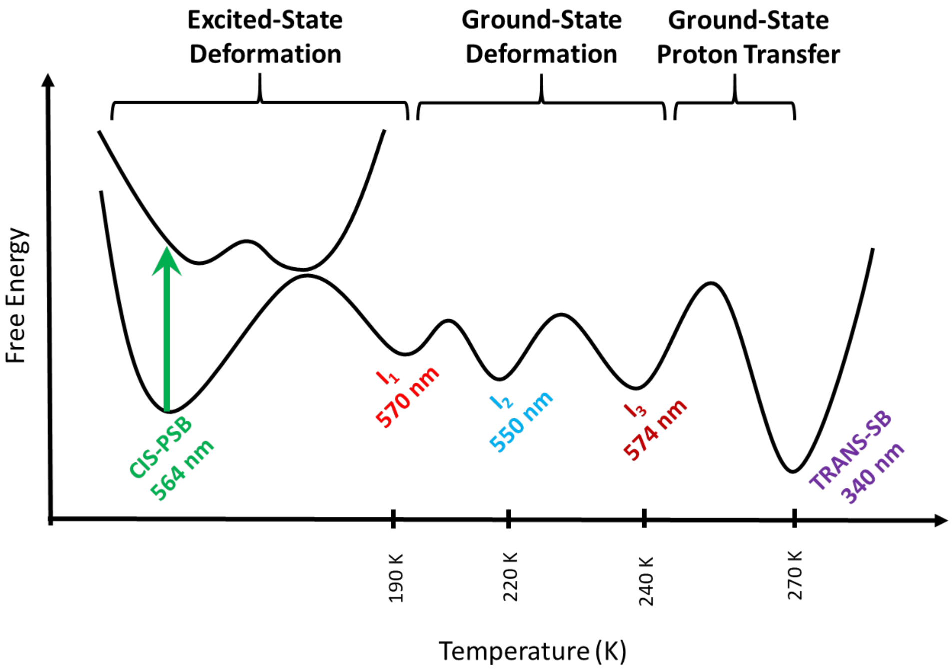


Figure I-26. Simplified Potential Energy Surface for the forward (top) and reverse (bottom) dynamics of **M1** initiated with UV excitation (forward) or green-light (reverse) excitation. Wavelengths are indications of the true peak wavelength for the intermediates (provided by Adam Jenkins).

motion (**M1**-L121E:A32Y). Elucidating the detailed mechanism of these systems requires a time-resolved experiment. As a starting point, we decided to use the **M1** system**,** where the imine bond undergoes photoisomerization for the time-resolved experiment. In collaboration with Prof. Larsen's group at UC Davis (Mr. Adam Jenkins), we were able to identify several intermediates using cryokinetics and ultrafast spectroscopy experiments. For **M1**, Mr. Adam Jenkins calculated the simple potential-energy surfaces of the ground-state intermediates for both forward (SB) and reverse (PSB) reaction directions (Figure I-26). The ultrafast data depicted one or more structural evolutions before a protonation (or deprotonation) reaction ensues. For the forward direction (SB excitation), the excited-state lifetime was calculated to be 30 ps. For the reverse direction (PSB

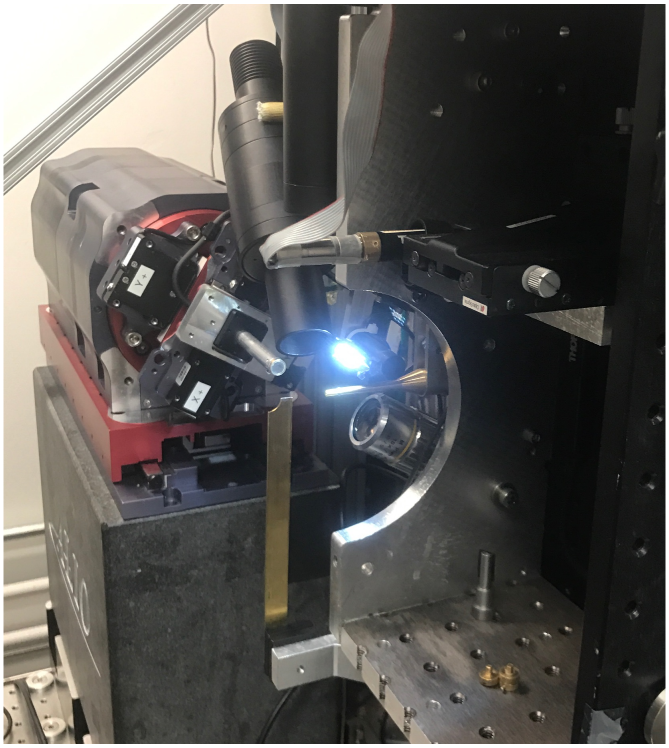


Figure I-27. The pump-probe method was used for the time-resolved x-ray experiment. The laser was used to initiate the reactions in the crystals, and the x-ray with microsecond-duration pulses was used to monitor the changes in the crystals. The x-ray and laser are positioned perpendicular to each other to increase the efficiency of the experiment. Photo credit: Prof. Geiger.

excitation), the excited-state kinetics were much faster (4.4 ps). This data revealed the possibility of observing intermediates during the photoisomerization of **M1**. The next step was to study the time-resolved x-ray experiment in **M1** crystals, where we could identify the structure of intermediates at atomic resolution. Two important features of **M1** crystals made them great choices for these experiments: first, the crystals are not mosaic; second, the reactions in these crystals are highly reversible when sequentially irradiated with green and UV light, and are accompanied by concomitant, drastic color changes offering a visible signature for each isomeric form. To perform the time-resolved experiment, the synchrotron resource for time-resolved X-ray science at BioCARS was chosen. Many **M1** crystals were grown in a sitting-drops for the first experiment (setup shown in Figure I-27). The "pump-probe" method was selected for the time-resolved x-ray crystallography of the **M1**-retinal complex. The UV laser was used to initiate the reaction in the crystals, and the X-ray with microsecond-duration pulses was used to monitor the reaction in the

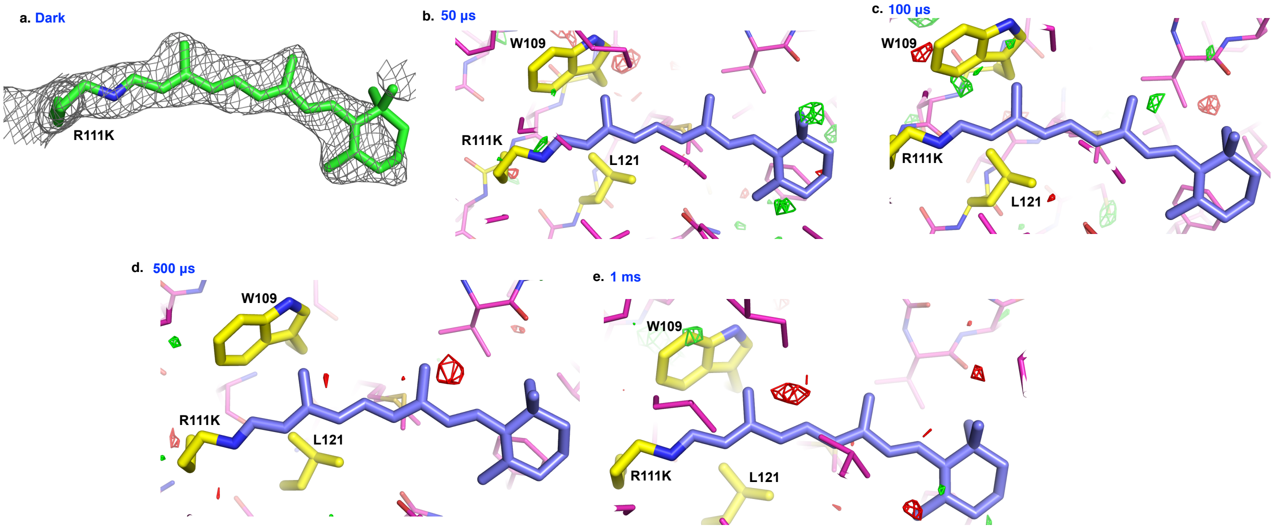


Figure I-28. a. A $2F_O$-F_C Electron density map of the dark state of **M1** collected at room temperature using the Laue method, contoured at 1σ for**.** The time-dependent difference maps were subtracted from the dark state: b. 50 μs, c. 100 μs, d. 500 μs, and e. 1 ms.

crystals. The X-ray and laser were positioned perpendicular to each other to increase the efficiency of the data collection. In contrast with a normal synchrotron experiment, where monochromatic X-rays are used and the crystal is rotated on a goniometer at 100 K, in this experiment, the crystal was located in a precisely stationary position, and a “pink” polychromatic X-ray beam is used. The result is Laue diffraction, which gives significantly more data on a single image. The data must also be collected at room temperature as cryo temperatures will completely freeze out any molecular motion.

The Laue diffraction experiment on the **M1** crystals was performed, and the data was kindly processed using the “linear analytical” integration method by Dr. Srajer from BioCARS. The initial data refinement showed promising preliminary results, with an excellent electron density for retinal (Figure **I**-28a). The next step was to investigate the changes occurring in the crystals during the photoisomerization using the pump-probe method. The 370 nm laser with a ns pulse width was chosen as the irradiation source (toward PSB formation-*cis* iminium). An LED (570 nm) illumination source was used to reset the crystals to the SB state. In this experiment, performed by Prof. Geiger and Dr. Srajer, several data points were collected (Dark, 50 μs, 100 μs, 0.5 ms, and 1ms). As shown in Figure **I**-28, the time- dependent difference map was obtained by subtraction: $\Delta F(hkl,t) = |\boldsymbol{F}(hkl, t)| - |\boldsymbol{F}(hkl, dark)|$. Additionally, the PDB model from the dark state was represented at each time point. Essentially, the presence of a positive and negative map in the vicinity of each atom demonstrates the movement of those atoms at each time point relative to the dark-state PDB model. As shown in Figure **I**-28, the change in the polyene is not significant, which leads to two hypotheses: first, the time scale we are looking at is not appropriate, and we should study the shorter time points where larger changes might be visible; or second, we simply need to increase the exposure time of laser irradiation, since the quantum yield of photoisomerization is not quite high enough in the crystal. These are the studies that will be performed in the future. However, these preliminary time-resolved experimental results, both in solution and crystal, demonstrate the potential of our engineered model system for studying retinal photoisomerization in a simpler system than membrane proteins. The advantage of this model system vs. membrane proteins is that our proteins can be easily expressed, crystallized, and studied using simple or time-resolved techniques.

I-6. Conclusion and future direction

In conclusion, an artificially-designed protein that recapitulates the photocycle of microbial rhodopsins, specifically and quantitatively photoisomerizing all-*trans*-retinal to 13-*cis*-retinal in a well-diffracting crystal, has been created. Because of its slow rate of isomerization, our engineered rhodopsin mimic can be exploited using relatively simple time-resolved spectroscopic and crystallographic methods. Our mimic also introduced an ideal system for time-resolved crystallography.

Using this system, a mechanism for protein photoswitching that involves no net conformational change of the chromophore was identified at atomic resolution. This observation, which would be difficult to observe using other techniques, opens the door to a novel strategy for the design and study of photoactive and photoswitchable protein systems. Our ultimate goal is to employ our system for time-resolved x-ray and FT-IR experiments to elucidate the intermediates of the isomerization and compare them with natural microbial rhodopsin. Another exciting avenue would be to explore the development of an isomerase, where the chromophore is hydrolyzed and released from the protein after isomerization. Exploring other isomerization events, such as mimicking eukaryotic rhodopsin isomerization where the 11-*cis* retinal isomerizes to all-*trans* retinal, using our model system, is another potential avenue of future investigation. Finally, if we can alter the conformation of the protein during photoisomerization, we can generate a new optogenetic tool in this system that can be exploited for a variety of protein-engineering applications.

I-7. Experimental [48]

I-7-1. Materials and methods

All-*trans* retinal, 13-*cis* retinal, and other chemicals used in this study were purchased from Sigma Aldrich unless otherwise specified. A BioLogic DuoFlow (BioRad) chromatography instrument was used for the protein purifications. Source Q and Fast Q anion exchange resins were purchased from GE Health Care. An Ultrasonic Homogenizer from Biologics, Inc. was used for sonication of the samples.

Table I-3. PCR protocol.

PCR program		
1x	94 °C	3 min
	94 °C	20 sec
20x	temperature 3-5 °C lower than primer melting temperature	50 sec
	72° C	4 min 30 sec
1x	72 °C	10 min
1x	10 ° C	5 min

Table I-4. PCR program.

PCR program		
1x	94 ^{o}C	3 min
	94 ^{o}C	20 sec
20x	temperature 3-5 ^{o}C lower than primer melting temperature	50 sec
	72^{o} C	4 min 30 sec (3 min 30 Sec for CRBPII)
1x	72 ^{o}C	10 min

I-7-2. List of primers:

The hCRABPII-pET17b plasmid previously described was employed for mutagenesis.[64]

L121W

Forward: 5'- GAACTGATCTGGACCATGACG-3'

Reverse: 5'-CGTCATGGTCCAGATCAGTTC-3'

L121E

Forward: 5'- GAACTGATCGAAACCATGACG-3'

Reverse: 5'-CGTCATGGTTTCGATCAGTTC-3'

L121Y

Forward: 5'- GAACTGATCTACACCATGACG-3'

Reverse: 5'- CGTCATGGTGTAGATCAGTTC-3'

L121Q

Forward: 5’- GAACTGATCCAGACCATGACG-3’

Reverse: 5’- CGT CAT GGT CTG GAT CAG TTC-3’

P39Y

Forward: 5'- GCAGCGTCCAAGTATGCAGTGG -3'

Reverse: 5'- CCACTGCTTGCTTGGACGCTGC - 3'

P39Q

Forward: 5’- GCAGCGTCCAAGCAAGCAGTGG-3’

Reverse: 5’- CCACTGCAAACTTGGACGCTGC-3’

A32Y

Forward: 5’-GAGGAAGATTTATGTGGCTGC-3’

Reverse: 5’-GCAGCCACATAAATCTTCCTC-3

T54V

Forward:5’-CTACATCAAAGTCTCCACCACCGTGCG-3’

Reverse: 5’- CGCACGGTGGTGGAGACTTTGATGTAG -3’

R111K

Forward: 5’-CCCAAGACCTCGTGGACCAAAGAACTGACCAACGATGGG-3’

Reverse: 5’-CCCATCGTTGGTCAGTTCTTTGGTCCACGAGGTCTTGGG-3’

R132Q:Y134F

Forward: 5’-GTTGTGTGCACCCAGGTCTTCGTCCG-3’

Reverse: 5’-CGGACGAAGACCTGGGTGCACACAAC-3’

R59Y

Forward: 5’-CCTCCACCACCGTGTACACCACAGAG-3’

Reverse: 5’- CTCTGTGGTGTACACGGTGGTGGAGG -3’

I-7-3. Transformation of PCR product

5 µL of PCR product was added to DH5α competent cells (50 µL) thawed on ice. The sample was kept 30 minutes in ice. After heat-shock at 42℃ for 45 seconds, 450 µL of LB was added to the sample. The sample and agar plate containing ampicillin (100 µg/mL) were incubated at 37℃ for 30 minutes. Then 100 µL of the solution was spread on the LB agar plate and incubated at 37℃ overnight. The PCR product was purified using the Promega DNA purification Kit.

I-7-4. Expression of hCRABPII variants

The gene was transformed into BL21 (DE3) pLysE *E. coli* competent cells. The same protocol mentioned for PCR was exploited, with the only difference being the addition of 100 µg/mL ampicillin and 28 µg/mL chloramphenicol as antibiotics. To inoculate 1 L of LB with ampicillin (100 µg/mL) and chloramphenicol (28 µg/mL), a single colony was used. After the cell culture was grown at 37℃ while shaking for 8 h, (0D~0.8-0.9) the overexpression was started by adding 1 mL of 1M IPTG solution into 1 L cell culture (overall concentration 1.0 mM of IPTG). The solution was shaken at 19℃ for 36 hours.[49, 53]

I-7-5. Protein isolation and purification

By centrifuging for 20 min at 5,000 rpm (4℃), the cells were collected. The supernatant was discarded and the cells were resuspended in 60 mL of Tris buffer (10 mM Tris.HCl, pH=8.0). Then ultrasonication was used to lyse the cells (Power 60%, 1 min x 3), and the sonicated mixture centrifuged at 16℃ (10,000 rpm, 20 min).

Purification by ion exchange chromatography using Q SepharoseTM, Fast Flow resin was next performed at 4 ℃. The protein was bound to the column by gravity flow, the column washed with 50 mL of 10 mM Tris. HCl buffer (pH=8.0). The bound protein was then eluted with 10 mM Tris. HCl, 150 mM NaCl, pH = 8.0, the pure fractions were desalted using EMD Millipore centriprep centrifugal units (cutoffs:10 kDa) with 10 mM Tris. HCl pH=8.0 buffer, three times, and loaded on a second anion exchange column (15Q, GE Health Sciences, BioLogic DuoFlow system). The Source Q program, as mentioned previously,[64] was used for the purification.

I-7-6. UV-vis measurements

UV-vis spectroscopy was performed using a Cary 300 and Cary 100 Bio Win UV UV-Vis spectrophotometer (Varian Instruments). The samples were prepared in a phosphate buffer purchased from Sigma Aldrich. The concentration of retinal was kept at a 0.5 equivalent of the protein concentration in all experiments. The PSB has absorptions corresponding to λ_{max} > 450 nm, while deprotonated imine peaks (SB) appear at λ_{max} ~370 nm.

The all-*trans* retinal extinction coefficient in ethanol is 48,000 $M^{-1}cm^{-1}$ at 368 nm, and the 13-*cis* retinal extinction coefficient in ethanol is 38,770 $M^{-1}cm^{-1}$ at 363 nm, as reported. [65] The protein extinction coefficient for different variants was calculated using the Gill and Hippel method.[66]

M1 ε_{280nm} = 22,400 $M^{-1}cm^{-1}$

M1-L121E ε_{280nm}= 22,283 $M^{-1}cm^{-1}$

M1-L121Q ε_{280nm}= 20,293 $M^{-1}cm^{-1}$

M1-L121E:P39Q ε_{280nm}= 16,776 $M^{-1}cm^{-1}$

M1-L121E:A32Y ε_{280nm}= 24,500 $M^{-1}cm^{-1}$

I-7-7. pK_a determination of mutants

The pK_a was determined by titration (absorbance vs. pH) and was fitted by the protocol previously described.[53] The following equation was used for fitting the data and calculating pK_a by employing KaleidaGraph:

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

I-7-8. Irradiation of protein/retinal complexes in solution

An Oriel Illuminator (Model 66142, Oriel Instruments) attached to a power supply (Model 668820, Oriel Instruments, 500 W Mercury (Xenon) lamp) was employed for the light irradiation of samples. A combination of two filters was utilized for all light irradiations. A glass filter (6 mm thickness) was utilized to filter UV light below 320 nm. The second filter for UV irradiations was a U-360 (UV) 2” square band-pass filter (center wavelength (CWL) = 360 nm, full width at half-maximum height (FWHM) = 45 nm, purchased from Edmund Optics). For the visible light green-light irradiations, a Y-50 2” square long-pass filter (cut-off position = 500 ± 6 nm, purchased from Edmund Optics) was used. The third filter was a 440 ± 20 nm UV-VIS Bandpass Filter purchased from Edmund Optics for blue light irradiation.[50]

I-7-9. Protein crystallization

The proteins were concentrated to ~20-25 mg/ml using EMD Millipore centriprep concentrators (cutoffs: 10 kDa), and ~4 equiv. of ligands (all-*trans* retinal or 13-*cis* retinal) were added in the dark. The vapor diffusion method was used for crystallization using 24 well plates in the dark (purchased from Hampton Research) with 1 ml solution in the reservoir. **M1**-L121E, **M1**-L121Q bound with all-*trans* and **M1**-L121E:A32Y bound with

13-*cis* retinal were crystallized under the following conditions: 20% PEG 3350 and 0.1 M DL-malic acid (titrated with NaOH to pH=6.0 and 6.5) at 4° C. For **M1**-L121W, **M1**-L121Y the crystals grew at 20% PEG 3350, 0.1 M sodium malonate (pH=6.0-6.5) 4° C. For **M1**-L121E:P39Q, the crystals grew at 20% PEG 3350, 0.1 M bis-tris propane (pH=6.0-6.5), 0.2 M sodium fluoride. All crystals appeared after 1-3 days and were then frozen using liquid nitrogen in a cryogenic buffer containing their mother liquor along with 20% glycerol in the dark. Diffraction data were collected at the Advanced Photon Source (APS) (Argonne National Laboratory IL) LS-CAT, (sector 21-ID-D,F, and G) using Eiger 9M, and MAR300 detectors, and 1.00Å wavelength radiation at 100K. The initial diffraction data was indexed, processed and scaled using the HKL2000 software package.[67] The structures were solved by molecular replacement using PHASER_ENREF_5[68] in PHENIX [69] and hCRABPII (PDB entry 4YFP) as a search model. The initial electron density map was produced using Phaser-MR in PHENIX. Model rebuilding, placement of water molecules, etc. were performed using COOT.[70] The structures were refined using the PHENIX program package. The chromophore was built using the COOT and PHENIX programs to generate restraints.

Table I-5. Crystallographic data collection and refinement statistics for various dark and light states of all-*trans* bound **M1**-L121E.

	M1-L121E bound with all-*trans* retinal in dark	**M1-L121E bound with all-*trans* retinal, 5 minutes irradiation with 399nm laser**	**M1-L121E bound with all-*trans* retinal, 30 seconds irradiation with 399 nm laser**	**M1-L121E bound with all-*trans* retinal, 30 seconds irradiation with 399 nm laser, 10 minutes incubation in dark**	**M1-L121E bound with all-*trans* retinal, 30 seconds irradiation with 399 nm laser, 25 minutes incubation in dark**
Space group	$P\ 3_12_1$	$P\ 3_12_1$	$P\ 3_12_1$	$P\ 3_12_1$	$P\ 3_12_1$
a(Å)	58.986	58.567	58.343	58.586	58.501
b(Å)	58.986	58.567	58.343	58.586	58.501
c (Å)	100.314	98.974	99.865	100.178	99.982
α(°)	90	90	90	90	90
β(°)	90	90	90	90	90
δ(°)	120	120	120	120	120
Molecules per Asymmetric Unit	1	1	1	1	1
Total reflection	405325	503530	777250	478475	463610
Unique Reflection	16473 (1618)	12474 (1214)	16010 (1552)	12273 (1205)	13681(1345)
Completeness (%)	99.92 (100.00)	99.92 (99.92)	99.48 (98.66)	99.83 (100.00)	99.54 (99.78)
Average I/σ	34.4 (2.15)	26 (1.87)	35.5 (3.3)	36.75 (3.80)	44.63 (3.3)
R_{merge}	0.05 (0.87)	0.078 (0.95)	0.047 (0.36)	0.067 (0.68)	0.065 (0.62)
Resolution (Å) (Last Shell)	35.79 - 1.90 (1.93 - 1.90)	45.14 - 2.07 (2.14 - 2.07)	28.0 -1.90 (1.97 - 1.90)	29.29-2.09 (2.16 - 2.09)	29.25 - 2.01 (2.01-2.08)
R_{work}/ R_{free}	19.50/22.74	19.31 /25.21	18.93/21.70	18.30/22.02	19.87/23.40
Bond Length (Å)	0.009	0.008	0.008	0.007	0.008
Bond Angle (°)	0.87	1.06	0.88	0.87	0.86
Average B factor	40.13	40.03	32.83	39.40	52.38
Number of water molecules	101	45	42	54	38
PDB ID	6MOP	6MQZ	6MQY	6MQW	6MQX

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

Table I-6. Crystallographic data collection and refinement statistics for 13-*cis* bound **M1**-L121E:A32Y, and all-*trans* bound **M1**-L121W, **M1**-L121Y, and **M1**-L121Q dark states.

	M1-L121E-A32Y bound with 13-*cis* retinal in dark	**M1-L121W bound with all-*trans* retinal in dark**	**M1-L121Y bound with all-*trans* retinal in dark**	**M1-L121Q bound with all-*trans* retinal in dark**
Space group	$P\ 3_12_1$	$P\ 3_12_1$	$P\ 3_12_1$	$P\ 3_12_1$
a(Å)	58.761	58.117	58.469	59.219
b(Å)	58.761	58.117	58.469	59.219
c (Å)	99.174	99.629	100.118	99.837
α(°)	90	90	90	90
β(°)	90	90	90	90
δ(°)	120	120	120	120
Molecules per Asymmetric Unit	1	1	1	1
Total reflection	556609	1093404	395603	552392
Unique Reflection	27796 (2743)	10327 (1019)	19270 (1890)	20945 (2022)
Completeness	99.77 (100)	98.99 (98.25)	99.94 (99.89)	99.84 (98.92)
Average I/σ	36 (2.18)	22.8 (1.6)	39.9 (3.6)	45 (2.4)
R_{merge}	0.065 (0.60)	0.077 (0.58)	0.044 (0.51)	0.055 (0.87)
Resolution (Å) (Last Shell)	35.51 - 1.58 (1.64 - 1.58)	27.90 - 2.20 (2.28 - 2.20)	35.60 - 1.79 (1.85 - 1.79)	35.77 - 1.75 (1.81 -1.75)
R_{work}/ R_{free} (%)	20.29/22.18	22.62/26.57	22.31/24.72	21.47/23.34
Bond Length (Å)	0.006	0.009	0.007	0.009
Bond Angle (°)	0.86	1.04	0.95	0.92
Average B factor	32.72	56.32	30.86	32.99
Number of water molecules	111	29	99	67
PDB ID	6MPK	6MOQ	6MOR	6MOV

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

Table I-7. Crystallographic data collection and refinement statistics for different light states of all-*trans* bound **M1**-L121Q.

	M1-L121Q bound with all-*trans* retinal, 5 minutes irradiation with laser (399 nm)	**M1-L121Q- bound with all-*trans* retinal, 10 minutes irradiation with laser (532 nm)**
Space group	$P\,3_12_1$	$P\,3_12_1$
a(Å)	59.075	59.238
b(Å)	59.075	59.238
c (Å)	99.814	99.393
α(°)	90	90
β(°)	90	90
δ(°)	120	120
Molecules per Asymmetric Unit	1	1
Total reflection	378799	545064
Unique Reflection	17237 (1682)	11843 (1167)
Completeness (%)	99.94 (100.00)	99.82 (98.90)
Average I/σ	40.6 (2.55)	19.3 (1.8)
R_{merge}	0.044 (0.734)	0.06 (0.51)
Resolution (Å) (Last Shell)	35.72 - 1.87 (1.937 - 1.87)	35.69 - 2.12 (2.20 - 2.12)
R_{work}/ R_{free} (%)	21.52/25.96	18.85/24.31
Bond Length (Å)	0.009	0.009
Bond Angle (°)	0.95	0.94
Average B factor	35.95	41.35
Number of water Molecules	50	43
PDB ID	6MQI	6MQJ

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

Table I-8. Crystallographic data collection and refinement statistics for dark and light states of all-*trans* bound **M1**-L121E:P39Q.

	M1:L121E :P39Q bound with all-*trans* in dark	**M1:L121E:P39Q bound with all-*trans* 5 minutes UV irradiation**
Space group	$P\ 3_12_1$	$P\ 3_12_1$
a(Å)	58.546	58.476
b(Å)	58.546	58.476
c (Å)	99.525	99.552
α(°)	90	90
β(°)	90	90
δ(°)	120	120
Molecules per Asymmetric Unit	1	1
Total reflection	258241	534615
Unique Reflection	10769 (1059)	6064 (570)
Completeness (%)	99.94 (99.81)	99.59 (97.43)
Average I/σ	14.5 (2.2)	11.25 (2)
R_{merge}	0.089 (0.64)	0.15 (0.89)
Resolution (Å) (Last Shell)	45.18 - 2.18 (2.26 - 2.18)	33.18 - 2.65 (2.74 - 2.65)
R_{work}/ R_{free}	20.35/25.01	19.12/27.34
Bond Length (Å)	0.009	0.010
Bond Angle (°)	0.94	1.01
Average B factor	31.76	32.46
Number of water molecules	25	12
PDB ID	6MOX	6MR0

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

REFERENCES

REFERENCES

1. Ernst, O. P.; Lodowski, D. T.; Elstner, M.; Hegemann, P.; Brown, L. S.; Kandori, H., Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chemical Reviews* **2014,** *114* (1), 126-163.

2. Kurihara, M.; Sudo, Y., Microbial rhodopsins: wide distribution, rich diversity and great potential. *Biophysics and Physicobiology* **2015,** *12*, 121-9.

3. Schotte, F.; Cho, H. S.; Kaila, V. R. I.; Kamikubo, H.; Dashdorj, N.; Henry, E. R.; Graber, T. J.; Henning, R.; Wulff, M.; Hummer, G.; Kataoka, M.; Anfinrud, P. A., Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography. *Proceedings of the National Academy of Sciences of the United States of America,* **2012,** *109* (47), 19256-19261.

4. Sudo, Y.; Ihara, K.; Kobayashi, S.; Suzuki, D.; Irieda, H.; Kikukawa, T.; Kandori, H.; Homma, M., A Microbial Rhodopsin with a Unique Retinal Composition Shows Both Sensory Rhodopsin II and Bacteriorhodopsin-like Properties. *Journal of Biological Chemistry* **2011,** *286* (8), 5967-5976.

5. Butt, H. J.; Fendler, K.; Bamberg, E.; Tittor, J.; Oesterhelt, D., Aspartic acids 96 and 85 play a central role in the function of bacteriorhodopsin as a proton pump. *The EMBO Journal* **1989,** *8* (6), 1657-1663.

6. Kandori, H., Ion-pumping microbial rhodopsins. *Frontiers in Molecular Biosciences* **2015,** *2*, 52.

7. Gordeliy, V. I.; Labahn, J.; Moukhametzianov, R.; Efremov, R.; Granzin, J.; Schlesinger, R.; Buldt, G.; Savopol, T.; Scheidig, A. J.; Klare, J. P.; Engelhard, M., Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* **2002,** *419* (6906), 484-487.

8. Govorunova, E. G.; Koppel, L. A., The road to optogenetics: Microbial rhodopsins. *Biochemistry-Moscow,* **2016,** *81* (9), 928-940.

9. Royant, A.; Nollert, P.; Edman, K.; Neutze, R.; Landau, E. M.; Pebay-Peyroula, E.; Navarro, J., X-ray structure of sensory rhodopsin II at 2.1-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America* **2001,** *98* (18), 10131-10136.

10. Kandori, H., Protein-Controlled Isomerization in Rhodopsins. In *Chemical Science of π-Electron Systems*, Akasaka, T.; Osuka, A.; Fukuzumi, S.; Kandori, H.; Aso, Y., Eds. Springer Japan: Tokyo, 2015; pp 695-713.

11. Hirai, T.; Subramaniam, S., Protein Conformational Changes in the Bacteriorhodopsin Photocycle: Comparison of Findings from Electron and X-Ray Crystallographic Analyses. *PLOS ONE* **2009,** *4* (6), e5769.

12. Nango, E.; Royant, A.; Kubo, M.; Nakane, T.; Wickstrand, C.; Kimura, T.; Tanaka, T.; Tono, K.; Song, C.; Tanaka, R.; Arima, T.; Yamashita, A.; Kobayashi, J.; Hosaka, T.; Mizohata, E.; Nogly, P.; Sugahara, M.; Nam, D.; Nomura, T.; Shimamura, T.; Im, D.; Fujiwara, T.; Yamanaka, Y.; Jeon, B.; Nishizawa, T.; Oda, K.; Fukuda, M.; Andersson, R.; Båth, P.; Dods, R.; Davidsson, J.; Matsuoka, S.; Kawatake, S.; Murata, M.; Nureki, O.; Owada, S.; Kameshima, T.; Hatsui, T.; Joti, Y.; Schertler, G.; Yabashi, M.; Bondar, A.-N.; Standfuss, J.; Neutze, R.; Iwata, S., A three-dimensional movie of structural changes in bacteriorhodopsin. *Science* **2016,** *354* (6319), 1552-1557.

13. Kato, H. E.; Zhang, F.; Yizhar, O.; Ramakrishnan, C.; Nishizawa, T.; Hirata, K.; Ito, J.; Aita, Y.; Tsukazaki, T.; Hayashi, S.; Hegemann, P.; Maturana, A. D.; Ishitani, R.; Deisseroth, K.; Nureki, O., Crystal structure of the channelrhodopsin light-gated cation channel. *Nature* **2012,** *482* (7385), 369-74.

14. Vogeley, L.; Sineshchekov, O. A.; Trivedi, V. D.; Sasaki, J.; Spudich, J. L.; Luecke, H., Anabaena Sensory Rhodopsin: A Photochromic Color Sensor at 2.0 Å. *Science* **2004,** *306* (5700), 1390-1393.

15. Maresca, J. A.; Keffer, J. L.; Miller, K. J., Biochemical Analysis of Microbial Rhodopsins. *Current Protocols in Microbiology* **2016,** *41*, 1F 4 1-1F 4 18.

16. Shibata, M.; Kandori, H., FTIR Studies of Internal Water Molecules in the Schiff Base Region of Bacteriorhodopsin. *Biochemistry* **2005,** *44* (20), 7406-7413.

17. Hatanaka, M.; Kashima, R.; Kandori, H.; Friedman, N.; Sheves, M.; Needleman, R.; Lanyi, J. K.; Maeda, A., Trp86 → Phe Replacement in Bacteriorhodopsin Affects a Water Molecule near Asp85 and Light Adaptation. *Biochemistry* **1997,** *36* (18), 5493-5498.

18. Otto, H.; Marti, T.; Holz, M.; Mogi, T.; Stern, L. J.; Engel, F.; Khorana, H. G.; Heyn, M. P., Substitution of amino acids Asp-85, Asp-212, and Arg-82 in bacteriorhodopsin affects the proton release phase of the pump and the pK_a of the Schiff base. *Proceedings of the National Academy of Sciences* **1990,** *87* (3), 1018-1022.

19. Nogly, P.; Weinert, T.; James, D.; Carbajo, S.; Ozerov, D.; Furrer, A.; Gashi, D.; Borin, V.; Skopintsev, P.; Jaeger, K.; Nass, K.; Båth, P.; Bosman, R.; Koglin, J.; Seaberg, M.; Lane, T.; Kekilli, D.; Brünle, S.; Tanaka, T.; Wu, W.; Milne, C.; White, T.; Barty, A.; Weierstall, U.; Panneels, V.; Nango, E.; Iwata, S.; Hunter, M.; Schapiro, I.; Schertler, G.; Neutze, R.; Standfuss, J., Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* **2018**, 361, eaat0094.

20. Palczewski, K., G., Protein-coupled receptor rhodopsin. *Annual review of biochemistry* **2006,** *75*, 743-67.

21. Zhou, X. E.; Melcher, K.; Xu, H. E., Structure and activation of rhodopsin. *Acta Pharmacologica Sinica* **2012,** *33* (3), 291-9.

22. 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. *Proceedings of the National Academy of Sciences of the Unitd States of America* **2009,** *106* (41), 17325-17330.

23. Tsukamoto, H.; Terakita, A., Diversity and functional properties of bistable pigments. *Photochemical & Photobiological Sciences* **2010,** *9* (11), 1435-1443.

24. Govorunova, E. G.; Sineshchekov, O. A.; Li, H.; Spudich, J. L., Microbial Rhodopsins: Diversity, Mechanisms, and Optogenetic Applications. *Annual Review of Biochemistry* **2017,** *86* (1), 845-872.

25. Bassolino, G.; Sovdat, T.; Soares Duarte, A.; Lim, J. M.; Schnedermann, C.; Liebel, M.; Odell, B.; Claridge, T. D. W.; Fletcher, S. P.; Kukura, P., Barrierless Photoisomerization of 11-cis Retinal Protonated Schiff Base in Solution. *Journal of the American Chemical Society* **2015,** *137* (39), 12434-12437.

26. Liu, R. S. H.; Asato, A. E., Photochemistry and synthesis of stereoisomers of vitamin-A. *Tetrahedron* **1984,** *40* (11), 1931-1969.

27. Schapiro, I., The origin of bond selectivity and excited-state reactivity in retinal analogues. *The Journal of Physical Chemistry A* **2016,** *120* (19), 3353-3365.

28. Ben-Nun, M.; Molnar, F.; Schulten, K.; Martínez, T. J., The role of intersection topography in bond selectivity of cis-trans photoisomerization. *Proceedings of the National Academy of Sciences of the United States of America,* **2002,** *99* (4), 1769-1773.

29. Jang, G. F.; Kuksa, V.; Filipek, S.; Bartl, F.; Ritter, E.; Gelb, M. H.; Hofmann, K. P.; Palczewski, K., Mechanism of rhodopsin activation as examined with ring-constrained retinal analogs and the crystal structure of the ground state protein. *The journal of Biological Chemistry* **2001,** *276* (28), 26148-53.

30. Mogi, T.; Marti, T.; Khorana, H. G., Structure-function studies on bacteriorhodopsin. IX. Substitutions of tryptophan residues affect protein-retinal interactions in bacteriorhodopsin. *The Journal of Biological Chemistry* **1989,** *264* (24), 14197-201.

31. Sharma, A. K.; Spudich, J. L.; Doolittle, W. F., Microbial rhodopsins: functional versatility and genetic mobility. *Trends in Microbiology* **2006,** *14* (11), 463-469.

32. Nango, E.; Royant, A.; Kubo, M.; Nakane, T.; Wickstrand, C.; Kimura, T.; Tanaka, T.; Tono, K.; Song, C.; Tanaka, R.; Arima, T.; Yamashita, A.; Kobayashi, J.; Hosaka, T.; Mizohata, E.; Nogly, P.; Sugahara, M.; Nam, D.; Nomura, T.; Shimamura, T.; Im, D.; Fujiwara, T.; Yamanaka, Y.; Jeon, B.; Nishizawa, T.; Oda, K.; Fukuda, M.; Andersson, R.; Bath, P.; Dods, R.; Davidsson, J.; Matsuoka, S.; Kawatake, S.; Murata, M.; Nureki, O.; Owada, S.; Kameshima, T.; Hatsui, T.; Joti, Y.; Schertler, G.; Yabashi, M.; Bondar, A. N.; Standfuss, J.; Neutze, R.; Iwata, S., A three-dimensional movie of structural changes in bacteriorhodopsin. *Science* **2016,** *354* (6319), 1552-1557.

33. Duñach, M.; Marti, T.; Khorana, H. G.; Rothschild, K. J., Uv-visible spectroscopy of bacteriorhodopsin mutants: substitution of Arg-82, Asp-85, Tyr-185, and Asp-212 results in abnormal light-dark adaptation. *Proceedings of the National Academy of Sciences of the United States of America* **1990,** *87* (24), 9873-9877.

34. Otto, H.; Marti, T.; Holz, M.; Mogi, T.; Stern, L. J.; Engel, F.; Khorana, H. G.; Heyn, M. P., Substitution of amino acids Asp-85, Asp-212, and Arg-82 in bacteriorhodopsin affects the proton release phase of the pump and the pK_a of the Schiff base. *Proceedings of the National Academy of Sciences of the United States of America* **1990,** *87* (3), 1018-1022.

35. Kandori, H., Role of internal water molecules in bacteriorhodopsin. *Biochimica et Biophysica Acta* **2000,** *1460* (1), 177-91.

36. Nogly, P.; Weinert, T.; James, D.; Carbajo, S.; Ozerov, D.; Furrer, A.; Gashi, D.; Borin, V.; Skopintsev, P.; Jaeger, K.; Nass, K.; Båth, P.; Bosman, R.; Koglin, J.; Seaberg, M.; Lane, T.; Kekilli, D.; Brünle, S.; Tanaka, T.; Wu, W.; Milne, C.; White, T.; Barty, A.; Weierstall, U.; Panneels, V.; Nango, E.; Iwata, S.; Hunter, M.; Schapiro, I.; Schertler, G.; Neutze, R.; Standfuss, J., Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* **2018**, eaat0094.

37. Govorunova, E. G.; Sineshchekov, O. A.; Janz, R.; Liu, X.; Spudich, J. L., Natural light-gated anion channels: A family of microbial rhodopsins for advanced optogenetics. *Science* **2015,** *349* (6248), 647-650.

38. Kouyama, T.; Nishikawa, T.; Tokuhisa, T.; Okumura, H., Crystal structure of the L intermediate of bacteriorhodopsin: evidence for vertical translocation of a water molecule during the proton pumping cycle. *Journal of Molecular Biology* **2004,** *335* (2), 531-546.

39. W Stoeckenius, a.; Bogomolni, R. A., Bacteriorhodopsin and related pigments of halobacteria. *Annual Review of Biochemistry* **1982,** *51* (1), 587-616.

40. Wagner, N. L.; Greco, J. A.; Ranaghan, M. J.; Birge, R. R., Directed evolution of bacteriorhodopsin for applications in bioelectronics. *Journal of the Royal Society, Interface* **2013,** *10* (84), 20130197.

41. Schobert, B.; Cupp-Vickery, J.; Hornak, V.; Smith, S. O.; Lanyi, J. K., crystallographic structure of the K intermediate of bacteriorhodopsin: Conservation of free

energy after photoisomerization of the retinal. *Journal of Molecular Biology* **2002,** *321* (4), 715-726.

42. Sass, H. J.; Buldt, G.; Gessenich, R.; Hehn, D.; Neff, D.; Schlesinger, R.; Berendzen, J.; Ormos, P., Structural alterations for proton translocation in the M state of wild-type bacteriorhodopsin. *Nature* **2000,** *406* (6796), 649-53.

43. Fodor, S. P. A.; Ames, J. B.; Gebhard, R.; Van den Berg, E. M. M.; Stoeckenius, W.; Lugtenburg, J.; Mathies, R. A., Chromophore structure in bacteriorhodopsin's N intermediate: implications for the proton-pumping mechanism. *Biochemistry* **1988,** *27* (18), 7097-7101.

44. Wang, T.; Sessions, A. O.; Lunde, C. S.; Rouhani, S.; Glaeser, R. M.; Duan, Y.; Facciotti, M. T., Deprotonation of D96 in bacteriorhodopsin opens the proton uptake pathway. *Structure* **2013,** *21* (2), 290-7.

45. Subramaniam, S.; Henderson, R., Crystallographic analysis of protein conformational changes in the bacteriorhodopsin photocycle. *Biochimica et biophysica acta (BBA) - Bioenergetics* **2000,** *1460* (1), 157-165.

46. Luecke, H.; Schobert, B.; Richter, H.-T.; Cartailler, J.-P.; Lanyi, J. K., Structure of bacteriorhodopsin at 1.55 Å resolution, Edited by D. C. Rees. *Journal of molecular biology* **1999,** *291* (4), 899-911.

47. Wickstrand, C.; Dods, R.; Royant, A.; Neutze, R., Bacteriorhodopsin: Would the real structural intermediates please stand up? *Biochimica et Biophysica Acta* **2015,** *1850* (3), 536-53.

48. Ghanbarpour, A.; Nairat, M.; Nosrati, M.; Santos, E. M.; Vasileiou, C.; Dantus, M.; Borhan, B.; Geiger, J. H., Mimicking Microbial Rhodopsin Isomerization in a Single Crystal. *Journal of the American Chemical Society* **2019,** *141* (4), 1735-1741.

49. Nosrati, M.; Berbasova, T.; Vasileiou, C.; Borhan, B.; Geiger, J. H., A Photoisomerizing Rhodopsin Mimic Observed at Atomic Resolution. *Journal of the American Chemical Society* **2016,** *138* (28), 8802-8808.

50. Berbasova, T.; Santos, E. M.; Nosrati, M.; Vasileiou, C.; Geiger, J. H.; Borhan, B., Light-Activated Reversible Imine Isomerization: Towards a Photochromic Protein Switch. *Chembiochem* **2016,** *17* (5), 407-414.

51. Yapici, I.; Lee, K. S. S.; Berbasova, T.; Nosrati, M.; Jia, X.; Vasileiou, C.; Wang, W.; Santos, E. M.; Geiger, J. H.; Borhan, B., “Turn-On” Protein Fluorescence: In Situ Formation of Cyanine Dyes. *Journal of the American Chemical Society* **2015,** *137* (3), 1073-1080.

52. Wang, W.; Geiger, J. H.; Borhan, B., The photochemical determinants of color vision. *Bioessays* **2014,** *36* (1), 65-74.

53. Berbasova, T.; Nosrati, M.; Vasileiou, C.; Wang, W.; Lee, K. S. S.; Yapici, I.; Geiger, J. H.; Borhan, B., Rational Design of a Colorimetric pH Sensor from a Soluble Retinoic Acid Chaperone. *Journal of the American Chemical Society* **2013,** *135* (43), 16111-16119.

54. Lee, K. S. S.; Berbasova, T.; Vasileiou, C.; Jia, X.; Wang, W.; Choi, Y.; Nossoni, F.; Geiger, J. H.; Borhan, B., Probing Wavelength Regulation with an Engineered Rhodopsin Mimic and a C15-Retinal Analogue. *Chempluschem* **2012,** *77* (4), 273-276.

55. Vasileiou, C.; Lee, K. S. S.; Crist, R. M.; Vaezeslami, S.; Goins, S. M.; Geiger, J. H.; Borhan, B., Dissection of the critical binding determinants of cellular retinoic acid binding protein II by mutagenesis and fluorescence binding assay. *Proteins: Structure, Function, and Bioinformatics* **2009,** *76* (2), 281-290.

56. Vaezeslami, S.; Mathes, E.; Vasileiou, C.; Borhan, B.; Geiger, J. H., The Structure of Apo-wild-type Cellular Retinoic Acid Binding Protein II at 1.4 Å and its Relationship to Ligand Binding and Nuclear Translocation. *Journal of Molecular Biology* **2006,** *363* (3), 687-701.

57. Crist, R. M.; Vasileiou, C.; Rabago-Smith, M.; Geiger, J. H.; Borhan, B., Engineering a Rhodopsin Protein Mimic. *Journal of the American Chemical Society,* **2006,** *128* (14), 4522-4523.

58. Assar, Z.; Nossoni, Z.; Wang, W.; Santos, Elizabeth M.; Kramer, K.; McCornack, C.; Vasileiou, C.; Borhan, B.; Geiger, James H., Domain-Swapped Dimers of Intracellular Lipid-Binding Proteins: Evidence for Ordered Folding Intermediates. *Structure*, **2016**, *24, 1590-1598*.

59. Wang, W.; Nossoni, Z.; Berbasova, T.; Watson, C. T.; Yapici, I., Tuning the electronic absorption of protein-embedded all-trans-retinal. *Science (American Association for the Advancement of Science) 338* (6112), 1340-1343.

60. Luecke, H.; Schobert, B.; Cartailler, J.-P.; Richter, H.-T.; Rosengarth, A.; Needleman, R.; Lanyi, J. K., Coupling photoisomerization of retinal to directional transport in bacteriorhodopsin11Edited by D. C. Rees. *Journal of Molecular Biology* **2000,** *300* (5), 1237-1255.

61. Bolze, C. S.; Helbling, R. E.; Owen, R. L.; Pearson, A. R.; Pompidor, G.; Dworkowski, F.; Fuchs, M. R.; Furrer, J.; Golczak, M.; Palczewski, K.; Cascella, M.; Stocker, A., Human Cellular Retinaldehyde-Binding Protein Has Secondary Thermal 9-cis-Retinal Isomerase Activity. *Journal of the American Chemical Society* **2014,** *136* (1), 137-146.

62. Rothschild, K. J.; Gray, D.; Mogi, T.; Marti, T.; Braiman, M. S.; Stern, L. J.; Khorana, H. G., Vibrational spectroscopy of bacteriorhodopsin mutants: Chromophore isomerization perturbs tryptophan-86. *Biochemistry* **1989,** *28* (17), 7052-7059.

63. Needleman, R.; Chang, M.; Ni, B.; Váró, G.; Fornés, J.; White, S. H.; Lanyi, J. K., Properties of Asp212----Asn bacteriorhodopsin suggest that Asp212 and Asp85 both participate in a counterion and proton acceptor complex near the Schiff base. *Journal of Biological Chemistry* **1991,** *266* (18), 11478-11484.

64. Nosrati, M.; Berbasova, T.; Vasileiou, C.; Borhan, B.; Geiger, J. H., A Photoisomerizing Rhodopsin Mimic Observed at Atomic Resolution. *Journal of the American Chemical Society* **2016,** *138* (28), 8802-8.

65. Maeda, T.; Van Hooser, J. P.; Driessen, C. A.; Filipek, S.; Janssen, J. J.; Palczewski, K., Evaluation of the role of the retinal G protein-coupled receptor (RGR) in the vertebrate retina in vivo. *Journal of Neurochemistry* **2003,** *85* (4), 944-56.

66. Gill, S. C.; von Hippel, P. H., Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry* **1989,** *182* (2), 319-26.

67. Otwinowski, Z.; Minor, W., [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **1997,** *276*, 307-326.

68. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D* **1997,** *53* (3), 240-255.

69. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D* **2010,** *66* (2), 213-221.

70. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* **2010,** *66*, 486-501.

CHAPTER II: CREATING A NEW CLASS OF PROTEIN CONFORMATIONAL SWITCHES

II-1. Protein conformational switches

Understanding the concept of a protein conformational switch is woven tightly with appreciating the meaning of allostery, "*the second secret of life*".[1] Classically, allostery is defined as transferring the effect of binding on one site to a remote site from where the binding ensues, accompanied by a concomitant change in the function.[1-2]

Protein conformational switches can be broadly defined as a family of allosteric proteins that undergo structural rearrangement upon detecting input signals, such as small molecules, proteins, light, hormones, changes in pH, temperature, the reduction potential of the environment, etc.[3] Protein conformational switches are ubiquitous in biology (Figure II-1) and play essential roles in almost all crucial biological processes including translation, transcription, signaling, cellular transport, enzymatic activity and

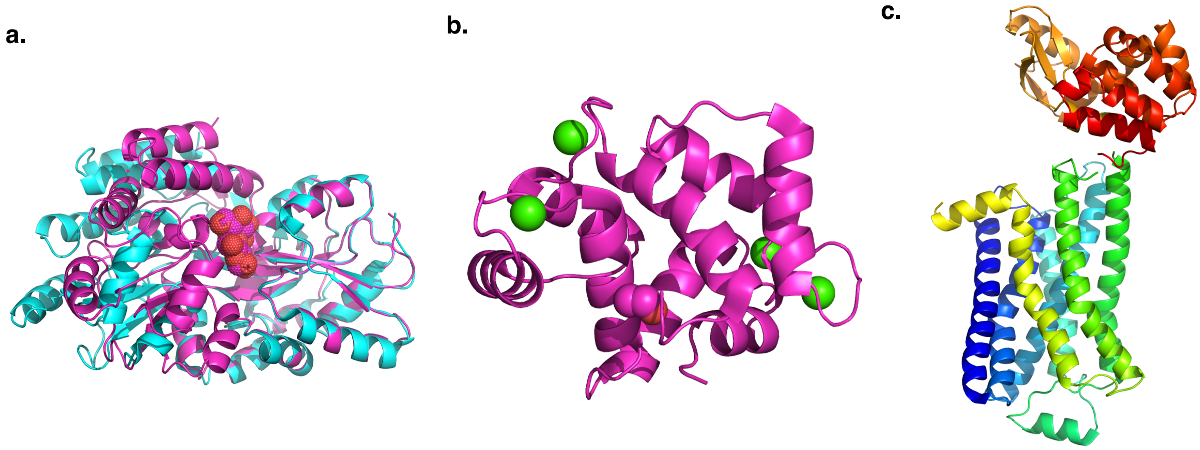


Figure II-1. Natural protein conformational switches. a. The overlay of apo (cyan) and holo (pink) maltose binding protein. b. The crystal structure of calcium-bound conformation of calmodulin. c. The GPCR transmembrane crystal structure.

others.[3-8] This means the presence of these proteins is vital for organisms to respond properly to the transitional environment in which they live.[1]

The ability to control the conformational readouts of a system makes these proteins highly interesting for various protein-engineering applications such as sensing and functional regulation.[7, 9-19] For example, these proteins have been used in the field of synthetic circuits where the arrangement of highly tunable and well-characterized modules can be managed to execute different complicated programs such as creating cellular memory, conducting cellular behavior, and converting analog to digital, wherein the continuous chemical gradient is converted into a digital output. [20-21]

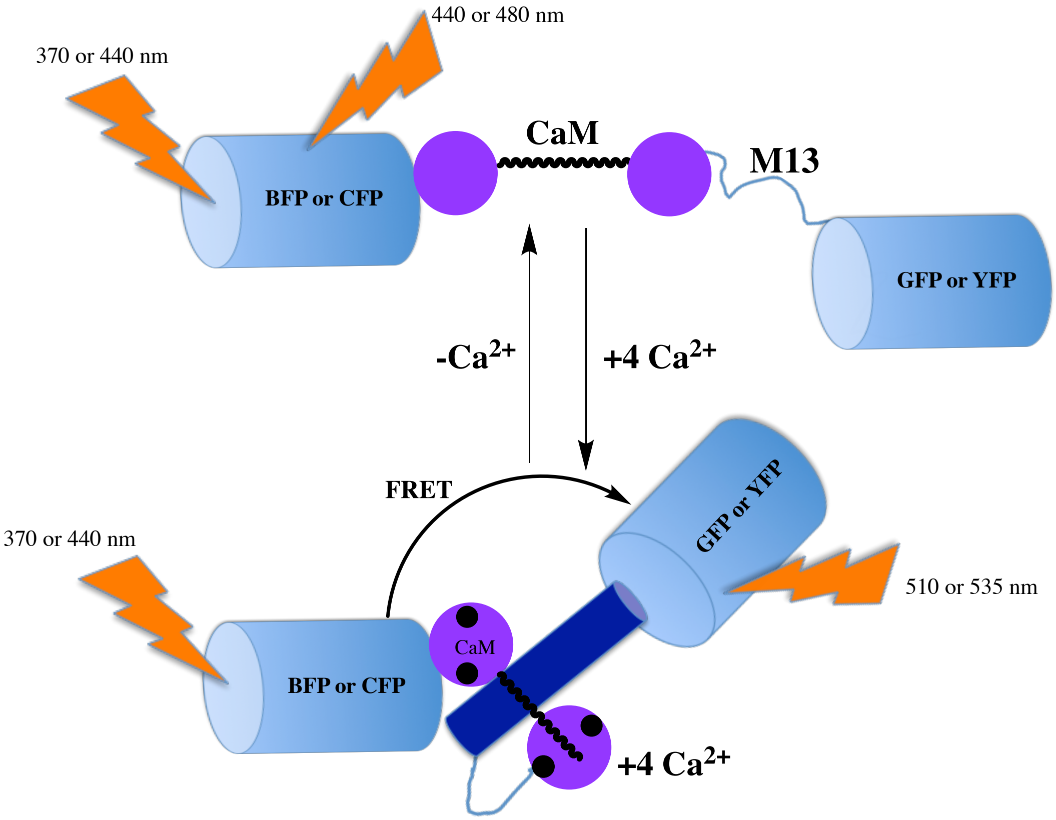


Figure II-2. Schematic picture demonstrating the FRET between GFPs can measure Ca^{2+}. The GFPs are drawn as cylinders. Top: calmodulin (CaM) without Ca^{2+} in the elongated conformation, disordered unbound M13 (calcium binding peptide) Bottom: The Ca^{2+}-bound calmodulin in the closed conformation and Ca^{2+}/M13 complex in the structured conformation.

Furthermore, these proteins have been employed as a sensor to identify and respond to certain signals. For instance, in a research study by Miyawaki *et al* [22] GFP and YFP were linked to the c- and n- termini of calmodulin. Upon sensing the calcium, calmodulin undergoes the conformational change leading to Förster resonance energy transfer (FRET), a distance-dependent fluorescence (Figure II-2). Therefore, the calcium concentration inside the cell can be a measurable readout throughout the change in FRET. The developments of these types of protein conformational switches leads to significant progress in developing novel calcium sensors, featuring the low-concentration detection of calcium release during the potential firing in neurons.[9, 23]

Another application of this strategy is to create novel affinity clamps by connecting two proteins through a flexible linker, which generates a new complex with a novel

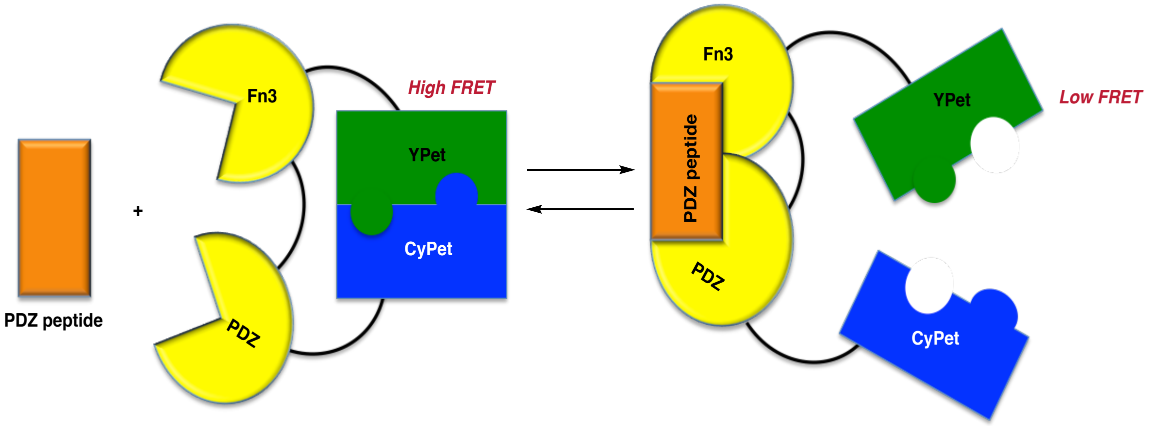


Figure II-3. A circularly-permutated PDZ domain is linked to a fibronectin type III (FN3)-based enhancer domain that detects the ligand-bound form of the PDZ domain. A significant conformational change ensues upon PDZ ligand binding with the concomitant formation of a sandwich complex that can be measured through the FRET between the two fluorescent proteins YPet and CyPet.

conformation upon sensing a ligand of interest. In this case, the signal can be monitored through intermolecular FRET activated through a large conformational change occurring from an open state to a closed state. Connecting a circularly-permutated PDZ to fibronectin through a flexible linker that detects the ligand-bound form of the PDZ domain is an example of this approach (Figure II-3).[24] Developing a maltose sensor by inserting the circularly-permutated GFP inside the domain of the maltose binding protein is another example of this strategy being used. In this case, upon sensing the maltose, a large protein conformational change eventuates, pulling on the circularly-permutated GFP and changing the fluorescence activity (Figure II-4).[3, 25]

In addition to a large conformational change occurring from well-structured protein domains, there are other examples where intrinsically disordered protein fragments have been employed. The sensing of a specific ligand causes the partial unfolding/folding states in the intrinsically disordered protein accompanied by a large change in the conformation, which is detectable using fluorescent proteins.[26-27] Synthetic protein conformational switches can be designed through alternative frame folding. [28-29] The ligand-binding site of the protein is replicated and linked to the opposite side of the protein to create a continuous polypeptide where the ligand binding causes reequilibration between native and circularly-permutated conformations. Eliciting a thrombin sensor that,

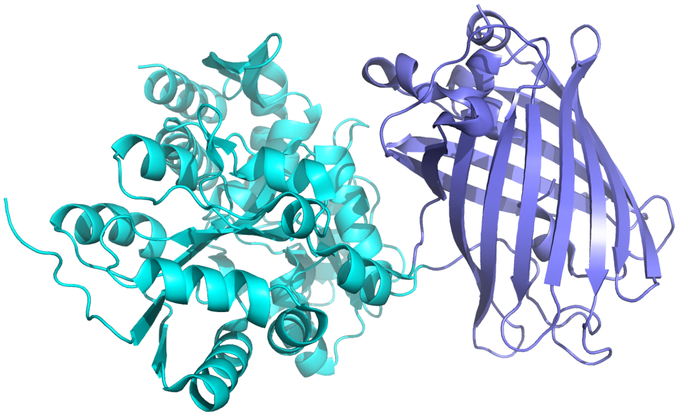


Figure II-4. Crystal structure of circularly-permutated GFP (purple) inserted inside the maltose-binding protein (cyan). PDB code: 3OSQ.

upon activation by light, turns the fluorescent output from green to yellow, is a successful example of this design approach.[30]

Protein conformational switches have been utilized as functional regulators to control and manage different cellular events. In most cases, the protein of interest is inserted inside the domain of the protein conformational switch, and the change in the conformation of the protein conformational switch leads to modifying the protein of interest's activity. This approach has been exercised extensively in eliciting enzyme-based protein switches with the ability to enlarge the signal transduction inside the cell, where the ability and compatibility of fluorescent-based sensors to identify and respond to certain cellular events are restrained.[3, 15, 31]

The creation of maltose-based β-lactamase with a large number of response times,[30] Figure II-5, protease-based proximity sensors of protein-protein interactions,[32-33] and allosteric protease receptors [34] are some examples where the domain-insertion strategy has been practiced.

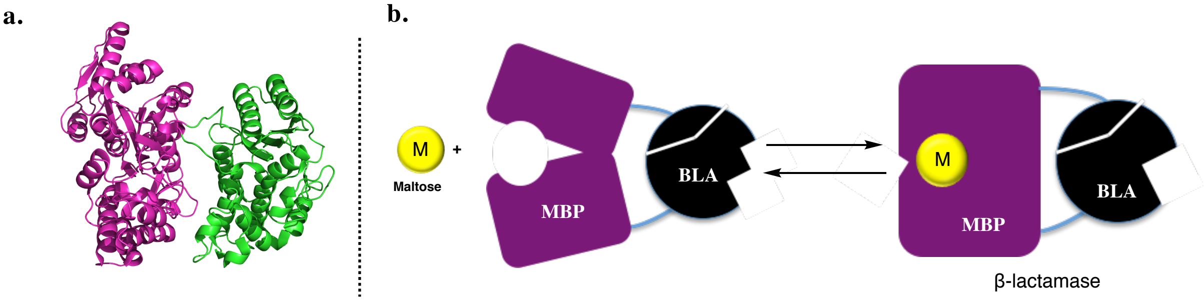


Figure II-5. Inserting maltose-binding protein (MBP) into a circularly permutated version of β-lactamase (BLA) giving rise to creation of a maltose sensor which regulates the activity of β-lactamase (BLA) as a function of maltose binding.

The idea of controlling function by light also led to creating new light-sensitive protein switches enabling spatio and temporal resolution in live cell imaging and fast response times in optogenetic applications.[18, 35-36] For the most part, this strategy has been performed using LOV domains [37] and BLUF domains where the blue light drives the flavin-based chromophore to form a covalent bond with an active cysteine inside the protein (Figure II-6) causing the unfolding of the c-terminal helix.[38] These proteins have been applied to create new classes of optogenetic molecular tools to turn on transcription factors and regulate neural cell circuitry. Wang et al demonstrated the hallmark of these studies, where a light-and-calcium-gated transcription factor (TF) system called FLARE (Fast Light and Activity-Regulated Expression) has been developed by reengineering the LOV domain, TEV protease, and calmodulin binding peptide. In this system, the TF, which is located in the plasma membrane, cannot perform the gene transcription of a reporter gene situated in the nucleus.[36] However, upon blue light treatment and in the presence of high calcium concentration, the TF is cleaved from the membrane and gets translocated to the nucleolus where it leads to the transcription of a reporter gene. This

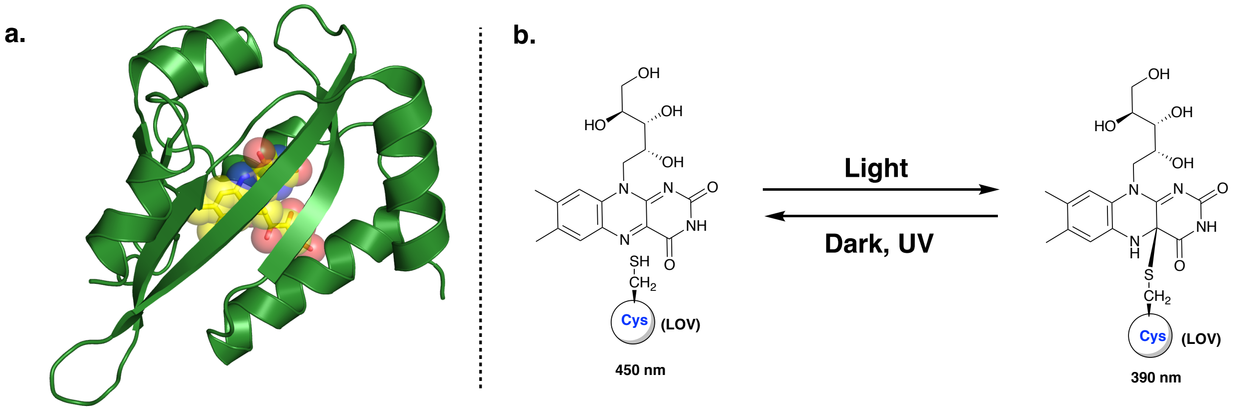


Figure II-6. a. The crystal structure of LOV domain in dark. PDB code: 3T50. b. LOV domain photocycle is detected by a ground state oxidize flavin generating a flavin-cysteine adduct upon blue-light irradiation. The adduct reset to the ground state through UV irradiation or dark adaptation.

strategy provides a powerful tool for optogenetics, where rhodopsin proteins [39] and azobenzene-based photoswitches are the only other alternatives.[40]

As mentioned above, natural protein conformational switches have been amply operated to create new molecular tools in synthetic biology; [3, 9, 11, 15, 29-31] however, there are few examples where a new protein conformational switch has been developed. Producing an allosteric site in an antibody using a computational- approach [41] and the de novo design of an allosteric metalloprotein through the mechanical strain of disulfide bonds[42] are some of the current efforts towards creating new allosteric behavior. So far, most attempts to create allosteric regulation centered on biophysical models comparing the two states — the *presence* and *absence* of the input signals — both concentrating on understanding the thermodynamic impetus of this phenomena. There is a paucity of information for grasping the molecular basis of allostery. Essentially, the connectivity in the protein scaffold leading to allosteric responses on a molecular level is poorly understood.[43-44] The goal of this chapter is to create allostery and elucidate the mechanism of that through atomic resolution crystal structures and rational mutagenesis using the domain-swapped (DS) dimer of human cellular retinol binding protein II (hCRBPII) as an engineering template. This study accentuates the design elements of contriving novel allosteric proteins based on high-resolution structural interrogation and evaluation in *a protein scaffold that has not been evolved to work as such.*

II-2. Using human cellular retinol binding protein II as an engineering scaffold

Our strategy to create a novel protein conformational switch is focused on using small cytosolic proteins belonging to intracellular Lipid Binding Protein (iLBP).[45,46] This protein family, as mentioned in the previous section, is responsible for trafficking hydrophobic ligands inside the cells. hCRBPII is responsible for carrying retinal and retinol within the cell.[47] These proteins structurally contain ten β-sheets and two α- helices embellishing the ligand binding site (Figure II-7). Very recently, Assar *et al* [48] identified several key mutations in this protein that give rise to a substantial amount of domain-swapped dimer. Furthermore, they elucidated the mechanism giving rise to DS dimer formation. A thorough definition and examples of domain-swapping are covered in Chapter III. Generally, domain swapping can be defined as a mechanism where two identical protein segments exchange part of their structures to generate dimer or higher-order oligomers. In our previous work,[48] we elaborated that a single mutation in hCRBPII leads to the breakage of the hydrogen bond between Y60 and E72, which then leads to destabilizing the closed monomer formation, increasing the lifetime of open-monomer intermediates, thereby increasing the population of the DS dimer of hCRBPII. Dr .Zahra Assar further confirmed that the WT hCRBPII can lead to negligible dimer formation, and fortuitously, she obtained the crystal structure of this protein as well.[48] Other than hCRBPII, Fatty Acid Binding Protein 5, another protein in this family that is a carrier for long-chain fatty acids and other endocannabinoids, has been recognized to generate a

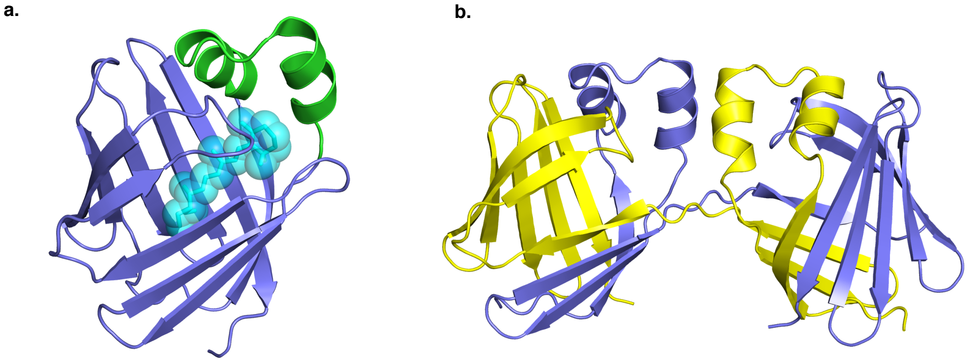


Figure II-7. a. The monomeric form of hCRBPII bound with retinol (shown with cyan). The two alfa-helices decorating the ligand binding site were highlighted with green. (PDB code: 4QZT) b. The domain-swapped dimer of hCRBPII. Each chain is color coded (PDB code: 4ZH9).

domain-swapped dimer.[49] In this particular case, the crystal structures revealed different protein conformations upon binding with two different ligands.[49]

Table II-1. Summary of apo and holo domain-swapped dimer crystal structures for several hCRBPII mutants.

Mutants	**Holo/ Apo**	**Space group**	**No. of Molecules**	**Symm./ Asymm.**
Q108K:T51D	Apo	*P*21212	1	Symmetric
Q108K:K40L: T51F	Apo	*P*21212	1	Symmetric
Q108K:K40L: T51W	Apo	*P*21212	1	Symmetric
Q108K:T51D:A28H	Apo	*P*21212	1	Symmetric
Q108K:T51D*	Holo	*P*212121	12	Asymmetric
Q108K:K40L: T51W	Holo	*P*21	12	Asymmetric
Q108K:K40L: T51F	Holo	*P*21	12	Asymmetric
Q108K:K40L:T51F:Y60A	Holo	*P*21212	1	Symmetric
WT-hCRBPII	Holo	*P*212121	12	Asymmetric
Q108K:T51D:A28C	Apo	*P*212121	2	Asymmetric
Q108K:T51D:A28C	Holo	*P*212121	4	Asymmetric
Q108K:T51D:A28H	Holo	*P*212121	12	Asymmetric
Q108K:K40L:T51H	Holo	*P*212121	12	Asymmetric
Q108K:K40L:T51N	Holo	*P*212121	12	Asymmetric

*The crystal structure of the holo Q108K:T51D domain-swapped dimer was obtained by Dr. Nosrati.

It has been shown that domain-swapping can sometimes alter the original functions of the protein by increasing the concentration of active/binding sites, generating binding sites in the interface of swapped units, and creating the possibility of allosteric regulation.[50] Having a large number of hCRBPII variants on hand, we decided to explore the possibility of creating allostery in hCRBPII mutants.

II-2-1. Crystallization of apo and holo retinal-bound variants of hCRBPII

In our previous work [48] two positions, Tyr60 residing in the hinge-loop region, and Glu72 which directly interacts with Tyr60, were found to promote DS dimerization when mutated. Two structural forms, symmetric and asymmetric DS dimers, of hCRBPII, were

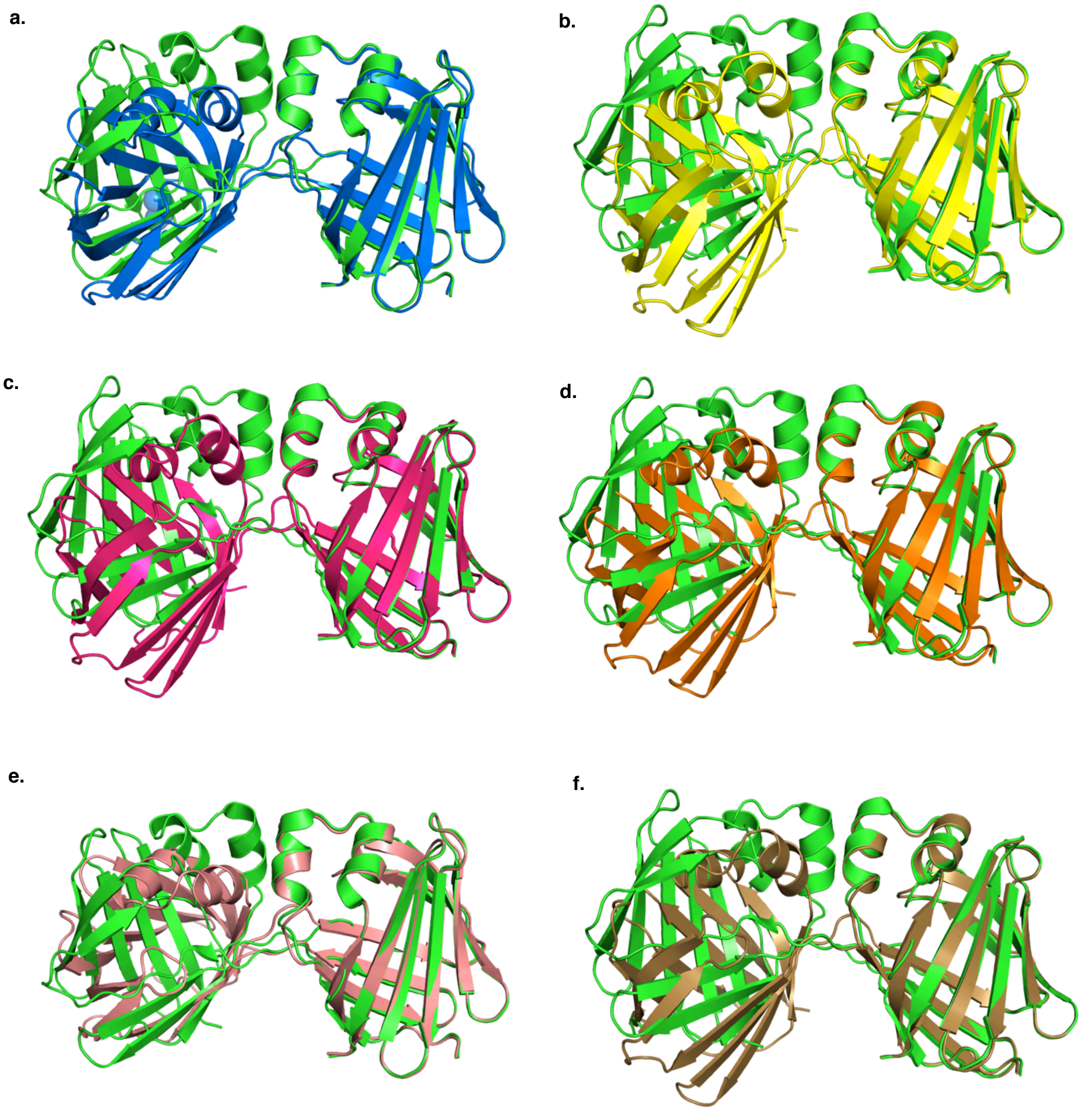


Figure II-8. Overlay of WT-hCRBPII (holo) DS dimer chains H and D shown by dark blue and WT-hCRBPII (apo) (PDB code: 4ZH9) shown by green. b. Overlay of WT-hCRBPII (holo) chains F and J shown by yellow and WT-hCRBPII (apo) shown by green. c. Overlay of WT-hCRBPII (holo) chains B and L shown by red and WT-hCRBPII (apo) shown by green. d. Overlay of WT-hCRBPII (holo) chains E and L shown by orange and WT-hCRBPII (apo) shown by green. e. Overlay of WT-hCRBPII (holo) chains A and K shown by pink and WT-hCRBPII (apo) shown by green. f. Overlay of WT-hCRBPII (holo) chains E and I is shown by brown and WT-hCRBPII (apo) shown by green.

observed, depending on the amino acid at position 60. While the wild-type, Y60L and E72A variants produced almost identical symmetric dimers, both the Q108K:K40L:Y60W and Y60W variants produced almost identical asymmetric dimers. Disappointingly however, we were unable to obtain any ligand-bound structures of these mutants. Dr. Nosrati and Dr. Santos identified additional mutational “hot spots” outside the hinge region that predominantly fold as domain-swapped dimers. We have now crystallized and determined the structures of five retinylidene PSB-bound variants, all of which produce mostly DS dimers, along with four of their apo counterparts. Furthermore, the crystal structure of the retinal-bound WT hCRBPII holo dimer has been determined (Figure II-8). The apo structures demonstrated a symmetric dimer as previously described [48] (Figure II-9a), though some differed in space group and crystal packing (Table 1). Together, the data indicates that the symmetric fold is the most common DS dimer structure. These results further indicate that the domain-swapped dimer has a dominant and robust structure in the absence of any perturbation, regardless of crystal packing.

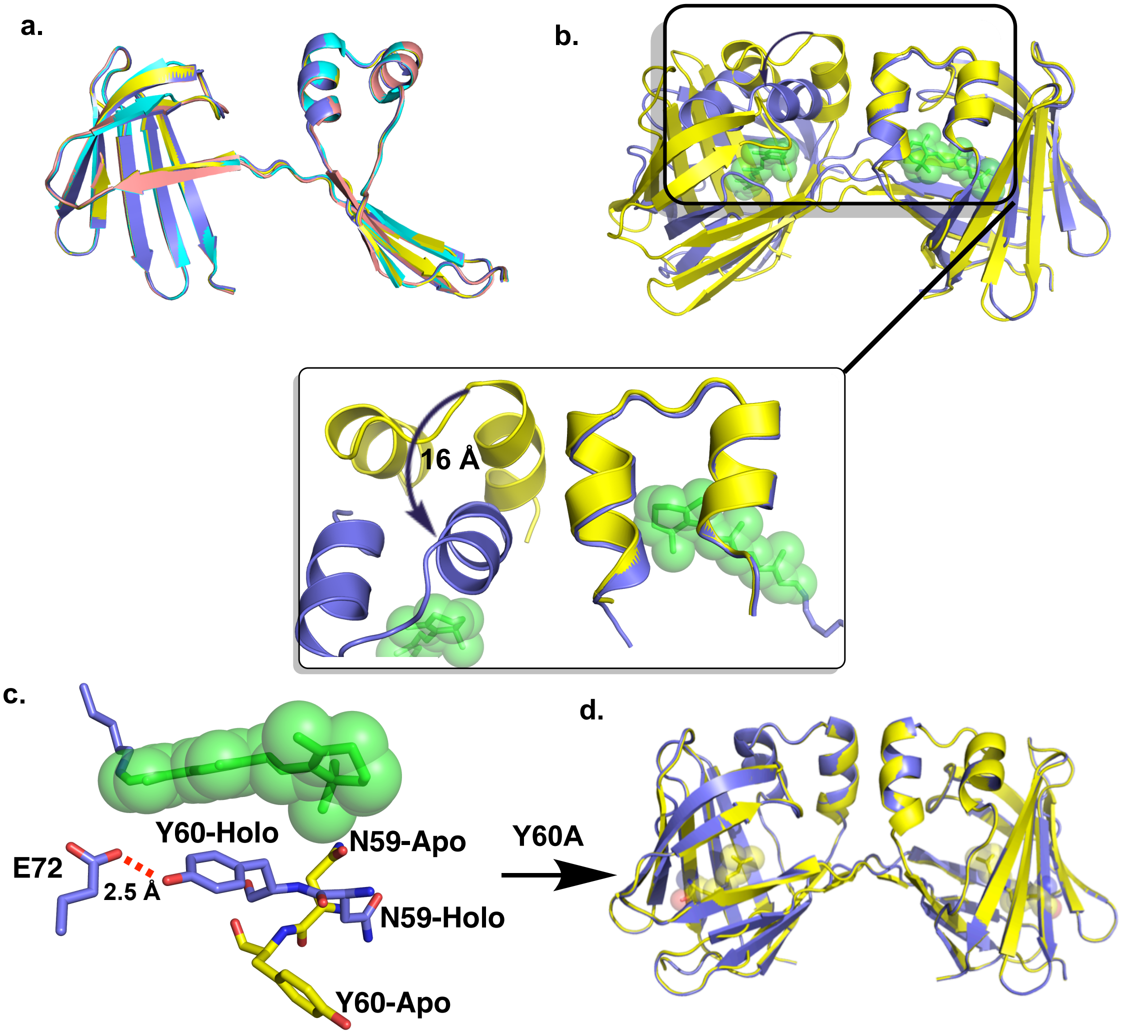


Figure II-9. a. Overlay of one chain each of four apo dimers: WT (cyan), Q108K:T51D (purple-blue), Q108K:K40L:T51F (yellow), and Q108K:K40L:T51W (pink), showing a relatively rigid conformation for apo dimers. b. Overlay of holo (purple) and apo (yellow) Q108K:T51D structures showing a large change in the relative orientations of the two domains upon ligand binding, leading to a 16 Å movement (inset) in the α-helix. Retinal (green) is shown in both sphere and stick representations. c. The retinal (green) binding pocket of apo (yellow C atoms) and retinal-bound (blue C atoms) Q108K:T51D. Allosteric conformational change is driven by the orientation of the Tyr60 and Asn59 sidechains. A flipped-in Asn59 would sterically clash with the bound retinylidene, leading to the flipped-out conformation of Asn59 and flipped-in conformation of Tyr60. All other atoms colored by type throughout: nitrogen (blue), oxygen (red). d. Overlay of apo Q108K:K40L:T51F (purple) and holo Q108K:K40L:T51F:Y60A (yellow) showing no change in the relative orientation of the two domains upon ligand binding.

In contrast, the structures of the retinylidene-bound dimers reveal that none lead to the symmetric structure cited above. In fact, the retinal-bound holo displayed a different relative orientation of the two domains in comparison with the apo symmetric dimers

(Figure 11-9b). Remarkably, most of the holo DS dimer crystal structures depicted four to twelve molecules in the asymmetric unit in contrast to the apo symmetric dimer that mostly crystallized with one molecule in the asymmetric unit, Table II-1.

However, the similarity between the relative orientations of the two dimer domains is much closer in the holo structures than the apo symmetric dimers (Figure II-10). The distance between the main-chain nitrogen atoms of Arg30, located in the ligand portal region, is much larger in the overlaid holo/apo structures of Q108K:K40L:T51F than in

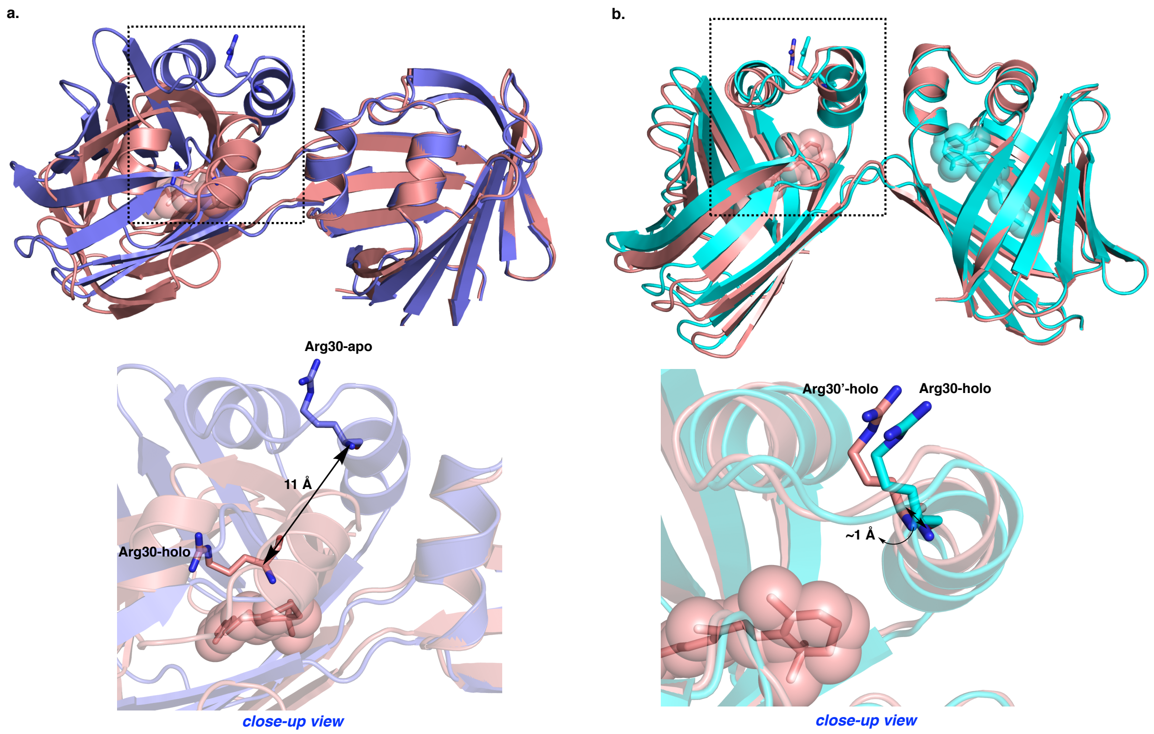


Figure II-10. Retinylidene-bound hCRBPII undergo dramatic conformational change upon ligand binding. a. An overlay of apo Q108K:K40L:T51F (**KLF**) (purple) and holo **KLF** (chains A and C, both shown in pink), showing the large motion of helix 1 upon ligand binding. b. An overlay of holo **KLF** (pink) and holo Q108K:T51D (cyan), showing the modest conformational difference in the two holo structures. The relative distance between C_α of Arg 30 in the overlaid holo/apo structures is 11.3 Å while the same distance in the overlaid holo structures is 1 Å.
*Bound retinal molecules in holo structures are shown as transparent spherical models. The complete dimer in the apo structures was generated by crystallographic two-fold symmetry operation.

the holo/holo Q108K:K40L:T51F and Q108K:T51D structures (~11 Å in apo/holo vs. ~1.0 Å in holo/holo).

II-2-2. Elucidating the mechanism of the protein switch

The major cause of this large relative twist in the two domains upon ligand binding involves the dispensation of Asn59 and Tyr60 (Figure II-9). In virtually all of the apo-symmetric dimers, the Asn59 side-chain is positioned inside the binding cavity while Tyr60 follows the anticipated β-strand register and is directed outside of the binding cavity. However, in the hCRBPII monomer (with more than 40 structures determined so far), Asn59 and Tyr60 are directed outside and inside of the binding pocket, respectively (Figure II-11).

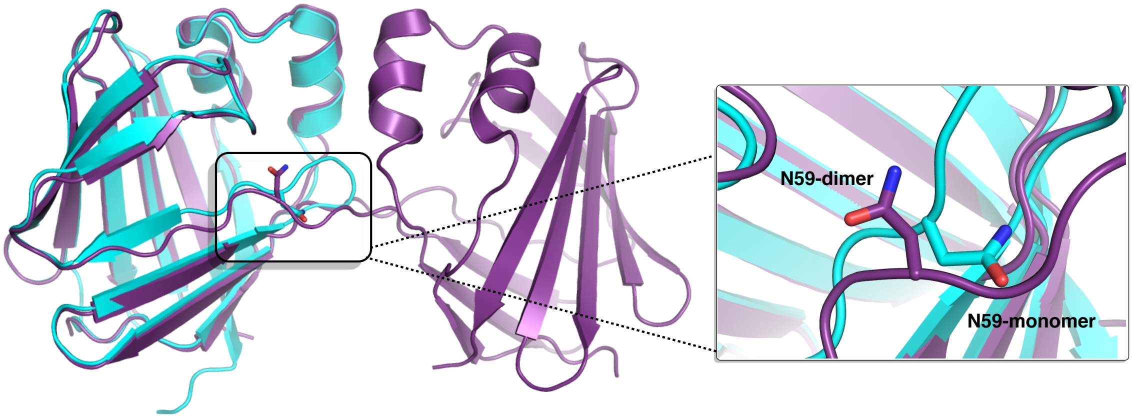


Figure II-11. Overlay of WT apo monomer (gray, PDB code: 2RCQ) and WT apo hCRBPII dimer (purple, PDB code: 4ZH9). The “flipped in” Asn59 conformation is required for the relative orientation of the two halves of an open monomer to adopt the orientation required to form the symmetric DS dimer.

Presumably, the flipped-in Asn59 conformation facilitates the relative orientation of the two halves of an open monomer required to form the symmetric domain- swapped dimer during folding (Figure II-11).[48] Conversely, the Asn59 side-chain is flipped out of the binding cavity in the seven holo dimer structures, similar to the monomer. The cause of this is clearly seen in Figure II-9, where an overlay of the Q108K:T51D holo and apo

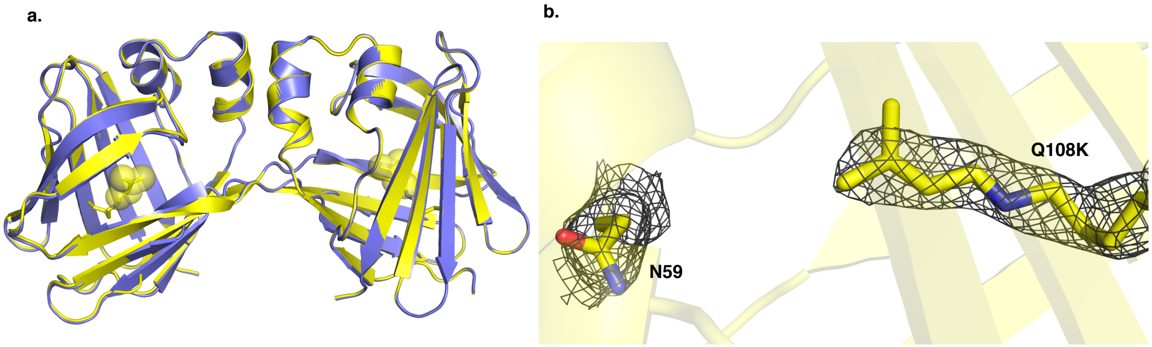


Figure II-12. a. The overlay of Q108K:K40L:T51F (apo) (purple) and Q108K:K40L:T51F:Y60A (holo) (yellow) shows no change in the relative orientation of two domains upon ligand binding. b. The retinal bound Q108K:K40L:T51F:Y60A DS dimer. The electron density is observed only for the first 5 carbons of retinal illustrating that Y60A mutation makes the chromophore disordered inside of the binding pocket. Therefore, Asn59 keeps its flipped-in conformation which does not lead to the change in the relative orientation of the two domains. *Bound retinal molecules in the holo structure are shown as transparent spherical models. The complete dimer in the symmetric DS dimers was generated by crystallographic two-fold symmetry operation.

dimers shows that the flipped-in Asn59 sterically impinges on the bound retinylidene. As a result, retinal binding sterically forces the Asn59 side-chain out of the binding cavity, leading to the large change in the relative orientation of the two domains. The flipped-out Asn59 in holo domain-swapped dimers also correlates with the Tyr60 conformation, in which the flipped-out Asn59 causes a flipped-in conformation for Tyr60, resulting in a low-barrier hydrogen bond between the side-chains of Tyr60 and Glu72, which influences the trajectory of the bound retinylidene (Figure II-9). Therefore, the interactions between the ligand and both Tyr60 and Asn59 act as the "allosteric spring" propelling the large conformational change of the two domains upon ligand binding. This same structural change is observed for the WT-hCRBPII domain-swapped dimer bound non-covalently with retinal (Figure II-8). The hypothesis that the allosteric conformational change between apo and holo dimers is driven solely by the orientation of the Tyr60 and Asn59 side-chains naturally gives rise to the prediction that the "allosteric spring" could be broken by the mutation of either of these residues.

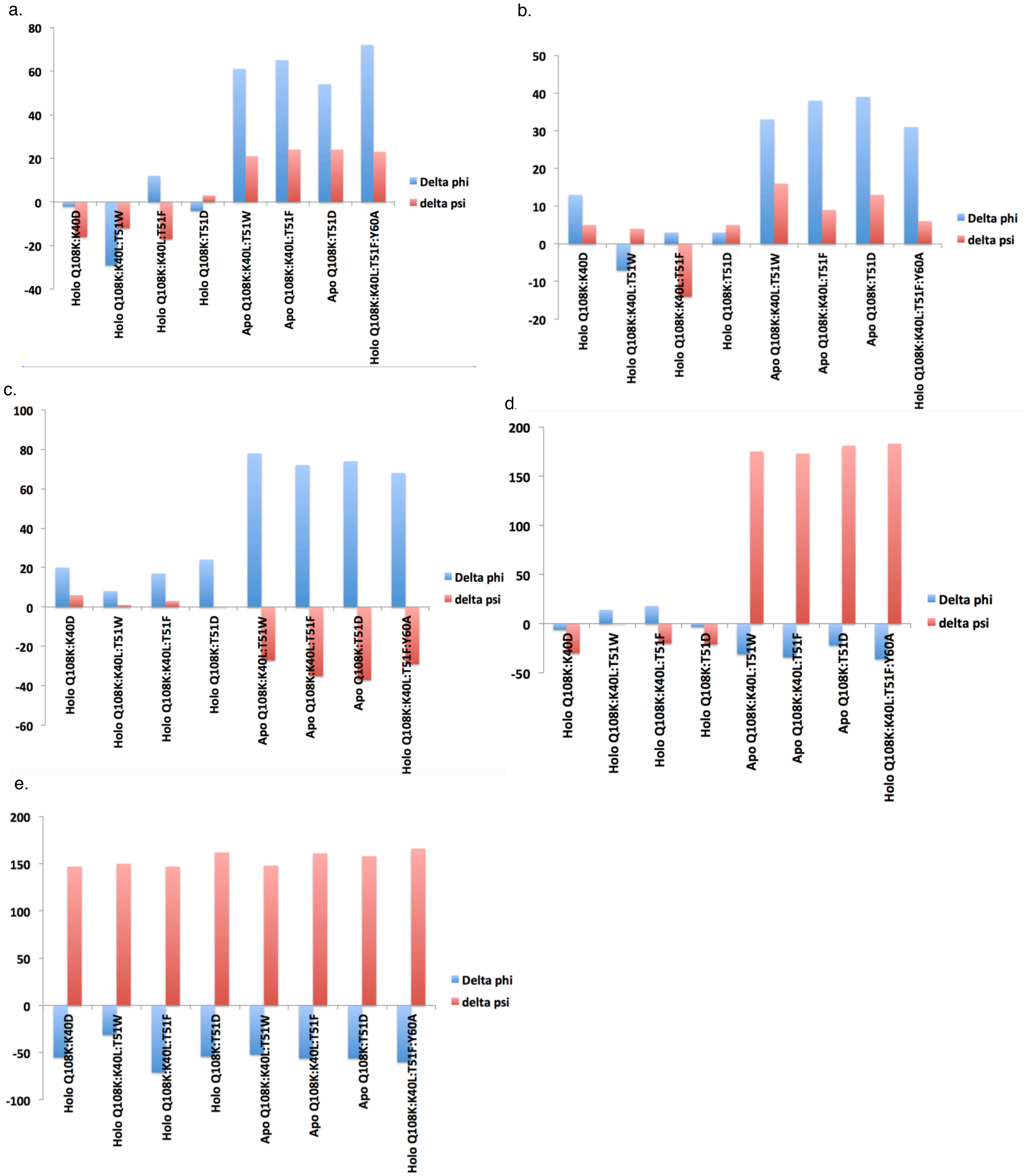


Figure II-13. The difference between phi/psi angles of holo and apo DS dimers and WT hCRBPII monomer phi/psi angles. For holo structures, one of the chains from each mutant has been considered. Phi angle differences are shown in blue and psi angle differences are shown in red. a. Tyr 60. b. Asn59. c. Arg58. d. Phe57. e. Thr60. As shown from a to d, a large difference in the phi/psi angles between holo and apo structures is observed.

The latter hypothesis was tested by mutating Tyr60 to Ala, resulting in the

Q108K:K40L:T51F:Y60A hCRBPII tetramutant. As predicted the retinylidene-bound holo dimer structure is very similar to the apo-symmetric dimers (Figure II-9d). In the other holo dimers, the flipped-in conformation of Tyr60 stabilizes the β-ionone ring and places it in a defined trajectory that sterically clashes with Asn59, resulting in its flipped-out conformation. However, the removal of this interaction (Y60A mutation) leads to ligand disorder, demonstrated by the loss of electron density for all but the four carbons closest to the imine bond (Figure II-12). This allows Asn59 to maintain its flipped-in conformation, resulting in a similar orientation for the two domains as compared to apo dimers, even with the ligand bound (Figure II-12). The observation above supports the central role played by Tyr60 and Asn59 in the mechanism for allosteric conformational change seen in hCRBPII domain-swapped dimers and further illustrates the symmetric apo dimer conformation as the default structure for the dimer in the absence of any perturbation. The rotation of the phi and psi angles of residues 58 to 60 is primarily responsible for both flipping Asn59 from inside to outside of the binding pocket and significantly changing the relative orientation of the two dimer domains (Figures II-9 and II-13). This illustrates the almost mechanical nature of the structural changes seen in the dimer.

II-3. Creating a disulfide bond as a trigger for allosteric behavior

With the structural and mechanical details regulating the conformational change well understood, we then pursued strategies for enlarging the library of conformational readouts available to the system. As shown in Figure II-14, the relative orientations of two α- helices in the holo versus apo domain-swapped dimers are proximal and undergo a large twist such that the introduction of a disulfide bond between them will necessarily alter the conformational readout of ligand binding, with the Ala28 position particularly well-positioned for the linkage. Our prediction is experimentally verified, as the Q108K:T51D:A28C variant yields a DS dimer with a disulfide bond between the Cys28 residues of the two chains. This was shown by both reducing and non-reducing SDS

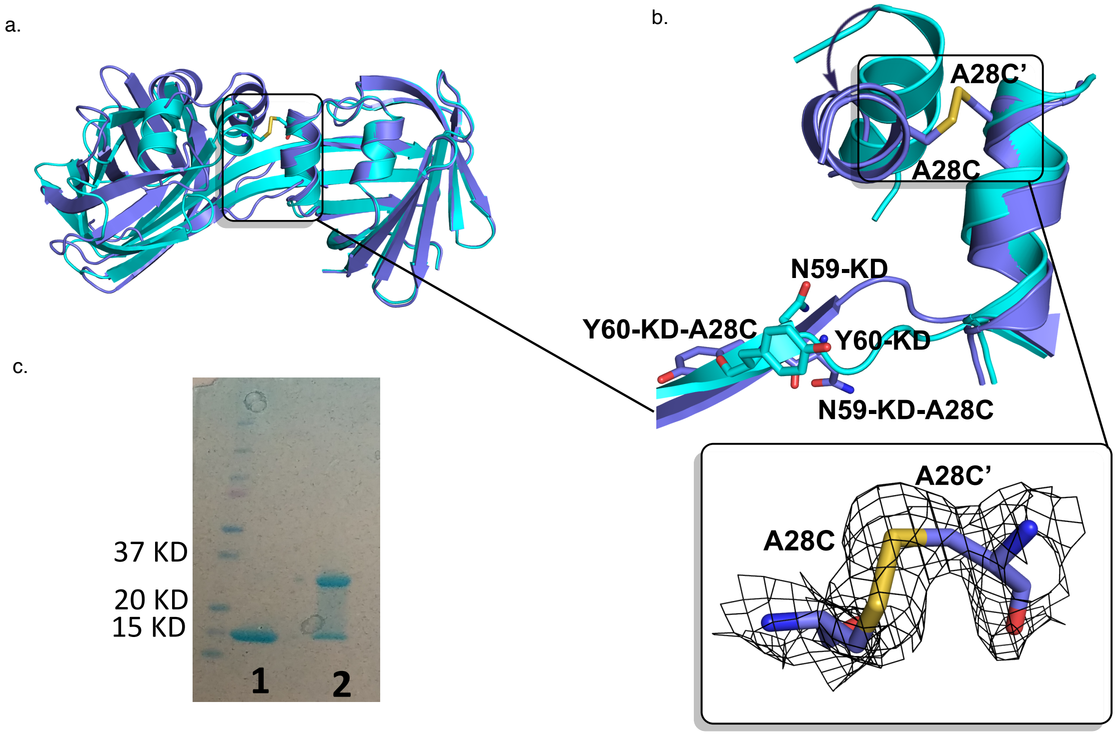


Figure II-14. Engineering a cross-link in the interface of two domains by A28C mutation: a. Overlay of Q108K:T51D-Apo (cyan) and Q108K:T51D:A28C-Apo (blue) showing the new conformation enforced by disulfide bond formation. b. A close-up showing the disulfide bond and the flipped-out conformation for Asn59 and flipped-in conformation for Tyr60. Inset: The disulfide bond electron density contoured at 1σ. c) SDS PAGE gel electrophoresis under reducing (1) and non-reducing (2) conditions, confirming disulfide bond formation in solution.

PAGE gel electrophoresis and crystal structure analysis (Figure II-14). The inter-subunit linkage causes the apo structure of Q108K:T51D:A28C to adopt a relative orientation of domains different from apo Q108K:T51D (Figure II-14a), resulting in a flipped-out conformation for Asn59 and a flipped-in conformation for Tyr60 in the Q108K:T51D:A28C variant. In contrast, the retinal-bound structure of this mutant revealed yet another unique and distinguishable conformation. This demonstrates the potential use of the DS dimer as a conformational molecular switch, where the switching can be tuned to give distinct conformational outcomes. The overlay of the apo and holo Q108K:T51D:A28C structures clearly shows the driving force behind this new conformational change to be the steric interactions between residues located in the α-helices and retinal that drive a conformational change in the opposite direction relative to the holo domain-swapped dimers. Additionally, the retinal dissociation constant for Q108K:T51D and Q108K:T51D:A28C was calculated using a fluorescent quenching assay showing that the

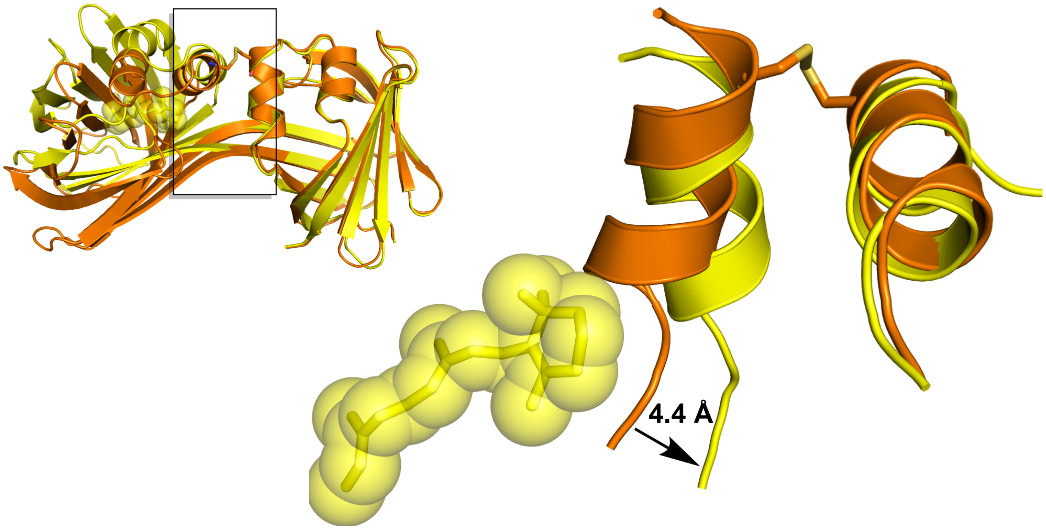


Figure II-15. An overlay of apo (orange) and holo Q108K:T51D:A28C (yellow), both featuring the Cys28 disulfide bond. On the right side, the bound retinal sterically forces the α-helices to move, inducing the ligand-induced conformational change.

disulfide bond has a minor effect on the retinal dissociation constant (Figure II-16). Collectively, four unique conformations are identified: apo and holo DS dimers and apo and holo disulfide-bonded DS dimers, with each revealing a distinct conformation (Figure II-17). Furthermore, we employed CD spectroscopy as a convenient tool to explore the conformational readouts of our system in solution (Figure II-18). In this assay, retinal binding causes an increase in the CD at 218 nm in the dimer (Q108K:T51D), whereas the monomer (Q108K:K40L:T51V) and Y60 mutated dimer (Q108K:K40L:T51F:Y60A) showed minor changes upon the addition of retinal. Furthermore, a decrease in CD absorption at 218 nm was observed for the disulfide-bonded dimer (Q108K:T51D:A28C) upon adding the reducing agent (1mM DTT). The additon of retinal after reducing the disufide bond causes an increase in CD absorption at 218 nm correlated with Q108K:T51D retinal addition. However, the unreduced forms of Q108K:T51D:A28C (dimer) and Q108K:K40L:Y60L:A28C (dimer), used as controls, led to minor changes in CD absorption at 218 nm with the addition of retinal and reducing agents, respectively.

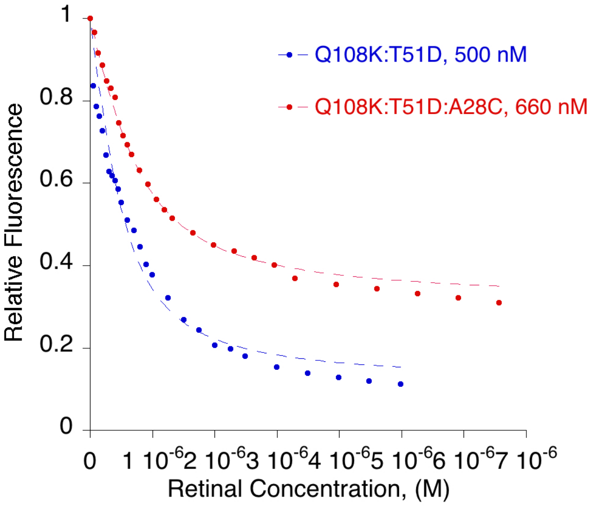


Figure II-16. Fluorescent quenching assay for Q108K:T51D and Q108K:T51D:A28C shows dissociation constants of 220 nM and 368 nM respectively for retinal.

These four conformations can be addressed via ligand binding and/or redox potential, resulting in a versatile protein switch. Conveniently, each conformation gives rise to a

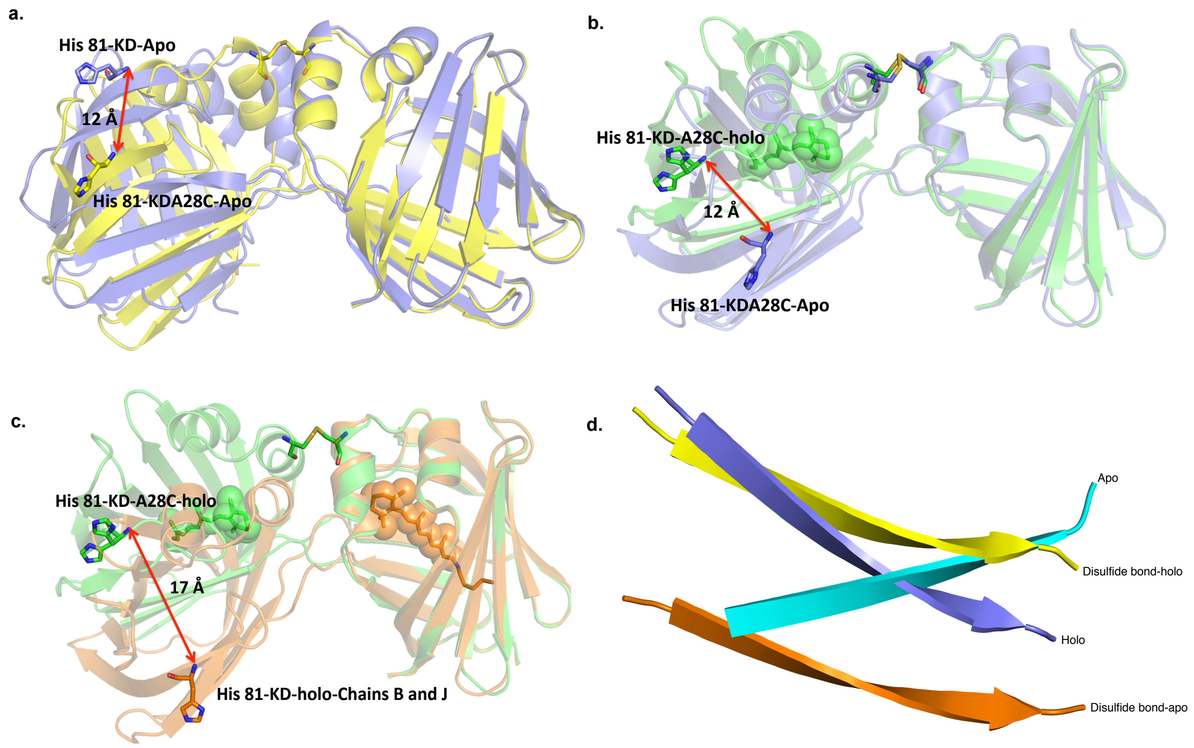


Figure II-17. Structural overlays of a. Q108K:T51D (KD) apo (yellow) versus KDA28C apo (blue). b. KD-A28C-holo (green) versus KDA28C-Apo (blue). c. KD-A28C-holo (green) versus KD-holo (orange). All of these overlays show a large conformational change originated from either a disulfide bond or ligand binding or both. d. A summary of conformations potentially addressable by ligand binding, redox environment, or both. Only the c-termini of the proteins are represented. The distance between apo versus holo retinal bound structures is 12 Å, apo versus disulfide bond-apo is 16 Å, apo versus disulfide bond holo is 7.5 Å, and the apo with disulfide bond versus holo with disulfide bond is 10 Å. The apo and holo structures are represented by Q108K:T51D (cyan and blue, respectively) and the apo and holo disulfide-bonded structures are represented by Q108K:T51D:A28C (orange and yellow, respectively).

change in the c-terminus (by as much as 10 Å), allowing the protein switch to be attached at the c-terminus instead of being inserted. This is in contrast to most naturally-occurring

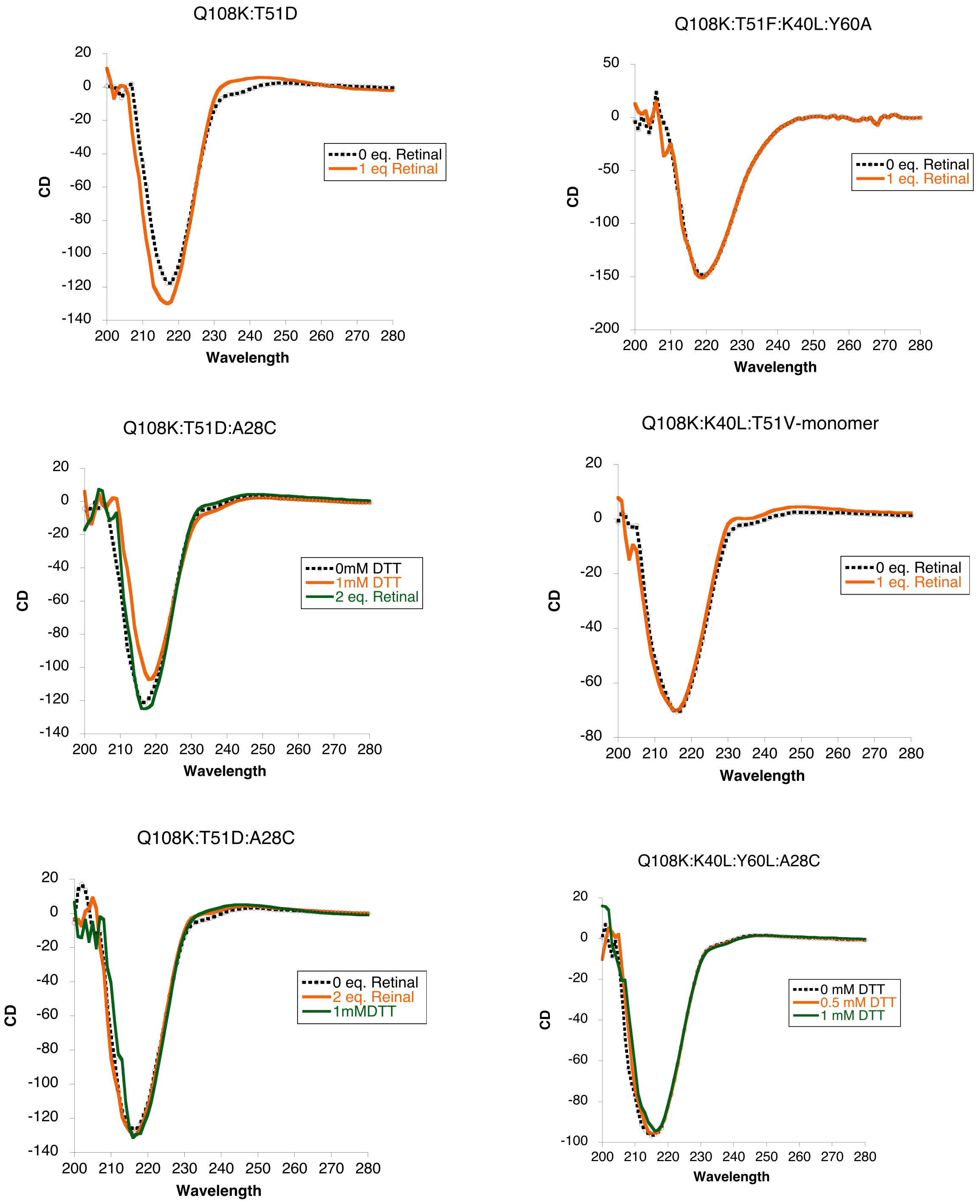


Figure II-18. The CD spectra of different variants: Q108:T51D dimer, Q108K:K40L:T51F:Y60A dimer, Q108K:T51D:A28C dimer, Q108K:K40L:T51V monomer, and Q108K:K40L:Y60L:A28C upon adding retinal or/and reducing agent.

conformational protein switches, where there is relatively little motion of the termini.

II-4. Generating an allosterically-regulated zinc binding site

A protein switch is only useful if the conformational change can be coupled with a change in function. We chose metal binding as a convenient and potentially applicable function to test in our system. To this end, a metal binding site was designed in the interface of the two domains because residues that are proximal in the apo conformation can undergo significant separation upon ligand binding, with the predicted result being a substantial decrease in metal binding affinity.

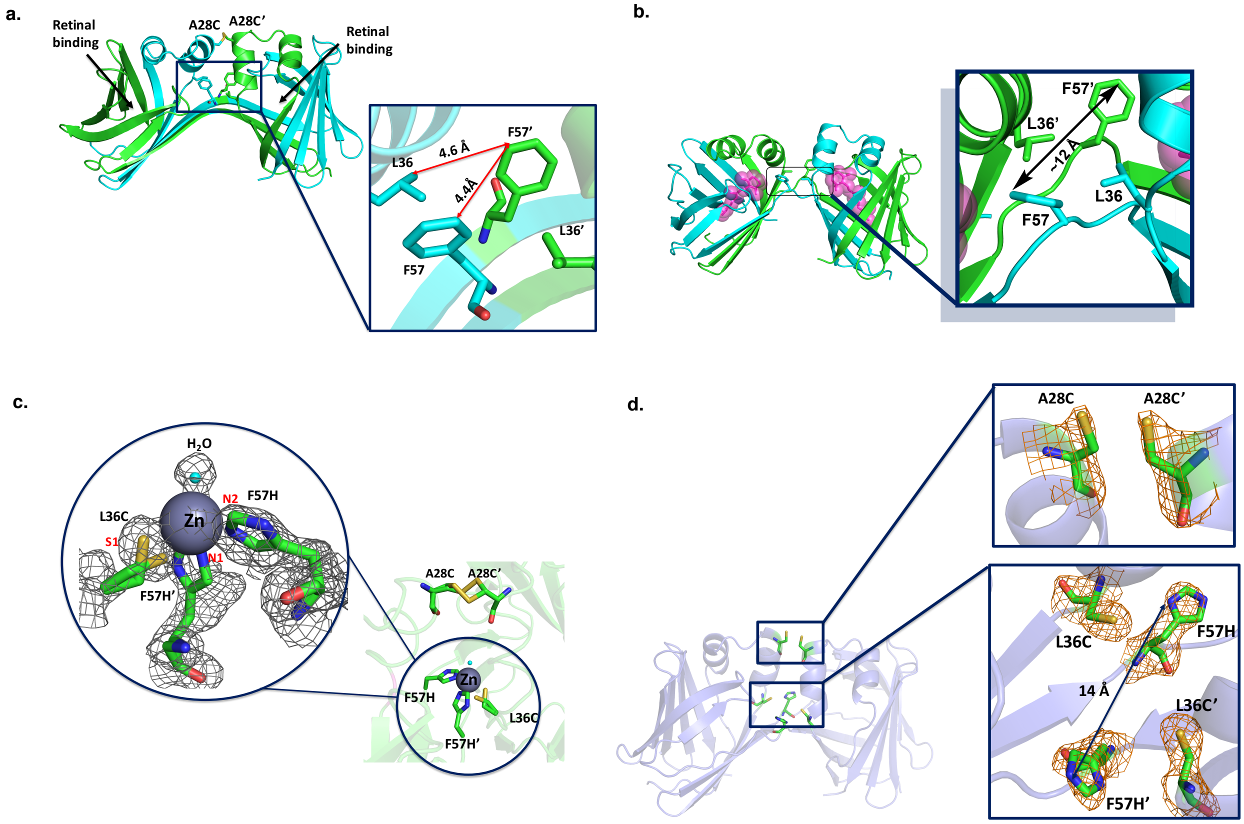


Figure II-19. a. Designing a new metal-binding site using Q108K:T51D:A28C crystal structure as a guide. The proximity of F57H and L36C is critical to providing the metal-binding site. b. The crystal structure of holo retinal-bound Q108K:T51D chains G and I, showing the large separation between two F57s from each chain upon ligand binding. c. Atomic resolution crystal structure of Q108K:T51D:A28C:L36C:F57H bound with zinc showing tetrahedral geometry for zinc coordinated with two histidines coming from two adjacent chains of dimer, a cysteine, and a water molecule. The electron density of the $2F_O$-F_C map is countered to 1σ. The conformations of F57H that did not contribute to the metal-binding site have been omitted for clarity.

Though a number of potential binding sites were investigated, Listed in Table II-2 and Table II-3, Leu36 and Phe57 were chosen due to their favorable relative orientation in the apo Q108K:T51D:A28C structure (Figures II-19a and II-19b), and their separation upon retinal binding (II-19b), resulting in the Q108K:T51D:A28C;L36C;F57H (**AG1**) variant. The metal ion affinity of apo **AG1** was evaluated using two assays, a radioactive ^{65}Zn-binding assay (with a help of Dr. Pinger, Prof. Spence's lab) and isothermal titration calorimetry (ITC) (collaboration with Prof. Jin) (Figure II-20). In both cases, a dissociation constant of 1.6 μM was obtained. In contrast, the metal ion affinity was substantially reduced through the addition of the ligand (all-*trans* retinal) (Figure II-20b). The addition of the reductant only moderately reduces the affinity either in the presence or absence of all-*trans* retinal (Figure II-20b). Two variants that do not contain the metal binding site (Q108K:T51D and Q108K:T51D:A28C) show more than one hundred-fold weaker metal ion affinity, demonstrating the specificity of metal binding to the engineered site. This demonstrates the allosteric control of metal binding in the DS dimer, as structurally predicted.

To verify the allosteric mechanism of Zn-bound **AG1** was determined, revealing a tetrahedrally coordinated zinc with one His57 from each protomer, one Cys36 and one water molecule define the coordination sphere of the bound zinc (Figure II-19c). The coordination distances are consistent with those found in other zinc-bound proteins (bond

Table II-2. Variants constructed in search of an allosteric metal binding site.

Variants	Results
Q108K:T51D:**T56C**	Insoluble expression
Q108K:T51D:**F57H**	The protein was expressed solubly but does not generate a metal-binding site in the dimer form.
Q108K:K40D:T53A:R58L:Q38F:Q4FV62E:**F57H**	Insoluble expression
Q108K:T51D:**F57C**	Insoluble expression
Q108K:T51D:**F57E**	Insoluble expression
Q108K:T51D:**F57C:S55C**	Insoluble expression (The in vitro refolding experiment also led to protein aggregation)
Q108K:T51D:**A28C**	The protein was expressed solubly but does not generate a metal-binding site in the dimer form shown by the crystal structure.
Q108K:T51D:**I32C**	The protein was expressed solubly but does not generate a metal-binding site in the dimer form shown by the crystal structure.
Q108K:T51D:**A28H**	The protein was expressed solubly but does not generate a metal-binding site in the dimer form shown by the crystal structure.
Q108K:T51D:**L36C**	Insoluble expression
Q108K:T51D:**I32H**	Insoluble expression (The in vitro refolding experiment also showed the formation of monomer, dimer, and trimer. However, the refolded protein was not stable during concentrating.
Q108K:K40L:**T51H**	The protein was expressed solubly but does not generate a metal-binding site in the dimer form shown by the crystal structure.
Q108K:**K40H:I42H**	Insoluble expression
Q108K:**K40H:I42H:T51H**	Insoluble expression (The in vitro refolding experiment also led to protein aggregation)
Q108K:T51D:**A28C:L36C**	The protein was expressed solubly but does not generate a metal-binding site in the dimer form shown by the crystal structure.
Q108K:T51D:**F27Y:A28H**	The protein was expressed solubly. The crystallization of dimer were not successful.
Q108K:T51D: **F27H:A28H**	The protein was expressed solubly.It led to more oligomer formation whose crystallization was unsuccessful.
Q108K:T51D:**A28H:D26H**	Insoluble expression
Q108K:T51D:**I32C:L36C**	Insoluble expression
Q108K:T51D:**I32C:F57H**	Insoluble expression
Q108K:T51D:**L36H:F57H**	Insoluble expression
Q108K:T51D:**A28C:F57H**	Insoluble expression
Q108K:T51D:**A28C:L36H**	Insoluble expression
Q108K:T51D:**I32C:L36H:F57H**	Insoluble expression
Q108K:T51D:**I32H:L36H:F57H**	Insoluble expression
Q108K:T51D:I32G:**L36C:F57H**	Insoluble expression
Q108K:T51D:A28C:I32A:**L36C**	Insoluble expression
Q108K:T51D:**A28C:I32H:L36H:F57H**	Insoluble expression
Q108K:T51D:**S55C:F57H:L36H**	Insoluble expression
Q108K:T51D:**S55H:F57H**	Insoluble expression

distances for Zn-S: 2.2 Å, Zn-N$_1$: 2.2 Å, Zn-N$_2$: 2.1 Å, and Zn-O: 2.1 Å.). It should be noted

Table II-3. Insertion in the hinge loop region and other variants constructed in search of an allosteric metal binding site.

Variants	Result
Q108K:K40L:T51F:**Y60H:E72C**	The protein was expressed solubly but does not generate a metal-binding site in the dimer form as shown by the crystal structure.
Q108K:T51D:**K31:H:H:I32**	Insoluble expression
Q108K:T51D:**K31:C:C:I32**	Insoluble expression
Q108K:T51D:**K31:H:H:H:I32**	Insoluble expression
Q108K:T51D:**F57:H:H:H:H:H:H:R58**	The protein was expressed solubly and was purified using nickel column but disappointingly it was not crystallized.
Q108K:T51D:A28C:**F57:H:R58**	The protein was expressed solubly but does not generate a metal-binding site in the dimer form as shown by the crystal structure.
Q108K:T51D:A28C:**F57:H:H:H:R58**	The protein was expressed solubly but does not generate a metal-binding site in the dimer form as shown by the crystal structure.
Q108K:T51D:A28C:**F57:H:H:H:H:R58**	The protein was expressed solubly.
Q108K:T51D:A28C:**F57:H:H:H:H:H:R58**	
Q108K:T51D:A28C:**F57:H:H:H:H:H:H:R58**	The protein was expressed solubly and it was crystallized in the apo form. Radioactive labelling zinc assay revealed a K_d: 20 nM to zinc.
Q108K:T51D:A28C:**F57:L:H:H:H:L:L:R58**	The protein was expressed solubly but Disappointingly it was not crystallized.
Q108K:T51D:A28C:**F57:C:C:C:C:C:C:R58**	Insoluble expression
Q108K:T51D:A28C :**R58:H:H:H:N59**	The protein was expressed solubly but does not generate a metal-binding site in the dimer form as shown by the crystal structure.
Q108K:T51D:A28C:**L36C:F57H**	The protein was expressed solubly and it was crystallized in the zinc-bound form. Radioactive labelling zinc assay and ITC revealed a K_d: 1.6μM to zinc.
Q108K:T51D:A28C:**L36C:F57H:N59D**	The protein was expressed solubly but does not generate a metal-binding site in the dimer form as shown by the crystal structure.
Q108K:T51D:A28C**:L36C:F57C**	The protein was expressed solubly but led to a large amount of trimer formation. The crystallization of dimer and trimer were not successful.

that the zinc geometry revealed here resembles that seen for many natural zinc binding proteins such as carbonic anhydrase. [51]

Further, the structure of the reduced disulfide form of **AG1** was also determined, showing the disulfide bond to be successfully reduced, but with little conformational change relative to metal-bound **AG1**, consistent with the modest change in metal affinity upon reduction (Figure II-19d).

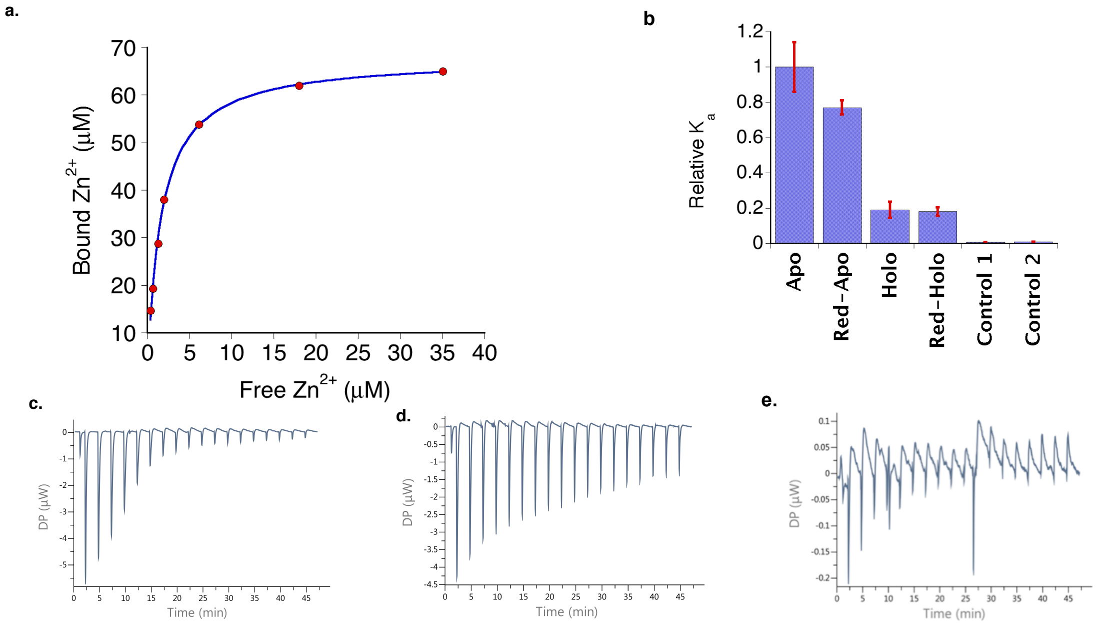


Figure II-20. a. The binding curve for apo **AG1** using a radioactive-labeled ^{65}Zn showed K_d=1.62 ± 0.14 µM. b. Comparison of association constants for different conformations calculated in 40 µM total $^{65}ZnCl_2$: 1) Apo; 2) in the presence of reducing agent only (Red-Apo), 3) in the presence of retinal (Holo). 4) in the presence of all-*trans* retinal and reducing agent (Red-Holo); As controls, Q108K:T51D (control 1) and Q108K:T51D:A28C (control 2) were used and led to a negligible association constant with Zn. c. ITC experiment shows the dissociation constant of 1.60 ± 0.22 µM for apo. d. The ITC experiment shows the dissociation constant of 29.3±11µM in the presence of all-*trans* retinal and reducing agent. e. Q108K:T51D:A28C was used as a control.

As shown in Figure II-21b, the overlay of the zinc-bound **AG1** and retinal-bound Q108K:T51D shows the extensive conformational change anticipated upon retinal binding, resulting in a rearrangement of the metal binding site to a conformation incompetent for metal binding (Figures II-21b and II-21c), consistent with the allosteric behavior detected in the metal binding assays ($K_{d\ (Holo\ +\ reduced\ form)}$: 29 ± 8 µM vs $K_{d\ (apo)}$:1.60 ± 0.22 µM). (Figure II-20).

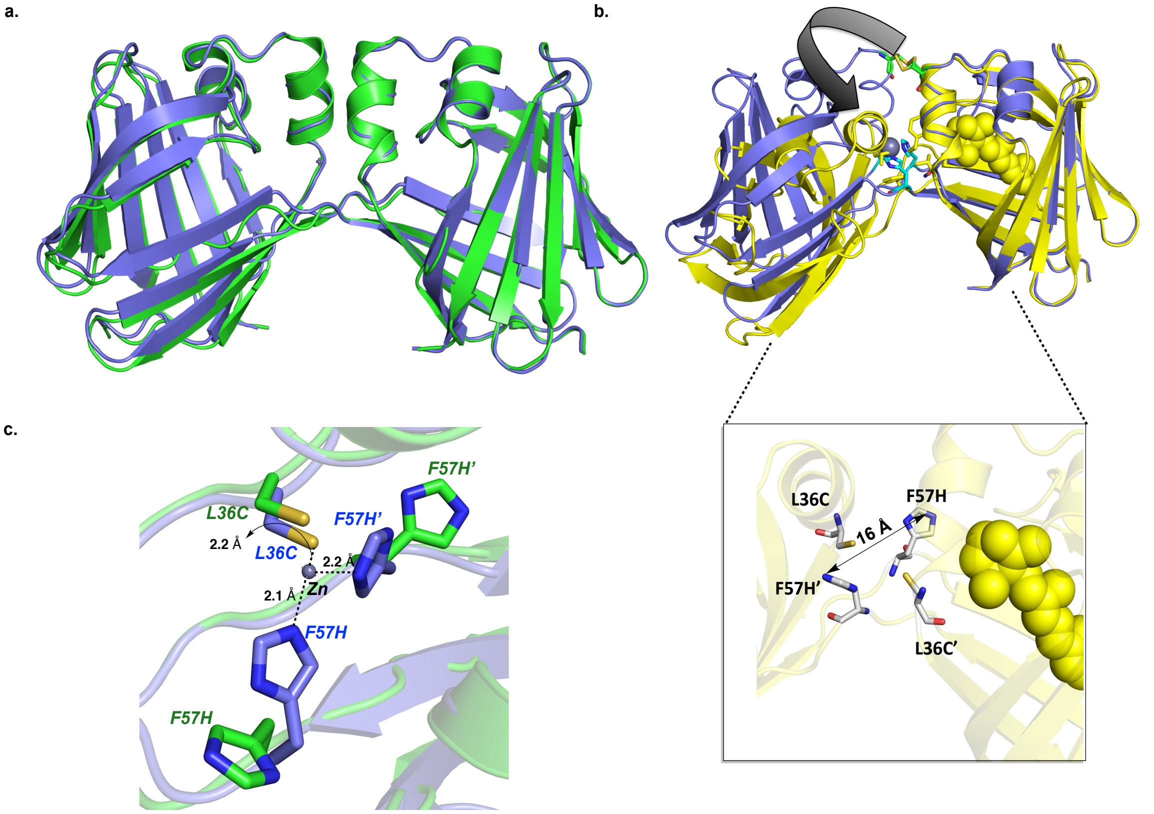


Figure II-21. a. The overlay of Q108K:T51D:A28C:L36C:F57H reduced form (green) and KD-Met (purple-blue) showing no significant change in the conformation upon reducing the disulfide bond. b. Overlay of KD-Met (yellow) and holo Q108K:T51D (purple-blue) crystal structures demonstrating a large change anticipated upon retinal binding, giving rise to reorganization of the metal binding site to a conformation which is not optimal for metal binding. F57 residues of each chain in Q108K:T51D structure have been mutated using PyMOL. c. Overlay of zinc-bound apo Q108K:T51D:A28C:L36C:F57H (purple) (K_d=1.60 ± 0.22 μM) with holo and reduced form structure (K_d=29 ± 8 μM) modeled using Q108K:T51D (green) crystal structure demonstrating a large change anticipated upon retinal binding, giving rise to reorganization of the metal binding site to a conformation which is not optimal for metal binding.

II-5. Using fatty acids as input signals

So far, we created a novel protein conformational switch that can be activated by retinal or the reduction potential of the environment. To further expand the number of input signals in our system, we decided to use fatty acids, which are highly crucial for regulating many biological processes, such as inflammation, signal transduction, energy production, etc. [52]. To create the binding site for fatty acids — inspired by fatty acid binding protein families and cellular retinoic acid binding protein II — Q108R:T51D was created

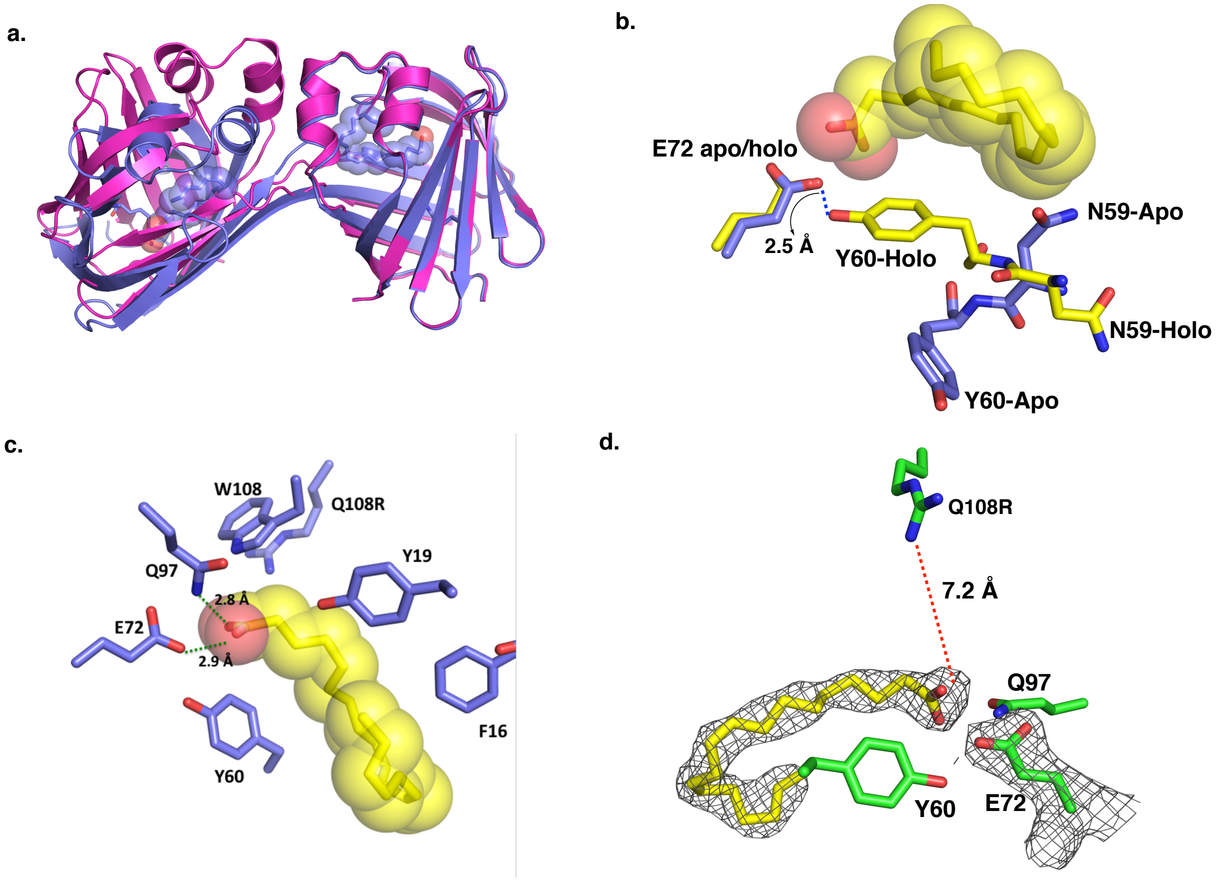


Figure II-22. a. The overlay of holo linoleic acid bound Q108R:T51D (purple) and Q108K:T51D (apo) domain-swapped dimer crystal structures. b. The same mechanism of dispensation of N59 and Y60 demonstrates that the large conformational change occurs upon fatty acid binding. c. The binding pocket of linoleic acid and the interacting residues were highlighted. d. The linoleic acid electron density was shown. (The $2F_O$-F_C map was countered to 1σ.) Q108R is far away from the carboxylic acid group of linoleic acid and it not involved in the binding event.

with the hope that Q108R could interact with fatty acids and stabilize them in the binding pocket (Figure I-22c). The crystal structures of Q108R:T51D bound with linoleic acid and arachidonic acid were obtained at 2.4 Å and 1.97 Å resolution. The overlay of these structures with an apo domain-swapped dimer revealed the large conformational change in the relative orientations of two dimer domains as observed for holo retinal-bound

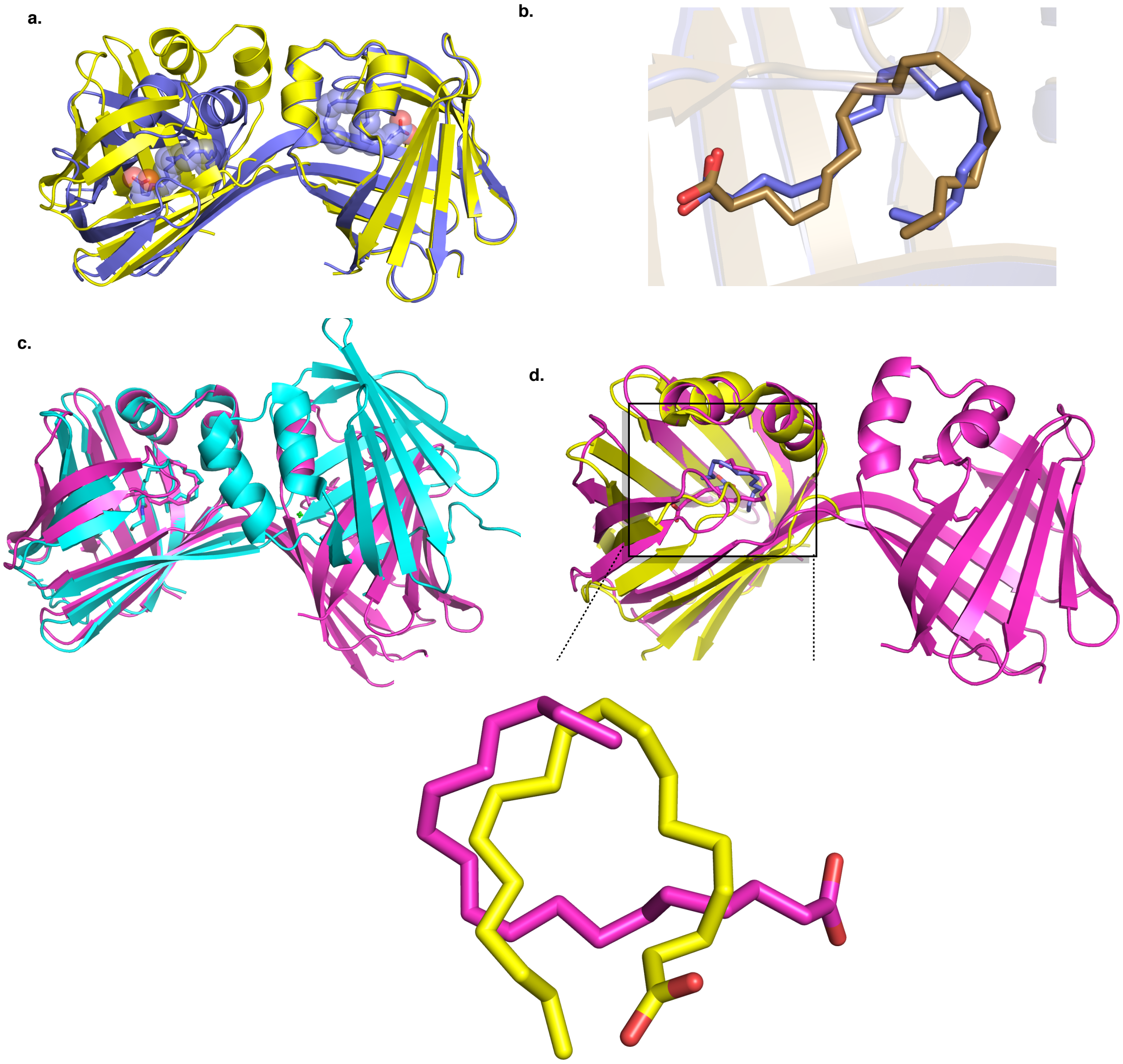


Figure II-23. a. Overlay of arachidonic acid bound Q108R:T51D crystal structure (purple) with apo Q108K:T51D (yellow) showing the large conformational change in the relative orientation of the two dimer units. b. Overlay of linoleic acid bound structure vs. arachidonic acid bound structure demonstrates a similar trajectory for these two fatty acids in the binding pocket of the Q108R:T51D DS dimer variant of hCRBPII. c. Overlay of Q108R:T51D crystal structure bound with linoleic acid (pink) with FABP5 bound with N-(2-hydroxyethyl) icosanamide (PDB code: 4AZR). d. Overlay of Q108R:T51D crystal structure bound with linoleic acid (pink) with FABP4 bound with linoleic acid (yellow), PDB code: 2Q9S.

domain-swapped dimers (Figure II-22a). The close-up overlay of apo and holo in the hinge region revealed the conformational switch follows the same mechanism as the retinal-bound holo dimer, where the dispensation of Asn59 and Tyr60 causes a large change in the relative orientations of the two dimer domains (Figure II-22b).

Furthermore, the crystal structure revealed a novel binding site for both fatty acids. As shown in Figures II-20C and II-20d, the structure of Q108R:T51D bound with linoleic acid shows that E72 and Y60 are interacting with the carboxylate group of fatty acids, and Q108R does not participate in fatty acid binding. This is in contrast with the typical fatty acid protein binding pocket, which shows that the caboxylate group of fatty acids usually interacts with a conserved Arg and Tyr.[53] The crystal structure of arachidonic acid bound with Q108R:T51D shows a similar conformational change and trajectory of fatty acid binding as cited above (Figure II-23). This observation begs the question of whether or not the WT monomeric form of hCRBPII, traditionally thought to merely bind with retinal and retinol,[54] also has an afffinity toward fatty acids. Since these proteins are largely expressed within the intestine,[54] they might have a secondary function to the one originally determined. In this case, a fluorescence-quenching competition assay was performed for the WT-hCRBPII-monomer, revealing that linoleic acid is able to replace retinal in the binding pocket in a concentration-dependent manner. Additionally, the crystal structure of WT-hCRBPII bound with linoleic acid was obtained at 1.6 Å resolution, further depicting the reproducibility of the binding pocket toward linoleic acid binding in WT-hCRBPII (The structure and data has not been shown here).

II-6. The observation of a ligand-induced “swap back” at atomic resolution

To expand the conformational diversity of our domain-swapped dimer, we used merocyanine, a synthetic fluorophore with aldehyde functionality (Figure II-24). Previously, Yapici *et al*. demonstrated new “turn-on” fluorescent protein tags using merocyanine, exciting at 594 nm and emitting at 619 nm upon generation of a protonated Schiff base with the monomeric hCRABPII.[55] Furthermore, in other unpublished results Dr. Santos described the photobleaching feature of this chromophore. She showed that during excitation, the fluorophore is photobleached leading to the loss of fluorescence; then, by re-adding the fluorophore, the protein rebinds with new molecules, and fluorescence is retrieved. If the bleaching resulted in loss of bound ligand, this feature could give us a convenient optogenetic with conformational switching driven by

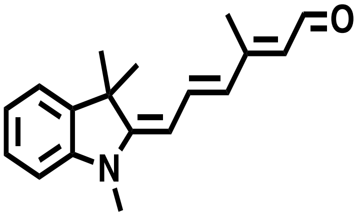


Figure II-24. The structure of merocyanine chromophore with aldehyde functionality designed in Prof. Borhan’s lab.

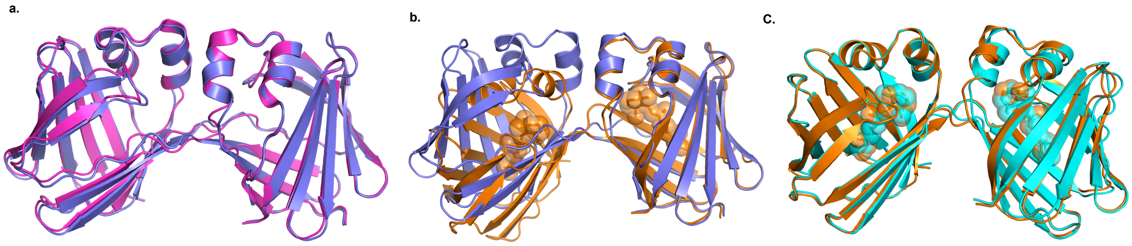


Figure II-25. a. Overlay of apo DS dimer of **A1** (purple) with WT-hCRBPII DS dimer (orange) showing a symmetric dimer (RMSD: 0.3.). b. Apo DS dimer of **A1** (purple) with holo retinal-bound DS dimer of **A1** (orange) revealing the large conformational change in the relative orientation of the two dimer domains. c. Overlay of holo retinal-bound DS dimer of **A1** (orange) with holo retinal-bound DS dimer of Q108K:T51D (cyan).

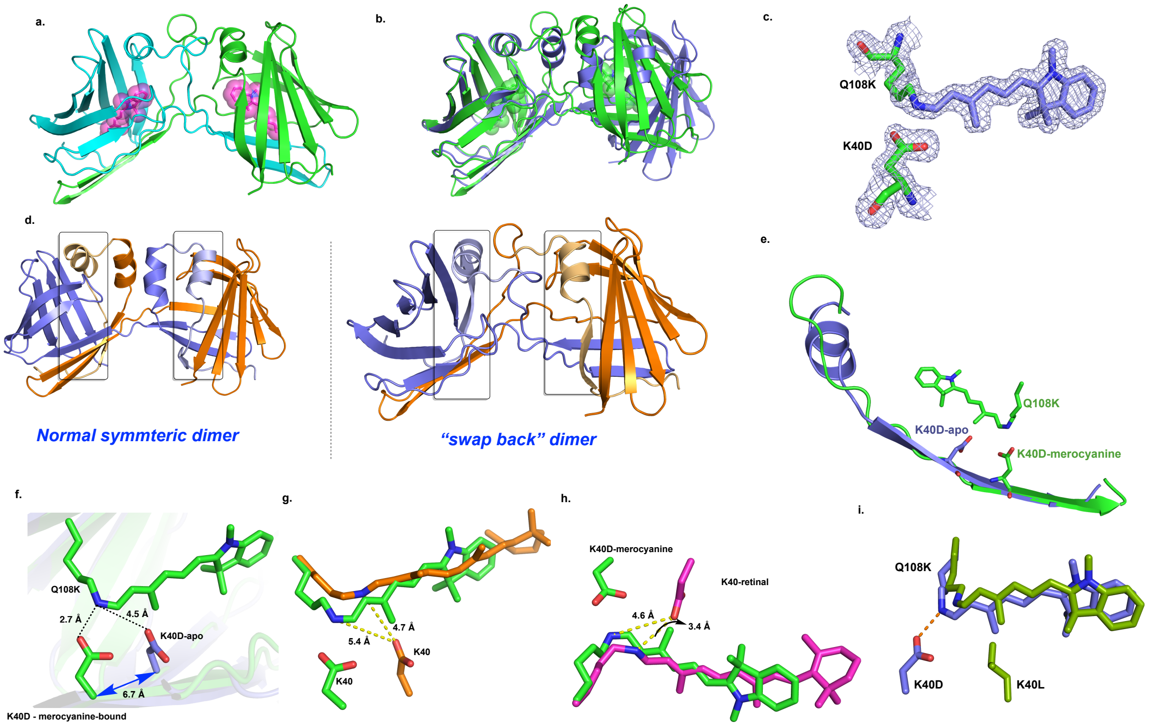


Figure II-26. a. The merocyanine-bound **A1** crystal structure showing the swapped-back DS dimer. Chains are shown in green and cyan. The merocyanine ligand is represented by a combination of sphere and stick representation (pink). b. The overlay of merocyanine-bound **A1** (green) and apo (purple-blue). c.The electron density highlighted for the ligand and K40D contoured to 1σ. d. Left: The normal domain-swapped dimer where half of the protein undergo the domain-swapping. The swapped open monomers showed with different colors. Right: The "swap back" dimer where the α-helix and beta strand 1 from each chain is swapped back highlighted with a box and light orange and purple colors. e. K40D has to move two residues along the strand to interact directly with the nitrogen of iminuim, giving rise to a change in the registry of the amino acids that make up strand 2 forcing the melting of the α-helix. f. Different positions of K40D in apo (purple-blue) vs. merocyanine-bound **A1** DS dimer (green) in the overlaid structures. g. The positions of K40D in merocyanine-bound **A1** DS dimer (green) vs. retinal-bound **A1** (orange), chain I. h. The positions of K40D in merocyanine-bound **A1** (green) vs. retinal-bound **A1** (pink), chain i. The overlay of merocyanine-bound **A1** DS dimer (green) vs. merocyanine-bound Q108K:K40L:T51V (purple blue).

chromophore bleaching, resulting in a return to the apo conformation. Therefore, we decided to screen the crystallization of the merocyanine ligand with many variants of hCRBPII that form domain-swapped dimers. The comprehensive screening led us to the following variants: Q108K:T51D and Q108K:R58L:T53A: K40D:Q38F:Q4F (**A1**). Unfortunately merocyanine-bound Q108K:T51D yielded poorly diffracting crystals (beyond 4 Å resolution). However, the crystal structures of apo, merocyanine bound and retinal bound **A**1 were determined successfully. The apo structure of the protein, as shown in Figure II-25a, depicted the expected apo symmetric dimer. Furthermore, the retinal-bound crystal structure showed 12 molecules in asymmetric unit and retinal binding led to a large twist in the relative orientations of the two dimer domains, similar to that observed for other retinal-bound dimers (Figures II-25b and II-25c).

In contrast, the crystal structure of **A1**/merocyanine (obtained at 1.81 Å) gave a surprisingly different structure. A new swapping occurred in the protein. In the normal domain-swapped dimer of hCRBPII, half of the protein (55-133) underwent domain swapping. Here, n-terminal residues (1-25) of the protein were swapped back, leading to a large conformational change in the protein (Figures II-26a-II-26d). Having the atomic-resolution crystal structures of this variant in our hands, we were interested in understanding the mechanism of this conformational change. As shown in Figures II-26e and II-26f, the overlay of the merocyanine-bound structure with retinal-bound holo dimers reveals that in the normal domain-swapped dimer, the K40 residue does not interact directly with the nitrogen of iminium. However, in the merocyanine-bound structure this residue moves two positions along the β-strand to create a direct hydrogen bond with the nitrogen of the iminium (2.7 Å), altering the registration of 26-46 which include β-strand 2 and α-helix 2. This shift in registry apparently causes the melting of α-helix (26-36) in the structure. Presumably, loss of helix/helix interactions due to melting the second helix would also facilitate the helix 1 and strand 1 swap back. The reason behind this abnormal behavior of the ligand can be seen in Figures II-26g and II-26h. The merocyanine ligand, which is a shorter ligand than all-*trans* retinal, adopts almost the same configuration of the imine bond for all the structures obtained so. In all of these conformations, merocyanine lacks the ability to interact directly with K40D, Figure II-26I. However, the position of the imine bond can be altered in retinylidiene structures, as seen in some chains of the all-*trans* retinal-bound **A1** crystal structure**,** where retinal generates a direct interaction with K40D, Figure II-26h. Another difference is that merocyanine is a highly-conjugated ligand in comparison with all-*trans* retinal, leading to a naturally high pK_a of the chromophore. Additionally, in the case of the **A1** variant, the binding pocket is more hydrophobic as a result of hydrophilic residue mutations, such as Q4F, Q38F, and T53A. Therefore, there is a high pK_a iminium in the merocyanine-bound structure — possessing a positive charge that is distant from the negatively-charged K40D — surrounded by many hydrophobic residues in the binding pocket. As mentioned earlier, merocyanine does not have the ability to alter the position of the imine bond to create a direct interaction with the counter ion. As a result, K40D has to move two residues along the strand, giving rise to a change in the registry of the amino acids that make up strand 2 forcing the melting of the α-helix. To investigate the effect of K40D, three new variants: **A1**-K40, **A1**-K40L, and **A1**-K40F, were constructed, all of which were dominantly expressed as monomers. This indicates that the presence of K40D is important for dimer formation.

Only the monomer of **A1**-K40F bound to merocyanine led to well-diffracting crystals. As shown in Figure II-27*a*, the crystal structure demonstrates that in the absence of K40D there is no change in registry of the β-strand. This is consistent with the idea that the change in registry is driven by the formation of the salt bridge between Asp40 and the Iminium. Instead, the chromophore is rotated almost 180° allowing the iminium nitrogen to form a cation-π interaction with Trp106. This cation-π interaction between iminium and Trp106 (or its equivalent) has been seen in several other PSB-forming ligand-bound structures such as Q108K:K40L:T51V:R58W:Y19W/retinal (PDBcode: 5FFH), Q108K:K40L:T51V:R58W:Y19W/retinal (PDB code: 5F7G), and Q108K:K40L:T51V:T53C:R58W:T29L:Y19W:A33W/retinal (PDB code: 4EDE). Finally, the ability to create such a large conformational change through merocyanine can be used to generate a new class of molecular switches in optogenetics, a critical application that should be pursued in the future.

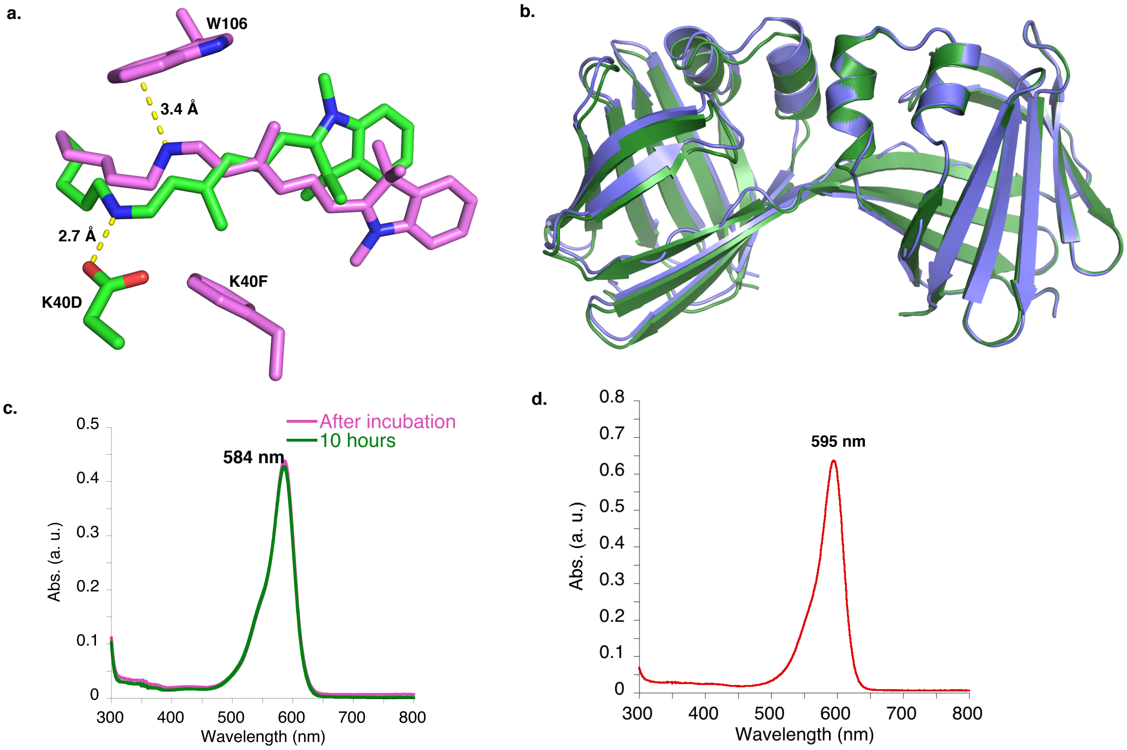


Figure II-27. a. Overlay of **A1** DS dimer bound with merocyanine (green) and **A1**-K40F monomer bound with merocyanine (pink) portraying a significant change in the trajectory of the ligand. b. Overlay of apo **A1** vs. **ThioFluor9** chromophore-bound **A1** DS dimer crystal structure shows no conformational change. c. UV. Vis spectrum of **A1** bound with the merocyanine ligand. The protonated Schiff base absorbs at 584 nm. d. UV. Vis spectrum of **A1**-K40F bound with the merocyanine ligand. The protonated Schiff base absorbs at 595 nm.

As shown above, retinal and merocyanine led to complete different conformations on **A1**. This characteristic would allow us to control the conformational state of one protein scaffold by different ligands (retinal and merocyanine). We also used the shorter ligands as a proof of principle — **ThioFluors**, thiophene-based fluorophores synthesized by Dr.

Santos (see Chapter IV for details) — where the ligand binding does not disturb the apo-symmetric dimer and therefore does not alter the protein conformation (Figure II-27b). This demonstrates the power of our protein conformational switch to “surf” various conformational outcomes upon the introduction of different input signals. The latter application can be especially pursued for certain *in vivo* studies when the protein conformational change is unnecessary at the time of the experiment. In this case, the goal is to merely monitor the protein of interest’s localization, which is linked to the protein conformational switch using fluorescence signals.

II-7. Hinge loop region: nature’s economic way to create conformationally-distinct proteins?

Nature creates functionally and conformationally-distinct proteins from a common scaffold through convergent or divergent evolution.[56] Domain-swapping, where two identical segments of a protein exchange to create dimers or higher-order oligomers, was originally speculated to be the natural mechanism for surfing different oligomerization states in proteins, and therefore regulating the final function.

Based on previous studies, variables leading to a large amount of domain-swapped oligomer formation include a change in pH, protein concentration, ligands, temperature, and certain mutations in the hinge loop region.[57,58]

Specifically, many research groups have demonstrated the hinge loop role as an effective tool for controlling the oligomerization states in proteins.[59] For example, studies of domain swapping in staphylococcal nuclease, suc1, ckshs1, CD2, and single-chain Fv showed that reducing the hinge-loop length causes destabilization of monomeric forms and promotion of domain swapped dimer formation. [60-64] Other studies revealed that the insertion of specific amino acids in the hinge region could favor domain swapping.[65]

However, to our knowledge, there are no examples showing, in a systematic manner, the effect of the insertion in the hinge loop region on the overall conformational states of the domain-swapped oligomers and their monomeric forms with atomic-resolution crystal structures. This study illustrates the essential role the hinge loop region plays to explore new conformational states without imposing a drastic change in the sequence of proteins. To investigate the role of the hinge region, we needed a suitable platform, meaning a protein scaffold that has a high tolerance toward mutations and can be easily expressed and crystallized.

II-7-1. Choosing a suitable construct

As mentioned earlier, the role of the engineered disulfide bond was highlighted as a tool to create new conformations in a protein. To investigate the role of amino acid insertion and to avoid changing the monomer/dimer ratio, we decided to use a disulfide-bonded construct as our initial template. The presence of the disulfide bond would further stabilize the scaffold and make it more stable towards mutation, but also, since the position of the disulfide bond is selected on top of the α-helices, it would not prevent conformational flexibility to produce and investigate different conformations in the dimer. The Q108K:T51D:A28C:L36C variant was chosen as a starting point. Q108K is a mutation that is required for Schiff base formation, T51D favors dimer formation, A28C generates the disulfide bond, and finally, L36C enforces the overall protein conformation as a normal symmetric dimer, as defined in our previous sections. This construct also gives rise to well-diffracting crystals, allowing us to observe the changes at atomic resolution.

II-7-2. “Phase problem" in the domain-swapped dimer

It’s important to highlight that in all of our domain-swapped dimer structures, domain swapping is the result of the loop connecting β-strands 3 and 4 of the monomer, leading to the formation of a single β-strand stretching the length of the dimer (Figure II-28). However, this results in a “phase problem” because, while the *odd* sidechains of β-

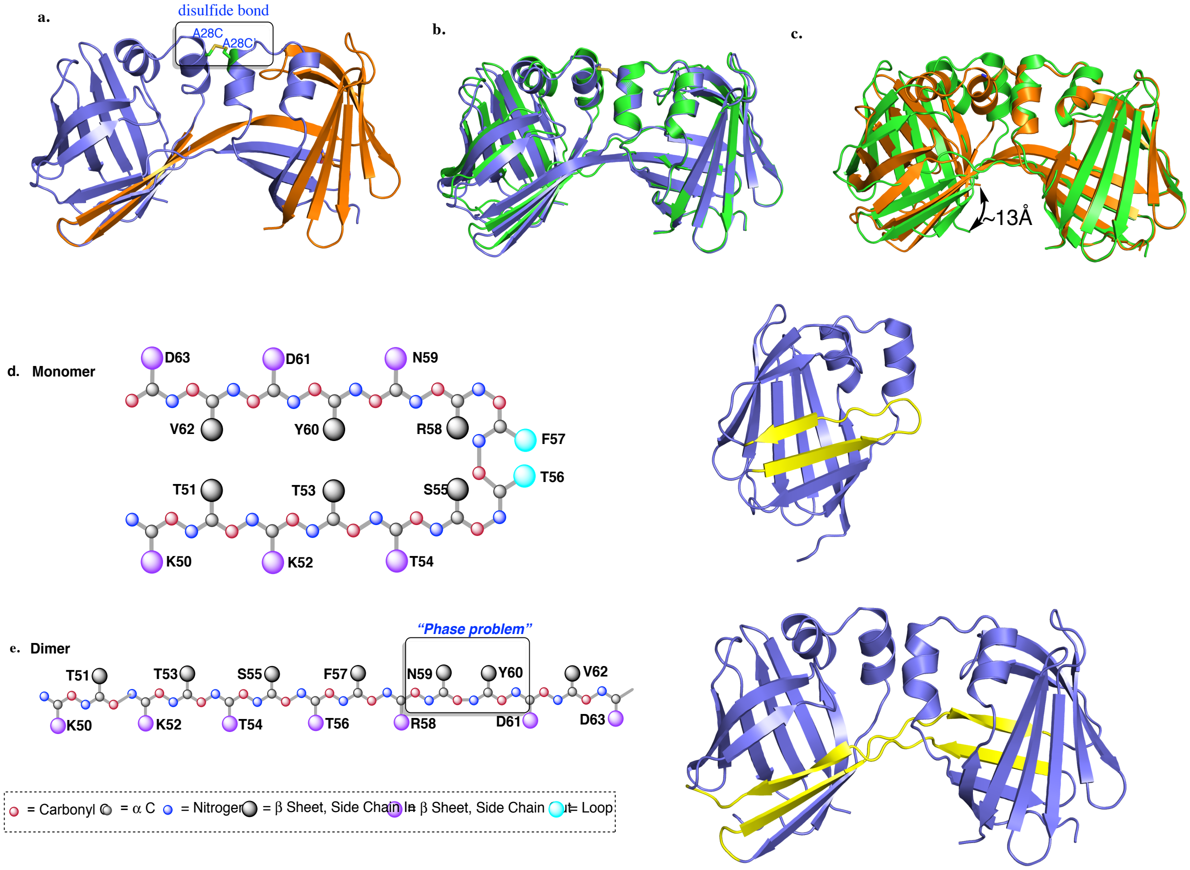


Figure II-28. a. Q108:T51D:A28C:L36C crystal structure. Each chain is shown in a different color and is connected by the disulfide bond on the top part of the α-helices. b. Overlay of Q108:T51D:A28C:L36C crystal structure (purple) with WT hCRBPII domain-swapped dimer (green) showing a similar conformation to the symmetric dimer for Q108:T51D:A28C:L36C. c. Overlay of Q108:T51D:A28C (orange) with WT hCRBPII dimer showing a drastic change in the relative orientation of the two dimer units, leading to a large change in the c-termini distance between the two structures. d. Arrangement of the residues connecting strands 3 and 4 in the monomer; left panel shows the crystal structure of the monomer where the aforementioned strands were highlighted. e. Arrangement of the residues connecting strands 3 and 4 in the dimer to make a continuous strand leading to a phase problem; left panel shows the crystal structure of the dimer where the aforementioned strands were highlighted.

strand 3 face inside the binding pocket, the *even* sidechains of β-strand 4 face outward. This means that the single β-strand formed in the domain-swapped dimer must, at some point, re-phase the strand, which would then put the even sidechains of the c-terminal half of the strand "in phase" with the odd-numbered sidechains of the first half.

II-7-3. Amino acid insertion: odd vs. even number of insertions

According to the "phase theory," we would expect the insertion to show a new trend in the overall structure of the protein, where the number of residues connecting the two strands is different from the original domain-swapped dimer.

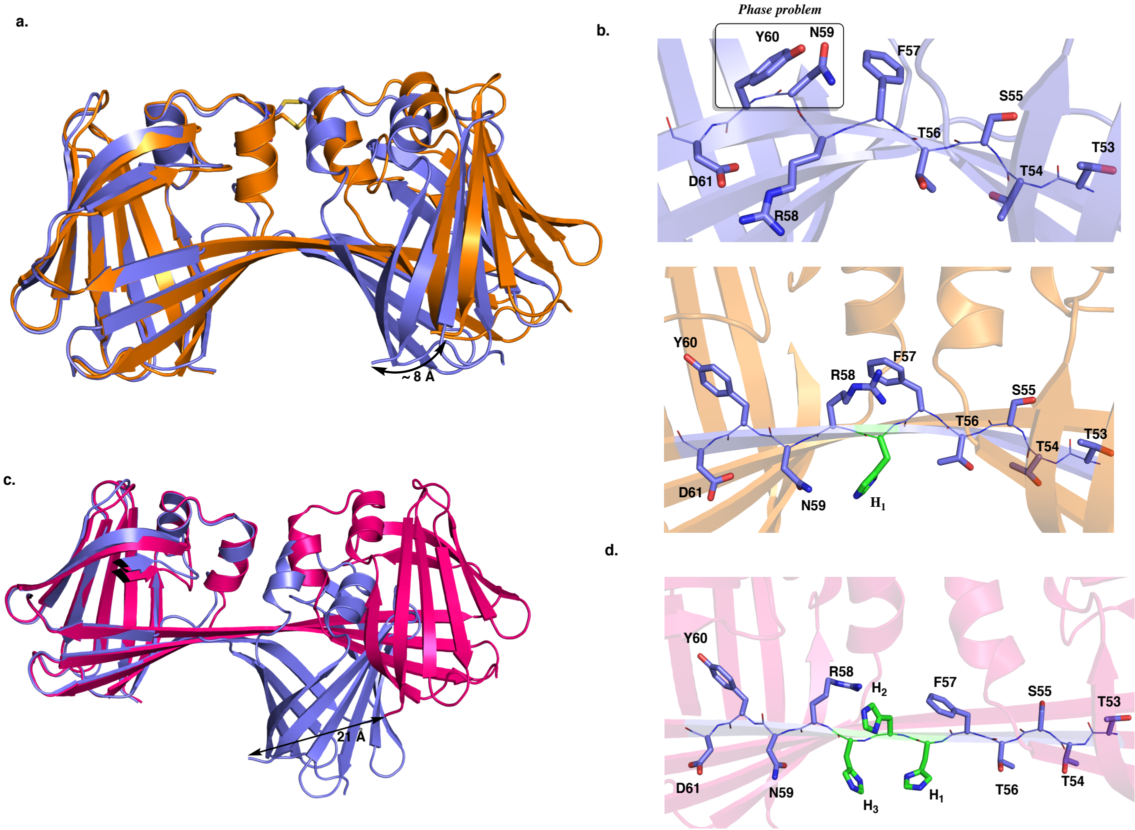


Figure II-29. a. Overlay of **1H** (orange) and Q108K:T51D:A28C:L36C (purple) crystal structures. The c-terminus has been moved ~ 8 Å after introducing His between F57 and R58. b. Top: The hinge loop regions and part of the two β-strands (3 and 4) were highlighted in Q108K:T51D:A28C:L36C where both Y60 and N59 are located inside the binding pocket and cause a "phase problem". Bottom: After insertion of one His between F57 and R58, a continuous β-strand is formed where every residue is "in phase". c. Overlay of **3H** (red) and Q108K:T51D:A28C:L36C (purple) crystal structures, representing a large movement of the c-termini and the relative orientation of the two dimer units. d. The "zoom-in" demonstration of the hinge region after the addition of three His, solving the "phase issue".

The result of the insertion could lead to two predictions: 1. Odd insertion would give rise to a continuous β-strand; even insertion would lead to loop formation between the two strands; 2. The number, nature, and location of the insertions would give rise to conformationally-distinct protein scaffolds, since they can influence the torsional angles of the key residues in the hinge region in different ways. In the initial effort, histidines were added between F57 and R58, the residues located in the middle of the hinge loop. Histidine residues were chosen for three reasons: first, they are soluble amino acids; second, their pK_as are pretty close to the PI of the protein (~5) so the insertion does not lead to a large change in the charge, and subsequently, the purification strategy that uses the ion exchange column; and third, during the amino acid insertion experiment, a novel metal-binding site can possibly be generated, which portrays how the new conformation can impose new functions in the protein. In our initial effort, two constructs were created: Q108K:T51D:A28C:L36C:F57:<u>H</u>:R58 (**1H**) and Q108K:T51D:A28C:L36C:F57:<u>H:H:H</u>:R58 (**3H**). These constructs were expressed and purified, and their crystal structures were determined.

As shown in Figure II-29, the crystal structures of these three constructs were overlaid with Q108K:T51D:A28C:L36C, which is structurally very similar to the apo-symmetric dimer. Adding a histidine between F57 and R58 creates an ideal β-strand connecting the two separated strands, fixing the phase problem mentioned above and satisfying the first prediction.

Upon adding this residue, as shown in Figures II-29a and II-29b, the relative orientations of the two dimer domains were altered, leading to 8 Å distance between the c-termini of the two structures (before and after insertion).

As shown in Figures II-29c and II-29d, the crystal structure of the **3H** insertion depicted

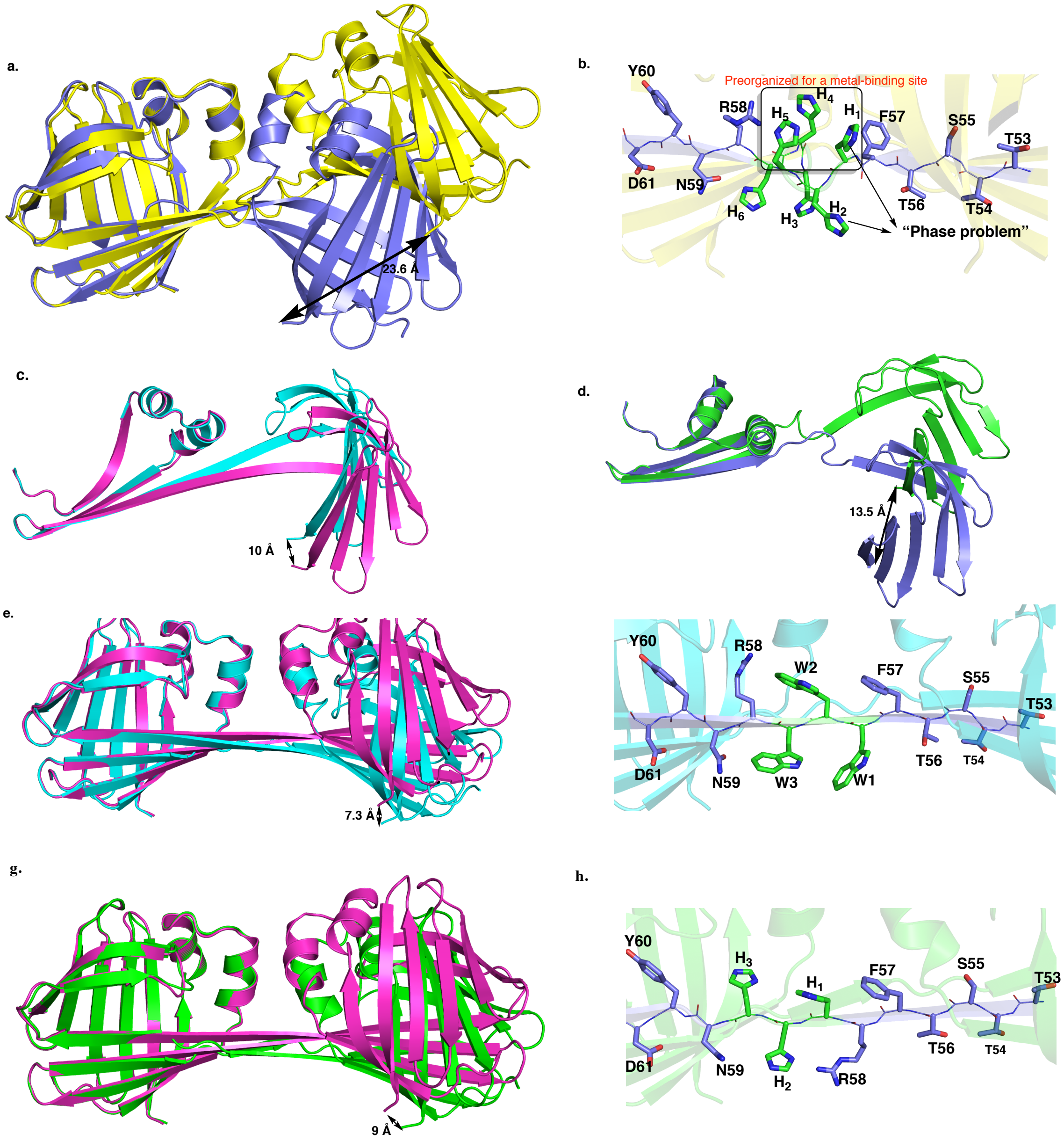


Figure II-30. Overlay of **6H** (yellow) and Q108K:T51D:A28C:L36C (purple) crystal structures. The c-terminus has been moved more than 23 Å after introducing six His between F57 and R58. b. Adding the even number of residues depicts the "phase problem," where the two neighboring His are outside the binding pocket and "out of phase." Also, three His are in the right conformation for generating a metal-binding site. c. Overlay of two open monomers of **3H** generating a domain-swapped dimer (chain A, pink, and chain B, cyan)**.** d. Overlay of two open monomers of 6H generating a domain-swapped dimer (chain A, green, and chain B, purple). e. Overlay of **3W** (cyan) and **3H** (pink) crystal structures, showing a novel conformation was created where the nature of amino acid insertion is different. f. Close-up view of the **3W** hinge region, where three Trp residues were inserted between F57 and R58 and every residue is "in phase." g. Overlay of **3H'** vs. **3H** crystal structures, demonstrating the position of insertion is also highly important in the creation of a new conformation in the protein. h. Close-up view of **3H'** hinge region, where three His residues were inserted between R58 and N59 and every residue is "in phase".

the same features of a continuous strand, formed as a result of an odd number of residues added in the β-sheet. Furthermore, the overlay of the **3H**-insertion vs. no-insertion structures demonstrated a larger conformational change in the relative orientation of the two dimer units. It almost looks as if the structure goes from a "*cis"* conformation, in the case of no insertion, to a "*trans"* form, in the case of **3H** insertion, giving rise to a drastic change in the c-termini of the protein (21 Å distance). Overall, the odd number of insertions satisfied the phase theory, as mentioned earlier, and it leads to a substantial change in the relative orientation of the two dimer units.

The next step was to test the effect of an even number of insertions. We tested this by introducing four histidines and six histidines between F57 and R58, leading to the formation of Q108K:T51D:A28C:L36C:F57:H:H:H:H:R58 (**4H**) and Q108K:T51D:A28C:L36C:F57:H:H:H:H:H:H:R58 (**6H**).

The four histidines did not produce high-quality crystals; however, the six histidines diffracted to 2.54 Å. As shown in Figure II-30b, the structure delineated the formation of loops in the hinge loop region where three histidines were preorganized to bind zinc.

Radioactive ^{65}Zn labeling (with the help of Dr. Pinger) further confirmed a 20.8 ± 3.8 nM dissociation constant with zinc. This shows that adding a certain number of histidines in the binding pocket can not only affect the conformation, but also leads to the creation of a new function in the protein.

Additionally, the overlay of this structure with no H insertion demonstrates a new conformational change in the relative orientation of the two dimer units, leading to a ~24 Å change in the c-terminal distance (Figure II-30a). The overlay of the two chains of **6H** generating the domain-swapped dimer revealed that structurally, the two open-monomer units are very different in comparison with the two open monomers forming the **3H** domain-swapped dimer (Figures II-30c and II-30d). This demonstrates that for a **6H** domain-swapped dimer, a large conformational change should occur in the c- and n-termini of the open monomers to form a dimer molecule, simply because more structural flexibility occurs as a result of loop formation in the hinge region vs. the strand in the **3-H** structure. Perhaps the presence of a disulfide bond would assist with the formation of a domain-swapped dimer, where the two open monomers are drastically different in their conformations. The in-depth screening to obtain a crystal — in the context of the Q108K:T51D:F57:H:H:H:H:H:H:R58 construct that does not provide the disulfide bond — was not successful.

II-7-4. The effect of the nature of amino acid insertion

The next step is to elucidate how the nature of amino acid insertion affects the overall conformation of the protein. This time, tryptophan was chosen as a bulkier amino acid that would be expected to achieve different conformational states of the protein. To test this hypothesis, three and six tryptophan residues were added to the Q108K:T51D:A28C:L36C construct to create Q108K:T51D:A28C:L36C:F57:WWW:R58 (**3W**) and Q108K:T51D:A28C:L36C:F57:W:W:W:W:WW:R58 (**6W**) constructs. The **3W** constructs were expressed solubly and successfully purified; however, **6W** did not lead to soluble expression. **3W** was also successfully crystallized, and its structure was determined at 2.38 Å resolution. As shown in Figures II-30e and II-30f, reminiscent of other odd insertions, this structure also showed a continuous β-strand formation as a result of solving the phase problem. However, it leads to a conformationally-distinct orientation of the two dimer units relative to one another, in comparison with **3H** and **6H**.

This further demonstrates the high capacity of the loop region to "surf" different conformational states in the protein, not only by the length of insertion, but also by the nature of the amino acid insertion in the hinge region.

II-7-5. Investigating the effect of the insertion position on the conformational state

The next question is how a change in the position of insertion can affect the conformations of these domain-swapped dimer proteins. To tackle this question, we decided to change the position of insertion to between R58 and N59 instead of between F57 and R58. The results of the insertion yields Q108K:T51D:A28C:L36C:R58:<u>H:H:H</u>:N59 (**3H'**) and Q108K:T51D:A28C:L36C:R58:<u>H:H:H:H:H:H</u>:N59 (**6H'**). The first construct was not solubly expressed; however, the second construct was successfully expressed and purified, and its crystal structure was determined at 2.19 Å resolution. As shown in Figures II-30g and II-30h, the overlay of the construct with **3-H** insertion between F57 and R58 demonstrated a novel conformation, which showed that in addition to the length and nature of the amino acid, the position of insertion also effects the overall conformation of the proteins.

II-7-6. The effect of insertion on the monomer fold

Our final question is how these insertions would affect the monomeric form of the protein. We added three histidines and six histidines to **WT**-hCRBPII, leading to F57:H:H:H:R58 (**WT-3H**) and F57<u>:H:H:H:H:H:H:</u>R58 (**WT-6H**) constructs. The constructs were purified and yielded substantial quantities of monomer (The dimer was not detected.). The crystal

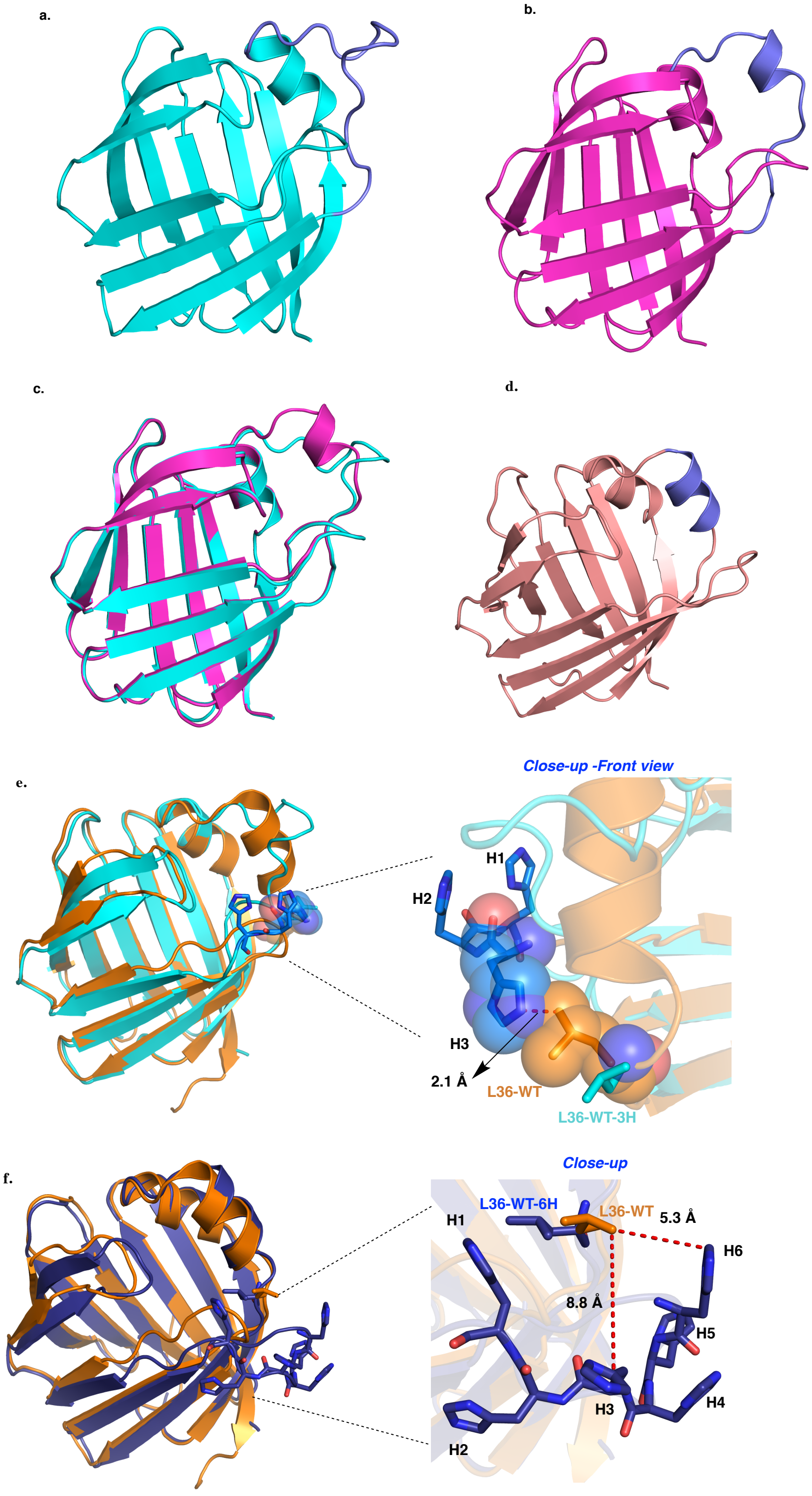


Figure II-31. a. Crystal structure of **WT-3H**, chain B. The melted α-helix is shown in purple. b. Crystal structure of **WT-3H,** chain A. The melted α-helix is shown in purple. c. Overlay of two chains of **WT-3H.** d. Crystal structure of **WT-6H.** e. Overlay of **WT-3H** (cyan), and **WT**-hCRBPII (orange) (PDB code: 2RCQ) The inserted histidines were shown with blue. f. Overlay of **WT-6H** (dark blue), and **WT**-hCRBPII (orange).

structures of both of these proteins were determined at atomic resolution (1.3 and 1.1 Å, respectively). As shown in Figures II-31a and II-31b, the crystal structure of **WT-3H** shows alpha helix 2 to be melted.

To our knowledge, this is the first report of such a structure in the iLBP family, in which an α-helix, which is responsible for covering the ligand-binding site, melted, and that this drastic change was observed at atomic resolution. The **6H** inserted structure represented a similar α-helix, as expected for hCRBPII, but with a different orientation in the loop region (Figure II-31d). As shown in the overlay of the wild type (PDB code: 2RCQ) vs. **WT-3H** and wild type vs. **WT-6H**, the orientation of the loop is highly different, and perhaps in **WT-3H**, this would require unfolding the α-helix where it sterically clashes with L36 in the α-helix conformation (Figure II-31e). However, the longer length of insertion gives rise to higher flexibility in the hinge loop region, which leads to no effect on the overall structure of the protein, Figure I31f.

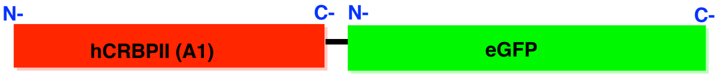
The take-home lesson from this study is that the hinge-loop region in proteins, especially those prone to domain swapping, can be used as a tool to explore different conformational states. This leads to a new, evolving scaffold with novel conformations and functional features produced only through economic changes in the protein sequence.

The effects of the number, location, and nature of amino acid insertion, as observed in atomic-resolution crystal structures, are quite drastic. This study highlights the high potency of the hinge-loop region to engineer novel proteins with new functions using a protein redesign strategy for protein engineering communities.

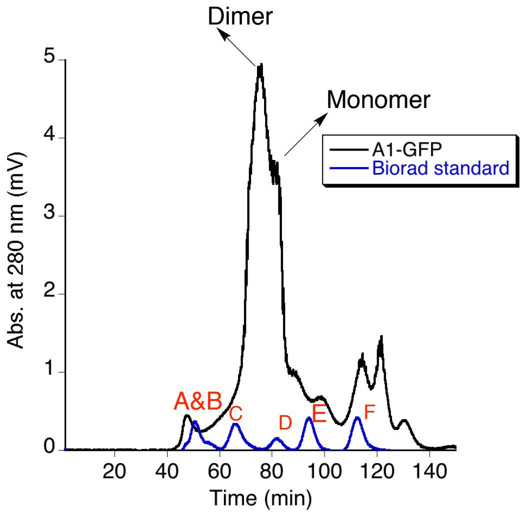
II-8. Using the hCRBPII domain-swapped dimer as a functional regulator

As mentioned in the introduction, section II-1, many natural protein conformational switches have been employed as functional regulators. Most of the efforts to create such systems, as cited earlier, are centered on domain insertion, where the circularly permutated versions of the target proteins are inserted inside the domains of the protein conformational switches, where the largest conformational changes occur. However, this strategy suffers from an important pitfall, in which many proteins might not fold properly when another protein is inserted inside their domains. As a result, this strategy may increase the risk of protein aggregation and complicate the protein design strategy.[66] Creating new protein conformational switches with a large conformational change in the c- and n-termini would address this problem. As mentioned in sections II-3 to II-6, we created a large conformational change in the c-termini of the protein upon the addition of retinal, fatty acid, merocyanine, and a change in the reduction potential of environment (Figures II-17d, II-22, II-26). The next step toward creating a functional regulator is to first investigate whether or not the domain-swapped dimer is still formed after fusing another protein to the c-termini of our protein conformational switch. To perform this investigation, a **A1**-GFP construct was created in pET17b. (The fusion construct of hCRBPII was made first by Dr. Santos, and a few mutations were imposed to create the **A1**-GFP construct). The protein was solubly expressed and, fortuitously, as mentioned in the experimental section, it was purified using ion exchange chromatography and size exclusion chromatography. The UV-vis spectrum of the fused protein and the green color of the purified protein reveal that the GFP expressed solubly. To elucidate whether hCRBPII expressed as a dimer or not, the biorad standard was carried out on the column and the retention times of standard proteins were compared to the **A1-GFP** retention time**.** As shown in Figure II-32, the result demonstrates dimer as well as monomer expression of **A1-GFP** (86 kDa) in the presence of n-terminal fused eGFP. This observation further highlights that the domain-swapped dimer is a robust scaffold and does not dissociate in the presence of a fusion protein in the c-terminus. Therefore, as a result of a large conformational change observed in the c-terminus, the dimer can be used as a convenient platform for functional switch-related applications.

II-9. Does the domain-swapped dimer of hCRBPII possess a novel functionality in humans?



a.



b.

Biorad gel filteraction standard components

Component		Molecular Weight*
Thyroglobulin (bovine)	A&B	670,000
γ-globulin (bovine)	C	158,000
Ovalbumin (chicken)	D	44,000
Myoglobin (horse)	E	17,000
Vitamin B_{12}	F	1,350

Figure II-32. a. EGFP that is fused to the c-terminus of **A1** hCRBPII. The gel filtration result demonstrates the production of the dimer in the presence of GFP when the retention time is compared to the biorad gel filtration standard. b. Biorad gel filtration standard components.

Recently allosteric regulation for some fatty acid binding proteins (FABPs) and cellular retinoic acid binding proteins (CRABPs), namely FABP4[53] and CRABPII,[67] was proposed. It was suggested in both FABP4 and CRABPII that ligand binding gives rise to a conformational change in the protein, resulting in a translocation of FABP4 and CRABPII to the nucleus. This then facilitates the delivery of the fatty acid and retinoic acid to their respective nuclear receptors.[68-69] Nonetheless, the crystal structures of the apo and holo iLBP monomers show small conformational variations upon ligand binding.[46, 70-]

[74] Based on the high-resolution crystal structures cited above, a large reproducible conformational change happens in the relative orientation of two dimer domains in the hCRBPII DS dimer upon ligand binding, demonstrating high sensitivity of this fold for ligand binding. Based on these findings, we wondered, based on the inherent allostery of the DS dimer, if the DS dimer fold could have some physiological relevance.

Subsequently we identified a single mutation, T51A that substantially increases the dimer/monomer ratio of hCRBPII. Intriguingly, the T51A variant is found in the human population, though no phenotype has been attributed to it thus far. We successfully determined the crystal structure of the dimer at 1.50 Å resolution and found it to form the same symmetric dimer every other apo-DS dimer adopts. This observation, along with the discovery of fatty acid binding for hCRBPII, as mentioned in section II-5, further suggest that the DS dimer of hCRBPII might have a distinct biological function in humans. The validation of such a mechanism will require further functional exploration of the iLBP family members *in vivo.*

II-10. Conclusion and future direction

Using the high-resolution structural evaluation and interrogation approach, we created a novel protein conformational switch that can be activated by ligand binding. We elucidated the mechanism of this protein conformational switch and then, by employing an engineered disulfide cross-link, we expanded the number of conformational readouts of the system. Furthermore, to define a function for the protein conformational switch, we generated a novel metal binding site in the interface of two dimer domains that can be regulated by the conformational readouts of the system. Using biochemical and biophysical techniques, we showed the structure and binding affinity of the metal in each conformational state of the protein. Additionally, we demonstrated the ability to drive the conformational change using fatty acids and synthetic chromophores. In the latter case, the effect of binding was investigated using a merocyanine fluorophore, where the second swapping observed at atomic resolution gave rise to the largest conformational change observed so far. The special nature of this chromophore as a photobleachable ligand introduces the possibility of creating novel optogenetic conformational switches. Further improvement of such a concept can be achieved by using retinal isomerization, where the two isomers can generate different conformational outcomes in the protein, or by using other light-sensitive ligands, such as azobenzene-based chromophores.

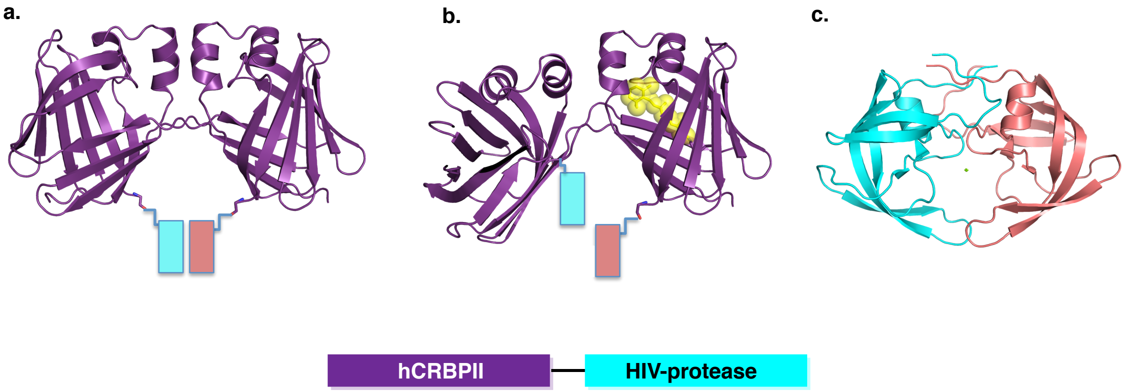


Figure II-33. The schematic design of HIV-protease fused to hCRBPII. a. Apo form of hCRBPII, where presumably HIV protease is in active form. b. Ligand-bound form of HCRBPII, which presumably causes a perturbation in the orientations of two dimer units. c. HIV protease crystal structure. PDB code: 2pc0.

We can take advantage of our highly-flexible protein conformational switch to make a new classes of functional regulators. For instance, a new construct has been made with the Q108K:T51D variant of hCRBPII in the n-terminus and HIV-protease in the c-terminus, connected by three Gly residues. HIV-protease is an important enzyme for the maturation of the HIV virus and functions as a homo dimer (Figure II-33). Through side-by-side protein/protein interaction, the active site of the enzyme is created when the two aspartates are in close proximity to one another. The connection of HIV-protease to the c-terminus of our protein (**hCRBPII-HIV**) may allow us to develop an allosterically regulated protease. Ligand binding would alter the conformations of our protein and subsequently the two monomeric forms of HIV-protease and perturb the active site of the enzyme. The **hCRBPII-HIV** construct would aid us in understanding if we can alter the function of an enzyme through our protein conformational switch. Additionally, this phenomenon can be investigated in the future by attaching our protein conformational switch to transcription factors, especially those creating homo dimers to elicit novel molecular tools in biology with higher flexibility of signal implementation.

II-11. Experimental Procedures

II-11-1. PCR protocol

Table II-4. PCR condition.

Template (DNA plasmid)	70 ng (× µl)
Primer forward	20 pmol (y µl)
Primer reverse	20 pmol (z µl)
dNTP	1 µl
10X pfu buffer	5 µl
pfu Turbo (DNA polymerase)	1 µl
DI water	43-(x+y+z) µl

Table II-5. PCR program.

	PCR program	
1x	94 °C	3 min
	94 °C	20 sec
20x	temperature 3-5 °C lower than primer melting temperature	50 sec
	72° C	4 min 30 sec (3 min 30 Sec for CRBPII)
1x	72 °C	10 min

II-11-2. List of PCR primers:

Site-directed mutagenesis was performed using K-CRBPII or KL-CRBPII-PET17b, which were generated in our previous work.[48, 75]

T51D

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

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

T51F

Forward: 5'-GGTGATAACTTCAAG**TTC**AAAACCACTAGCAC-3'

Reverse: 5'-GTGCTAGTGGTTTT**GAA**CTTGAAGTTATCACC-3'

T51W

Forward: 5'- GGTGATAACTTCAAG**TGG**AAAACCACTAGCAC -3'

Reverse: 5'- GTGCTAGTGGTTTT**TCC**ACTTGAAGTTATCACC -3'

A28C

Forward: 5’- GATATTGATTTT**<u>TGC</u>**ACCCGCAAG-3’

Reverse: 5’- CTTGCGGGT**<u>GCA</u>**AAAATCAATATC-3’

A28H

Forward: 5’- GATATTGATTTT**<u>CAT</u>**ACCCGCAAG-3’

Reverse: 5’- CTTGCGGGT**<u>ATG</u>**AAAATCAATATC-3’

Y60A

Forward: 5’-CACATTCCGCAAC**<u>GCT</u>**GATGTGGATTTCAC-3’

Reverse: 5’-GTGAAATCCACATC**<u>AGC</u>**GTTGCGGAATGTG-3’

F57H

Forward: 5’-ACCACTAGCACA**<u>CAT</u>**CGCAAC-3’

Reverse: 5’-GTTGCG**<u>ATG</u>**TGTGCTAGTGGT -3’

L36C

Forward: 5’-GCAGTACGT**<u>TGT</u>**ACTCAGACG -3’

Reverse: 5’-CGTCTGAGT**<u>ACA</u>**ACGTACTGC -3’

II-11-3. Hinge loop insertion primers and PCR protocol

Site-directed mutagenesis was performed using Q108K:T51D:A28C:L36C-CRBPII-PET17b. The optimized primers for Q5 site directed mutagenesis were designed using the NEBase Changer.

F57<u>:H</u>:R58

Forward: 5’- CATCGCAACTATGATGTGGATTTC-3’

Reverse: 5’- GAATGTGCTAGTGGTTTTTG-3’

F57<u>:H:H:H</u>:R58

Forward: 5’- TCATCGCAACTATGATGTGGATTTC-3’

Reverse: 5’- TGATGGAATGTGCTAGTGGTTTTTG-3’

F57:H:H:H:H:H:H:R58

Forward: 5’- CACCACCACCGCAACTATGATGTGGATTTC-3’

Reverse: 5’- ATGATGATGGAATGTGCTAGTGGTTTTTG -3’

F57:W:W:W:R58

Forward: 5’- GTGGCGCAACTATGATGTGGATTTC -3’

Reverse: 5’- CACCAGAATGTGCTAGTGGTTTTTG-3’

F57:W:W:W:W:W:W:R58

Forward: 5’- TGGTGGTGGCGCAACTATGATGTGGATTTC-3’

Reverse: 5’- CCACCACCAGAATGTGCTAGTGGTTTTTG-3’

R58:H:H:H:N59

Forward: 5’- TCATAACTATGATGTGGATTTCACTGTTG-3’

Reverse: 5’- TGATGGCGGAATGTGCTAGTGGT-3’

R58:H:H:H:H:H:H:N59

Forward: 5’-CATCATCATAACTATGAT-3’

Reverse: 5’-ATGATGATGGCGGAATGTGCTAGTGGT-3’

II-11-4. Transformation of PCR product

DH5α competent cells (50 µL) were thawed on ice, and then 8 µL of PCR product was added to the sample. After 30 minutes incubation on ice, the samples were heat-shocked at 42℃ for 45 seconds and 450 µL of LB was added. Then the mixture solution was incubated at 37℃ for 30 minutes. The cells were spread on an LB agar plate containing ampicillin (100 µg/mL) and incubated at 37℃ overnight.

II-11-5. DNA Purification

A single colony was utilized to inoculate 5 mL of LB solution containing ampicillin (100 μg/mL). The cell culture was grown at 37℃ overnight. Then it was centrifuged at 13,000 rpm for one minute. DNA isolation was performed by employing Promega Wizard® Plus SV Miniprep (A1330). The sample was then sequenced by the Research Technology Support Facility at Michigan State University. A T7 primer was used for sequencing.

II-11-6. Expression of hCRBPII mutants

For the expression, the gene was transformed into BL21 (DE3) pLysS (InvitrogenTM) *E. coli* competent cells. A protocol similar to the one mentioned for the transformation of PCR products was employed. The only difference was the use of ampicillin 100 μg/mL as antibiotic. To inoculate 1 L of LB with ampicillin (100 μg/mL), a single colony was used. After the cell culture was grown at 37 ° C while shaking for 8 hours, the overexpression was induced by adding 1 mL of 1M IPTG solution into 1 L cell culture (overall concentration 1.0 mM of IPTG). The solution was shaken at 16℃ for 36 hours. Q108K:T51D:A28C and Q108R:T51D were expressed overnight at 23 ° C and 16 ° C respectively.

II-11-7. Protein isolation and purification

The cells were harvested by centrifuging for 20 min at 5,000 rpm at 4℃. After removal of the supernatant, the cells were resuspended in 60 mL of Tris buffer (10 mM Tris. HCl, pH=8.0). Then the resuspended cells lysed by ultrasonication (Power 80%, 1 min x 3) and centrifuged at 4℃ (10,000 rpm, 20 min).

The ion exchange chromatography Q SepharoseTM Fast Flow resin, was used for purification of proteins at 4℃. Further purification was performed with Fast Protein Liquid

Chromatography. In the first step of purification with Q Sepharose Fast Flow resin (GE Health Sciences), the bound protein was washed with 10 mM Tris.HCl, pH = 8.0 and then eluted with 10 mM Tris. HCl, 150 mM NaCl, pH = 8.0. The pure fractions desalted using concentrators against 10 mM Tris pH=8.0 buffer, and loaded on a second anion exchange column (15Q, GE Health Sciences, BioLogic DuoFlow system). The fractions were separately concentrated at 4°C to 8 mg/mL and further purified by size exclusion chromatography (Superdex 75 16/760 column) in a buffer containing 10 mM Tris. HCl, 150 mM NaCl, pH = 8.0.

II-11-8. Crystallization, data collection, and refinement

Each mutant protein was concentrated to between 6 to 8 mg/mL in a buffer containing 10 mM Tris.HCl, 150 mM NaCl, pH = 8.0. Crystals were grown using the hanging drop vapor diffusion method using 1 μL of protein solution and 1 μL of crystallization solution in the drop and 1 mL of crystallization solution in the reservoir. Q108K:K40L:T51F and Q108K:K40L:T51W apo crystals were obtained in the presence of 2 eq. retinol. However, the overall conformations of these mutants show the normal apo symmetrical dimer, and no evident electron density for retinol has been observed.

For the crystallization of holo DS dimer, proteins were incubated with 4 equivalents of retinal. All of the crystallization trials with retinal were wrapped in aluminum foil to protect the light sensitive chromophore. The crystallization was done under following conditions: solution of 25-30% PEG 4000, 0.1 M sodium acetate pH=4.0-4.8, 0.1 M ammonium acetate. The crystals were soaked in a cryoprotectant solution (30% PEG 4000, 0.1 M sodium acetate pH = 4.5, 0.1 M ammonium acetate with 20% glycerol), flash frozen in loops in liquid nitrogen and stored in a liquid nitrogen Dewar prior to data collection.

For crystallization of zinc bound Q108K:T51D:A28C:L36C:F57H, 7.8 mg/ml (252 μM) of the protein was incubated with 540 μM zinc acetate dihydrate (purchased from Fluka). After about 20 minutes of incubation, the crystallization was performed. The protein crystals grew after 2 days at 0.2 M ammonium acetate, 0.1 M Bis-Tris:HCl, pH 5.5, and 25% PEG 3350 and 0.1 M Bis-Tris:HCl, pH 5.5, 25% PEG 3350.

To obtain the reduced form crystals of Q108K:T51D:A28C:L36C:F57H, the same concentration of the protein (252μM) was incubated with 540 μM of iron (II) sulfate hepta hydrate for 5 minutes (purchased from J.T. Baker Chemical Company). The protein crystals grew after 2-4 days at 0.1 M Bis-Tris:HCl, pH 5.5, 25% PEG3350.

Fatty acid bound crystal structures

For the crystallization of Q108R:T51D and WT, 8-10 mg/ml of protein was incubated with 4 equivalents of fatty acid (either linoleic and arachidonic acid). The best crystals grew using a crystallization solution of 25-30% PEG 4000, 0.1 M sodium acetate pH=4.0-4.8, 0.1 M ammonium acetate, 28% PEG 400.

A1 crystal structures

The crystal structures of apo and holo retinal bound **A1**, as well as Q108K:T51D (~8mg/ml) was crystallized with 4 equivalent of merocyanine under 25-30% PEG 4000, and 0.1 M sodium acetate pH=4.0-4.8, 0.1 M ammonium acetate. The merocyanine bound crystal of **A1** was obtained by incubation of 4 equivalent of ligand with 10 mg/ml protein under two conditions using robot screening: 0.2 M calcium acetate, 0.1 M MES:NaOH, pH=6, 20% PEG 8000 and 0.2 M calcium chloride, 0.1 M HEPES, NaOH pH=7.5. **A1**-K40F crystals bound with merocyanine grew at 4 M sodium formate.

Hinge loop insertion mutants crystallization conditions

Q108K:T51D:A28C:L36C, 1H, **6H**, **3W** WT-**3H**, and WT-**6H** were crystallized under 25-30% PEG 4000, 0.1 M sodium acetate pH=4.0-4.8, 0.1 M ammonium acetate. The **3H** crystals grew under following conditions: 200 μl of protein (~11.7 mg/ml) was incubated with 500 μM Cobalt (II) acetate tetrahydrate and crystallized with 0.2 M ammonium acetate, 0.1 M Bis-Tris: HCl, pH 6.5, and 25% PEG 3350. The 3H’ and 6H also crystallized under 0.2 M sodium Fluoride, 20% PEG 3350.

For Diffraction data were collected at the Advanced Photon Source (APS) (Argonne National Laboratory IL) LS-CAT, (sector 21-ID-D, F, G) using a MAR300 and MAR350 detectors, 1.00Å wavelength radiation at 100K. Also, for the zinc bound protein anomalous data was collected above the zinc edge. The initial diffraction data were indexed, processed and scaled using the HKL2000 software package.[5] The structures were solved by molecular replacement using PHASER in PHENIX and hCRBPII (PDB entry 2RCQ) as a search model.[6] The initial electron density map was produced Phaser-MR in PHENIX.[7, 8] Model rebuilding, placement of water molecules etc. were done using COOT.[9] The structures were refined using the PHENIX program packages. The chromophore was built using the JLIGAND program to generate restraints.[10,11]

II-11-9. UV-vis measurements

UV-vis spectroscopy was performed using a Cary 300 and Cary 150 Bio WinUV UV-Vis spectrophotometer (Varian Instruments). The samples were prepared in phosphate buffer. The experiment was performed using 0.5 equivalents of retinoid. The protonated Schiff Base (PSB) has Absorptions corresponding to λ_{max}> 450 nm, while deprotonated imine peaks (SB) shows up at λ_{max} ~368 nm. The unbound retinal absorbs at λ_{max}~380 nm. All-*trans*-retinal (purchased from Sigma, ε = 48,000 M^{-1} cm^{-1}, λ_{max} =380 nm.), Q108K:T51D ε_{280nm} =57,050 $M^{-1}cm^{-1}$, Q108K:K40L:T51F ε_{280nm}= 64,540 $M^{-1}cm^{-1}$; Q108K:K40L:T51F=75,400 $M^{-1}cm^{-1}$; Q108K:K40D ε_{280nm}=58,690 $M^{-1}cm^{-1}$; Q108K:T51D:A28C:L36C:F57H; ε_{280nm} = 51,485 $M^{-1}cm^{-1}$; Q108K:T51D:A28C ε_{280nm} = 56,720 $M^{-1}cm^{-1}$, Q108K:K40L:Y60L:A28C ε_{280nm} =49,397 $M^{-1}cm^{-1.}$

II-11-10. Fluorescence quenching assay

The assay was done as previously reported.[76] 0.01% gelatin containing PBS (4 mM NaH_2PO_4, 16 mM Na_2HPO_4, 150 mM NaCl, pH=7.3) was used as a buffer. 3 mL of 500 nM protein in PBS buffer was added to the cell. The sample was excited at 283 nm with a slit width about 1.5 nm. The fluorescence is measured at the peak maximum, about 345 nm. The chromophore was added at different equivalents from a ~1.5 mM stock (in DMSO) sample maintained in the dark. The results are plotted as concentration of chromophore versus relative fluorescence intensity.

II-11-11. CD spectroscopy

CD spectra were recorded on a JASCO J-810 spectropolarimeter. The protein concentrations were adjusted in PBS buffer (purchased from Sigma Aldrich). The ethanolic stock solution of retinal and PBS stock solution of DTT were prepared. The CD changes (mdeg) were monitored upon adding retinal and DTT by using CD spectrometer. Protein concentrations were as follows:

Q108K:T51D (dimer): 4.76 μM, Q108K:K40L:T51V (monomer): 5.30 μM, Q108K:T51D:A28C (dimer): 4.05 μM, Q108K:K40L:T51F:Y60A (dimer): 5.6 μM, and Q108K:K40L:Y60L:A28C (dimer): 2.3 μM.

II-11-12. Radioactive $^{65}ZnCl_2$ labeling assay

For ultrafiltration zinc-binding experiments, radioactive $^{65}ZnCl_2$ (radionuclide purity: 99%) was purchased from PerkinElmer and was diluted with buffer and non-radioactive $ZnCl_2$ (10 mM Tris. HCl/150 mM NaCl, pH 7.0) to make a radioactive 259 μM $ZnCl_2$ stock solution.

Samples were prepared in 1.5 mL centrifugation tubes to be 250 μM total and contain 50.0 μM protein and varying concentrations of $^{65}ZnCl_2$ in buffer. To create a saturation-binding curve, the $^{65}Zn^{2+}$ concentrations were varied from 10-100 μM. For statistical comparisons of the relative binding affinities between different modifications of proteins for $^{65}Zn^{2+}$, samples were made with 40 μM $^{65}Zn^{2+}$ and 50 μM protein, with three to five replicates of each. Then, 220 μL of each sample was pipetted into an ultrafiltration centrifugation device with a 10 kDa MWCO size-exclusion membrane. The samples were centrifuged at 8000 rpm for 3 minutes to allow ~20 μL of ultrafiltrate to collect in the bottom of the tube, the concentration of which is considered the free-Zn^{2+} in the sample. The volume separated from the sample is low enough so as to not disturb the equilibrium of the binding system. Then, a 15 μL aliquot of ultrafiltrate sample was transferred into a separate tube for analysis. Calibration standards containing 15 μL of known Zn^{2+} concentrations were also prepared in separate tubes. Then, all standards and samples were analyzed by counting the emitted gamma radiation at (1115 keV) for 5 minutes per sample on a PerkinElmer 2480 Automatic Gamma Counter. The concentration of Zn^{2+} in the ultrafiltrate was quantified by comparing the counts-per-minute (CPM) for each sample to an external standards calibration curve. The concentration of Zn^{2+} in the ultrafiltrate was considered to be the free-Zn^{2+} concentration in the sample, while the bound-Zn^{2+} concentration was calculated as the free-Zn^{2+} subtracted from the total-Zn^{2+}. For the saturation-binding curve, free-Zn^{2+} vs bound-Zn^{2+} was plotted and analyzed via non-linear regression software (SigmaPlot 13.0) to calculate the equilibrium binding constant (K_d) and stoichiometry (B_{max}).

For comparing the different binding affinities between proteins, the global affinity constant, or nK_a of the binding system, was calculated from the free concentration of ligand. Reduced **AG1** (**Red-apo**) was prepared by adding 1mM BME in the buffer (10mM Tris.HCl, 150 mM NaCl, pH=7.0). Holo **AG1** (**Holo**) was obtained by adding 2 equivalents of ethanolic solution of all-*trans* retinal to **AG1**. Holo/reduced **AG1 (Red-holo)** then was obtained by adding 2 equivalents of ethanolic solution of retinal to the reduced **AG1**.

II-11-13. ITC experiment

Isothermal titration calorimetry (ITC) experiments were carried out using a PEAQ-ITC calorimeter (Malvern). To determine the K_d of apo **AG1** to Zn^{+2}, 1mM zinc acetate was titrated into the cell containing ~50 μM **AG1** proteins.

Titration experiments were performed at 25 °C. Both the protein and ligand solutions were centrifuged for 2 min at 10,000 g before each titration. In a typical experiment, the reference power was set at 10 μcal/sec, an initial 0.2 μl injection (disregarded in analyses) was followed by 19, 2-μl injections of titrant (Δt = 2.5 min). Experiments and controls with each combination were repeated at least once. The thermograms were blank-subtracted and analyzed using the MicroCal PEAQ ITC analysis software and fitted to single-site binding models.

Table II-6. X-ray crystallography data and refinement statistics of the holo dimer Q108K:T51D, Q108K:K40L:T51F, Q108K:K40L:T51W structures.

	Q108K:T51D	**Q108K:K40L:T51F**	**Q108K:K40L:T51W**
Space group	$P2_12_12_1$	$P\,2_1$	$P\,2_1$
a(Å)	66.19	36.48	36.33
b(Å)	73.79	63.28	63.14
c (Å)	353.40	117.78	118.22
α(°)	90	90	90
β(°)	90	96	96
δ(°)	90	90	90
Molecules per Asymmetric Unit	12	4	4
Total reflections	4184309	1580011	1638156
Unique	99384	2164	29781
Completeness(%	91.86 (82.88)[a]	98.1(96.6)[a]	96.25 (89.85) [a]
Average I/σ	22	32.2(2.5)[a]	13.6 (2.2)[a]
R_{merge}(%)	5 (80)[a]	6.8 (40.0) [a]	4.5 (40.0)[a]
Resolution (Å)	45.46 - 2.061	14.87 - 2.17	39.17-2.11
(Last Shell)	(2.13 - 2.06)[a]	(1.85 - 1.80) [a]	(2.183-2.11)[a]
R_{work}/ R_{free} (%)	22.23/27.35	20.49/ 29.13	19.70/27.13
RMSD from ideal value			
Bond Length (Å)	0.009	0.008	0.008
Bond Angle (°)	1.13	1.12	1.19
Average B factor	40.00	51.40	38.3
Number of water molecules	650	174	228
PDB ID	6E5S	6MLB	6MKV

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

Table II-7. X-ray crystallographic data and refinement statistics for holo dimer hCRBPII Q108K:K40L:T51F:Y60A Q108K:T51D:A28H, hCRBPII-WT, and Q108K:T51D:A28C mutants.

	Q108K:K40L:T51F:Y60A	**Q108K:T51D:A28H**	**hCRBPII WT**	**Q108K:T51D:A28C**
Space group	$P\ 2_12_1\ 2$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a(Å)	36.57	65.87	65.82	67.74
b(Å)	60.55	73.99	74.4	72.06
c (Å)	63.76	349.99	352.55	112.25
α(°)	90	90	90	90
β(°)	90	90	90	90
δ(°)	90	90	90	90
Molecules per Asymmetric Unit	1	12	12	4
Total reflection	846912	1539077	6589507	1021462
Unique Reflection	11665	49017	24088	15658
Completeness	99.72 (99.38)[a]	90.86 (90.57)[a]	83.47 (91.81)	99.96 (100.00)
Average I/σ	16.1(2.2)[a]	12 (2.07)	13 (2.20)[a]	11.5 (2.0)[a]
R_{merge}(%)	6.0 (74)	9.8 (69)	11.2 (93.5)	14.6 (88.5)
Resolution (Å)	31.88 -1.90	14.99 - 2.60	14.99 - 3.30	9.36 - 2.70
(Last Shell)	(1.97 - 1.90)[a]	(2.69 - 2.60)[a]	(3.42 - 3.30)[a]	(2.80 - 2.70)[a]
R_{work}/ R_{free} (%)	21.04/25.59	21.0/30.04	25.7/30.52	21.99-29.23
RMSD from ideal value				
Bond Length (Å)	0.007	0.009	0.002	0.004
Bond Angle (°)	0.82	1.11	0.51	0.89
Average B factor	44.16	38.5	74.6	40.08
Number of water molecules	43	171	89	39
PDB ID	6E6L	6MCU	6MCV	6E7M

---

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

Table II-8. X-ray crystallographic data and refinement statistics for apo dimer hCRBPII Q108K:T51D, Q108K:K40L:T51F, Q108K:K40L:T51W, and Q108K:T51D:A28H.

	Q108K:T51D	**Q108K:K40L :T51F**	**Q108K:K40L: T51W**	**Q108K:T51D :A28H**
Space group	$P2_12_12_1$	$P\ 2_1\ 2_12$	$P\ 2_1\ 2_12$	$P\ 2_1\ 2_12$
a(Å)	60.218	63.80	36.89	36.7
b(Å)	63.757	36.72	60.86	60.27
c (Å)	36.739	60.51	64.16	63.40
α(°)	90	90	90	90
β(°)	90	90	90	90
δ(°)	90	90	90	90
Molecules per Asymmetric Unit	1	1	1	1
Total reflection	540742	1023617	1145094	195446
Unique Reflection	16231	19468	6023	101134
Completeness (%)	99.71 (97.45)[a]	99.90 (100.00)[a]	99.90 (100.00)[a]	99.91(99.30)[a]
Average I/σ	32 (1.0)	23.8 (2.8)[a]	11(2.9)[a]	25 (2.3)[a]
R_{merge} (%)	8.5 (94)	5.7(43)	9.1 (53)	8.9 (73)
Resolution (Å) (Last Shell)	36.74 - 1.70 (1.76 1.70)[a]	28.12 - 1.96 (2.036 - 1.96)	32.08 - 2.4 (2.49- 2.40)[a]	36.73 - 1.99 (2.06-1.99)[a]
R_{work}/ R_{free} (%)	19.58/23.30	20.30/26.88	24.94/17.8	21.74/27.64
RMSD from ideal value				
Bond Length (Å)	0.012	0.007	0.008	0.008
Bond Angle (°)	1.31	1.05	1.12	0.92
Average B factor	36.70	38.20	48.01	34.23
Number of water molecules	60	86	49	30
PDB ID	6E5E	6E50	6E51	6E5Q

---

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

Table II-9. X-ray crystallographic data and refinement statistics for apo dimer Q108K:T51D:A28C, zinc-bound Q108K:T51D:A28C:L36C:F57H, and reduced Q108K:T51D:A28C:L36C:F57H hCRBPII mutants.

	Q108K:T51D:A28C	**Q108K:T51D:A28C: L36C:F57H Zinc bound**	**Q108K:T51D: A28C:L36C:F57H Reduced form**
Space group	$P2_12_12_1$	$P2_12_12_1$	$P\ 2\ 2_12_1$
a(Å)	30.86	60.42	36.97
b(Å)	68.18	61.76	60.55
c (Å)	130.04	72.66	63.78
α(°)	90	90	90
β(°)	90	90	90
δ(°)	90	90	90
Molecules per Asymmetric Unit	2	2	1
Total reflection	685586	4246024	366691
Unique Reflection	8997	32802 (3193)	5891 (501)
Completeness (%)	99.94 (100.00)[a]	96.19 (95.68)	99.5(99.7)
Average I/σ	10.63 (2.30)	22 (2.30)	22.6 (2.30)
R_{merge} (%)	18 (81)	15.4 (180)	6.6 (47)
Resolution (Å) (Last Shell)	36.58 - 2.6 (2.69 - 2.6)[a]	37.12 - 1.64 (1.70 - 1.64)[a]	31.98 - 2.40 (2.48 - 2.40)[a]
R_{work}/ R_{free} (%)	24.18/30.54	20.30/24.29	18.29/25.92
RMSD from ideal value			
Bond Length (Å)	0.003	0.007	0.007
Bond Angle (°)	0.75	0.93	1.03
Average B factor	26.70	25.50	57.39
Number of water molecules		211	12
PDB ID	6E5R	TBD	TBD

---

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

Table II-10. X-ray crystallography data and refinement statistics of the apo dimer **1H**, **3H**, **6H**, and **3W** mutants of hCRBPII.

	1H	**3H**	**6H**	**3W**
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	I4
a(Å)	39.71	36.93	35.80	140.65
b(Å)	56.58	69.98	76.58	140.65
c (Å)	136.197	106.34	110.52	63.95
α(°)	90	90	90	90
β(°)	90	90	96	90
δ(°)	90	90	90	90
Molecules per Asymmetric Unit	2	2	2	4
Total reflections	2496404	1156335	580019	725079
Unique Reflection	22274 (2048)	21400 (2032)	10638 (947)	25561(2499)
Completeness (%)	98.15 (91.58)[a]	98.51(95.71)[a]	98.48 (91.59)[a]	99.83 (98.85)[a]
Average I/σ	33.3 (2.52)	21.7 (2.4)	13.76 (2.01)	30.2 (2.81)
R_{merge} (%)	10.2 (93.8)	7.3 (68.6)	9.0 (45.8)	8.8 (70.1)
Resolution(Å)	38.12-1.96	33.24- 1.92	36.18 - 2.52	44.48-2.37
(Last Shell)	(2.03 -1.96)	(1.991 - 1.92)	(2.612 - 2.52)	(2.45 - 2.37)
R_{work}/ R_{free} (%)	22.26/27.15	20.57/24.92	18.13/25.07	20.08/28.01
RMSD from ideal value				
Bond Length (Å)	0.007	0.009	0.008	0.009
Bond Angle (°)	0.81	1.01	0.93	0.94
Average B factor	37.80	33.01	37.35	50.07
Number of water molecules	67	52	44	35
PDB ID	TBD	TBD	TBD	TBD

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

Table II-11. X-ray crystallographic data and refinement statistics for apo hCRBPII **3H'**, **WT-3H**, **WT-6H**, and Q108K:T51D:A28C:L36C hCRBPII mutants.

	3H'	**WT-3H**	**WT-6H**	**Q108K:T51D: A28C:L36C**
Space group	$P\ 2_12_1 2_1$	$P2_1$	P_1	$P2_1$
a(Å)	36.95	48.52	29.61	36.50
b(Å)	69.90	47.81	36.93	60.12
c (Å)	107.25	53.22	68.51	63.07
α(°)	90.00	90.00	73.44	90.00
β(°)	90.00	92.61	83.92	93.70
δ(°)	90.00	90.00	66.06	90.00
Molecules per Asymmetric Unit	2	2	2	2
Total reflection	3054247	2700610	728017	851776
Unique Reflection	14150 (1272)	55387 (5299)	101926 (9598)	17684 (1739)
Completeness (%)	94.80 (87.36)	96.44 (92.43)	93.17 (87.33)	92.63 (91.57)
Average I/σ	11.5 (2.8)	3.3 (2.1)	113 (3.3)	18.03 (2.8)
R_{merge} (%)	7.5 (26.4)	6.2 (62.1)	6.0 (50.7)	7.0 (42.0)
Resolution (Å) (Last Shell)	34.94 - 2.19 (2.27 - 2.19)	29.09 - 1.319 (1.37 - 1.32)	27.06- 1.08 (1.12 - 1.08)	31.47-1.98 (2.052 - 1.98)
R_{work}/ R_{free} (%)	19.84/27.17	21.68/19.42	21.35/20.70	24.86/29.77
Bond Length (Å)	0.008	0.005	0.005	0.008
Bond Angle (°)	0.84	0.77	0.80	1.05
Average B factor	23.86	17.17	15.10	31.30
Number of water molecules	53	339	295	30
PDB ID	TBD	TBD	TBD	TBD

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

Table II-12. X-ray crystallographic data and refinement statistics for domain-swapped dimer of **A1** mutant in apo, retinal bound, and merocyanine bound forms.

	A1-Apo	**A1/retinal**	**A1/merocyanine**
Space group	$P2_12_12$	$P2_12_12_1$	$P2_12_12_1$
a(Å)	60.78	65.51	60.15
b(Å)	64.03	73.80	60.63
c (Å)	36.80	350.12	76.95
α(°)	90	90	90
β(°)	90	90	90
δ(°)	90	90	90
Molecules per Asymmetric Unit	1	12	2
Total reflection	765009	2700610	1971397
Unique Reflection	4903 (477)	55387 (5299)	25564 (2465)
Completeness (%)	99.94 (99.79)	96.44 (92.43)	97.53 (95.35)
Average I/σ	24.7 (2.44)	13.8 (1.6)	25.2 (2.3)
R_{merge} (%)	10 (81)	15 (180)	11.3 (144.2)
Resolution (Å)	36.8 - 2.57	48.99-3.031	47.39-1.81
(Last Shell)	(2.66 - 2.57)	(3.14-3.03)	(1.88 - 1.81)
R_{work}/ R_{free} (%)	21.78/27.92	23.55/32.56	18.84/22.98
Bond Length (Å)	0.008	0.006	0.008
Bond Angle (°)	1.09	0.62	0.93
Average B factor	51.82	59.01	25.03
Number of water molecules	10	4	233
PDB ID	TBD	TBD	TBD

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

Table II-13. X-ray crystallographic data and refinement statistics for Q108R:T51D mutant bound with fatty acids.

	Q108R:T51D /Linoleic acid	**Q108R:T51D/ Arachidonic acid**
Space group	$P2_12_12_1$	$P2_12_12_1$
a(Å)	64.425	63.90
b(Å)	72.802	72.16
c (Å)	353.76	353.50
α(°)	90	90
β(°)	90	90
δ(°)	90	90
Molecules per Asymmetric Unit	12	12
Total reflection	3432087	5903084
Unique Reflection	57772 (5222)	55387 (5299)
Completeness (%)	86.83 (79.48)	96.44(92.43)
Average I/σ	24.7 (13.8)	33.63 (2.15)
R_{merge} (%)	7.0 (51.2)	78 (183)
Resolution (Å)	43.49 - 2.40	47.84 -1.96
(Last Shell)	(2.48 - 2.40)	(2.034 - 1.96)
R_{work}/ R_{free} (%)	19.42/26.91	22.21/26.34
Bond Length (Å)	0.008	0.008
Bond Angle (°)	0.98	0.93
Average B factor	36.25	39.72
Number of water molecules	134	22
PDB ID	TBD	TBD

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

REFERENCES

REFERENCES

1. Fenton, A. W., Allostery: an illustrated definition for the 'second secret of life'. *Trends in Biochemical Sciences* **2008,** *33* (9), 420-5.

2. Raman, S., Systems approaches to understanding and designing allosteric proteins. *Biochemistry* **2018,** *57* (4), 376-382.

3. Ha, J. H.; Loh, S. N., Protein conformational switches: from nature to design. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2012,** *18* (26), 7984-99.

4. Andrianantoandro, E.; Basu, S.; Karig, D. K.; Weiss, R., Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology* **2006,** *2* (1).

5. Silvaroli, J. A.; Arne, J. M.; Chelstowska, S.; Kiser, P. D.; Banerjee, S.; Golczak, M., Ligand binding induces conformational changes in human cellular retinol-binding protein 1 (CRBP1) revealed by atomic resolution crystal structures. *The Journal of Biological Chemistry* **2016,** *291* (16), 8528-40.

6. Meister, G. E.; Joshi, N. S., An engineered calmodulin-based allosteric switch for Peptide biosensing. *Chembiochem : a European journal of chemical biology* **2013,** *14* (12), 1460-7.

7. Ke, W.; Laurent, A. H.; Armstrong, M. D.; Chen, Y.; Smith, W. E.; Liang, J.; Wright, C. M.; Ostermeier, M.; van den Akker, F., Structure of an Engineered β-Lactamase Maltose Binding Protein Fusion Protein: Insights into Heterotropic Allosteric Regulation. *PLOS ONE* **2012,** *7* (6), e39168.

8. Choi, J. H.; Zayats, M.; Searson, P. C.; Ostermeier, M., Electrochemical activation of engineered protein switches. *Biotechnology and Bioengineering* **2016,** *113* (2), 453-456.

9. DeGrave, A. J.; Ha, J.-H.; Loh, S. N.; Chong, L. T., Large enhancement of response times of a protein conformational switch by computational design. *Nature Communications* **2018,** *9* (1), 1013.

10. Kim, J. R.; Ostermeier, M., Modulation of effector affinity by hinge region mutations also modulates switching activity in an engineered allosteric TEM1 β-lactamase switch. *Archives of Biochemistry and Biophysics* **2006,** *446* (1), 44-51.

11. Banala, S.; Aper, S. J. A.; Schalk, W.; Merkx, M., Switchable Reporter Enzymes Based on Mutually Exclusive Domain Interactions Allow Antibody Detection Directly in Solution. *ACS Chemical Biology* **2013,** *8* (10), 2127-2132.

12. Teasley Hamorsky, K.; Ensor, C. M.; Wei, Y.; Daunert, S., A Bioluminescent Molecular Switch For Glucose. *Angewandte Chemie International Edition* **2008,** *47* (20), 3718-3721.

13. Dueber, J. E.; Yeh, B. J.; Chak, K.; Lim, W. A., Reprogramming Control of an Allosteric Signaling Switch Through Modular Recombination. *Science* **2003,** *301* (5641), 1904-1908.

14. Guntas, G.; Mansell, T. J.; Kim, J. R.; Ostermeier, M., Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (32), 11224-11229.

15. Ostermeier, M., Engineering allosteric protein switches by domain insertion. *Protein Engineering, Design and Selection* **2005,** *18* (8), 359-364.

16. Baird, G. S.; Zacharias, D. A.; Tsien, R. Y., Circular permutation and receptor insertion within green fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** *96* (20), 11241-6.

17. Grünewald, F. S., Periplasmic binding proteins in biosensing applications. In *Advances in Chemical Bioanalysis*, Matysik, F.-M., Ed. Springer International Publishing: Cham, 2014; pp 205-235.

18. Akerboom, J.; Carreras Calderon, N.; Tian, L.; Wabnig, S.; Prigge, M.; Tolo, J.; Gordus, A.; Orger, M. B.; Severi, K. E.; Macklin, J. J.; Patel, R.; Pulver, S. R.; Wardill, T. J.; Fischer, E.; Schuler, C.; Chen, T. W.; Sarkisyan, K. S.; Marvin, J. S.; Bargmann, C. I.; Kim, D. S.; Kugler, S.; Lagnado, L.; Hegemann, P.; Gottschalk, A.; Schreiter, E. R.; Looger, L. L., Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Frontiers in Molecular Neuroscience* **2013,** *6*, 2.

19. Ambroggio, X. I.; Kuhlman, B., Design of protein conformational switches. *Current Opinion in Structural Biology* **2006,** *16* (4), 525-30.

20. Mukherji, S.; van Oudenaarden, A., Synthetic biology: understanding biological design from synthetic circuits. *Nature Reviews Genetics* **2009,** *10*, 859.

21. Lormeau, C.; Rybiński, M.; Stelling, J., Multi-objective design of synthetic biological circuits**This publication is supported by the Swiss National Science Foundation as part of the NCCR Molecular Systems Engineering. *IFAC-PapersOnLine* **2017,** *50* (1), 9871-9876.

22. Miyawaki, A.; Llopis, J.; Heim, R.; McCaffery, J. M.; Adams, J. A.; Ikura, M.; Tsien, R. Y., Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. *Nature* **1997,** *388* (6645), 882-887.

23. Sulmann, S.; Dell'Orco, D.; Marino, V.; Behnen, P.; Koch, K.-W., Conformational changes in calcium-sensor proteins under molecular crowding conditions. *Chemistry – A European Journal* **2014,** *20* (22), 6756-6762.

24. Koide, S.; Huang, J., Generation of high-performance binding proteins for peptide motifs by affinity clamping. *Methods Enzymology* **2013,** *523*, 285-302.

25. Marvin, J. S.; Schreiter, E. R.; Echevarría, I. M.; Looger, L. L., A genetically encoded, high-signal-to-noise maltose sensor. *Proteins: Structure, Function, and Bioinformatics* **2011,** *79* (11), 3025-3036.

26. Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J., The ensemble nature of allostery. *Nature* **2014,** *508* (7496), 331-9.

27. Garcia-Pino, A.; Balasubramanian, S.; Wyns, L.; Gazit, E.; De Greve, H.; Magnuson, R. D.; Charlier, D.; van Nuland, N. A.; Loris, R., Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell* **2010,** *142* (1), 101-11.

28. Mitrea, D. M.; Parsons, L. S.; Loh, S. N., Engineering an artificial zymogen by alternate frame protein folding. *Proceedings of the National Academy of Sciences the United States of America* **2010,** *107* (7), 2824-2829.

29. Ha, J.-H.; Loh, S. N., Construction of Allosteric Protein Switches by Alternate Frame Folding and Intermolecular Fragment Exchange. In *Synthetic Protein Switches: Methods and Protocols*, Stein, V., Ed. Springer New York: New York, NY, 2017; pp 27-41.

30. Do, K.; Boxer, S. G., GFP variants with alternative β-strands and their application as light-driven protease sensors: a tale of two tails. *Journal of the American Chemical Society* **2013,** *135* (28), 10226-9.

31. Ha, J.-H.; Karchin, J. M.; Walker-Kopp, N.; Castañeda, C. A.; Loh, S. N., Engineered domain swapping as an on/off Switch for protein function. *Chemistry & Biology* **2015,** *22* (10), 1384-1393.

32. Stein, V.; Alexandrov, K., Protease-based synthetic sensing and signal amplification. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** *111* (45), 15934-9.

33. Greenwood, C.; Ruff, D.; Kirvell, S.; Johnson, G.; Dhillon, H. S.; Bustin, S. A., Proximity assays for sensitive quantification of proteins. *Biomolecular Detection and Quantification* **2015,** *4*, 10-6.

34. Canto, I.; Soh, U. J.; Trejo, J., Allosteric modulation of protease-activated receptor signaling. *Mini Reviews in Medicinal Chemistry* **2012,** *12* (9), 804-11.
35. Deisseroth, K., Optogenetics. *Nature Methods* **2011,** *8* (1), 26-29.

36. Wang, W.; Wildes, C. P.; Pattarabanjird, T.; Sanchez, M. I.; Glober, G. F.; Matthews, G. A.; Tye, K. M.; Ting, A. Y., A light- and calcium-gated transcription factor for imaging and manipulating activated neurons. *Nature Biotechnology* **2017,** *35* (9), 864-871.

37. Pudasaini, A.; El-Arab, K. K.; Zoltowski, B. D., LOV-based optogenetic devices: light-driven modules to impart photoregulated control of cellular signaling. *Frontiers in Molecular Biosciences* **2015,** *2*, 18.

38. Anderson, S.; Dragnea, V.; Masuda, S.; Ybe, J.; Moffat, K.; Bauer, C., Structure of a novel photoreceptor, the BLUF domain of AppA from Rhodobacter sphaeroides. *Biochemistry* **2005,** *44* (22), 7998-8005.

39. Glock, C.; Nagpal, J.; Gottschalk, A., Microbial Rhodopsin Optogenetic Tools: Application for Analyses of Synaptic Transmission and of Neuronal Network Activity in Behavior. *Methods in Molecular Biology (Clifton, N.J.)* **2015,** *1327*, 87-103.

40. Beharry, A. A.; Woolley, G. A., Azobenzene photoswitches for biomolecules. *Chemical Society Reviews* **2011,** *40* (8), 4422-37.

41. Khersonsky, O.; Fleishman, S. J., Incorporating an allosteric regulatory site in an antibody through backbone design. *Protein Science* **2017,** *26* (4), 807-813.

42. Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A., De novo design of an allosteric metalloprotein assembly with strained disulfide bonds. *Journal of the American Chemical Society* **2016,** *138* (40), 13163-13166.

43. Mitternacht, S.; Berezovsky, I. N., Binding leverage as a molecular basis for allosteric regulation. *PLOS Computational Biology* **2011,** *7* (9), e1002148.

44. Burger, W. A. C.; Sexton, P. M.; Christopoulos, A.; Thal, D. M., Toward an understanding of the structural basis of allostery in muscarinic acetylcholine receptors. *The Journal of General Physiology* **2018,** *150* (10), 1360-1372.

45. Schaap, F. G.; van der Vusse, G. J.; Glatz, J. F., Evolution of the family of intracellular lipid binding proteins in vertebrates. *Molecular and Cellular Biochemistry* **2002,** *239* (1-2), 69-77.

46. Vaezeslami, S.; Mathes, E.; Vasileiou, C.; Borhan, B.; Geiger, J. H., The Structure of Apo-wild-type cellular retinoic acid binding protein II at 1.4 Å and its relationship to ligand binding and nuclear translocation. *Journal of Molecular Biology* **2006,** *363* (3), 687-701.

47. Nossoni, Z.; Assar, Z.; Yapici, I.; Nosrati, M.; Wang, W.; Berbasova, T.; Vasileiou, C.; Borhan, B.; Geiger, J., Structures of holo wild-type human cellular retinol-binding

protein II (hCRBPII) bound to retinol and retinal. *Acta Crystallographica. Section D, Biological Crystallography* **2014,** *70* (Pt 12), 3226-32.

48. Assar, Z.; Nossoni, Z.; Wang, W.; Santos, E. M.; Kramer, K.; McCornack, C.; Vasileiou, C.; Borhan, B.; Geiger, J. H., Domain-swapped dimers of intracellular lipid-binding proteins: evidence for ordered folding intermediates. *Structure* **2016,** *24* (9), 1590-8.

49. Sanson, B.; Wang, T.; Sun, J.; Wang, L.; Kaczocha, M.; Ojima, I.; Deutsch, D.; Li, H., Crystallographic study of FABP5 as an intracellular endocannabinoid transporter. *Acta Crystallographica Section D* **2014,** *70* (2), 290-298.
50. Bennett, M. J.; Eisenberg, D., The evolving role of 3D domain swapping in proteins. *Structure* **2004,** *12* (8), 1339-1341.

51. Liang, J. Y.; Lipscomb, W. N., Binding of substrate CO_2 to the active site of human carbonic anhydrase II: a molecular dynamics study. *Proceedings of the National Academy of Sciences of the United States of America* **1990,** *87* (10), 3675-9.

52. Das, U. N., Biological significance of essential fatty acids. *The Journal of the Association of Physicians of India* **2006,** *54*, 309-19.

53. Gillilan, R. E.; Ayers, S. D.; Noy, N., Structural basis for activation of fatty acid-binding protein 4. *Journal of Molecular Biology* **2007,** *372* (5), 1246-60.

54. Napoli, J. L., Cellular retinoid binding-proteins, CRBP, CRABP, FABP5: Effects on retinoid metabolism, function and related diseases. *Pharmacology & Therapeutics* **2017,** *173*, 19-33.

55. Yapici, I.; Lee, K. S.; Berbasova, T.; Nosrati, M.; Jia, X.; Vasileiou, C.; Wang, W.; Santos, E. M.; Geiger, J. H.; Borhan, B., "Turn-on" protein fluorescence: in situ formation of cyanine dyes. *Journal of the American Chemical Society* **2015,** *137* (3), 1073-80.

56. Gilson, A. I.; Marshall-Christensen, A.; Choi, J.-M.; Shakhnovich, E. I., The Role of Evolutionary Selection in the Dynamics of Protein Structure Evolution. *Biophysical Journal* **2017,** *112* (7), 1350-1365.

57. Liu, Y.; Eisenberg, D., 3D domain swapping: As domains continue to swap. *Protein Science* **2002,** *11* (6), 1285-1299.

58. Bennett, M. J.; Schlunegger, M. P.; Eisenberg, D., 3D DOMAIN SWAPPING - A Mechansim for Oligomer Assembly. *Protein Science* **1995,** *4* (12), 2455-2468.
59. Nandwani, N.; Surana, P.; Udgaonkar, J. B.; Das, R.; Gosavi, S., Amino-acid composition after loop deletion drives domain swapping. *Protein Science* **2017,** *26* (10), 1994-2002.

60. Green, S. M.; Gittis, A. G.; Meeker, A. K.; Lattman, E. E., One-step evolution of a dimer from a monomeric protein. *Nature Structural Biology* **1995,** *2* (9), 746-751.

61. Rousseau, F.; Schymkowitz, J. W. H.; Wilkinson, H. R.; Itzhaki, L. S., Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proceedings of the National Academy of Sciences of the United States of America* **2001,** *98* (10), 5596-5601.

62. Seeliger, M. A.; Schymkowitz, J. W. H.; Rousseau, F.; Wilkinson, H. R.; Itzhaki, L. S., Folding and association of the human cell cycle regulatory proteins ckshs1 and ckshs2. *Biochemistry* **2002,** *41* (4), 1202-1210.

63. Parker, M. J.; Dempsey, C. E.; Hosszu, L. L. P.; Waltho, J. P.; Clarke, A. R., Topology, sequence evolution and folding dynamics of an immunoglobulin domain. *Nature Structural Biology* **1998,** *5* (3), 194-198.

64. Arndt, K. M.; Muller, K. M.; Pluckthun, A., Factors influencing the dimer to monomer transition of an antibody single-chain Fv fragment. *Biochemistry* **1998,** *37* (37), 12918-12926.

65. Rousseau, F.; Schymkowitz, J. W. H.; Itzhaki, L. S., The unfolding story of three-dimensional domain swapping. *Structure* **2003,** *11* (3), 243-251.

66. Kanwar, M.; Wright, R. C.; Date, A.; Tullman, J.; Ostermeier, M., Protein switch engineering by domain insertion. *Methods in Enzymology* **2013,** *523*, 369-388.

67. Sessler, R. J.; Noy, N., A ligand-activated nuclear localization signal in cellular retinoic acid binding protein-II. *Molecular Cell* **2005,** *18* (3), 343-53.

68. Gillilan, R. E.; Ayers, S. D.; Noy, N., Structural Basis for Activation of Fatty Acid-binding Protein 4. *Journal of Molecular Biology* **2007,** *372* (5), 1246-1260.

69. Sessler, R. J.; Noy, N., A Ligand-Activated Nuclear Localization Signal in Cellular Retinoic Acid Binding Protein-II. *Molecular Cell* **2005,** *18* (3), 343-353.

70. Nossoni, Z.; Assar, Z.; Yapici, I.; Nosrati, M.; Wang, W.; Berbasova, T.; Vasileiou, C.; Borhan, B.; Geiger, J., Structures of holo wild-type human cellular retinol-binding protein II (hCRBPII) bound to retinol and retinal. *Acta Crystallographica Section D* **2014,** *70* (12), 3226-3232.

71. Folli, C.; Calderone, V.; Ramazzina, I.; Zanotti, G.; Berni, R., Ligand Binding and Structural Analysis of a Human Putative Cellular Retinol-binding Protein. *Journal of Biological Chemistry* **2002,** *277* (44), 41970-41977.

72. Aalten, D. M. F. v.; Findlay, J. B. C.; Amadei, A.; Berendsen, H. J. C., Essential dynamics of the cellular retinol-binding protein evidence for ligand-induced conformational changes. *Protein Engineering* **1995,** *8* (11), 1129-1135.

73. Silvaroli, J. A.; Arne, J. M.; Chelstowska, S.; Kiser, P. D.; Banerjee, S.; Golczak, M., Ligand binding induces conformational changes in human cellular retinol-binding protein 1 (CRBP1) revealed by atomic resolution crystal structures. *Journal of Biological Chemistry* **2016,** *291* (16), 8528-8540.

74. Malpeli, G.; Stoppinf, M.; Zapponi, M. C.; Folli, C.; Berni, R., Interactions with retinol and retinoids of bovine cellular retinol-binding protein. *European Journal of Biochemistry* **1995,** *229* (2), 486-493.

75. Wang, W.; Nossoni, Z.; Berbasova, T.; Watson, C. T.; Yapici, I.; Lee, K. S.; Vasileiou, C.; Geiger, J. H.; Borhan, B., Tuning the electronic absorption of protein-embedded all-*trans*-retinal. *Science* **2012,** *338* (6112), 1340-3.

76. Vasileiou, C.; Lee, K. S. S.; Crist, R. M.; Vaezeslami, S.; Goins, S. M.; Geiger, J. H.; Borhan, B., Dissection of the critical binding determinants of cellular retinoic acid binding protein II by mutagenesis and fluorescence binding assay. *Proteins: Structure, Function, and Bioinformatics* **2009,** *76* (2), 281-290.

CHAPTER III: DISCOVERY OF A DOMAIN-SWAPPED TRIMER AS A NOVEL FOLD FOR THE ILBP FAMILY

III-1. Oligomerization in biology: an efficient regulator for many biological processes

Oligomerization is perhaps one of the most efficient natural tools to create functionally distinct proteins from a common fold through evolutionary mechanisms.[1] Oligomerization manifests its critical roles in many biological processes, such as signaling,[2-3] enzymatic activities,[4] and cargo transportation,[5-6] where the self-associations of the proteins trigger different biological pathways. There are many services the oligomerization of proteins offers in cellular events, such as enhancing protein stability, inventing novel active sites or offering specificity to existing active centers, tuning the accessibility of proteins, etc.[1, 7-10]

Although oligomerization provides incredible advantages and diversity in protein functions, in some cases, it can also lead to protein aggregation and unwanted pathogenic frameworks.[11-12] Due to the dual aspects of oligomerization in cellular events, thorough study of the proteins subjected to these processes is crucial. However, through a simple comparison of the number of proteins that have been structurally determined so far vs. the number of known oligomeric forms that have been reported, we can conclude that there is a dearth of information in this area. The lack of biophysical data and high-resolution crystal structures of the oligomeric forms of proteins is a missing link in the investigation of the function-structure relationship in protein oligomerization.[10] [1] These types of studies are essential due to the nature of oligomerization that can be exploited in the field of protein engineering to broaden the ability of reprogramming functions and conformations in proteins. In this chapter, we will first mention previous efforts that have been performed to understand the oligomerization states of proteins. Then we will use our structural-based protein design approach to discover and elucidate the mechanism of oligomerization in our model system using hCRBPII as a template.

III-2. Domain swapping

Domain-swapping, where two protein domains trade identical parts of their structures to create dimers or higher-order oligomers, is one of the mechanisms leading to self-associations and oligomer formation in proteins (Figure III-1). [13-15] In most cases, instead of the small, polypeptide part of a protein called the "hinge loop" turning to create a closed monomer, it straightens and adopts the elongated conformation in the domain-

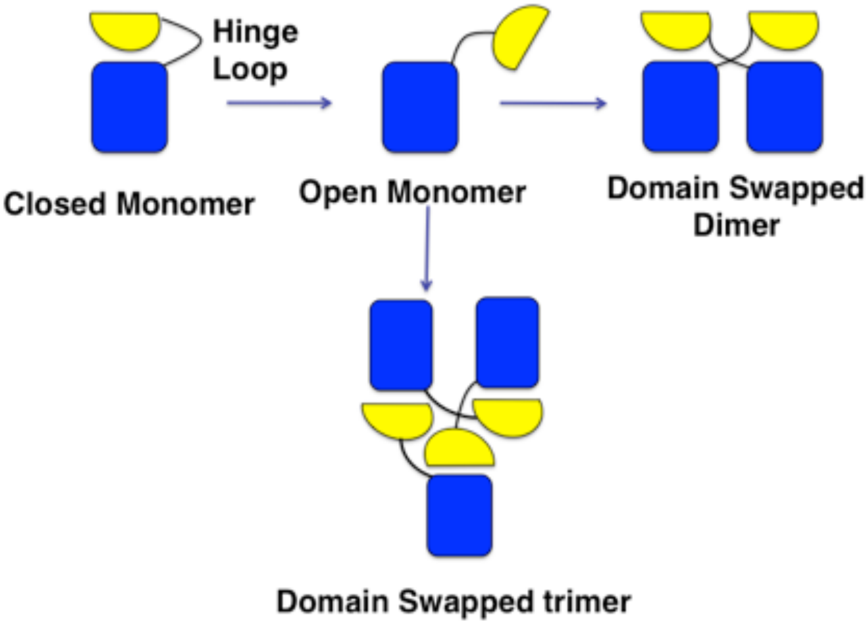


Figure III-1. Proposed mechanism of domain-swapping. Instead of turning and making a closed monomer, the hinge loop region elongates in the open monomer conformation. If two or more open monomers with proper C- and N- termini conformations find each other, they can generate dimers, trimers, or higher-order oligomers.

swapped from. The majority of the proteins that undergo domain-swapping merely exchange small parts of their secondary structures with identical molecules.[14, 16] However, there are several examples demonstrating that half of the protein is subjected to swapping.[17-18]

Domain swapping can ensue in two geometrically distinct modes: first, through the mutual or cyclical fashion to produce dimers, trimers, and higher-order oligomers; and second, in an open-ended or "run-away" fashion, where each protein molecule exchanges a domain with a complementary part of the neighboring molecule (Figure III-2). The latter

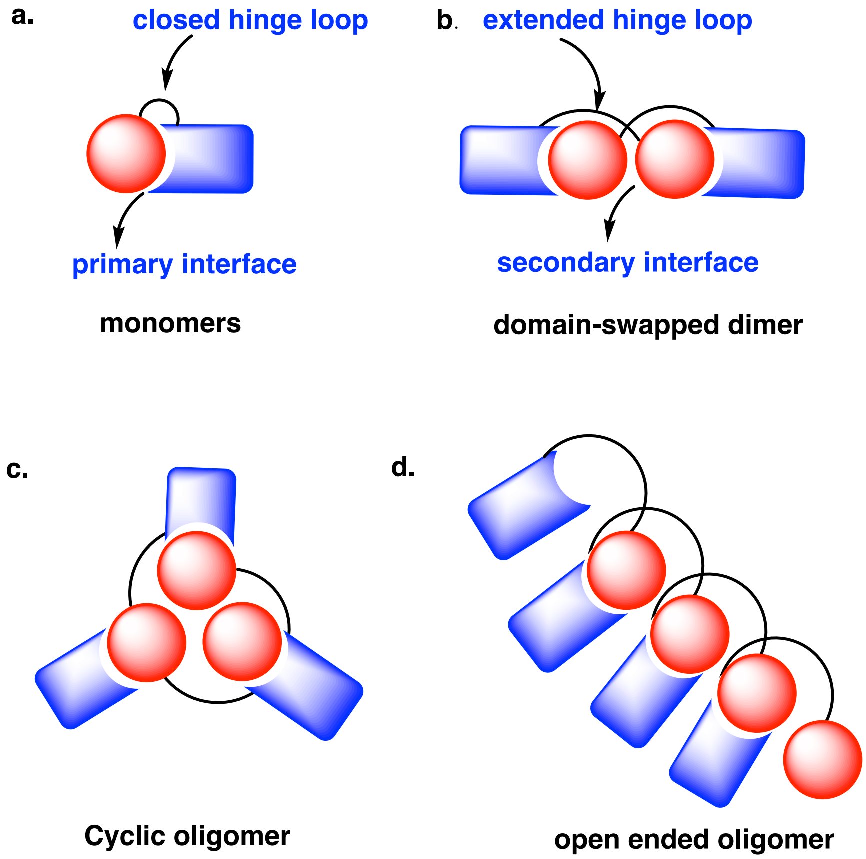


Figure III-2. Schematic demonstration of different domain-swapped forms. a. Closed monomer. The hinge loop region and primary interface are highlighted. b. Domain-swapped dimer. The extended conformation of the hinge loop and secondary interface created by two open monomers is represented. c. Cyclical oligomerization. d. Open-ended or "run-away" oligomerization.

mechanism of swapping is common in fibril formation although it has not been proven to be the origin of that.[14]

There are many variables playing roles to favor domain swapping, such as certain mutations in the hinge loop region, lyophilization, temperature, pH of the environment, protein concentration, ligand binding, etc.[1, 13, 15, 19-22] Understanding these variables is crucial for controlling the domain-swapping mechanism. Below we will discuss some of the well-known domain-swapped dimers and the factors giving rise to domain-swapping.

III-3. Reviewing several important examples of domain-swapped proteins

Diphteria toxin (DT), the protein that causes diphtheria, [23] is the first example of bonafide domain-swapped dimers that were structurally characterized (Figure III-3a).[24] DT performs three tasks in the course of cell intoxication using three folding segments: catalytic domain (C domain), transmembrane domain (T domain), and receptor domain (R domain). DT is endocytosed from coated pins when the R domain joins the cell-surface receptor. The protein then experiences a conformational change in the low-pH environment of the endosome and is transferred to the endosomal membrane. Finally, the C domain is transferred to the cytosol, which leads to the termination of protein synthesis through the deactivation of the eukaryotic elongation factor 2 (EF-2).[22, 24-26]
DT was first found to undergo domain swapping at low pH, mimicking the physiological endosomic environment (Figure III-3b). Additionally, based on the crystal structure obtained from the complex formation of DT and its receptor, it was shown that the DT domain-swapped dimer exists, in the presence of receptor binding, at neutral pH as well. [25,22] This study provides one of the best examples of how the domain-swapping mechanism leads to altered the protein function in biological systems.

Another excellent example of 3D domain-swapping is ribonucleases (RNase A), a bovine pancreatic enzyme whose exclusive function is RNA hydrolysis.[27] This protein family is dominantly monomeric (Figure III-3c), with only a few RNases (BS-RNase) found in the dimeric form in nature.[28] The dimeric conformation bestows proteins with incredible biological features, such as allosteric control and an increased concentration of active sites, among others.[29] The BS-RNase dimer formation prompted researchers to characterize the domain-swapped dimer of the RNase A protein. Eventually, the lyophilization of RNase A (in the presence of acetic acid) led to the formation of the RNase domain-swapped dimer. One of the interesting features of RNase-A is its ability to generate two different domain-swapped dimers. One domain-swapped dimer was formed as a result of the N-terminal α- helices swapping (Figure III-3d),[20] while another domain-swapped structure was formed where the c-terminal β-strands were exchanged (Figure III-3e).[30-31]

Later on, Park and Raines *et al.* discovered N-terminal domain swapping in physiological conditions.[32] Aside from the domain-swapped dimer form of RNase A, Liu et al. could identify two domain-swapped trimers that are formed through “linear” or “open-ended swapping” and cyclic domain-swapping (Figure III-3f).[31] Further studies demonstrated the formation of a domain-swapped tetramer for BS RNase through the incubation of the protein with acetic acid, followed by lyophilizing and re-dissolving the powder in “benign buffer”.[33-34]

The next example of domain-swapping is the immunoglobulin-binding B1 domain of streptococcal protein G (Figures III-3g and III-3h).[35] The structural analysis of the domain-swapped dimer structure with the wildtype monomer illuminates that Phe34 is a key amino acid triggering monomer-dimer conversions.[35-36] The mutation of this residue to Ala causes a WT monomeric structure. Additionally, some variants of GB1 lead to the domain-swapped dimer resulting in fibril formation.[37]

Notably, Cyanovirin-N, typified as CV-N, is another protein producing domain-swapped dimers, in which half of the protein swaps. This protein has become an excellent

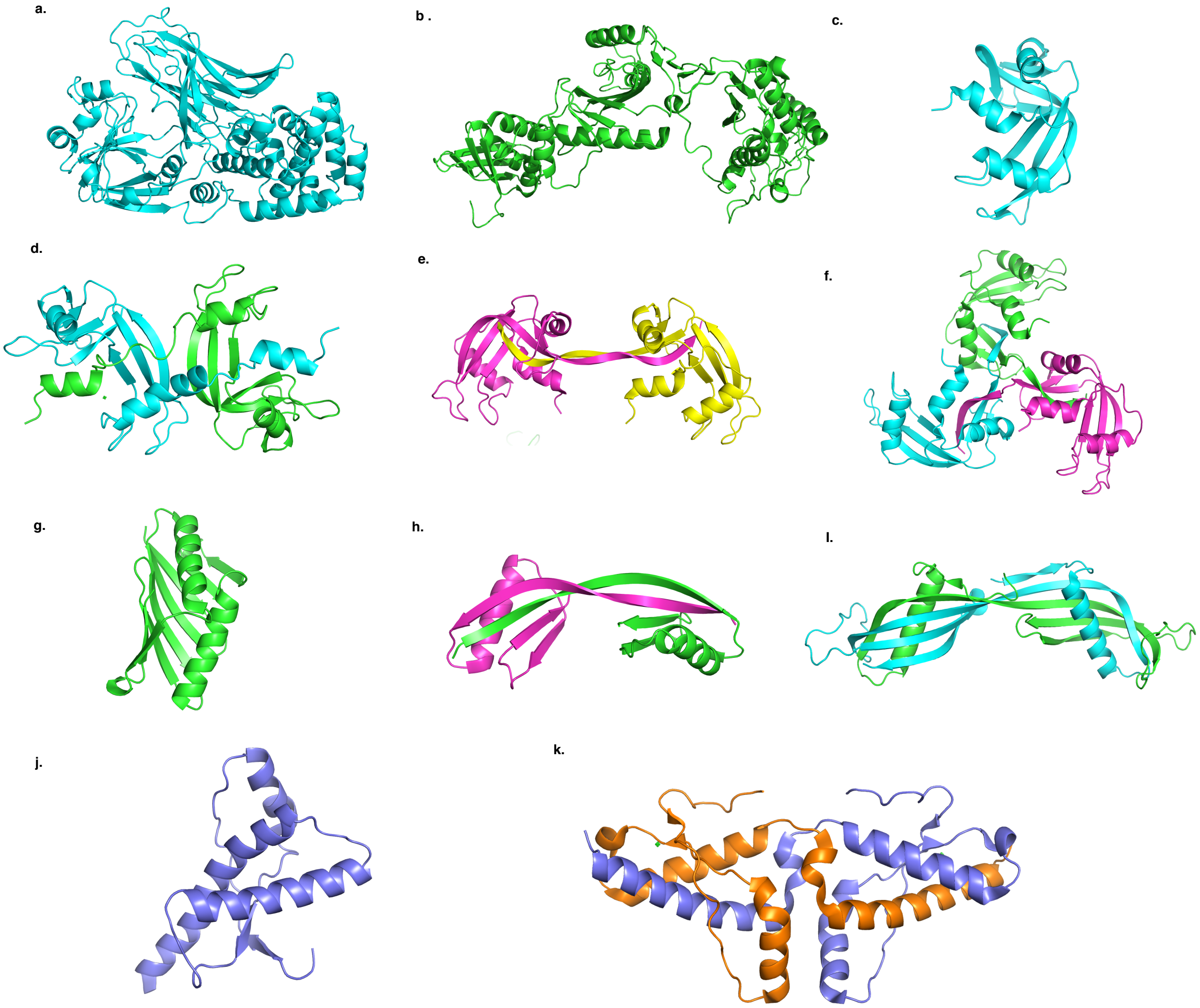


Figure III-3. Several examples of domain-swapped proteins. a. Diphteria toxin monomeric form (PDB code: 1MDT). b. Diphteria toxin domain-swapped dimer form (PDB code: 1DDT). c. Monomeric structure of RNase A (PDB code: 1RTB). d. RNaseA N-terminal domain-swapped dimer crystal structure (PDB code: 1A2W). e. RNaseA C-terminal domain-swapped dimer crystal structure (PDB code: 1F0V). f. RNaseA C-terminal cyclical domain-swapped trimer crystal structure (PDB code: 1JS0). g. GB1 monomeric structure (PDB code: 1Q10). h. GB1 monomeric structure (PDB code: 3V3X). i. amyloid-like cystatin domain-swapped dimer (PDB code:1TIJ). j. Human prion protein monomer (PDB code: 1QM0). k. Human prion protein domain-swapped dimer (PDB code: 1I4M).

candidate for deactivating enveloped viruses, such as Influenza, Ebola, Hepatitis C, HIV, etc.[38-40] The amount of domain-swapped dimer formation varies based on the experimental conditions. In neutral conditions, the monomer and domain-swapped dimer forms are in equilibrium with one another. Interestingly, the domain-swapped dimer was crystallized.[40] It has been shown that the domain-swapped-dimer is a kinetically-trapped form of the protein in high concentrations, and in physiological conditions, it reverts back to a more stable monomeric form.[41] The concurrence of the dimer and monomer in solution at normal conditions demonstrates the low barrier of interconversion between the two different forms of the protein. This equilibrium can be altered by certain mutations in the hinge loop region. For instance, the P51G mutation significantly stabilized the monomeric form.[22, 42]

However, since the WT dimer is not stable in the physiological conditions, the biological function of the domain-swapped dimer are still elusive. [22]

As mentioned earlier, there are examples of domain-swapped proteins where the domain-swapping can trigger unwanted protein aggregations and pathogenic structures [43]. The hallmarks of these disease-related proteins are cystatin[44] and prion,[45] whose domain-swapped dimer forms have been confirmed using crystal structures (Figures III-3i, III-3j, and III-3k). In contrast to the initial models of fibril amyloid formation that suggested the proteins are transformed entirely to β-sheet conformations from their native structures,[46] these proteins indicated that many structural features of the native forms of the proteins are preserved during fibril amyloid formation.[47] These proteins further show that fibril formation can be generated under ambient conditions *in vivo* and does not need harsh conditions with a high concentration of denaturant, as is usually employed for in vitro fibril formation. Additionally, it was proposed that only a segment of the protein (mostly the terminal or middle sections of the domains) undergoes the β-stack formation, and the rest of the polypeptide chain usually remains intact in its native form. As shown above, many variables contribute to domain-swapped dimer formation in proteins, and domain-swapping can alter the functions of proteins and in some cases, lead to disease-related structures, such as Parkinson's disease, diabetes, Huntington's disease, Alzheimer's disease, etc. [36, 48-50] Our next step is to understand the domain-swapping mechanism in the iLBP family.

III-4. Domain swapping in the iLBP family: a novel tool for protein engineering

As mentioned in the previous Chapter, Assar and Nossoni *et al* [17] reported that several variants of hCRBPII lead to large populations of domain-swapped dimers in which

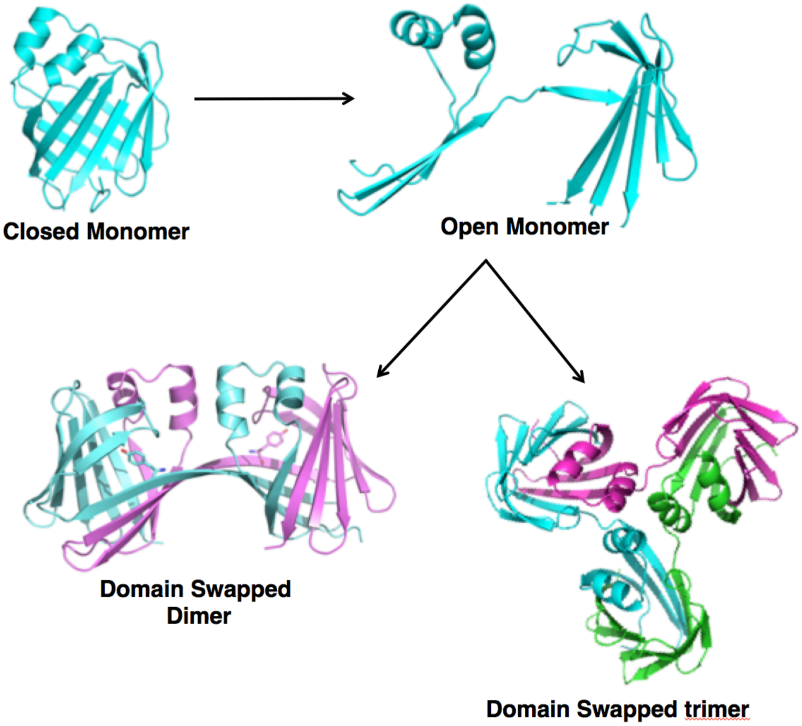


Figure III-4. The formation of domain-swapped dimer (previous study) and trimer (current study) based on open-monomer theory in hCRBPII.

half of the protein undergoes domain swapping. Furthermore, it has been reported that FABP5,[51] another member of the iLBP family, produces a similar domain-swapped dimer, as observed for hCRBPII.

In our previous study,[17] we demonstrated that the domain-swapped dimer is a kinetically trapped form of the protein that occurs as a result of increasing the lifetime of open-monomer intermediates through certain mutations in the hinge region. Our previous work[17] demonstrates the importance of the Y60 and E72 hydrogen bond in stabilizing the monomeric fold. The removal of the hydrogen bond between Y60 and E72 through mutations of either of these residues led to an increase in the dimer population as a result of enhancing the lifetime of open-monomer intermediates (Figure III-4). This is in contrast to the folding studies of this protein family conducted extensively by Gierasch's lab[52] (mainly focused on human cellular retinoic acid binding protein I (hCRABPI)), in which they showed that early β-barrel closure, driven by hydrophobic interaction, is pivotal in the folding pathway of this family, as it impedes protein aggregation and off-pathway oligomer formations such as dimers.[17]

As also shown in Chapter II, we brought the domain-swapped dimer to another level by illuminating the allosteric behavior of the hCRBPII domain-swapped dimer upon ligand binding, where a large conformational change was observed at atomic resolution. Through protein engineering, a new class of protein conformational switches was created that can be activated by retinal or reduction potential of the environment. Additionally, a novel metal-binding site was designed that can be tuned by the conformational readouts of the system. Furthermore, other studies have shown the power of domain-swapping that can be exploited in different protein engineering applications such as triggering the domain swapping by ligand binding,[53] generating a protein functional regulator using domain-swapped form of ribose binding protein,[54] controlling protein splicing through domain-swapping,[55] designing an optogenetic tool to regulate protein/protein interactions using domain-swapped form of photoactive yellow protein variant.[56] These examples demonstrate how domain-swapping can be employed in protein engineering to produce new functions in the protein, an ability which is not attainable in the monomeric form.

III-5. Discovery of the domain-swapped trimer

The first observation of an hCRBPII domain-swapped trimer was made during the purification of the Q108K:Q4F:Q38F:K40L:T53A:R58L (**A1**) variant, which eluted at two

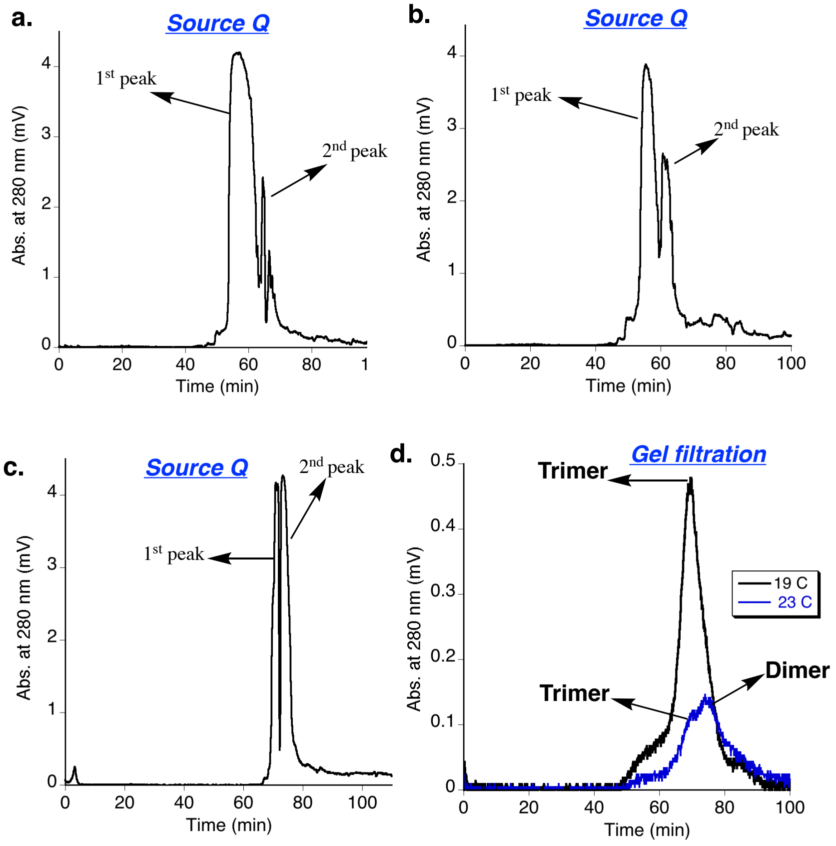
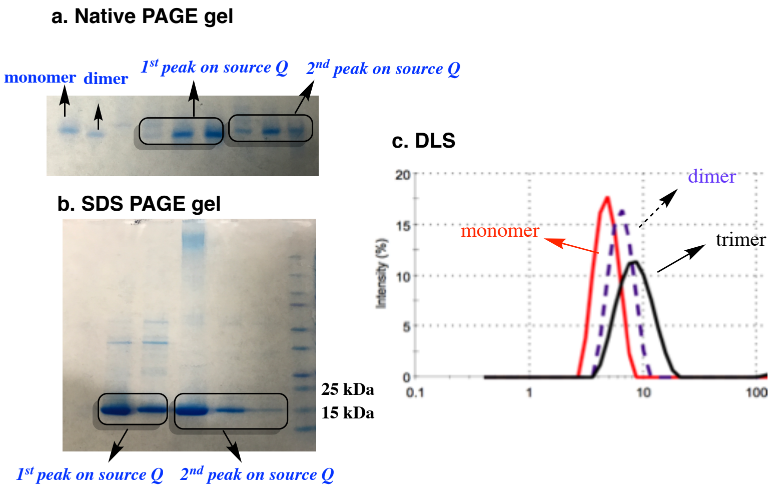


Figure III-5. Source Q and gel filtration chromatograms for: a. **A1**: Q108K:K40L:T53A:Q38F:R58L, 23°C, 20 hours expression. b. **A1**, 19°C, 40 hours expression. c. **A1**: V62E, 19°C, 40 hours expression. d. Gel filtration chromatogram of high-salt-eluted protein samples for **A1** (corresponding to the second peak in source Q) expressed at 19°C, 40 hours (black) vs. **A1** expressed at 23°C for 20 hours (blue).

distinct salt concentrations, with the first characteristic of the domain-swapped dimer (Figure III-5a).

Further analysis using gel filtration, NativePAGE gel, Dynamic Light Scattering (DLS), and X-ray diffraction revealed that the species eluting at higher salt was an hCRBPII domain-swapped trimer (Figures III-5, III-6, and III-7a). Structural analysis of the apo trimer of **A1** showed that two trimer molecules (six chains) existed in the asymmetric unit. Each trimer contained three structurally similar domains, related to each other by pseudo C_3 symmetry (RMSD~0.4) (Figure III-7). Additionally, the overlay of the open monomers of the trimer containing chains A, B, and D showed that these three open monomers are similar to each other, aside from a small change in the relative orientation of the two dimer domains (average RMSD~0.6) (Figure III-7). However, more structural


a. Native PAGE gel

b. SDS PAGE gel

c. DLS

Figure III-6. a. Native PAGE gel for 1st and 2nd peaks shown in the source Q column. b. SDS PAGE gel for 1st and 2nd peaks shown in the source Q column. c. Dynamic light scattering for monomer, dimer, and trimer (Q108K:K40L:T51V was used as a monomer for comparison) .

flexibility is observed between the open monomers of chain C, E, and F. In other words, while the ABD trimer has pseuo-three fold symmetry, the CEF trimer is significantly more asymmetric.

Table III-1. Ratios of monomer, dimer, and trimer formations in different variants.

Mutant	Trimer	Dimer	Monomer	Expression condition
A1	5%	95%	<1%	23°C - 20 hours
A1	30%*	67%	2%	19°C - 40 hours
A1-V62E	58%	42%	-	19°C - 40 hours
KD	22%	75%	3%	19°C - 40 hours
KLK	17%	70%	13%	19°C - 40 hours
KD-A28C	-	99%	-	23°C - 20 hours
KD-I32C	95%	5%	-	23°C - 20 hours
KD-A28C:I32C	34%	66%	-	23°C - 20 hours
A28C	-	-	100%	23°C - 20 hours
I32C	20%	40%	40%	23°C - 20 hours
A1-I32C	91%	9%	-	19°C - 40 hours
KDA28W**	100%	0	0	23°C - 20 hours
KDI32M:L36C	64%	34%	-	23°C - 20 hours
KDI32S	28%	15%	57%	19°C - 40 hours

** The structure of the mutant was not obtained. It could be side by side protein oligomerization.

Once the domain-swapped trimer was discovered, we subsequently identified and determined the structures of five other domain-swapped trimer-forming variants, including **A1**, Q108K:K40L:T53A:Q38F:R58L:V62E (**A1**-V62E), Q108K:T51D (**KD**), Q108K:K40L:T51K (**KLK**), Q108K:K40L:T53A:Q38F:R58L:F57H (trimer and dimer) (**A1**-

F57H), and Q108K:T51D:A28C:I32C (**KD**-A28C:I32C). In all cases, similar trimeric structures are evident, though there are significant differences between them. For example, the KD domain-swapped trimer contains 12 molecules (four distinct trimers; structure not shown here) in the asymmetric unit and shows more structural flexibility and heterogeneity than that seen in **A1**. This is in contrast to virtually all apo dimer structures (ten in all), which form very similar symmetric structures (the average RMSD between open monomers is ~0.3 Å). The conclusion to be drawn from this is that the domain-swapped trimer is more flexible than the domain-swapped dimer and is capable of adopting a number of distinct conformations. Additionally, as shown in Figure III-8, the overlay of the trimer's open

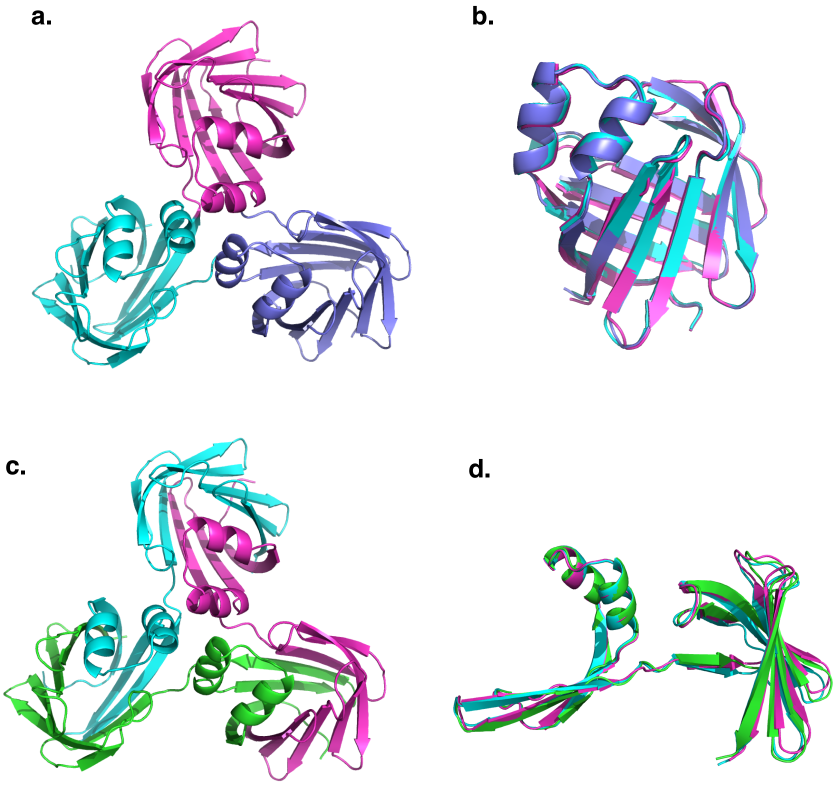


Figure III-7. a. The crystal structure of domain-swapped trimer of **A1** (chains A, B, and D). The close monomers highlighted with different colors. b. The overlay of three closed monomers show similar conformations for all of them. c. Domain-swapped trimer structure of **A1** (the same chains as mentioned above) where the open monomers made up trimer are highlighted by different colors. d. Overlay of the three open monomers demonstrated the similar conformations for chains A, B, and D.

monomer with the open monomers of the apo symmetric dimer, apo asymmetric dimer, and holo dimer reveals that the relative orientation of the c- and n-termini of the open monomers in the trimer form is very different from the apo and holo dimers, demonstrating the large displacement in the orientations of the c- and n-termini (relative to the symmetric, domain-swapped dimer) required to generate a domain-swapped trimer (Figure III-8). Furthermore, the torsional angle comparison between the dimer and trimer reveals that the torsional angles of N59, located on the β-strand, that runs between the domains in the trimer are significantly different from the apo dimer, which may be one of the factors

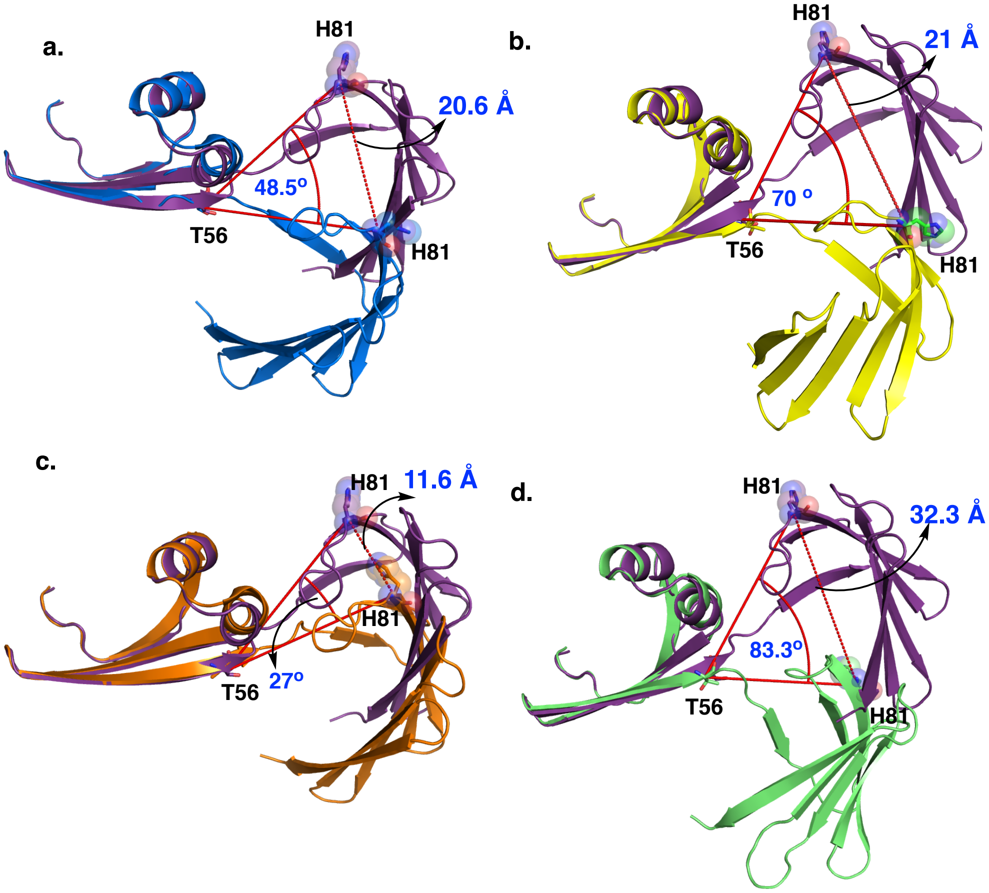


Figure III-8. Overlay of open monomer of apo domain-swapped trimer **A1** (purple), chain D, with: a. open monomer of apo WT hCRBPII domain-swapped dimer (blue); b. open monomer of Q108K:K40L:Y60W, chain A, apo asymmetric domain-swapped dimer (yellow); c. open monomer of Q108K:K40L:Y60W, chain B, apo asymmetric domain-swapped dimer (yellow); and d. open monomer of Q108K:K40L:T51F holo domain-swapped dimer, chain A (green).

contributing to this substantial conformational variation (Figure III-9). Based on our previous work, we proposed a folding pathway for hCRBPII that involved a common “open monomer” intermediate, where the n- and c-termini of the protein partially fold, followed by closure to either the monomer, dimer, or now trimer, depending on the concentration and propensity of the polypeptide to adopt an orientation consistent with the oligomeric form (Figure III-4).

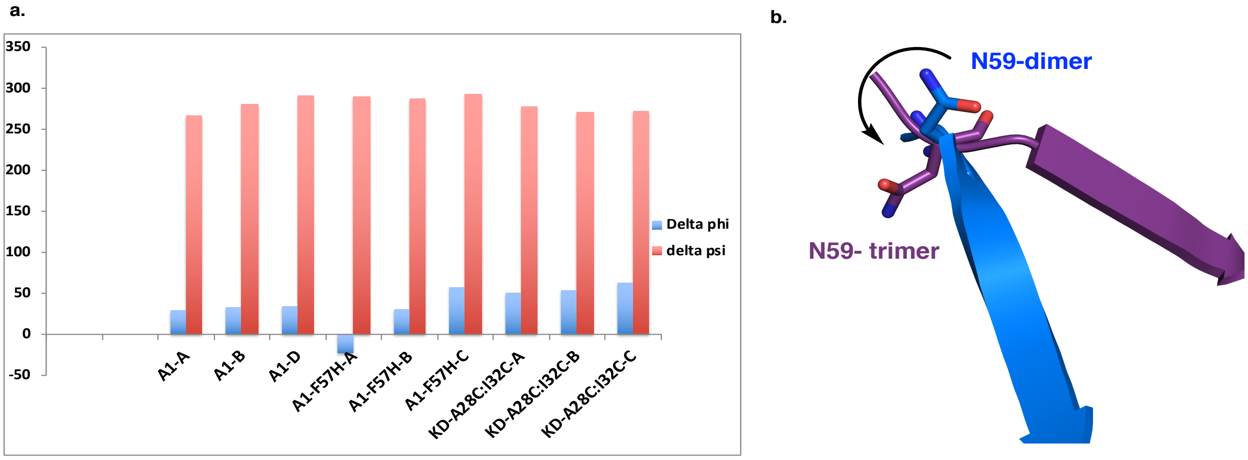


Figure III-9. a. Subtraction of N59 torsional angles for different trimer structures vs. the apo symmetric dimer. b. Overlay of dimer (blue) vs. trimer (purple) demonstrates that by rotating around N59, the dimer structure can be converted to trimer and vice versa.

Therefore, the mutations that provide a higher concentration of open monomers are expected to lead to a higher amount of dimer or trimer formation. To ponder this hypothesis, we used Q108K:T51D (**KD**), which was previously reported to yield a substantial amount of domain-swapped dimers. The protein’s monomer yield was very low. The expression of this protein at 19 degrees for ~40 hours depicted a large amount of domain-swapped dimers, along with a trimer peak (Table III-1). The **KD** and **A1** variants

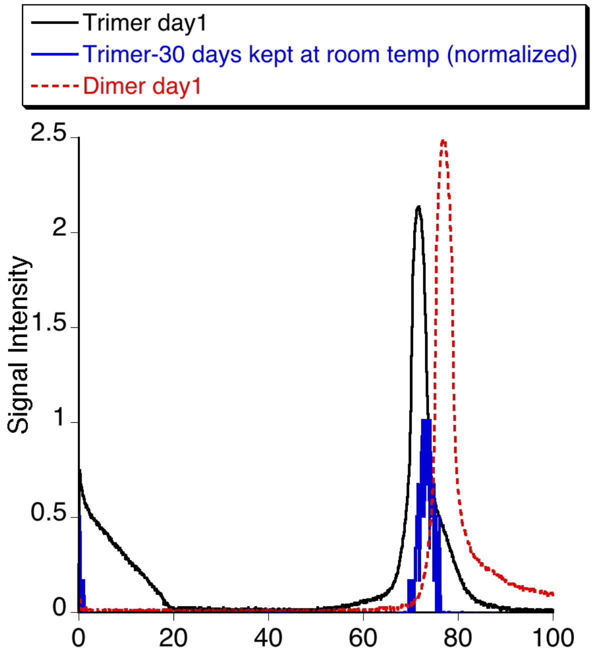


Figure III-10. Gel filtration chromatogram of Q108K:T51D dimer and trimer. The trimer was kept at room temperature and reloaded to Gel filtration column to examine the stability of trimer. The intensity for the trimer kept for 30 days at room temperature was normalized.

only have the Q108K mutation in common. Additionally, the expression of a substantial number of mutants of hCRBPII containing Q108K showed that this mutation does not play a significant role in creating domain-swapped species. However, **A1**, **A1**-V62E, **KD**, and Q108K:K40L:T51K (**KLK**) all contain a number of mutations, such as K40D and T53A in **A1**, T51D in KD, and T51K in **KLK**, that lead to a significant amount of domain-swapped dimers. Adding a V62E mutation to the **A1** construct resulted in more than 50% trimer, which is another example of the aforementioned effect. So far, all the constructs leading to trimer formation also make a substantial amount of domain-swapped dimers. Therefore, the formation of an open monomer in higher concentration might cause trimer formation as a byproduct, a process which is under the control of open-monomer concentration. Also remarkably the trimer form was stable after about a month kept at room temperature confirmed with gel filtration (Figure III-10).

III-6. Stability of dimer vs. trimer

Next, the thermal stability and structural analysis of the dimer vs. trimer was conducted. (We are very thankful to Cayman Chemical and their employee, Dr. Assar, who aided us in performing this assay.) In terms of stability, as shown in Table III-2, the melting point of the **KD** dimer was calculated to be about 39˚ C and the trimer around 59˚C. For **A1**, the melting point was calculated to be about 58˚ C for the trimer, whereas the dimer led to aggregation during the melting point experiment. Therefore, the trimer appears to be thermodynamically more stable than the dimer in these variants. A careful look at the structure of the domain-swapped trimer reveals the presence of an additional hydrophobic core at the center of the trimer structure. This additional interacting surface may provide more stability compared to the dimer. All three Phe57 residues in the trimer,

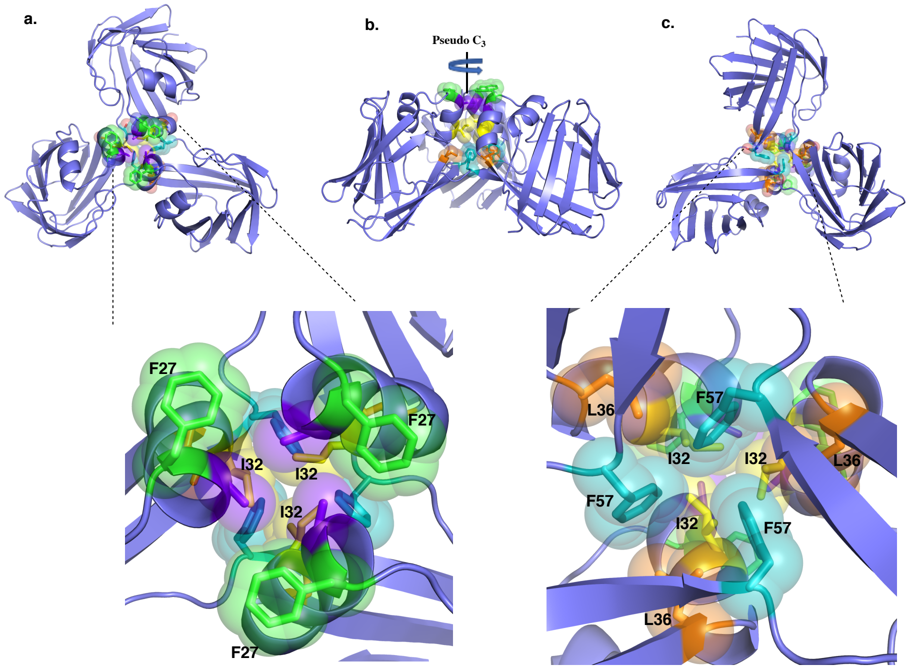


Figure III-11. Hydrophobic core generated by hydrophobic residues at the center of the domain-swapped trimer. F27 from each chain is shown in green, I32 in yellow, L36 in orange, and F57 in cyan. The pseudo-C_3 symmetry brings residues from the different chains close together.

together with Leu36, Ileu32, and Phe27, provide the surface for this additional hydrophobic core (Figure III-11).

Table III-2. Thermal stability assay for different variants.

Variants	Trimer M.P. (˚C)	Dimer M.P. (˚C)	Monomer M.P. (˚C)
A1	58	Aggregated	~ No monomer*
KD	59	39	~ No monomer*
KD-A28C	No trimer*	51.5	~No monomer*
KD-I32C	46	Noisy	No monomer*
KD-A28C:I32C	Noisy	48.5	No monomer*
A1-F57H	Aggregated	45.5	46
A1-F57H (Zn)	45, 61	43.5	41

*The monomer or trimer populations were too low to be confidently isolated for performing the assay.

III-7. Effect of ligand binding on trimer structure

In chapter II, we showed that in the domain-swapped dimer, a large conformational change in the relative orientations of the two domains occurs upon retinal binding. Unfortunately, in contrast to many other efforts, the retinal-bound structure of the domain-swapped trimer was not obtained. Instead, we used a thiophene-based chromophore with aldehyde and azetidine functionalities (**ThioFluor3**) (check chapter IV for further information), designed and synthesized in Dr. Borhan’s lab as a similar ligand to retinal for crystallization purposes. Using the chromophore shown in Figures III-12a and III-12b, with **A1** variant, we obtained a 2.54 Å resolution structure showing ligand binding in of the three binding sites, but with relatively different trajectories of **ThioFluor3** (Figure III-12b).

Interestingly, as shown in Figure III-12a, ligand binding does not alter the conformation of

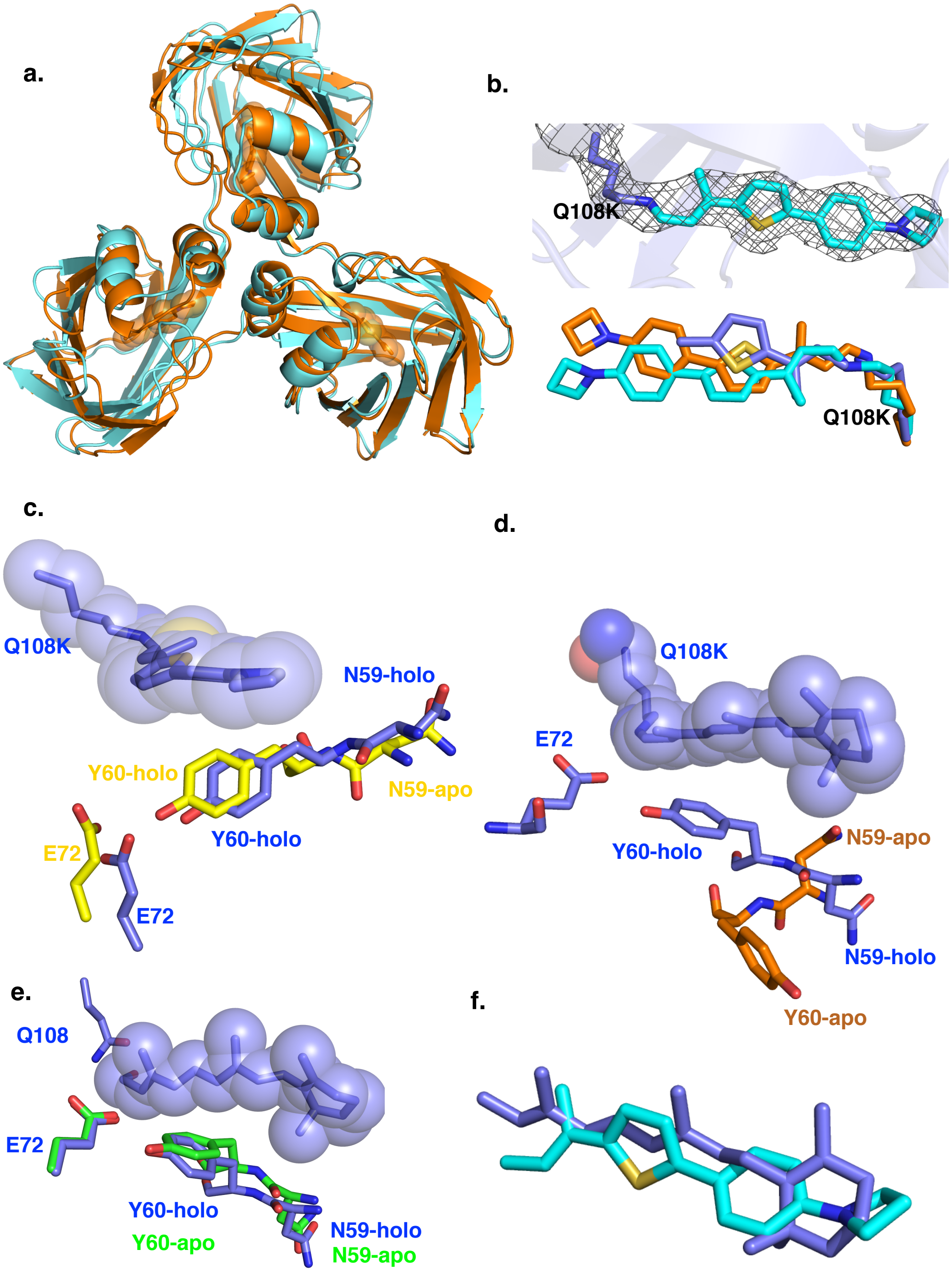


Figure III-12. a. Overlay of apo (cyan) and holo (orange) ligand-bound trimer (**A1** variant). b. Top: electron density of the ligand in one of the trimer chains. Bottom: overlay of ligands in three different chains. c. Overlay of apo (yellow) and holo ligand-bound trimer of **A1** variant (purple). N59 is pointing outside the binding cavity in both apo and holo structures where it does not sterically impinge on the chromophore. d. Overlay of apo (orange) and retinal-bound domain-swapped dimer (purple). N59 in apo conformation sterically clashes with the retinal trajectory. The result of this steric interaction is the "flipped-out" conformation for N59 in holo. e. Overlay of apo (green) and holo (purple) retinol-bound structure of WT-hCRBPII (PDB codes: 2RCQ and 2RCT). f. Overlay of retinylidene Q108K:T51D dimer (purple) with the synthesized chromophore used for trimer crystallization with **A1** variant (cyan).

the domain-swapped trimer considerably. As mentioned in chapter II, in the apo domain-swapped dimer (Figure III-12d), Asn59 is “flipped in” and Tyr60 is “flipped out” of the binding pocket. Asn59, in its flipped-in conformation, would sterically clash with the bound ligand, resulting in the flipped-out conformation for Asn59 and flipped-in conformation for Tyr60 within the holo dimer. The dispensation of these two residues causes a large conformational change within the holo domain-swapped dimer. However, as shown in Figures III-12c and III-12e, Asn59 in both the apo domain-swapped trimer and monomer is pointing outside of the binding cavity and Tyr60 is pointing inside. Asn59, in the flipped-out conformation, does not sterically impinge on the bound chromophore. Therefore, ligand binding does not affect the conformation of Asn59, and subsequently, Tyr60. Instead, there is a minor conformational change in the domain-swapped trimer that occurs upon ligand binding. This seems to indicate that the domain-swapped trimer might not be an allosterically regulated form of hCRBPII.

III-8. A cross-subunit disulfide bond drives trimerization

Though we have identified mutational "hot spots" that lead to domain-swapped trimers during expression, control over the folding pathway is lacking as invariably significant amounts of domain-swapped dimers are also formed. We therefore turned to a strategy previously exploited in controlling both domain-swapped dimer formation and allosteric behavior— the appropriate introduction of disulfide crosslinks that will bias the folding pathway toward domain-swapped trimer formation. In Chapter II, we showed that the introduction of cross-subunit disulfide linkages between Cys28 residues results in new allosteric behavior in the domain-swapped dimer and in essentially exclusive formation of domain-swapped dimers (Table III-1). Inspection of the domain-swapped trimer structure led us to conclude that a disulfide linkage in the middle of the α-helices may increase domain-swapped trimer formation, resulting in the **KD**-I32C and **A1**-I32C variants. The I32C position in the middle of the α-helices would lead to decreasing the flexibility of open monomers made up the dimer and presumably disfavoring this form. Gratifyingly, both variants, **KD**-I32C and **A1**-I32C, produced domain-swapped trimers almost exclusively. Similarly, the addition of the I32C mutation in the dimer-favored

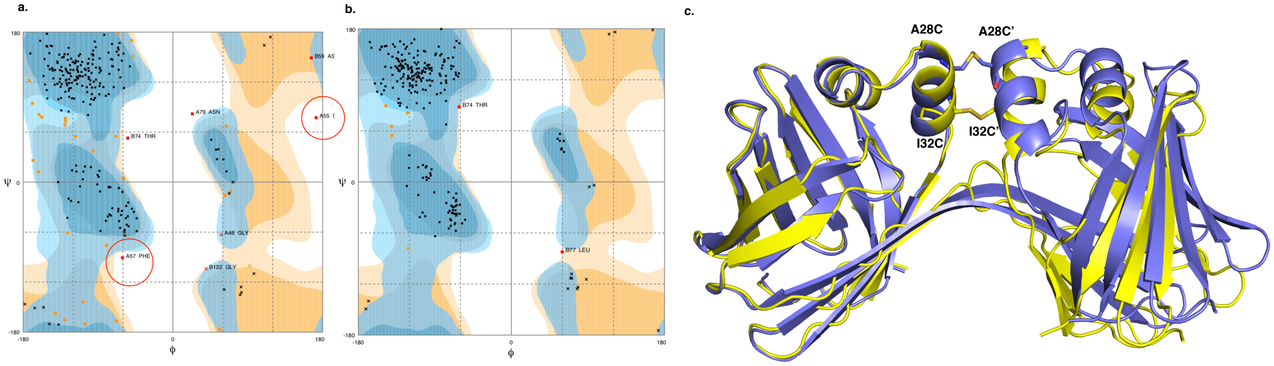


Figure III-13. Ramachandran plots for a. Q108K:T51D:I32C and b. Q108K:T51D:A28C. Q108:T51D:I32C shows some of the residues as outliers. c. Overlay of Q108K:T51D:A28C (purple-blue) vs. Q108K:T51D:I32C (yellow) crystal structures.

Q108K:T51D:A28C construct again increased trimer formation (Table III-1). Even in the context of the wild-type background, which strongly favors monomers and produces no trimers, the addition of the I32C single mutation resulted in a mixture of monomer (40%), dimer (40%), and trimer (20%) (Table III-1). In this case, we could also obtain the monomer crystal structure.

A comparison of the **KD**-A28C and **KD**-I32C domain-swapped dimer structures shows that the overall structures are very similar, with both adopting the apo symmetric dimer domain orientation. However, the Cys32 crosslink enforces an unfavorable Ramachandran outlier for residue Ser55 (Figure III-13a), indicating decreased stability of the dimer due to the Cys32 crosslink. In contrast, the **KD**-A28C:I32C domain-swapped trimer structure shows inter-subunit Cys32 and Cys28 crosslinks comfortably accommodated in the center of the structure (Figure III-14). This demonstrates how the domain-swapped folding pathway can be manipulated by strategic placement of disulfide crosslinks, allowing control over the folding product.

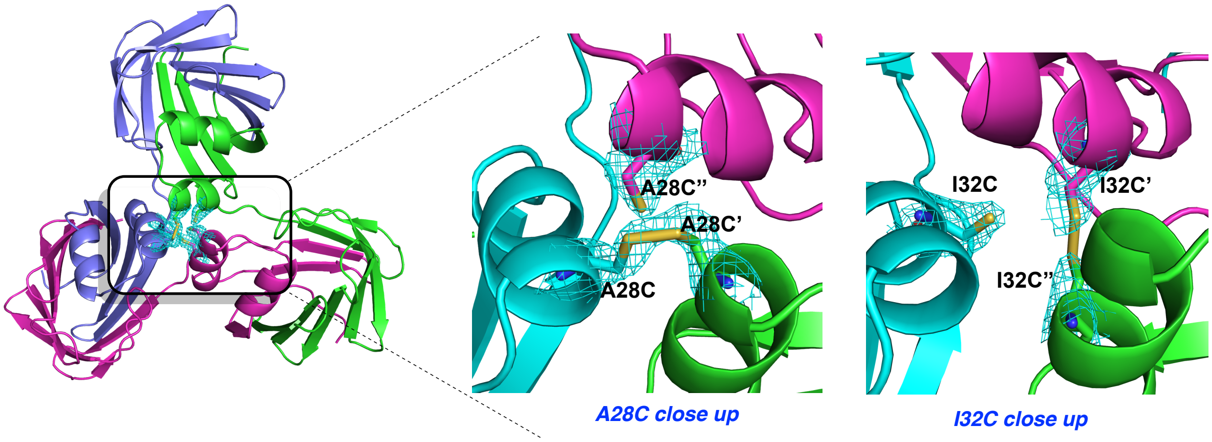


Figure III-14. The crystal structure of the domain-swapped trimer of **KD**:A28C:I32C showing the disulfide bond formation between different chains.

III-9. Introduction of a metal binding site in the domain-swapped trimer

As mentioned earlier, a characteristic feature of the domain-swapped trimer is the pseudo-C_3 symmetry axis running through the buried surface between the three domains of the trimer. The three Phe57 residues are located in the center of the pseudo-C_3 symmetry axis (Figure III-12). Previous studies demonstrated that domain-swapped oligomeric structures could result in binding sites for ions. For example, the domain-swapped trimer of RNase harbors a sulfate ion binding site,[57] and the domain-swapped trimer of CksHs2 has a divalent metal ion bound in the trimer interface.[58]

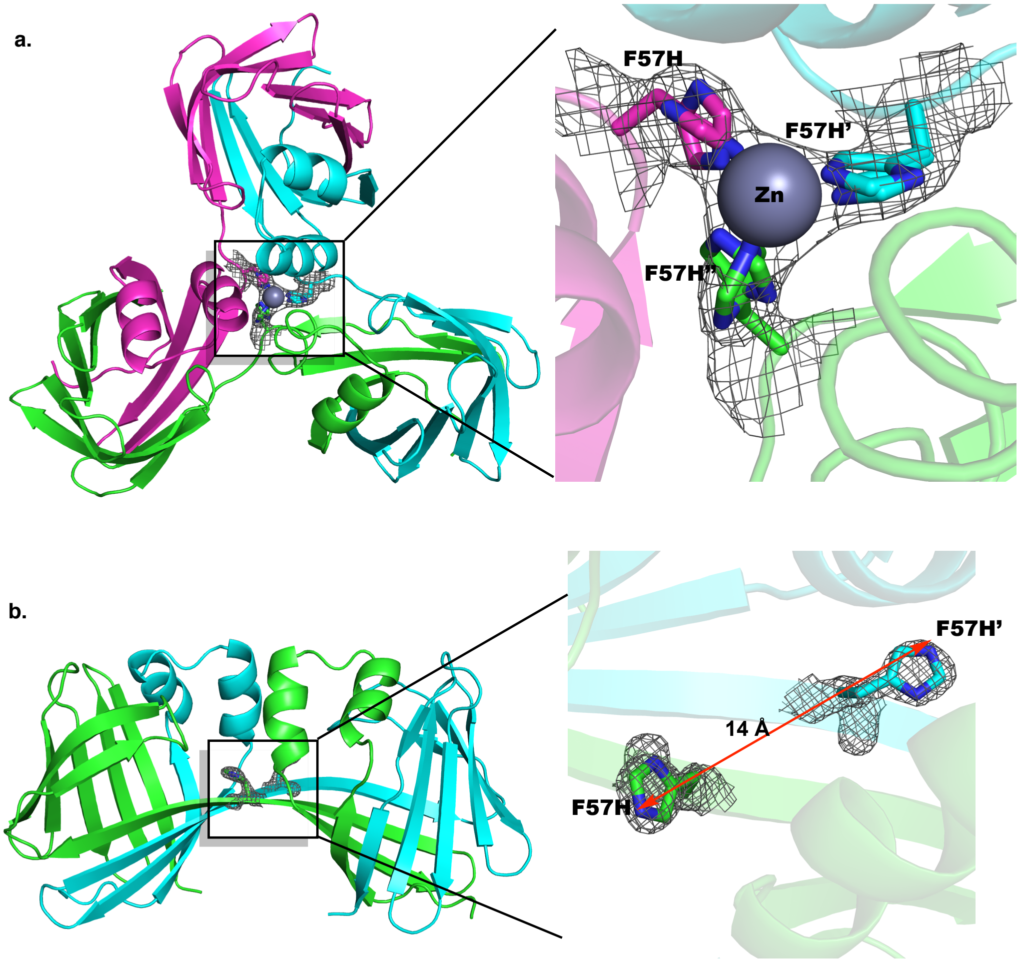


Figure III-15. The zinc-bound apo **A1**-F57H mutant: a. The $2F_o$-F_c map shows the presence of zinc coordinated with three histidines (contoured to 1σ.). b. The domain-swapped dimer of **A1**-F57H shows two histidines are far away from each other to generate the metal-binding site (contoured to 1σ.).

In the cases mentioned above, the structures possess metal-binding residues in the interface; however, the hCRBPII domain-swapped trimers and dimers have hydrophobic interfaces devoid of metal-binding moieties. We therefore attempted to create a metal-binding site by introduction of the F57H mutation to the domain-swapped trimer-forming **A1** variant to give the **A1**-F57H variant. Unfortunately, **A1**-F57H significantly reduced the quantity of trimers (<5%). This is probably because the trimer interface is less hydrophobic, and therefore less able to promote trimer formation. Therefore, we decided to add metal ion to the expression media in an effort to create a template for, and thereby promote, the trimer folding pathway. To our satisfaction, the addition of zinc acetate (100 μM) resulted in substantial quantities of trimers (15%), though monomers and dimers were also produced. The crystal structure of the Zn^{2+}-bound **A1**-F57H domain-swapped trimer revealed a single trimer in the asymmetric unit, perhaps indicating a less-flexible trimer relative to the other, structurally-heterogeneous trimer structures previously obtained, Figure III-15a. In this structure, as shown in Figure III-15a, zinc was coordinated with the three His57 residues coming from different chains to generate a trigonal pyramidal structure. Although zinc usually makes a tetrahedral structure, trigonal pyramidal (TP) zinc-centered structures are not rare in biology. For example, Fructose-1,6-bisphosphatase,[59] an essential regulatory enzyme in gluconeogenesis, and NADP-Dependent Alcohol dehydrogenase[60] both have zinc-centered TP structures. Furthermore, the domain-swapped dimer crystal structure of **A1**-F57H was obtained and showed that the two F57H residues are too far away from each other (~14 Å) to generate a metal-binding site, Figure III-15b. However, given the conditions of crystallization for the dimer, containing sodium malonate, the aforementioned structure for the dimer might be due to the lack of available zinc for the dimer (malonate can act as a multidentate ligand and generate a zinc complex). The anomalous difference map obtained from the zinc-bound trimer crystal collected using X-ray near the zinc absorption edge confirmed the presence of zinc bound in the center of the trimer with a 8σ anomalous difference peak, Figure III-16a. We were also successful in determining the zinc binding affinity (through collaboration with Dr. Pinger from Dr. Spence's lab) using a radioactive-zinc labeling assay. As shown in Figure III-16b, the K_d was found to be 0.70 ±0.21 μM for **A1**-F57H. Also, the melting points for this variant were calculated in the presence and the absence of zinc. As shown in Table III-2, the data further depicts the role of zinc in stabilizing the structure of the domain-swapped trimer protein where the absence of zinc causes protein aggregation during the thermal stability assay. These results suggest a novel protein fold in the iLBP family that could be used for a variety of protein engineering applications where the metal-binding site is required for triggering a special function in the protein.

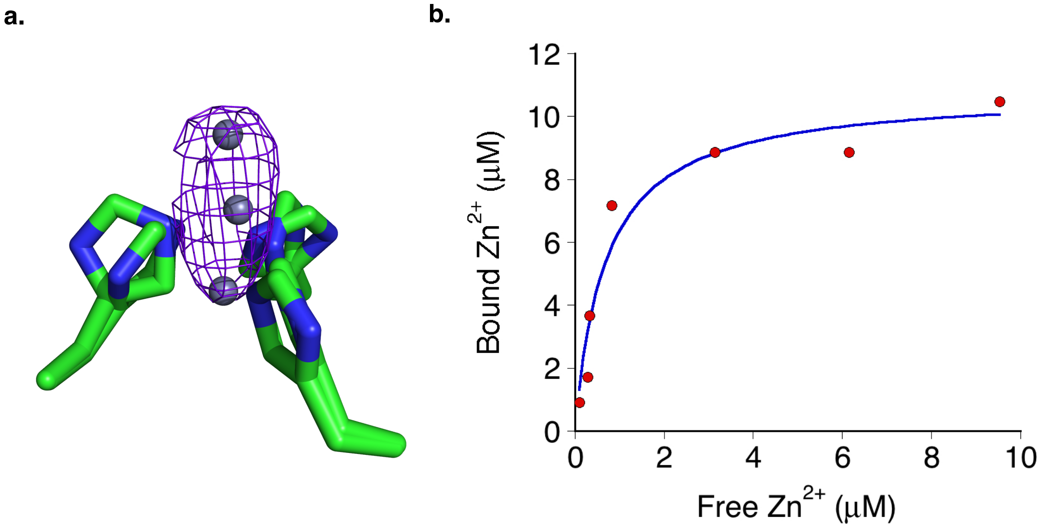


Figure III-16. a. The anomalous difference map collected above zinc x-ray absorption edge contoured to 8σ. The saturation-binding curve for **A1**-F57H showing the K_d: 0.70 ±0.21 μM.

III-10. Using “designed disulfide bond” to favor the domain-swapped oligomeric forms of other iLBP family members

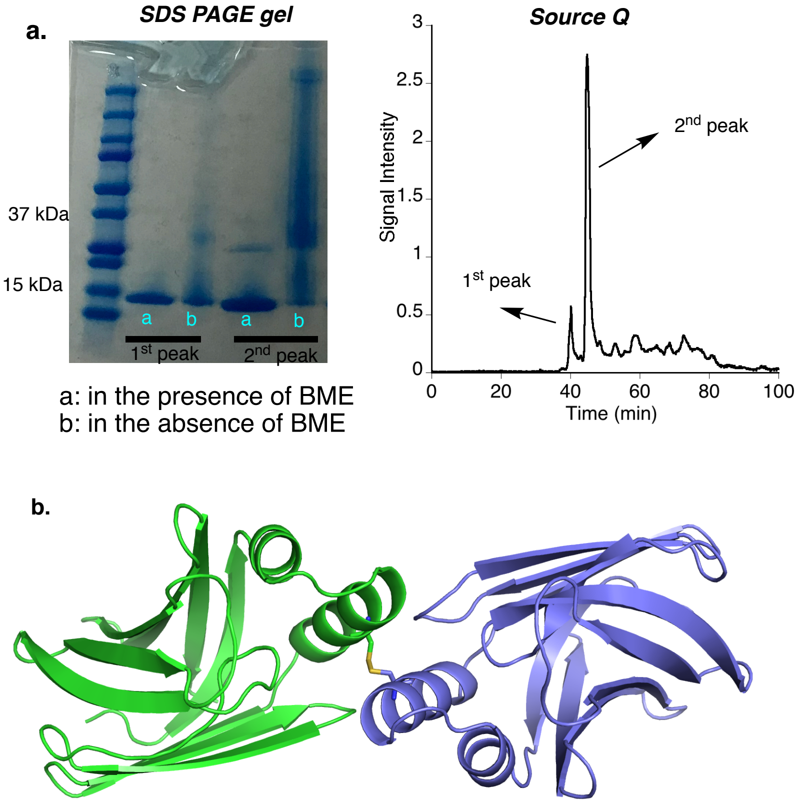


Figure III-17. a. Left: The gel electrophoresis under the reducing and no-reducing conditions. Right:. Source Q chromatogram of I31C showing two peaks. b. The crystal structure of I31C showing two close monomers connected by cross linked disulfide bridge.

The ability to favor different oligomeric forms through a designed disulfide bond piqued our attention and led us to pursue the same strategy and apply the mutation providing the disulfide bond in the equivalent position in CRABPII, another protein in iLBP, possessing about 40% sequence similarity with hCRBPII. The I31C mutation was made (equivalent to I32C in hCRBPII) and solubly expressed. The SDS PAGE gel demonstrated the presence of disulfide bond by comparing the results under reducing and non-reducing gel SDS PAGE gel electrophoresis, Figure III-17a. To our delight, the crystal structure of I31C variant was obtained at 2.1 Å resolution, Figure III-17c. The structure depicted two close monomers connected to each other through the disulfide bond (I31C from each monomer) forming a side-by-side dimer. This is in contrast to hCRBPII that the same mutation led to domain-swapped dimer and trimer formation. This observation is consistent by Gierasch's lab folding experiment [52] (as mentioned in III-4) and it further shows different proteins in the same family can have completely various folding pathways.

III-11. Conclusion and future direction

In conclusion, the first domain-swapped trimer in the iLBP protein family has been reported. The x-ray crystallographic data, along with melting-point measurements, showed the trimer structure has more thermal stability in comparison with the domain-swapped dimer. Furthermore, for the first time, an engineered metal-binding site and a disulfide bond have been used to control the oligomerization state in domain-swapped oligomers. The domain-swapped trimer provides a new structural template for protein design applications, including applications involving metal binding and ligand binding.

In the future, further mutational studies can be done to detect other potential oligomerization states in the hCRBPII protein. This system could also be used as a potential candidate for studying the folding pathway in the iLBP family, where the high-resolution crystal structures of various folds, along with biochemical assays, would illuminate how domain-swapped dimers and trimers can evolve from the monomeric form. Finally, the methods used for regulating oligomerization states in hCRBPII can be performed in other protein families, such as CRBPI, FABP4, and FABP5, to unravel different conformational states in those proteins.

III-12. Experimental procedures

III-12-1. DNA and protein purification

The same procedure mentioned in the experimental section of chapter II was used here.

III-12-2. PCR condition for hCRBPII

The same procedure explained in the experimental section of chapter II was used here.

III-12-3. List of primers

Site-directed mutagenesis was performed using K-CRBPII or KL-CRBPII-PET17b, which were generated in our previous work.[1]

K40D

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

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

T51D

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

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

K40L

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

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

Q108K

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

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

Q4F

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

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

R58L

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

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

R58L-57H

Forward: 5'-AGCACACACCTGAACTATGATGTG-3'

Reverse: 5'-CACATCATAGTTCAGGTGTGTGCT-3’

I32C

Forward: 5'-AACCGCAAG**TGT**GCAGTACGT-3’

Reverse: 5’-ACGTACTGC**ACA**CTTGCGGGT-3’

A28C

Forward: 5’- GATATTGATTTT**TGC**ACCCGCAAG-3’

Reverse: 5’- CTTGCGGGT**GCA**AAAATCAATATC-3’

V62E

Forward: 5’- AACTATGAT**GAG**GATTTCACTGTTGGAGTA-3’

Reverse: 5’- TACTCCAACAGTGAAATC**CTC**ATCATAGTT-3’

III-12-4. Transformation of PCR product

DH5α competent cells (50 μL) were thawed on ice and then 5 μL of PCR product was added to the sample. The sample was kept 30 minutes in ice; then the sample was heat-shocked at 42℃ for 45 seconds and 450 μL of LB was added. The sample was then incubated at 37℃ for 30 minutes. The cells were spread on a LB agar plate containing ampicillin (100 μg/mL) and incubated at 37℃ overnight.

III-12-5. DNA purification

The same procedure mentioned in the experimental section of chapter II was used here.

III-12-6. Expression of hCRBPII mutants

For the expression, the gene was transformed into BL21 (DE3) *E. coli* competent cells. A similar protocol mentioned for the transformation of PCR products was employed. To inoculate 1 L of LB with ampicillin (100 µg/mL), a single colony was used. After the cell culture was grown at 37℃ while shaking for~ 8 h, the overexpression was induced by adding 1 mL of 1M IPTG solution into the 1 L cell culture (overall concentration 1.0 mM of IPTG). The solution was shaken at mentioned temperature and cited time.

III-12-7. Protein isolation and purification

The cells were harvested by centrifuging for 20 min at 5,000 rpm at 4℃. After removal of the supernatant, the cells were resuspended in 60 mL of Tris buffer (10 mM Tris.HCl, pH=8.0). Then the resuspended cells lysed by ultrasonication (Power 60%, 1 min x 3) and centrifuged at 16℃ (10,000 rpm, 20 min).

The ion exchange chromatography using Q SepharoseTM, Fast Flow resin was used for the purification of proteins at 4℃. Further purification was performed with Fast Protein Liquid Chromatography. In the first step of purification with Q Sepharose Fast Flow resin (GE Health Sciences), the bound protein was first washed with 10 mM Tris-HCl, pH = 8.0 and then eluted with 10 mM Tris. HCl, 150 mM NaCl, pH = 8.0, and the pure fractions were desalted by dialysis against a 10 mM Tris. HCl pH = 8.0 buffer and loaded on a second anion exchange column (15Q, GE Health Sciences, BioLogic DuoFlow system). The low and high salt fractions were separately concentrated at 4°C to 5-8 mg/mL and further purified by size exclusion chromatography (Superdex 75 16/75 column, GE Health Sciences, packed in-house) in a buffer containing 10 mM Tris, 150 mM NaCl, pH = 8.0.

III-12-8. Crystallization, data collection and refinement

Each mutant protein was concentrated to between 6 to 8 mg/mL in a buffer containing 10 mM Tris, 150 mM NaCl, pH = 8.0. Crystals were grown via the hanging drop vapor diffusion method using 1 µL of protein solution and 1 µL of crystallization solution in the drop and 1 mL of crystallization solution in the reservoir. For the crystallization of the holo DS trimer **A1/ThioFluor3**, proteins were incubated with four equivalents of **ThioFluor3**. The best crystals for **A1**, **KLK**, **KD** grew using a crystallization solution of 25-30% PEG4000, 0.1 M sodium acetate pH 4.5-4.8, 0.1 M ammonium acetate. **A1**-F57H crystals grew under the following conditions: 0.2 M ammonium iodide, 20% PEG 3350. I31C (CRABPII) crystals grew under following conditions: 0.1 M sodium Malonate, 20% PEG 3350. The crystals were soaked in a cryoprotectant solution (the original conditions + 20% glycerol), flash frozen in loops in liquid nitrogen, and stored in a liquid nitrogen Dewar prior to data collection. Diffraction data was collected at the Advanced Photon Source (APS) (Argonne National Laboratory IL) LS-CAT, (sector 21-ID-D,F,G) using MAR300 and MAR350 detectors, 1.00 Å wavelength radiation at 100 K. The initial diffraction data was indexed, processed and scaled using the HKL2000 software package. The structures were solved by molecular replacement using PHASER in PHENIX and hCRBPII (PDB entry 2RCQ) as a search model. The initial electron density map was produced by Phaser-MR in PHENIX. Model rebuilding, placement of water molecules, etc. were done using COOT. The structures were refined using the PHENIX program packages. The chromophore was built using COOT and PHENIX to generate restraints.

III-12-9. Radioactive ^{65}Zn labelling assay

$^{65}ZnCl_2$ was purchased from (radionuclide purity: 99%) PerkinElmer. The $ZnCl_2$ was diluted in 18 μM distilled H_2O to make a radioactive 100.0 μM $^{65}ZnCl_2$ stock solution. Samples were prepared in 1.5 mL centrifugation tubes to contain a total volume of 210 μL and contain 15.0 μM protein and varying concentrations of $ZnCl_2$ in buffer. To create a saturation binding curve, the Zn^{2+} concentrations were varied from 1-15 μM. Then, 200 μL of each sample were pipetted into an ultrafiltration centrifugation device with a 10 kDa MWCO size-exclusion membrane (Amicon Ultra 0.5 mL centrifugal filters). The samples were centrifuged at 8,000g for 3 minutes to allow 6-15 μL of ultrafiltrate to collect in the bottom of the tube. The concentration of Zn^{2+} in the ultrafiltrate is considered the free-Zn^{2+} in the sample, as described by the literature. The volume separated from the sample is small enough so as to not disturb the equilibrium of the binding system (<8% of the volume of the bulk sample). Then, a 9.0 μL aliquot of ultrafiltrate sample was transferred into a separate tube for analysis. Calibration standards containing 9.0 μL of known radioactive Zn^{2+} concentrations were also prepared in separate tubes. Then, all standards and samples were analyzed by counting the emitted gamma radiation at (1115 keV) for 5 minutes per sample on a PerkinElmer 2480 Automatic Gamma Counter. The concentration of Zn^{2+} in the ultrafiltrate was quantified by comparing the counts-per-minute (CPM) for each sample to an external standards calibration curve. The concentration of Zn^{2+} in the ultrafiltrate was considered to be the free-Zn^{2+} concentration in the sample, while the concentration of protein-bound Zn^{2+} was calculated as the free-Zn^{2+} subtracted from the total-Zn^{2+}. The concentrations of free Zn^{2+} vs protein-bound Zn^{2+} for each sample were plotted and analyzed via non-linear regression software

(SigmaPlot 13.0) to create a saturation binding curve. This curve allowed the calculation of the equilibrium binding constant (K_d) and stoichiometry (B_{max}).

Table III-3. X-ray crystallographic data and refinement statistics for apo trimer Q108K:T51D:A28C:I32C, **A1** and Zinc-bound **A1**-F57H hCRBPII mutants.

	Q108K:T51D:A28C: I32C	**Q108K: K40L: T53A: Q38F: R58L (A1) apo**	**Zinc-bound A1-F57H**
Space group	$C222_1$	$P\ 2_12_12_1$	$P\ 2_12_12_1$
a(Å)	87.93	67.05	60.51
b(Å)	160.49	72.68	74.45
c (Å)	86.33	164.80	94.34
α(°)	90	90	90
β(°)	90	90	90 90
δ(°)	90	90	90
Molecules per Asymmetric Unit	3	6	3
Total reflection	1917398	2352250	2067465
Unique Reflection	35625	18585	14425 (1402)
Completeness (%)	99.11 (94.26)	99.49 (94.94)	99.87 (99.79)
Average I/σ	34.8	23	27(2.3)
R_{merge}	5	17	0.1 (1.1)
Resolution (Å)	29.39 - 2.10	14.97 - 2.90	46.96 - 2.55
(Last Shell)	(2.18- 2.10)	(2.99 - 2.89)	(2.64 - 2.55)
R_{work}/ R_{free} (%)	21.04 /24.85	32.16/25.46	32.36/25.80
RMS Bond Length (Å)	0.008	0.001	0.011
Bond Angle (°)	0.92	1.49	1.62
Average B factor	48.04	58.70	54.56
Number of water molecules	110	16	14
PDB ID	TBD	TBD	TBD

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

Table III-4. X-ray crystallographic data and refinement statistics for trimer hCRBPII mutants Q108K:T51D, **A2**, and Q108K:K40L:T51K.

	Q108K: T51D	**Q108K: K40L: T53A: Q38F: R58L:V62E (A2)**	**Q108K: K40L:T51K**
Space group	$P2_1$	$P22_12_1$	$P2_12_12_1$
a(Å)	84.56	37.12	69.83
b(Å)	111.27	101.03	78.47
c (Å)	101.83	109.47	178.89
α(°)	90	90	90
β(°)	112.89	90	90
δ(°)	90	90	90
Molecules per Asymmetric Unit	12	3	6
Total reflection	2373724	555975	
Unique Reflection	42592 (4247)	10793 (1015)	20546 (1916)
Completeness (%)	98.99 (99.79)	99.24 (93.81)	99.39 (94.38)
Average I/σ	22.5 (2.5)	17.1(2.1)	13.1 (2.3)
R_{merge}	0.098 (0.687)	0.119 (0.739)	0.2 (1.1)
Resolution (Å) (Last Shell)	15 - 2.80 (2.90 - 2.80)	45.87 - 2.79 (2.89 -2.79)	47.48 - 2.99 (3.09-2.99)
R_{work}/ R_{free} (%)	33.80/27.28	29.81/24.42	33.35/25.49
RMS Bond Length (Å)	0.014	0.011	0.022
RMS Bond Angle (°)	1.96	1.72	2.38
Average B factor	70.91	56.58	48.15
Number of water molecules	77	38	12
PDB ID	TBD	TBD	TBD

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

Table III-5. X-ray crystallographic data and refinement statistics for apo dimer hCRBPII Q108K:T51D:I32C and **A1**-F57H mutants.

	Q108K: T51D:I32C	**(A1-F57H) apo**
Space group	$P6_3$	$P2_12_12_1$
a(Å)	133.28	60.09
b(Å)	133.28	63.84
c (Å)	51.16	72.83
α(°)	90	90
β(°)	90	90
δ(°)	120	90
Molecules per Asymmetric Unit	2	2
Total reflection	2082961	902279
Unique Reflection	8637 (843)	31490 (3092)
Completeness (%)	98.82 (98.71)	99.62 (99.84)
Average I/σ	13.8 (2.1)	43.8 (2.2)
R_{merge}	0.17(1.40)	0.06 (0.90)
Resolution (Å) (Last Shell)	33.20- 3.20 (3.32- 3.20)	46.35 - 1.70 (1.76 - 1.70)
R_{work}/ R_{free} (%)	23.48/26.17	30.6-27.07
RMS Bond Length (Å)	0.01	0.006
RMS Bond Angle (°)	1.17	0.89
Average B factor	73.69	30.96
Number of water molecules	14	74
PDB ID	TBD	TBD

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

Table III-6. X-ray crystallographic data and refinement statistics for apo side-by-side dimer of hCRABPII I31C mutant.

	I31C
Space group	C2*1*
a(Å)	118.18
b(Å)	61.57
c (Å)	71.42
α(°)	90
β(°)	122
δ(°)	90
Molecules per Asymmetric Unit	3
Total reflection	1184723
Unique Reflection	22923 (2158)
Completeness (%)	98.90 (93.05)
Average I/σ	29.06 (2.15)
R_{merge}	6.2 (62)
Resolution (Å) (Last Shell)	33.17-2.17 (2.25 - 2.17)
R_{work}/ R_{free} (%)	21.75/30.98
RMS Bond Length (Å)	0.008
RMS Bond Angle (°)	0.96
Average B factor	54.84
Number of water molecules	14
PDB ID	TBD

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

REFERENCES

REFERENCES

1. Marianayagam, N. J.; Sunde, M.; Matthews, J. M., The power of two: protein dimerization in biology. *Trends in Biochemical Sciences* **2004,** *29* (11), 618-25.

2. Smulski, C. R.; Decossas, M.; Chekkat, N.; Beyrath, J.; Willen, L.; Guichard, G.; Lorenzetti, R.; Rizzi, M.; Eibel, H.; Schneider, P.; Fournel, S., Hetero-oligomerization between the TNF receptor superfamily members CD40, Fas and TRAILR2 modulate CD40 signalling. *Cell Death & Disease* **2017,** *8* (2), e2601.

3. Needham, S. R.; Roberts, S. K.; Arkhipov, A.; Mysore, V. P.; Tynan, C. J.; Zanetti-Domingues, L. C.; Kim, E. T.; Losasso, V.; Korovesis, D.; Hirsch, M.; Rolfe, D. J.; Clarke, D. T.; Winn, M. D.; Lajevardipour, A.; Clayton, A. H.; Pike, L. J.; Perani, M.; Parker, P. J.; Shan, Y.; Shaw, D. E.; Martin-Fernandez, M. L., EGFR oligomerization organizes kinase-active dimers into competent signalling platforms. *Nat Commun* **2016,** *7*, 13307.

4. Fatmi, M. Q.; Chang, C. E., The role of oligomerization and cooperative regulation in protein function: the case of tryptophan synthase. *PLoS Comput Biol* **2010,** *6* (11), e1000994.

5. Sato, K.; Nakano, A., Oligomerization of a cargo receptor directs protein sorting into COPII-coated transport vesicles. *Molecular Biology of the Cell* **2003,** *14* (7), 3055-63.

6. Klingenberg, M., Membrane protein oligomeric structure and transport function. *Nature* **1981,** *290* (5806), 449-54.

7. Hashimoto, K.; Panchenko, A. R., Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proceedings of the National Academy of Sciences* **2010,** *107* (47), 20352-20357.

8. Hurtley, S. M.; Helenius, A., Protein Oligomerization in the Endoplasmic Reticulum. *Annual Review of Cell Biology* **1989,** *5* (1), 277-307.

9. Fallas, J. A.; Ueda, G.; Sheffler, W.; Nguyen, V.; McNamara, D. E.; Sankaran, B.; Pereira, J. H.; Parmeggiani, F.; Brunette, T. J.; Cascio, D.; Yeates, T. R.; Zwart, P.; Baker, D., Computational design of self-assembling cyclic protein homo-oligomers. *Nature chemistry* **2017,** *9* (4), 353-360.

10. Gabizon, R.; Friedler, A., Allosteric modulation of protein oligomerization: an emerging approach to drug design. *Frontiers in Chemistry* **2014,** *2*, 9.

11. Choi, M. L.; Gandhi, S., Crucial role of protein oligomerization in the pathogenesis of Alzheimer's and Parkinson's diseases. *The FEBS journal* **2018,** *285* (19), 3631-3644.

12. Wilkaniec, A.; Czapski, G. A.; Adamczyk, A., Cdk5 at crossroads of protein oligomerization in neurodegenerative diseases: facts and hypotheses. *Journal of Neurochemistry* **2016,** *136* (2), 222-33.

13. Bennett, M. J.; Schlunegger, M. P.; Eisenberg, D., 3D domain swapping: a mechanism for oligomer assembly. *Protein Science* **1995,** *4* (12), 2455-68.

14. Rousseau, F.; Schymkowitz, J. W.; Itzhaki, L. S., The unfolding story of three-dimensional domain swapping. *Structure* **2003,** *11* (3), 243-51.

15. Rousseau, F.; Schymkowitz, J.; Itzhaki, L. S., Implications of 3D domain swapping for protein folding, misfolding, and function. In *Protein Dimerization and Oligomerization in Biology*, Mattews, J. M., Ed. **2012**; Vol. 747, pp 137-152.

16. Koharudin, L. M.; Liu, L.; Gronenborn, A. M., Different 3D domain-swapped oligomeric cyanovirin-N structures suggest trapped folding intermediates. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110* (19), 7702-7.

17. Assar, Z.; Nossoni, Z.; Wang, W.; Santos, E. M.; Kramer, K.; McCornack, C.; Vasileiou, C.; Borhan, B.; Geiger, J. H., Domain-swapped dimers of intracellular lipid-binding proteins: evidence for ordered folding intermediates. *Structure* **2016,** *24* (9), 1590-8.

18. Barrientos, L. G.; Louis, J. M.; Botos, I.; Mori, T.; Han, Z.; O'Keefe, B. R.; Boyd, M. R.; Wlodawer, A.; Gronenborn, A. M., The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure* **2002,** *10* (5), 673-86.

19. Rousseau, F.; Schymkowitz, J. W.; Wilkinson, H. R.; Itzhaki, L. S., Intermediates control domain swapping during folding of p13suc1. *The Journal of Biological Chemistry* **2004,** *279* (9), 8368-77.

20. Liu, Y.; Hart, P. J.; Schlunegger, M. P.; Eisenberg, D., The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1 Å resolution. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** *95* (7), 3437-42.

21. O'Neill, J. W.; Kim, D. E.; Johnsen, K.; Baker, D.; Zhang, K. Y., Single-site mutations induce 3D domain swapping in the B1 domain of protein L from peptostreptococcus magnus. *Structure* **2001,** *9* (11), 1017-27.

22. Rousseau, F.; Schymkowitz, J.; Itzhaki, L. S., Implications of 3D domain swapping for protein folding, misfolding and function. *Advances in Experimental Medicine and Biology* **2012,** *747*, 137-52.

23. Zhang, J.; Wei, H.; Guo, X.; Hu, M.; Gao, F.; Li, L.; Zhang, S., Functional verification of the diphtheria toxin A gene in a recombinant system. *Journal of Animal Science and Biotechnology* **2012,** *3* (1), 29.

24. Bennett, M. J.; Eisenberg, D., Refined structure of monomeric diphtheria toxin at 2.3 A resolution. *Protein Science* **1994,** *3* (9), 1464-75.

25. Zhukova Ia, F.; Dzhedzhula, H. H.; Radavs'kyi Iu, L., [The structural-functional organization of the diphtheria toxin]. *Mikrobiolohichnyi Zhurnal (Kiev, Ukraine : 1993)* **1997,** *59* (3), 91-101.

26. Bennett, M. J.; Choe, S.; Eisenberg, D., Refined structure of dimeric diphtheria toxin at 2.0 A resolution. *Protein Science* **1994,** *3* (9), 1444-63.

27. Liu, Y.; Hart, P. J.; Schlunegger, M. P.; Eisenberg, D., The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** *95* (7), 3437-3442.

28. Kim, J. S.; Soucek, J.; Matousek, J.; Raines, R. T., Catalytic activity of bovine seminal ribonuclease is essential for its immunosuppressive and other biological activities. *The Biochemical Journal* **1995,** *308 (Pt 2)* (Pt 2), 547-50.

29. Vitagliano, L.; Adinolfi, S.; Sica, F.; Merlino, A.; Zagari, A.; Mazzarella, L., A potential allosteric subsite generated by domain swapping in bovine seminal ribonuclease. *Journal of Molecular Biology* **1999,** *293* (3), 569-77.

30. Miller, K. H.; Marqusee, S., Propensity for C-terminal domain swapping correlates with increased regional flexibility in the C-terminus of RNase A. *Protein Science* **2011,** *20* (10), 1735-44.

31. Miller, K. H.; Karr, J. R.; Marqusee, S., A hinge region *cis*-proline in ribonuclease A acts as a conformational gatekeeper for C-terminal domain swapping. *Journal of Molecular Biology* **2010,** *400* (3), 567-78.

32. Park, C.; Raines, R. T., Dimer formation by a "monomeric" protein. *Protein Science* **2000,** *9* (10), 2026-33.

33. Gotte, G.; Mahmoud Helmy, A.; Ercole, C.; Spadaccini, R.; Laurents, D. V.; Donadelli, M.; Picone, D., Double Domain Swapping in Bovine Seminal RNase: Formation of Distinct N- and C-swapped Tetramers and Multimers with Increasing Biological Activities. *PLOS ONE* **2012,** *7* (10), e46804.

34. Adinolfi, S.; Piccoli, R.; Sica, F.; Mazzarella, L., BS-RNase tetramers: an example of domain-swapped oligomers. *FEBS Lett* **1996,** *398* (2-3), 326-32.

35. Byeon, I. J.; Louis, J. M.; Gronenborn, A. M., A protein contortionist: core mutations of GB1 that induce dimerization and domain swapping. *Journal of Molecular Biology* **2003,** *333* (1), 141-52.

36. Gronenborn, A. M., Protein acrobatics in pairs--dimerization via domain swapping. *Current Opinion in Structural Biology* **2009,** *19* (1), 39-49.

37. Louis, J. M.; Byeon, I. J.; Baxa, U.; Gronenborn, A. M., The GB1 amyloid fibril: recruitment of the peripheral beta-strands of the domain swapped dimer into the polymeric interface. *Journal of Molecular Biology* **2005,** *348* (3), 687-98.

38. Dey, B.; Lerner, D. L.; Lusso, P.; Boyd, M. R.; Elder, J. H.; Berger, E. A., Multiple antiviral activities of cyanovirin-N: blocking of human immunodeficiency virus type 1 gp120 interaction with CD4 and coreceptor and inhibition of diverse enveloped viruses. *Journal of Virology* **2000,** *74* (10), 4562-9.

39. Matei, E.; Basu, R.; Furey, W.; Shi, J.; Calnan, C.; Aiken, C.; Gronenborn, A. M., Structure and Glycan Binding of a New Cyanovirin-N Homolog. *The Journal of Biological Chemistry* **2016,** *291* (36), 18967-76.

40. Barrientos, L. G.; Gronenborn, A. M., The highly specific carbohydrate-binding protein cyanovirin-N: structure, anti-HIV/Ebola activity and possibilities for therapy. *Mini Reviews in Medicinal Chemistry* **2005,** *5* (1), 21-31.

41. Botos, I.; Wlodawer, A., Cyanovirin-N: a sugar-binding antiviral protein with a new twist. *Cellular and molecular life sciences : CMLS* **2003,** *60* (2), 277-87.

42. Barrientos, L. G.; Matei, E.; Lasala, F.; Delgado, R.; Gronenborn, A. M., Dissecting carbohydrate-Cyanovirin-N binding by structure-guided mutagenesis: functional implications for viral entry inhibition. *Protein Engineering, Design & Selection : PEDS* **2006,** *19* (12), 525-35.

43. Fontaine, S. N.; Brown, D. R., Mechanisms of prion protein aggregation. *Protein and Peptide Letters* **2009,** *16* (1), 14-26.44. Perlenfein, T. J.; Mehlhoff, J. D.; Murphy, R. M., Insights into the mechanism of cystatin C oligomer and amyloid formation and its interaction with beta-amyloid. *The Journal of Biological Chemistry* **2017,** *292* (27), 11485-11498.

45. Yang, S.; Levine, H.; Onuchic, J. N.; Cox, D. L., Structure of infectious prions: stabilization by domain swapping. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **2005,** *19* (13), 1778-82.

46. Rambaran, R. N.; Serpell, L. C., Amyloid fibrils: abnormal protein assembly. *Prion* **2008,** *2* (3), 112-7.

47. Shameer, K.; Sowdhamini, R., Functional repertoire, molecular pathways and diseases associated with 3D domain swapping in the human proteome. *Journal of Clinical Bioinformatics* **2012,** *2* (1), 8.

48. Jaskolski, M., 3D domain swapping, protein oligomerization, and amyloid formation. *Acta Biochimica Polonica* **2001,** *48* (4), 807-27.

49. Carey, J.; Lindman, S.; Bauer, M.; Linse, S., Protein reconstitution and three-dimensional domain swapping: benefits and constraints of covalency. *Protein Science* **2007,** *16* (11), 2317-33.

50. Schlunegger, M. P.; Bennett, M. J.; Eisenberg, D., Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Advances in Protein Chemistry* **1997,** *50*, 61-122.

51. Sanson, B.; Wang, T.; Sun, J.; Wang, L.; Kaczocha, M.; Ojima, I.; Deutsch, D.; Li, H., Crystallographic study of FABP5 as an intracellular endocannabinoid transporter. *Acta Crystallographica Section D* **2014,** *70* (2), 290-298.

52. Gershenson, A.; Gierasch, L. M.; Pastore, A.; Radford, S. E., Energy landscapes of functional proteins are inherently risky. *Nat Chem Biol* **2014,** *10* (11), 884-891.

53. Karchin, J. M.; Ha, J. H.; Namitz, K. E.; Cosgrove, M. S.; Loh, S. N., Small Molecule-Induced Domain Swapping as a Mechanism for Controlling Protein Function and Assembly. *Scientific Reports* **2017,** *7*, 44388.

54. Ha, J. H.; Karchin, J. M.; Walker-Kopp, N.; Castaneda, C. A.; Loh, S. N., Engineered Domain Swapping as an On/Off Switch for Protein Function. *Chem Biol* **2015,** *22* (10), 1384-93.

55. Aranko, A. S.; Oeemig, J. S.; Kajander, T.; Iwai, H., Intermolecular domain swapping induces intein-mediated protein alternative splicing. *Nat Chem Biol* **2013,** *9* (10), 616-22.

56. Reis, J. M.; Burns, D. C.; Woolley, G. A., Optical control of protein-protein interactions via blue light-induced domain swapping. *Biochemistry* **2014,** *53* (30), 5008-16.

57. Zegers, I.; Deswarte, J.; Wyns, L., Trimeric domain-swapped barnase. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** *96* (3), 818-822.

58. Tang, Y.; Reed, S. I., The Cdk-associated protein Cks1 functions both in G1 and G2 in Saccharomyces cerevisiae. *Genes & Development* **1993,** *7* (5), 822-832.

59. Zhang, Y.; Liang, J. Y.; Huang, S.; Ke, H.; Lipscomb, W. N., Crystallographic studies of the catalytic mechanism of the neutral form of fructose-1,6-bisphosphatase. *Biochemistry* **1993,** *32* (7), 1844-1857.

60. Ramaswamy, S.; Eklund, H.; Plapp, B. V., Structures of Horse Liver Alcohol Dehydrogenase Complexed with NAD+ and Substituted Benzyl Alcohols. *Biochemistry* **1994,** *33* (17), 5230-5237.

CHAPTER IV: ELUCIDATING THE MECHANISM OF HCRBPII-BASED FLUORSCENT PROTEIN TAGS USING HIGH-RESOLUTION CRYSTAL STRUCTURES

IV-1. Fluorescent protein tags

Fluorescence microscopy facilitates the *in vitro* studies of numerous crucial biological processes, including protein trafficking,[1-3] protein-protein interactions,[4-5] enzymatic activities,[6-8] signaling,[9] and others.[10] The major tools in investigating the aforementioned processes are synthetic dyes that are post-translationally incorporated into the proteins[11-13] and different fluorescent proteins.[14] Although synthetic dyes can provide valuable labeling tools to target various organelles or proteins, this method suffers from the multiple washing steps needed to enhance the signal contrast. However, this problem can be solved using fluorescent protein tags, which do not require the exogenous chromophore for activation. Notably, fluorescent proteins can be genetically encoded to assist the labeling of a specific protein.

The hallmark of this class of proteins is Green Fluorescent Protein (GFP),[15] whose ligand is formed autocatalytically through the cyclization of three proximal residues (Ser65-Tyr66-Gly67) located inside the protein (Figure IV-1).[16-17]

Since the discovery of GFP (almost 60 years ago), this protein has been engineered to cover different colors ranging the full visible spectrum suitable for seven-color imaging of different cellular organelles.[18] However, one of the pitfalls of GFP is that it demands molecular oxygen to maturate its intrinsic fluorophore, which makes it futile for studies that should be performed in a low-oxygen or anaerobic environment.[19]

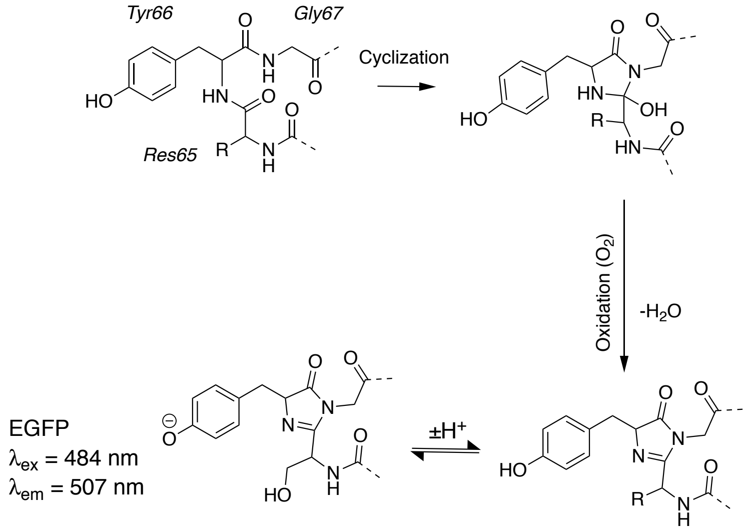
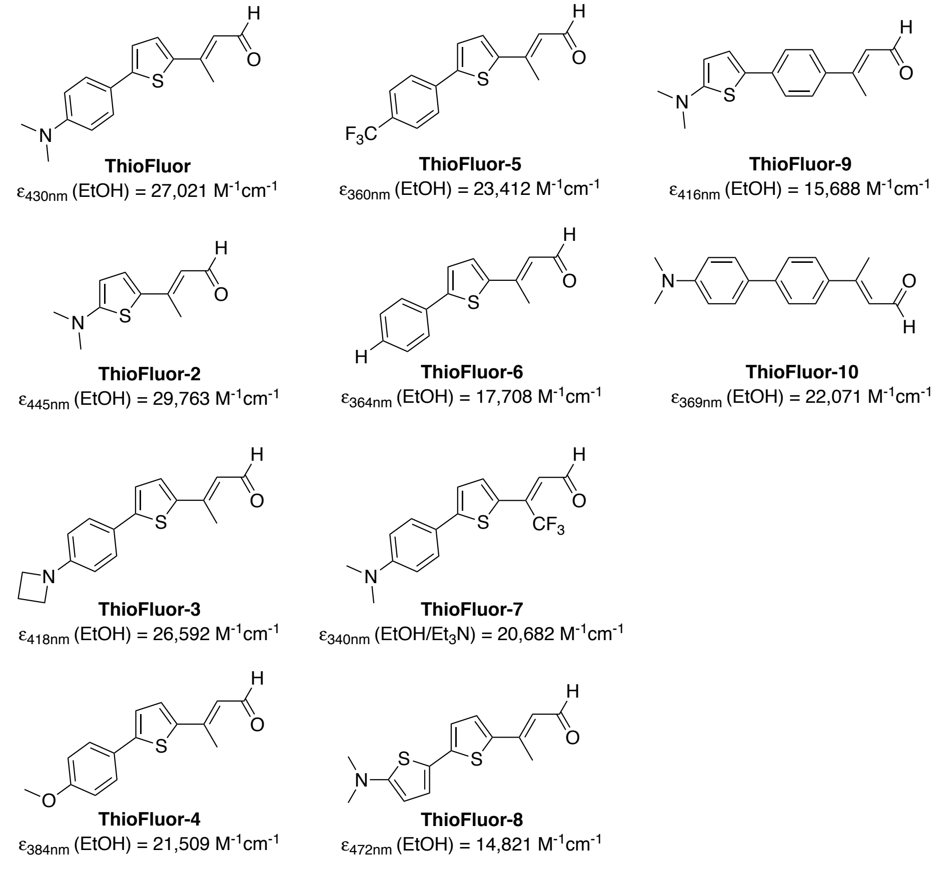


Figure IV-1. The maturation of the GFP chromophore contains cyclization, oxidation, and dehydration steps to generate an emissive ligand.

To address this problem and to increase the system's flexibility, Prof. Borhan's lab developed many fluorogenic chromophores [20-22] with aldehyde functionality that can ligate to the protein through protonated Schiff base (PSB) formation (in the same fashion as retinal) with an active Lys that was engineered inside the binding cavity (Q108K in hCRBPII and R111K in CRABPII).[20, 23-25] Having demonstrated that the absorption maximum of all-*trans*-retinal bound hCRBPII can be rationally tuned over 220 nm through electrostatic perturbations,[25] the focus is now on developing new fluorogenic chromophores whose emission can span the entire visible and near-IR spectrum. However, developing such a system requires understanding the microenvironment of the protein and managing the protein/ligand interactions. This can be facilitated using high-resolution crystal structures where the interaction can be visualized at atomic level. In this chapter, we will describe the power of using high-resolution protein crystallography and synthetic organic chemistry to create, interrogate, and examine novel functionalities in a new class of fluorescent protein tags.

IV-2. Elucidating the electronic effects on the photophysics of hCRBPII-based fluorescent proteins

To elucidate the structure and electronic relationship of the chromophore in the rhodopsin mimic, Dr. Santos and Dr. Gholami from Prof. Borhan's lab designed and synthesized a series of thiophene-based chromophore derivatives with aldehyde functionality, as shown in Figure IV-1. (Please see their Ph.D. theses for further information.)


ThioFluor
ε_{430nm} (EtOH) = 27,021 $M^{-1}cm^{-1}$

ThioFluor-5
ε_{360nm} (EtOH) = 23,412 $M^{-1}cm^{-1}$

ThioFluor-9
ε_{416nm} (EtOH) = 15,688 $M^{-1}cm^{-1}$

ThioFluor-2
ε_{445nm} (EtOH) = 29,763 $M^{-1}cm^{-1}$

ThioFluor-6
ε_{364nm} (EtOH) = 17,708 $M^{-1}cm^{-1}$

ThioFluor-10
ε_{369nm} (EtOH) = 22,071 $M^{-1}cm^{-1}$

ThioFluor-3
ε_{418nm} (EtOH) = 26,592 $M^{-1}cm^{-1}$

ThioFluor-7
ε_{340nm} (EtOH/Et_3N) = 20,682 $M^{-1}cm^{-1}$

ThioFluor-4
ε_{384nm} (EtOH) = 21,509 $M^{-1}cm^{-1}$

ThioFluor-8
ε_{472nm} (EtOH) = 14,821 $M^{-1}cm^{-1}$

Figure IV-2. ThioFluor analogs synthesized by Prof. Borhan's lab to investigate structure-property correlations of thiophene-based fluorophores with aldehyde functionality inside the binding pocket of the protein.

The electron-donating groups were created on by substituting the *para-N,N*-dimethyl aniline found in **ThioFluor** with azetidine (**ThioFluor-3**) and methoxy (**ThioFluor-4**); the electron-withdrawing containing chromophore was synthesized with trifluoromethyl in the para position (**ThioFluor-5**); and hydrogen was used as a reference (**ThioFluor-6**). The electron acceptor property of the iminium was altered by generating **ThioFluor-7**, in which a trifluoromethyl group was introduced in the β-position relative to the aldehyde. Additionally, to demonstrate the effect of conjugation across the π system, **ThioFluor-8, ThioFluor-9,** and **ThioFluor-10** were synthesized. **ThioFluor-2**, as the shortest chromophore, was created to make a cyanine-reminiscent dye structure with high conjugation. To launch this study, we needed a protein platform to investigate the effects of various derivatizations on **ThioFluor** in terms of protein/chromophore behavior. We employed Q108K:K40L:T51V:T53S:R58W:Y19W:L117E (**L1**:L117E) as a template that can generate a high pK_a form of chromophore-iminium through the acidic residue

Table IV-1. Spectroscopic properties of **ThioFluor**-PSB analogs coupled with **L1**:L117E. (The following data was prepared by Dr. Santos.)

Ligand	**λ_{abs}**	**λ_{em}**	**Φ**	**ε ($M^{-1}cm^{-1}$)**	**pK_a**
ThioFluor-5	423	-	-	23,530	10.9
ThioFluor-6	424	502	-	13,820	10.6
ThioFluor-4	466	543	-	29,878	10.9
ThioFluor-10	489	675	0.06	7,050	11.3
ThioFluor-2	515	536	0.13	50,578	11.3
ThioFluor-3	548	679	0.27	21,223	10.9
ThioFluor	558	673	0.31	15,091	11.2
ThioFluor-9	562	700	0.02	15,406	11.1
ThioFluor-8	641	715	0.09	22,948	11.2
ThioFluor-7	714	780	-	22,517	4.3

(L117E) interaction with the iminium nitrogen. Furthermore, the presence of aromatic residues such as R58W and Y19W rigidifies the chromophore inside the binding cavity resulting in a well-ordered chromophore in the protein binding cavity (Figure IV-3). Using the aforementioned construct, Dr. Santos measured the Spectroscopic properties of the **ThioFluor**-PSB analogs, as enumerated in Table IV-1. Also, as shown in Figure IV-3, the crystal structures of the **ThioFluor**-PSB analogs were obtained at atomic resolution. Fortuitously, as expected, a well-ordered chromophore, characterized by the high quality of the electron density nestled in the binding pocket of the protein (Figure IV-4).The absorption wavelength of each PSB-analogue portrayed the relative correlation between the maximum wavelength and degree of conjugation affected by the electron-withdrawing and electron-donating substituents, (Table IV-1). For instance, the most blue-shifted absorption wavelengths for chromophore/protein complexes were seen for

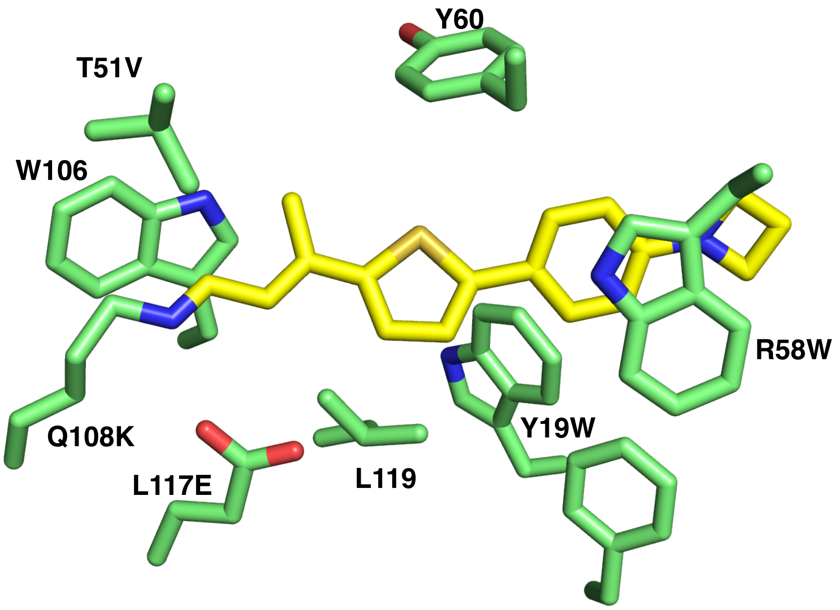


Figure IV-3. ThioFluor3 bound with the **L1**:L117E variant of hCRBPII. The aromatic residues, Y19W, R58W, and W106, and other proximal amino acids are highlighted in the binding pocket.

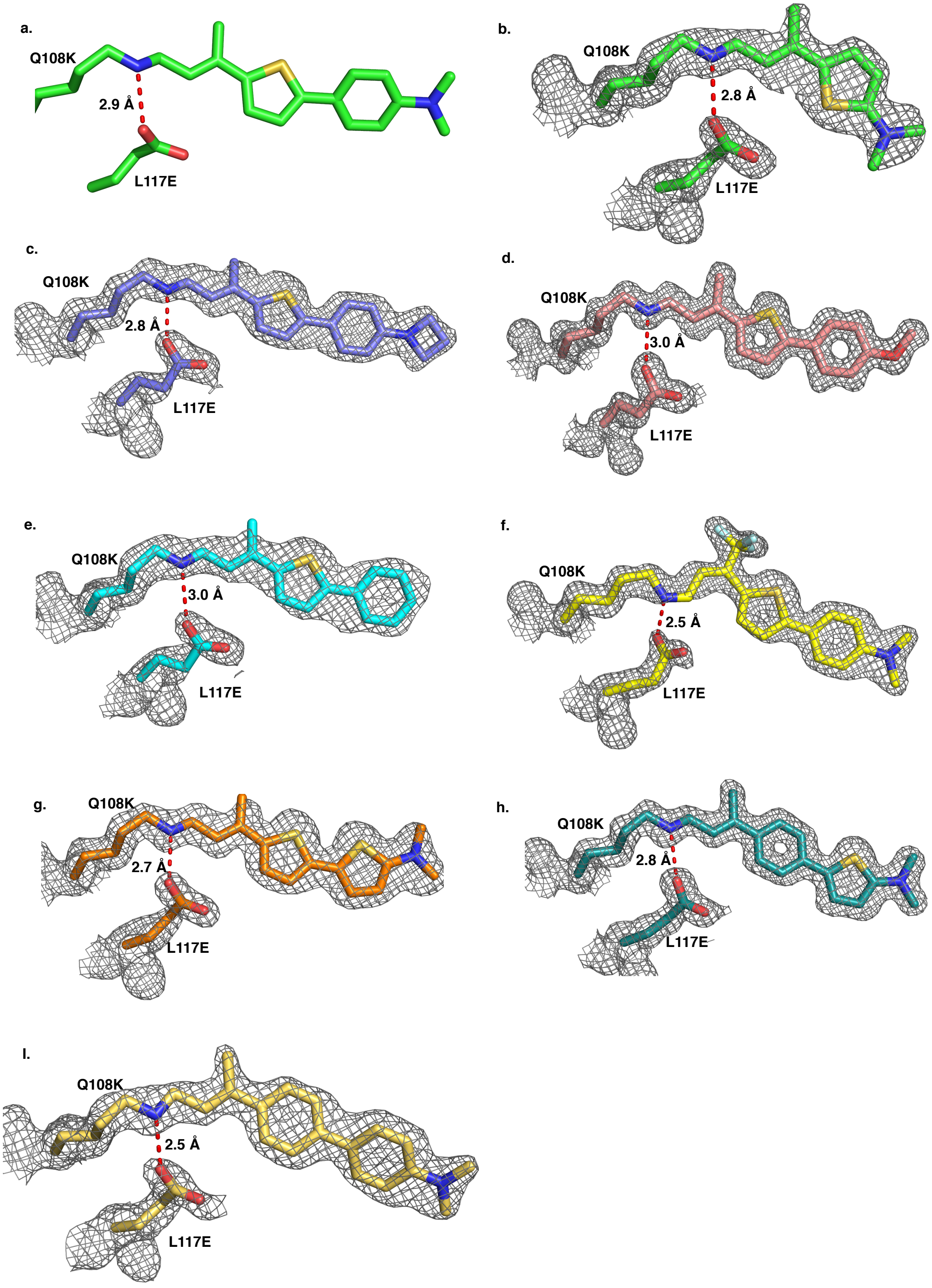


Figure IV-4. Crystal structures of Q108K:K40L:T51V:T53S:R58W:Y19W:L117E (**L1**:L117E) bound with **ThioFluor** analogs. a. **L1**:L117E crystal structure bound with ThioFluor. b. **ThioFluor2**. c. **ThioFluor3**. d. **ThioFluor4.** e. **ThioFluor6**. f. **ThioFluor7**. g. **ThioFluor8.** h. **ThioFluor9.** i. **ThioFluor10.**

ThioFluor-5 (λ_{abs} = 423 nm) and **ThioFluor-6** (λ_{abs} = 424 nm), containing the trifluoromethyl group and hydrogen atom, respectively. Also, as shown in Table IV-I, the iminiums of these two complexes are non-emissive, but they both have a high pK_a. Unfortunately, attempts at crystallizing **ThioFluor5** were not successful. However, the crystal structure of **ThioFluor-6** was obtained at atomic resolution. As shown in Figure IV-6e, the crystal structure depicts *trans*-iminium with L117E directly interacting with the nitrogen of iminium, explaining the high pK_a of the chromophore.

Table IV-1 shows that **ThioFluor-4** contains an electron-donating methoxy group that reveals a more red-shifted absorption wavelength (λ_{abs} = 466 nm) compared with hydrogen or trifluoromethyl. The high-resolution crystal structure of this mutant shows the same features as observed for **ThioFluor6**, such as *trans*-imine and L117E conformations, as mentioned above. The overlay of this structure with **ThioFluor6** illustrates the similar trajectory of both chromophores (Figure IV-5). Therefore, the origin of the aforementioned difference in absorption wavelength should be exclusively due to the electron-donating character of the methoxy group. Nonetheless, the pigmented chromophore is still non-emissive (Table IV-1), highlighting the lack of ability of the methoxy group to create a push-pull system like the *N, N*-dimethyl group (**ThioFluor**).

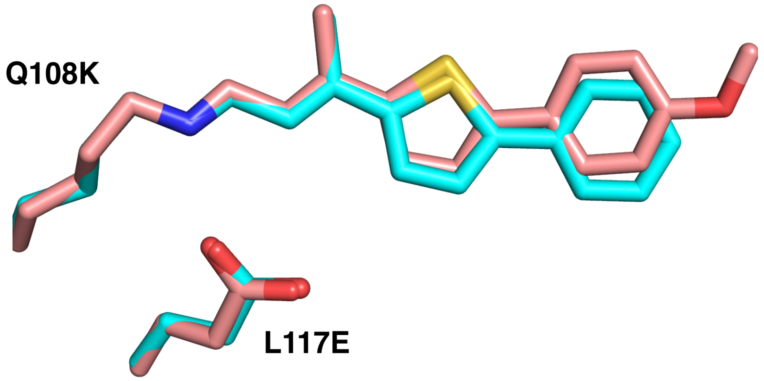


Figure IV-5. Overlay of **L1**:L117E bound with **ThioFluor6** (cyan) and **L1**:L117E bound with **ThioFluor4** (pink). (RMSD: 0.1)

The azetidine-based chromophore, **ThioFluor-3,** and the *N, N*-dimethyl chromophore, **ThioFluor,** depict larger red-shifted absorption wavelengths compared with **ThioFluor6**, which only possesses a hydrogen atom in the para position.

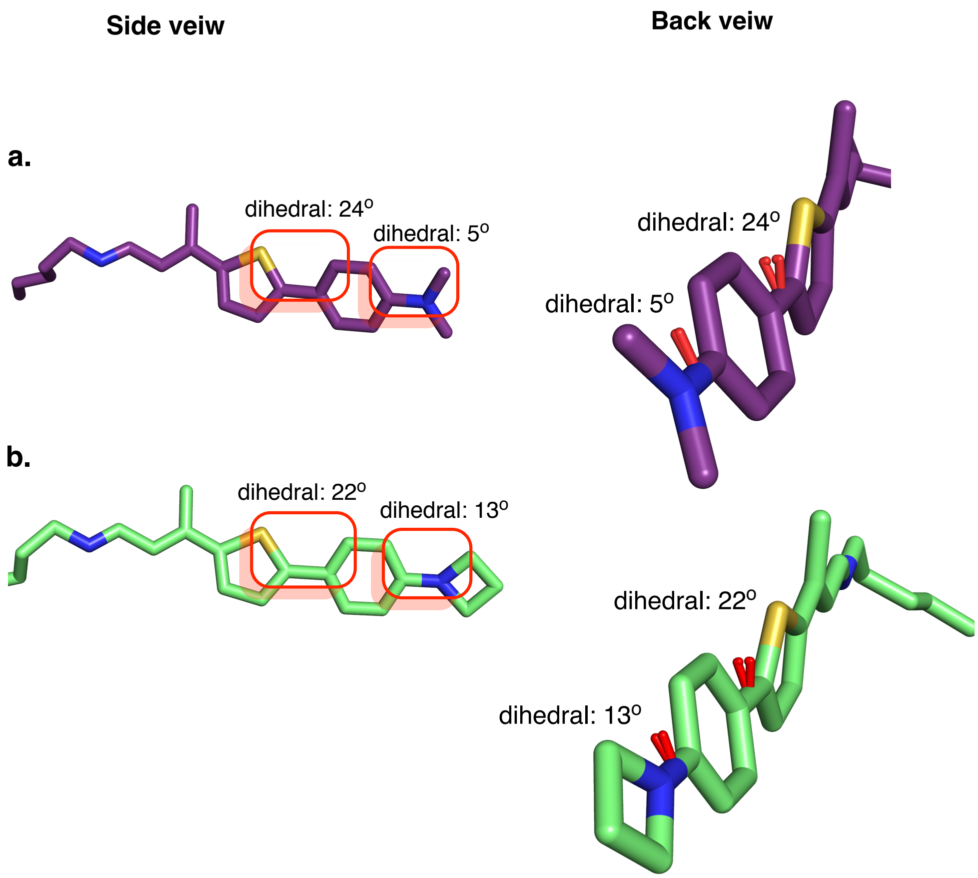


Figure IV-6. a. **L1**:L117E crystal structure bound with **ThioFluor** with two presentations. The crystal structure depicted the almost-planarity of the N,N-dimethyl group with the chromophore's backbone. b. **L1**:L117E crystal structure bound with **ThioFluor3**. The crystal structure depicted the almost-planarity of the azetidine group with the chromophore's backbone.

Looking at the crystal structures of both chromophores, **ThioFluor** and **ThioFluor-3** demonstrates that the origin of this change in absorption is solely due to the change in the electronic structure of the chromophore as a result of fixing two electron-donating groups in the para position that essentially extend the conjugation. Notably, the crystal structure portrays the near-planarity of both para-substituted analogs with the rest of chromophore's backbone, confirming the cited hypothesis. The planarity further explains the higher quantum yields for these two chromophores, as shown in Table IV-1.

The placement of the trifluoromethyl group in the β-position, **ThioFluor7** (λ_{abs} = 714 nm)**,** to the aldehyde gave a more red-shifted absorption wavelength in comparison with **ThioFluor** (λ_{abs} = 558 nm). The high-resolution crystal structure obtained

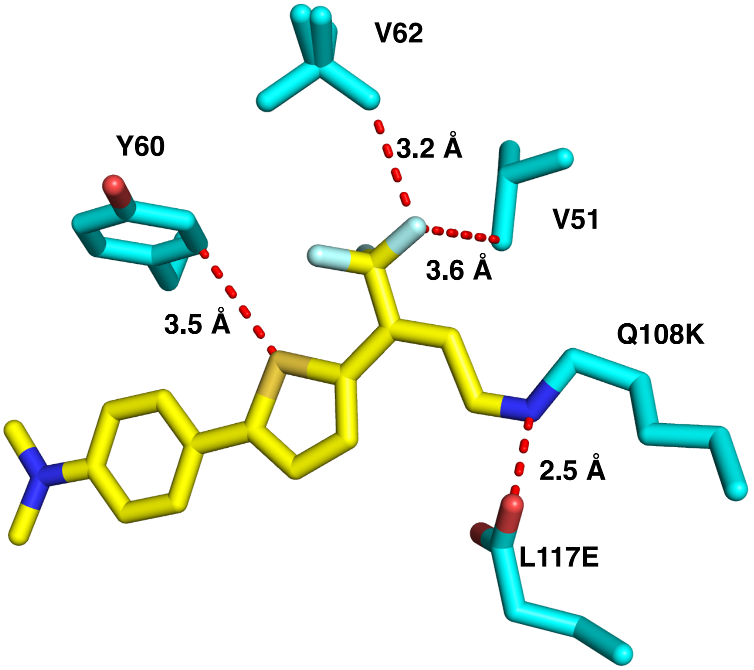


Figure IV-7. The binding pocket of **ThioFluor7** bound with **L1**:L117E demonstrated the interacting residues. Two C-H…F interactions trifluoromethyl group and V62 and V6, S… π interaction which can be seen in all of the **ThioFluor** analogs with **L1**:L117E. and finally he low barrier hydrogen bond with L117E and nitrogen of iminium.

for **ThioFluor7** shows a *cis* double bond in the structure, with an overall trajectory similar to **ThioFluor**. Interestingly, the chromophore is not as emissive as **ThioFluor**, which might either be due to the *cis* conformation of the double bond or the presence of the trifluoromethyl group substitution. However, these hypotheses should be confirmed or ruled out through future experiments by synthesizing the *trans* isomer or photochemically producing the *trans* form from the *cis* isomer. As shown in Figure IV-7, the chromophore also made S…π, C-H…F, and salt bridge interactions with Y60, V62, and L117E, respectively.

As mentioned above, the bathochromic shifts were obtained either by installing electron-donating groups in the para position (**ThioFluor** and **ThioFluor3**) or introducing the electron-withdrawing group in the β-position relative to the aldehyde functional group (**ThioFluor7**). The next step was to investigate how changes in aromaticity can alter the photophysics of the chromophore. **ThioFluor-8** was used to fulfil this purpose. As shown in the structure of the chromophore (Figure IV-4g), the phenyl ring was substituted with a thiophene, which maintains less resonance energy in comparison with phenyl. Therefore, the less-aromatic character of thiophene is expected to facilitate breaking the aromaticity and enhancing resonance through the π system. This would lead to lowering the energy gap between the HOMO and LUMO and a subsequent red-shifted absorption wavelength.

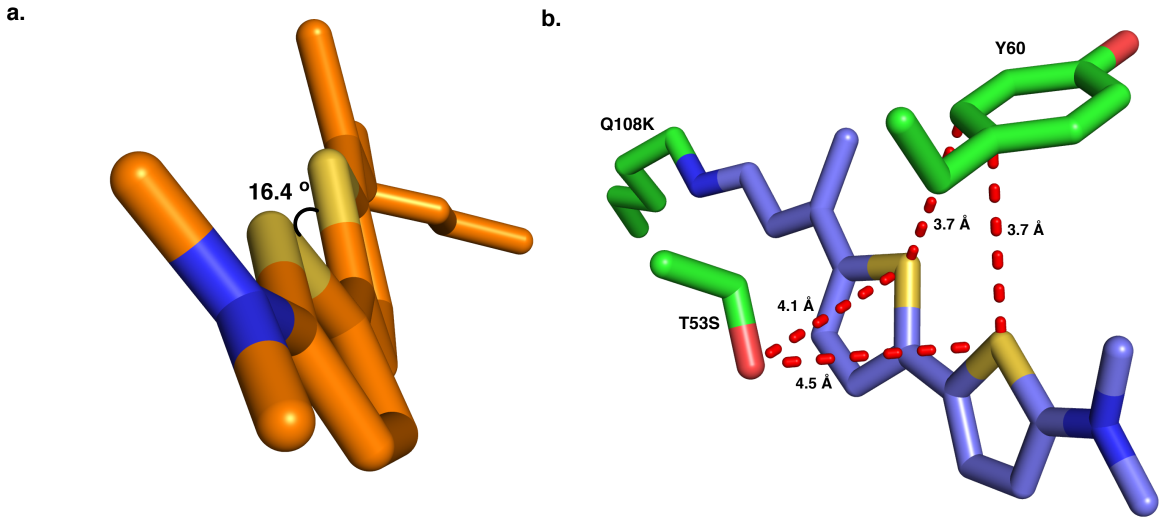


Figure IV-8. a. Approximate coplanarity of the two thiophenes in **ThioFluor8**. b. Position of two thiophenes provides a suitable conformation for metal-binding. This ligand can be used in the future as a metal sensor, where metal-binding can tune the absorption and emission properties of **ThioFluor8.** The residues that are in proximity with the two rings are highlighted. Metal binding site that bridges chromophore and protein can be envisioned by mutation of these residues.

As depicted in Table IV-1, **ThioFluor-8** bound with the protein showed λ_{abs} = 641 nm, which is much more red-shifted than **ThioFluor**. The high-resolution crystal structures of this chromophore were obtained at 1.6 Å resolution, demonstrating the near coplanarity between the two thiophene rings (Figure IV-8a). Therefore, the red-shifted absorption wavelength can be simply attributed to the electronic effects cited above. Another interesting feature of the chromophore, as shown in Figure IV-8b, is that the thiophene groups point in the same direction within the crystal structure. This trajectory is optimized to bind metals in a bidentate fashion. The metal-binding can further tune the fluorescence efficiency of the ligand by altering the aromaticity of the thiophenes. As a proof of principle, KAuCl4 was added to **ThioFluor8/L1**:L117E complex led to formation of short lived 545 nm species as shown in Figures IV-9a and IV-9b. The spectra of the controls including **ThioFluor8/L1**:L117E, as well as **ThioFluor/L1**:L117E delineated little changes upon

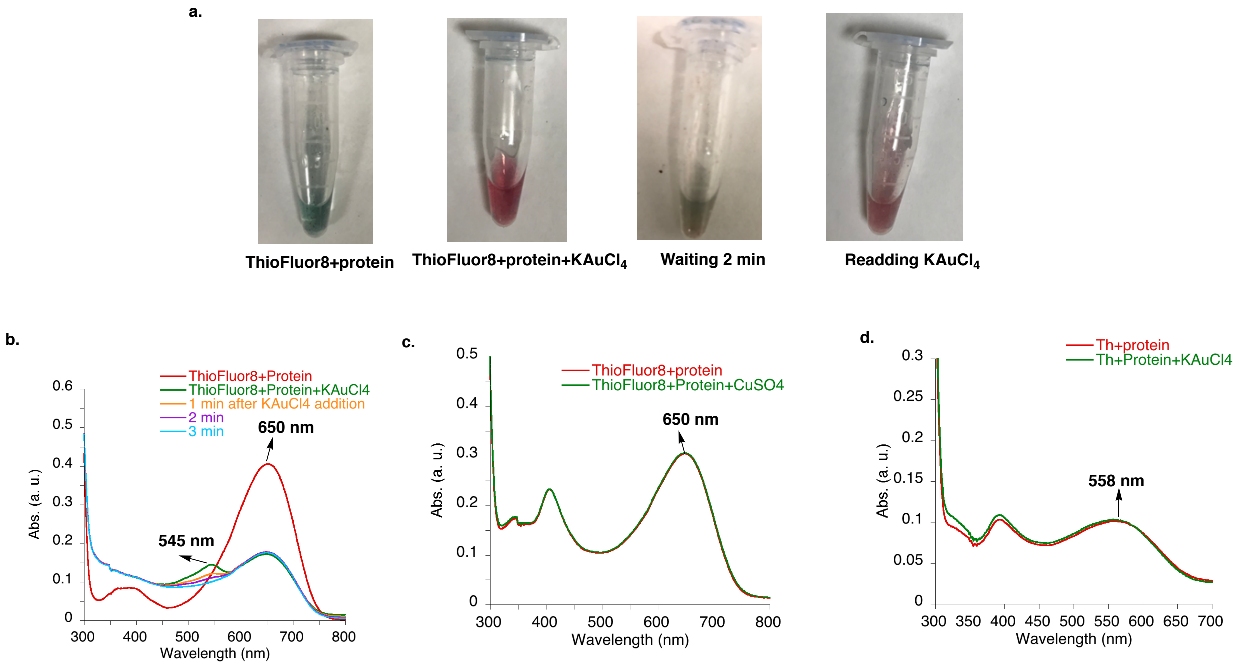
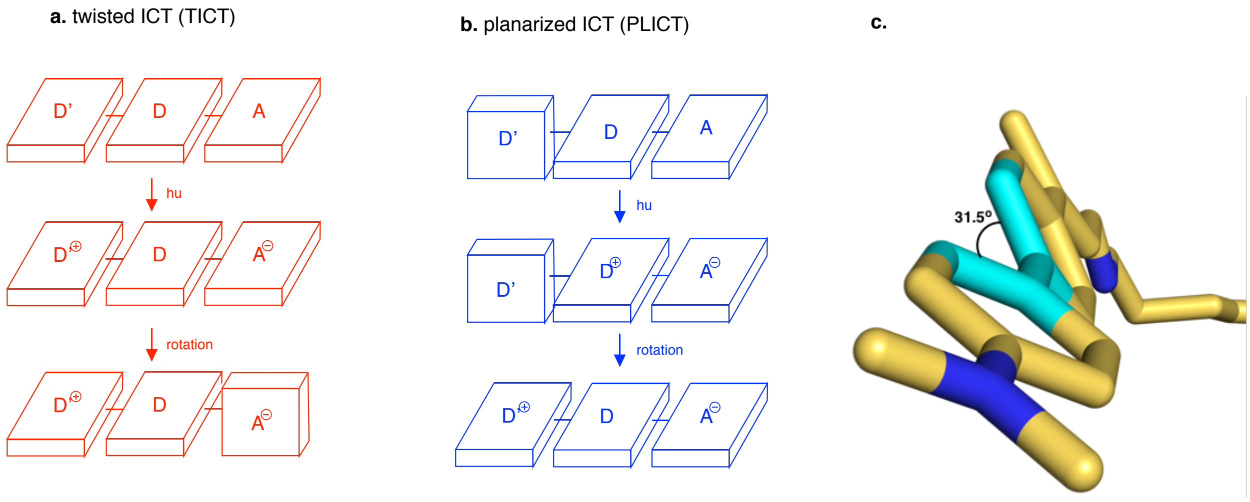


Figure IV-9. a. Color changes upon: iminium formation of **ThioFluor8/L1**:L117E (Protein concentration ~35 μM - 0.5 equivalent **ThioFluor** was added., Adding 100 μM KAuCl4 to the mixture of **ThioFluor8/L1**:L117E, incubating the previous mixture for two minutes, readding 100 μM KAuCl4. b. The UV-vis spectrum of **ThioFluor8/L1**:L117E pre and post 100 μM KAuCl4 addition. c. The UV-vis spectrum of **ThioFluor8/L1**:L117E pre and post 100 μM CuSO4 addition as a control. d. The UV-vis spectrum of **ThioFluor/L1**:L117E pre and post 100 μM KAuCl4 addition as a control.

$CuSO_4$ and $KAuCl_4$ additions respectively. Although the complexation of protein with Au^{3+} was not stable, it shows a promising preliminary result of metal/ligand/protein interactions in an unoptimized binding pocket causing a hypsochromic shift in the absorption wavelength of **ThioFluor8/L1**:L117E complex. This approach can be further pursued by imposing novel mutations in the binding pocket that can create third or fourth coordination sites and increase the selectivity of the binding site for a specific metal, accomplished by both the ligand and the protein binding site. Specifically, to optimize this interaction, mutations were made on the Y60 and T53 (Y60C and T53C) positions in an ongoing project.

Next, the replacement of thiophene with phenyl was investigated to yield **ThioFluor-10** featuring two adjacent phenyl groups**.** In this case, a blue shift was observed in the absorption wavelength (λ_{abs} = 489 nm) comparable with **ThioFluor** (λ_{abs} = 558 nm), yet with a significant Stokes shift of 186 nm between the excitation and emission wavelengths (Table IV-1).



a. twisted ICT (TICT)

b. planarized ICT (PLICT)

c.

Figure IV-10. Representation of the geometric arrangement of the donor and acceptor groups upon excitation in: a. twisted intramolecular charge transfer (TICT) and b. planarized intramolecular charge transfer (PLICT) c. Dihedral angle between the two phenyl rings in **ThioFluor10** measured 31.5º.

The crystal structure (at 1.8 Å resolution, Figure IV-4i) showed the *trans*-imine and other features similar to those observed for previous structures. The large change in absorption may be due to a decrease in delocalization of positive charge on the molecule as a result of enhancing the aromaticity by placing two phenyl rings instead of a phenyl and thiophene ring in **ThioFluor**.

Looking at the crystal structure, the torsional angles between the two phenyl groups was 31.5^{o}, resulting in the donor and acceptor groups being between planar and perpendicular orientations. The excitation of the chromophore similar to **ThioFluor** could lead to an intramolecular charge transfer (ICT) from donor to acceptor,[26-28] with an ensuing rotation around the single bond linking the donor and acceptor groups. This reorientation in the excited state can proceed to achieve either a complete perpendicular conformation, which is characteristic of a twisted intramolecular charge transfer (TICT), or give rise to a planar orientation (PLICT), which could explain the large Stokes shifts observed (Figure IV-11). Although ruling out either of these mechanisms is not possible

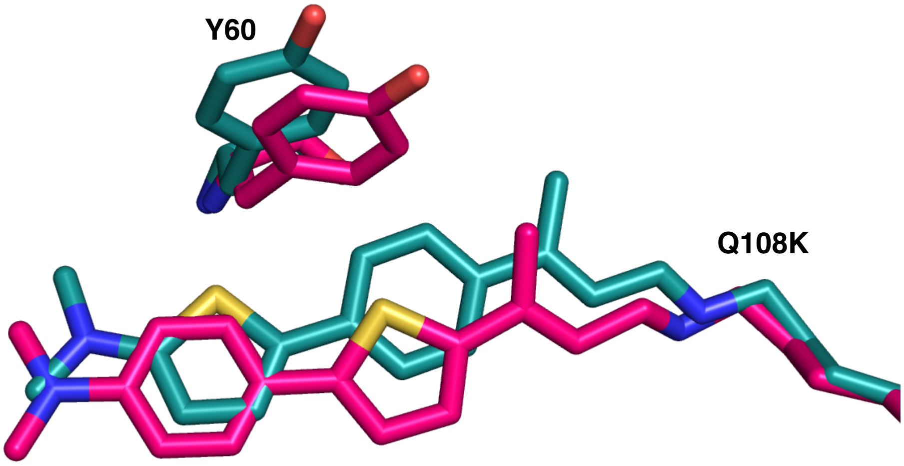


Figure IV-11. Overlay of **L1**:L117E crystal structures bound with **ThioFluor** (pink) and **ThioFluor9** (green), showing a small change in trajectory, which affects the position of Y60.

with the ground state crystal structure, the TICT mechanism is more correlated with a low quantum yield (0.06), as listed in Table IV-1 for **ThioFluor-10**.

Switching the order of the thiophene and phenyl groups (**ThioFluor-9)** did not lead to a significant change in the absorption wavelength of the chromophore. The crystal structure did not show a large difference from **ThioFluor**, as shown in Figure IV-11.

Lastly, the crystal structure of the shortest chromophore (Figure IV-4b) surprisingly revealed a similar trajectory of the chromophore in the binding cavity, signifying the roles of aromatic residues in rigidifying the even-shorter chromophore (Figure IV-12). Additionally, for the first time, **ThioFluor2** depicts a new conformation of thiophene. As shown in Figure IV-12b, the reason behind this new conformation might be due to the steric buttressing effect of Y60 (which can impinge on the sulfur atom of thiophene), along with a higher flexibility of **ThioFluor2** in the binding pocket.

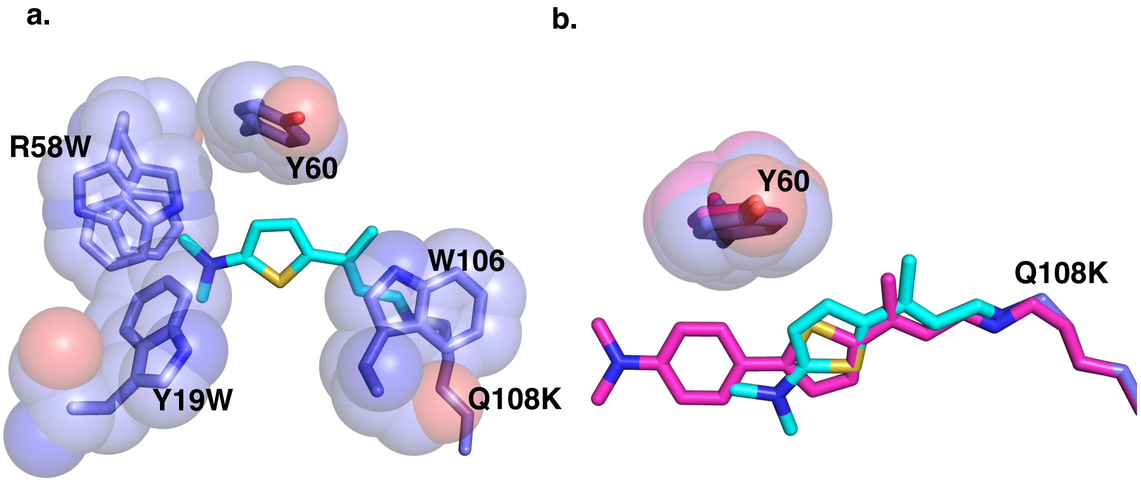


Figure IV-12. a. Binding pocket of **L1**:L117E bound with **ThioFluor2**, depicting how the aromatic residues sandwich the chromophore in the binding pocket. b. Overlay of **L1**:L117E bound with **ThioFluor2** (cyan) and **ThioFluor5** (pink) crystal structures, showing a new conformation of thiophene for **ThioFluor2.**

Since all of the crystal structures (other than **ThioFluor-7**, which shows a *cis* double bond) have a similar trajectory and imine configuration, we can confidently conclude that the electronic effects are a significant factor contributing to the results observed in the absorption maxima, Stokes shifts, and quantum yields of the **L1**:L117E/ **ThioFluor** analog complexes.

IV-3. Using high-resolution protein crystallography to optimize a ratiometric pH sensor

Prof. Borhan's lab was successful in creating ratiometric pH sensors, a prominent application of engineered fluorescent protein tags. The first attempt at creating such a system was achieved in hCRABPII, where two variants of the protein with more than 100 nm difference in absorption wavelength, as well as distinct pK_as, were utilized for the absorption-based pH sensing. [24] Borhan's lab further created a novel fluorescence-based ratiometric pH sensor in hCRBPII, using a julolidine fluorophore. [22] This design relies on a titratable amino acid side chain, L117E, that upon acidification, becomes protonated and alters the absorption of fluorophore, leading to a subsequent change in emission intensity at the same excitation. The mechanism of L117E controlling the photophysics of the chromophore is based on localizing and delocalizing the positive charge on the polyene, when L117E is deprotonated and protonated, respectively. [22]

Although this approach reveals a concentration-independent pH measurement, it suffers from the fact that julolidine is not cell-permeable.

To address this problem, Dr. Santos employed the same system, this time using **ThioFluor** as a chromophore. As shown in Figure IV-13, the acidification of the **L1**:L117E/**ThioFluor** complex as a template led to a bathochromic change in wavelength from 558 nm to 614 nm. The plot of pH vs. absorption of **L1**:L117E/**ThioFluor** showed two pK_as, 6.6. and 11.2, possibly related to L117E and iminium, respectively (Table IV-2). Although the presence of L117E increased the rate of PSB formation and showed a ratiometric pH change with a fluorescent readout, it had the disadvantage of lowering the extinction coefficient by almost half.

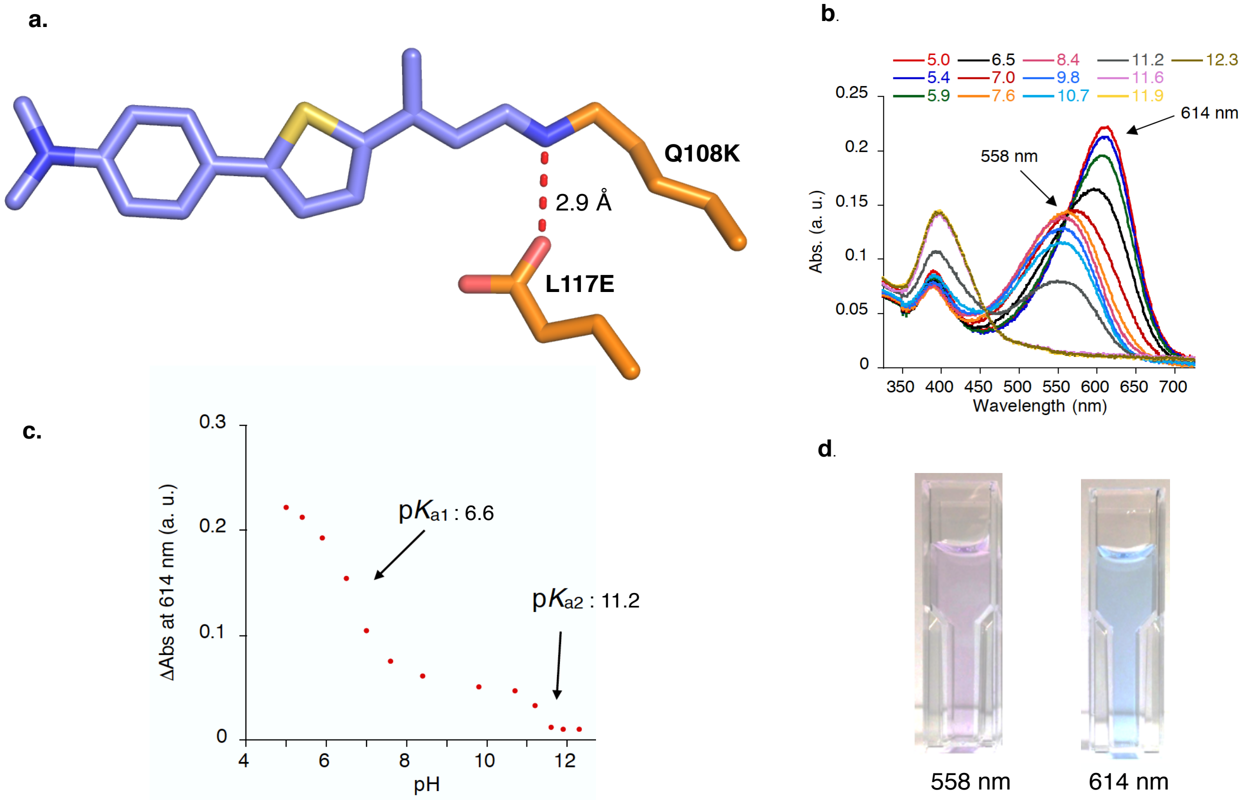


Figure IV-13. Crystal structure of **ThioFluor** bound with **L1**:L117E, showing how L117E acts as a counter-ion for the nitrogen of iminium. b. UV-vis spectrum of pH titration for the **L1**:L117E/**ThioFluor** complex. c. Plot of abs. vs. pH reveals two apparent pK_as, 6.6 and 11.2 for L117E and iminium, respectively. d. Color changes upon acidification of the **L1**:L117E/**ThioFluor** complex. (The data for UV-vis. and pK_a was provided by Dr. Santos.)

To increase the extinction coefficient of the **L1**:L117E / **ThioFluor** complex, A33W was introduced to **L1**:L117E to create **L1**:L117E:A33W. The presence of A33W led to a sufficient increase in the extinction coefficient (40,000 $M^{-1}cm^{-1}$) in a protonated form of L117E, while in the deprotonated state of L117E, the extinction was almost the same as **L1**:L117E (15,000 $M^{-1}cm^{-1}$). Furthermore, A33W insertion did not alter the quantum yield, resulting in a two-fold increase in brightness.

The crystal structure of **L1**:L117E:A33W bound with **ThioFluor** was successfully determined at 1.3 Å resolution. Interestingly, the structure revealed a *cis*–iminium in contrast with the **L1**:L117E/**ThioFluor** structure. As shown in Figure IV-14, the overlay of **L1**:L117E/**ThioFluor** and **L1**:L117E:A33W/ **ThioFluor** shows that the bulky A33W forces the molecule more deeply into the binding pocket, resulting in the formation of the *cis*-iminium (Figure IV-14a). The deeper penetration into the binding cavity than **L1**:L117E/**ThioFluor**, may enhance the electrostatic interaction of the chromophore and protein (Figure IV-14b). The pK_a titration illustrated two pK_as, 11.3 and 8, corresponding

Table IV-2. Effect of L117D versus L117E in the development toward a single fluorescent protein ratiometric sensor.

hCRBPII mutant	λ_{abs}[b]	λ_{abs}[c]	ΔAbs	pK_a[d]	pK_a[e]	Φ[f]	$t_{1/2}$ (min)[g]	E.R.[h]
L1:L117E	558	614	56	11.2	6.6	0.31	82	7
L1:L117E:A33W	539	628	89	11.3	8.0	0.28	241	-
L1:L117D	525	618	93	10.9	6.7	0.30	14	21
L1:L117D:Q4F	529	623	94	11.0	5.7	0.31	8	16

L1 = Q108K:K40L:T51V:T53S:R58W:Y19W. b. λ_{max} of absorbance at basic pH. c.λ_{max} of absorbance at acidic pH. d. p*K*a corresponding to the iminium. e. p*K*a corresponding to L117D/E. f. Quantum yield measured at ~ pH 7.2. g. Kinetic measurements were performed at 23°C with 20 μM protein and 0.5 equiv ligand. PSB formation was monitored by UV-Vis at λ_{max} for each complex over time at ~ pH 7.2. Data was then fit to second order kinetics. h. ER = the highest emission ratio collected as a result of excitation at 514 nm and 633 nm. The data for this table were provided by Dr. Santos.

to iminium and glutamate, respectively (Table IV-2). In this case, the high pK_a of glutamate could be due to the *cis*-iminium formation, where the nitrogen of iminium is not in direct interaction with the *cis*-iminium. Also, in comparison with **L1**:L117E, a larger change between the two iminium absorption maxima was obtained upon acidification (89 nm vs. 56 nm). This might be due to the presence of A33W in the mouth of the binding cavity, which leads to the further delocalization of the charge on the chromophore by dispelling some of the ordered water molecules out of the binding cavity, as previously reported for

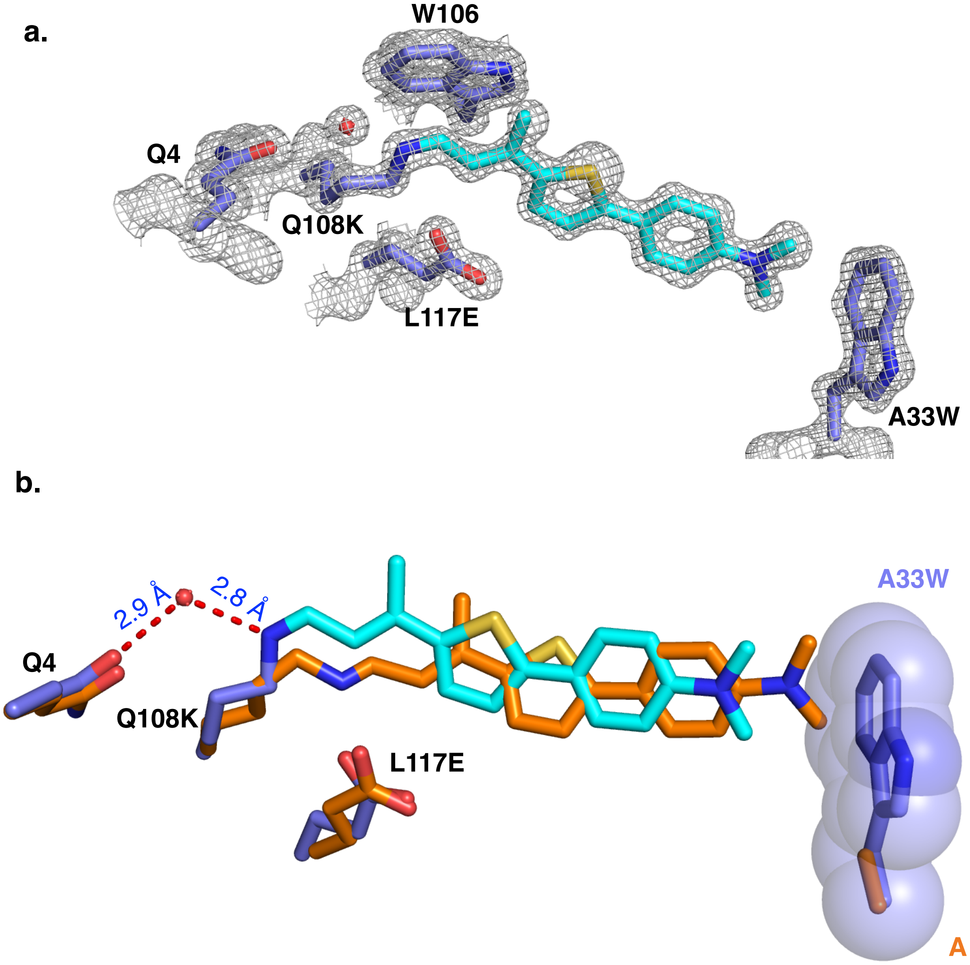


Figure IV-14. The crystal structure of **L1:**L117E:A33W/**ThioFluor** depicts the *cis*-iminium *cis*-iminium in **L1**:L117E**:**A33W/**ThioFluor** further stabilized through a water mediated hydrogen bond between Q4 and iminium (the electron density is contoured to 1σ.).
b. The Overlay of **L1:**L117E/**ThioFluor**(orange) vs. **L1**:L117E**:**A33W/**ThioFluor** (cyan) revealed the presence of A33W sterically pushes the chromophore and locates it deeper in the binding pocket.

retinylidene variants.[25] Although this construct had the advantage of higher brightness and a larger separation between the two iminium maxima, the rate of PSB formation was slower than that for **L1**:L117E (Table IV-2).

To further optimize the system, Dr. Santos investigated the effect of L117D vs.

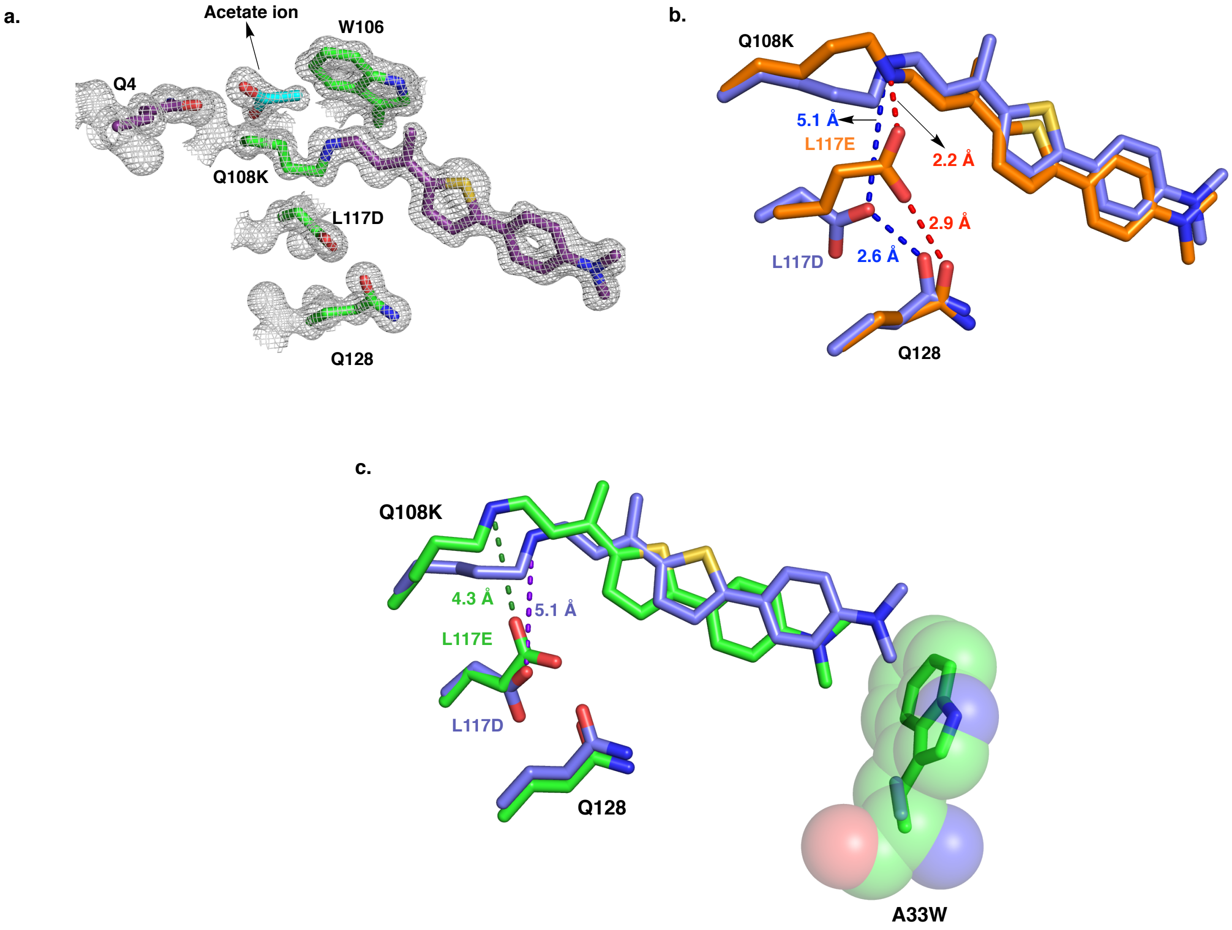


Figure IV-15. a. Crystal structure of **L1:**L117D/ThioFluor shows a cis-*iminium* configuration. Unfortunately, in this case the acetate ion coming from the crystallization conditions occupied the water molecule site between Q4 and the nitrogen of imine. b. Overlay of **L1:**L117D/**ThioFluor** (purple) and **L1:**L117E/**ThioFluor** depicts a different conformation of L117D vs. L117E, which is very far from the nitrogen of iminium. c. Overlay of two *cis*-iminium structures coming from **L1:**L117D/**ThioFluor** (purple) and **L1**:L117E**:**A33W/**ThioFluor** highlights the importance of A33W in locating the chromophore deeper in the binding pocket.

L117E. **L1:**L117D (**L1**:L117D) was created and the pK_a titration was performed, showing similar effects on the absorption wavelength of the imine upon acidification. The two pK_as were also measured at 11.0 and 7.0 for iminium and aspartate respectively. The crystal structure of this construct was achieved at 1.0 Å resolution, demonstrating the *cis*-iminium formation (Figure IV-15a). The overlay of this structure with L117E depicts the large change in the side chain of Lys108, which is a further indication of the different isomeric forms of imine. Additionally, L117D, in contrast to L117E, did not interact with the nitrogen of iminium; instead making a hydrogen bond with Gln128 (Figure IV-15b). This could either be due to the *cis* vs. *trans* configuration of iminium in L117D vs. L117E, leading to an increased distance between the L117D position and the nitrogen of iminium, or it could be due to the crystallization pH (4.0-4.5) leading to the protonation of L117D and a change in its conformation. To find an answer to this question, we needed to obtain the crystal structure of L117D containing a construct with the *trans* conformation of the imine bond. As shown in Figure IV-13, the *cis*-iminium in **L1**:L117D is stabilized through a water-mediated hydrogen bond with Q4. The hydrogen bond interaction between Q4 and the nitrogen of the *cis*-iminium was removed though the Q4F mutation to yield a **L1:**L117D:Q4F construct that presumably would lead to the formation of the *trans* form of iminium by destabilizing the *cis*-iminium isomer. The crystal structure of the **L1**:L117D:Q4F/**ThioFluor** complex was determined at 1.6 Å resolution, Figure IV-16. The presence of the Q4F mutation caused a one-unit decrease in the pK_a of L117D (which makes this construct useful for physiological pH-sensing applications (working pH 5.7 to 7.7) (Table IV-2).

As shown in Figure IV-15, the structure depicts a *trans*-iminium configuration with the exact same conformation of L117D as observed in **L1**:L117D, where L117D interacted directly with Q128. This observation suggests that this conformation of L117D might correspond to the protonated form of L117D, which is produced as a result of a low pH of crystallization or L117D conformation remains in the same way in different pHs and remotely affects the chromophore photophysics. Unfortunately, attempts to grow the crystals in basic pH or soaking the crystals in higher pH cryogenic buffers were not

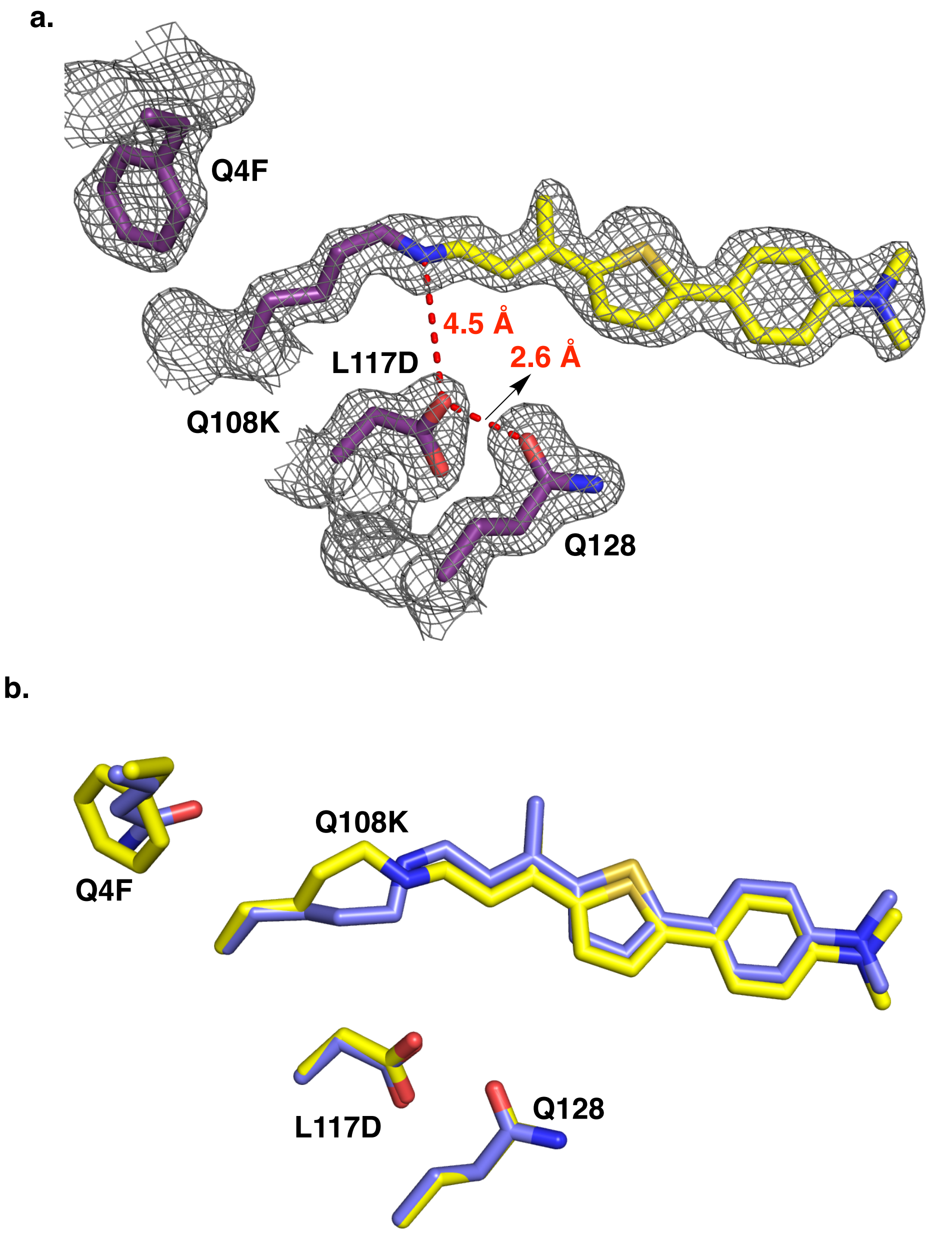


Figure IV-16. a. The crystal structure of **L1:**L117D:Q4F/**ThioFluor** demonstrates the *trans*-imine configuration and the same conformation of L117D as observed for **L1:**L117E/**ThioFluor**. b. The overlay of **L1:**L117D:Q4F/**ThioFluor** (yellow) and **L1:**L117E/**ThioFluor** (purple). Interestingly the position of nitrogen of imine is almost the same for both structures.
In the pH of crystallization (~4.5) both aspartate and iminium would be anticipated to be protonated.

successful. In the latter case, the structure did not show any tangible changes in different pHs of soaking. However, this hypothesis should be further confirmed by neutron diffraction, where the visualization of hydrogen atoms' positions on L117D and the iminium nitrogen in different pHs can be accomplished.

IV-4. Elucidating the mechanism of excited-state proton transfer using high-resolution crystal structures

Excited-state proton transfer (ESPT) is a mechanism that largely ensues through intermolecular or intramolecular hydrogen bonding. Consider a ground state of A1–H···A2 (where A1 and A2 are electronegative atoms, such as N or O), in which two π-conjugated moieties or molecules are joined together through hydrogen bonding. Upon excitation of the system with light, and as a result of ESPT, a new state is achieved where the hydrogen atom transfers from one moiety to another to generate an A1···H–A2 system. In this case, single or dual fluorescence can be observed for the ESPT system,

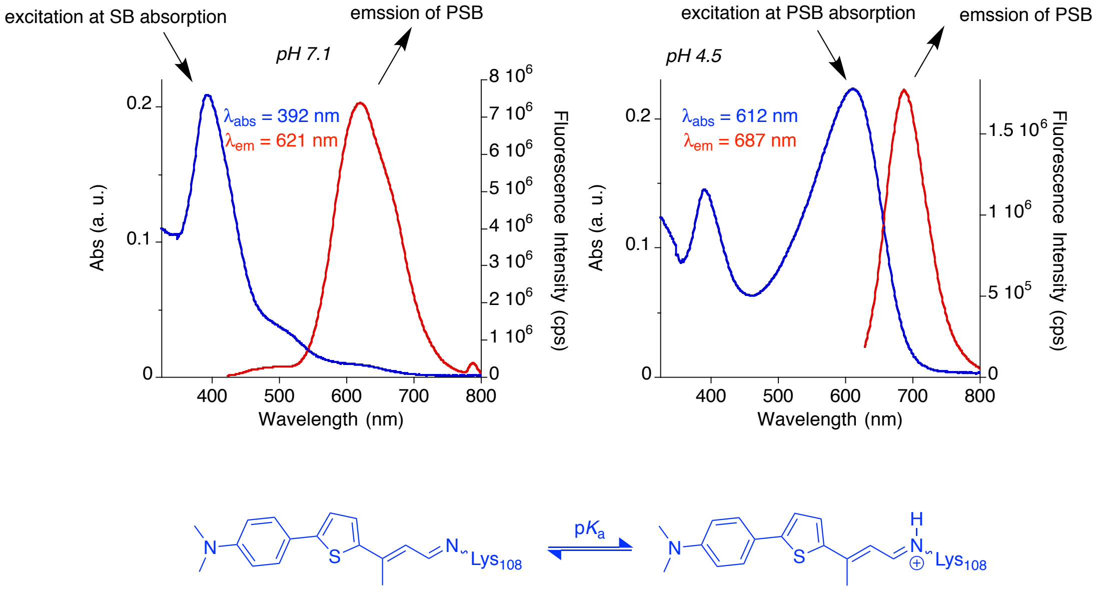


Figure IV-17. UV-vis and emission spectra of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W/**ThioFluor** SB (pH 7.1) and PSB (pH:4.5). The data were provided by Dr. Santos.

hinging upon the energy barrier. [29] This mechanism has been observed in many biological processes, such as photosystem II,[30] green fluorescent proteins,[31] bacteriorhodopsin, [32] and others.[29]

The first observation of ESPT in the hCRBPII binding pocket was made by Prof. Borhan's lab. Dr. Santos showed that in certain mutants, when the protein/chromophore excited at the Schiff base (SB) wavelength, two bands appear. One of the bands corresponds to the SB emission and the other is more similar to the PSB emission. Further studies demonstrated that the placement of acidic residues in certain positions within the vicinity of the chromophore causes this dual emission, which leads to a large Stokes shift between the excitation and emission wavelength or ESPT. Additional exploration of this topic revealed some criteria for an ideal ESPT system: the chromophore should generate the imine, not iminium, with an active Lys 108, and the pK_a of the chromophore should be kept low at all times to maintain the SB. To elucidate the mechanism of ESPT, after screening many variants and conditions for crystallization, we successfully obtained high resolution structures of two hCRBPII variants, bound to two thiophene-based chromophores that showed a high yield of ESPT (Figure IV-17), one in the monomeric fold and the other in the trimeric fold of hCRBPII.

As shown in Figure IV-18, the crystal structure of **ThioFluor** bound to Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W (**L2**) was determined at 1.6 Å resolution. The crystal structure depicted four molecules in the asymmetric unit. Interestingly, the electron density for chain A was correlated with the *trans*-imine*,* chain B was refined as a mixture of the *cis* and *trans*-imine with about 0.5 occupancy for each, and chain C corresponded with the *cis*-imine. The overlay of these three structures reveals similar

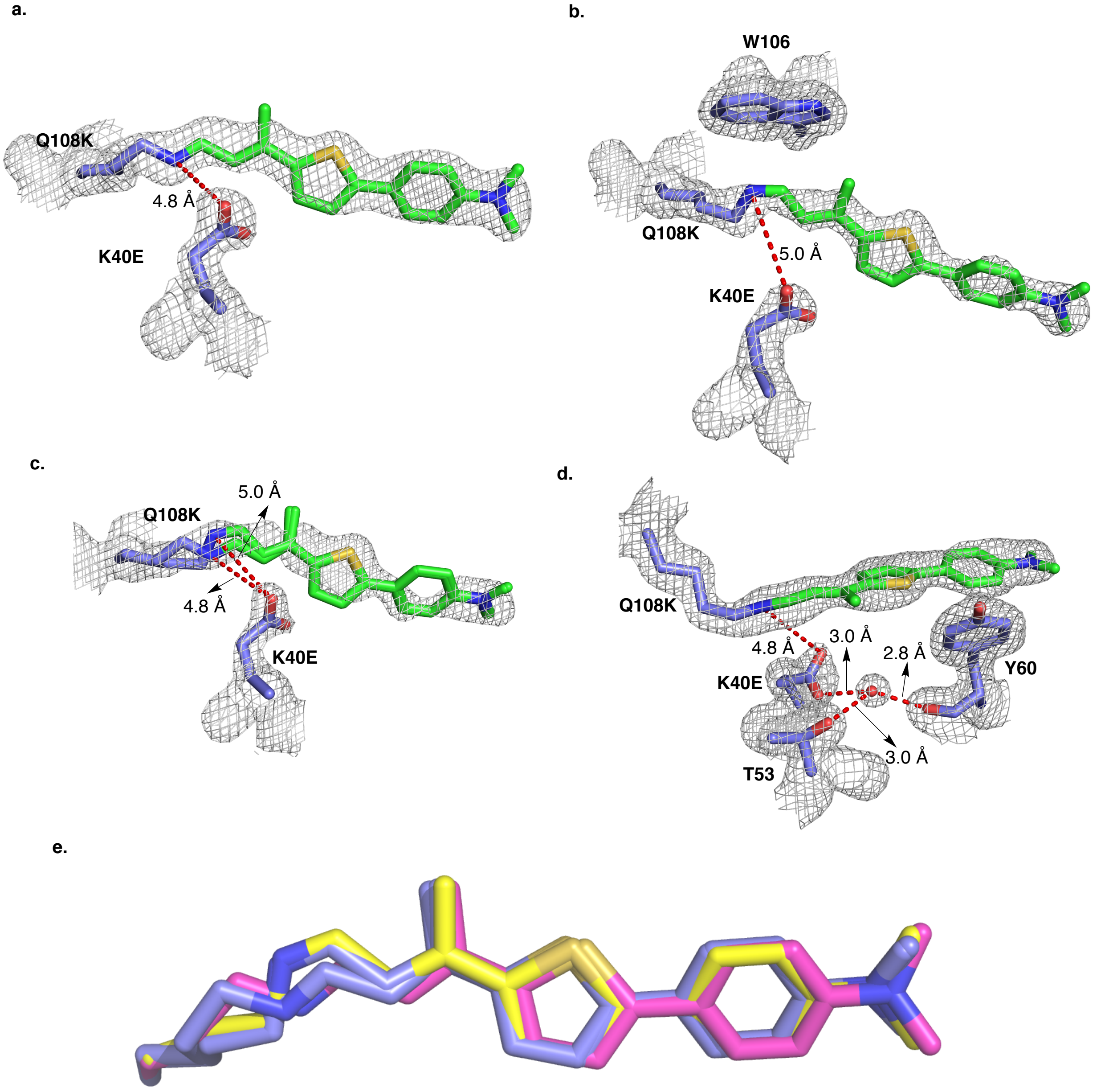


Figure IV-18. Crystal structure of Q108K:K40E:T53A:R58W:Q38F:Q4F:Y19W **(L2/ThioFluor** complex shows four molecules in the asymmetric units, where the electron density of the chromophore in three chains indicates ligand binding. a. **ThioFluor** bound to the protein in chain A depicts the *trans* imine. b. **ThioFluor** bound to the protein in chain C depicts an electron density more correlated with a *cis* iminine. c. **ThioFluor** bound to the protein in chain B depicts an electron density more correlated with a mixture of *cis* and *trans* iminine. d. K40E conformation that is not interacting with the nitrogen of imine. It is expected that this conformation would cause a suppressed p*Ka* for imine, leading to an SB population. K40E is held in this conformation through a water-molecule interaction with the T53 and Y60 main chain. The electron density is contoured to 1σ for all the structures.

trajectories of the chromophore for each chain (Figure IV-18). Remarkably, in all of these

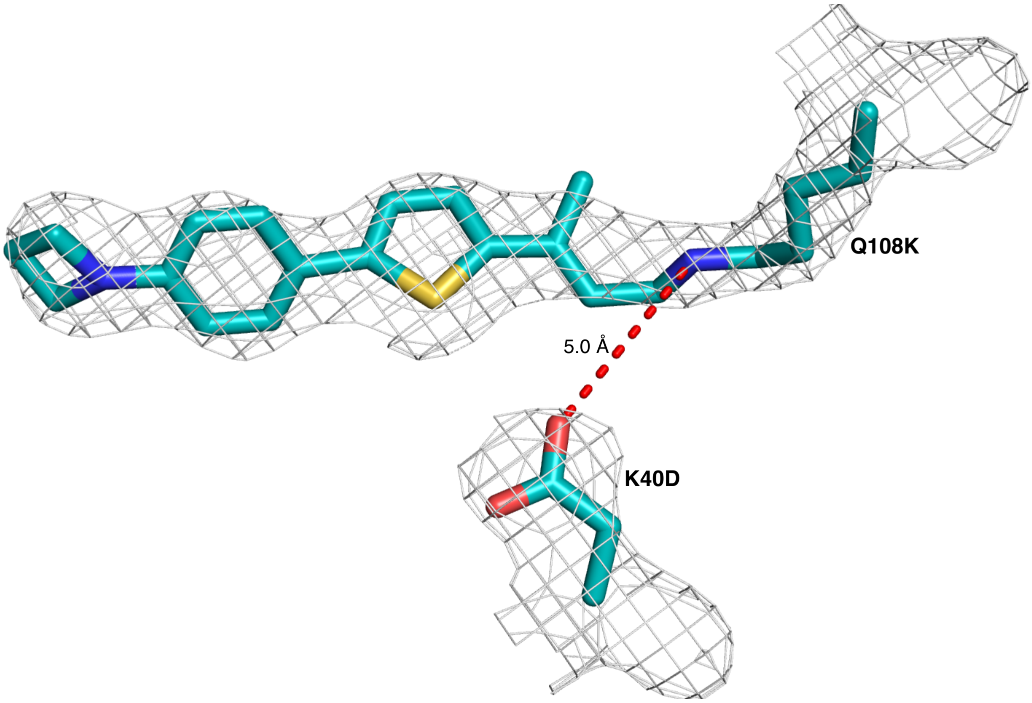


Figure IV-19. Crystal structure of **ThioFluor3** bound with a trimeric form of Q108K:K40D:T53A:R58L:Q4F:Q38F reveals a K40 conformation that is far away from the nitrogen of iminium.

chains, K40E does not interact directly with the nitrogen of iminium, which can explain the low pK_a of the imine. As mentioned earlier, the ground-state imine should be kept at a low pK_a for an efficient ESPT, a requirement that is clearly met in the atomic-resolution crystal structure of the **L2**/**ThioFluor** complex. How is K40E, an acidic residue, kept away from the nitrogen of imine? The crystal structure signifies that an ordered water molecule, mediated between Thr51, the oxygen atom of the Tyr60 main chain, and the nitrogen atom of the Ala53 main chain, holds K40E in that specific conformation. The crystal structure of Q108K:K40D:T53A:R58L:Q4F:Q38F trimer-bound **ThioFluor3** elucidates a similar scenario (see chapter III for further information) where K40D is located in a remote position from the nitrogen of imine.

Additionally, the crystal structure of another chromophore-bound complex showing the ESPT was obtained (the chromophore was **FR1** and was synthesized by Dr. Wei Sheng). Consistent with **ThioFluor** bound proteins demonstrarting ESPT, **FR1** bound Q108K:K40H:T53A:R58L:Q4F:Q38F (**L3**) variant of hCRBPII showed an ESPT but with

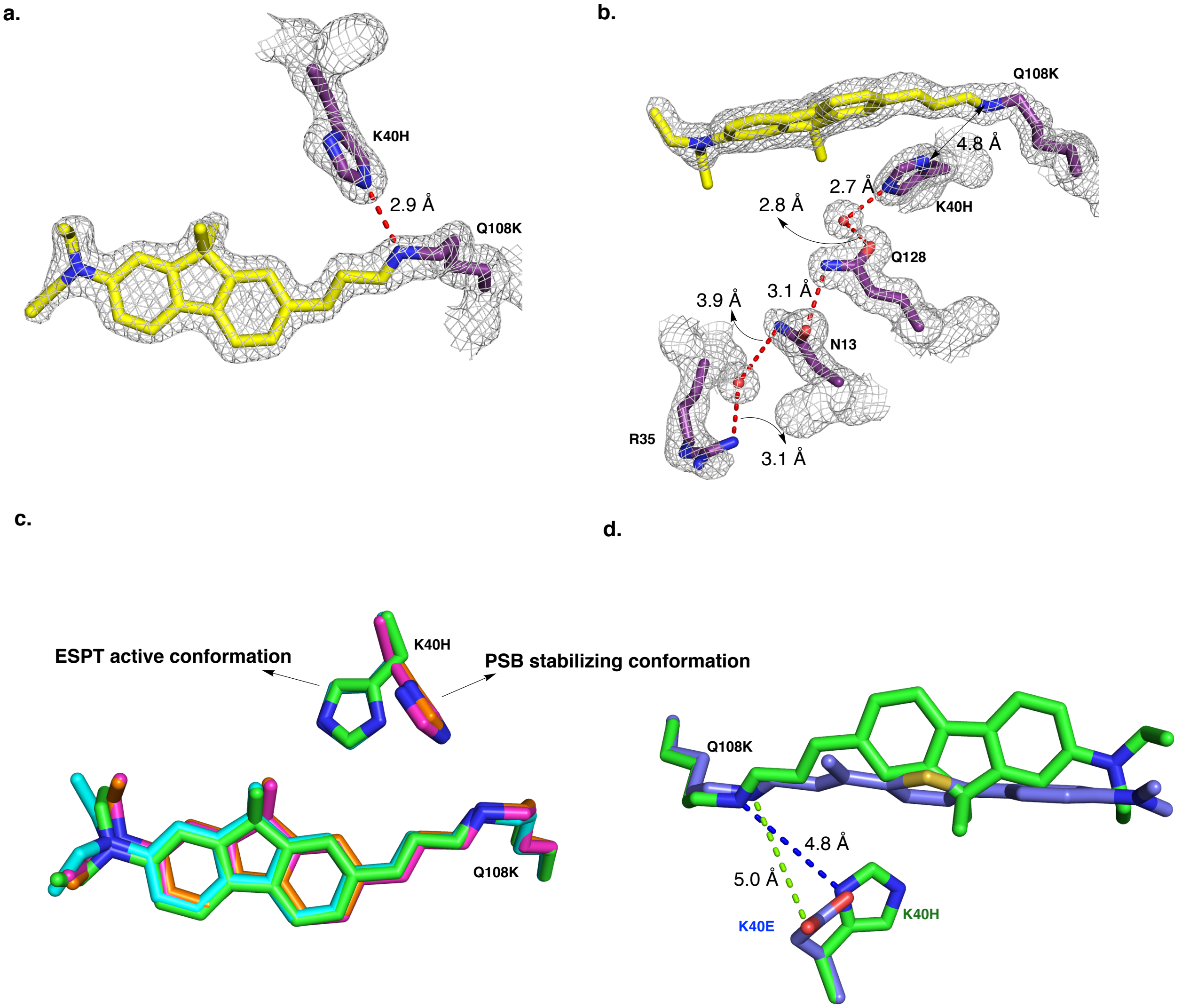


Figure IV-20. Crystal structure of Q108K:K40H:T53A:R58L:Q4F:Q38F (**L3**) bound with **FR1** shows four molecules in the asymmetric unit that are fully occupied by **FR1**. a. Chains A and B show the same conformation of K40H, in which it directly interacts with the nitrogen of iminium. b. Chains C and D show the same conformation of K40H, in which it does not interact with the nitrogen of iminium (4.8 Å distance). Instead, K40H gets locked into this conformation by hydrogen bond networks containing the residues highlighted in the Figure. c. Left: Overlay of chain A and chain C, indicating that the only difference between the two chains is a different conformation of K40H. The K40H conformation that is closer to the nitrogen of imine would be expected to stabilize the iminium, and the other conformation would be an active ESPT ground-state conformation. Right: Overlay of **L3**/FR1 (green) and **L2/ThioFluor** (purple) crystal structures further confirms the ESPT active conformer of K40H.

a low quantum yield in comparison with **L2/ThioFluor** (Figure IV-21). Additionally, in contrast to the **L2/ThioFluor** complex that dominantly forms SB in the ground state, **L3/FR1** shows some population of PSB as well as SB. The crystal structure of the **L3/FR1** complex at 1.6 Å resolution depicts the presence of four molecules in the asymmetric unit with the same trajectory for the chromophore (Figure IV-19) in all four. However, as shown in Figure IV-19, the conformation of K40H varies in the different chains. In two chains (A and B), K40H maintains the interact ion with the nitrogen of iminium, and in the other two chains (C and D), K40H rotates away from the nitrogen of the iminium, generating a water-mediated hydrogen bond network with Q128, N13, and R35. It is expected that the K40H with the conformation proximal to the nitrogen of the iminium would stabilize the positive charge on the nitrogen of iminium and lead to PSB formation (Figure IV-20c). This observation is also is in agreement with the UV-vis spectrum of the **L3/FR1** complex that depicts the mixture of SB and PSB in the ground state, Figure IV-21. As mentioned earlier,

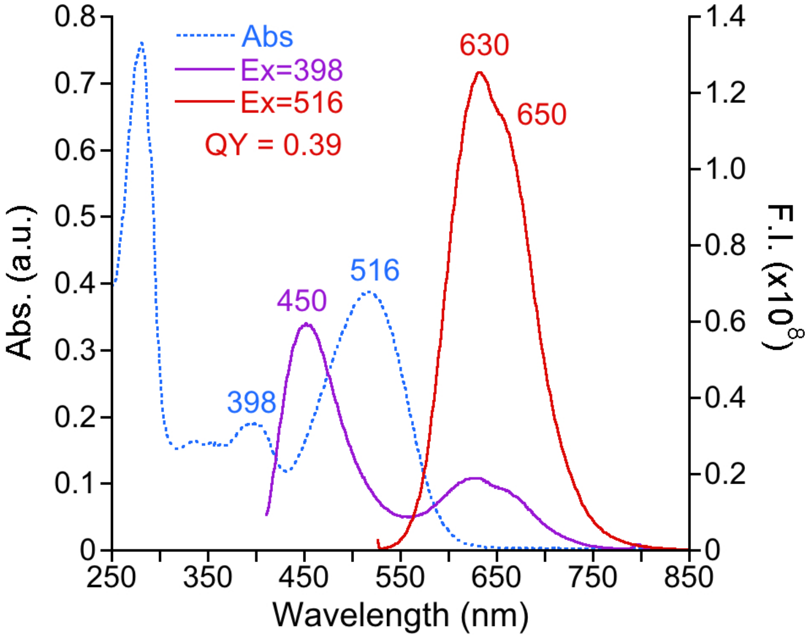


Figure IV-21. UV-vis and emission spectra of **L3/FR1** complex. The data were provided by Dr. Sheng.

the K40H conformation which is proximal to the iminium is not favored for ESPT, because the chromophore is expected to have a high pK_a and already be protonated in the ground state. On the other hand, in chains C and D, where the K40H is in a distal position relative to the nitrogen of iminium, it would be expected that the nitrogen of imine maintains a suppressed pK_a. The latter conformation of K40H is anticipated to be an ideal conformation that gives ESPT. The overlay of **L2/ThioFluor** and **L3**/**ThioFluor** further proves the aforementioned points, in that K40E is always distant from the nitrogen of imine when **L1**/**ThioFluor** has a higher quantum yield of ESPT, Figure IV-20d.

Based on the high-resolution crystal structures offered above and biophysical data, we propose the key residues (such as K40E, K40H, and K40D) should be kept distant from the nitrogen of iminium in the ground state where they maintain the low pK_a imine. However**,** in response to the excitation of chromophore and altering its p$K_{a,}$ a transient proton translocation to the imine ensues, eventuating in a dual emission and what is called ESPT.

IV-5. Conclusion and future direction

In this chapter, several crystal structures for fluorophore bound hCRBPII variants were obtained. Through using atomic resolution crystal structures the microenvironment of the chromophore inside the protein binding pocket was visualized and the mechanism of hCRBPII fluorescent protein tags in pH ratiometric sensor and ESPT were discussed. The ability to use high-resolution protein crystallography to elucidate the mechanism of the engineered hCRBPII makes this protein a unique and well-studied engineered protein that can be further employed for a variety of protein-engineering applications.

The high-resolution crystal structures of **ThioFluor** analogs, as well as their low mosaicity, provide great opportunities to employ them for time resolved and neutron diffraction experiments. The ability to capture the intermediates of the ESPT process with time-resolved x-ray crystallography would provide valuable knowledge for the future design and improvement of fluorescent protein tags. Furthermore, neutron diffraction experiments would give us a clear picture of hydrogen atoms distribution in the protein binding pocket in different pHs.

IV-6. Experimental procedures

IV-6-1.ThioFluor analog synthesis

Please see Dr. Elizabeth M. Santos and Dr. Hadi Gholami's PhD theses.

IV-6-2. FR1 synthesis

Please see Dr. Wei Sheng's PhD thesis.

IV-6-3. Protein expression and purification

The same protocol mentioned in Chapter II was used here with this modification that the expression was performed at OD~0.9-1.0 at 23°C for ~20 hours.

IV-6-4. Crystallization conditions

Each mutant protein was concentrated between 6 and 8 mg/mL in a buffer containing 10 mM Tris, 150 mM NaCl, pH = 8, and was incubated with four equivalent chromophores. Crystals were grown via the hanging drop vapor diffusion method using 1 μL of protein solution and 1 μL of crystallization solution in the drop and 1 mL of crystallization solution in the reservoir.

The best crystals grew using a crystallization solution of 25-30% PEG 4000, 0.1 M sodium acetate, pH=4.0-4.8, 0.1 M ammonium acetate. The crystals were soaked in a cryoprotectant solution (30% PEG 4000, 0.1 M sodium acetate pH = 4.5, 0.1 M ammonium acetate with 20% glycerol), flash frozen in loops in liquid nitrogen, and stored in a liquid nitrogen Dewar prior to data collection.

Table IV-3. X-ray crystallographic data and refinement statistics for hCRBPII **L1**:L117E bound with **ThioFluor2** , **ThioFluor3**, and **ThioFluor4** analogs.

	ThioFluor2-L1:L117E	**ThioFluor3-L1:L117E**	**ThioFluor4-L1:L117E**
Space group	*P* 21	*C*21	*C*21
a(Å)	29.26	29.41	29.54
b(Å)	66.72	66.65	66.45
c (Å)	63.95	64.15	63.86
α(°)	90.00	90.00	90.00
β(°)	90.78	90.81	90.38
δ(°)	90.00	90.00	90.00
Molecules per asymmetric unit	2	2	2
Total reflection	1799743	685606	1468011
Unique reflection	37540 (3548)	15679 (1524)	30407 (2898)
Completeness (%)	96.34 (91.23)	96.92 (93.30)	96.91 (93.15)
Average I/σ	37.3 (2.1)	28.97(4.3)	17.2 (2.2)
R_{merge}	0.069 (0.950)	0.047(0.250)	0.080 (0.534)
Resolution (Å) (last shell)	26.74-1.51 (1.56-1.51)	32.07-1.607 (1.65-1.61)	24.81- 1.28 (1.33- 1.28)
R_{work}/ R_{free} (%)	19.69/22.88	19.43/22.03	16.75/18.32
RMS bond length (Å)	0.008	0.006	0.008
RMS bond angle (°)	0.93	0.89	1.06
Average B factor	24.17	24.53	13.82
Number of water molecules	265	39	1.06
PDB ID	TBD	TBD	TBD

Table IV-4. X-ray crystallographic data and refinement statistics for hCRBPII **L1**:L117E variants bound with **ThioFluor6** , **ThioFluor7**, and **ThioFluor8** analogs.

	ThioFluor6- L1:L117E	**ThioFluor7- L1:L117E**	**ThioFluor8- L1:L117E**
Space group	$P2_1$	*C*21	*C*21
a(Å)	29.30	29.35	29.30
b(Å)	66.84	67.10	66.47
c (Å)	64.07	64.02	64.28
α(°)	90.00	90.00	90.00
β(°)	90.75	90.37	90.37
δ(°)	90.00	90.00	90.00
Molecules per asymmetric unit	2	1	1
Total reflection	1589460	633336	372414
Unique reflection	30514 (2883)	25044 (2501)	14611 (1373)
Completeness (%)	96.47 (92.67)	96.04 (96.19)	97.20 (92.45)
Average I/σ	37.8 (2.3)	31.4 (3.16)	28.6 (2.5)
R_{merge}	0.058 (0.679)	0.042 (0.365)	0.069 (0.990)
Resolution (Å) (last shell)	33.42-1.62 (1.68-1.62)	29.72 - 1.37 (1.42 -1.37)	29.52 -1.64 (1.70- 1.64)
R_{work}/ R_{free} (%)	22.05/24.64	18.87/20.77	17.26/21.89
Bond length (Å)	0.008	0.007	0.008
Bond angle (°)	0.96	0.85	0.91
Average B factor	25.41	18.20	21.11
Number of water molecules	121	134	66
PDB ID	TBD	TBD	TBD

[a] Values in parentheses refer to the last resolution shell.
(**L1**:L117E: Q108K:K40L:T53S:R58W:Y19W:L117E)

Table IV-5. X-ray crystallographic data and refinement statistics for hCRBPII variants bound with **ThioFluor** analogs.

	ThioFluor9-L1:L117E	**ThioFluor10-L1:L117E**	**ThioFluor-L1:L117E**
Space group	*C*21	*P*1	*C*21
a(Å)	29.396	29.40	29.618
b(Å)	66.751	66.74	66.614
c (Å)	64.067	64.22	64.119
α(°)	90.00	89.73	90.00
β(°)	90.30	90.28	90.95
δ(°)	90.00	90.10	90.00
Molecules per asymmetric unit	2	4	1
Total reflection	287594	984824	661219
Unique reflection	19317 (1927)	44915 (4209)	18119 (1752)
Completeness (%)	93.38 (92.91)	96.48 (89.43)	98.07 (95.10)
Average I/σ	41.29 (4.8)	24.07 (2.6)	2.16 (34.4)
R_{merge}	0.041(0.482)	0.065 (0.614)	0.05 (0.478)
Resolution (Å) (last shell)	33.38 - 1.48	33.37 - 1.79	33.31 - 1.54 (1.59 - 1.54)
R_{work}/ R_{free} (%)	17.52/19.35	21.21/24.78	17.72/21.21
Bond length (Å)	0.008	0.008	0.006
Bond angle (°)	0.94	0.83	0.85
Average B factor	17.29	26.03	23.78
Number of water molecules	45	330	84
PDB ID	TBD	TBD	TBD

[a] Values in parentheses refer to the last resolution shell.
(**L1**:L117E: Q108K:K40L:T53S:R58W:Y19W:L117E)

Table IV-6. X-ray crystallographic data and refinement statistics for hCRBPII variants **L1**:L117E, **L1**:L117D, and **L1**:L117D:Q4F bound with **ThioFluor** analogs.

	ThioFluor L1:L117E:A33W	**ThioFluor- L1:L117D**	**ThioFluor- L1:L117D:Q4F**
Space group	*C*21	*C*21	*C*21
a(Å)	28.918	29.233	29.196
b(Å)	66.617	66.641	67.337
c (Å)	63.262	63.939	64.025
α(°)	90	90	90
β(°)	90.981	90.474	89.311
δ(°)	90	90	90
Molecules per asymmetric unit	1	1	1
Total reflection	851159	1199572	1096273
Unique reflection	32216 (3049)	63103 (5744)	16377 (1601)
Completeness (%)	99.10 (94.18)	93.09 (85.35)	99.90 (99.38)
Average I/σ	35.5 (2.1)	44.3(1.5)	3.04 (36.6)
R_{merge}	0.059 (0.44)	0.059 (0.727)	0.14 (1.9)
Resolution (Å) (last shell)	33.31- 1.258 (1.303- 1.258)	29.55-0.991 (1.027- 0.991)	29.8 - 1.599 (1.657- 1.599)
R_{work}/ R_{free} (%)	16.95/18.59	19.17/21.46	18.57/21.37
Bond length (Å)	0.008	0.005	0.005
Bond angle (°)	0.95	0.88	0.84
Average B factor	16.71	16.73	25.98
Number of water molecules	179	191	74
PDB ID	TBD	TBD	TBD

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

Table IV-7. X-ray crystallographic data and refinement statistics for hCRBPII **L2**, **L3**, and Q108K:K40D:T53A:R58L:Q38F:Q4F variants bound with **ThioFluor** and **FR1**, and **ThioFluor3** analogs.

	ThioFluor-L2	**FR1-L3**	**ThioFluor3** **Q108K:K40D:T53A:R58L:Q38F:Q4F-trimer**
Space group	*P*1	*P*1	$P22_12_1$
a(Å)	36.766	36.687	36.932
b(Å)	54.99	54.515	102.41
c (Å)	69.805	63.931	108.478
α(°)	109.338	89.996	90
β(°)	97.797	86.63	90
δ(°)	102.603	77.686	90
Molecules per asymmetric unit	4	4	3
Total reflection	970285	950433	1077503
Unique reflection	54659 (5195)	55438 (5085)	14902 (1453)
Completeness (%)	96.19 (91.67)	85.90 (79.16)	99.89 (99.83)
Average I/σ	24.6(1.4)	22.50 (2.12)	18.88 (1.42)
R_{merge}	0.044 (0.51)	0.046 (0.53)	0.097 (1.251)
Resolution (Å) (last shell)	34.95 -1.671 (1.731 -1.671)	32.04 - 1.594 (1.651- 1.594)	34.74-2.5 (2.589-2.5)
R_{work}/ R_{free} (%)	18.38/22.94	19.67/23.27	20.76/27.97
Bond length (Å)	0.007	0.008	0.008
Bond angle (°)	0.86	1.18	1.24
Average B factor	29.66	24.08	50.39
Number of water molecules	405	235	77
PDB ID	TBD	TBD	TBD

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

REFERENCES

REFERENCES

1. Shintani, T.; Reggiori, F., Fluorescence microscopy-based assays for monitoring yeast Atg protein trafficking. *Methods Enzymology* **2008,** *451*, 43-56.

2. Botelho, H. M.; Uliyakina, I.; Awatade, N. T.; Proença, M. C.; Tischer, C.; Sirianant, L.; Kunzelmann, K.; Pepperkok, R.; Amaral, M. D., Protein traffic disorders: an effective high-throughput fluorescence microscopy pipeline for drug discovery. *Scientific Reports* **2015,** *5*, 9038.

3. Gaietta, G.; Deerinck, T. J.; Adams, S. R.; Bouwer, J.; Tour, O.; Laird, D. W.; Sosinsky, G. E.; Tsien, R. Y.; Ellisman, M. H., Multicolor and electron microscopic imaging of connexin trafficking. *Science* **2002,** *296* (5567), 503-507.

4. Massoud, T. F.; Paulmurugan, R.; De, A.; Ray, P.; Gambhir, S. S., Reporter gene imaging of protein-protein interactions in living subjects. *Current Opinion in Biotechnology* **2007,** *18* (1), 31-7.

5. Xing, S.; Wallmeroth, N.; Berendzen, K. W.; Grefen, C., Techniques for the Analysis of Protein-Protein Interactions in Vivo. *Plant physiology* **2016,** *171* (2), 727-58.

6. Bigelow, C. E.; Vishwasrao, H. D.; Frelinger, J. G.; Foster, T. H., Imaging enzyme activity with polarization-sensitive confocal fluorescence microscopy. *Journal of Microscopy* **2004,** *215* (Pt 1), 24-33.

7. Welder, F.; McCorquodale, E. M.; Colyer, C. L., Proteinase assay by capillary electrophoresis employing fluorescence-quenched protein-dye conjugates. *Electrophoresis* **2002,** *23* (11), 1585-90.

8. Turunen, P.; Rowan, A. E.; Blank, K., Single-enzyme kinetics with fluorogenic substrates: lessons learnt and future directions. *FEBS Letters* **2014,** *588* (19), 3553-3563.

9. Svoboda, K. K. H.; Reenstra, W. R., Approaches to studying cellular signaling: a primer for morphologists. *The Anatomical Record* **2002,** *269* (2), 123-139.

10. Süel, G., Chapter thirteen - Use of Fluorescence Microscopy to Analyze Genetic Circuit Dynamics. In *Methods in Enzymology*, Voigt, C., Ed. Academic Press: 2011; Vol. 497, pp 275-293.

11. Raymo, F. M., Photoactivatable Synthetic Dyes for Fluorescence Imaging at the Nanoscale. *The Journal of Physical Chemistry Letters* **2012,** *3* (17), 2379-85.

12. Combs, C. A., Fluorescence microscopy: a concise guide to current imaging methods. *Current protocols in neuroscience* **2010,** *Chapter* 2, Unit2.1.

13. Tsien, R. Y.; Waggoner, A., Fluorophores for Confocal Microscopy. In *Handbook of Biological Confocal Microscopy*, Pawley, J. B., Ed. Springer US: Boston, MA, 1995; pp 267-279.

14. Billinton, N.; Knight, A. W., Seeing the Wood through the Trees: A Review of Techniques for Distinguishing Green Fluorescent Protein from Endogenous Autofluorescence. *Analytical biochemistry* **2001,** *291* (2), 175-197.

15. Patterson, G. H.; Knobel, S. M.; Sharif, W. D.; Kain, S. R.; Piston, D. W., Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophysical Journal* **1997,** *73* (5), 2782-2790.

16. Ma, Y.; Sun, Q.; Zhang, H.; Peng, L.; Yu, J.-G.; Smith, S. C., The mechanism of cyclization in chromophore maturation of green fluorescent protein: a theoretical study. *The Journal of Physical Chemistry B* **2010,** *114* (29), 9698-9705.

17. Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D., Understanding GFP chromophore biosynthesis: controlling backbone cyclization and modifying post-translational chemistry. *Biochemistry* **2005,** *44* (6), 1960-1970.

18. Chudakov, D. M.; Lukyanov, S.; Lukyanov, K. A., Fluorescent proteins as a toolkit for in vivo imaging. *Trends Biotechnol* **2005,** *23* (12), 605-13.

19. Jensen, E. C., Use of Fluorescent Probes: Their effect on cell biology and limitations. *The anatomical record* **2012,** *295* (12), 2031-2036.

20. Sheng, W.; Nick, S. T.; Santos, E. M.; Ding, X.; Zhang, J.; Vasileiou, C.; Geiger, J. H.; Borhan, B., A near-infrared photoswitchable protein–fluorophore tag for no-wash live cell imaging. *Angewandte Chemie International Edition* **2018,** *57* (49), 16083-16087.

21. Yapici, I.; Lee, K. S.; Berbasova, T.; Nosrati, M.; Jia, X.; Vasileiou, C.; Wang, W.; Santos, E. M.; Geiger, J. H.; Borhan, B., "Turn-on" protein fluorescence: in situ formation of cyanine dyes. *Journal of the American Chemical Society* **2015,** *137* (3), 1073-80.

22. Berbasova, T.; Tahmasebi Nick, S.; Nosrati, M.; Nossoni, Z.; Santos, E. M.; Vasileiou, C.; Geiger, J. H.; Borhan, B., A genetically encoded ratiometric pH probe: wavelength regulation-inspired design of pH indicators. *Chembiochem : a European Journal of Chemical Biology* **2018,** *19* (12), 1288-1295.

23. Wang, W.; Geiger, J. H.; Borhan, B., The photochemical determinants of color vision: revealing how opsins tune their chromophore's absorption wavelength.

BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology **2014,** *36* (1), 65-74.

24. Berbasova, T.; Nosrati, M.; Vasileiou, C.; Wang, W.; Lee, K. S.; Yapici, I.; Geiger, J. H.; Borhan, B., Rational design of a colorimetric pH sensor from a soluble retinoic acid chaperone. *Journal of the American Chemical Society* **2013,** *135* (43), 16111-9.

25. Wang, W.; Nossoni, Z.; Berbasova, T.; Watson, C. T.; Yapici, I.; Lee, K. S.; Vasileiou, C.; Geiger, J. H.; Borhan, B., Tuning the electronic absorption of protein-embedded all-trans-retinal. *Science* **2012,** *338* (6112), 1340-3.

26. Haberhauer, G.; Gleiter, R.; Burkhart, C., Planarized Intramolecular Charge Transfer: A Concept for Fluorophores with both Large Stokes Shifts and High Fluorescence Quantum Yields. *Chem-Eur J* **2016,** *22* (3), 971-978.

27. Sasaki, S.; Drummen, G. P. C.; Konishi, G., Recent advances in twisted intramolecular charge transfer (TICT) fluorescence and related phenomena in materials chemistry. *J Mater Chem C* **2016,** *4* (14), 2731-2743.

28. Zhang, Z. Y.; Zhang, G. Q.; Wang, J. X.; Sun, S. S.; Zhang, Z. Z., The mechanisms of Large Stokes Shift and Fluorescence Quantum Yields in anilino substituted Rhodamine analogue: TICT and PICT. *Comput Theor Chem* **2016,** *1095*, 44-53.

29. Zhou, P.; Han, K., Unraveling the detailed mechanism of excited-state proton transfer. *Acc Chem Res* **2018,** *51* (7), 1681-1690.

30. Weinberg, D. R.; Gagliardi, C. J.; Hull, J. F.; Murphy, C. F.; Kent, C. A.; Westlake, B. C.; Paul, A.; Ess, D. H.; McCafferty, D. G.; Meyer, T. J., Proton-Coupled Electron Transfer. *Chemical Reviews* **2012,** *112* (7), 4016-4093.

31. Tonge, P. J.; Meech, S. R., Excited state dynamics in the green fluorescent protein. *Journal of Photochemistry and Photobiology A: Chemistry* **2009,** *205* (1), 1-11.

32. Hayashi, S.; Tajkhorshid, E.; Schulten, K., Structural changes during the formation of early intermediates in the bacteriorhodopsin photocycle. *Biophys J* **2002,** *83* (3), 1281-97.