# Protein Design and Engineering Using the Fluorescent Non-canonical Amino Acid L-(7-hydroxycoumarin-4-yl)ethylglycine

by

Patrick R. Gleason

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2020 by the
Graduate Supervisory Committee:

Jeremy H. Mills, Chair
Sidney M. Hecht
Petra Fromme
Nicholas Stephanopoulos

ARIZONA STATE UNIVERSITY

December 2020

ABSTRACT

Proteins are, arguably, the most complicated molecular machines found in nature. From the receptor proteins that decorate the exterior of cell membranes to enzymes that catalyze the slowest of chemical reactions, proteins perform a wide variety of essential biological functions. A reductionist view of proteins as a macromolecular group, however, may hold that they simply interact with other chemical species. Notably, proteins interact with other proteins, other biological macromolecules, small molecules, and ions. This in turn makes proteins uniquely qualified for use technological use as sensors of said chemical species (biosensors).

Several methods have been developed to convert proteins into biosensors. Many of these techniques take advantage of fluorescence spectroscopy because it is a fast, non-invasive, non-destructive and highly sensitive method that also allows for spatiotemporal control. This, however, requires that first a fluorophore be added to a target protein. Several methods for achieving this have been developed from large, genetically encoded autofluorescent protein tags, to labeling with small molecule fluorophores using bioorthogonal chemical handles, to genetically encoded fluorescent non-canonical amino acids (fNCAA). In recent years, the fNCAA, L-(7-hydroxycoumarin-4yl)ethylglycine (7-HCAA) has been used in to develop several types of biosensors.

The dissertation I present here specifically addresses the use of the fNCAA L-(7-hydroxycoumarin-4-yl)ethylglycine (7-HCAA) in protein-based biosensors. I demonstrate 7-HCAA's ability to act as a Förster resonance energy transfer (FRET) acceptor with tryptophan as the FRET donor in a single protein containing multiple tryptophans. I the describe efforts to elucidate—through both spectroscopic and structural characterization—

interactions within a 7-HCAA containing protein that governs 7-HCAA fluorescence. Finally, I present a top-down computational design strategy for incorporating 7-HCAA into proteins that takes advantage of previously described interactions. These reports show the applicability of 7-HCAA and the wider class of fNCAAs as a whole for their use of rationally designed biosensors.

# DEDICATION

To Christina. Your unending belief in my potential is a constant source of the fuel that has driven our progress over the last five years. This work is as much a testament to your sacrifice, patience, and hard work as it is to mine. I hope you are as proud of us as I am.

To Alarick. You are my greatest accomplishment. I am so excited to see what the future holds for you and I am jealous of your potential. I love you immensely.

ACKNOWLEDGMENTS

To espouse a long-standing cliché, science is not done in a vacuum. The work in these pages would not be possible without the collaborations of several people who have taken the time to either mentor me or be mentored by me.

First, I would like to thank Professor Jeremy Mills. It is often difficult to find mentor/mentee relationships that work well for both individuals. Here, Jeremy was the perfect mentor for me. I need a considerable amount of independence to explore my own ideas and to perform my own techniques. Jeremy provided that independence, allowing my failures and successes to be my own.

I would like to thank two of my mentees also. First Bethany (Beth) Kolbaba Kartchner. A good mentor grows as their mentee grows. Beth raised the bar for my patience, tolerance, and acceptance. Because of her, I learned to approach situations with new perspectives that a younger version of me would not have. I applaud her willingness to strive for more and her appetite for learning. I have no doubt that her future in science is bright. Second, I would like to thank Dominic (Dom) Grisingher. I cannot thank Dom enough. He is the only undergraduate I have ever worked with who understood that when I say, “You are either innately good at this or you have to work harder than everyone else.”, I don’t think anyone is “innately good” at anything. In a world where it seems like everyone wants to spend their time telling you how good they are at this or that, Dom understood that I don’t care. He understood that I only care about the next thing we are going to accomplish. I would dare to argue that I have spent more time, working late at nights and on the weekends, with Dom over the last 5 years than with my wife or son. The only thing I have left to say to him is, don’t stop committing 100% to the things you choose to do.

I would also like to thank Dr. Nathan Henderson and Chad Simmons for their mentorship. I am consistently humbled by their knowledge of structural biology. I strive (and falter) to their level. The discussion I have had with each of them have only made me a better scientist. To that end, I will also thank Patrick (O'Malley) Kelly. As a friend and colleague, he helped me by listening to ideas, plans, and frustrations. Lunches will likely not be the same going forward. Additionally, I would like to thank Dr. Noureddine Fahmi. His mentorship in the techniques associated with organic chemistry is priceless. I have learned more over the last two years working with him on various amino acid syntheses than I could have ever hoped to have. I sincerely appreciate his mentorship and friendship. As both of us embark on new adventures in our careers, I wish him nothing but the best of luck.

I also need to thank my parents Frederick (Rick) and Cindy Gleason, my sister Stephanie, and my in-laws Carla and Mark Mayer. Their added support through the stresses of graduate school only worked to lessen my burdens. Finally, I thank my wife, Christina, and son, Alarick. This wasn't worth doing if it weren't for either of you. I believe there is nothing I can say or do that would appropriately express my gratitude. I love you both immensely.

TABLE OF CONTENTS

Page

APPENDIX Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF ABBREVATIONS

Abbreviation

2-(N-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)amino)-2-deoxyglucose ..... (2-NBDG)

4-biphenyl-L-phenylalanine ..... (4PBA)

4 parameter logistic function ..... (4PL)

7-hydroxycoumarin ..... (7-HC)

L-(7-hydroxycoumarin-4-yl)ethylglycine ..... (7-HCAA, X)

Autofluorescent proteins ..... (aFP)

Alanine ..... (Ala, A)

β-mercaptoethanol ..... (BME)

Calmodulin ..... (CaM)

Circular Dichroism ..... (CD)

Column Volume ..... (CV)

Dalton ..... (Da)

Dihydrofolate reductase ..... (DHFR)

Enhanced cyan fluorescent protein ..... (eCFP)

Excited State Proton Transfer ..... (ESPT)

Fragmented Anti-body ..... (FAB)

Fluorescein arsenical hairpin binder-ethanedithiol ..... (FlAsh)

Fluorescent non-canonical amino acid ..... (fNCAA)

Förster Resonance Energy Transfer ..... (FRET)

Green fluorescent protein ..... (GFP)

Glutamine ..... (Gln, Q)

Glutamic acid ..... (Glu, E)

G-protein coupled receptor ..... (GPCR)

Ion Exchange Chromatography ..... (IEC)

Isoleucine ..... (Ile, I)

isopropyl-β-D-thiogalactoside ..... (IPTG)

Isothermal calorimetry ..... (ITC)

KiloDalton ..... (kDA)

Leucine ..... (Leu, L)

Lysine ..... (Lys, K)

Molar ..... (M)

Methionine ..... (Met, M)

Millimolar ..... (mM)

Non-canonical amino acid ..... (NCAA)

Nickel-nitrilotriacetic acid ..... (Ni-NTA)

Nanomolar ..... (nM)

Nanometer ..... (nm)

Optical Density ..... (OD)

Periplasmic binding protein ..... (PBP)

Polymerase chain reaction ..... (PCR)

Protein Databank ..... (PDB)

Phenylalanine ..... (Phe, F)

Picomolar ..... (pM)

Revolutions per minute ..... (RPM)

Streptavidin ..... (SAV)

Serine ..... (Ser, S)

Tris Buffered Saline ..... (TBS)

Thermal melting/Melting Temperature ..... ($T_M$)

Tryptophan ..... (Trp, W)

Tyrosine ..... (Tyr, Y)

Ultra-violet ..... (UV)

Volts ..... (V)

Yellow fluorescent protein ..... (YFP)

Micromolar ..... (μM)

# PREFACE

This is an exciting time to be a protein scientist. Our ability to use computers to 'quickly' sample both sequence and conformational space has unlocked a new age of designer proteins. On top of that, the fact that chemical biologists and organic chemists continue to develop amino acids with side chains not found in nature opens the possibilities of encoding many novel chemical functionalities into proteins. My goal is to combine these two worlds. I want to understand how proteins interact with these new amino acids so that I can design new proteins that contain them and subsequently use these engineered proteins to accomplish heretofore challenging goals in the study of biochemical systems.

The work presented in this dissertation is the culmination of 5 years of research and represents a concerted effort by members of the Mills lab to understand the effects of protein environments on the fluorescent non-canonical amino acid L-(7-hydroxycoumarin-4-yl)ethylglycine (7-HCAA). Specifically, we have been interested in understanding through both spectroscopic and structural analysis, the interactions governing changes in 7-HCAA fluorescence in protein environments. These interactions can then be used as input for the computational design of protein-based fluorescent biosensors using protein design software, such as the Rosetta Macromolecular Design Suite.

Each chapter of this dissertation represents a reproduction of a journal article (either already published or in preparation). While I am the lead author on each of these publications, I would be remiss to say that I completed all of the work myself. For each of these papers, I have had help along the way.

**Chapter 2** relies heavily on the published article: **Gleason, P. R.**; Kelly, P. I.; Grisingher, D. W.; Mills, J. H. An Intrinsic FRET Sensor of Protein–Ligand Interactions.

*Org. Biomol. Chem.* **2020**, *18* (21), 4079–4084. https://doi.org/10.1039/D0OB00793E. However, changes were made for this thesis in order to have a cohesive document from one chapter to the next. The original concept for this work was my own and I performed all aspects of this work from molecular cloning to protein expression and purification to assay development and execution. Dominic came on to the project in order to help with some of the molecular cloning and Patrick Kelly came on to the project to assist with protein expression and assay execution. I made all of the figures and both Jeremy and I wrote the manuscript. I've presented it here, largely, intact but with modifications (i.e. rearrangement of figures from the supplemental information) to better fit with the over thesis document.

**Chapter 3** is a version of a journal article in preparation. All aspects of the work presented in this chapter were performed by me. Additional authors to be listed on the final paper are Bethany Kolboba Kartchner, J. Nathan Henerson, Chad R. Simmons, Noureddine Fahmi, and Jeremy H. Mills. Bethany helped with molecular cloning and protein expression, Nathan and Chad each provided mentorship in and helped with protein crystallography and the associated data collection experiments, Chad also solely collected data for several of the first datasets and Nour synthesized the small molecule, biotinamide, for use in one of our experimental assays. All of the figures for this work have been produced by me and the words presented in this document are my own with edits by Jeremy Mills.

**Chapter 4** is a version of a journal communication in preparation. The work presented in this chapter is wholly my own, however, for the final journal communication there will be the additional authors, Dominic W. Grisinger and Jeremy H. Mills. Dominic

was brought in at the end of the project to help with protein expression and purification under my guidance. As with all of my projects, the figures and words are my own with edits and comments from Jeremy.

**Appendix A** contains the abstracts and author lists from, as well as a brief description of my contributions to three separate publications that I am listed as an author of.

# CHAPTER 1.

# PROTEINS AND THE FLUORESCENT NON-CANONICAL AMINO ACID L-(7-HYDROXYCOUMARIN-4-YL)ETHYLGLYCINE

## 1.1 Introduction to Proteins as Biosensors

Detection of analytes is fundamental to the study of biological systems. Measuring the location, quantity, and specificity of a target protein for different analytes is useful for assessing both the nature and progression of diseases, physiological relevance of a particular chemical species, and for reporting on environmental quality. Not surprisingly, through the use of proteins, living systems have evolved the means for taking stock of both internal and external analytes. Unlike other macromolecules, such as nucleic acids which store information, carbohydrates which provide signaling motifs, or lipids which enclose and protect the cellular system, the function of proteins is to interact with other chemical species. This is primarily facilitated by the fact that the 20 canonical amino acids used to synthesize proteins have variable steric and electrostatic complexity compared with other biological macromolecules. As such, proteins possess the ability to bind analytes with a high degree of specificity and selectivity not readily obtainable with other systems.

The aforementioned properties of proteins have allowed them to function as biosensors in a plethora of diagnostic assays, but often only after extensive engineering. Nonetheless, proteins have found use in a number of applications that themselves rely on analytical methods including colorimetry, mass spectrometry, and electrochemistry. However, especially in recent years, no analytical method is likely more ubiquitous in biochemical analyses than fluorescence-based techniques. This is because fluorescence assays have high levels of sensitivity, don’t require expensive analytical equipment, and

are typically developed using low complexity techniques (e.g. tagging a protein of interest with a fluorescent protein).

Protein fluorescence assays can be grouped into two main methods: 1) the use of autofluorescent protein (aFP) fusion proteins or 2) the use of chemically reactive side chains to site specifically label a protein with a small organic dye. While the use of aFPs, such as green fluorescent protein (GFP) and its analogues, is abundant in the literature, they can suffer from several drawbacks. Specifically, they are large in size (~25 kDa or larger), can only be placed at the termini without substantial protein engineering, and are typically always fluorescent. Moreover, making such large protein-fusions introduces the possibility of the aFP tag interfering with protein function. While the second method of post-translationally modifying reactive residues is generally considered less perturbative, it also suffers from limitations. Three of the most notable limitations are the facts that the target residue needs to be accessible to the dye, that multiple reactive residues can convolute experimental assays, and that labeling with more than one fluorophore requires orthogonal labeling techniques.

In recent years, a set of genetically encoded fluorescent non-canonical amino acids[1–6] (fNCAA) whose emission wavelength fall in the visible range have been developed (Figure 1.1). Because such "fNCAAs" are co-translationally incorporated into proteins, they are not limited to placement at surface exposed residues or the protein termini, a common limitations to other frequently used methods of fluorescently labeling proteins.[7] These fNCAAs are small in size—not much larger than any of the canonical amino acids—and can effectively be placed anywhere in a protein. This alleviates many of the constraints imposed by aFPs and chemical labeling techniques. However, the manner in which these

fNCAAs could potentially be used will depend on the surrounding protein environment. This is not only from the perspective of a given site being able to accommodate the steric bulk of an fNCAA, but also because residues that surround the fluorescent side chains could be used to tune their spectroscopic properties. These common features suggest that fNCAAs could represent excellent starting points for the development of novel, protein-based fluorescent biosensors.

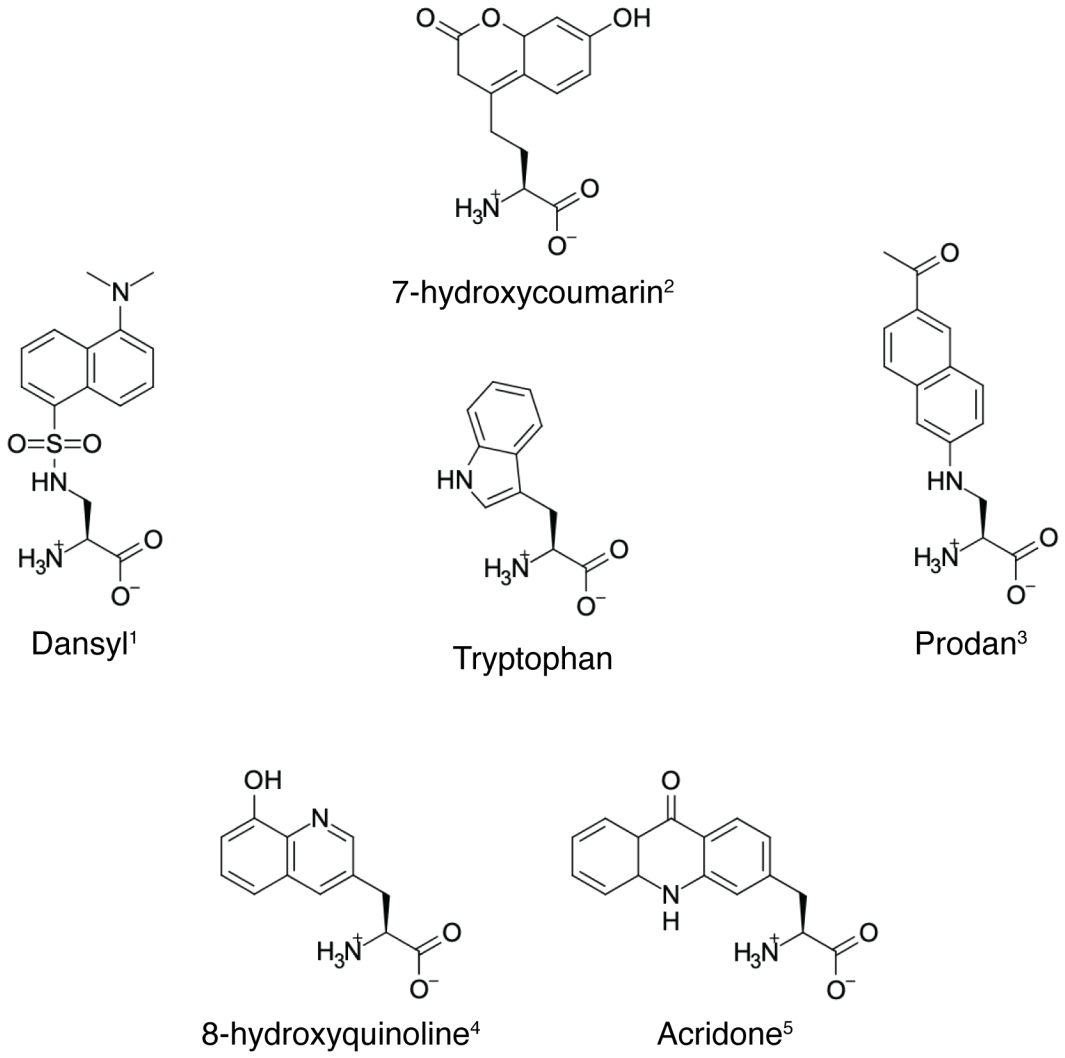
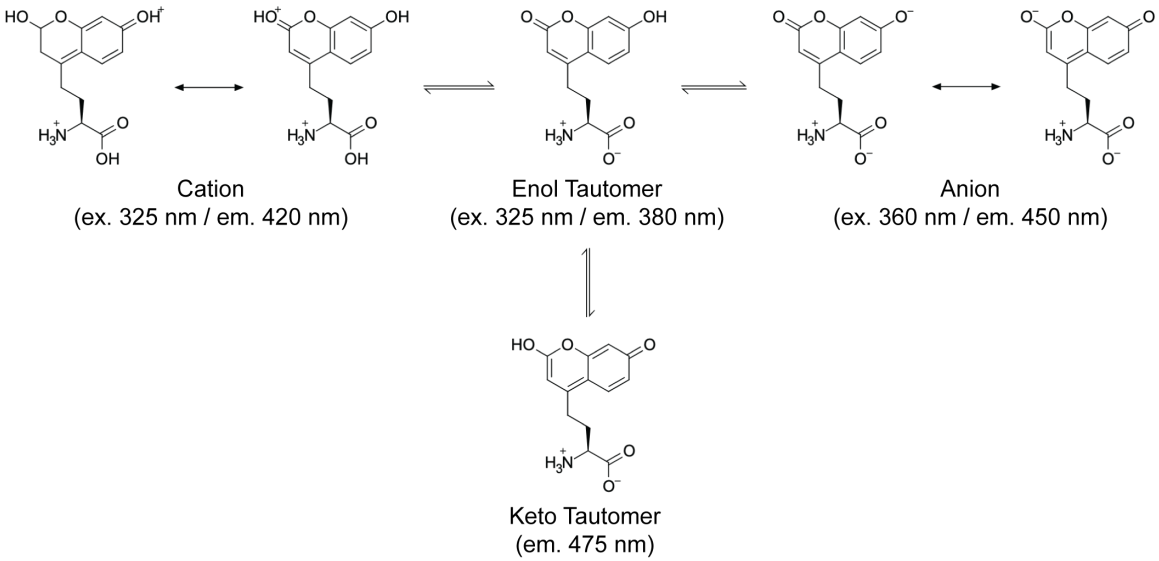



**Figure 1.1.** Chemical structures of the genetically encoded fluorescent non-canonical amino acids.[1–5]

## 1.2 Properties of the Small Molecule 7-Hydroxycoumarin

Among the fNCAAs, L-(7-hydroxycoumarin-4-yl)ethylglycine[2] (7-HCAA, Figure 1.1), has found extensive use in the development of fluorescent protein-based sensors. This is likely because the 7-hydroxycoumarin (7-HC) moiety found in its side chain has several tunable features that can modulate its emission spectrum. Perhaps the most comprehensive set of experimental studies that were aimed at elucidating these properties were performed by Tetsuo Moriya in the 1980s.[8–10] Moriya's seminal work[8] showed that the 7-HC can exist as a number of tautomeric or ionic species in the ground state: an enol tautomeric, an anionic, a cationic, and a keto tautomeric species (Scheme 1.1). Each of these chemical species has a unique absorption/emission spectrum that are favored in different environments.

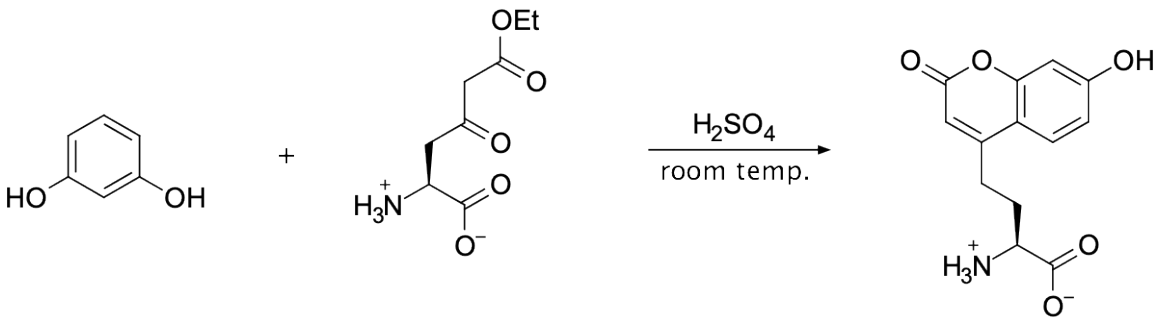



**Scheme 1.1.** Tautomeric and ionized forms of L-(7-hydroxycoumarin-4-yl)ethylglycine and their excitation/emission wavelengths. Values are from small molecule studies performed in aqueous solutions.[8] Unlike the enol tautomer, the keto tautomer is not a photoacid but a photobase and its emission spectrum is dependent on the protonation of the excited state of the anionic species.

In addition to the various ground state 7-HC species, Moriya definitively showed that 7-HC is also a photoacid where the phenolic $pK_a$ drops from ~7.8 in the ground state to ~0.4 in the excited state.[8] When excited in aqueous solvents, the phenolic proton of the neutral enol tautomer is rapidly shuttled to a water molecule in a process termed excited state proton transfer (ESPT). This immediately generates the anionic, phenolate species in the excited state. The emission maximum of this phenolate species occurs at 450 nm, while if ESPT is blocked and emission comes instead from the neutral enol species, the emission maximum shifts to 380 nm.[10–12]

### 1.3 Literature Review of L-(7-hydroxycoumarin-4-yl)ethylglycine

The synthesis of 7-HCAA via the von Pechmann reaction (Scheme 1.2) and its first use as a fluorescent indicator in synthesized peptides by Garbay and co-workers showed the viability of coumarin based fNCAAs in biological contexts.[13] Since then it has been incorporated into peptides and full length proteins through both *in vitro* and *in vivo* techniques.[2,14] This has led to several studies that use 7-HCAA to report on ligand binding,[15–17] protein-protein interactions,[18–20] enzyme-substrate binding,[21,22] protein folding,[23] and changes in tyrosine phosphorylation state.[24] While the excitation/emission maxima (325/450 nm) of 7-HCAA may not be ideal for *in vivo* microscopy based studies (although it has been used in such[13,25,26]), it has proved very effective for *in vitro* studies that utilize methods such as Förster Resonance Energy Transfer (FRET),[14,16,17,23,27–30] shifts in excitation maxima,[24,31] and quenching of the 7-HCAA by the protein environment.[19–22]




**Scheme 1.2.** Synthesis of L-(7-hydroxycoumarin-4-yl)ethylglycine.[13]

Of the previously listed techniques, FRET has been the most widely used. For a detailed explanation of FRET and its benefits see the introduction of chapter 2. Skerra and co-workers were the first to use 7-HCAA in a FRET-based assay. In this study, 7-HCAA served asa donor for the cyan autofluorescent protein, eCFP.[29] In this work, the researchers incorporated 7-HCAA on the surface of eCFP and showed that the close proximity of the fluorophores was sufficient to allow resonance energy transfer to occur. Unfortunately, the simplicity of this system is not directly useful as a biosensor; although it does show the versatility of 7-HCAA in a biological setting. Perhaps a more useful display of 7-HCAA with an aFP was that of Lee and co-workers.[15] They placed the 7-HCAA in a periplasmic binding protein (PBP) that was then fused with yellow fluorescent protein (YFP). Upon ligand-binding the PBP undergoes a conformational change that brings the 7-HCAA closer to YFP increasing the FRET efficiency. This was then used to perform high-throughput screening of engineered proteins in order to find PBP mutants with higher affinity for the leucine.

Interestingly enough, the use of 7-HCAA as a FRET donor is not restricted to aFPs only. Tian and co-workers have produced two works in recent years.[16,30] The first incorporates a 7-HCAA into a protein that also contained an Alexa-488 dye, which was

incorporated as a post-translational modification.[16] This allowed ligand-dependent conformational changes to be monitored in an abscisic acid receptor protein. Their second report labeled a G protein coupled receptor (GPCR) with both the 7-HCAA and a fluorescein dye using the well-documented FlAsh protocol.[30] Again they were able to monitor conformational changes in different GPCR subunits. What might be surprising is that neither of these works used direct FRET measurements. Instead, they relied on changes in FRET lifetimes. Although this is sound in both theory and experimental procedure, fluorescence lifetime measurements require expensive equipment and need to be operated and maintained by experienced personnel, thus reducing the general applicability of this technique. While the use of 7-HCAA as a donor for both aFPs and small molecule organic dyes has been shown to be a viable means for measuring conformational changes in proteins, it still suffers from the same limitations described earlier. Namely aFPs produce large fusion proteins that may interfere with protein function and post-translational dyes are limited in their placement.

Another way to take advantage of the benefits of FRET is to use the 7-HCAA as a FRET acceptor for other amino acids that act as the FRET donor. Hecht and co-workers first did this by incorporating the 7-HCAA with a series of fNCAAs containing different isomeric terphenyl moieties through *in vivo* protein synthesis.[14,28] This allowed for the measurement of both activity and inhibition in the enzyme dihydrofolate reductase (DHFR). Additionally, this work showed that the protein's structure tolerated each fNCAA only at specific locations in the protein. This is likely due to the fact that the terphenyl moieties are rather large in size compared with canonical amino acids such as tryptophan and the terphenyl side chain may cause protein folding issues due to steric clashes with

other residues in the protein. These results have bearing on possible incorporation sites for all fNCAAs.

In order to circumvent challenges associated with the incorporation of large fNCAAs, one might use 7-HCAA as a FRET acceptor that is paired with tryptophan found in protein of interest, which would serve as a FRET donor. Several studies have successfully used this method to create FRET sensors.[17,18] Chapter 2 of this thesis describes the use of tryptophan and 7-HCAA as a FRET pair in a single protein for the use as a glucose sensor.[17]

As noted previously, FRET is not the only means by which to convert 7-HCAA into a useful fluorescent sensor. Amaro et al. were able to successfully measure protein hydration by deconvoluting the 7-HCAA fluorescence spectra of two haloalkane dehalogenases.[31] By comparing the intensities of the enol and keto tautomers along with the anion and "complexed" species, they were able to establish a ratiometric analysis that reports on the hydration of specific sites in the proteins. Although this method shows a unique use of the 7-HCAA, this does not likely represent a general approach for measuring protein hydration. Location of the 7-HCAA will surely be a limiting factor (i.e. placement of the 7-HCAA may cause protein folding issues) and the protein environment can have a number of distinct effects on 7-HCAA fluorescence as will be further discussed in Chapter 3 of this thesis.

Another example of utilizing the ratiometric analysis of at least two of the 7-HC species is the work performed by Lacey et al.[24] In this study, the 7-HCAA was site-specifically incorporated near a tyrosine phosphorylation site in the protein STAT3. In the unphosphorylated protein, the neutral 7-HC species is the predominant species; upon

phosphorylation, however, anionic 7-HC predominates. This allowed the researchers to accurately measure phosphorylation states of the protein by comparing the fluorescence intensities at 450 nm when excited at 325 nm (neutral excitation max) and 360 nm (anion excitation max).

Several other studies have been performed in which the mechanism for changes in 7-HCAA fluorescence aren't as clear. For example, Spies and co-workers placed the 7-HCAA into a glutamate racemase in order to measure ligand and inhibitor binding.[22] Their primary explanation for the changes in the fluorescence intensity centers on changes in the solvation of the 7-HC moiety. This hypothesis is supported with molecular dynamics simulations. However, because emission is exclusively reported at 450 nm, it is unclear whether or not ESPT is simply precluded in this protein, or whether a shift in the tautomeric state of 7-HCAA is responsible for the observed changes in fluorescence. Although additional characterization (e.g. collecting full emission spectra in the bound and unbound states) could have provided insight into this question, these studies were not performed. Similarly, Mendes et al. report the use of 7-HCAA in probing allosteric binding of nucleotides by an aspartate transcarbamoylase; however, the mechanism through which changes in 7-HCAA fluorescence occur was not explored.[21] It should be noted, however, that understanding the mechanism of 7-HCAA fluorescence changes was likely beyond the scope of both of these studies; rather the goal was likely to simply engineer a new sensor.

In contrast to the aforementioned studies, Henderson et al.[20] solved several crystal structures of the 7-HCAA containing antibody fragment (Fab) 5c8*.[19] These structures provide the first insights into how interactions between proteins and 7-HCAA can affect the fNCAA's fluorescent properties. Henderson suggests that a rapid back-transfer of the

phenolic proton after an ESPT event provides thermal relaxation of the 7-HCAA excited state is the most likely mechanism for quenching. This idea is bolstered by the fact that many other photoacids have shown similar back-transfer events.[32]

**1.4 Conclusion**

Given the large amount of literature on the use of 7-HCAA in protein-based biosensors, it is only natural to start to take a more directed approach to their development. In this work I will present three ways of using the 7-HCAA as a biosensor that exploit the various mechanisms for modulating 7-HCAA fluorescence: First, I will describe the detection of glucose using a modified hexokinase system that utilizes the fact that 7-HCAA is a FRET acceptor for tryptophan.[17] This represents the first time that a 7-HCAA is used as a FRET acceptor in a protein that contains multiple tryptophans and suggests that 7-HCAA can be used as a non-perturbative fluorophore that has little effect on the binding capacity, while still serving as an indicator of small molecule binding.

Second, I will describe a structural study that provides additional information regarding how protein environments can tune the fluorescent output of 7-HCAA. Both spectroscopic and structural characterization of several 7-HCAA containing streptavidin mutants are reported in this study. In this work, emission spectra from the 7-HCAA were observed that have no precedent in the literature. Furthermore, structural data provide insight into how to 7-HCAAs fluorescent properties are modified by local environments. These data could find great use in the rational design of biosensors for future studies.

Lastly, I will present the computational design and experimental characterization of a two-state calcium-sensing biosensor that incorporates specific structural information obtained by Henderson et al.[20] This work highlights our ability to design new biosensors

using a computationally-based, top-down approach. This study provides a roadmap as to how structural and spectroscopic data can serve as inputs for computational design approaches that result in rationally designed biosensors in which 7-HCAA serves as the primary reporter of conformational changes in a protein.

# CHAPTER 2.

# AN INTRINSIC FRET SENSOR OF PROTEIN-LIGAND INTERACTIONS

## 2.1 Introduction

As previously mentioned in chapter 1, many biosensors that incorporate a 7-HCAA utilize FRET as the detection method. FRET is the non-radiative transfer of energy between donor and acceptor fluorophores. FRET-based techniques have proven to be exceptionally useful tools in biology, especially for the study of protein folding[23,33–35], protein-protein[5,36] and protein-ligand[14,37,38] interactions, and the dynamic fluctuations associated with protein function.[39] One advantage of FRET is that the efficiency of energy transfer—and therefore the fluorescent output of the acceptor fluorophore—is inversely proportional to the distance between the donor and acceptor to the sixth power.[40] Because small changes in distance between the donor and acceptor fluorophores can produce large changes in FRET efficiencies, FRET is well-suited to study processes that are associated with conformational changes in proteins. Furthermore, the use of FRET in ligand binding assays has been shown to be a useful alternative to other techniques such as ITC or fluorescence polarization.[37]

All protein-based FRET assays require that the donor and acceptor fluorophores are introduced at defined locations within a target protein.[41] One method for achieving this is to append genetically encoded fluorescent proteins to the N- and C- termini of the protein under study[42] but this is only useful if the termini are appropriately spaced. Alternatively, small molecule fluorophores can be covalently attached to the target protein post-translationally, but this often necessitates extensive protein engineering involving two orthogonal labeling methods.[43]

A possible solution to these challenges would be to use a genetically encoded fNCAA—which is directly incorporated within the target protein during protein expression—as a FRET donor or acceptor[14,18,44,45]. In FRET, the emission spectrum of the donor fluorophore must overlap with the excitation spectrum of the acceptor to form a FRET pair. This raises the intriguing possibility that an fNCAA with appropriate photochemical properties could serve as one component of a FRET pair with a naturally occurring amino acid. The excitation spectrum of the fNCAA L-(7-hydroxycoumarin-4-yl)ethylglycine (7-HCAA)[2] has extensive spectral overlap with the emission spectrum of the naturally occurring amino acid tryptophan (Trp, Figure 2.1) suggesting that these amino acids might constitute a viable FRET pair.

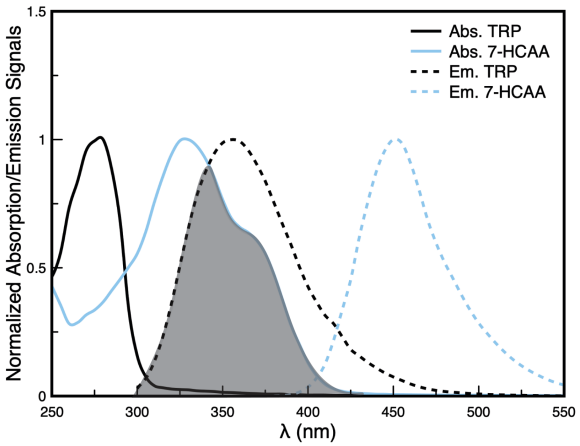



**Figure 2.1.** Absorption (solid lines) and emission (dashed lines) spectra for tryptophan (black lines) and 7-HCAA (light blue lines) at pH 7.5 showing the overlap integral in gray. Spectra are normalized to the $\lambda_{max}$ for each individual spectrum.

The potential of this approach has previously been studied by a number of groups. Chung and co-workers developed a FRET-based sensor of protein-ligand interactions using the biotin-binding protein streptavidin.[37] In this study, a series of coumarin derivatives were chemically conjugated to biotin and native Trp residues in SA were used as donor fluorophores; FRET was only observed in the biotin-SA complex. Alternatively, Hecht and

co-workers have developed FRET-based sensors in which two fNCAAs are incorporated within the same protein. In these studies, either 4-biphenyl-L-phenylalanine[14] (4BPA) or cyanotryptophan variants[46] served as donor fluorophores for 7-HCAA. According to the authors, the use of fNCAAs with spectral properties distinct from Trp was intended to circumvent the fact that many proteins contain multiple native Trp residues. More recently, Wang and co-workers used the Trp/7-HCAA FRET pair as a reporter of protein-protein interactions.[18] 7-HCAA was genetically encoded within β-arrestin, which does not contain any Trp residues. Complexation of β-arrestin with FAB30 (an antibody fragment that binds β-arrestin) resulted in FRET between Trp residues in FAB30 and 7-HCAA in the antigen. While other fNCAAs (e.g. acridonylalanine[5]) may have the potential to form a FRET pair with Trp, the spectral overlap between Trp and 7-HCAA (Figure 2.1) is far greater than for other fNCAAs. However, three-component FRET systems have been developed in which Trp serves as a donor fluorophore for the 7-methoxycoumarin moiety, which in turn serves as a donor fluorophore for acridonylalanine.[44]

This previous work establishes that Trp and 7-HCAA constitute a useful FRET pair. However, because these approaches employed multiple fNCAAs,[14,46] chemically conjugated fluorophores,[44] fluorescently modified ligands,[47] or target proteins that lack Trp,[18] these techniques may not be broadly applicable. In an effort to expand the utility of the Trp/7-HCAA FRET pair, we sought to determine if 7-HCAA can serve as a FRET acceptor in proteins containing multiple Trp residues. Here we report the rational design and characterization of a FRET-based sensor of ligand binding using 7-HCAA in a protein with three native Trps. Because our approach uses a single genetically encoded fNCAA, it has the potential to circumvent some of the aforementioned challenges.

## 2.2 Results and Discussion

**Design of a Genetically Encoded iFRET Sensor.** Our efforts began by identifying a protein that exhibits a significant conformational change upon substrate binding. Using the PyMOL molecular viewing software,[48] we examined a series of proteins in the Protein Structural Change Database[49] that undergo "coupled domain motions" during a ligand binding event. We found a hexokinase enzyme from the thermophilic archaea *Sulfolobus tokodaii*[50,51] that appeared highly suitable for this study. First, this hexokinase experiences a large conformational change upon binding hexose sugars including glucose. Second, three of the four native Trps are localized in C-terminal lobe of the protein (Figure 2.2) and the distance between the N- and C-terminal lobes decreases upon glucose binding. This suggests that incorporation of 7-HCAA at an appropriate site in the N-terminal lobe could result in changes in FRET efficiency upon glucose binding.

In FRET, the Förster distance is a characteristic of a given donor acceptor pair that defines the distance of separation at which half of the possible FRET efficiency will occur. For donors and acceptors separated by the Förster distance, small changes in inter-fluorophore distance will result in large changes in acceptor fluorescence.[41] We searched for residues that were greater than the reported Forster distance of 25.6 Å[37] away from the three tryptophans in the apo state (PDB ID: 2e2n) but which moved closer to the tryptophans in the holo state (PDB ID: 2e2o).

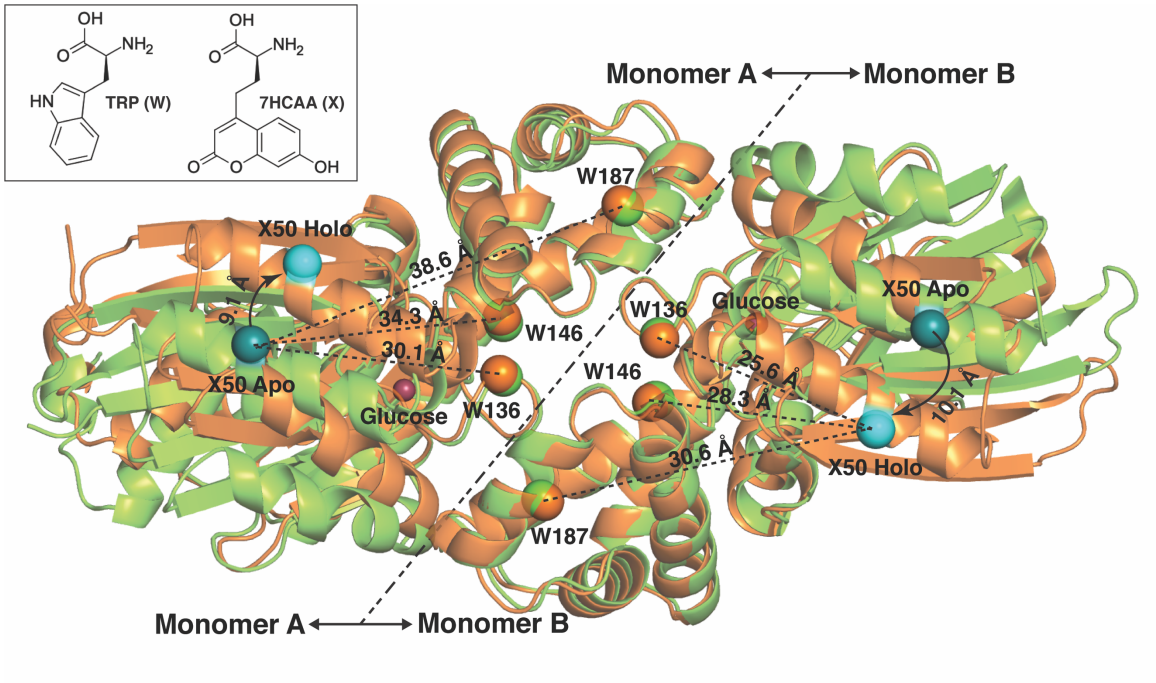



**Figure 2.2.** Models of the hexokinase dimer in the apo (green; PDB 2e2n) and holo (orange; PDB 2e2o) are shown superimposed with one another. Residue E50, which is mutated to 7-HCAA in this study, is shown as a dark teal or cyan sphere in the apo and bound forms, respectively. Three native tryptophan residues (W136, W146, and W187) are identified with orange and green spheres that correspond to their positions in the apo and holo structures, respectively. Distances between residue 50 and each tryptophan in monomer A are for the apo state and those depicted in monomer B are for the holo state. Spheres representing the glucose molecules are coloured burgundy. The differences in distances between the X50 in the apo and holo states in monomer A (9.1 Å) and monomer B (10.1 Å) can be attributed to the anisotropy of the crystal structures. Chemical structures of tryptophan (Trp, W) and L-(7-hydroxycoumarin-4-yl)ethylglycine (7-HCAA, X) are shown in an inset.

We identified residue E50 as a potential site of 7-HCAA incorporation. The Cα atom of E50 moves 4.5, 6.0 and 8.0 Å closer to the Cα atoms of W136, W146, and W187, respectively. Furthermore, E50 is surface-exposed, which suggests that substitution of this residue with 7-HCAA should have little effect on protein function. We also identified residue K257, which is surface-exposed and falls on the C-terminal lobe of hexokinase. Because the C-terminal lobe does not move upon glucose binding (Figure 2.3), no change in acceptor fluorescence was expected for the K257X mutant (where 'X' represents 7-HCAA) irrespective of whether or not glucose was bound.

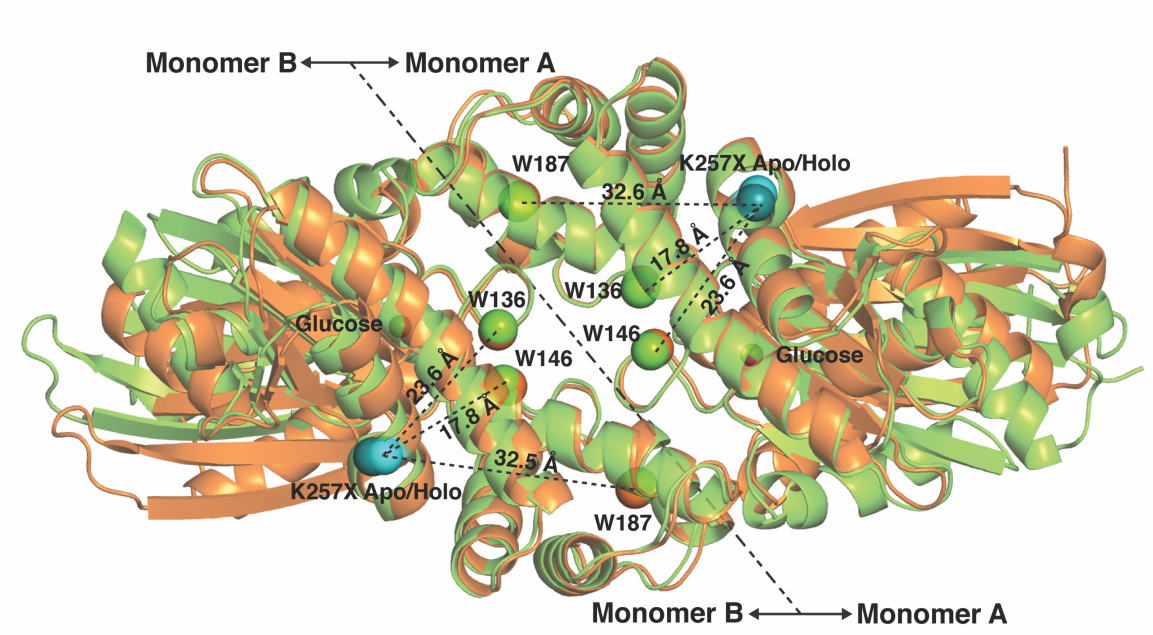



**Figure 2.3.** Models of the K257X hexokinase mutant dimer in the apo (green; PDB 2e2n) and holo (orange; PDB 2e2o) are shown superimposed with one another.[51] Residue K257, is shown as a dark teal or cyan sphere in the apo and holo forms, respectively. Three native tryptophan residues (W136, W146, and W187) are identified with orange and green spheres that correspond to their positions in the apo and holo structures, respectively. Distances between residue 257 and each tryptophan are displayed to indicate that residue 257 does not move in relation to any tryptophan upon glucose binding.

As noted previously, this hexokinase variant has four native Trp residues: Three Trps are located in the C-terminal lobe at positions 136, 146, and 187 and one is located in the N-terminal lobe at position 76. To avoid effects related to the presence of a Trp in the same lobe as the 7-HCAA, we sought to substitute the Trp at position 76 with a distinct amino acid. A BLAST sequence analysis identified a homologous protein from *S. acidocaldarius* that had 65% sequence homology to the *S. tokodaii* hexokinase, but which contained tyrosine at position 76. All studies were therefore performed using a hexokinase variant (termed HexY) that contained a W76Y mutation.

**Biochemical Characterization of 7-HCAA Modified Hexokinases.** To generate HexY and 7-HCAA-containing HexY variants, a gene encoding the wild type hexokinase was ordered from Integrated DNA Technologies (Coralville, IA) and was cloned into a

pET29b expression vector. Mutation of Trp 76 to Tyr was carried out using site-directed mutagenesis to generate the HexY construct. Using the newly constructed HexY gene as a template, the codons for residues E50 and K257 were individually mutated to the amber stop codon (TAG) again using site directed mutagenesis. 7-HCAA incorporation was achieved using the amber codon suppression methodology developed by Schultz and co-workers.[7]

Briefly, sequence confirmed E50X and K257X mutants were co-transformed into BL21 Star (DE3) cells along with the pEVOL-CouRS plasmid, which contains an orthogonal tRNA with an anticodon loop that is specific to the amber stop codon and an aminoacyl tRNA synthetase that acylates this tRNA with 7-HCAA. The E50X and K257X mutants were then expressed in the absence and presence of 7-HCAA. Fluorescent bands consistent with the full-length proteins were observed with SDS-PAGE analysis only when 7-HCAA was included in the expression medium (Figure 2.4). Because the parent hexokinase exists as a disulfide-mediated homodimer in its native state, we also analyzed each of the three proteins on an SDS-PAGE gel in the absence and presence of the reducing agent β-mercaptoethanol (BME). All three proteins showed a decrease in apparent size in the presence of BME suggesting that dimerization was not impeded by 7-HCAA incorporation (Figure 2.5).

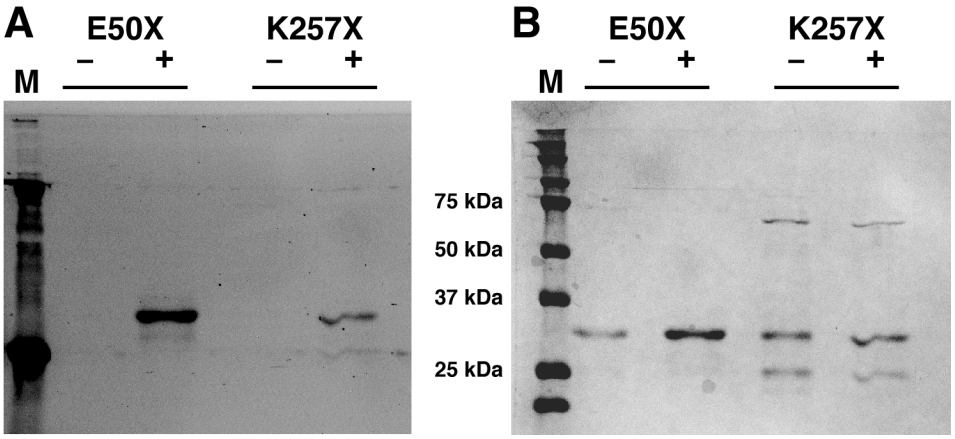



**Figure 2.4.** Reducing SDS-PAGE analysis of the expression of the E50X and K257X hexokinase mutants is shown. Visualization of the proteins was achieved by (A) UV light illumination and (B) coomassie stain. Protein expression of both mutants was carried out in the absence (-) and presence (+) of 1 mM 7-HCAA. Full-length protein expression is only expected in the presence of 7-HCAA. The presence of full-length protein in the (-) lanes in panel B can be attributed to amber suppression with a mis-acylated $tRNA_{CUA}$ and likely contain tyrosine at position 50 or 257 instead of 7-HCAA. In our experience, mis-acylation of the $tRNA_{CUA}$ with canonical amino acids occurs more frequently when NCAAs are not present during expression. Based on spectroscopic analysis (comparison of $A_{280}$ and $A_{325}$) of the mutant proteins, the amber codons in E50X and K257X were suppressed by 7-HCAA at ~92% and 86% efficiencies, respectively.

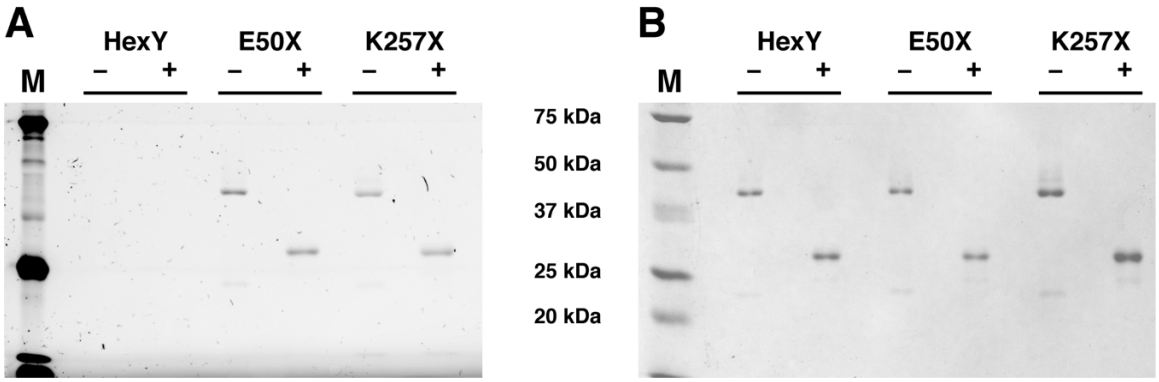



**Figure 2.5.** SDS-PAGE analysis of HexY and the E50X and K257X mutants under non-reducing and reducing conditions. Lanes marked with a minus sign contain no β-mercaptoethanol (BME) while those with a plus sign contained 1 mM BME. Visualization of the purified proteins was achieved with illumination by UV light (A) and coomassie stain (B). In the absence of BME, proteins with molecular weights corresponding to the disulphide-linked dimer of HexY are observed while monomeric species are observed in the presence of BME.

To confirm that the addition of 7-HCAA HexY had not negatively affected protein folding, we collected circular dichroism (CD) spectra of HexY and both mutants in Tris buffered saline (TBS, 50 mM Tris-HCl, pH 7.5, 150 mM NaCl). No differences in the CD spectra of the E50X or K257X mutants were observed relative to the parent protein (Figure 2.6A). We further explored the effects of 7-HCAA incorporation by conducting thermal

melts for all three proteins (Figure 2.6B) in TBS. Although the $T_M$ of both mutants was observed to be slightly lower than HexY (96.0 °C for HexY, 94.2 °C for E50X, and 93.7 °C for K257X, Figure 2.6B, Table 2.1), all variants were considered to be well folded in the subsequent assays, which were conducted at 25 °C. Finally, glucose binding of each mutant was analyzed in TBS with fluorescence polarization using 2-(N-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)amino)-2-deoxyglucose (2-NBDG), a fluorescently labeled glucose analogue. Mixing 2-NBDG with HexY, E50X and K257X resulted in an increase of polarized emission relative to the small molecule suggesting that the ability to bind glucose was maintained in all variants (Table 2.2).

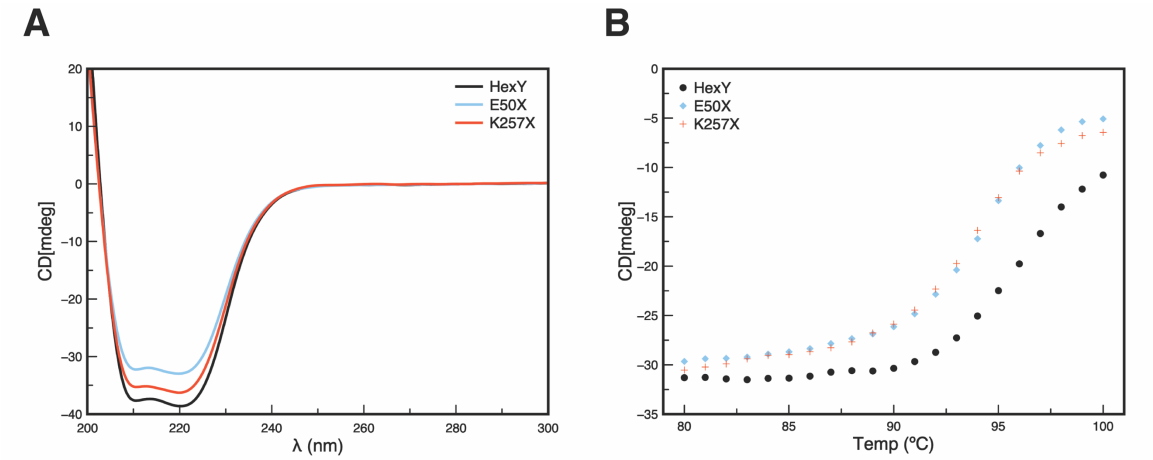



**Figure 2.6**. (A) CD spectra of HexY (black line), E50X (light blue line), and K257X (red line) (~10 µM) at pH 7.5. Spectra are the average of three readings. (B) Melting curve of HexY (black circles), E50X (light blue diamonds), and K257X (red crosses) at 222 nm. All data was processed and smoothed using the manufacturer's software.

**Table 2.1.** Parameters and Statistics for apparent $T_M$ curve fit.

| | W76Y | E50X | K257X |
|---|---|---|---|
| **Best-fit values** | | | |
| **Bottom** | -31.28 mDeg | -28.96 mDeg | -29.57 mDeg |
| **Top** | -7.831 mDeg | -3.147 mDeg | -4.744 mDeg |
| $T_M$ | 96.02 ºC | 94.19 ºC | 93.66 ºC |
| **HillSlope** | 48.53 | 48.28 | 45.16 |
| **log $T_M$** | 1.982 | 1.974 | 1.972 |
| **Span** | 23.45 mDeg | 25.82 mDeg | 24.83 mDeg |
| **95% CI (profile likelihood)** | | | |
| **Bottom** | -31.40 to -31.16 mDeg | -29.35 to -28.60 mDeg | -30.02 to -29.14 mDeg |
| **Top** | -8.549 to -7.017 mDeg | -4.208 to -1.871 mDeg | -5.832 to -3.409 mDeg |
| $T_M$ | 95.86 to 96.19 ºC | 93.95 to 94.47 ºC | 93.39 to 93.98 ºC |
| **HillSlope** | 45.96 to 51.21 | 42.96 to 54.17 | 39.60 to 51.34 |
| **log $T_M$** | 1.982 to 1.983 | 1.973 to 1.975 | 1.970 to 1.973 |
| **Goodness of Fit** | | | |
| **Degrees of Freedom** | 17 | 17 | 17 |
| **R squared** | 0.9996 | 0.9979 | 0.9974 |
| **Sum of Squares** | 0.4544 | 3.523 | 4.150 |
| **Sy.x** | 0.1635 | 0.4552 | 0.4941 |
| **Number of points** | | | |
| **# of X values** | 63 | 63 | 63 |
| **# Y values** | 21 | 21 | 21 |

**Table 2.2.** Fluorescence anisotropy of HexY mutants using 2-(N-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)amino)-2-deoxyglucose (2-NBDG).

| Free Dye / Construct | Fluorescence Polarization |
|---|---|
| 2-NBDG | 0.065 +/- 0.007 |
| HexY | 0.090 +/- 0.007 |
| E50X | 0.078 +/- 0.007 |
| K257X | 0.087 +/- 0.010 |

**Spectroscopic Analysis.** Initial FRET studies of the E50X and K257X mutants were carried out in TBS at a pH of 7.5, which corresponds to the previously reported pH for maximal enzymatic activity.[50] Both the E50X and K257X mutants showed emission at 450 nm when excited at 280 nm (Figure 2.7A). Alternatively, HexY (which does not contain 7-HCAA) showed no emission at 450 nm upon excitation at 280 nm (Figure 2.7A). Upon the addition of glucose, the E50X mutant showed a 47.4% increase in fluorescence

intensity when excited at 280 nm (Figure 2.7A); calculated FRET efficiencies were 15.3% and 19.2% for the apo and holo states, respectively. Alternatively, the addition of glucose to the K257X mutant did not result in a change in emission at 450 nm when excited at 280 nm (Figure 2.7A).

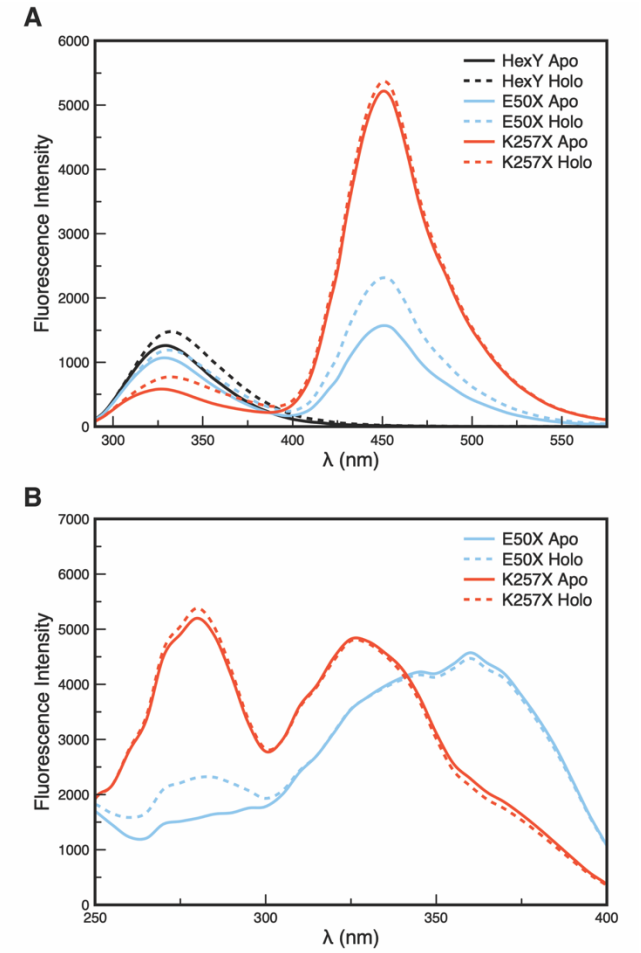



**Figure 2.7.** Fluorescence spectra of HexY constructs (~10 μM) in the absence (apo, solid lines) and presence (holo, dashed lines) of 10 mM glucose. (A) Fluorescence spectra ($\lambda_{ex}$ = 280 nm) for HexY (black lines), E50X (light blue lines) and K257X (red lines). (B) Excitation scan from 250-400 nm ($\lambda_{em}$ = 450 nm) at pH 7.5.

To confirm that changes in FRET efficiencies had given rise to the observed changes in fluorescence, an excitation scan was carried out at an emission wavelength of 450 nm. In the E50X mutant, the increase in fluorescence upon glucose binding was due to an increase in excitation at 280 nm (Figure 2.7B), suggesting that increased FRET from

tryptophan was the origin of this effect. Alternatively, no change in the excitation spectrum of K257X was observed between the apo and bound states (Figure 2.7B). Furthermore, no change in fluorescence was detected when the 7-HCAA was directly excited at 325 or 360 nm, corresponding to the excitation maxima of the neutral and phenolate forms, respectively (Figure 2.8).

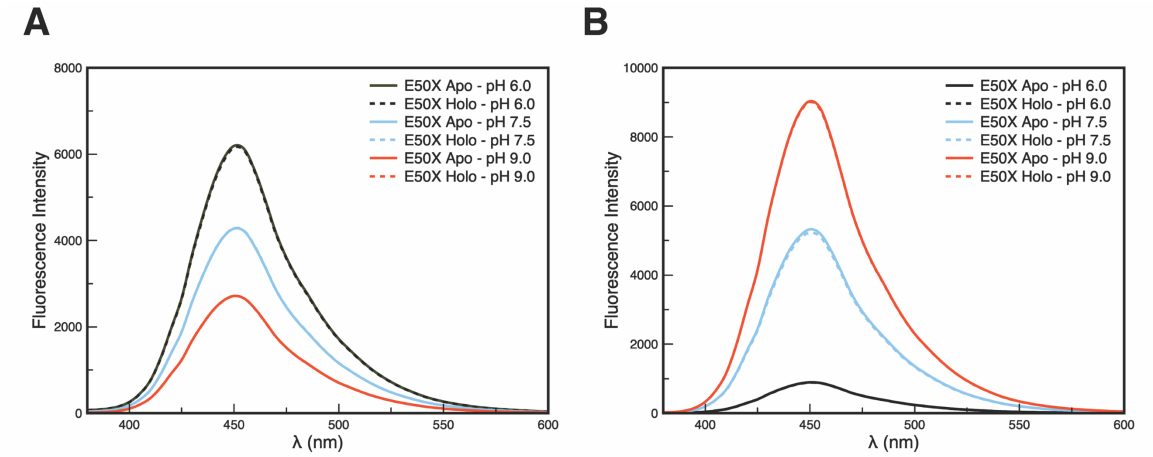



**Figure 2.8.** Fluorescence spectra for E50X in the absence (apo) and presence (holo) of 10 mM glucose for pH 6.0 (black lines), pH 7.5 (light blue lines), and pH 9.0 (red lines) excited with either (A) 325 nm or (B) 360 nm wavelength light.

Despite the fact that the K257X mutant showed no change in FRET efficiency in the presence of glucose, it did show a significant increase (232%) in overall fluorescence intensity at 450 nm relative to the E50X mutant when excited at 280 nm (Figure 2.7A). A 45.8% decrease in tryptophan fluorescence at 330 nm was also observed in the K257X mutant relative to the E50X mutant. These data are likely a consequence of the fact that position 257 is only 17.8 Å from one of the three tryptophans in HexY (Figure 2.3) while the closest tryptophan to E50X in the holo form is 25.6 Å away. Thus, the dramatic increase in fluorescence observed in the K257X mutant is likely a consequence of its proximity to the tryptophan at position 146.

As mentioned above, 7-HCAA contains a phenol (with a reported $pK_a$ of 7.8[8]) that can exist in a neutral or anionic state. Spectroscopic analysis of the E50X and K257X mutants revealed ratios of neutral and anionic 7-HCAA species that differed from those expected at pH 7.5 (compare Figures 2.1 with Figure 2.9A). Namely, the absorption spectrum of the E50X mutant confirmed the presence of both 7-HCAA species while the neutral species predominated in the K257X mutant (Figure 2.9). This suggests that the local environment surrounding 7-HCAA at positions 50 and 257 altered the p$K_a$ of the 7-HCAA phenols. It is possible that ligand binding could also induce changes in the local environment surrounding the 7-HCAA moiety that could, in turn, lead to changes in fluorescence intensity in a FRET independent manner. However, because the absorbance spectra of both the E50X and K257X mutants showed no differences between the apo and holo forms (Figure 2.9), this was not the case.

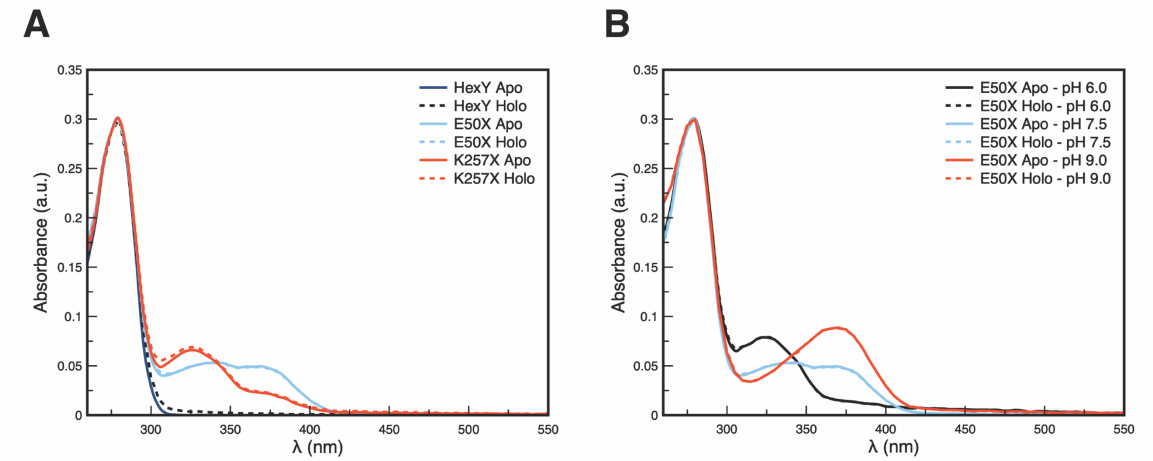



**Figure 2.9.** (A) Absorbance spectra of HexY (black lines) and the E50X (light blue lines) and K257X (red lines) mutants are shown in the absence (solid lines) and presence (dashed lines) of 10 mM glucose. (B) Absorbance spectra of the E50X mutant are shown at pH 6.0, 7.5, and 9.0 in the absence (solid lines) and presence (dashed lines) of 10 mM glucose. Spectra are scaled to absorbance at 280 nm.

Because the neutral and anionic forms of 7-HCAA possess distinct photochemical properties (e.g. excitation maxima of 325 and 360 nm, respectively), it is likely that one

species may be a better FRET acceptor for Trp than the other. To examine this, we again excited E50X at 280 nm in the presence and absence of glucose at pHs of 6.0 (50 mM Bis-Tris, pH 6.0, 150 mM NaCl) and 9.0 (50 mM Tris-HCl, pH 9.0, 150 mM NaCl). The absorbance spectrum of E50X at pH 6.0 showed only neutral 7-HCAA, while at pH 9.0, only the phenolate form was observed (Figure 2.9B). At pH 6.0, a smaller increase (28.4%) in emission at 450 nm than was observed at pH 7.5 but the signal was increased relative to the other pH values regardless of the presence of glucose (Figure 4). At pH 9.0, a fluorescence intensity increase of 64.9% was observed (Figure 4), which represents the largest change in fluorescence observed for E50X upon glucose binding. However, the overall fluorescence intensities of the apo and holo forms of E50X at pH 9.0 were less than those observed at pH 6.0 (Figure 2.10). Excitation scans at an emission wavelength of 450 nm again indicated that increases in fluorescence at these pHs were strictly due to changes in excitation at 280 nm (Figure 2.11). No change in fluorescence intensity at 450 nm was observed when 7-HCAA was excited at 325 or 360 nm (Figure 2.8) at pH 6.0 or 9.0.

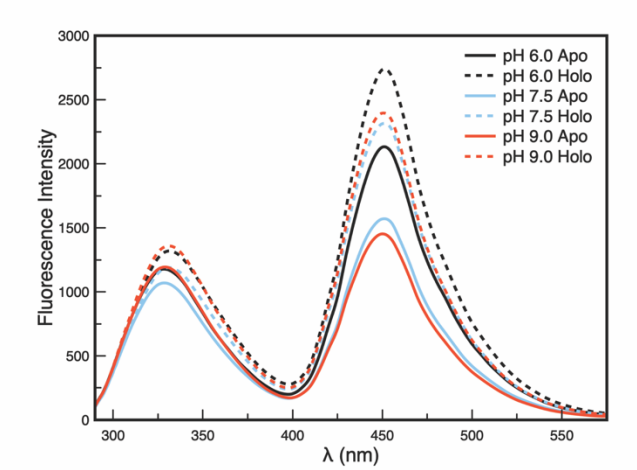



**Figure 2.10.** Fluorescence spectra of E50X ($\lambda_{ex}$ = 280 nm) in the absence (solid lines) and presence (dashed lines) of 10 mM glucose at various pH values are shown. Data were collected at pH 6.0 (black lines), pH 7.5 (light blue lines), and pH 9.0 (red lines).

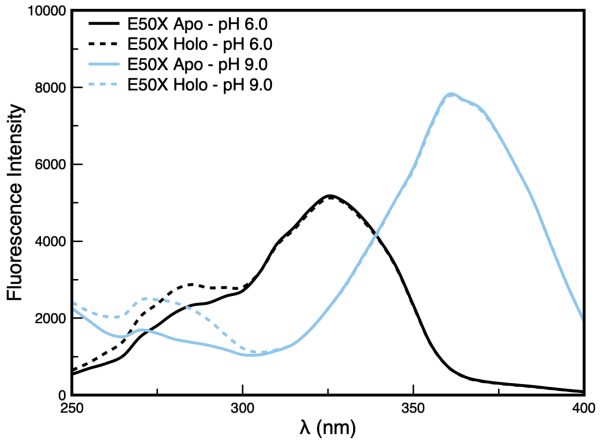



**Figure 2.11.** Excitation scan of 450 nm emission at pH 6.0 (black lines) and pH 9.0 (light blue lines). These data show that increases in emission at 450 nm are localized to 280 nm excitation and can be associated with a FRET event from tryptophan.

These data can be rationalized in terms of both the overlap integrals of Trp and 7-HCAA in the neutral and anionic forms and the absorbance spectra of the two 7-HCAA species. As seen in Figure 2.12, neutral 7-HCAA absorbs substantially more 280 nm light than the anionic form, but its overlap integral with Trp is greatly reduced. This is consistent with the observation that the fluorescence intensity at pH 6.0 is much higher than that at pH 9.0 irrespective of the presence of glucose (Figure 2.10). However, because the overlap integral between Trp and neutral 7-HCAA is reduced relative to the anionic form, the change in intensity due to FRET is also reduced at pH 6.0 relative to pH 9.0. The fluorescence data obtained at pH 7.5 are also consistent with these observations. The absorbance spectrum of E50X at pH 7.5 indicates a mixed population of neutral and anionic species, but the anion predominates. Thus, although the emission spectra of the E50X mutant at pH 7.5 are highly similar to those observed at pH 9.0, the presence of a small amount of neutral 7-HCAA at this pH results in an overall increase in fluorescence intensity in the apo form and a smaller increase in intensity due to FRET upon substrate binding. Although the highest fluorescence intensity was observed at pH 6.0, the fact that a

reproducible FRET-based change in fluorescence intensity due to ligand binding was observed at pH 7.5 suggests that this approach should be useful for studying protein function in physiologically relevant conditions.

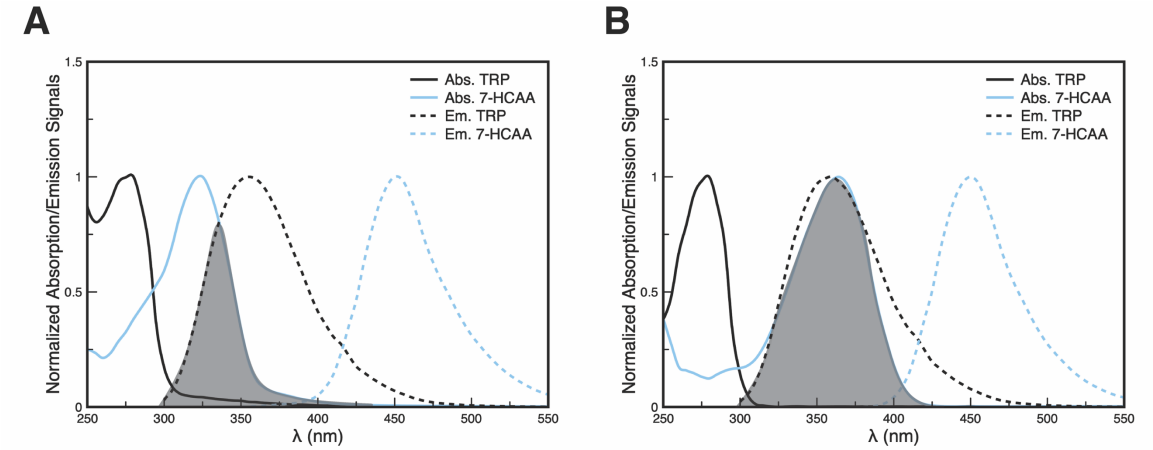



**Figure 2.12.** Absorption (solid lines) and emission (dashed lines) spectra for tryptophan (black lines) and 7-HCAA (light blue lines) at (A) pH 6.0 and (B) pH 9.0 showing the overlap integral in grey. Spectra are normalized to the $\lambda_{max}$ for each individual spectrum.

The aforementioned data also suggest the possibility of tuning the properties of this system via careful choice of the site of 7-HCAA incorporation. In the event that high fluorescent signal is desired at the expense of sensitivity, 7-HCAA might be placed in proximity to acidic residues that would favor the neutral form. Alternatively, if high sensitivity is desired irrespective of emission intensity, 7-HCAA can be incorporated in the vicinity of basic residues, which would favor the anionic form.

**Glucose Binding Affinity of E50X.** Finally, we hoped to explore whether the Trp/7-HCAA pair could be used to directly study protein function. Given the dependence of hexokinase activity on pH,[50] we assessed the $K_d$ of the enzyme at various pHs using FRET. Namely, E50X fluorescence was measured in the presence of glucose at concentrations ranging from 100 nM – 10 mM and in buffers (described above) with pH values of 6.0, 7.5, and 9.0 (Figure 2.13). A sigmoidal increase in fluorescence at 450 nm

was observed with increasing glucose concentration. A non-linear least squares fit of these data gave an apparent $K_d$ of 48.7 µM at pH 7.5 (Table 2.3), which is in good agreement with a previously reported $K_m$ value of 50 µM.[50] At pH 6.0 the apparent $K_d$ was 77.0 µM while at pH 9.0 it was 753.2 µM (Table 2.3). These data suggest the utility of this method for studying protein function without the need for fluorescently labeled substrates.

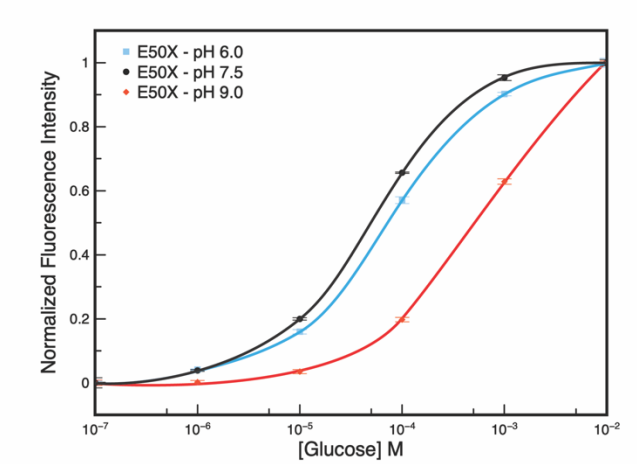



**Figure 2.13.** Glucose dependent response of E50X at a pH of 6.0 (black circles and line; $K_d$ = 77.0 µM), pH of 7.5 (light blue squares and line; $K_d$ = 48.8 µM), and pH 9.0 (red diamonds and line; $K_d$ = 753.2 µM). Normalized fluorescence data ($\lambda_{ex}$ = 280 nm, $\lambda_{em}$ = 450 nm) for final glucose concentrations of 100 nM, 1 µM, 10 µM, 100 µM, 1 mM, 10 mM.

**Table 2.3.** Parameters and Statistics for apparent $K_d$ curve fit.

| | E50X - pH 6.0 | E50X - pH 7.5 | E50X - pH 9.0 |
|---|---|---|---|
| **Best-fit values** | | | |
| **Bottom** | 2132 RFU | 1659 RFU | 1409 RFU |
| **Top** | 2611 RFU | 2320 RFU | 2490 RFU |
| ***K*$_d$** | 76.99 µM | 48.76 µM | 653.2 µM |
| **HillSlope** | 0.8409 | 0.8631 | 0.7717 |
| **log *K*$_d$** | -4.114 | -4.312 | -3.123 |
| **Span** | 479.3 | 661.3 | 1081 |
| **95% CI (profile likelihood)** | | | |
| **Bottom** | 2116 to 2146 RFU | 1645 to 1672 RFU | 1401 to 1417 RFU |
| **Top** | 2587 to 2640 RFU | 2302 to 2341 RFU | 2454 to 2533 RFU |
| ***K*$_d$** | 58.75 to 102.7 µM | 41.37 to 57.48 µM | 664.2 to 870.4 µM |
| **HillSlope** | 0.6794 to 1.056 | 0.7618 to 0.9778 | 0.7147 to 0.8314 |
| **log *K*$_d$** | -4.231 to -3.988 | -4.383 to -4.240 | -3.178 to -3.060 |
| **Goodness of Fit** | | | |
| **Degrees of Freedom** | 17 | 17 | 17 |
| **R squared** | 0.9922 | 0.9971 | 0.9991 |
| **Sum of Squares** | 6058 | 4517 | 2282 |
| **Sy.x** | 18.88 | 16.30 | 11.58 |
| **Number of points** | | | |
| **# of X values** | 21 | 21 | 21 |
| **# Y values** | 21 | 21 | 21 |

## 2.3 Conclusion

In this study we demonstrated for the first time that an fNCAA containing a 7-hydroxycoumarin moiety can be used as a FRET acceptor in proteins with multiple native Trp residues. When incorporated at an appropriate site, 7-HCAA is minimally perturbative to protein function and has the ability to report on ligand binding events that are coupled to conformational changes in a target protein. Our approach represents a direct path to metabolite sensing that obviates the need for protein labeling with multiple fluorophores, the use of multiple fNCAAs or fluorescently labeled ligands. Although 7-HCAA's short excitation and emission wavelengths suggest that this FRET pair is most suited to applications that are performed *ex vivo*, techniques such as two-photon excitation could make it possible to conduct *in vivo* studies. Thus, this system may find immediate use in high-throughput drug screening, monitoring metabolite flux, or any other applications for

which a minimally invasive FRET pair that reports on protein functions coupled to conformational changes is desired.

## 2.4 Materials and Methods

**Reagents.** 7-HCAA was obtained from Bachem (Torrance, CA) and 2-(N-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)amino)-2-deoxyglucose (2-NBDG) was obtained from Cayman Chemical Company (Ann Arbor, MI). Commercially available reagents were used without further purification.

**Molecular Cloning, Protein Expression, & Purification.** The wild-type hexokinase gene from *Sulfolobus tokodaii*[50,51] was obtained from Integrated DNA Technologies (Coralville, IA) cloned into a pET29b(+) vector (Novagen) using Gibson assembly (New England Biolabs). First, the W76Y mutation was made with overlap extension PCR and Gibson Assembly using the wild-type gene as a template. The presence of the mutation was confirmed using Sanger sequencing (Genewiz; Plainfield, NJ). The E50X and K257X mutations were each made separately using overlap extension PCR followed by Gibson assembly. Sequence confirmed pET29b(+) plasmids containing the E50X and K257X mutations were transformed into BL21* (DE3) cells (Invitrogen) along with the pEVOL-CouRS[2,52] plasmid. The pEVOL-CouRS plasmid contains an orthogonal aminoacyl tRNA synthetase / tRNA pair that includes a tRNA with an anti-codon loop specific to the amber codon (termed $tRNA_{CUA}$) and a CouRS aminoacyl tRNA synthetase that can acylate the $tRNA_{CUA}$ with 7-HCAA. Because pET29b(+) contains a kanamycin resistance marker and pEVOL-CouRS contains a chloramphenicol resistance marker, BL21 Star(DE3) (Invitrogen; Carlsbad CA) cells containing either the E50X or K257X hexokinase mutant were plated on dual-antibiotic plates (kanamycin, 50 μg / mL and

chloramphenicol, 34 μg / mL) and were allowed to incubate overnight at 37 ºC. Alternatively, BL21* (DE3) cells containing HexY were cultured on plates containing kanamycin alone.

A single colony was used to inoculate 5 mL of 2xYT media and cultures were incubated overnight at 37 ºC with 250 rpm shaking to an $OD_{600}$ of ~6.0. The cultures were then used to inoculate in 50 mL of TB media supplemented with 100 mM Tris-HCl, pH 7.0. Arabinose (0.2% w/v) and the non-canonical amino acid, 7-HCAA (1 mM final concentration), were added to each culture and incubated for 1 hour at 37 ºC with 250 rpm shaking. This incubation period prior to HexY overexpression serves to induce production of CouRS. Expression of the HexY mutants was induced by addition of isopropyl-β-D-thiogalactoside to a final concentration of 1 mM. The cultures were then incubated between 8 and 16 hours at 25 ºC, with shaking at 180 rpm. Cells were harvested via centrifugation (4,200 xg for 10 minutes), resuspended in lysis buffer (25 mM Tris-HCl, pH 8.0, 10 mM NaCl) and frozen. Cells were thawed and lysozyme (1 mg / mL final concentration; Sigma Aldrich) and NaCl (500 mM final concentration) were added. Cells were then lysed via sonication for 5 minutes and centrifuged at 20,000 xg for 20 minutes to remove cell debris.

All proteins contained C-terminal 6x histidine tags and were purified on immobilized metal ion chromatography using a nickel-nitrilotriacetic acid resin (Ni-NTA, HisTrap FF, GE Healthcare). After loading on Ni-NTA, contaminant proteins were removed by washing the column with 5 column volumes (CV) of Ni-NTA buffer A (25 mM Tris-HCl pH 8.0, 20 mM imidazole and 500 mM NaCl) followed by 5 CV of 90% Ni-NTA buffer A with 10% Ni-NTA buffer B (25 mM Tris-HCl pH 8.0, 500 mM imidazole and 150 mM NaCl). Proteins were then eluted with five column volumes of 100% Ni-NTA

buffer B. Eluted fractions that were fluorescent under UV illumination were consolidated, diluted 10x with anion exchange (IEC) wash buffer A (25 mM Tris-HCl pH 8.0, 10 mM NaCl) and further purified on IEC resin (HiTrap Q FF, GE Healthcare). The column was washed with five CV of IEC wash buffer A, followed by five CV of 90% IEC buffer A and 10% IEC buffer B (25 mM Tris-HCl pH 8.0, 500 mM NaCl). Proteins were eluted with five column volumes of 100% IEC buffer B. Again, protein fractions that were found to be fluorescent were consolidated and concentrated in a 10 kDa molecular weight cutoff spin column (Amicon) to 500 µL. The concentrated protein was then injected onto a size-exclusion column (Superdex 200 increase 10/300 gl, GE Healthcare) and eluted from the column with sizing buffer (50 mM Tris-HCl pH 9.0, 150 mM NaCl). Fractions that were fluorescent were collected, consolidated, and concentrated to a volume of ~1 mL on a 10 kDa molecular cutoff spin column (Amicon) and refrigerated at 4 ºC for use in experimental procedures.

**Reducing Plus/Minus 7-HCAA SDS-PAGE.** Single colonies of the E50X or K257X mutants of HexY were used to inoculate 5 mL cultures of 2xYT medium containing chloramphenicol and kanamycin. After growth overnight at 37 °C, 100 µL of confluent culture was used to inoculate two 5 mL cultures of each mutant. 7-HCAA (in 200 mM NaOH) was added to a final concentration of 1 mM in one of the two culture tubes containing cells harboring the E50X or K257X expression plasmids; an equivalent amount of 200 mM NaOH was added to the other tube. Protein expression was carried out as described above to the point of cell lysis. Rather than purifying the proteins, 100 µL of cell culture was removed from the 5 mL overnight cultures and centrifuged at 8,000 xg for 10 minutes. Pellets were lysed as previously described. Proteins were then purified using a

gravity Ni-NTA column. 500 µL of Ni-NTA slurry (Qiagen; Hilden, Germany) was placed in a micro spin column (Bio-Rad; Hercules, CA). The resin was washed with 5 CV water and 5 CV Ni-NTA buffer A. The lysed cell supernatant was added to the column, washed with 5 CV of N-NTA buffer A and eluted with 5 CV of Ni-NTA buffer B. 15 µL of sample was mixed with 5 µL of 4x Laemmli buffer with β-mercaptoethanol (BME; Sigma Aldrich; St. Louis, MO) and brought to a final volume of 20 µL with water. The sample was heated to 95 ºC for 5 minutes and 15 µL of sample was loaded into each well of a 4% stacking/15% resolving polyacrylamide gel and run at 100 V for 90 minutes. Visualization was again carried out using the GelDoc XR followed by Coomassie staining.

**Non-Reducing/Reducing SDS-PAGE.** 15 µg of each of the purified HexY mutants was mixed with 5 µL of 4x Laemmli buffer with BME and brought to a final volume of 20 µL with water. The sample was heated to 95 ºC for 5 minutes and 15 µL of sample was loaded into each well of a 4% stacking/15% resolving polyacrylamide gel and run at 100 V for 90 minutes. Visualization was again carried out using the GelDoc XR followed by Coomassie staining.

**Fluorescence Polarization Assay.** 90 µL of protein sample (~10 µM) was mixed with 10 µL of 2-NBDG (1 mM) (Cayman Chemical Company; Ann Arbor, MI) in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 50 mM Bis-Tris pH 6.0, 150 mM NaCl. Polarization data were collected on a HORIBA Nanolog fluorimeter using 440 nm excitation, 520 nm emission, and 5 nm slit widths for both. Data was analyzed using the fluoressence software package, version 3.5.1.991 (HORIBA Jobin Yvon).

**Spectroscopic Analysis.** Spectroscopic experiments were performed in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 50 mM Bis-Tris pH 6.0, 150 mM NaCl, or 50 mM Tris-

HCl, pH 9.0, 150 mM NaCl using a 1 cm quartz cuvette (Starna Cells; Atascadero, CA) on a SpectraMax M5 spectrophotometer (Molecular Devices; San Jose, CA). Each glucose-free mutant was concentrated to a volume of 100 μL and then diluted to a final absorbance value of 0.05 at 325 nm for pH 7.5, 0.1 at 325 nm for pH 6.0, and 0.1 at 360 nm for pH 9.0. The samples were then divided into separate tubes into which glucose (10 mM final concentration) or an equivalent volume of buffer was added. Absorbance spectra were taken from 200 nm to 750 nm at 5 nm intervals, again using a SpectraMax M5 spectrophotometer. Fluorescence excitation spectra at ($\lambda_{em}$ = 450 nm) were collected with excitation wavelengths ranging from 250 nm to 400 nm at 5 nm intervals. Fluorescence emission spectra were collected by exciting the proteins at either 280 nm, 325 nm or 360 nm and measuring the fluorescence intensity from 380 nm to 600 nm in 5 nm intervals. All spectra represent the average of three readings.

**Calculation of Fluorescence Intensity Changes and FRET Efficiencies for E50X.** The increase in fluorescence intensity from E50X apo to E50X holo was calculated using Equation 1. ΔI is the calculated change in fluorescence intensity and $E50X_{holo}$ and $E50X_{apo}$ represent the absolute fluorescent intensities of this mutant in the presence and absence of glucose, respectively. FRET efficiencies of E50X in the apo and holo forms were calculated using Equations 2 and 3, respectively. In these equations, $E50X_{apo}$, $E50X_{holo}$, $HexY_{apo}$, and $HexY_{holo}$ represent the fluorescent intensities of these proteins in the specified forms at 330 nm.

$$\Delta I = \frac{E50X_{holo} - E50X_{apo}}{E50X_{apo}} \quad \textbf{(Eq. 1)}$$

$$E = 1 - \left(\frac{E50X_{apo}}{HexY_{apo}}\right) \quad \textbf{(Eq. 2)}$$

$$E = 1 - \left(\frac{E50X_{holo}}{HexY_{holo}}\right) \quad \textbf{(Eq. 3)}$$

**E50X Glucose Dependent Response.** 10 µL of 1 M glucose was serial diluted in sterile water to concentrations of 100 mM, 10 mM, 1 mM, 100 µM, 10 µM and 1 µM. 10 µL of each glucose concentration was then mixed with 90 µL of E50X (10 µM) that had been dialyzed into the buffers listed above at pHs of 6.0, 7.5, and 9.0; final glucose concentrations were: 10 mM, 1 mM, 100 µM, 10 µM, 1 µM and 100 nM. Samples were excited at 280 nm and fluorescence was measured at 450 nm using a 1 cm quartz cuvette (Starna Cells; Atascadero, CA) on a SpectraMax M5 spectrophotometer (Molecular Devices; San Jose, CA). All data were normalized to the highest value at each pH and are the average of three readings. Curve fitting was performed with the Prism software package (Graphpad; San Diego, CA) using the 4-parameter logistic (4PL) non-linear regression function (Eq. 4). $K_d$ values derived from these fits are listed in Table 2.2.

$$f(x) = d + \frac{a-d}{1+\left(\frac{x}{c}\right)^b} \quad \textbf{(Eq. 4)}$$

**Circular Dichroism.** 300 µL of protein sample (~10 µM) was placed into a 1 mm quartz cuvette (Starna Cells; Atascadero, CA) in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 50 mM Bis-Tris pH 6.0, 150 mM NaCl. Multi-wavelength CD measurements were taken every 1 nm from 300 to 200 nm. For melting curves, sample was incubated with 5 mM BME (Sigma Aldrich; St. Louis, MO) at room temperature for 1 hour prior to the reading. Measurements were taken from 80 to 100 ºC at 222 nm wavelength. All measurements were taken on a Jasco J-815 CD Spectrometer (Jasco Inc., Easton, MD) and processed with the Jasco Spectra Manager V1.54.03 software package. Curve fitting was performed with the Prism software package (Graphpad; San Diego, CA); the 4-parameter logistic (4PL)

non-linear regression function (Eq. 4) was used again. $T_M$ values derived from these fits are listed in Table 2.3.

# CHAPTER 3.

# STRUCTURAL INSIGHTS INTO THE ORIGIN OF ALTERED SPECTROSCOPIC PROPERTIES OF A FLUORESCENT NON-CANONICAL AMINO ACID IN RESPONSE TO SMALL MOLECULE BINDING

## 3.1 Introduction

Although FRET-based approaches have found great use in the development of fluorescent, protein-based sensors of substrate binding, the limitation that two distinct fluorophores must be introduced in a target protein at well-defined sites still exists. Although the approach described in the previous chapter suggests the promise of generating sensitive FRET pairs using 7-HCAA and a naturally occurring amino acid, it seems to be limited to proteins in which extensive conformational changes occur as a consequence of substrate binding. An alternative approach might be to genetically encode a single fNCAA whose fluorescent properties would be altered directly by a substrate binding event. Although this approach would obviate the need to add two fluorophores to a protein, a current limitation exists in that only scant information regarding how fluorophores like 7-HCAA respond to changes in their local environments. Thus, in this chapter, I describe our efforts to directly ascertain how interactions between the protein and the 7-HCAA can modify the fluorescence output of the 7-HCAA.

To achieve this, a 7-HCAA was genetically encoded at five positions in the biotin-binding protein streptavidin (SAV). The photochemical properties of each mutant were assessed in the presence and absence of biotin and revealed biotin-dependent changes in fluorescence in four of the five mutant proteins. Structural characterization of three of these mutants was also achieved, which provided direct evidence of the manner in which the

chemical environment surrounding 7-HCAA alters its spectral properties. The results of this study should therefore prove useful in future efforts to develop novel protein-based fluorescent sensors of analyte binding.

### 3.2 Results

**Selection of sites of 7-HCAA incorporation.** To choose sites of 7-HCAA incorporation, we first analyzed a high-resolution (0.95 Å) structure of SAV bound to biotin (PDB ID: 3ry2).[53] Our sole selection criterion was the proximity of a candidate residue to the biotin binding site. Although it is difficult to confidently predict the orientation 7-HCAA would adopt at a given position in SAV, we reasoned that substitution of 7-HCAA at sites suitably close to the biotin binding pocket could result in biotin-dependent alterations of the chemical environments in the vicinity of the fNCAA. We were not concerned as to whether the distinct chemical environments were a result of changes in 7-HCAA conformation due to biotin binding or due to the introduction of new chemical functional groups that were contributed directly by biotin itself.

Inspection of the 3ry2 structure identified five residues on the "lip" of the biotin binding pocket (L25, L110, S112, W120, and L124; Figure 3.1) that were not directly involved in hydrogen bonding interactions with biotin. Although the side chains of residues 45-50 are in close proximity to the biotin substrate, they fall on the flexible "loop 3/4", which adopts "open" and "closed" conformations in the absence and presence of biotin, respectively; these residues were excluded from consideration.

7-HCAA incorporation was achieved using the Amber codon suppression methodology developed by Schultz and co-workers.[7] Full-length protein expression was observed for all five mutants only when 7-HCAA was present in the growth media (Figure

3.2), suggesting that protein expression is dependent on the presence of 7-HCAA. Mutant proteins were refolded from inclusion bodies using a modified version of a previously reported protocol[54] and further purified using immobilized metal ion, anion exchange, and size exclusion chromatography (see materials and methods for full purification details).

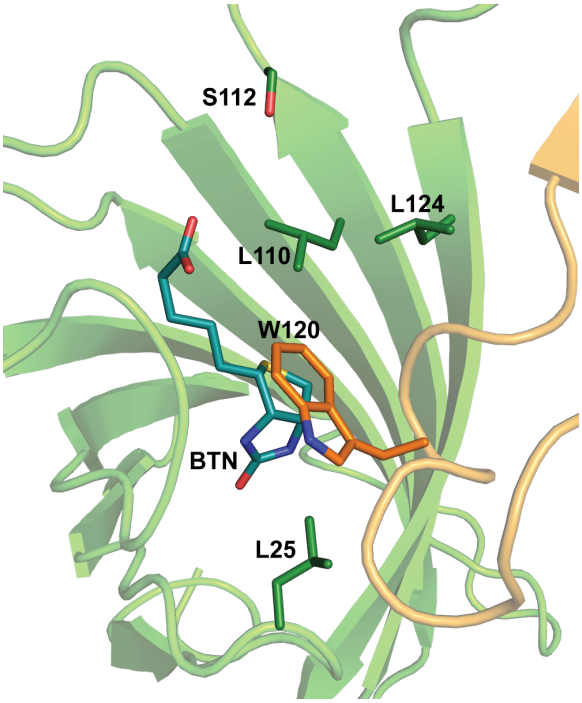



**Figure 3.1.** Sites of 7-HCAA incorporation. The structure of one subunit of the streptavidin tetramer (Green) and an adjacent subunit (Orange) are shown in cartoon form (PDB ID: 3ry2).[53] Biotin (teal) and the side chains of the five residues that were substituted with 7-HCAA are shown as sticks. Residue W120 (orange sticks) lies on an adjacent chain of streptavidin but interacts with the biotin bound to the primary subunit displayed in this figure.

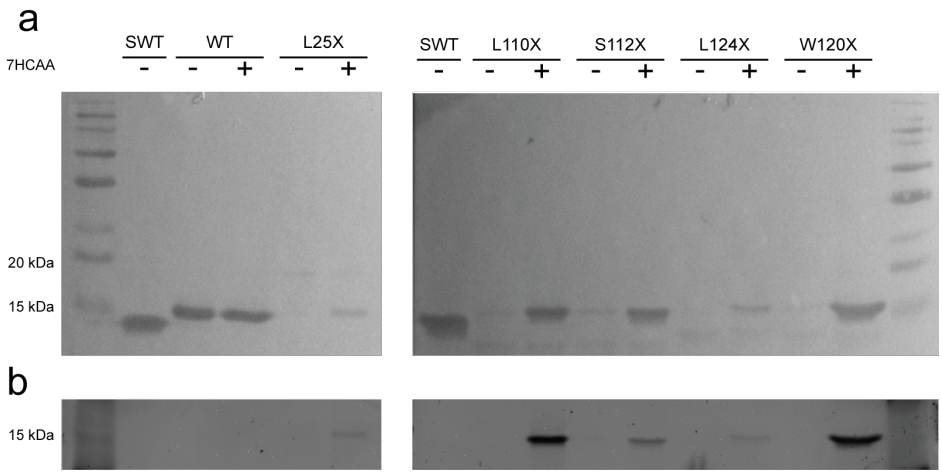



**Figure 3.2.** Streptavidin Mutant +/- 7-HCAA Expression. (a) Western blot and (b) fluorescent image of whole cell *E. coli* lysates for each streptavidin mutant either in the presence or absence of 7-HCAA. 1 μg of commercially purchased wild-type streptavidin (SWT) without a 6x-his tag was used as a standard. Faint bands can be detected in the L25X, L110X, S112X, L124X, and W120X without 7-HCAA suggesting that a fraction of the protein can be expressed with tyrosine in lieu of 7-HCAA. Additionally, faint bands below each of the full-length L110X, S112X, L124X, and W120X bands is indicative of truncated protein in which the stop codon was not fully suppressed.

**Spectroscopic and structural characterization of 7-HCAA containing SAV mutants.** In order to explore the effect of biotin binding on the spectral properties of 7-HCAA, we collected absorbance and fluorescence spectra of each SAV variant in the absence (apo-form) and presence (holo-form) of saturating amounts (100 μM) of biotin. At neutral pH, the 7-HC moiety exhibits two absorbance maxima at ~325 nm and ~360 nm that correspond to the neutral and anionic species, respectively (Figure 2a inset). Because the pKa of 7-HC is ~7.8,[8] the phenol and phenolate species should have both been present under the assay conditions (20 mM Tris-HCl, pH 7.0, 150 mM NaCl). For this study, all proteins were excited at 340 nm, which corresponds to the isosbestic point of 7-HCAA. When possible, structural data were also collected for each protein variant. The spectroscopic and structural data from all mutants assayed in this study are described below.

**L25X Characterization.** The L25X mutant (where X signifies mutation of the native residue to 7-HCAA) shows a fluorescence increase of 40% in the presence of biotin (Table 3.1; Figure 3.2b). Unfortunately, protein expression yields were too low and the protein was too unstable in solution to investigate the structural basis for this change in fluorescence. However, since a single absorbance maximum at 325 nm was observed for the L25X mutant in both the apo- and holo-forms (Figure 3.2b inset), it appears that 7-HCAA exists primarily in the neutral form regardless of the presence or absence of biotin. In spite of the lack of structural data for this mutant, we observe that the change in emission is not derived from changes in protonation state. If this had been the case, there would be changes in the absorbance spectrum, yet here, the same absorbance spectrum is present regardless of the presence of biotin.

**Table 3.1.** Spectroscopic data for SAV mutants.

| Mutant | Abs. Max (apo / holo) | Δ Abs. Intensity (holo : apo) | Em. Max (apo / holo) | Δ Em. Intensity (holo : apo) |
|---|---|---|---|---|
| **7-HCAA (pH 7.0)** | 326 | N/A | 452 | N/ A |
| **L25X** | 325 / 325 | -3% | 453 / 455 (+2 nm) | +40% |
| **L110X** | 324 / 324 | 12% | 449 / 435 (-14 nm) | -44% |
| **S112X** | 331 / 325 (-6 nm) | 10% | 451 / 455 (+4 nm) | +72% |
| **W120X** | 329 / 332 (+3 nm) | -18% | 453 /433 (-20 nm) | +135% |
| **L124X** | N/A | N/A | 450 /450 | +6% |

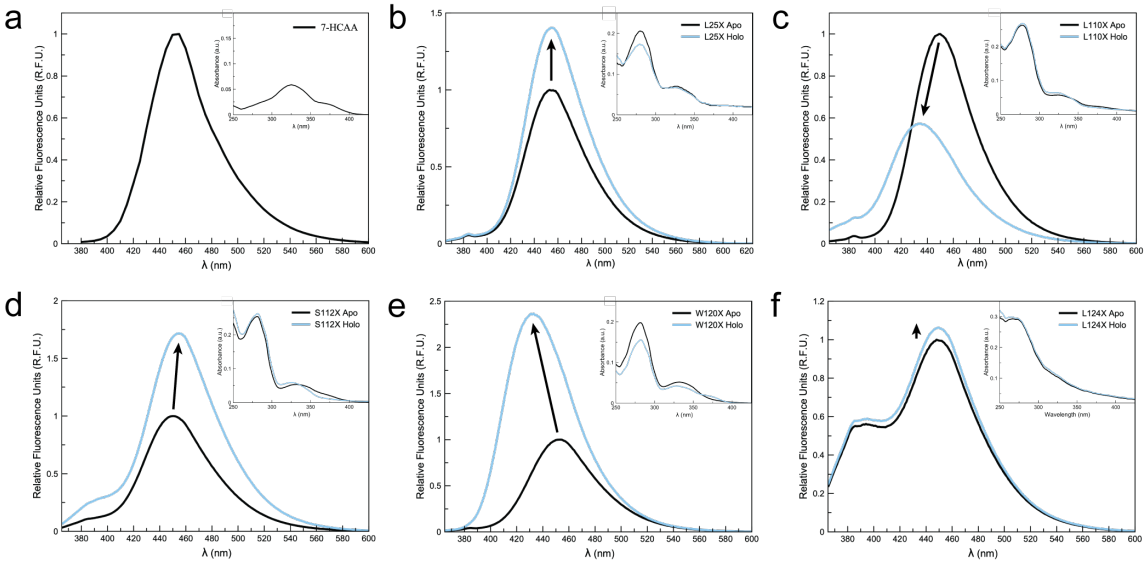



**Figure 3.3.** Normalized fluorescence and absorbance spectra of the (a) 7-HCAA, (b) L25X, (c) L110X, (d) S112X, (e) W120X, and (f) L124X mutants. Fluorescence spectra of the apo and holo forms of each mutant are shown in black and cyan lines, respectively. The fluorescence intensity of the apo form of each protein was set to unity in each case. Absorption spectra for each mutant are shown in the insets; again, black and cyan lines represent the apo and holo forms of SAV, respectively. Studies were carried out at ~10 uM protein with 100 μM biotin at an excitation wavelength of 340 nm; all spectra are the average of three independent experiments.

**L110X Characterization.** For the L110X mutant, a 44% decrease in emission intensity with an associated 14 nm blue shift in emission maximum (449 nm to 435 nm; Table 1, Figure 3.2c) was observed with the addition of biotin. Although the absorption spectra for both the apo- and holo-forms of L110X are consistent with a mixed population of the neutral and anionic 7-HCAA species (Figure 3.2c inset), the addition of biotin increases the population of the neutral species by ~12%.

Crystal structures of the L110X mutant for both apo- and holo-forms were solved to 1.55 Å and 2.10 Å respectively. In the apo-form (Figure 3.3a), 7-HCAA occupies the biotin-binding pocket and the 7-HC side chain is in van der Waals contact with residues W120 and L124; the Cβ atoms of S88 and S112 pack against the opposite side of the coumarin ring. Clear electron density is observed for the biotin molecule in the holo structure (Figure 3.3b) and alignment of this structure with the 3ry2 structure from the PDB

indicates that the bound biotins are superimposable (Figure 3.4). The carboxylate of biotin packs directly against the pyrone ring of 7-HCAA in a planar orientation, which forces the 7-HC side chain deeper into the SAV binding pocket (Figure 3.3c). This slight (1.10 Å) movement of the coumarin moiety is accommodated by backbone rearrangements in the vicinity of residues L124 and W120.

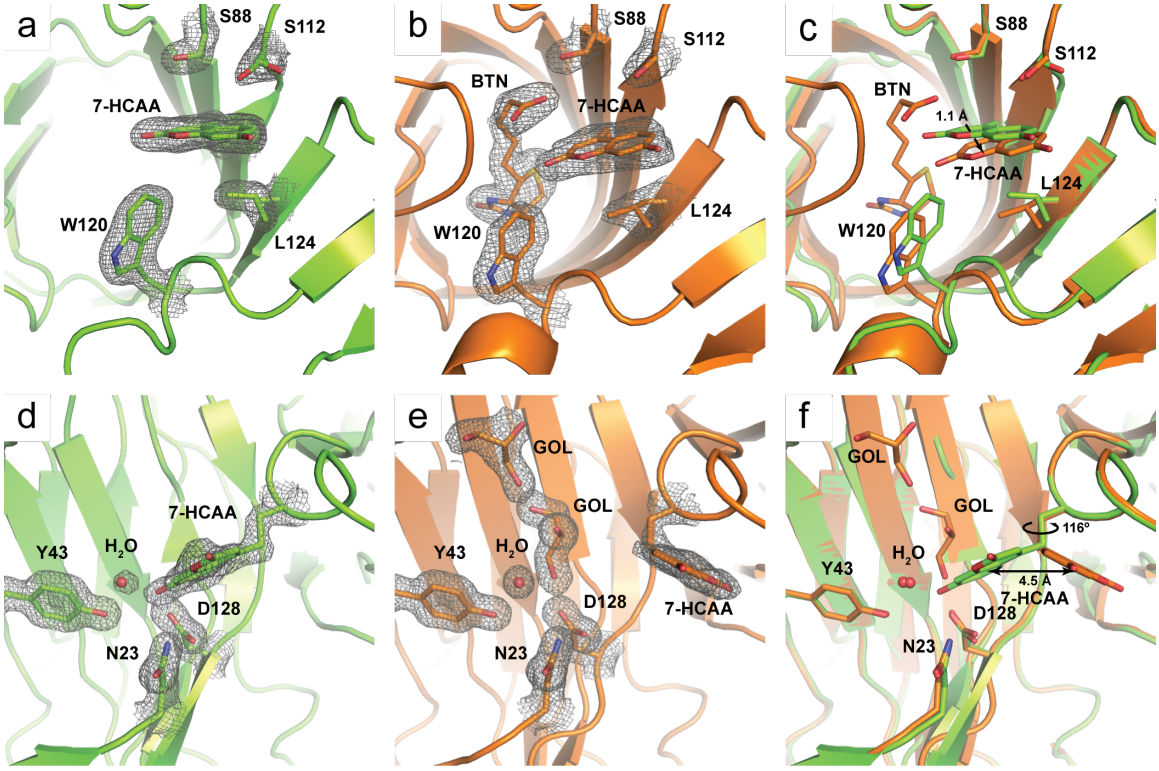



**Figure 3.4.** Structures of (a) L110X apo (PDB ID: 6udb), (b) L110X holo (PDB ID: 6udc), (c) an overlay of L110X apo (green) and holo (orange), (d) W120X apo (PDB ID: 6ud1), (e) W120X plus glycerol (PDB ID: 6ud6), and (f) an overlay of W120X apo (green) and holo (orange). The 7-HCAA residue, the biotin (teal) in L110X holo, the glycerol (teal) in W120X plus glycerol, and those residues surrounding and interacting with the 7-HCAA are shown as sticks. Electron density around each residue is shown as a $2F_O$-$F_C$ map contoured to 1 σ.

**Table 3.2.** Crystallographic statistics.

| Structure | L110X Apo | L110X Holo | S112X Holo | W120X Apo | W120X Glycerol |
|---|---|---|---|---|---|
| Data collection | | | | | |
| Space group | P 1 2 1 1 | P 1 2 1 | I 2 2 2 | I 4 | I 4 1 |
| Cell dimensions<br>a, b ,c (Å)<br>α, β, γ (deg) | 46.46, 85.56, 58.10<br>90.000, 99.0, 90.00 | 50.81, 98.28, 52.65<br>90.00, 112.3, 90.00 | 46.209, 93.110, 104.369<br>90.000, 90.000, 90.000 | 57.351, 57.351, 172.320<br>90.000, 90.000, 90.000 | 57.519, 57.519, 173.026<br>90.000 90.000 90.000 |
| Total Reflections | 614,572 | 340,473 | 956,126 | 956,126 | 2,652,590 |
| Unique Reflections | 64,975 | 27,711 | 40,032 | 40,032 | 44,646 |
| Resolution (Å) | 50.00-1.55<br>(1.58-1.55) | 50.00-2.10<br>(2.14-2.10) | 50.00-1.55<br>(1.58-1.55) | 50.00-1.55<br>(1.58-1.55) | 50.00-1.50<br>(1.53-1.50) |
| I/σ(I) | 10.9 (1.3) | 13.75 (1.9) | 32.83 (2.36) | 32.83 (2.36) | 36.00 (2.14) |
| $R_{meas}$ | 0.062 (0.743) | 0.160 (0.894) | 0.115 (1.047) | 0.115 (1.047) | 0.077 (1.105) |
| $R_{pim}$ | 0.030 (0.364) | 0.076 (0.442) | 0.036 (0.341) | 0.036 (0.341) | 0.021 (0.310) |
| $CC_{1/2}$ | (0.938) | .725 (0.582) | 0.999 (0.794) | 0.999 (0.794) | 1.000 (0.845) |
| $R_{merge}$ | -- | -- | -- | -- | -- |
| Completeness (%) | 99.6 (93.6) | 97.1 (92.6) | 98.75 (99.2) | 97.9 (99.9) | 100.0 (100.0) |
| Redundancy | 2.1 (2.0) | 2.2 (2.0) | 5.2 (4.6) | 5.2 (4.6) | 6.8 (6.4) |
| Refinement | | | | | |
| Resolution (Å) | 42.78-1.55 | 49.14-2.10 | 43.08-1.55 | 43.08-1.55 | 29.65-1.50 |
| $R_{work}$ | 0.166 | 0.194 | 0.174 | 0.173 | 0.126 |
| $R_{free}$ | 0.201 | 0.256 | 0.212 | 0.197 | 0.167 |
| Number of protein atoms | -- | -- | -- | -- | -- |
| Number of solvent atoms | -- | -- | -- | -- | -- |
| Average B factors, Å$^2$ | -- | -- | -- | -- | -- |
| Protein atoms, Å$^2$ | -- | -- | -- | -- | -- |
| Solvent atoms, Å$^2$ | -- | -- | -- | -- | -- |
| rmsd bond lengths (Å) | 0.0221 | 0.0161 | 0.0197 | 0.0245 | 0.0307 |
| rmsd bond angles (deg) | 2.147 | 1.919 | 2.051 | 2.546 | 2.437 |

*The values in parentheses indicate statistics for the highest resolution shell.

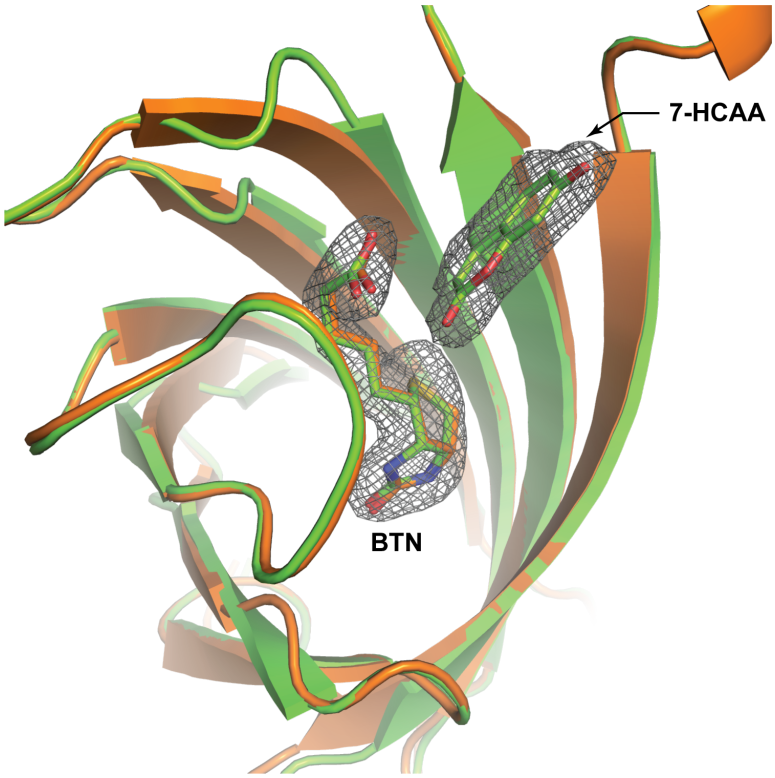



**Figure 3.5.** Overlay of L110X chain A holo (green) with 3RY2 chain A (orange) showing the placement of biotin in both structures. The RMSD of the biotin in the L110X crystal structure to the biotin in 3RY2 was 0.149 Å. Electron density around the L110X biotin shown as a $2F_O$-$F_C$ map contoured to 1 σ.

Given the similar chemical environments that are present in the apo- and holo-forms of L110X, it is likely that the close packing interactions between the biotin carboxylate and 7-HCAA contribute to the observed alterations to the fluorescent spectra. In an effort to explore this, we synthesized the biotin analogue, biotinamide, which replaces biotin's negatively charged carboxylate with a neutral acetamide group. A fluorescence spectrum of L110X in the presence of biotinamide showed a 42% increase in emission intensity and a 10 nm blue shift in emission maximum (Figure 3.6).

**S112X characterization.** Addition of biotin to S112X resulted in a 71% increase in fluorescence intensity (Figure 2d). It should also be noted that there is a slight increase in ~380 nm emission in the holo-form, which could arise from interactions that partially block ESPT (Figure 2d). An absorbance spectrum of the S112X mutant indicated that mixed populations of neutral and anionic 7-HCAA are present in both the apo- and holo-forms.

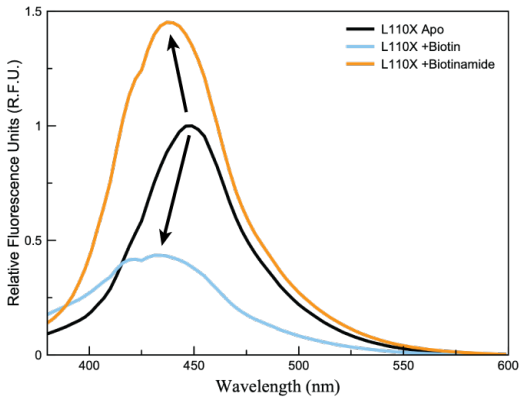



**Figure 3.6.** Normalized fluorescence spectroscopy for L110X in the apo (black), plus biotin (cyan) and plus biotinamide (orange) (~10 uM protein with 100 µM biotin/biotinamide) at 340 nm excitation. Insets: UV-Vis spectroscopy for each respective mutant in the apo- (black) and holo- (cyan) forms. Spectra are the average of three experiments.

Despite repeated attempts, we were unable to obtain crystals of the apo S112X mutant. However, a structure of the S112X mutant in the holo form was solved to 1.8 Å

resolution; both the biotin and 7-HCAA are fully resolved in this structure (Figure 3.6). Again, the position of biotin in this structure is consistent with that observed in the wild type structure (PDB ID: 3ry2).[53] In this structure, depending on the monomer, the 7-HCAA occupies two separate positions. In chain A, the phenol of the 7-HCAA points away from the biotin binding pocket and appears reasonably solvent exposed. In chain B, however, the phenol is pointed toward the biotin binding pocket of chain A and away from bulk solvent (Figure 3.6).

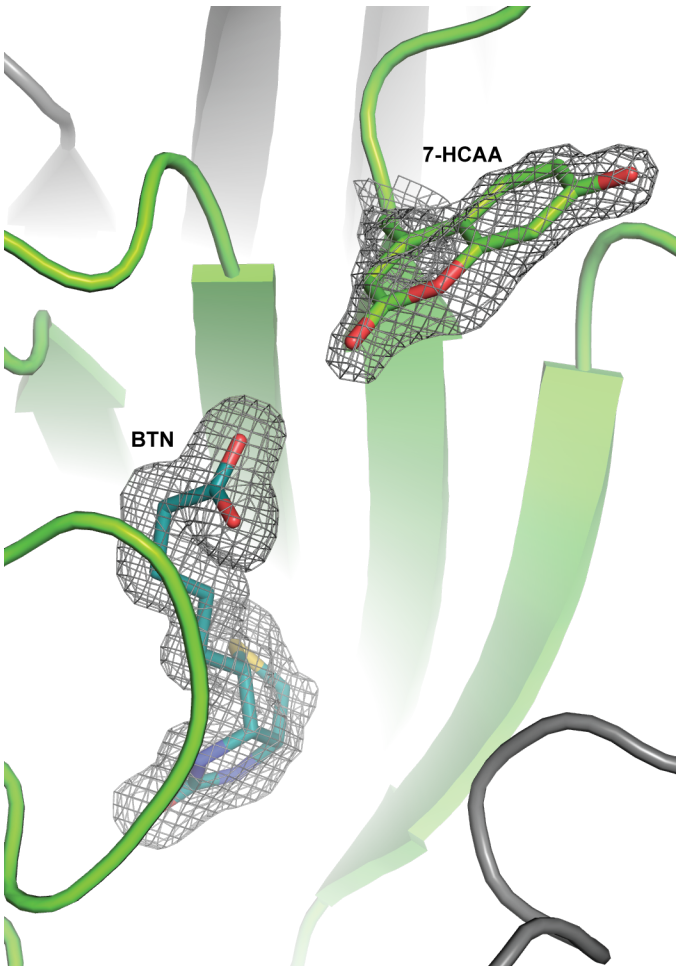



**Figure 3.7.** Crystal structure of S112X chain A (green) with biotin bound (teal). The other monomeric chains are colored in gray. Electron density around each residue is shown as a $2F_O$-$F_C$ map contoured to 1 σ.

Analysis of this structure in the context of the tetrameric SAV complex reveals a number of potential interactions that could give rise to the observed absorption spectra. For example, a population of 7-HCAA may for a hydrogen bond with K121 that may block ESPT to bulk solvent. It is possible that such interactions could exist in solution but not in the crystallized protein lattice due to slight structural rearrangements that occur during the crystallization process. Therefore, while our structural data does offer an explanation of the overall increase at 450 nm, explanations for certain aspects of the emission spectra are not generally apparent from the structure.

**W120X characterization.** In the presence of biotin, the emission intensity of W120X increases 135% (Figure 3.2e), which represents the largest biotin-dependent change in fluorescence observed in this study. Similar to the L110X mutant, a blue shift in emission maximum from 453 nm (apo-form) to 433 nm (holo-form) is observed.

We initially solved a structure of the apo-form of the W120X mutant to 1.17 Å resolution at pH 3.5. However, only weak electron density was observed for the 7-HCAA residue, which appeared to adopt multiple conformations. In one of the conformations, the 7-HCAA appeared to form hydrogen bond with D128; the absorbance spectra of this mutant (collected at pH 7.0) are consistent with the presence of this interaction (Figure 3.2e, inset). We rationalized that the crystallization conditions (pH 3.5) likely produced a significant population of protonated D128 that would preclude this hydrogen bonding interaction. Thus, crystals obtained at pH 3.5 were soaked in buffers with increasing pH. We ultimately collected diffraction data to 1.55 Å resolution from a crystal that had been soaked to pH 5.5; electron density for 7-HCAA is clearly visible in this structure (Figure 3.3d). Notably, an anomalously short hydrogen bond (< 2.4 Å) is observed between the

phenolic oxygen of 7-HCAA and the carboxylate of residue D128 (Figure 3.7); this hydrogen bond is apparent in both chains of the asymmetric unit. Additionally, an extensive hydrogen bonding network (Figure 3.7) is observed in the region around the phenol of the coumarin. that includes an ordered water molecule, and the phenol of tyrosine 43.

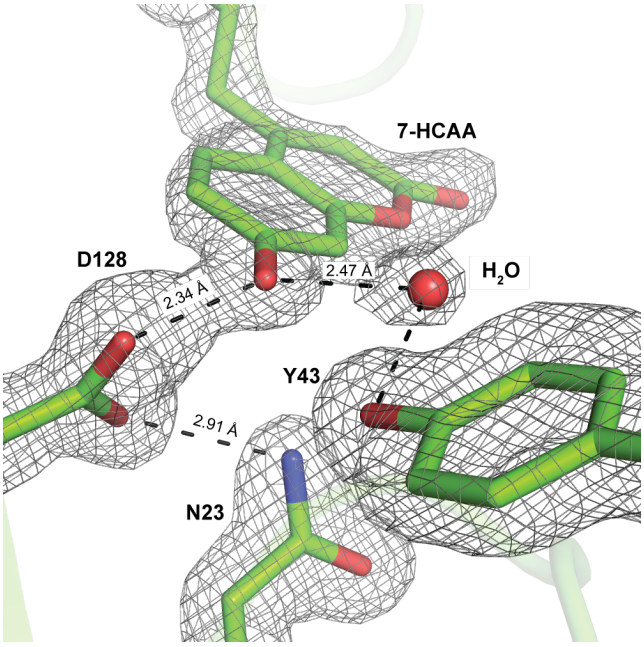



**Figure 3.8.** Zoomed in view of the crystal structure of the W120X apo 7-HCAA hydrogen bonding network. Distances are the averages of the two monomeric streptavidin chains in the asymmetric unit. The short hydrogen bond distances shown here may be caused by uncertainty in the position of the 7-HCAA. Electron density around each residue is shown as a $2F_O$-$F_C$ map contoured to 1 σ.

Several attempts at either co-crystallization of W120X with biotin or soaking W120X apo-form crystals with biotin were unsuccessful. However, we noted that a previously solved structure of the W120X mutant (1.50 Å, pH 5.5) contained two glycerol molecules, which had been used as a cryoprotectant, were present in the biotin binding pocket (Figure 3.3e). Importantly, the glycerol molecules served to displace the 7-HCAA from the binding pocket. Namely, a rotation of 116° about $\chi_3$ of the 7-HCAA resulted in a 4.9 Å displacement of 7-HCAA from its buried position in the apo protein (Figure 3.3f). A spectrum of the W120X in the presence of 50% glycerol showed an 11% increase in

fluorescence at 450 nm. When overlaid with the 3ry2 structure, the bound glycerol molecules directly superimpose with a number of atoms in the biotin molecule (Figure 3.9). Notably, loop3-4, which closes over the binding site when biotin binds is found in the open position in the glycerol bound structure.

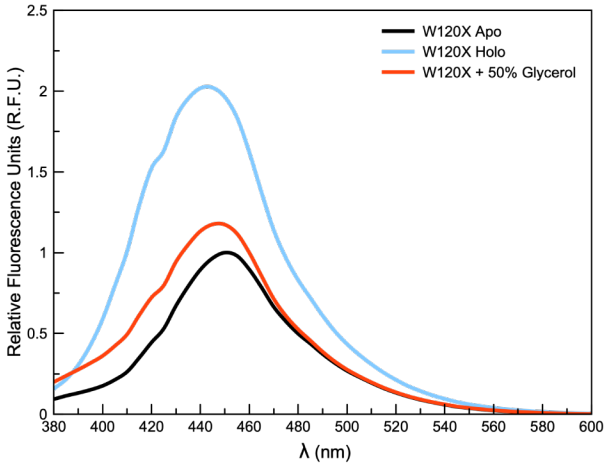



**Figure 3.9.** Normalized fluorescence spectroscopy for W120X in the apo- (black) and holo- (cyan) forms (~10 uM protein with 50% glycerol) at 340 nm excitation.

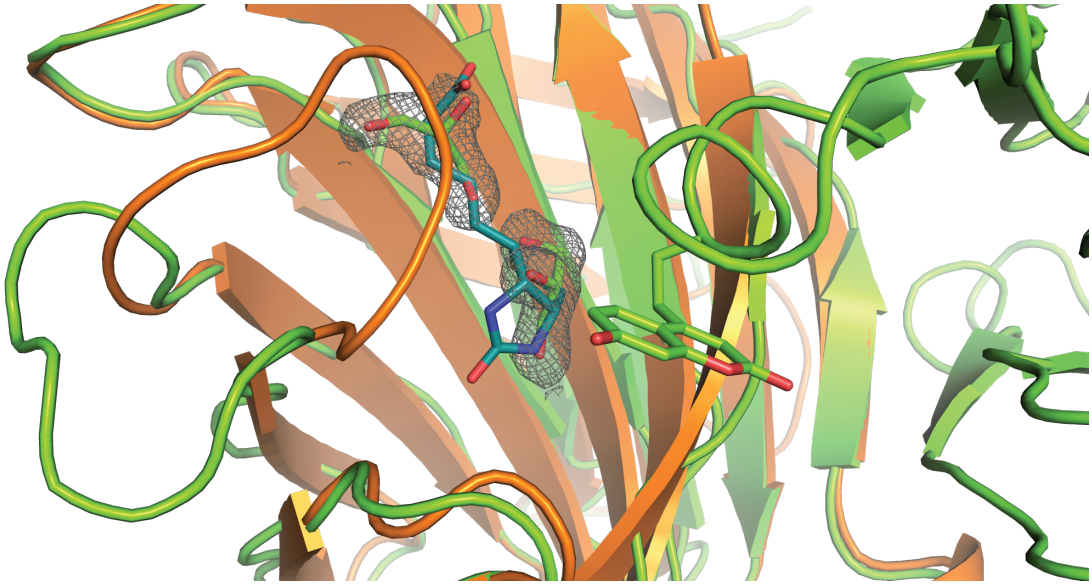


**Figure 3.10.** Crystal structure of W120X plus glycerol (green) overlaid on PDB 3RY2 (orange) showing the location of the glycerol molecules in relation to the known biotin (teal) location. Also shown is loop3-4 in the open position in the W120X plus glycerol + glycerol fluorescence spectra. Additionally, the biotin itself may be contributing to these differences as well. Electron density around each residue is shown as a $2F_O$-$F_C$ map contoured to 1 σ.

**L124X characterization.** Both the absorbance and emission spectra of the L124X mutant exhibit features distinct from those observed in all other variants. The emission spectrum of L124X contains two major peaks at 380 nm and 450 nm (Figure 3.2f). Although the predominant emission maximum is observed at 450 nm, emission at 380 nm suggests that ESPT from 7-HCAA has been partially blocked in this mutant. Although increases in emission intensity are observed at both 380 nm and 450 nm upon biotin binding, the magnitude of this change in emission is substantially smaller than those observed for all other variants. The absorbance spectra of both the apo- and holo-forms of L124X are also distinct from other mutants in that they are essentially linear from 300 nm to 450 nm (Figure 3.2f inset). Due to low protein expression yields and the formation of protein aggregates in solution, characterization of this protein was limited to spectroscopic analysis alone**.**

**Affinity of the streptavidin mutants for biotin.** Although structural characterization of the biotin-bound forms for three of the five mutants indicated that the introduction of 7-HCAA had not disrupted the wild type biotin binding interactions, it seemed likely that the fNCAA could have altered the affinity of the mutant proteins for biotin. To test this, we measured the apparent $K_d$ for each of the stable SAV mutants (L110X, S112X, and W120X) using the changes in the fluorescence intensity of 7-HCAA. Unfortunately, the red shifts of L110X and W120X make it difficult to identify a linear trend for their affinity for biotin. However, by comparing the ratio of emission at 450 nm (the emission maximum of the apo-form) to emission at 435 nm (the emission maximum of the holo-form), we can derive apparent $K_d$s for biotin binding (Figure 3.10). A non-linear

regression fit of these data suggests an apparent $K_d$ of 568.6 nM for L110X, 1.6 μM for S112X and 6.2 nM for W120X (statistics for the fits are provided in Table 3.3). These apparent dissociation constants suggest that the introduction of the fluorophore does result in a substantial loss in affinity relative to the wild-type protein. However, the fact that dissociation constants in the nM range are still observed suggests that strong binding interactions can remain even with the addition of a fNCAA in proximity to the binding site.

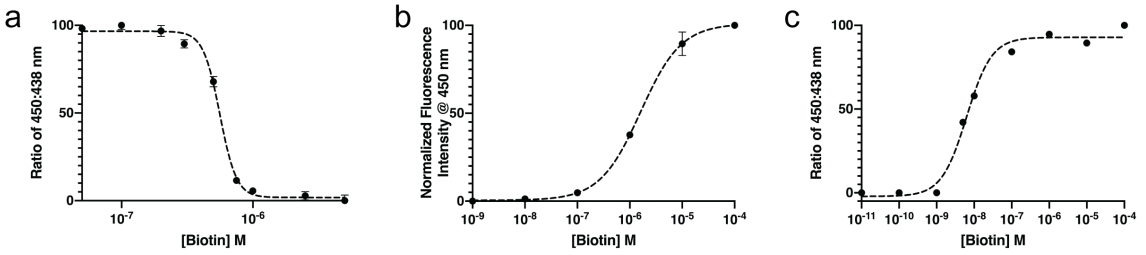



**Figure 3.11.** Binding analysis for the (a) L110X, (b) S112X, and (c) W120X SAV mutants. For the L110X and W120X mutants, the affinity for biotin was determined by using the ratio of 450 nm to 438 nm emission. For S112X, the single emission wavelength of 450 nm was used. Some error bars too small to be visible were omitted.

**Table 3.3** Parameters and statistics for L110X, S112X, and W120X apparent $K_d$ curve fits.

| | **L110X** | **S112X** | **W120X** |
|---|---|---|---|
| **Best-fit values** | | | |
| **Botom** | 1.83 RFU | 0.4543 | -2.149 |
| **Top** | 96.62 RFU | 101 | 92.86 |
| ***K<sub>d</sub>*** | 560.3 nM | 1.61 μM | 6.28 nM |
| **HillSlope** | -6.857 | 1.12 | 1.301 |
| **log *K<sub>d</sub>*** | -6.252 | -5.794 | -8.202 |
| **Span** | 94.79 | 100.5 | 95.01 |
| **95% CI (profile likelihood)** | | | |
| **Botom** | -0.8138 to 4.371 RFU | -1.734 to 2.576 RFU | -12.81 to 7.383 RFU |
| **Top** | 95.11 to 98.96 RFU | 97.36 to 104.9 RFU | 85.12 to 101.4 RFU |
| ***K<sub>d</sub>*** | 543.0 to 576.8 nM | 1.38 to 1.90 μM | 4.03 to 1.06 nM |
| **HillSlope** | N/D | 0.9523 to 1.358 | 0.6759 to 2.567 |
| **log *K<sub>d</sub>*** | -6.265 to -6.239 | -5.861 to -5.722 | -8.394 to -7.976 |
| **Goodness of Fit** | | | |
| **Degrees of Freedom** | 26 | 14 | 5 |
| **R squared** | 0.9947 | 0.9969 | 0.9891 |
| **Sum of Squares** | 303 | 95.97 | 161.3 |
| **Sy.x** | 3.414 | 2.618 | 5.68 |
| **Number of points** | | | |
| **# of X values** | 30 | 18 | 9 |
| **# Y values analyzed** | 30 | 18 | 9 |

### 3.3 Discussion

The primary goal of this study was to gain a deeper understanding of how chemical environments surrounding the coumarin-containing fNCAA 7-HCAA could alter its photochemical properties. To do this, 7-HCAA was substituted for five native amino acids in streptavidin, which were then characterized spectroscopically and, when possible, structurally. Four of the five 7-HCAA-containing SAV mutants generated in this study exhibited substantial biotin-dependent changes to their fluorescence spectra. Included among the responses were both increases and decreases in fluorescence emission intensity as well as both red- and blue-shifted emission maxima. Structural analysis of three of the mutants provided insight into the atomic-level interactions that may be responsible for the observed changes in fluorescence while the remaining two proteins, L25X and L124X, proved recalcitrant to structural analysis in both the apo and holo forms. Nonetheless, the structural data collected for the L110X, S112X and W120X mutants suggest that both residues in SAV and the biotin itself can interact with 7-HCAA in ways that alter its fluorescent properties.

In the absence of structural data, it is difficult to confidently predict the orientations adopted by the 7-HCAA in either the apo- or holo-forms. That said, when considered in combination with structural characterization of wild type SAV, spectroscopic data can provide clues to the environments experienced by the fNCAA in the apo- and holo-forms. In the L25X mutant, the Cα—Cβ vector of the native leucine residue points in the direction of the biotin binding pocket. It is possible that the 7-HCAA adopts a rotamer that places the coumarin side chain within this pocket in the apo protein. However, because the

absorbance spectra of both the apo- and holo-form of the 7-HCAA shows a primarily neutral molecule (compared with the small molecule 7-HCAA, compare Figure 3.3a inset with Figure 3.3b inset) we can be confident that the molecule is interacting with the protein in some fashion. Also, because the absorbance spectrum of L25X doesn't change on biotin binding (Figure 3.3b, inset) it is likely not a change in the position of the 7-HCAA that causes the fluorescence increase. Namely, if a more buried 7-HCAA orientation were adopted by this mutant in the apo-form, a change in its absorbance would likely occur upon biotin binding as a consequence of displacement from its original environment by biotin or an alteration of its chemical environment by functional groups present on the substrate. Without structural data, it is difficult to rationalize the 40% increase in fluorescence intensity that is observed in this mutant on biotin binding. However, given the proximity of this residue to the bound substrate, it is possible that slight alterations in 7-HCAA position may occur as a consequence of the introduction of biotin, which in turn result in the observed increase in fluorescence.

Although structural characterization of the L124X mutant was precluded by the protein's instability, the featureless absorbance spectra observed for both the apo- and holo-forms suggest that a 7-HCAA photodimer may have been formed. Structural analysis of the SAV tetramer indicates that the L124 residue in one subunit is only 4.3 Å away from its symmetry mate in the adjacent subunit. Upon substitution of L124 to 7-HCAA, it seems possible that the planar 7-hydroxycoumarin moieties could directly interact with one another through π-stacking interactions that would place the 7-HC side chains in van der Waals contact. Importantly, the nearly linear absorbance spectra obtained for the L124X mutant in the apo- and holo-forms are consistent with those in previous reports of coumarin

dimers.[55] The fluorescence spectra of the apo and bound states also appear to be consistent with a photoinduced dimer of 7-HCAA. Namely, for the dimerization reaction to occur, both 7-HCAA side chains would need to adopt orientations that were not buried within the biotin binding pocket. This feature would in turn limit the chances of the fNCAA to be responsive to the biotin binding event. It is unfortunate that the weak expression of the L124X mutant precluded structural analysis, which would have provided more insight into the nature of this intriguing mutant.

For the S112X mutant, the increase in fluorescence with little change in emission max (450 nm to 455 nm) is consistent with other reports of 7-HCAA fluorescence changes.[19] It is likely that the observed increase in fluorescence arises from the disruption of interactions that are present in the apo form, but which are dissolved due to a structural rearrangement that occurs upon biotin binding. However, without structural information little can be said about the apo-form interactions that may be causing the quenching phenomenon. It should also be noted that there is a slight increase in ~380 emission in the holo-form that is consistent with a reduction in the ability of 7-HC to perform ESPT. This change in the population of 7-HCAA that can undergo ESPT may be due to the close proximity of the biotin carboxylate. Namely the increased negative charge in that area may alter the $pK_a$ of the excited state such that a small population of 7-HCAA does not undergo ESPT.

Unlike the other three mutants, both L110X and W120X show large changes in fluorescence and significant blue shifts in the emission maxima upon addition of biotin. To the best of our knowledge, the observation of a 7-HC species that emits near 435 nm has not been previously described in the literature. Given the structural data we collected for

each of these mutants, we rationalized that the electronic changes in the environment directly around the 7-HCAA give rise to the observed data. Namely, the carboxylate of the biotin in the case of L110X and the backbone carbonyls (Glu116 and Ala117) in the case of W120X. This aligns well with the work of Callis and Burgess[56] where they showed that changes in electrostatic potential along the $L_a$ axis of the indole ring of tryptophan was responsible for changes in the tryptophan fluorescence. In that work, they suggested that this was due to an internal Stark Effect associated with the specific protein environment surrounding the tryptophan.

In the case of the 7-HC, the transition state dipole axis runs from the $C\eta_2$ atom of the phenol ring to the lactone oxygen (Figure 3.11); directly where the biotin carboxylate is resting in the solved L110X holo-structure (Figure 3.3c). Because biotinamide didn't affect the 435 nm emission, it can be assumed that having a polar side chain along the transition dipole moment would cause a similar effect. Given that only biotin induced quenching and that the biotin carboxylate is analogous to both aspartic and glutamic acid, it is reasonable to believe that this interaction can be designed into other protein sensors and used to monitor changes by both quenching the 7-HC moiety and ratiometrically measuring 435/450 nm fluorescence.

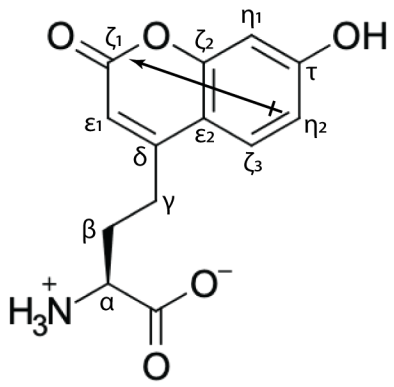



**Figure 3.12.** Transition dipole moment of 7-HCAA showing the direction of change for the electron density from the ground state to the fluorescence emitting excited state of 7-HCAA.

For the W120X mutant, we attribute quenching to the particularly short hydrogen bond between the 7-HCAA phenol and D128 in the adjacent SAV monomer (Figure 3.9). This short hydrogen bond is reminiscent of other short hydrogen bonds formed in proteins between acidic residues and conjugated phenols.[57,58] It has previously been suggested that a short hydrogen bond between an acidic residue and the 7-HCAA could be responsible for quenching via backward transfer of a proton.[20,32,58–62] While we were unable to obtain a biotin bound structure for the W120X mutant, we can infer several potential reasons why we see both an increase in fluorescence and the shift to 435 nm from the glycerol bound structure. Firstly, the increase in fluorescence can be attributed to moving the 7-HC moiety out to solvent (Figure 3.3e, f) similar to the increase in the S112X mutant. Secondly, several carbonyl oxygens (residues 116 and 117) and amide groups (residues 119 and 120) from the protein backbone are within close distance to the 7-HCAA, again invoking a Stark Effect for the change in emission max.

### 3.4 Conclusion

Despite a wealth of fluorescent tools that have proven their use in the study biological systems, the ability to develop fluorescent sensors of protein-small molecule

interactions remains a challenge. Many current approaches for achieving this rely on FRET-based approaches that require at least two site-specific fluorescent labels and are often best suited to study proteins in which large conformational changes accompany the binding event.[5,14,17,37] Fluorescent non-canonical amino acids possess a number of properties (e.g. small sizes and an increased versatility with respect to sites of incorporation) that suggest their potential for use in these studies. However, a lack of characterization of these fluorophores in proteins likely limits our ability to routinely use these tools in this way.

The results reported in this study provide an atomic level understanding of the manner in which protein environments respond to the incorporation of the versatile fNCAA 7-HCAA and the manner in which changes in local environments affect the fluorescent properties of this fluorescent amino acid. The spectroscopic and structural characterizations reported herein could be used in future efforts aimed at rationally designing new fluorescent sensors of protein-ligand interactions or will, at a minimum, provide insight into sites of fNCAA incorporation that might lead to new protein-based fluorescent of metabolite binding.

### 3.5 Materials and Methods

**Protein Expression & Purification.** Wild-type core streptavidin and DNA oligos for making each Amber stop codon mutation were ordered from Integrated DNA Technologies, Inc. Wild-type core streptavidin was cloned into a pET29b vector using Gibson assembly. Mutations were made via overlap extension PCR and Gibson assembly. BL21 Star (DE3) cells were transformed with both a pEVOL-CouRS plasmid containing a chloramphenicol resistance marker and two copies of the evolved CouRS synthetase and a

pET29 plasmid containing a kanamycin resistance marker and the mutant streptavidins. The cells were incubated overnight in 2xYT media at 37 ºC, 250 rpm to $OD_{600}$ ~5.0, centrifuged for 10 minutes at 4200x g, and resuspended in fresh 2xYT media. Arabinose (0.2% w/v final) and the non-canonical amino acid, 7-HCAA (1 mM final), were added to each culture and incubated for 1 hour at 37 ºC 250 rpm in order to express the CouRS synthetase. Induction of the streptavidin was performed by adding isopropyl-β-D-thiogalactoside (1 mM final). The cultures were then incubated between 8 and 16 hours at 30 ºC, 180 rpm. Cells were harvested via centrifugation and lysed in cell lysis buffer (25 mM Tris-HCl, pH 8.0, 0.1% Triton-X 100, 3 mM β-mercaptoethanol). Inclusion bodies were collected via centrifugation at 20,000 xg for 30 minutes and washed 3-4 times with wash buffer (25 mM Tris-HCl, 150 mM NaCl, 0.1% Triton-X 100) followed by solubilization in 6M Guanidine-HCl. Solubilized protein was refolded in refolding buffer (25 mM Tris-HCl, 20 mM imidazole, 500 mM NaCl) by adding the solubilized protein dropwise to the fastest part of a rapidly spinning solution of refolding buffer. Protein was then purified via a C-terminus 6x His-tag on a nickel-NTA resin (Wash Buffer: 25 mM Tris-HCl pH 8.0, 20 mM imidazole, 500 mM NaCl; Elution Buffer: 25 mM Tris-HCl pH 8.0, 500 mM imidazole, 150 mM NaCl), followed by anion exchange (Wash Buffer: 25 mM Tris-HCl pH 8.0, 10 mM NaCl; Elution Buffer: 25 mM Tris-HCl pH 8.0, 500 mM NaCl) and finally size-exclusion chromatography (25 mM Tris-HCl pH 7.0, 150 mM NaCl).

**SDS-PAGE.** 5 mL cultures for each mutant were grown overnight in 2x YT media at 37 ºC and 250 rpm in the presence of 50 μg/mL (final) kanamycin and 34 μg/mL (final) chloramphenicol. The cultures were then centrifuged at 4,200 rpm and the media was

decanted. Each pellet was resuspended in 5 mL of fresh 2x YT media containing 0.2% arabinose (final), 50 µg/mL (final) kanamycin and 34 µg/mL (final) chloramphenicol. The cultures were split into two separate cultures and 1 mM (final) 7-HCAA was added to only one culture for each mutant. The cultures were then incubated at 37 ºC and 250 rpm shaking for 1 hour. 1mM (final) IPTG was added and the cultures were incubated at 30 ºC and 225 rpm for 16 hours. 15 µL of each culture and 1 µL of 1mg/mL (1 µg final) Streptavidin-wt (Sigma) was mixed with 15 µL 4x Laemmli loading buffer and incubated at 95 ºC for 10 minutes followed by 25 ºC for 5 minutes. 15 µL of sample was loaded into each well of a 4% stacking / 15% resolving polyacrylamide gel and run at 125 V for 90 minutes.

**Western Blotting.** An acrylamide gel was equilibrated in 25 mL of transfer buffer (25 mM Tris, pH 8.3, 192 mM glycine, 20% methanol, 0.1% SDS) for 60 minutes. Protein from the acrylamide gel was transferred to a nitrocellulose membrane using a semi-dry transfer apparatus (Bio-Rad) until the pre-stained ladder was completely transferred to the membrane (30 minutes at 25 V). The membrane was removed from the apparatus and incubated with a TBS-T blocking solution (25 mM Tris, pH 8.0, 150 mM NaCl, 0.1% v/v Tween-20, 5% w/v powdered nonfat milk) for 60 minutes. The membrane was then washed 3 times for 5 minutes each in 25 mL of TBS-T (25 mM Tris, pH 8.0, 150 mM NaCl, 0.1% v/v Tween-20) on an orbital shaker. The membrane was then incubated in TBS-T with HRP-conjugated anti-streptavidin antibodies (1:10000 dilution) at room temperature for 1 hour on an orbital shaker. The membrane was then washed with 25 mL TBS-T 3 times for 5 minutes each on an orbital shaker. A 1x solution of DAB solution (Pierce) was created immediately before use by adding 2.5 mL of 10x DAB with 22.5 mL of Stable Peroxide

Substrate Buffer. Finally, the membranes were incubated for 30 minutes in the 1x DAB solution, air-dried, and imaged.

**Spectroscopic Analysis**. All spectroscopic experiments were performed in a Tris-Buffered Saline (TBS) solution (25 mM Tris-HCl pH 7.0, 150 mM NaCl) using a 1 cm quartz cuvette. Each apo mutant was concentrated to 100 µL and then diluted to a final absorbance of 0.05 at 340 nm. The samples were then split and either biotin (100 µM final concentration, dissolved in 30% DMSO) or TBS was added to each. A SpectraMax M5 spectrophotometer was used to take absorbance spectra from 200 nm to 750 nm at 1 nm intervals in triplicate. A Horiba Nanolog fluorimeter was used to measure relative fluorescence intensities. Both the excitation and emission slit widths were set to 2 nm and the spectra were taken from 365 nm to 625 nm in triplicate with excitation at 340 nm.

**Crystallization.** A CrystalMation Phoenix (Rigaku) was used to screen a 10 mg/mL solution (10 mM HEPES, pH 5.5, 75 mM NaCl) of both apo and holo L110X, S112X, and W120X against three sitting drop vapor diffusion crystal screening libraries (Hampton Research, Crystal HT, Index HT, and PEG/Ion HT). Each screen contains 96 conditions and each condition was tested twice (v/v ratios of protein to reservoir drop) for a total of 576 total conditions per protein in 200 and 300 nl drop sizes. Conditions that produced crystals after 5 days were then recapitulated using larger volume sitting drop vapor diffusion. Streptavidin L110X apo and holo crystals were grown with a reservoir solution of 0.1 M Bis-Tris, pH 6.5, 25% w/v polyethylene glycol 3350. Streptavidin S112X holo, and W120X apo crystals were grown with a reservoir solution of 0.1M citric acid, pH 3.5, 3.0 M NaCl. Each drop contained 2 µL of protein solution mixed with 2 µL of reservoir solution, crystals were grown until no new growth was visible (approximately 1

week). Both the S112X and W120X crystals were soaked in a new reservoir solution of 0.1M citric acid, pH 5.5, 3.0 M NaCl. Three 2 μL drop exchanges were performed over the course of 12 hours and the crystals were allowed to soak overnight. The cryoprotectant was 25% PEG for L110X, S112X and the apo crystal for W120X. 25% glycerol was used as cryoprotectant for the W120X glycerol bound crystal.

**Data Collection and Structure Determination.** Diffraction data from the L110X apo and W120X apo crystals were collected at the Berkeley Center for Structural Biology (BCSB) from the Advanced Light Source (beamline 5.0.2) on a Dectris Pilatus3 6M detector. Data from the L110X holo and W120X glycerol bound crystals were collected at the Argonne National Laboratory Advanced Photon Source (beamline 19-ID) on a Dectris Pilatus 6M detector. Diffraction data from the S112X holo crystals were collected at the Stanford Synchrotron Radiation LightSource (beamline BL9-2) on a Dectris Pilatus 6M detector.

Crystals were flash frozen in liquid nitrogen prior to data collection at 100 K. Data were indexed, refined, integrated, and scaled using the HKL-3000 software package. Both structures were solved by molecular replacement using Phaser19 using an all glycine model of streptavidin with loops removed as the search model (PDB ID: 1swt). All models were refined using Refmac520 and model building was carried out with the program Coot. The chemical description of 7-HCAA was taken from Henderson et al.[63] All structural figures were made with the PyMOL molecular graphics software.[48] Structures and all supporting data have been deposited in the protein databank; PDB IDs: 6uc3, 6ud1, 6ud6, 6udb, 6udc.

# CHAPTER 4.

# COMPUTATIONAL DESIGN OF A FLUORESCENT BIOSENSOR USING A GENETICALLY ENCODED NON-CANONICAL AMINO ACID

## 4.1 Introduction

The results of the structural studies presented in chapter 3 and another published structural study performed in our lab by Henderson et al.[20] suggest that the fluorescence of 7-HCAA can be directly affected by the side chains of canonical amino acids. In the Henderson et al. study, an X-ray crystal structure (PDB ID: 6bjz) of a 7-HCAA-containing fragmented antibody (FAB; 5c8) that exhibits an increase in fluorescence when it binds its antigen CD40L (a protein expressed on the surface of T-cells) was solved.[19,20] These structural data highlighted hydrogen bonding and hydrophobic packing interactions between residues in 5c8 and 7-HCAA that are disrupted upon antigen binding.

Interestingly, both my work and the Henderson et al. study contained a structure where a putative short hydrogen bond (Figure 3.8) is implicated in the quenching mechanism. Although, at the onset of this study, it was unclear as to whether the Glu in the FAB acted simply as a positioning handle to help stabilize the 7-HCAA's orientation or if it also contributed to quenching. It was later suggested in the final document by Henderson et al.[20] that the short distance between the phenol of the 7-HCAA and the carboxylic acid of the Glu could allow for rapid proton transfer back to the 7-HCAA after excited state proton transfer occurred (ESPT) resulting in thermal relaxation of the excited-state rather than the emission of a photon; consistent with other reports of photoacids.[32]

It was also thought that π-stacking between a tryptophan and the 7-HCAA in the 5c8 structure may also be contributing to the quenching in that construct. The possibility

that Trp can act as a quencher of various organic dyes has been previously studied[64]. To see if Trp has a similar quenching influence on 7-HCAA, I performed a Stern-Volmer analysis where the fluorescence of the 7-HCAA is measured in relation to varying amounts of tryptophan. A Stern-Volmer analysis will indicate whether short-distance (collisional) interactions between a fluorophore and another molecule will produce quenching of the fluorophore.[65] By adding increasing the amount of potential quencher, one should see a decrease in fluorescence that correlates linearly with the concentration of potential quencher. My analysis did, in fact, show a concentration-dependent quenching of the 7-HCAA by Trp (Figure 4.1). All of these data raise the intriguing possibility that 7-HCAA could serve as a reporter of protein function if a similar local environment consisting of a short hydrogen bond and π-stacked tryptophan could be engineered into the protein of choice.

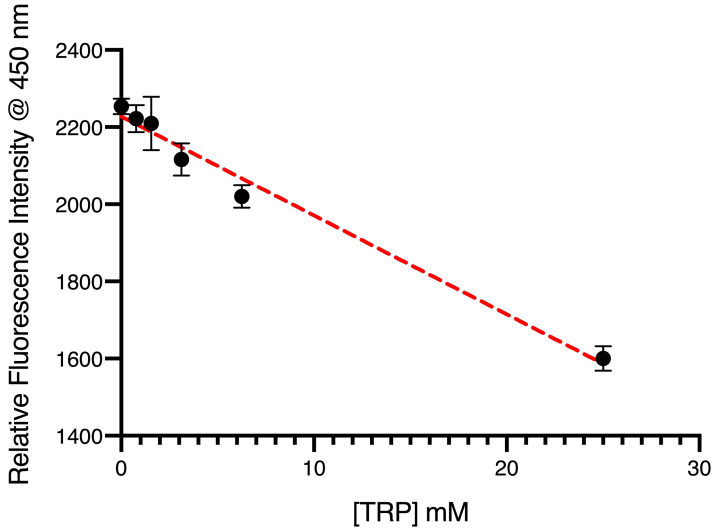



**Figure 4.1.** Stern-Volmer analysis of 7-HCAA with varying concentrations of tryptophan. This analysis indicates that tryptophan is an effective quencher of 7-HCAA.

In this chapter, I use the Rosetta macromolecular design suite (hereafter known as Rosetta) to engineer a fluorescent protein-based sensor of calcium binding using 7-HCAA.

Five designs (CaM-d1, CaM-d2, CaM-d3, CaM-d4, and CaM-d5) were selected for protein expression, of which, two (CaM-d1 and CaM-d5) showed increases in fluorescence in the presence of calcium. Only one design (CaM-d5) showed characteristic circular dichroism (CD) spectra consistent with biologically active calmodulin. Additionally, spectroscopic, fluorescence lifetime and CD data for a set of knockout mutants for components of the CaM-d5 theosite support the hypothesis that the designed interactions are responsible for the reported changes in 7-HCAA fluorescence.

**4.2 Results and Discussion**

**Design Strategy.** To accurately place amino acid sides chains in relation to each other in Rosetta, I first had to develop what is known as a "theosite" or rather a set of geometrical measures (e.g., distances, angles and torsion angles between atoms) that define the proximity of the amino acid side chains in relation to each other. Specifically, for this study, we were interested in incorporating the aforementioned Trp/7-HCAA π-stacking interaction and a short hydrogen bond between the 7-HCAA phenol and an acidic residue (either glutamic or aspartic acid). I, therefore, developed an idealized geometric arrangement of the 7-HCAA, Trp, and Glu that could be computationally designed into an existing protein structure using the Rosetta macromolecular software suite (Figure 4.2).[66] The theosite is comprised of two components: the indole ring of Trp in a π-stacking interaction with the coumarin moiety of 7-HCAA (both perpendicular and parallel interactions were allowed) and the carboxylic acid of a Glu residue (or aspartic acid, Asp, D) within 2.5 Å +/- 0.25 Å hydrogen bonding distance of the 7-HCAA phenol (Figure 4.2).

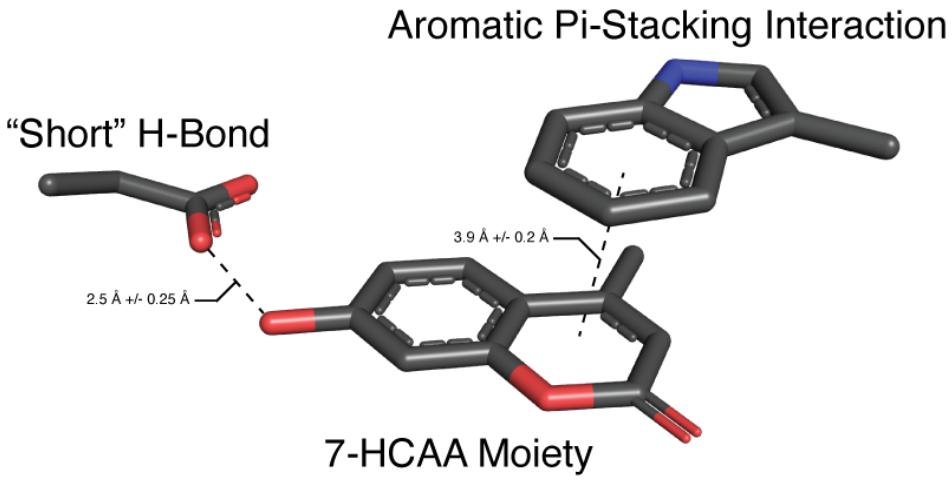



**Figure 4.2.** Idealized geometric constraints used for RosettaMatch algorithm.

Many times computational protein design will use libraries of protein X-ray crystal structures as "scaffolds" to identify protein backbones that are amenable to the theosite geometries.[67] In this case, however, we wanted to specifically design the theosite into an arbitrarily large conformational change on a single protein. Such a conformational change could then be exploited to move components of the theosite farther away or closer to each other upon ligand binding/removal. The model protein calmodulin (CaM) undergoes such a conformational change upon calcium-binding.

In the calcium unbound state (apo), CaM is a highly dynamic protein that comprises two domains (N-terminal and C-terminal) that rotate in relation to each other around a flexible hinge region; in the calcium bound (holo) state, the hinge region adopts a rigid alpha-helical structure. One disadvantage of using CaM for design studies is that the dynamic nature of the apo state has yet to be crystallized. As a consequence, there are no apo CaM X-ray structures in the protein databank (PDB) to use for design purposes. This also means that several low energy conformations that could exist for the apo state; making it difficult to favor any one state over another. We would hope that any chosen design

preferences one of the states over all others to ensure that equilibrium favors a quenched 7-HCAA. To overcome these challenges, we identified a CaM NMR structure (PDB ID: 1cfc) that contained 25 different structures or "states". Using the molecular viewing software PyMOL,[48] we separated these states into individual files and created a structure library (Figure 4.3).

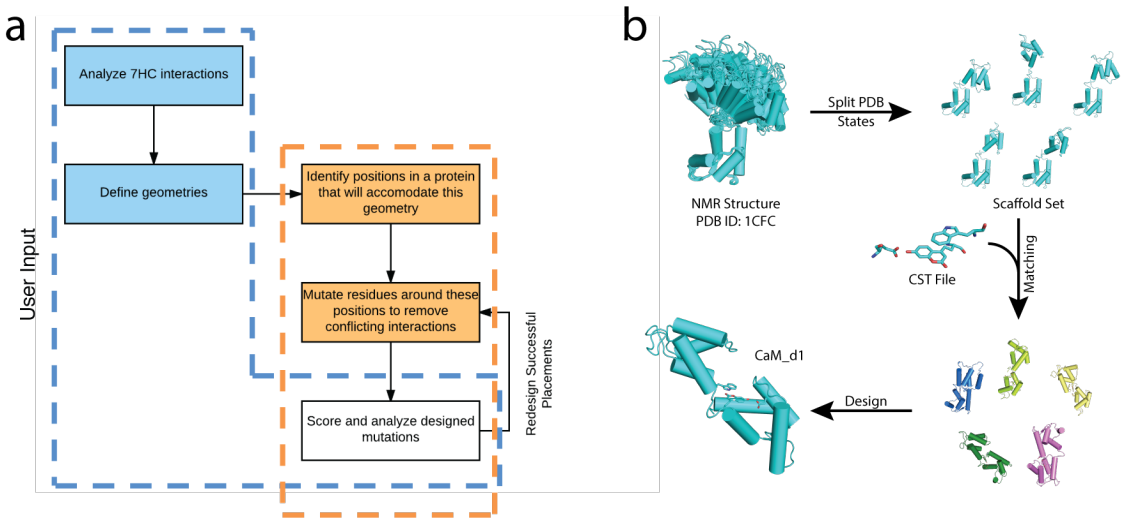



**Figure 4.3.** (a) Schematic of overall computational design process and (b) process specific to the computational design of a calmodulin-based biosensor using an NMR structure (PDB ID: 1cfc).

RosettaMatch[68] was then used to identify which CaM states allowed the incorporation of the theosite without disrupting the backbone. Each "match" was then processed with RosettaDesign[66,67] to introduce stabilizing mutations to the 7-HCAA, Trp, or Glu. Designs were analyzed and filtered on how well the geometric constraints were satisfied in that particular design and five designs—CaM-d1-d5 (Table 4.1)—were selected for expression. We also used RosettaDesign to model each of the designed residues into the holo state (PDB ID: 1cll).

**Table 4.1.** CaM Design Mutations.

| Construct | Mutations |
|---|---|
| CaM-d1 | M37E, M77X, S82W, L40T, Q42T, M73A, and K76I |
| CaM-d2 | S82X, I86V, E115W, L117T, and M146E |
| CaM-d3 | L5X, Q9W, F13E, M73A, and M77A. |
| CaM-d4 | S82X, I86A, M110T, E115W, L117E, and M146E. |
| CaM-d5 | Q9E, S82T, E84L, I85A, L113X, E115G, L117W, and M146S. |

**Initial Design Characterization.** Upon purification (see Materials and methods), each design was subjected to an initial spectroscopic screen to identify any changes in 7-HCAA fluorescence due to the presence of calcium. Of the five designs, one design (CaM-d2; Figure 4.4b) showed no significant change in fluorescence and two designs (CaM-d3 and CaM-d4; Figure 4.4c and 4.4d) showed decreases in fluorescence of 36.84 and 17.16% respectively. The final two designs both showed increases in fluorescence of 38.20% for CaM-d1 and 106.22% for CaM-d5 (Figure 4.4a and 4.4e) upon calcium binding.

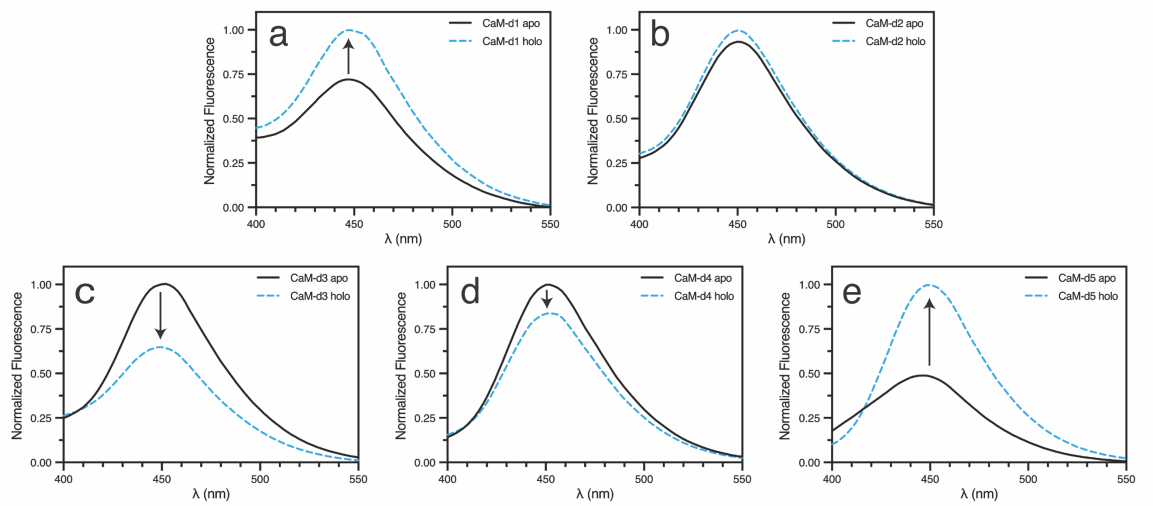



**Figure 4.4.** Emission scans for (a) CaM-d1, (b) CaM-d2, (c) CaM-d3, (d) CaM-d4 and (e) CaM-d5 in the absence (apo) and presence of 10 mM $CaCl_2$. Fluorescence values are normalized to the species (either apo or holo) with the highest fluorescence intensity. The excitation wavelength was 325 nm.

We hypothesized that CaM-d2, CaM-d3, and CaM-d4's inability to show increases in fluorescence was due to either overall protein misfolding or misorientation of the design residues. To assess whether each design was properly folded, we collected circular

dichroism (CD) spectra for both the parent protein (CaM) and each of the designs. CaM-d2 and CaM-d3 each appeared to have a significant degree of disorder in the apo state (Figure 4.5c and 4.5d) but not in the holo. This is consistent with the notion that the designed residues are disrupting the apo state making both of these unsuccessful designs. While CaM-d1 and CaM-d4 were more structurally stable than CaM-d2 or CaM-d3 (Figure 4.5b and 4.5e), each had a larger than expected change in alpha-helical content (as measured by the change at 208 and 222 nm) upon binding calcium that is inconsistent with CaM (Figure 4.5a). This suggests that there are indeed changes to the apo structure that affect the secondary structure to some degree. In the case of CaM-d4, both the inconsistency with the CD data and the apparent decrease in fluorescence ruled it out as a successful design. In contrast to the other four designs, CD spectra for CaM-d5 were consistent with a well-folded primarily alpha-helical protein (Figure 4.5f).

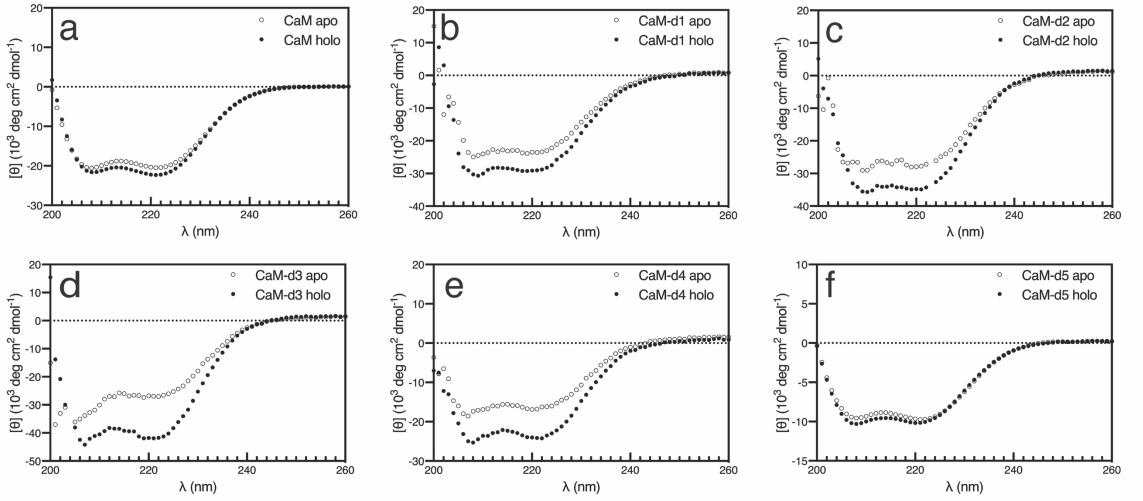



**Figure 4.5.** Circular dichroism spectra for (a) CaM, (b) CaM-d1, (c) CaM-d2, (d) CaM-d3, (e) CaM-d4 and (f) CaM-d5 in the absence (apo) and presence (holo) of 10 mM $CaCl_2$.

Because CaM-d1, unlike CaM-d4, actually produced an increase in fluorescence, we continued to examine it as a possibly successful design. However, during experimental procedures, we noticed that CaM-d1 exhibited inconsistent changes in fluorescence. We

also noted that CaM-d1 was less stable in solution (25 mM Tris-HCl, pH 7.0, 150 mM NaCl, 100 mM KCl; TBS-KCl) than either CaM or CaM-d5. Due to this observation, the previously discussed differences in CD spectra between CaM and CaM-d1 (compare Figures 4.5a and 4.5b) and the fact that the designed residues were located at the hinge region of CaM, we hypothesized that this design also likely disrupted the dynamic states of the apo protein.

To test this hypothesis, we performed thermal melts on both the apo and holo forms of CaM, CaM-d1**,** and CaM-d5 (Figure 4.6). The melting curves of apo and holo CaM differ dramatically in that the apo state melts to completion by 90 ºC while the holo state is not known to fully melt under standard melting conditions (Figure 4.6a).[69] In the case of CaM-d1, however, both states appeared to fully melt by ~70 ºC (Figure 4.6b) suggesting that CaM-d1 did not successfully switch conformations with the addition of calcium; CaM-d5, on the other hand, had similar melting curves to that of CaM (Figure 4.6c), confirming that it was indeed a fully functional CaM variant. Because of the disruption to the structure of CaM-d1, it was decided that it too was not a successful design and it was removed from further experimental characterization.

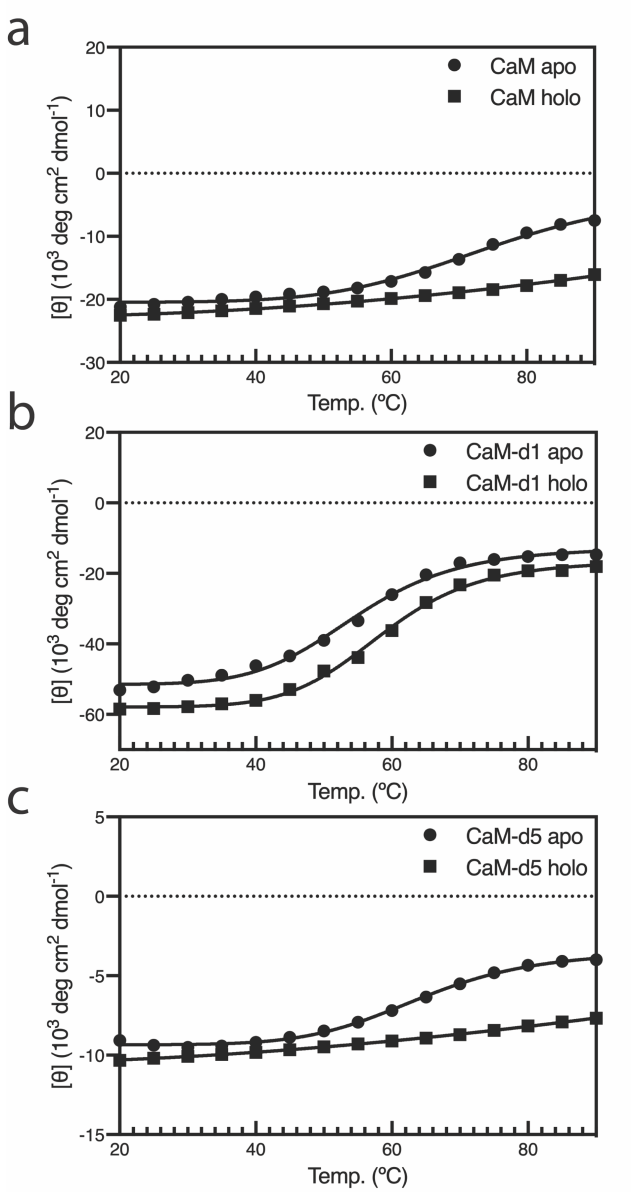



**Figure 4.6.** Thermal melting curves recorded by circular dichroism spectroscopy at 222 nm for (a) CaM, (b) CaM-d1 and (c) CaM-d5 in the absence (apo) and presence (holo) of 10 mM $CaCl_2$.

**Specific CaM-d5 Characterizations.** We also rationalized that if CaM-d5 was indeed a successful design, we would be able to use the change in 7-HCAA fluorescence to measure its apparent $K_d$ for calcium. The $K_d$ of CaM for calcium has been reported as ~ 3 μM.[70] Since we have previously used the 7-HCAA to measure an apparent $K_d$ in the μM range[17] it seemed likely that we could use the fluorescence intensity of the 7-HCAA to

measure CaM-d5's affinity for $Ca^{2+}$. CaM-d5 fluorescence was measured in the absence and presence of calcium at concentrations ranging from 10 nM to 10 mM in TBS-KCl. A sigmoidal increase in fluorescence at 450 nm was observed with increasing calcium concentration giving an apparent Kd of 3.01 μM (Figure 4.7; Table 4.2) which is consistent with a previously reported value.[70]

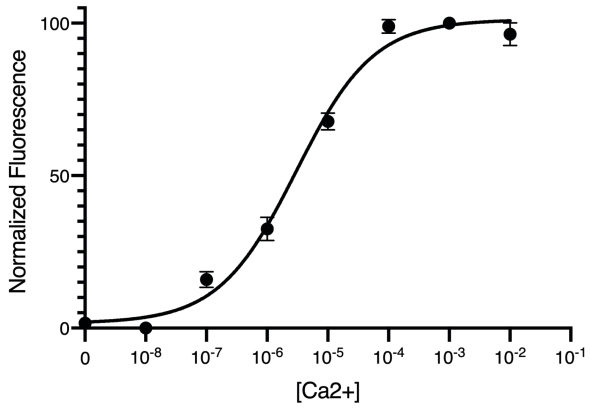



**Figure 4.7.** CaM-d5 specificity for calcium determined by relative fluorescence. Some error bars too small to be visible were omitted.

Having established CaM-d5 as a functioning variant of CaM, we set out to verify that the observed changes in fluorescence were a result of the design through mutational analysis. CaM-d5 consists of eight mutations to the parent protein: Q9E, S82T, E84L, I85A, L113X, E115G, L117W, and M146S (where X represents 7-HCAA). Since our design strategy was derived from the work of Henderson et al.[20] – where a Glu and a TRP residue are thought to be responsible for quenching – we chose to make four mutants (CaM-d5 E9A, E9D, E9Q, and W117I) that specifically addressed the ability for E9 or W117 to quench 7-HCAA fluorescence.

**Table 4.2.** Parameters and statistics for CaM-d5 apparent $K_d$ curve fit.

| **Best-fit values** | |
|---|---|
| **Top** | 101.2 RFU |
| **Bottom** | 1.563 RFU |
| **log $K_d$** | -5.522 |
| **HillSlope** | 0.6812 |
| **$K_d$** | 3.01 µM |
| **Span** | 99.62 |
| **95% CI (profile likelihood)** | |
| **Top** | 97.19 to 105.5 RFU |
| **Bottom** | -3.307 to 5.926 RFU |
| **log $K_d$** | -5.682 to -5.368 |
| **HillSlope** | 0.5538 to 0.8390 |
| **$K_d$** | 2.081 µM to 4.282 µM |
| **Goodness of Fit** | |
| **Degrees of Freedom** | 20 |
| **R squared** | 0.9894 |
| **Sum of Squares** | 433.3 |
| **Sy.x** | 4.654 |
| **Number of points** | |
| **# of X values** | 24 |
| **# Y values analyzed** | 24 |

**Investigation of Glu8 and Trp117's role in CaM-d5 quenching.** As previously mentioned, the 7-hydroxycoumarin (7-HC) moiety is a photoacid with a $pK_a$ of ~7.8 in the ground state and ~1.5 in the excited state. In polar environments, the low $pK_a$ of the excited state causes the 7-HC to spontaneously transfer a proton from the phenol to a nearby Brønsted base; a process known as excited-state proton transfer (ESPT). In aqueous solvents, the base is typically a water molecule.[8,11,12] Relaxation of the excited-state—typically through the emission of a photon at 450 nm—back to the ground state allows for the proton to be transferred back. If ESPT is blocked, then the emission max undergoes a

bathochromic shift to 380 nm. No design in this study contained any significant amount of 380 nm emission confirming that a proton is transferred from the excited state of the 7-HC to a nearby base.

It was suggested by Henderson et al.[20] that in the FAB, ESPT was occurring from the 7-HCAA to a nearby Glu residue and that an ultrafast transfer of the proton back to the 7-HC phenol could result in quenching of the excited state through thermal relaxation. As previously mentioned, the FAB crystal structure showed a 2.5 Å hydrogen bond between the Glu and the 7-HCAA. In the apo form of CaM-d5, RosettaDesign predicted Glu9 to have a similar 2.5 Å hydrogen bond with the phenol of 7-HCAA.

To measure the validity of this, we subjected the designed apo state to ONIOM calculations at the BL3YP/321-G level of theory for the high layer (residues 7-HCAA, Glu9, and Trp117) and the UFF level of theory for the low layer using Gaussian09.[71] This calculation placed the Glu8/7-HCAA hydrogen bond at 2.4 Å further suggesting that these distances were reasonable given the design model. To test the contribution of E9 quenching in CaM-d5, the mutations E9A (alanine), E9D (aspartic acid), and E9Q (glutamine) were chosen to knock out the hydrogen bonding potential, lengthen the hydrogen bond, and maintain the hydrogen bond while removing the acidic component of the side chain respectively.

Spectroscopic characterization of the mutants showed that CaM-d5 produces the largest change in fluorescence compared with all other mutants (Figure 4.8). The apo forms of E9A, E9D, and E9Q showed increases in fluorescence intensity of 103.01, 69.05, and 143.59% respectively compared with CaM-d5 apo (Figure 4.9). If E9 is participating in quenching the 7-HCAA (rather than just acting as structural support), elimination of the

hydrogen bond with E9A would be expected to increase the fluorescence intensity in the apo state as shown (compare Figure 4.9 CaM-d5 apo and CaM-d5 E9A apo). The fact that E9D doesn't have as dramatic of a change in intensity as E9A suggests that the hydrogen bond is still formed in this variant but that the quenching mechanism is not as efficient as in E9. Because back-transfer of the hydrogen bond would require nearby Brønsted base, we might expect E9A and E9Q to have similar intensities. The fact that E9Q is 20% brighter than E9A suggests that the 7-HC moiety may be adopting a different conformation in E9A where the 7-HC forms a hydrogen bond to another nearby residue (e.g. E7). The holo forms of E9A, E9D, and E9Q mutants also showed lower intensity increases compared with CaM-d5$_{holo}$ (Figure 4.8).

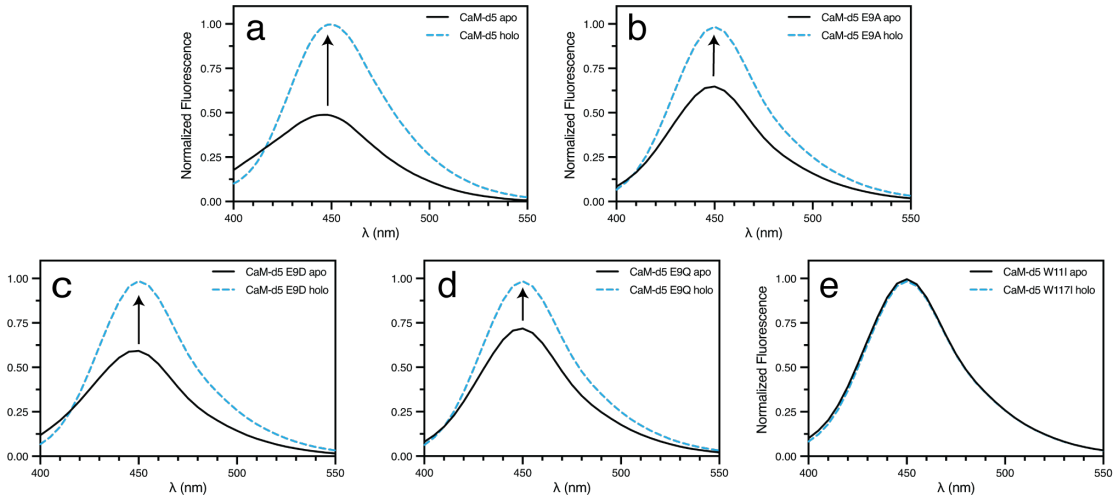



**Figure 4.8.** Emission scans for (a) CaM-d5, (b) CaM-d5 E9A, (c) CaM-d5 E9D, (d) CaM-d5 E9Q and (e) CaM-d5 W117I in the absence (apo) and presence of 10 mM $CaCl_2$. Fluorescence values are normalized to the species (either apo or holo) with the highest fluorescence intensity. The excitation wavelength was 325 nm.

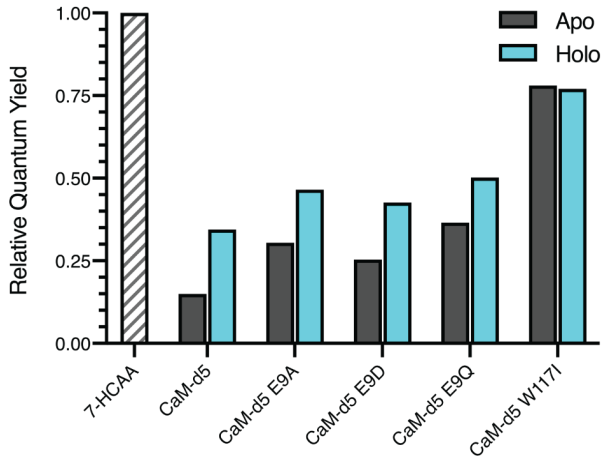



**Figure 4.9.** Relative quantum yield measurements for 10 µM of CaM-d5, CaM-d5 E9A, CaM-d5 E9D, CaM-d5 E9Q and CaM-d5 W117I—as measured by absorbance of the 7-HCAA at 325 nm—in the absence (apo) and presence (holo) of 10 mM $CaCl_2$ compared with 10 µM of 7-HCAA in TBS-KCl (25 mM Tris-HCl, pH 7.0, 150 mM NaCl, 100 mM KCl).

The RosettaDesign model of CaM-d5 predicts that residue 9, regardless of the mutant, moves significantly farther away from the 7-HCAA in the holo form (37.6 Å measured from Cα to Cα). The observation that $E9_{holo}$ and $E9D_{holo}$ have less of an increase in intensity compared with $E9A_{holo}$ or $E9Q_{holo}$ indicates that there may be a long-range component to quenching 7-HCAA by the acidic residues possibly through an extensive water network.

Furthermore, $W117_{apo}$ is predicted to be in a perpendicular pi-stacking orientation with the 7-HCAA and this mutant showed a dramatic overall increase in the fluorescence compared with CaM-d5. There was no significant change in fluorescence between the apo and holo forms suggesting that W117 is significantly quenching the apo state of CaM-d5. The overall quantum yield of W117I is equivalent to ~78% of the free small molecule 7-HCAA, suggesting that the tryptophan plays a significant role in quenching (Figure 4.9).

Thermal melting ($T_M$) curves for each mutant in the apo state were also examined and compared with CaM wild-type protein (Figure 4.10; Table 4.3). CaM-d5 had a $T_M$ of

59.39 ºC compared with 61.09 ºC for CaM. This suggests very little disruption in the overall stability of the apo state by the designed residues in CaM-d5. Of the mutants to CaM-d5: E9A, E9D, and W117I all indicated less stability in the protein with melting temperatures of 55.94, 56.63, and 54.67 ºC respectively. These data suggest that removal of the putative hydrogen bond between E9 and the 7-HCAA or the removal of the putative pi-stacking interaction is detrimental to the overall stability of CaM-d5. The E9Q mutant, on the other hand, restores stability with a $T_M$ of 59.63 ºC.

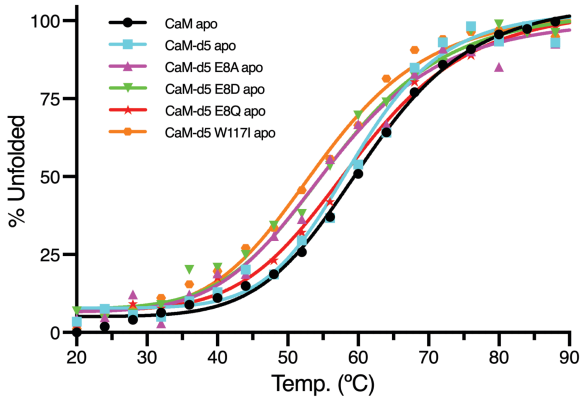



**Figure 4.10.** Thermal melting curves recorded by circular dichroism spectroscopy at 222 nm for CaM, CaM-d5, CaM-d5 E9A, CaM-d5 E9D, CaM-d5 E9Q, and CaM-d5 W117I in the absence (apo) of $CaCl_2$ in TBS-KCl (25 mM Tris-HCl, pH 7.0, 150 mM NaCl, 100 mM KCl).

**Table 4.3.** Melting temperatures for CaM, CaM-d5, CaM-d5 E9A, CaM-d5 E9D, CaM-d5 E9Q, and CaM-d5 W117I in the absence (apo) of $CaCl_2$ in TBS-KCl (25 mM Tris-HCl, pH 7.0, 150 mM NaCl, 100 mM KCl).

| Construct | $T_M$ (ºC) |
|---|---|
| CaM | 61.09 |
| CaM-d5 | 59.39 |
| CaM-d5 E9A | 55.94 |
| CaM-d5 E9D | 56.63 |
| CaM-d5 E9Q | 59.63 |
| CaM-d5 W117I | 54.67 |

Finally, fluorescence lifetime data for each of the CaM-d5 mutants was examined and compared with the 7-HCAA small molecule (Figure 4.11; Table 4.4). The small molecule 7-HCAA has a single fluorescent exponential decay time of 5.90 ns. Each mutant,

regardless of whether it was the apo or holo form, showed a bi-modal set of decay times. The "fast" decay time was fairly consistent among the E9 mutants centering around 1.5 ns in the apo state (Table 4.4) while the W117I mutant showed a decay time approximately twice as long (3.27 ns). The "slower" decay time for each mutant is consistent with the small molecule 7-HCAA (Table 4.4). These data suggest that W117 is more impactful on the quenching (which is consistent with the steady state fluorescence spectra in Figure 4.8 and Figure 4.9) of the 7-HCAA than E9.

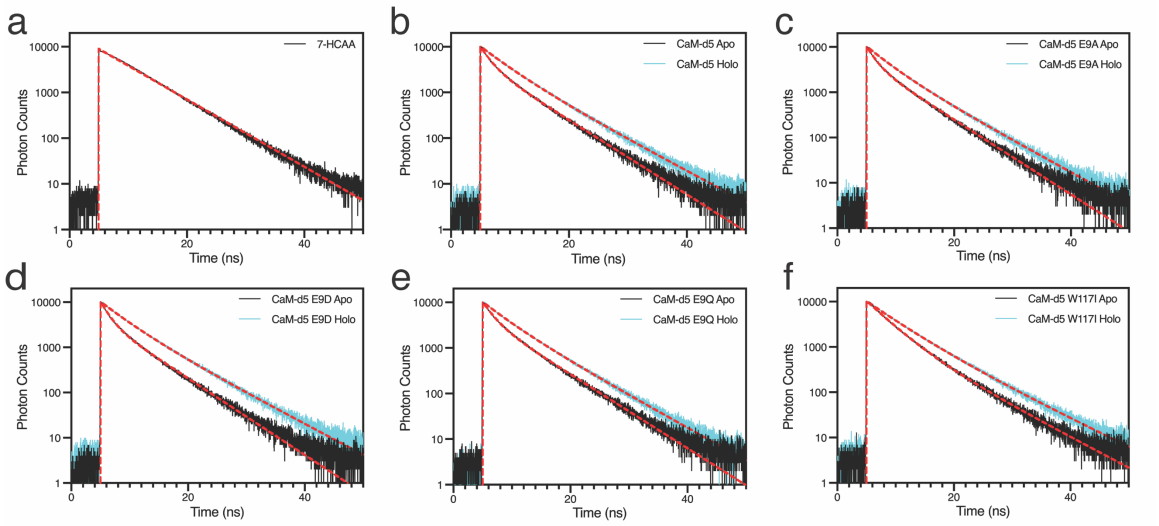



**Figure 4.11.** Fluorescent lifetime scans for (a) CaM-d5, (b) CaM-d5 E9A, (c) CaM-d5 E9D, (d) CaM-d5 E9Q and (e) CaM-d5 W117I in the absence (apo) and presence of 10 mM $CaCl_2$. The excitation wavelength was 325 nm with emission detection at 450 nm.

**Table 4.4.** CaM-d5 mutant fluorescence lifetime.

| Construct | Apo T1 (ns) | Apo T2 (ns) | Holo T1 (ns) | Holo T2 (ns) |
|---|---|---|---|---|
| CaM-d5 | 1.6216 | 5.3367 | 3.5353 | 6.3166 |
| CaM-d5 E9A | 1.4623 | 5.2663 | 3.1855 | 6.1544 |
| CaM-d5 E9D | 1.4140 | 5.1282 | 3.4932 | 6.3740 |
| CaM-d5 E9Q | 1.6818 | 5.3545 | 3.3578 | 6.0836 |
| CaM-d5 W117I | 3.2690 | 6.4188 | 4.2215 | 7.0141 |
| 7-HCAA | 5.8988 | -- | -- | -- |

## 4.3 Conclusion

The results reported here represent the first description of a computationally designed biosensor where the fluorescent output is controlled by atomic level interactions

between amino acids. While most protein design studies require the designs be characterized using structural methods, in this case it wasn't such characterization wasn't plausible. Instead, we have consistently shown through a variety of techniques that indeed the designed residues are contributing to the quenching of the 7-HCAA. This opens the door for the design of new biosensors of ligand-protein interaction through precision engineering.

### 4.4 Materials and Methods

**Computational design methods:** Computational methods were adapted from those described in Renfew et al.[72], Richter et al.[67], and Mills et al.[73]

**RosettaMatch:** First, we developed Rosetta parameters files for a virtual 7-hydroxycoumarin (V7H.params) and the L-(7-hydroxycoumarin-4yl) ethylglycine (7-HC.params). The virtual 7-HC can be used as an imaginary ligand for initially matching the 7-HCAA to. This methodology allows us to use the RosettaMatch algorithm to quickly identify sites where the 7-HCAA can fit without clashing with the backbone using primary matching. The rest of the theosite residues can then be matched using secondary matching. Next we developed a constraint file (7-HC.cst) based on the theosite geometry we identified from PDB 6bjz. And finally, matching was carried out with a bash script (matching.sh) that loops through each of the 25 CaM NMR states using a Rosetta flags file (general_match.flags), a Rosetta positions file (CaM_relaxed.pos) and a Rosetta gridlig file (CaM_relaxed.gridlig).

**Rosetta Params File for the Virtual 7-HC (V7H.params):**

```
NAME V7H
IO_STRING V7H Z
TYPE LIGAND
AA UNK
ATOM  CZ2  VIRT  VIRT
```

```
ATOM  CE2  VIRT  VIRT
ATOM  CD   VIRT  VIRT
ATOM  CE1  VIRT  VIRT
ATOM  CZ1  VIRT  VIRT
ATOM  OH1  VIRT  VIRT
ATOM  OH2  VIRT  VIRT
ATOM  X1   VIRT  VIRT
ATOM  CZ3  VIRT  VIRT
ATOM  CH2  VIRT  VIRT
ATOM  CT   VIRT  VIRT
ATOM  CH1  VIRT  VIRT
ATOM  OI   VIRT  VIRT
ATOM  X2   VIRT  VIRT
BOND  CD    CE1
BOND  CD    CE2
BOND  CD    X1
BOND  CE1   CZ1
BOND  CE2   CZ2
BOND  CE2   CZ3
BOND  CZ1   OH1
BOND  CZ1   OH2
BOND  CZ1   X1
BOND  CZ2   CH1
BOND  CZ2   OH1
BOND  CZ2   X1
BOND  CZ3   CH2
BOND  CH1   CT
BOND  CH2   CT
BOND  OI    CT
BOND  CH1      X2
BOND  X2       CH2
NBR_ATOM  CZ2
NBR_RADIUS 3.656923
ICOOR_INTERNAL    CZ2     0.000000    0.000000    0.000000   CZ2    CE2    CD
ICOOR_INTERNAL    CE2     0.000000  179.999999    1.397726   CZ2    CE2    CD
ICOOR_INTERNAL    CD      0.000000   61.054756    1.476951   CE2    CZ2    CD
ICOOR_INTERNAL    CE1    -0.668101   62.497014    1.343762   CD     CE2    CZ2
ICOOR_INTERNAL    CZ1     0.036514   55.942975    1.470990   CE1    CD     CE2
ICOOR_INTERNAL    OH1     0.878190   63.104613    1.382945   CZ1    CE1    CD
ICOOR_INTERNAL    OH2  -179.775270   58.803175    1.225876   CZ1    CE1    OH1
ICOOR_INTERNAL    X1    179.116328  123.544702    1.414880   CZ1    CE1    OH2
ICOOR_INTERNAL    CZ3   179.995663   62.685988    1.402267   CE2    CZ2    CD
ICOOR_INTERNAL    CH2    -0.017870   59.090910    1.400047   CZ3    CE2    CZ2
ICOOR_INTERNAL    CT     -0.196397   59.994819    1.388908   CH2    CZ3    CE2
ICOOR_INTERNAL    CH1     0.239061   59.473111    1.386404   CT     CH2    CZ3
ICOOR_INTERNAL    OI    179.877910   62.064103    1.363088   CT     CH2    CH1
ICOOR_INTERNAL    X2    179.116328  123.544702    1.414880   CH1    CT     OI
```

## Rosetta Params File for 7-HCAA (7-HC.params):

```
NAME 7-HC
IO_STRING 7-HC X
TYPE POLYMER
AA UNK
ROTAMER_AA GLU
ATOM  N   Nbb  NH1  -0.56
ATOM  CA  CAbb CT1  -0.04
ATOM  C   CObb C    0.67
ATOM  O   OCbb O    -0.50
ATOM  CB  CH2  CT2  -0.13
ATOM  CG  CH2  CT2  -0.13
ATOM  CD  aroC CA   -0.07
ATOM  CE1 aroC CA   -0.07
ATOM  CE2 aroC CPT  -0.07
ATOM  CZ1 COO  CA   0.67
ATOM  CZ2 aroC CPT  -0.07
ATOM  CZ3 aroC CA   -0.07
ATOM  OH1 OOC  OS   -0.71
ATOM  OH2 OOC  OB   -0.71
```

```
ATOM  CH1 aroC CA   -0.07
ATOM  CH2 aroC CA   -0.07
ATOM  CT  aroC CA   -0.07
ATOM  OI  OH   OH1  -0.61
ATOM  H   HNbb H    0.48
ATOM  HA  Hapo HB   0.14
ATOM 1HB  Hapo HA   0.14
ATOM 2HB  Hapo HA   0.14
ATOM 1HG  Hapo HA   0.14
ATOM 2HG  Hapo HA   0.14
ATOM 1HE1 Haro HP   0.16
ATOM 1HZ3 Haro HP   0.16
ATOM 1HT3 Haro HP   0.16
ATOM 1HT4 Haro HP   0.16
ATOM 1HK  Hpol H    0.48
ATOM  X1  VIRT VIRT
BOND  N    CA
BOND  N    H
BOND  CA   C
BOND  CA   CB
BOND  CA   HA
BOND  C    O
BOND  CB   CG
BOND  CB  1HB
BOND  CB  2HB
BOND  CG   CD
BOND  CG  1HG
BOND  CG  2HG
BOND  CD   CE1
BOND  CD   CE2
BOND  CE1  CZ1
BOND  CE1 1HE1
BOND  CE2   X1
BOND  CE2  CZ2
BOND  CE2  CZ3
BOND  CZ1  OH1
BOND  CZ1  OH2
BOND  CZ2  CH1
BOND  CZ2  OH1
BOND  CZ3  CH2
BOND  CZ3 1HZ3
BOND  CH1  CT
BOND  CH1 1HT3
BOND  CH2  CT
BOND  CH2 1HT4
BOND  OI   CT
BOND  OI  1HK
LOWER_CONNECT N
UPPER_CONNECT C
CHI 1  N    CA   CB   CG
CHI 2  CA   CB   CG   CD
CHI 3  CB   CG   CD   CE1
CHI 4  CH2  CT   OI  1HK
PROTON_CHI 4 SAMPLES 2 0 180 EXTRA 1 20
NBR_ATOM  CB
NBR_RADIUS 11.045592
FIRST_SIDECHAIN_ATOM  CB
PROPERTIES PROTEIN L_AA
ICOOR_INTERNAL    N       0.000000    0.000000    0.000000   N     CA    C
ICOOR_INTERNAL    CA      0.000000  180.000000    1.446845   N     CA    C
ICOOR_INTERNAL    C       0.000000   72.586913    1.537478   CA    N     C
ICOOR_INTERNAL   LOWER -136.769138   58.123619    1.380354   N     CA    C
ICOOR_INTERNAL   UPPER  124.287033   65.804611    1.381715   C     CA    N
ICOOR_INTERNAL    O     -56.508386   56.506227    1.229365   C     CA    N
ICOOR_INTERNAL    CB   -121.675264   69.564721    1.533092   CA    N     C
ICOOR_INTERNAL    CG   -176.379613   67.974162    1.536032   CB    CA    N
ICOOR_INTERNAL    CD    176.113190   62.788045    1.517216   CG    CB    CA
ICOOR_INTERNAL    CE1    -1.814124   55.933267    1.343392   CD    CG    CB
ICOOR_INTERNAL    CZ1  -179.889013   55.959236    1.471403   CE1   CD    CG
ICOOR_INTERNAL    OH1     0.852830   63.111809    1.382812   CZ1   CE1   CD
ICOOR_INTERNAL    CZ2    -1.115446   59.013932    1.406814   OH1   CZ1   CE1
```

```
ICOOR_INTERNAL    CE2     0.521496   58.414164    1.397528   CZ2   OH1   CZ1
ICOOR_INTERNAL    X1      0.000000   61.000000    1.006000   CE2   CD    CE1
ICOOR_INTERNAL    CZ3  -179.626138   62.715720    1.402779   CE2   CZ2   OH1
ICOOR_INTERNAL    CH2     0.034026   59.094164    1.400474   CZ3   CE2   CZ2
ICOOR_INTERNAL    CT     -0.187129   59.999749    1.388940   CH2   CZ3   CE2
ICOOR_INTERNAL    CH1     0.156993   59.460676    1.386731   CT    CH2   CZ3
ICOOR_INTERNAL   1HT3   179.944714   58.241631    1.084347   CH1   CT    CH2
ICOOR_INTERNAL    OI    179.982710   62.072634    1.362753   CT    CH2   CH1
ICOOR_INTERNAL   1HK    178.800297   71.437402    0.971533   OI    CT    CH2
ICOOR_INTERNAL   1HT4  -179.879122   59.977525    1.086946   CH2   CZ3   CT
ICOOR_INTERNAL   1HZ3  -179.983813   58.528353    1.086677   CZ3   CE2   CH2
ICOOR_INTERNAL    OH2  -179.750920   58.828675    1.225586   CZ1   CE1   OH1
ICOOR_INTERNAL   1HE1   179.832692   57.392308    1.084318   CE1   CD    CZ1
ICOOR_INTERNAL   1HG   -122.444890   72.063496    1.098054   CG    CB    CD
ICOOR_INTERNAL   2HG   -115.745364   72.157141    1.097455   CG    CB   1HG
ICOOR_INTERNAL   1HB   -122.194526   70.108847    1.098011   CB    CA    CG
ICOOR_INTERNAL   2HB   -116.602226   71.551488    1.097743   CB    CA   1HB
ICOOR_INTERNAL    HA   -119.872500   70.730397    1.095858   CA    N     CB
ICOOR_INTERNAL    H     171.176963   62.294655    1.012488   N    LOWER  CA
```

## RosettaMatch Constraint File (7-HC.cst):

```
#BLOCK 1: VIRTUAL LIGAND PLACEMENT:
CST::BEGIN
  TEMPLATE::   ATOM_MAP: 1 atom_name: OI CT CH1
  TEMPLATE::   ATOM_MAP: 1 residue3:  V7H

  TEMPLATE::   ATOM_MAP: 2 atom_name: OH2 CZ1 OH1
  TEMPLATE::   ATOM_MAP: 2 residue3:  7-HC

  CONSTRAINT:: distanceAB:    7.10    0.10   80.00  1       1
  CONSTRAINT::    angle_A:   29.40    1.00   10.00  360.00  1
  CONSTRAINT::    angle_B:   31.00    1.00   10.00  360.00  1
  CONSTRAINT::  torsion_A:    0.00    1.00   10.00  360.00  1
  CONSTRAINT:: torsion_AB:    0.00    1.00   10.00  360.00  1
  CONSTRAINT::  torsion_B:    0.00    1.00   10.00  360.00  1

#BLOCK 2: PI-STACKING INTERACTIONS:
CST::BEGIN
  TEMPLATE::   ATOM_MAP: 1 atom_name: X1 CD CE1
  TEMPLATE::   ATOM_MAP: 1 residue3:  V7H

  TEMPLATE::   ATOM_MAP: 2 atom_type: aroC,
  TEMPLATE::   ATOM_MAP: 2 residue1:  FWY

  CONSTRAINT:: distanceAB:    3.90    0.20   80.00  0       1
  CONSTRAINT::    angle_A:   90.00   20.00   10.00  180.00  1
  CONSTRAINT::    angle_B:   90.00   20.00   10.00  180.00  1
  CONSTRAINT::  torsion_A:   90.00   20.00   10.00  180.00  1
  CONSTRAINT::  torsion_B:   90.00   20.00   10.00   90.00  1

  ALGORITHM_INFO:: match
    SECONDARY_MATCH: DOWNSTREAM
  ALGORITHM_INFO::END
CST::END

#BLOCK 3: HYDROGEN BONDING TO 7-HYDROXYCOUMARIN PHENOL:
CST::BEGIN
  TEMPLATE::   ATOM_MAP: 1 atom_name: OI CT CH2
  TEMPLATE::   ATOM_MAP: 1 residue3:  7-HC

  TEMPLATE::   ATOM_MAP: 2 atom_type: OOC ,
  TEMPLATE::   ATOM_MAP: 2 residue1:  DE

  CONSTRAINT:: distanceAB:    2.75    0.25   80.00  0       1
  CONSTRAINT::    angle_A:  180.00   20.00   10.00  180.00  1
  CONSTRAINT::    angle_B:  120.00   20.00   10.00  120.00  1

  ALGORITHM_INFO:: match
```

```
    SECONDARY_MATCH: UPSTREAM_CST 1
  ALGORITHM_INFO::END
CST::END
```

### RosettaMatch Flags File (general_match.flags):

```
-bump_tolerance 0.5
-chemical:exclude_patches D_AA
-consolidate_matches 1
-enumerate_ligand_rotamers
-ex1
-ex2
-euclid_bin_size 0.5
-euler_bin_size 6
-in:ignore_unrecognized_res
-match:filter_colliding_upstream_residues
-match:filter_upstream_downstream_collisions
-match_grouper SameSequenceAndDSPositionGrouper
-match:updown_collision_tolerance 0.3
-match:upstream_residue_collision_tolerance 0.95
-packing
-only_enumerate_non_match_redundant_ligand_rotamers
-out::file::output_virtual
-output_matches_per_group 1
-output_matchres_only false
-output_format PDB
-use_input_sc
```

### RosettaMatch Bash Script (matching.sh):

```
for i in $(ls ../inputs/CaM_00*.pdb);
        do <PATH TO ROSETTA>/main/source/bin/match.macosclangrelease \
        -s $i \
        -match:lig_name V7H \
        -extra_res_fa ../params/V7H.params ../params/7-HC.params \
        -match:geometric_constraint_file ../inputs/7-HC.cst \
        -match:scaffold_active_site_residues ../inputs/CaM_relaxed.pos \
        -active_site_definition_by_gridlig ../inputs/CaM_relaxed.gridlig \
        -match:grid_boundary ../inputs/CaM_relaxed.gridlig \
        @../flags/general_match.flags;
done
```

### RosettaDesign:

Once a set of matches was identified, we created a RosettaDesign xml file (design.xml) that utilizes a typical Rosetta enzyme design protocol.[67] Design was carried out with a bash script (design.sh) that loops through each match using a Rosetta flags file (general_design.flags), the matching constraint file (7-HC.cst) and each of the params files (V7H.params and 7-HC.params).

### RosettaDesign XML File (design.xml):

```
<ROSETTASCRIPTS>
        <TASKOPERATIONS>
```

```
        <DetectProteinLigandInterface name="dsgn_cuts_on" cut1="6" cut2="8"
        cut3="10" cut4="12" design="1"/>

        <DetectProteinLigandInterface name="dsgn_cuts_off" cut1="6"
        cut2="8" cut3="10" cut4="12" design="0"/>
</TASKOPERATIONS>

<SCOREFXNS>
        <ScoreFunction name="tal2014" weights="talaris2014"/>
</SCOREFXNS>

<MOVERS>
        #Add constraints to file

        <AddOrRemoveMatchCsts name="addcst" cst_instruction="add_new"/>

        <AddOrRemoveMatchCsts name="rmvcst" cst_instruction="remove"
        keep_covalent="1"/>

        <AddOrRemoveMatchCsts name="addprg"
        cst_instruction="add_pregenerated"/>

        #Optimize the pose per the cst file

        <EnzRepackMinimize name="cstopt" scorefxn_minimize="tal2014"
        cst_opt="1" design="0"repack_only="0" fix_catalytic="0"
        minimize_rb="1" minimize_bb="1" minimize_sc="1" minimize_lig="1"
        min_in_stages="1" cycles="1" task_operations="dsgn_cuts_off"/>

        #Design and repacking around the catalytic residues; keep the
        catalytic residues fixed in this instance.

        <EnzRepackMinimize name="dsgn" scorefxn_minimize="tal2014"
        cst_opt="0"design="1"repack_only="0" fix_catalytic="1"
        minimize_rb="1" minimize_bb="1" minimize_sc="1" minimize_lig="0"
        min_in_stages="1" backrub="0" cycles="1"
        task_operations="dsgn_cuts_on"/>

        #Minimize after each design.

        <EnzRepackMinimize name="min" scorefxn_minimize="tal2014"
        cst_opt="0" design="0"repack_only="0" fix_catalytic="1"
        minimize_rb="1 " minimize_bb="1" minimize_sc="1" minimize_lig="1"
        min_in_stages="1" backrub="0" cycles="1"
        task_operations="dsgn_cuts_off"/>

        #Perform a final repacking step.

        <EnzRepackMinimize name="rpkmin" scorefxn_minimize="tal2014"
        cst_opt="0" design="0" repack_only="1" fix_catalytic="0"
        minimize_rb="1" minimize_bb="1" minimize_sc="1" minimize_lig="0"
        min_in_stages="1" backrub="0" cycles="1"
        task_operations="dsgn_cuts_off"/>

        <GenericMonteCarlo name="multi_cstopt" mover_name="cstopt"
        scorefxn_name="tal2014" trials="10" sample_type="low"
        temperature="0.6" drift="1" recover_low="1" preapply="0"/>

        <GenericMonteCarlo name="multi_dsgn" mover_name="dsgn"
        scorefxn_name="tal2014" trials="10" sample_type="low"
        temperature="0.6" drift="0" recover_low="1" preapply="0"/>

        <GenericMonteCarlo name="multi_min" mover_name="min"
        scorefxn_name="tal2014" trials="10" sample_type="low"
        temperature="0.6" drift="0" recover_low="1" preapply="0"/>

        <GenericMonteCarlo name="multi_rpkmin" mover_name="rpkmin"
        scorefxn_name="tal2014" trials="10" sample_type="low"
        temperature="0.6" drift="0" recover_low="1" preapply="0"/>
```

```
                <ParsedProtocol name="des_min">
                        <Add mover="multi_dsgn"/>
                        <Add mover="multi_min"/>
                </ParsedProtocol>

                <ParsedProtocol name="finmin_rpkmin">
                        <Add mover="multi_min"/>
                        <Add mover="multi_rpkmin"/>
                </ParsedProtocol>
        </MOVERS>

        <PROTOCOLS>
                <Add mover="addcst"/>
                <Add mover="multi_cstopt"/>
                <Add mover=”des_min”/>
                <Add mover=”des_min”/>
                <Add mover=”des_min”/>
                <Add mover="dsgn"/>
                <Add mover="rmvcst"/>
                <Add mover="finmin_rpkmin"/>
        </PROTOCOLS>
</ROSETTASCRIPTS>
```

**RosettaDesign Flags File (general_design.flags):**

```
-run:preserve_header
-output_virtual false
-use_input_sc
-no_his_his_pairE
-score::hbond_params correct_params
-lj_hbond_hdis 1.75
-lj_hbond_OH_donor_dis 2.6
-linmem_ig 10
-nblist_autoupdate true
-in:ignore_unrecognized_res
-out::overwrite
-restore_talaris_behavior
```

**RosettaDesign Bash Script:**

```
for i in $(ls ../matches/*.pdb);
        do <PATH TO ROSETTA>/main/source/bin/rosetta_scripts.macosclangrelease \
        -nstruct 1 \
        -jd2:ntrials 1 \
        -parser:protocol design.xml \
        -database <PATH TO ROSETTA>/main/database/ \
        -out::overwrite -s $i \
        -extra_res_fa ../params/V7H.params ../params/7-HC.params \
        -enzdes::cstfile ../inputs/7-HC.cst \
        @../flags/general_design.flags
done
```

**Reagents.** 7-HCAA was obtained from Bachem (Torrance, CA). Commercially available reagents were used without further purification.

**Molecular Cloning, Protein Expression, & Purification.** A vector containing the wild-type calmodulin gene (CaM) from H. sapien was generously provided as a gift from the Hariadi lab at Arizona State University. It was subsequently amplified by PCR and

subcloned into a pET29b(+) vector (Novagene) using Gibson assembly (New England Biolabs). Mutations for each design were generated by overlap extension PCR and Gibson assembly using the CaM gene as a template. The mutations for each design were confirmed with Sanger sequencing (Genewiz: Plainfield, NJ). The sequenced confirmed expression plasmids for CaM and each design were then transformed into BL21* (DE3) cells (Invitrogen; Carlsbad CA) along with the pEVOL-CouRS plasmid.[2,52] The pEVOL-CouRS plasmid contains an orthogonal aminoacyl tRNA synthetase / tRNA pair that includes a tRNA with an anti-codon loop specific to the amber codon (termed $tRNA_{CUA}$) and a CouRS aminoacyl tRNA synthetase that can acylate the $tRNA_{CUA}$ with 7-HCAA. Because pET29b(+) contains a kanamycin resistance marker and pEVOL-CouRS contains a chloramphenicol resistance marker, BL21* (DE3) cells containing any of the designs were plated on dual-antibiotic plates (kanamycin, 50 μg / mL and chloramphenicol, 34 μg / mL) and were allowed to incubate overnight at 37 ºC. Alternatively, BL21* (DE3) cells containing CaM were cultured on plates containing kanamycin alone.

Proteins were expressed as previously described.[17] Briefly, a single colony was used to inoculate 5 mL of 2xYT media and cultures were incubated overnight at 37 ºC with 250 rpm shaking until they reached an $OD_{600}$ of ~6.0. The cultures were then used to inoculate 50 mL of 2xYT media and incubated at 37 ºC with 250 rpm shaking until an $OD_{600}$ of 1 was reached. 7-HCAA (1 mM final concentration) and Arabinose (0.2% final concentration) were then added to each of the design cultures and the cultures were incubated for 1 hour at 37 ºC. Cultures were then incubated between 24 and 48 hours at 25 ºC with 180 rpm shaking. Cells were harvested with centrifugation (4,200 xg for 15 minutes), resuspended in lysis buffer (25 mM Tris-HCl, pH 8.0, 10 mM NaCl, 3 mM BME)

and frozen. Cell pellets were then thawed and PMSF (1 mM final concentration), lysozyme (1 mg/mL final concentration), and NaCl (500 mM final concentration) were added. The cell suspension was then subjected to sonication for 5 minutes (1 second on, 2 seconds off, 20 Hz, total 15 minutes) and centrifuged at 50,000 xg for 20 minutes to remove cell debris.

CaM and each of the designs contain a C-terminal 6x histidine tag. Therefore, protein was purified using a nickel-nitrilotriacetic acid resin (Ni-NTA, HisTrap FF, GE Healthcare). Contaminant proteins were removed by washing the column with 5 column volumes (CV) of Ni-NTA buffer A (25 mM Tris-HCl pH 8.0, 20 mM imidazole and 500 mM NaCl) followed by immediate elution with 100% Ni-NTA buffer B (25 mM Tris-HCl pH 8.0, 500 mM imidazole and 150 mM NaCl). The need for immediate elution was due to the low binding affinity of CaM to the Ni-NTA resin, typically we started seeing elution at just 3% Ni-NTA buffer B. Eluted fractions that were found to be fluorescent using a SpectraMax M5 spectrophotometer (Molecular Devices, San Jose, CA) were combined and diluted 10x with anion exchange (IEC) wash buffer A (25 mM Tris-HCl pH 8.0, 10 mM NaCl). Proteins were then further purified on IEC resin (HiTrap Q FF, GE Healthcare). The column was washed with five CV of IEC wash buffer A, followed by five CV of 50% IEC buffer A and 50% IEC buffer B (25 mM Tris-HCl pH 8.0, 500 mM NaCl). Proteins were eluted with five CV of 100% IEC buffer B. Again, protein fractions that were found to be fluorescent were consolidated and concentrated in a 10kDa molecular weight cutoff spin column (Amicon) to 500 µL. The concentrated protein was then injected onto a size-exclusion column (Superdex 200 increase 10/300 gl, GE Healthcare) and eluted from the column with sizing buffer (50 mM Tris-HCl pH 9.0, 150 mM NaCl). Fractions that were fluorescent were collected, consolidated, and concentrated to a volume of ~1 mL on a 10

kDa molecular cutoff spin column (Amicon) and refrigerated at 4 ºC for use in experimental procedures.

**Spectroscopic Analysis.** All spectroscopic experiments were performed in 20 mM Tris-HCl pH 7.0, 150 mM NaCl, 100 mM KCl (TBS-KCl) using a 1 cm quartz cuvette (Starna Cells; Atascadero, CA) using a SpectraMax M5 spectrophotometer. Each design was concentrated or diluted as necessary to a final absorbance value of 0.05 at 325 nm. The samples were then divided into separate tubes and either 10 µL of $CaCl_2$ (10 mM final concentration) or 10 µL of buffer was added. Absorbance spectra were taken from 250 nm to 550 nm at 5 nm intervals, again using a SpectraMax M5 spectrophotometer. Fluorescence emission spectra were collected by exciting the proteins at 325 nm and measuring the fluorescence intensity from 380 nm to 550 nm in 5 nm intervals. All spectra represent the average of three readings.

**Circular Dichroism.** 300 µL of protein sample (10 µM) was placed into a 1 mm quartz cuvette (Starna Cells; Atascadero, CA) in TBS-KCl buffer. Multi-wavelength CD measurements were taken every 1 nm from 260 to 200 nm. For melting curves. Measurements were taken from 20 to 90 ºC at 222 nm wavelength. All measurements were taken on a Jasco J-815 CD Spectrometer (Jasco Inc., Easton, MD) and processed with the Jasco Spectra Manager V1.54.03 software package. Curve fitting was performed with the Prism software package (Graphpad; San Diego, CA); the 4-parameter logistic (4PL) non-linear regression function with variable slope (Eq. 1). $T_M$ values derived from these fits are listed in Table S2.

$$f(x) = d + \frac{a-d}{1+\left(\frac{x}{c}\right)^b} \qquad \textbf{(Eq. 1)}$$

**CaM-d5 Calcium Dependent Response.** 10 µL of 1 M $CaCl_2$ was serial diluted in sterile water to concentrations of 100 mM, 10 mM, 1 mM, 100 µM, 10 µM and 1 µM. 10 µL of each $CaCl_2$ concentration was then mixed with 90 µL of CaM-d5 (10 µM) in TBS-KCl buffer to give concentrations of: 10 mM, 1 mM, 100 µM, 10 µM, 1 µM, 100 nM and 10 nM. Samples were excited at 325 nm and fluorescence was measured at 450 nm using a 1 cm quartz cuvette (Starna Cells; Atascadero, CA) on a SpectraMax M5 spectrophotometer. All data were normalized to the highest value and are the average of three readings. Curve fitting was performed with the Prism software package (Graphpad; San Diego, CA) using the 4-parameter logistic (4PL) non-linear regression function (Eq. 1). $K_d$ values derived from these fits are listed in Table 4.2.

# CHAPTER 5.

# OVERVIEW AND OUTLOOK

The work I presented in this thesis centers around the fNCAA 7-HCAA for use in protein-based biosensors. The small size and the ability to be genetically encoded at nearly any location are two ways fNCAAs offer advantages over other technologies such as autofluorescent fusion proteins (aFP) or post-translationally added fluorophores. Previous research using any of the five genetically encoded fNCAAs (see Figure 1.1) covers the spectrum of fluorescent techniques, including Förster resonance energy transfer (FRET), changes in fluorescence lifetime measurements, and changes in fluorescence spectra based on solvatochromism to name a few. However, these works fail to investigate direct ways of controlling the fluorophore through specific well-defined interactions with other amino acids in the protein. Instead, they rely on trial and error to find optimal placements for the fluorophore or make assumptions regarding the underlying mechanisms that control the fluorescence output. My work, on the other hand, has been to better understand, at least for 7-HCAA, how to control the fluorophore in a protein environment in order to have provide general control of the fluorescent output.

The first project I presented in this thesis, was a FRET sensor for monitoring glucose concentration. What made this work unique compared with other FRET based sensors is that it was the first work to show the use of tryptophan (Trp) as a FRET donor for a 7-HCAA FRET acceptor within the same protein. This was an important step forward in the development of FRET sensors because both the acceptor and donor were genetically encoded with emission from the acceptor in the visible range. This allows for the

stochiometric labeling of the protein at locations that are optimal for FRET and with easy detection of the FRET event.

FRET efficiencies are well-known to be sensitive to changes in the distance between the donor and acceptor fluorophores. However, the orientation of the transition dipole moment between the donor and acceptor can have a large impact as well. This is generally overlooked when developing FRET sensors because the fluorophores are typically aFPs or post-translationally added dyes. Both of these labeling methods tend to be at locations (the termini of the protein for aFPs or surface exposed residue for post-translational modifications) that allow free rotation of the fluorophore. Free rotation of the fluorophore will tend the orientation factor to 2/3 making it a constant in the FRET equation in most applications.

Being able to label a protein of interest at locations other than the termini or surface presents an opportunity to favor optimal orientation factors over optimal distance. I am therefore currently developing software based on the Rosetta protein design suite that can accurately calculate FRET efficiencies between the 7-HCAA and native tryptophan residues in a protein of interest that doesn't strictly rely on distance. This is only possible if the fluorophores are constrained with little freedom of rotation, such as can be done by genetically encoding the fluorophores. We expect this work to be a major jump in how FRET labeling is performed.

The second project I presented in this document, explored the structural basis for changes in fluorescence of the 7-HCAA when a ligand was present. For this work we chose to utilize the well-known streptavidin/biotin system because of the streptavidin protein's high affinity for the ligand biotin. Our lab is the first to present structures of proteins

containing a 7-HCAA. Each of these structures elucidated different interactions between the protein and the 7-HCAA or the biotin and 7-HCAA that shed light on the mechanism for changes in fluorescence. For example, in the case of L110X (see Figure 3.2 and Figure 3.3) where upon ligand binding fluorescence was quenched, we were able to identify that it was the carboxylate of the biotin sitting planar to the pyranone ring of the 7-HCAA that induced quenching. What's interesting about this project in general is that when we know what interactions drive changes in fluorescence, we can take those interactions and use computational design methods to build them into new sensors. Understanding these interactions at the atomic level gives us more control for future biosensor design projects.

In the final project presented in chapter 4 of this work, I utilized interactions that had been identified by a post-doc in our lab, Nathan Henderson, between a genetically encoded 7-HCAA, a glutamic acid and a tryptophan to computationally design a calcium biosensor using the protein calmodulin. This represents the first computationally designed biosensor using an fNCAA and is important because it shows just how we can take interactions from protein crystal structures (i.e. those interactions described in chapter 3) and use them in design in a predictable fashion. This opens the door for the design of new biosensors that are more sensitive than previously possible. The next step for this work is to move away from model systems like calmodulin to build novel biosensors.

In all, my work has been a cohesive examination into the use of 7-HCAA in the development of protein-based biosensors. However, this work is just the beginning and in general represents a set of "proof-of-concepts" that can be greatly expanded on. By researching, designing and engineering novel biosensors with a high degree of fluorescence output control, this work has provided foundational groundwork.

REFERENCES


(1) Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. A Genetically Encoded Fluorescent Amino Acid. *Proc. Natl. Acad. Sci.* **2006**, *103* (26), 9785–9789. https://doi.org/10.1073/pnas.0603965103.

(2) Wang, J.; Xie, J.; Schultz, P. G. A Genetically Encoded Fluorescent Amino Acid. *J. Am. Chem. Soc.* **2006**, *128* (27), 8738–8739. https://doi.org/10.1021/ja062666k.

(3) Soo Lee, H.; Guo, J.; Lemke, E. A.; Dimla, R. D.; Schultz, P. G. Genetic Incorporation of a Small, Environmentally Sensitive, Fluorescent Probe into Proteins in Saccharomyces Cerevisiae. *J. Am. Chem. Soc.* **2009**, *131* (36), 12921–12923. https://doi.org/10.1021/ja904896s.

(4) Soo Lee, H.; Spraggon, G.; Schultz, P. G.; Wang, F. Genetic Incorporation of a Metal-Ion Chelating Amino Acid into Proteins as a Biophysical Probe. *J. Am. Chem. Soc.* **2009**, *131* (7), 2481–2483. https://doi.org/10.1021/ja808340b.

(5) Speight, L. C.; Muthusamy, A. K.; Goldberg, J. M.; Warner, J. B.; Wissner, R. F.; Willi, T. S.; Woodman, B. F.; Mehl, R. A.; Petersson, E. J. Efficient Synthesis and in Vivo Incorporation of Acridon-2-Ylalanine, a Fluorescent Amino Acid for Lifetime and Förster Resonance Energy Transfer/Luminescence Resonance Energy Transfer Studies. *J. Am. Chem. Soc.* **2013**, *135* (50), 18806–18814. https://doi.org/10.1021/ja403247j.

(6) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. Designing Logical Codon Reassignment-Expanding the Chemistry in Biology. *Chem. Sci.* **2015**, *6* (1), 50–69. https://doi.org/10.1039/c4sc01534g.

(7) Xiao, H.; Schultz, P. G. At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code. *Cold Spring Harb. Perspect. Biol.* **2016**, *8* (9), 1–19. https://doi.org/10.1101/cshperspect.a023945.

(8) Moriya, T. Excited-State Reactions of Coumarins in Aqueous Solutions. I. The Phototautomerization of 7-Hydroxycoumarin and Its Derivative. *Bull. Chem. Soc. Jpn.* **1983**, *56* (1), 6–14. https://doi.org/10.1246/bcsj.56.6.

(9) Moriya, T. Excited-State Reactions of Coumarins in Aqueous Solutions. VI. Fluorescence Quenching of 7-Hydroxycoumarins by Chloride Ions in Acidic Solutions. *Bull. Chem. Soc. Jpn.* **1988**, *61* (3), 753–759. https://doi.org/10.1246/bcsj.61.753.

(10) Moriya, T. Excited-State Reactions of Coumarins. VII. The Solvent-Dependent Fluorescence of 7-Hydroxycoumarins. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 1873–1886. https://doi.org/10.1246/bcsj.61.1873.

(11) Georgieva, I.; Trendafilova, N.; Aquino, A.; Lischka, H. Excited State Properties of 7-Hydroxy-4-Methylcoumarin in the Gas Phase and in Solution. A Theoretical Study. *J. Phys. Chem. A* **2005**, *109* (51), 11860–11869. https://doi.org/10.1021/jp0524025.

(12) Georgieva, I.; Trendafilova, N.; Aquino, A. J. A.; Lischka, H. Excited-State Proton Transfer in 7-Hydroxy-4-Methylcoumarin along a Hydrogen-Bonded Water Wire. *J. Phys. Chem. A* **2007**, *111* (1), 127–135. https://doi.org/10.1021/jp0662202.

(13) Brun, M. P.; Bischoff, L.; Garbay, C. A Very Short Route to Enantiomerically Pure Coumarin-Bearing Fluorescent Amino Acids. *Angew. Chemie - Int. Ed.* **2004**, *43* (26), 3432–3436. https://doi.org/10.1002/anie.200454116.

(14) Chen, S.; Fahmi, N. E.; Wang, L.; Bhattacharya, C.; Benkovic, S. J.; Hecht, S. M. Detection of Dihydrofolate Reductase Conformational Change by FRET Using Two Fluorescent Amino Acids. *J. Am. Chem. Soc.* **2013**, *135* (35), 12924–12927. https://doi.org/10.1021/ja403007r.

(15) Ko, W.; Kim, S.; Lee, H. S. Engineering a Periplasmic Binding Protein for Amino Acid Sensors with Improved Binding Properties. *Org. Biomol. Chem.* **2017**, *15* (41), 8761–8769. https://doi.org/10.1039/c7ob02165h.

(16) Zhou, P.; Lv, P.; Yu, L.; Liu, S.; Zhang, L.; Tian, C. Fluorescence Lifetime Based Distance Measurement Illustrates Conformation Changes of PYL10-CL2 upon ABA Binding in Solution State. *Chinese Chem. Lett.* **2019**, *30* (5), 1067–1070. https://doi.org/10.1016/j.cclet.2019.01.020.

(17) Gleason, P. R.; Kelly, P. I.; Grisingher, D. W.; Mills, J. H. An Intrinsic FRET Sensor of Protein–Ligand Interactions. *Org. Biomol. Chem.* **2020**, *18* (21), 4079–4084. https://doi.org/10.1039/D0OB00793E.

(18) Huang, S.-M.; Yang, F.; Cai, B.-Y.; He, Q.-T.; Liu, Q.; Qu, C.-X.; Han, M.-J.; Kong, W.; Jia, Y.-L.; Li, F.; Yu, X.; Sun, J.-P.; Wang, J. Genetically Encoded Fluorescent Amino Acid for Monitoring Protein Interactions through FRET. *Anal. Chem.* **2019**, *91* (23), 14936–14942. https://doi.org/10.1021/acs.analchem.9b03305.

(19) Mills, J. H.; Lee, H. S.; Liu, C. C.; Wang, J.; Schultz, P. G. A Genetically Encoded Direct Sensor of Antibody-Antigen Interactions. *ChemBioChem* **2009**, *10* (13), 2162–2164. https://doi.org/10.1002/cbic.200900254.

(20) Henderson, J. N.; Simmons, C. R.; Fahmi, N. E.; Jeffs, J. W.; Borges, C. R.; Mills, J. H. Structural Insights into How Protein Environments Tune the Spectroscopic Properties of a Non-Canonical Amino Acid Fluorophore. *Biochemistry* **2020**, *59* (37), 3401–3410. https://doi.org/10.1021/acs.biochem.0c00474.

(21) Mendes, K. R.; Martinez, J. A.; Kantrowitz, E. R. Asymmetric Allosteric Signaling

in Aspartate Transcarbamoylase. *ACS Chem. Biol.* **2010**, *5* (5), 499–506. https://doi.org/10.1021/cb9003207.

(22) Dean, S. F.; Whalen, K. L.; Spies, M. A. Biosynthesis of a Novel Glutamate Racemase Containing a Site-Specific 7-Hydroxycoumarin Amino Acid: Enzyme-Ligand Promiscuity Revealed at the Atomistic Level. *ACS Cent. Sci.* **2015**, *1* (7), 364–373. https://doi.org/10.1021/acscentsci.5b00211.

(23) Saraogi, I.; Zhang, D.; Chandrasekaran, S.; Shan, S. O. Site-Specific Fluorescent Labeling of Nascent Proteins on the Translating Ribosome. *J. Am. Chem. Soc.* **2011**, *133* (38), 14936–14939. https://doi.org/10.1021/ja206626g.

(24) Lacey, V. K.; Parrish, A. R.; Han, S.; Shen, Z.; Briggs, S. P.; Ma, Y.; Wang, L. A Fluorescent Reporter of the Phosphorylation Status of the Substrate Protein STAT3. *Angew. Chemie* **2011**, *123* (37), 8851–8855. https://doi.org/10.1002/ange.201102923.

(25) Charbon, G.; Brustad, E.; Scott, K. A.; Wang, J.; Løbner-Olesen, A.; Schultz, P. G.; Jacobs-Wagner, C.; Chapman, E. Subcellular Protein Localization by Using a Genetically Encoded Fluorescent Amino Acid. *ChemBioChem* **2011**, *12* (12), 1818–1821. https://doi.org/10.1002/cbic.201100282.

(26) Charbon, G.; Wang, J.; Brustad, E.; Schultz, P. G.; Horwich, A. L.; Jacobs-Wagner, C.; Chapman, E. Localization of GroEL Determined by in Vivo Incorporation of a Fluorescent Amino Acid. *Bioorganic Med. Chem. Lett.* **2011**, *21* (20), 6067–6070. https://doi.org/10.1016/j.bmcl.2011.08.057.

(27) Cottam Jones, J. M.; Harris, P. W. R.; Scanlon, D. B.; Forbes, B. E.; Brimble, M. A.; Abell, A. D. Fluorescent IGF-II Analogues for FRET-Based Investigations into the Binding of IGF-II to the IGF-1R. *Org. Biomol. Chem.* **2016**, *14* (9), 2698–2705. https://doi.org/10.1039/c5ob02110c.

(28) Chen, S.; Fahmi, N. E.; Bhattacharya, C.; Wang, L.; Jin, Y.; Benkovic, S. J.; Hecht, S. M. Fluorescent Biphenyl Derivatives of Phenylalanine Suitable for Protein Modification. *Biochemistry* **2013**, *52* (47), 8580–8589. https://doi.org/10.1021/bi401275v.

(29) Kuhn, S. M.; Rubini, M.; Müller, M. A.; Skerra, A. Biosynthesis of a Fluorescent Protein with Extreme Pseudo-Stokes Shift by Introducing a Genetically Encoded Non-Natural Amino Acid Outside the Fluorophore. *J. Am. Chem. Soc.* **2011**, *133* (11), 3708–3711. https://doi.org/10.1021/ja1099787.

(30) Shi, P.; Zhang, Y.; Lv, P.; Fang, W.; Ling, S.; Guo, X.; Li, D.; Liu, S.; Sun, D.; Zhang, L.; Liu, D.; Zheng, J. S.; Tian, C. A Genetically Encoded Small-Size Fluorescent Pair Reveals Allosteric Conformational Changes of G Proteins upon Its Interaction with GPCRs by Fluorescence Lifetime Based FRET. *Chem. Commun.*

**2020**, *56* (51), 6941–6944. https://doi.org/10.1039/d0cc02691c.

(31) Amaro, M.; Brezovský, J.; Kováčová, S.; Sýkora, J.; Bednář, D.; Němec, V.; Lišková, V.; Kurumbang, N. P.; Beerens, K.; Chaloupková, R.; Paruch, K.; Hof, M.; Damborský, J. Site-Specific Analysis of Protein Hydration Based on Unnatural Amino Acid Fluorescence. *J. Am. Chem. Soc.* **2015**, *137* (15), 4988–4992. https://doi.org/10.1021/jacs.5b01681.

(32) Tolbert, L. M.; Solntsev, K. M. Excited-State Proton Transfer: From Constrained Systems to "Super" Photoacids to Superfast Proton Transfer. *Acc. Chem. Res.* **2002**, *35* (1), 19–27. https://doi.org/10.1021/ar990109f.

(33) Miyake-Stoner, S. J.; Miller, A. M.; Hammill, J. T.; Peeler, J. C.; Hess, K. R.; Mehl, R. A.; Brewer, S. H. Probing Protein Folding Using Site-Specifically Encoded Unnatural Amino Acids as FRET Donors with Tryptophan. *Biochemistry* **2009**, *48* (25), 5953–5962. https://doi.org/10.1021/bi900426d.

(34) Schuler, B.; Eaton, W. A. Protein Folding Studied by Single-Molecule FRET. *Curr. Opin. Struct. Biol.* **2008**, *18* (1), 16–26. https://doi.org/10.1016/j.sbi.2007.12.003.

(35) Davis, K. B.; Zhang, Z.; Karpova, E. A.; Zhang, J. Application of Tyrosine-Tryptophan Fluorescence Resonance Energy Transfer in Monitoring Protein Size Changes. *Anal. Biochem.* **2018**, *557* (April), 142–150. https://doi.org/10.1016/j.ab.2018.07.022.

(36) Kenworthy, A. K. Imaging Protein-Protein Interactions Using Fluorescence Resonance Energy Transfer Microscopy. *Methods* **2001**, *24* (3), 289–296. https://doi.org/10.1006/meth.2001.1189.

(37) Kim, J. H.; Sumranjit, J.; Kang, H. J.; Chung, S. J. Discovery of Coumarin Derivatives as Fluorescence Acceptors for Intrinsic Fluorescence Resonance Energy Transfer of Proteins. *Mol. BioSyst.* **2014**, *10* (1), 30–33. https://doi.org/10.1039/C3MB70323A.

(38) Lee, M. M.; Peterson, B. R. Quantification of Small Molecule–Protein Interactions Using FRET between Tryptophan and the Pacific Blue Fluorophore. *ACS Omega* **2016**, *1* (6), 1266–1276. https://doi.org/10.1021/acsomega.6b00356.

(39) Parks, J. W.; Stone, M. D. Coordinated DNA Dynamics during the Human Telomerase Catalytic Cycle. *Nat. Commun.* **2014**, *5* (1), 4146. https://doi.org/10.1038/ncomms5146.

(40) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Lakowicz, J. R., Ed.; Springer US: Boston, MA, 2006. https://doi.org/10.1007/978-0-387-46312-4.

(41) Berney, C.; Danuser, G. FRET or No FRET: A Quantitative Comparison. *Biophys. J.* **2003**, *84* (6), 3992–4010. https://doi.org/10.1016/S0006-3495(03)75126-1.

(42) Pollok, B. A.; Heim, R. Using GFP in FRET-Based Applications. *Trends Cell Biol.* **1999**, *9* (2), 57–60. https://doi.org/10.1016/S0962-8924(98)01434-2.

(43) Patel, M. J.; Yilmaz, G.; Bhatia, L.; Biswas-Fiss, E. E.; Biswas, S. B. Site-Specific Fluorescence Double-Labeling of Proteins and Analysis of Structural Changes in Solution by Fluorescence Resonance Energy Transfer (FRET). *MethodsX* **2018**, *5*, 419–430. https://doi.org/10.1016/j.mex.2018.03.006.

(44) Ferrie, J. J.; Ieda, N.; Haney, C. M.; Walters, C. R.; Sungwienwong, I.; Yoon, J.; Petersson, E. J. Multicolor Protein FRET with Tryptophan, Selective Coumarin-Cysteine Labeling, and Genetic Acridonylalanine Encoding. *Chem. Commun.* **2017**, *53* (80), 11072–11075. https://doi.org/10.1039/C7CC05492K.

(45) Wang, Z.; Talukder, P.; Hecht, S. M.; Chen, S. Fluorescent CD4 Probe for Potential HIV-1 Gp120 Protein Detection. *Bioorganic Med. Chem. Lett.* **2015**, *25* (6), 1182–1185. https://doi.org/10.1016/j.bmcl.2015.01.071.

(46) Talukder, P.; Chen, S.; Roy, B.; Yakovchuk, P.; Spiering, M. M.; Alam, M. P.; Madathil, M. M.; Bhattacharya, C.; Benkovic, S. J.; Hecht, S. M. Cyanotryptophans as Novel Fluorescent Probes for Studying Protein Conformational Changes and DNA-Protein Interaction. *Biochemistry* **2015**, *54* (51), 7457–7469. https://doi.org/10.1021/acs.biochem.5b01085.

(47) Ghisaidoobe, A.; Chung, S. Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy Transfer Techniques. *Int. J. Mol. Sci.* **2014**, *15* (12), 22518–22538. https://doi.org/10.3390/ijms151222518.

(48) The PyMOL Molecular Graphics System Version 1.8 Schrödinger LLC. The PyMOL Molecular Graphics System, Version 1.8. **2015**.

(49) Amemiya, T.; Koike, R.; Kidera, A.; Ota, M. PSCDB: A Database for Protein Structural Change upon Ligand Binding. *Nucleic Acids Res.* **2012**, *40* (D1), 554–558. https://doi.org/10.1093/nar/gkr966.

(50) Nishimasu, H.; Fushinobu, S.; Shoun, H.; Wakagi, T. Identification and Characterization of an ATP-Dependent Hexokinase with Broad Substrate Specificity from the Hyperthermophilic Archaeon Sulfolobus Tokodaii. *J. Bacteriol.* **2005**, *188* (5), 2014–2019.

(51) Nishimasu, H.; Fushinobu, S.; Shoun, H.; Wakagi, T. Crystal Structures of an ATP-Dependent Hexokinase with Broad Substrate Specificity from the Hyperthermophilic Archaeon Sulfolobus Tokodaii. *J. Biol. Chem.* **2007**, *282* (13), 9923–9931. https://doi.org/10.1074/jbc.M610678200.

(52) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An Enhanced System for

Unnatural Amino Acid Mutagenesis in E. Coli. *J. Mol. Biol.* **2010**, *395* (2), 361–374. https://doi.org/10.1016/j.jmb.2009.10.030.

(53) Le Trong, I.; Wang, Z.; Hyre, D. E.; Lybrand, T. P.; Stayton, P. S.; Stenkamp, R. E. Streptavidin and Its Biotin Complex at Atomic Resolution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (9), 813–821. https://doi.org/10.1107/S0907444911027806.

(54) Howarth, M.; Chinnapen, D. J. F.; Gerrow, K.; Dorrestein, P. C.; Grandy, M. R.; Kelleher, N. L.; El-Husseini, A.; Ting, A. Y. A Monovalent Streptavidin with a Single Femtomolar Biotin Binding Site. *Nat. Methods* **2006**, *3* (4), 267–273. https://doi.org/10.1038/nmeth861.

(55) Buckup, T.; Dorn, J.; Hauer, J.; Härtner, S.; Hampp, N.; Motzkus, M. The Photoinduced Cleavage of Coumarin Dimers Studied with Femtosecond and Nanosecond Two-Photon Excitation. *Chem. Phys. Lett.* **2007**, *439* (4–6), 308–312. https://doi.org/10.1016/j.cplett.2007.03.076.

(56) Callis, P. R.; Burgess, B. K. Tryptophan Fluorescence Shifts in Proteins from Hybrid Simulations: An Electrostatic Approach. *J. Phys. Chem. B* **1997**, *101* (46), 9429–9432. https://doi.org/10.1021/jp972436f.

(57) Oltrogge, L. M.; Boxer, S. G. Short Hydrogen Bonds and Proton Delocalization in Green Fluorescent Protein (GFP). *ACS Cent. Sci.* **2015**, *1* (3), 148–156. https://doi.org/10.1021/acscentsci.5b00160.

(58) Anderson, S.; Crosson, S.; Moffat, K. Short Hydrogen Bonds in Photoactive Yellow Protein. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (6), 1008–1016. https://doi.org/10.1107/S090744490400616X.

(59) Harris, C. M. Acid-Base Properties of 1-Naphthol.1 Proton-Induced Fluorescence Quenching. *J. Phys. Chem.* **1980**, *84* (11), 1366–1371. https://doi.org/10.1021/j100448a016.

(60) Pines, E.; Huppert, D.; Agmon, N. Geminate Recombination in Excited-State Proton Transfer Reactions: Numerical Solution of the Debye-Smoluchowski Equation with Backreaction and Comparison with Experimental Results. *J. Chem. Phys.* **1988**, *88* (9), 5620–5630. https://doi.org/10.1063/1.454572.

(61) Pines, E.; Fleming, G. R. Self Quenching of 1-Naphthol. Connection between Time-Resolved and Steady-State Measurements. *Chem. Phys.* **1994**, *183* (2–3), 393–402. https://doi.org/10.1016/0301-0104(94)00098-0.

(62) Leiderman, P.; Genosar, L.; Huppert, D.; Shu, X.; Remington, S. J.; Solntsev, K. M.; Tolbert, L. M. Ultrafast Excited-State Dynamics in the Green Fluorescent Protein Variant S65T/H148D. 3. Short- and Long-Time Dynamics of the Excited-

State Proton Transfer. *Biochemistry* **2007**, *46* (43), 12026–12036. https://doi.org/10.1021/bi7009053.

(63) Henderson, J. N.; Simmons, C. R.; Fahmi, N. E.; Jeffs, J. W.; Borges, C. R.; Mills, J. H. Structural Insights into How Protein Environments Tune the Spectroscopic Properties of a Non-Canonical Amino Acid Fluorophore. *Biochemistry* **2020**, acs.biochem.0c00474. https://doi.org/10.1021/acs.biochem.0c00474.

(64) Doose, S.; Neuweiler, H.; Sauer, M. A Close Look at Fluorescence Quenching of Organic Dyes by Tryptophan. *ChemPhysChem* **2005**, *6* (11), 2277–2285. https://doi.org/10.1002/cphc.200500191.

(65) Gehlen, M. H. The Centenary of the Stern-Volmer Equation of Fluorescence Quenching: From the Single Line Plot to the SV Quenching Map. *J. Photochem. Photobiol. C Photochem. Rev.* **2020**, *42*, 100338. https://doi.org/10.1016/j.jphotochemrev.2019.100338.

(66) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y. A.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popović, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P. Rosetta3; 2011; pp 545–574. https://doi.org/10.1016/B978-0-12-381270-4.00019-6.

(67) Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS One* **2011**, *6* (5), 1–12. https://doi.org/10.1371/journal.pone.0019230.

(68) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D. New Algorithms and an in Silico Benchmark for Computational Enzyme Design. *Protein Sci.* **2006**, *15* (12), 2785–2794. https://doi.org/10.1110/ps.062353106.

(69) Brzeska, H.; Venyaminov, S. V.; Grabarek, Z.; Drabikowski, W. Comparative Studies on Thermostability of Calmodulin, Skeletal Muscle Troponin C and Their Tryptic Fragments. *FEBS Lett.* **1983**, *153* (1), 169–173. https://doi.org/10.1016/0014-5793(83)80141-0.

(70) Crouchl, T. H.; Klee, C. B. Positive Cooperative Binding of Calcium to Bovine Brain Calmodulin. *Biochemistry* **1980**, *19* (16), 3692–3698. https://doi.org/10.1021/bi00557a009.

(71) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P.

Hratchian, J. V. Ort, and D. J. F. Gaussian 09. Gaussian, Inc.: Wallingford CT 2016.

(72) Renfrew, P. D.; Choi, E. J.; Bonneau, R.; Kuhlman, B. Incorporation of Noncanonical Amino Acids into Rosetta and Use in Computational Protein-Peptide Interface Design. *PLoS One* **2012**, *7* (3). https://doi.org/10.1371/journal.pone.0032637.

(73) Mills, J. H.; Khare, S. D.; Bolduc, J. M.; Forouhar, F.; Mulligan, V. K.; Lew, S.; Seetharaman, J.; Tong, L.; Stoddard, B. L.; Baker, D. Computational Design of an Unnatural Amino Acid Dependent Metalloprotein with Atomic Level Accuracy. *J. Am. Chem. Soc.* **2013**, *135* (36), 13393–13399. https://doi.org/10.1021/ja403503m.

(74) Koh, M.; Yao, A.; Gleason, P. R.; Mills, J. H.; Schultz, P. G. A General Strategy for Engineering Noncanonical Amino Acid Dependent Bacterial Growth. *J. Am. Chem. Soc.* **2019**, *141* (41), 16213–16216. https://doi.org/10.1021/jacs.9b08491.

(75) Bu, G.; Parrish, S.; Gleason, P. R.; Nielsen, D. R.; Nannenga, B. L. Heterologous Expression and Purification of the Bicarbonate Transporter BicA from Synechocystis Sp. PCC 6803. *Protein Expr. Purif.* **2020**, *175* (105716), 105716. https://doi.org/10.1016/j.pep.2020.105716.

(76) Gleason, P. R.; Nannenga, B. L.; Mills, J. H. Rapid Structural Analysis of a Synthetic Non-Canonical Amino Acid by Microcrystal Electron Diffraction. *Front. Mol. Biosci.* **2020**.

# APPENDIX A

# ADDITIONAL PUBLICATIONS

A. Introduction

In addition to the work that I presented as chapters in this thesis, I have also had the pleasure of either publishing in collaboration with other labs or on smaller side projects. Below the reader will find the following three works: **(1)** *A General Strategy for Engineering Noncanonical Amino Acid Dependent Bacterial Growth*, **(2)** *Heterologous expression and purification of the bicarbonate transporter BicA from Synechocystis sp. PCC 6803*, and **(3)** *Rapid Structural Analysis of a Synthetic Non-Canonical Amino Acid by Microcrystal Electron Diffraction.* A synopsis of each these works will be presented as the title of the published work, the author list, the abstract from the published manuscript, and a brief description of my contribution to the project.

B. A General Strategy for Engineering Noncanonical Amino Acid Dependent Bacterial Growth[74]

Minseob Koh,[†,#] Anzhi Yao,[†,#] Patrick R. Gleason,[‡] Jeremy H. Mills,[‡] and Peter G. Schultz[*,†]

*†Department of Chemistry and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States*

*‡School of Molecular Sciences and The Biodesign Center for Molecular Design and Biomimetics, Arizona State University, Tempe, Arizona 85281, United States*

Abstract

Synthetic auxotrophy in which bacterial viability depends on the presence of a synthetic amino acid provides a robust strategy for the containment of genetically modified organisms and the development of safe, live vaccines. However, a simple, general strategy

to evolve essential proteins to be dependent on synthetic amino acids is lacking. Using a temperature-sensitive selection system, we evolved an *Escherichia coli* (*E. coli*) sliding clamp variant with an orthogonal protein-protein interface, which contains a Leu273 to p-benzoylphenyl alanine (pBzF) mutation. The *E. coli* strain with this variant DNA clamp has a very low escape frequency ($<10^{-10}$), and its growth is strictly dependent on the presence of pBzF. This selection strategy can be generally applied to create NCAA dependence of other organisms with DNA clamp homologues.

Contribution

For this work, Professor Mills and I collaborated with the Schultz lab at the Scripps Research Institute. The Schultz lab was able to engineer (through directed evolution techniques) a DNA sliding clamp that was dependent on the NCAA p-benzoylphenyl alanine (pBzF), thus making the organism (in this case *E. coli*) dependent on pBzF. Our goal and contribution to this work was to facilitate structural models for two of their clamp homologues (named h1 and h5) that was consistent with the experimental data.

To accomplish this, we utilized the Rosetta macromolecular design suite. Because the DNA sliding clamp is C2 symmetric obligate homodimers, we utilized Rosetta's symmetric design algorithms which require only one of the monomers as input. By using only one chain from a high-resolution crystal structure (PDB ID: 4ovf, chain B; 2.05 Å resolution) we were able to accurately recapitulate the dimer complex to a 0.46 Å root mean squared deviation (RMSD) over all backbone atoms. Models of the h1 and h5 mutants were created through *in silico* mutagenesis, symmetrized, and scored using the Rosetta score function.

We found, in our models, that the mutations that had arisen through directed evolution had both reduced steric hinderance and had stabilized a favorable conformation of the pBzF. Specifically, in both the h1 and h5 variants, mutations to leu108 (cys in h1 and ala in h5) as well as the conserved mutations V70A and M97L provided space for the accommodation of the bulky phenyl group of pBzF.

While these models are qualitative in nature, it is hard to imagine a more appropriate structural model. This work helps to establish a method for developing organisms that are dependent on synthetic compounds. These organisms can then be used is various technologies (e.g. the development of benign bacterial species that can be used as "vaccines").

## C. Heterologous Expression and Purification of the Bicarbonate Transporter BicA from *Synechocystis sp.* PCC 6803[75]

Guanhong Bu,[a,b] Sydney Parrish,[a] Patrick R. Gleason,[c,d] David R. Nielsen,[a] Brent L. Nannenga[a,b,*]

[a]Chemical Engineering, School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, 85287, USA

[b]Center for Applied Structural Discovery, The Biodesign Institute, Arizona State University, Tempe, AZ, 85287, USA

[c]School of Molecular Sciences, Arizona State University, Tempe, AZ, 85287, USA

[d]Center for Molecular Design and Biomimetics, The Biodesign Institute, Arizona State University, Tempe, AZ, 85278, USA

Abstract

The high-flux/low-affinity cyanobacterial bicarbonate transporter BicA is a member of sulfate permease/solute carrier 26 (SulP/SLC26) family and plays a major role in cyanobacterial inorganic carbon uptake. In order to study this important membrane protein, robust platforms for over-expression and protocols for purification are required. In this work we have optimized the expression and purification of BicA from strain *Synechocystis sp.* PCC 6803 ($BicA_{6803}$) in Escherichia coli. It was determined that expression with C43 (DE3) Rosetta2 at 37 ºC produced the highest levels of over-expressed $BicA_{6803}$ relative to other strains screened, and membrane solubilization with n-dodecyl-β-D-maltopyranoside facilitated the purification of high levels of stable and homogenous $BicA_{6803}$. Using these expression and purification strategies, the final yields of purified BicA were 6.5 ± 1.0 mg per liter of culture.

Contribution

For this work, I collaborated with Professor Nannenga's student Guanhong Bu from Arizona State University. They required circular dichroism (CD) data to support the stability and secondary structure of the BicA6803 protein. This was the first description of the optimized expression and purification for this protein. CD data can be used to estimate how well a protein is folded and the amount of secondary structure in the protein (i.e. how much of the protein is α-helical). Here, I was able to show that the protein was not only well-folded but also primarily comprised of α-helices. This was consistent with descriptions of this protein in previously reported literature. My contributions were necessary for the successful publication of this paper as the Nannenga lab lacked the ability to answer reviewer comments without a CD spectrum.

## D. Rapid Structural Analysis of a Synthetic Non-canonical Amino Acid by Microcrystal Electron Diffraction[76]

Patrick R. Gleason[1,2], Brent L. Nannenga[3,4*], and Jeremy H. Mills[1,2*]

[1]School of Molecular Sciences, Arizona State University, Tempe, AZ 85287 USA

[2]Center for Molecular Design and Biomimetics, The Biodesign Institute, Arizona State University, Tempe, AZ, 85287 USA

[3]Chemical Engineering, School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ 85287 USA

[4] Center for Applied Structural Discovery, The Biodesign Institute, Arizona State University, Tempe, AZ, 85287 USA

### Abstract

Structural characterization of small molecules is a crucial component of organic synthesis. In this work, we applied microcrystal electron diffraction (MicroED) to analyze the structure of the product of an enzymatic reaction that was intended to produce the unnatural amino acid 2,4-dihydroxyphenylalanine (24DHF). Characterization of our isolated product with nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS) suggested that an isomer of 24DHF had been formed. Microcrystals present in the isolated product were used to determine its structure to 0.62 Å resolution, which confirmed its identity as 2-amino-2-(2,4-dihydroxyphenyl)propanoic acid (24DHPA). Moreover, the MicroED structural model indicated that both enantiomeric forms of 24DHPA were present. Notably, the entire structure determination process including setup, data collection, and refinement was completed in approximately one hour. The MicroED data not only bolstered previous results obtained using NMR and MS but

also immediately provided information about the stereoisomers present in the product, which is difficult or impossible to achieve using NMR or MS, respectively. Our results therefore demonstrate that MicroED methods can provide useful structural information on timescales that are only slightly longer than many commonly used analytical methods and could be added to the existing suite of small molecule structure determination tools in future studies.

Contribution

As the lead author of this work, I not only performed all of the experiments, but I also contributed significantly to the manuscript. The results presented in this paper were not expected. I had several peer-reviewed papers that suggested I could synthesize a resorcinol NCAA using a tyrosine phenol lyase enzyme from *Citrobacter freundii*. Interestingly, at no-point was I able to isolate the desired product. However, upon switching protocols to one that overloaded the system with the initial reactants I discovered I had synthesized an alpha-substituted resorcinol containing NCAA. With the help of Brent L. Nannenga, we were able to solve the crystal structure using micro electron diffraction techniques. While our initial results were disheartening, our ability to use the undesirable product of our reaction to highlight new techniques in rapid structure analysis was welcomed.