THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF CHEMISTRY

The Evolution of Allosteric Regulation in Chorismate Mutase

EVAN WILLIAM NELSON
SPRING 2022

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Premedicine
with honors in Chemistry

Reviewed and approved* by the following:

David Boehr, Ph.D.
Associate Professor of Chemistry
Thesis Supervisor

Elizabeth Elacqua, Ph.D.
Assistant Professor of Chemistry
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

Allostery, or "action at a distance," is important for the function of many enzymes, and can be thought of as the ways in which proteins communicate with their environment and express their cellular needs. Here, we use *Saccharomyces cerevisiae* chorismate mutase (ScCM) as a model system for studying allostery. ScCM is a homodimeric enzyme that is allosterically inhibited by tyrosine, allosterically activated by tryptophan, and displays substrate cooperativity. ScCM has sequence and structural similarity to other fungal and plant chorismate mutase enzymes, although not all of these enzymes are allosterically regulated in the same fashion. To gain information about the sequence determinants of allosteric regulation, I employed bioinformatic analyses to study sequence differences between fungal and plant chorismate mutases. However, the high degree of similarity between these enzymes did not provide much insight into their different allosteric regulation patterns. These results suggested that differences in allosteric regulation were not encoded only according to conserved amino acid residues, and more in-depth analysis (e.g. amino acid covariance analysis) may be required. I also further explored the ScCM T226I variant, which previous studies indicated was "locked" in the more active "R" conformation, using nuclear magnetic resonance spectroscopy to gain more insight into the solution characteristics of this variant. These studies indicated that contrary to the results of previous studies, the introduction of the T226I mutation causes structural changes (e.g. misfolding) to the ScCM protein that are observable in NMR spectra.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Throughout the span of my academic career, it has been my family that has served as the foundation of all successes I have achieved. My mother, Jodi, and my father, Bill, have presented me with the incredible gift of knowledge while instilling in me moral and ethical values that will serve me a lifetime. Without my parents, I would have nothing.

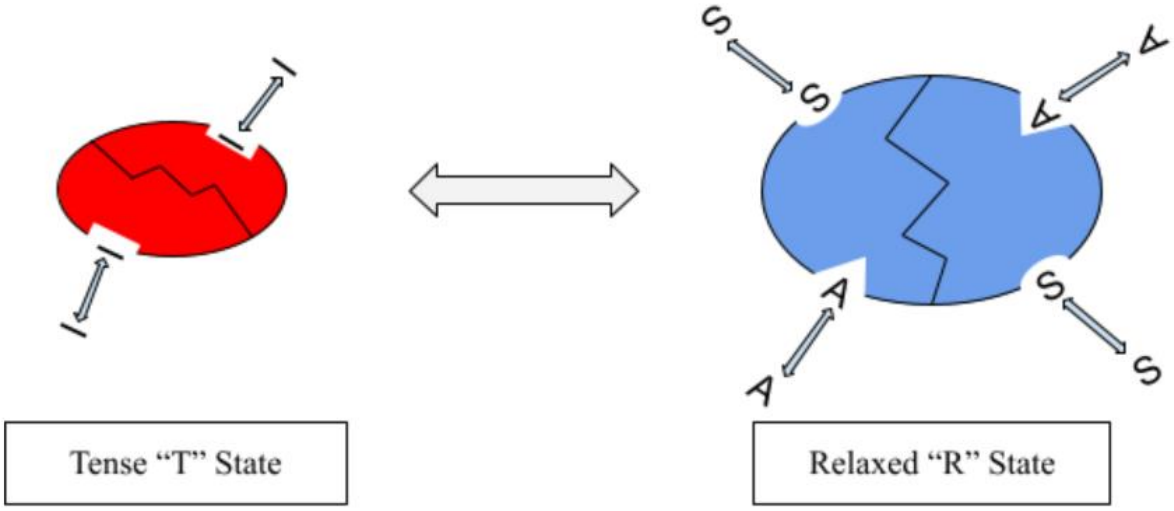
In addition to my parents, I have had the great fortune of being surrounded by other family members and friends who have helped shape the man I have become. In particular, I would like to acknowledge my younger brother, Dylan, for his influences on my maturation. Although he is my junior by five years, growing up with him has served as an incredible learning experience. I hope that I may always embody what constitutes the ideal older brother.

Finally, I would like to acknowledge each professor and mentor I met while studying at Penn State. During my time as an undergraduate researcher, I have had the great pleasure of working alongside David Boehr, Ph.D. and doctoral student Dennis Winston. Without their tutelage, I would never have been capable of creating this thesis.

Chapter 1 : Introduction

Allostery is essential to enzyme regulation and can be described as when remote sites of a protein are energetically connected to generate a functional response (1). Allostery is observed in a host of biological processes, including metabolism, gene regulation, signal transduction, and catalysis. The early models of allostery include the Monod-Wyman-Changeux (MWC) model and the Koshland–Némethy–Filmer (KNF) model, in which they attempt to provide an explanation of the allosteric functioning of multimeric proteins (1).

The MWC model is also called the "concerted model,' and it describes two interconvertible states that are referred to as the "T" (tense) and "R" (relaxed) states (2). The "T" and "R" function in thermal equilibrium with all subunits of an oligomeric protein sharing the same state (see **Figure 1.1**).



*I is allosteric inhibitor, S is substrate, A is allosteric activator

Figure 1.1: MWC model in which there is a conformational equilibrium between the less active "T" state and more active "R" state

The KNF model is also called the "sequential model," and it describes subunits in one oligomeric protein that can adopt different states (3). The binding of the first allosteric effector results in a local conformation change in one subunit, which then forces conformational change in neighboring subunits (see **Figure 1.2**). Affinity for the allosteric effector in the other binding sites is experienced, which results in cooperativity of binding (3).

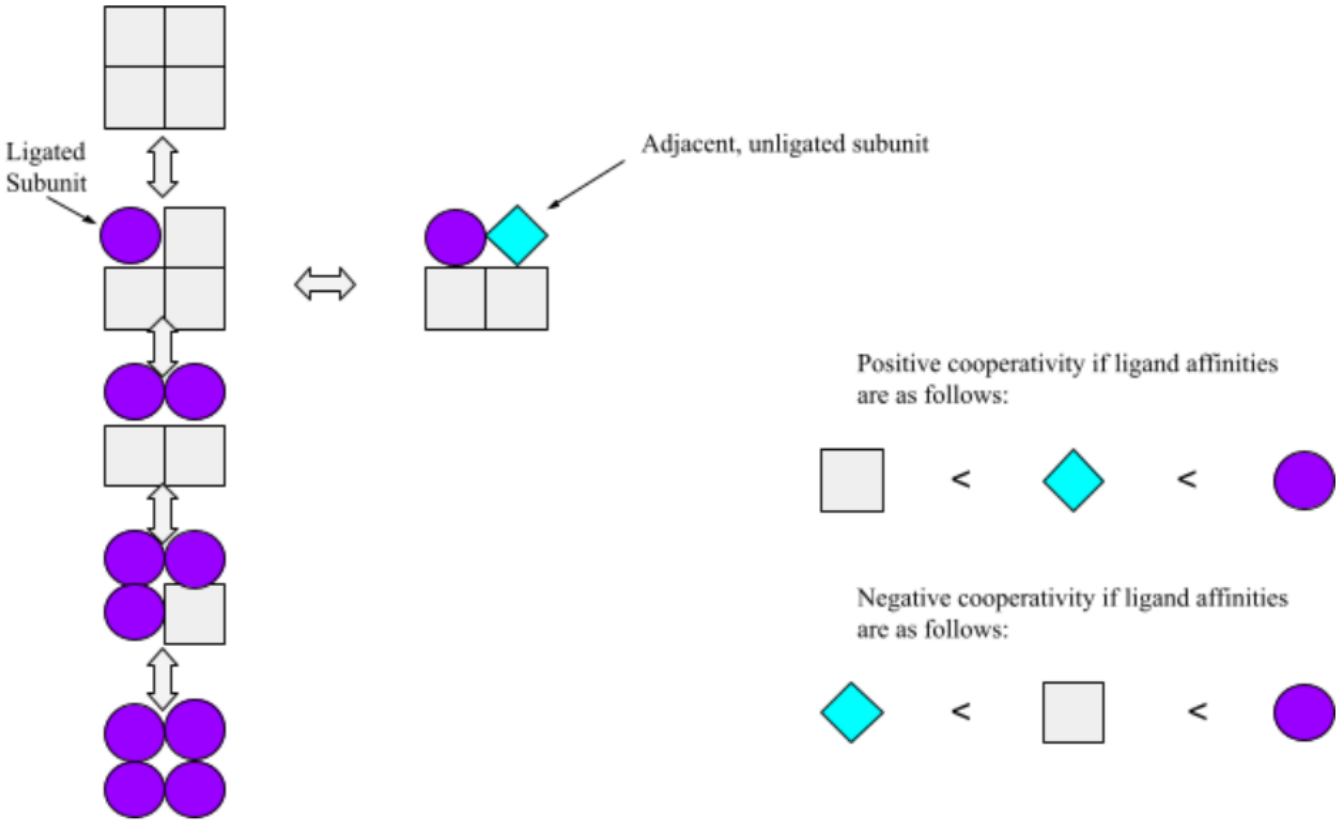


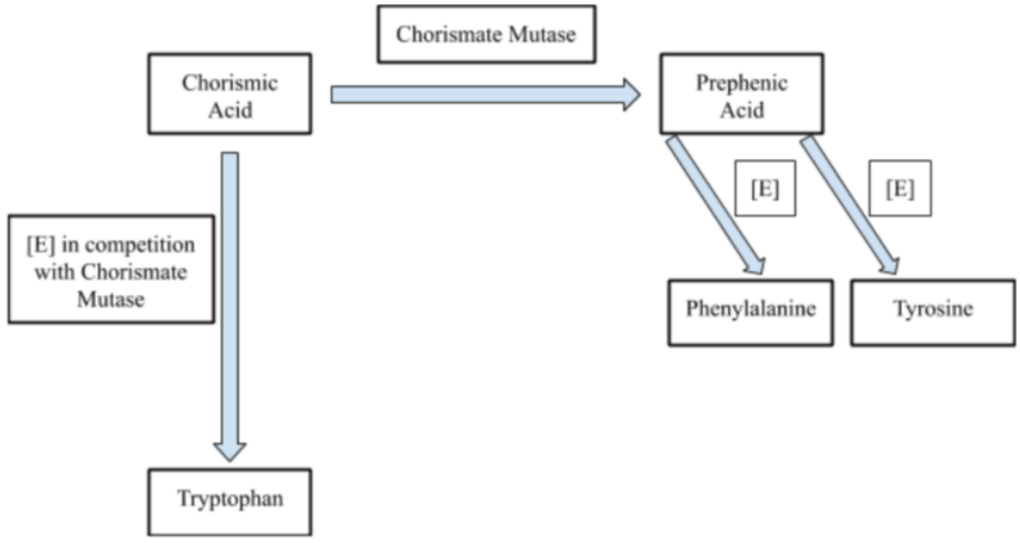
Figure 1.2: KNF model in which individual subunits can be in a "T" or "R" state

One model system to better understand allostery is the chorismate mutase enzyme from *Saccharomyces cerevisiae*, which I will term ScCM. ScCM exhibits a range of allosteric behavior, including positive homotropic cooperativity (i.e. binding of chorismate substrate at one active site

enhances binding of chorismate at the other active site) and negative heterotropic cooperativity (i.e. binding of tyrosine at the allosteric site inhibits chorismate binding). The relatively small size of ScCM compared to other allosteric enzymes (ScCM is only ~ 60 kDa) also makes it more easily tractable via nuclear magnetic resonance (NMR) spectroscopy.

Chorismate mutase (CM) is responsible for facilitating the enzymatic conversion of chorismic acid to prephenic acid (see **Figure 1.3**). Prephenic acid can be subsequently transformed into phenylalanine and tyrosine (see **Figure 1.3**). It then makes metabolic sense that tyrosine (or Tyr) is an allosteric inhibitor of ScCM (and phenylalanine, or Phe, is an allosteric inhibitor for other CM enzymes). Chorismate is also used by the enzyme anthranilate synthase towards production of the other aromatic amino acid, tryptophan. Tryptophan (or Trp) allosterically activates ScCM via the same allosteric binding site as Tyr.

.



*Competing enzyme is anthranilate synthase. This enzyme makes anthranilic acid from chorismic acid, which then is transformed into tryptophan by various tryptophan enzymes.

Figure 1.3: Progression of biochemical reactions involving chorismate mutase and related enzymes

There are different classes of chorismate mutases, the AroH class and the AroQ class. For the purposes of this study, the AroQ class will serve as the main focus. In particular, ScCM is an AroQβ chorismate mutase. The quaternary structure of this type of chorismate mutase is a homodimer composed of two symmetrical 30 kDa monomers (4) (see **Figure 1.4**). Each of these monomers is composed of ten α-helices, two 3_{10} -helices, four loop components in the region where the effector binds, and a disordered loop (L220s loop). The L220s loop encompasses residues 221 through 226 of the enzyme amino acid sequence. Relatively strong van der Waals interactions are what keep the two monomers together, leading to a dimer the shape of a bipyramid (4).

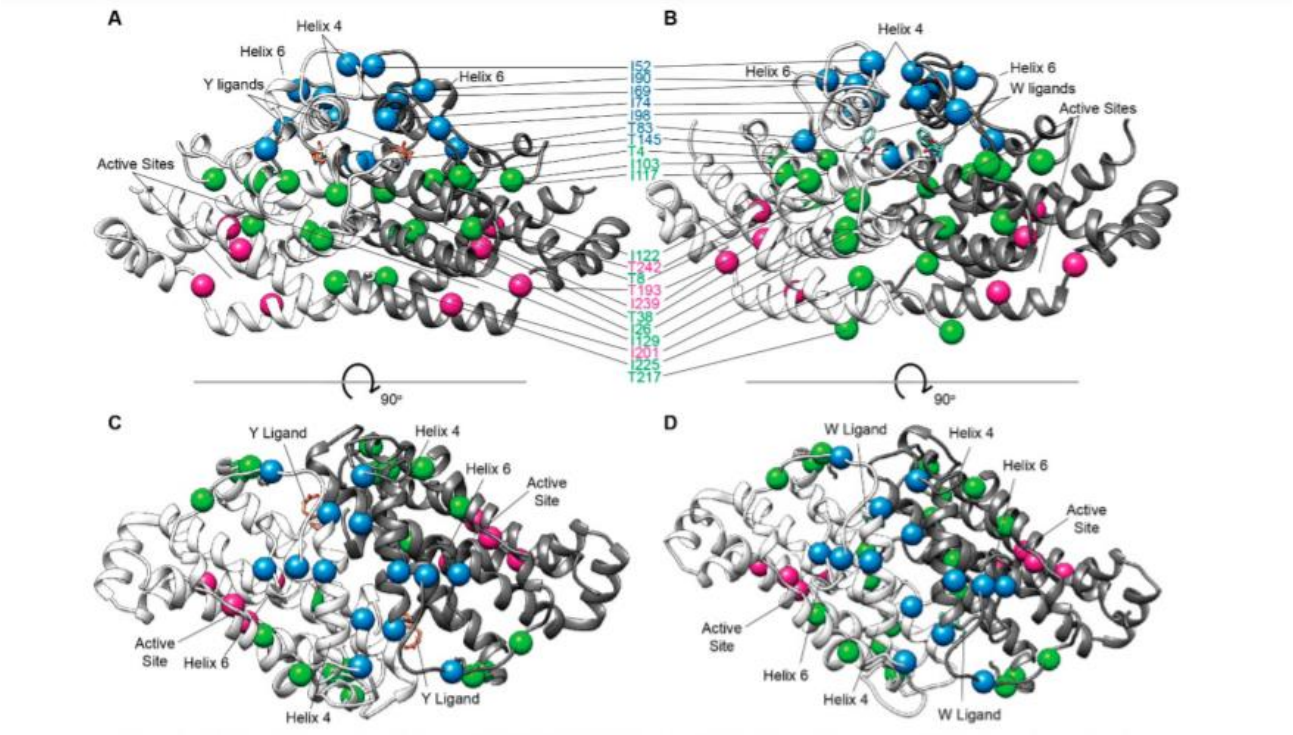


Figure 1.4: Structures of the "T" and "R" states of ScCM. A and C. X-ray crystal structures of the allosterically inhibited "T" conformation in which Tyr is bound. B and D. X-ray crystal structures of the allosterically activated "R" conformation in which Trp is bound. Adapted with permission from reference (18). Copyright (2019) American Chemical Society.

X-ray crystal structure analysis has suggested that ScCM follows a simple MWC-type model with "T" and "R" states (see **Figure 1.4** (18)). However, studies – especially those studies that focus in on the L220s loop region of ScCM – have indicated that a simple R-T conformation change with ScCM may not accurately describe the conformational changes that actually transpire (20). In particular, the I225T/T226I ScCM variant behaves like wild-type ScCM when allosteric effector Trp is present, but it is not inhibited by Tyr, which indicates that a destabilized "T" conformation does not allow for a stabilized "R" conformation when lacking allosteric effector (20). Additionally, Tyr/Trp both display negative homotropic cooperativity (i.e. binding of one Tyr inhibits the binding of another Tyr, so that ScCM bound with only one Tyr is populated in solution), which the MWC model cannot explain (18).

Other evidence against the MWC model has come from relaxation dispersion NMR experiments, which measure conformational exchange on the microsecond-millisecond timescale. On this timescale, studies were not consistent with ScCM fluctuating between "T" and "R" states (21).

Thus, we have sufficiently established that ScCM serves as an appropriate model system for understanding allostery and exhibits a range of allosteric behaviors. The MWC model also does not quite explain all of the allosteric behaviors exhibited by ScCM, and so a more complex model is necessary. It is also known that other fungal and plant CM enzymes are allosterically regulated differently. In Chapter 2, I performed bioinformatic analysis to gain more insight into possible sequence determinants that differentiate allosteric behaviors. This analysis was performed during the COVID-19 pandemic, as I had limited access to the lab space. In Chapter 3, I conducted NMR investigations of the T226I variant to better understand why this variant is "locked" into the active "R" conformation. These forms of experimentation ultimately helped contribute to a growing

understanding of ScCM as an allosterically-regulated enzyme and provided insight into the evolution of ScCM's allosteric behaviors.

Chapter 2 : Bioinformatic analysis of chorismate mutase enzymes to better understand allosteric mechanisms

2.1 Introduction

This chapter focuses on the results of bioinformatic pursuits, and in particular, the bioinformatics study of chorismate mutase enzymes of different species of yeasts and plants. Chorismate mutases are allosterically-regulated enzymes that serve as appropriate model enzymes for bioinformatic study. ScCM, in particular, displays allosteric behaviors, including positive homotropic cooperativity and negative heterotropic cooperativity. Although ScCM displays these forms of cooperativity, not all organisms' chorismate mutases behave in the same fashion. Trp, Phe, and Tyr need not always act as activators and inhibitors of chorismate mutase; some organisms' CM enzymes are impacted by effector binding while others are left largely unimpacted. Most of the CMs studied here are regulated by Phe and Tyr binding (e.g. *Arxula adeninivorans, Candida maltosa,* and others), however there are a select few CM enzymes that are not regulated by binding of Phe or Tyr (e.g. *Candida albicans, Candida tropicalis,* and others).

In this chapter, I attempted to identify amino acid sequence determinants of these allosteric patterns through bioinformatic analyses. The yeast and plant chorismate mutases were separated into different groups for analysis, as these species-specific enzymes display differences in allosteric behaviors. A closer look was taken at different residues and enzyme/substrate interactions to determine different modes of enzyme regulation. In particular, focus was directed to those residues that are in contact with ligands, including Trp, Tyr, and the transition state analog (TSA), which is 8-hydroxy-2-oxa-bicyclo[3.3.1]non-6-ene-3,5-dicarboxylic acid. This analysis

was accomplished by using multiple sequence alignments; multiple sequence alignments are alignments of three of more biochemical sequences that allow for inference of evolutionary relationships between the sequences included. Exploring the organism-specific regulation of chorismate mutase has proven to be a necessary step in garnering a greater understanding of this important enzyme.

2.2 Methods

Species (yeasts and plants) for sequence analysis were chosen after careful review of the existing literature (14, 19). Sequences for these chorismate mutases were then acquired using UniProt (8) and NCBI Protein Blast (9). Following acquisition of sequences, multiple sequence alignments were carried out via use of NCBI Protein Blast (9) using standard parameters. The results were analyzed using the residues important for ScCM; these important residues were determined via review of the literature (6, 16, 17).

2.3 Results

2.3.1 Determination of important chorismate mutase residues

The important residues for ScCM were determined using a combination of literature (6, 16, 17), database resources (10, 11), and structural analyses. Here, I used the PDB protein databank accessions 2CSM, 3CSM, and 4CSM. 2CSM represents the Tyr-bound "T" state, 3CSM represents the Trp-bound "R" state bound with transition state analog (TSA), and 4CSM has Tyr and TSA bound. Residues deemed important for binding Tyr and Trp at the allosteric site of ScCM (**Table**

2.1) differ from those residues deemed important in binding the TSA at the active site of ScCM (**Table 2.2**). These important residues were determined based upon the distance of polar interactions between ScCM and Tyr/Trp/TSA. Each of these important residues can be viewed spatially within the ScCM enzyme model as well (see **Figure 2.1**).

Table 2.1: Important ScCM residues for interacting with tyrosine and tryptophan at the allosteric site based on PDB 2CSM and PDB 3CSM. This table was created, in part, using data provided by reference (16).

ScCM Residue	Tyr/Trp Atom	Tyr Distance (Å)	Trp Distance (Å)
Arg-75 (N atom)	O	2.7 and 3.2	5.0 and 5.0
Arg-76 (N atom)	OH	3.3	
Asn-138 (O atom)	N	3.8	3.3
Asn-139 (O atom)	N	3.5	3.2
Gly-141 (N atom)	O	2.8	3.0
Ser-142 (N atom)	O	2.8	3.0
Ser-142 (O atom)	N	2.4	2.8
Thr-145 (O atom)	OH	3.1	

Table 2.2: Important ScCM residues for interacting with transition state analog (TSA) at the active site based on PDB 3CSM and PDB 4CSM.

ScCM Residues for TSA Contacts of 3CSM	ScCM Residues for TSA Contacts of 4CSM
Arg-157	Arg-16
Lys-168	Arg-157
Ile-192	Lys-168
Asn-194	Asn-194
Glu-198	Glu-198
Lys-243	Ile-239
	Lys-243

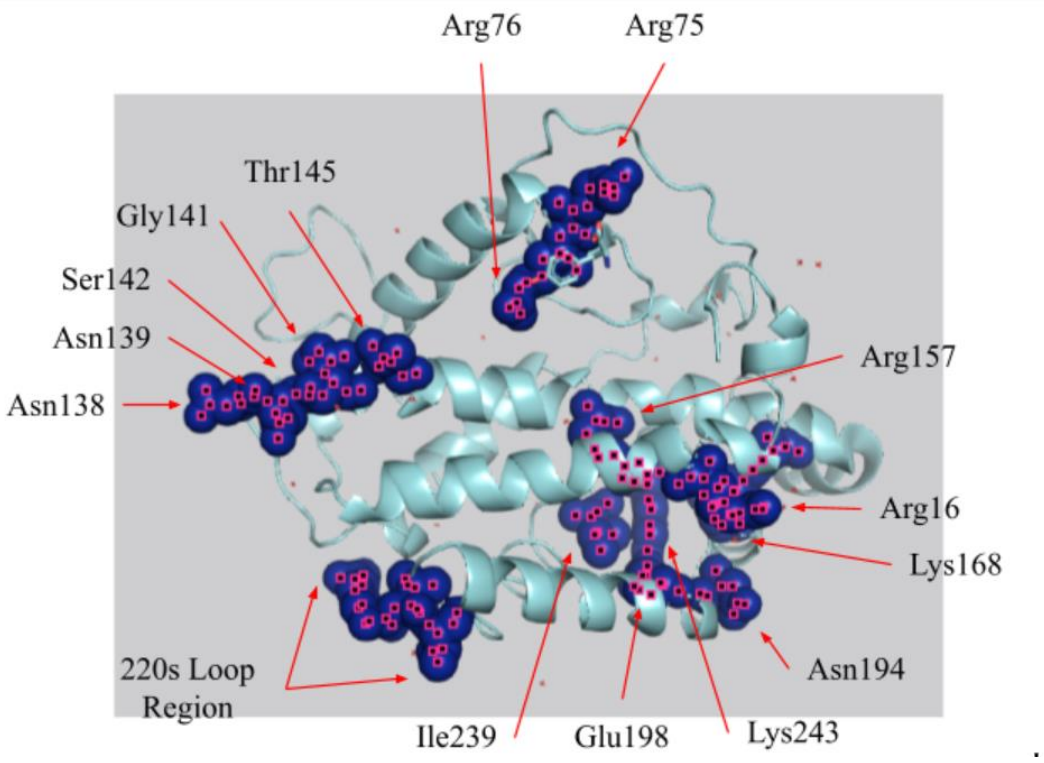


Figure 2.1: Three-dimensional structure of ScCM with important residues labeled

2.3.2 Acquisition and characterization of yeast chorismate mutase sequences

In order to conduct multiple sequence alignments for the chorismate mutase sequences of different organisms, it was necessary to determine the sequences of different organisms' CMs. Sequences were selected for three major reasons: 1) each of these yeasts contains only one form of the chorismate mutase enzyme, and that enzyme is located solely within the cytosol (19); 2) each of these yeasts' chorismate mutases has existing data relating to their activities in the presence of Phe and Tyr (19); 3) each of these organisms possesses a chorismate mutase enzyme whose sequence is readily accessible. Based on these criteria, I collected the following sequences using UniProt (8) and NCBI Protein Blast (9): *Arxula adeninivorans, Candida albicans, Candida maltosa, Candida parapsilosis, Candida tropicalis, Candida utilis, Debaryomyces hansenii, Hansenula fabianii, Hansenula polymorpha, Kluyveromyces marxianus, Pichia guilliermondii/Meyerozyma guillermondii, Saccharomyces cerevisiae, Saccharomycopsis capsularis, Schizosaccharomyces octosporus, Sporobolomyces salmonicolor, Yarrowia lipolytica.* The FASTA sequence of each organism's chorismate mutase was recorded in a large document for future use in multiple sequence alignments.

For the organisms whose sequence was obtained, most have chorismate mutases with activities that are impacted by binding of Phe or Tyr. Of the included chorismate mutases, the following organisms' enzymes are inhibited or activated by Phe or Tyr binding: *Arxula adeninivorans, Candida maltosa, Candida parapsilosis, Candida utilis, Hansenula fabianii, Hansenula polymorpha, Kluyveromyces marxianus, Pichia guilliermondii/Meyerozyma guillermondii, Saccharomyces cerevisiae, Saccharomycopsis capsularis, Schizosaccharomyces octosporus, Sporobolomyces salmonicolor, Yarrowia lipolytica.* Thus, the following organisms'

enzymes are not inhibited or activated by Phe or Tyr binding: *Candida albicans, Candida tropicalis, Debaryomyces hansenii*.

2.3.3 Multiple sequence alignment analysis for yeast chorismate mutases

The residues we identified to be most important for ScCM functioning (**Tables 2.1-2.2**) were used as a focal point for analyzing multiple sequence alignments. These alignments were performed using NCBI Protein Blast (9). To effectively organize the data obtained from these complex alignments (see **Appendix A**), a table was generated (**Table 2.3**). Cells that contain "*" indicate that the sequence for that organism's chorismate mutase did not have a residue at the given position.

Table 2.3: Multiple sequence alignments for all yeast sequences

	Residue at ScCM Location													
	Allosteric Site Residues							**Active Site Residues**						
Organism	**Arg 75**	**Arg 76**	**Asn 138**	**Asn 139**	**Gly 141**	**Ser 142**	**Thr 145**	**Arg 16**	**Arg 157**	**Lys 168**	**Asn 194**	**Glu 198**	**Ile 239**	**Lys 243**
CM enzymes that are regulated by Phe and/or Tyr														
A. adeninivorans	R	R	E	N	G	S	I	R	R	K	N	E	I	K
C. maltosa	R	R	E	N	G	S	T	R	R	K	N	E	I	K
C. parapsilosis	R	R	E	N	G	S	T	R	R	K	N	E	I	K
C. utilis	R	R	E	N	G	S	T	R	R	K	N	E	I	K
H. fabianii	R	R	E	N	G	S	V	R	R	K	N	E	I	K
H. polymorpha	R	R	D	N	G	S	M	R	R	K	N	E	I	K
K. marxianus	R	R	E	N	G	S	T	R	R	K	N	E	I	K
P./M.guillermondii	R	R	E	N	G	S	V	R	R	K	N	E	I	K
S. capsularis	R	R	E	N	G	S	L	R	R	K	N	E	I	K
S. octosporus	R	R	D	N	G	S	V	R	R	K	D	E	I	K
S. salmonicolor	R	R	D	N	G	S	T	*	R	K	K	E	I	K
Y.lipolytica	R	R	E	N	G	S	V	R	R	K	N	E	I	K
CM enzymes that are NOT regulated by Phe and/or Tyr														
C. albicans	R	R	E	N	G	S	T	R	R	K	N	E	I	K
C. tropicalis	R	R	E	N	G	S	T	R	R	K	N	E	I	K
D. hansenii	R	R	E	N	G	S	V	R	R	K	N	E	I	K

Upon analysis of the multiple sequence alignments for those yeast species whose chorismate mutases are impacted by Phe/Tyr binding and those species whose chorismate mutases are not impacted by Phe/Tyr binding (**Table 2.3**), it became apparent that there is a high degree of conservation at most residue locations.

Of the residue locations (**Table 2.3**), it is important to discuss some residues that are of particular importance to ligand binding, and thus, allosteric regulation. In ScCM, Arg75 associates

with Tyr via hydrogen bonding. After aligning all sequences, it was shown that all other yeasts' chosrismate mutases possessed an Arg at position 75 as well. The exact same is true for Arg76, which is another residue that hydrogen bonds to Tyr in ScCM.

Another set of residues important to consider is the Glu23, Arg157, and Tyr234 grouping. While these residues are numerically distant in primary amino acid sequence, they are close spatially. In ScCM, Glu23 forms two interactions: 1) Glu23 associates with Arg157 via hydrogen bonding; 2) Glu23 associates with Tyr234 in the active site via hydrogen bonding when Tyr is bound, but not when Trp is bound. These Glu23 interactions with other ScCM amino acids contribute to the three-dimensional structure of ScCM, and thus, it was important to study conservation at these locales. After analyzing the multiple sequence alignments, there is total conservation at Glu23 (all Glu except for the chorismate mutase of *Schizosaccharomyces octosporus*, whose sequence was missing a residue at this locations), there is total conservation at Arg157, and there is total conservation at Tyr234. This high level of conservation allows us to make the assumption that these interactions are necessary for proper folding and functioning of all chorismate mutases, regardless of the species from which they are derived.

Thr145 is yet another important residue location, as it is responsible for associating with Tyr via hydrogen bonding in ScCM. Due to the important nature of this residue, it was analyzed alongside supplemental information about whether the enzyme is impacted by Phe and Tyr binding (**Table 2.4**). Interestingly, there was not total conservation. This lack of conservation is quite intriguing, and Thr145 may serve as a key residue in future bioinformatic studies (e.g homology modeling and normal mode analysis) of chorismate mutase enzymes. It is likely that these future studies would be necessary given the apparent lack of any pattern present in the multiple sequence alignment for Thr145.

Table 2.4: Lack of conservation displayed in the Thr145 alignment

Organism	Thr145	Possibility of H Bond	Inhibited by Phe?	Inhibited by Tyr?
CM enzymes that are regulated by Phe and/or Tyr				
A. adeninivorans	I	No	Yes	No
C. maltosa	T	Yes	No	Yes
C. parapsilosis	T	Yes	No	Yes
C. utilis	T	Yes	Yes	Yes
H. fabianii	V	No	Yes	Yes
H. polymorpha	M	No	Yes	Yes
K. marxianus	T	Yes	Yes	No
P./M.guillermondii	V	No	Yes	Yes
ScCM	T	Yes	No	Yes
S. capsularis	L	No	Yes	Yes
S. octosporus	V	No	Yes	Yes
S. salmonicolor	T	Yes	Yes	Yes
Y.lipolytica	V	No	No (Activated above 100% activity)	No (Activated above 100% activity)
CM enzymes that are NOT regulated by Phe and/or Tyr				
C. albicans	T	Yes	No	No
C. tropicalis	T	Yes	No	No
D. hansenii	V	No	No	No

The 220s loop region in ScCM is functionally important, and in particular, the Thr226 residue provides insight into allosteric regulation. Thr226, when it is substituted for an Ile, locks ScCM into an active "R" state. But while the L220s loop region is important for catalysis of chorismic acid to prephenic acid, there is not much conservation among these residues (**Table 2.5**). When trying to find a clear pattern in the multiple sequence alignments of the 220s loop region (**Table 2.5**), such a pattern was unable to be elucidated. This may indicate that the importance of the 220s loop region is greatest in ScCM, and other chorismate mutases do not rely as heavily upon this region for functioning.

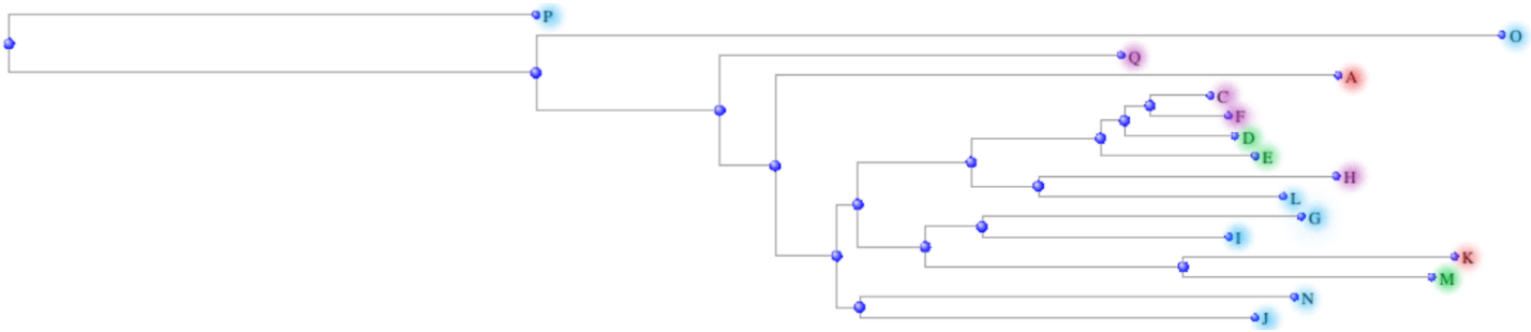
Table 2.5: Multiple sequence alignments of yeast sequences in 220s loop region

	Residue at ScCM Location						
Organism	**Ser 220**	**Gly 221**	**Glu 222**	**Arg 223**	**Arg 224**	**Ile 225**	**Thr 226**
CM enzymes that are regulated by Phe and/or Tyr							
A. adeninivorans	W	S	Q	G	N	I	S
C. maltosa	*	*	Q	S	K	I	E
C. parapsilosis	Y	G	Q	S	K	V	K
C. utilis	W	S	P	S	K	V	K
H. fabianii	L	A	Q	G	K	V	N
H. polymorpha	F	T	Q	S	K	V	K
K. marxianus	Q	G	D	K	K	I	T
P./M.guillermondii	Y	S	Q	S	K	V	R
S. capsularis	T	T	Q	S	K	V	K
S. octosporus	T	K	T	D	K	I	N
S. salmonicolor	E	G	P	M	R	V	D
Y.lipolytica	W	S	Q	G	K	V	D
CM enzymes that are NOT regulated by Phe and/or Tyr							
C. albicans	F	G	Q	S	K	V	K
C. tropicalis	F	G	Q	S	K	V	K
D. hansenii	Y	S	Q	S	K	I	Q

The high degree of conservation among yeast sequences regardless of how the enzyme is impacted by effector binding provides little insight into the biochemical basis for differences in effector binding. However, this extremely high degree of residue consistency does eliminate sequence as the primary determining factor for enzyme-effector specificity.

The similarities in sequences were further analyzed via the creation of a phylogenetic tree (see **Figure 2.2**) using the NCBI Protein Blast Cobalt phylogenetic tree widget (12). Species that are closest together on phylogenetic trees share a recent common ancestor, which implies that

we can use this tree as a means of studying evolutionary changes that resulted in changes to chorismate mutase. While the motivation for generating this phylogenetic tree was to explore similarities in enzyme sequence in the hopes that these similarities would match with groupings according to allosteric behavior, this was not accomplished with complete success. Upon looking at the tree (see **Figure 2.2**), one can see that there are some areas in which CM enzymes are grouped according to regulation behavior, but there are many exceptions to this rule.



*blue = inhibited by both Phe and Tyr; red = inhibited by Phe only; green = inhibited by Tyr only; purple = not inhibited

Figure 2.2: Phylogenetic tree for yeast sequences groups CM enzymes based on sequence similarity, but only partially according to regulation patterns

2.3.4 Acquisition and characterization of plant chorismate mutase sequences

Following analysis of the yeast chorismate mutase sequences, a small sample of plant CM sequences was studied. Plant chorismate mutase enzymes were chosen for analysis after receiving less-than-satisfactory results when using yeast CMs as study subjects. It was our hope that plant CMs could provide us with more sequence diversity so that we could derive more knowledge about the species-specific allosteric differences in CM enzymes. The methodology of study was identical to the procedure used for the yeast species. The following plant species were obtained from the literature (14) using the same selection criteria as were used for the yeasts, and their sequences were found via UniProt (15): *Arabidopsis thaliana* AtCM1 - allosterically regulated, *Arabidopsis*

thaliana AtCM2 – unregulated; *Arabidopsis thaliana* AtCM3 - allosterically regulated; *Physcomitrella patens* PpCM1/PpCM2 - both allosterically regulated; *Selaginalla moellendorffii* SmCM - allosterically regulated; *Amborella trichopoda* AmtCM1/AmtCM2 - both allosterically regulated.

2.3.5 Multiple sequence alignment analysis for plant chorismate mutases

Sequence alignments (**Tables 2.6-2.7**) were performed for the plant chorismate mutase sequences. This was accomplished in the same fashion as the yeast enzyme analysis. Mostly, widespread conservation was observed, and any deviations were analyzed.

Table 2.6: Multiple Sequence Alignments for Plant Sequences

	Residue at ScCM Location													
	Allosteric Site Residues							**Active Site Residues**						
Organism	**Arg 75**	**Arg 76**	**Asn 138**	**Asn 139**	**Gly 141**	**Ser 142**	**Thr 145**	**Arg 16**	**Arg 157**	**Lys 168**	**Asn 194**	**Glu 198**	**Ile 239**	**Lys 243**
CM enzymes that are allosterically regulated														
A. thaliana AtCM1	G	R	G	N	G	S	V	R	R	K	F	E	M	K
A. thaliana AtCM3	D	R	G	N	G	S	L	R	R	K	Y	E	M	K
P. patens PpCM1/ PpCM2	R	R	G	N	G	S	V	R	R	K	F	E	M	K
S.moellendorffii SmCM	R	R	G	N	G	S	A	R	R	K	F	E	M	K
A. trichopoda AmtCM1/AmtCM2	G	R	G	N	G	S	V	R	R	K	Y	E	M	K
CM enzymes that are NOT allosterically regulated														
A. thaliana AtCM2	G	R	G	N	P	S	A	R	R	K	F	E	I	K

Table 2.7: Multiple sequence alignments for plant sequences within 220s loop region

	Residue at ScCM Location						
Organism	**Ser 220**	**Gly 221**	**Glu 222**	**Arg 223**	**Arg 224**	**Ile 225**	**Thr 226**
CM enzymes that are allosterically regulated							
A. thaliana AtCM1	M	E	V	Y	K	I	S
A. thaliana AtCM3	*	D	S	Y	K	I	Q
P. patens PpCM1/ PpCM2	*	E	R	Y	K	I	D
S.moellendorffii SmCM	*	A	L	Y	K	V	E
A. trichopoda AmtCM1/AmtCM2	K	G	N	R	K	I	D
CM enzymes that are NOT allosterically regulated							
A. thaliana AtCM2	*	G	K	Y	K	V	D

As mentioned previously, hydrogen bonding was used as an analytical parameter for the sequence alignments, as the allosteric effectors interact with the chorismate mutase enzyme via hydrogen bonds. Among the Arg16, Arg76, Asn139, Ser142, Arg157, Lys168, Glu198, and Lys243 residues, complete conservation was observed amongst all plant chorismate mutases. The Asn138, Thr145, and Ile225 residues displayed results such that each plant residue showed inability for the sidechain to hydrogen bond while the ScCM residues showed ability of its sidechain to hydrogen bond. Gly141, Ile192, Ile225, and Ile239 residues displayed results such that both the plant and ScCM residues showed inability for the sidechain to hydrogen bond. The Arg223, Arg224, and Thr226 residues displayed results such that both the plant and ScCM residues showed ability for the sidechain to hydrogen bond. The Arg75, Asn194, Gly221, and Glu222 residues showed results that were not as simple to analyze; these residues showed ScCM and plant residues that were split between ability and inability to hydrogen bond. The proline of *A. thaliana* AtCM2 at Gly141 is intriguing, as proline is more conformationally

restricted compared to glycine; the Gly141 residue is located in a helix secondary structure in ScCM. Finally, Ser220 seems to be an interesting residue, as it was not present in the majority of plant sequences. The reason for this absence is not clear, but one likely reason is that these plant CM enzymes have 220s loops of differing lengths.

2.4 Conclusions

While the findings of this bioinformatics study did not provide any data to suggest that the sequences of the yeast and plant chorismate mutases differ greatly, they do provide some insight into the regulation of this enzyme on a species-specific basis. The high degree of sequence conservation coupled with differences in impacts of effector binding suggests that there must be some sort of link between the structure and function of this enzyme that goes beyond simple discrepancies in residue sequence. The multiple sequence alignments performed were insufficient to explain the differences in regulation of the different chorismate mutases.

In the future, it may be advantageous to employ homology modeling to compare structures of the chorismate mutase enzymes using the Swiss-Model Expasy tool (13). Additionally, it could prove helpful to perform normal mode analysis to predict and understand the motions and mechanisms of the different variations of chorismate mutase; normal mode analysis can provide information on the array of equilibrium modes available to a system. It is possible that these forms of analysis could provide a greater understanding of the regulation of this enzyme in its many forms.

Chapter 3 : Solution-state characteristics of the R-locked T226I variant of ScCM

3.1 Introduction

ScCM is an allosterically-regulated enzyme that has been proposed to follow an MWC-type mechanism of allostery with “R” state and “T” state conformations. However, results from mutagenesis experiments in the L220s loop region have challenged this model (20). For instance, an I225T/T226I ScCM variant still behaves like wild-type ScCM when allosteric effector Trp is present, but it is not inhibited by Tyr, which indicates that a destabilized “T” conformation does not allow for a stabilized “R” conformation when lacking allosteric effector (20).

The majority of work in this chapter is centered upon the ScCM T226I variant. This T226I variant is of particular interest, as when the threonine located at the 226 position is changed to isoleucine, it has been observed to cause the enzyme to remain “locked” in the active “R” conformation, which renders the enzyme immune to the inhibitory effects of Tyr binding. Thus, to gain more insight into the allosteric behavior of the T226I variant, our goal was to characterize this variant in solution using circular dichroism (CD) and NMR spectroscopy. These methods provide insight into the secondary and tertiary structure of a protein, respectively.

One complication is that the protein construct used in the lab carries an extra C-terminal residue (i.e. Thr257), which is leftover following proteolytic cleavage of the hexahistidine tag. As the lab had previously observed the methyl resonances of threonine residues, Thr257 was converted to serine (T257S) to remove this NMR signal. My early studies suggested that addition of this amino acid substitution might interfere through some unknown mechanism with my studies of the T226I substitution. As such, I ended up studying three variants: T257S, T226I/T257S and

T226I (with threonine at position 257). My studies demonstrated that the T226I substitution (with either threonine or serine at position 257) led to more substantial structural changes than were anticipated, calling into question previous kinetic results for this variant.

3.2 Materials and Methods

3.2.1 Expression of ScCM variants in *E. coli* cells

Plasmids expressing ScCM were transformed into $CaCl_2$ competent *E.coli* BL21(DE3*) using standard procedures. Along with the gene for ScCM (or its variants), the plasmid also contained an ampicillin-resistance cassette and an IPTG (isopropyl ß-D-1-thiogalactopyranoside)-inducible promoter. Here, the protein was expressed with a C-terminal hexahistidine-tag and linker that is cleavable via TEV protease.

Following transformation, cells were then plated onto Luria-Bertani (LB) agar with ampicillin (50 µg/mL) with variable volumes of cells (5 µL, 10 µL, 15 µL). The plated cells were then allowed to grow overnight (~ 16 hours) at 37°C. Individual colonies were then used to inoculate 50 mL of NZCYM media with ampicillin (50 µg/mL) and shaken at 200 RPM (rotations per minute) overnight (~16 hours) at 37°C. Portions of the overnight culture (20 mL) were then removed and added to two flasks of 1 L NZCYM media with ampicillin (50 µg/mL). These large cultures were allowed to shake at 250 RPM and 37°C until OD_{600} (optical density at 600 nm) readings of ~ 0.800 were obtained. The cultures were then centrifuged at 4,000 RPM and 20°C for 45 minutes, and the NZCYM media was decanted. The cell pellets were resuspended in 1 L of M9 H_2O media (minimal media without glucose: 700 mL H_2O, 100 µL of 1 M $CaCl_2$, 7.098 g Na_2HPO_4, 3.0 g KH_2PO_4, 0.5 g NaCl, 3.0 g NH_4Cl, 2 mL of 1 M $MgSO_4$, 10mL 100x MEM

vitamin mix, 100 mg/L ampicillin, H_2O to bring to volume) and the centrifugation process was repeated. After centrifugation and decanting of the M9 H_2O media, the pellet was resuspended once more, but this time in 1 L of M9 D_2O media (minimal media with deuterated glucose: 700 mL D_2O, 100 µL of 1 M $CaCl_2$, 12.0 g Na_2HPO_4, 6.0 g KH_2PO_4, 0.5 g NaCl, 3.0 g NH_4Cl, 2 mL of 1 M $MgSO_4$, 3.0 g of ^2H glucose, 10mL 100x MEM vitamin mix, 100 mg/L ampicillin, 1 mL 1000x metals mix, D_2O to bring to volume). This M9 D_2O and cell solution was allowed to shake at 250 RPM and 37°C for 1 hour. After an hour, deuterated glycine (200 mg) and $^{13}CH_3$ threonine (100 mg) were added. Forty minutes after the addition of these two chemicals, $^{13}CH_3$ α-ketobutyric acid (140 mg) was added. Twenty minutes after the addition of this chemical, IPTG (250 mg) was added. Finally, this culture was allowed to shake at 250 RPM and 25°C overnight (~ 16 hours). Cells were harvested by centrifugation, M9 D2O media was decanted, and cells were frozen at -20°C until protein purification.

3.2.2 Protein purification of ScCM and variants

Cell paste was resuspended in lysis buffer (10 mL/g of cell paste); the total volume for the lysis buffer [Nickel Column Buffer, 1 mM PMSF (phenylmethylsulfonyl fluoride as a serine protease inhibitor)] was at least 35 mL to facilitate proper lysis of the cells. The cells were sonicated on ice at 75% power, 1 second on and 1 second off, for 1 minute; this was completed 15 times. Following sonication, the cells were centrifuged at ~ 25,000*g for 25 minutes. A nickel affinity column (Ni-NTA) was then prepared (~ 2 mL of resin per gram of cell paste) and equilibrated before applying sample and collecting fractions. The following solutions were used to collect fractions: 3 column volumes (CVs) Nickel Column Buffer (20 mM potassium phosphate at pH 7, 100 mM NaCl) with 10 mM imidazole, 3 CVs Nickel Column Buffer with 65 mM

imidazole, 3 CVs Nickel Column Buffer with 75 mM imidazole, 3 CVs Nickel Column Buffer with 85 mM imidazole, 3 CVs Nickel Column Buffer with 100 mM imidazole, 3 CVs Nickel Column Buffer with 500 mM imidazole. After collecting fractions, an SDS PAGE gel was run to identify which fractions contained pure ScCM protein.

Fractions containing pure ScCM protein were combined and spin concentrated until ~ 50 mL of protein solution remained. The approximate concentration of protein could then be determined using an absorbance reading at 280 nm (extinction coefficient was 25,830 $M^{-1}cm^{-1}$). Lyophilized TEV protease (1 mg of TEV for every 4 mg of ScCM) was added to the protein solution and dialyzed against 1 L of Nickel Column Buffer II (Nickel Column Buffer with 10 mM imidazole at pH 8) overnight to remove the hexahistidine tag. The following day the sample was dialyzed again against a fresh 1 L of Nickel Column Buffer II for three hours to ensure removal of imidazole. The sample was then applied to a Ni-NTA column that was pre-equilibrated with Nickel Column Buffer II. The column was further washed with Nickel Column Buffer II with 50 mM and 500 mM imidazole. Samples from fractions were analyzed by SDS PAGE to check for purity. Appropriate fractions were collected, concentrated to between 300 mM and 700 mM, and dialyzed against 500 mM potassium phosphate at pH 6.8, 0.2% sodium azide.

3.2.3 Analysis of protein variants using NMR and CD spectroscopy

Following purification of ScCM variants, protein NMR spectra were collected to determine the conformation of each enzyme variant. ^{1}H-^{13}C SOFAST-HMQC NMR spectra of Ile and Thr methyl-labeled chorismate mutase were collected using the following parameters: temperature of 298 K, field strength of 600 MHz, 160 t1 indirect points, spectral width of 8.715 ppm in indirect

dimension, central frequency in indirect dimension of 10.252, ^{1}H excitation center frequency of 0.78 ppm, and ^{1}H excitation bandwidth of 3 ppm.

NMR spectroscopy for proteins uses multidimensional NMR experiments in order to acquire structural information about the protein. In theory, each molecule within the protein contains a nucleus that experiences a unique electronic atmosphere, which means that each molecule should also have its own unique chemical shift. Band-Selective Optimized Flip Angle Short Transient (SOFAST) Heteronuclear Multiple Quantum Coherence (HMQC) is a technique especially designed for time-conscious acquisition of 2D heteronuclear NMR spectra. SOFAST-HMQC experiments use shaped ^{1}H pulses that influence only some proton spins, not all, so that their spin-lattice relaxation (T_1) is enhanced via dipolar interactions with the unimpacted proton spins (22). Additionally, this experiment succeeds in achieving maximal sensitivity for frequent repetition of pulse sequence by using Ernst angle excitation (the angle that provides the highest signal-to-noise ratio possible in a given quantity of time) (22).

As needed, circular dichroism (CD) was conducted to detect relative abundance of helices and sheets. CD spectra of ScCM were collected using a Jasco J-1500 CD spectrometer, thermostatted at 23°C, and a 1-mm pathlength quartz CD cuvette (Jasco J/0556). Samples were scanned from 260 nm to 185 nm, with the following instrument settings: 1 nm bandwidth, 0.5 nm data pitch, 50 nm/min scan rate, 4 s average time. Three scans were acquired and averaged for each condition.

3.3 Results

3.3.1 Purification and NMR characterization of ScCM T257S are consistent with previous results

As I had no previous experience in production and purification of isotopically-labeled proteins, it was determined that I should try to repeat results from Dennis Winston (my mentor) and Scott Gorman, Ph.D. (a previous graduate student in the lab). Following protein expression, cells were lysed and then passed over the first nickel affinity column so that the C-terminal His-tagged ScCM would be retained. The tag was then cleaved with TEV protease and the protein was passed across the second nickel affinity column so that the tag would be retained and the desired protein is found in the flow-through.

Upon inspection of SDS PAGE gels, it appears that this ScCM T257S sample was successfully purified (see **Figures 3.1-3.2**). The first nickel column was performed successfully, as evidenced by the SDS PAGE gel for this nickel column (see **Figure 3.1**). In this gel, the His-tagged ScCM T257S protein eluted by way of a wash concentrated with 500 mM imidazole (see **Figure 3.1**). This is typical and expected.

The SDS PAGE gel that resulted from the second nickel column (see **Figure 3.2**) continued to prove that proper purification was achieved. The ScCM T257S protein eluted in washes that contained 10 mM imidazole and 50 mM imidazole; the TEV protease appeared to have successfully cleaved the hexaHis tag (see **Figure 3.2**).

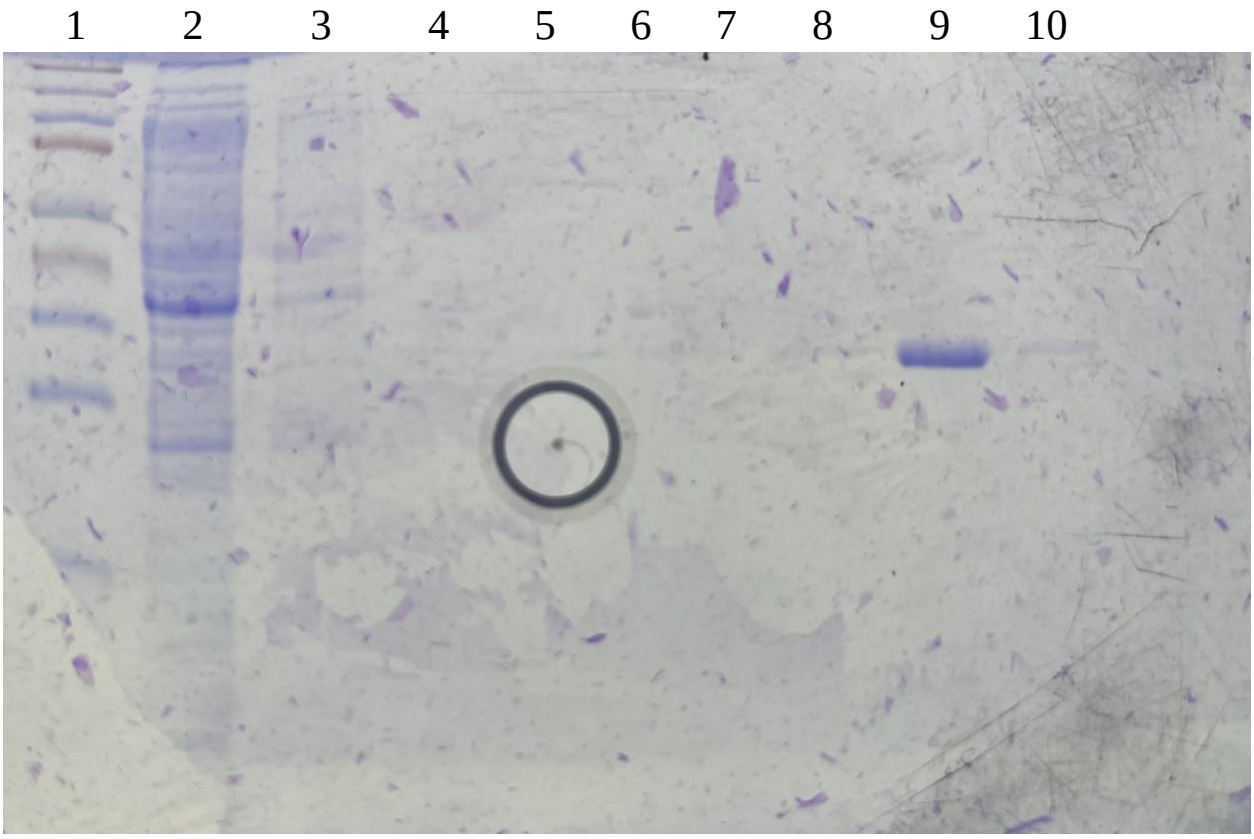


Figure 3.1: Purification of the T257S variant of ScCM across the first nickel affinity column as analyzed by SDS PAGE. Lanes 9 and 10 indicate a single protein that was retained by the nickel affinity column but then eluted using a higher imidazole concentration (500 mM).

Lane 1 – ladder; Lane 2 – flowthrough from supernatant; Lane 3 – flowthrough from Nickel Column Buffer with 65 mM imidazole total and Triton detergent; Lane 4 – flowthrough from Nickel Column Buffer with 65 mM imidazole total; Lane 5 – flowthrough from Nickel Column Buffer with 75 mM imidazole total; Lane 6 – flowthrough from Nickel Column Buffer with 85 mM imidazole total; Lane 7 – flowthrough from Nickel Column Buffer with 100 mM imidazole total; Lane 8 – flowthrough from Nickel Column Buffer with 500 mM imidazole total; Lane 9 – flowthrough from Nickel Column Buffer with 500 mM imidazole total; Lane 10 – flowthrough from Nickel Column Buffer with 500 mM imidazole total

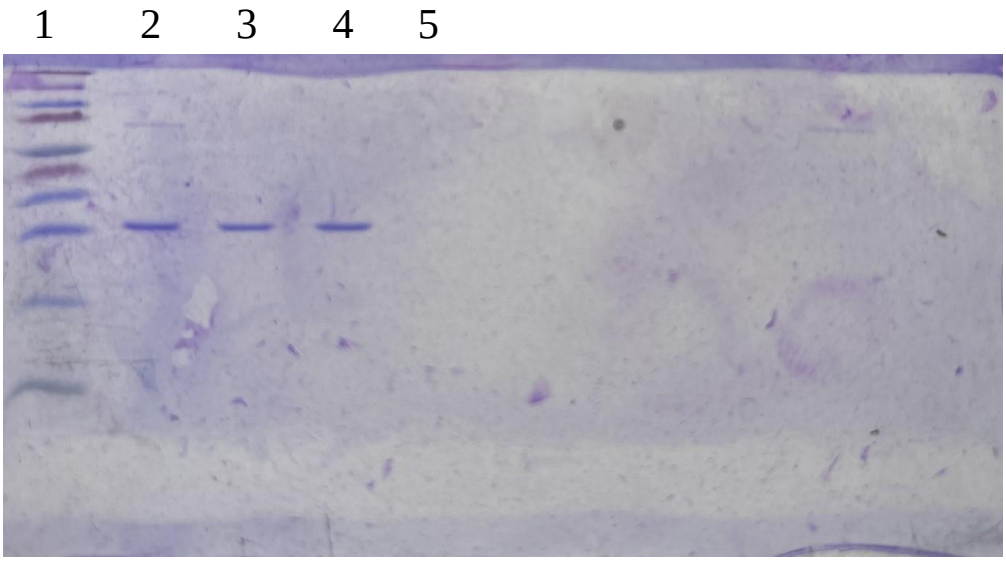


Figure 3.2: Purification of the T257S variant of ScCM across the second nickel affinity column as analyzed by SDS PAGE. Lanes 2-5 indicate a single protein that was not retained by the nickel affinity column, suggesting proper TEV proteolytic cleavage of the C-terminal hexahistidine tag.

Lane 1 – ladder; Lane 2 – flowthrough from sample from first column (Day One of purification); Lane 3 – flowthrough from four washes with Nickel Column Buffer II with 10 mM imidazole total; Lane 4 – flowthrough from Nickel Column Buffer II with 50 mM imidazole total; Lane 5 – flowthrough from Nickel Column Buffer II with 500 mM imidazole total

Following protein purification, I collected a SOFAST-HMQC ^{1}H-^{13}C NMR spectrum (see **Figure 3.3**). This NMR spectrum was typical of the "wild-type" ScCM T257S, with narrow, defined peaks. This spectrum (see **Figure 3.3**) was compared to a previously-collected spectrum of the T257S variant (see **Figure 3.4**). The comparison indicates that the ScCM T257S variant was successfully expressed and purified while possessing the proper tertiary and quaternary structures. It is also helpful to look at the spatial locations of these T257S spectrum peaks within a three-dimensional model of ScCM (see **Figure 3.5**).

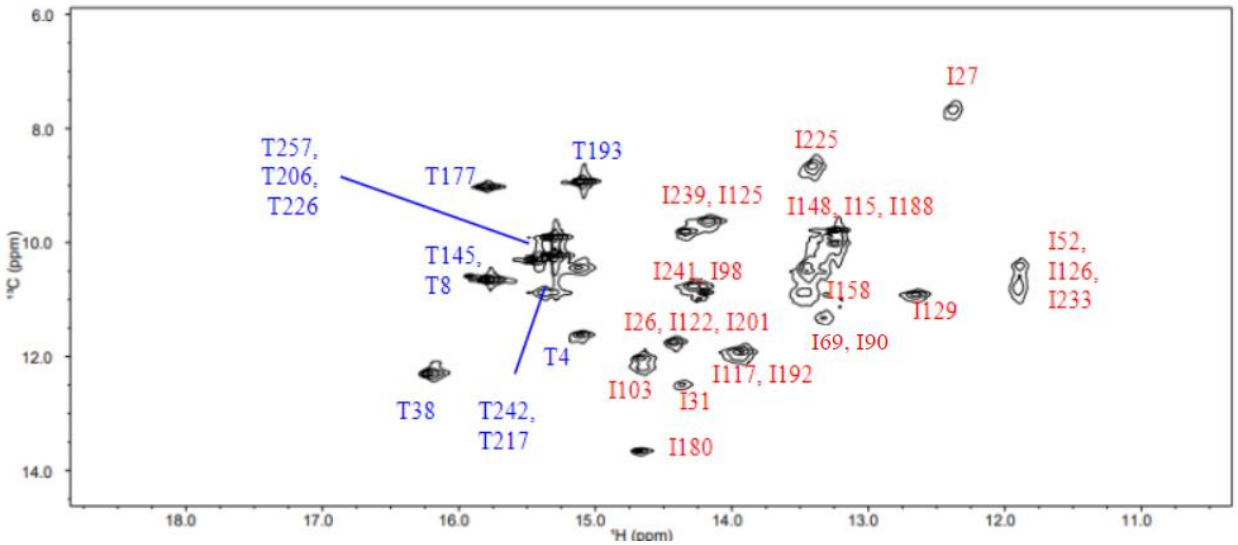


Figure 3.3: NMR spectrum for ScCM T257S indicates typical tertiary and quaternary structure for this variant.

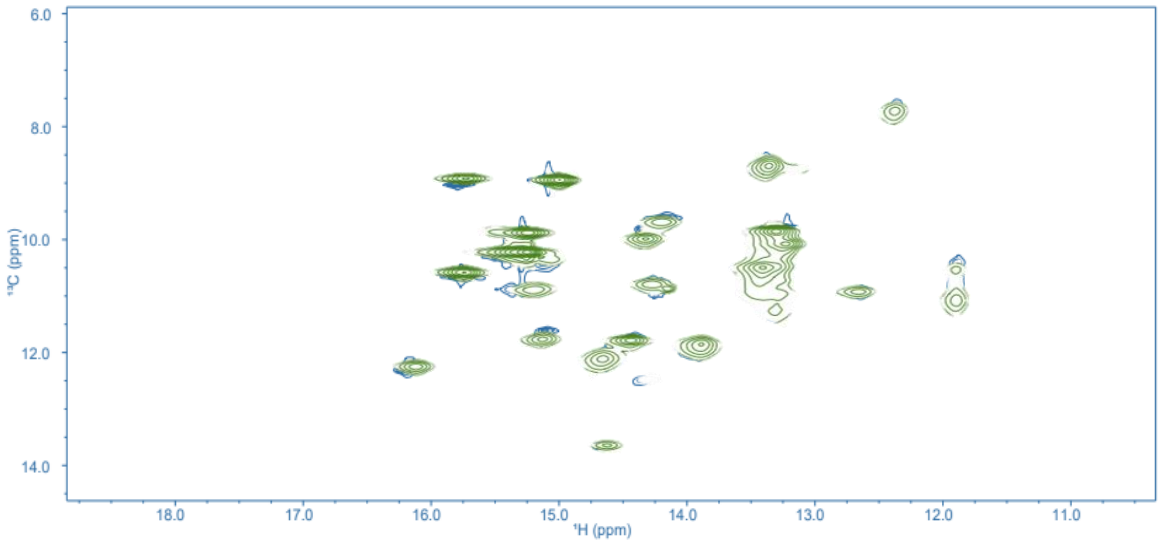


Figure 3.4: NMR spectrum for ScCM T257S (in blue) overlaid with previous T257S spectrum (in green) generated by Dennis Winston during past experimentation. This overlay confirms that the T257S variant purified in this chapter possesses typical tertiary and quaternary structures.

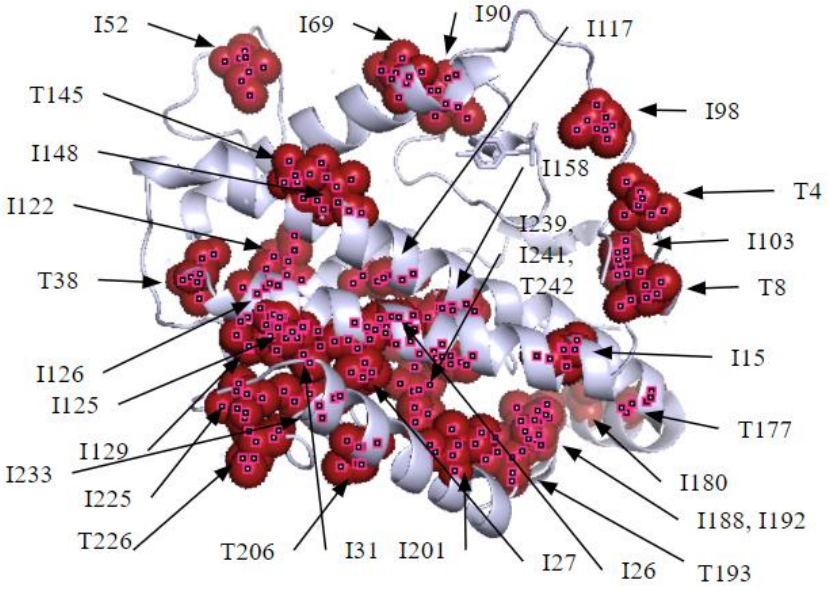


Figure 3.5: Three-dimensional structure of ScCM with NMR spectrum residues labeled

Thus, it was established that the results of previous studies could be replicated without issue. This proof of repeatability was essential for bolstering validity of the results that follow in subsequent sections of this chapter.

3.3.2 ScCM T226I/T257S does not fold properly

I then expressed and purified the T226I/T257S variant of ScCM to compare against the results from the T257S variant. However, the protein NMR spectrum (via the ^{1}H-^{13}C SOFAST-HMQC NMR experiment) was vastly different from before, suggesting that the protein was at least partially misfolded or, perhaps, there was some other protein contaminant (see **Figure 3.6**).

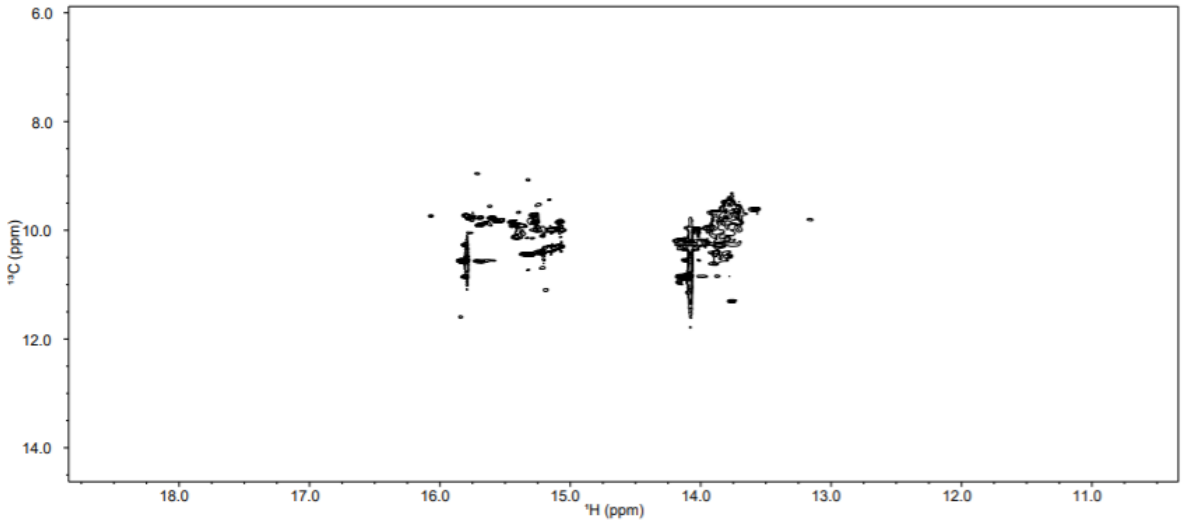


Figure 3.6: NMR spectrum for ScCM T226I/T257S is substantially different from that of ScCM T257S, suggesting protein misfolding or issues with protein purification.

Given the atypical nature of the NMR spectrum (see **Figure 3.6**), circular dichroism was performed. The results – in both graphic and numeric forms – indicate that the ScCM T226I/T257S protein was composed of a mixture of structural components (see **Figure 3.7**). The ScCM protein has traditionally been composed of almost all helix. However, for this protein sample, CD showed ~ 18% helix, ~ 31% sheet, ~ 14% turn, and ~ 37% other.

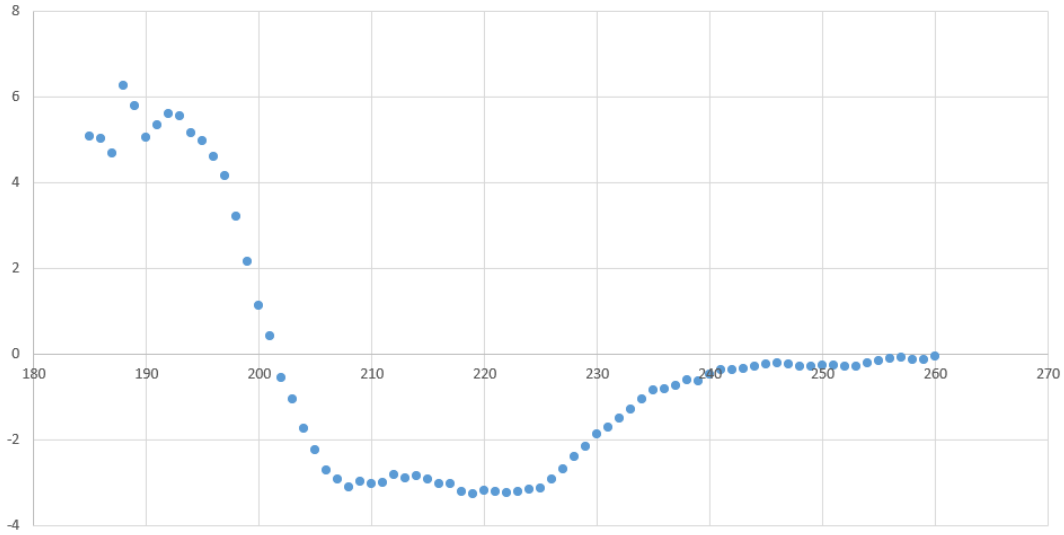


Figure 3.7: Circular dichroism results for ScCM T226I/T257S show misfolding or protein contamination.

This variant was expressed and purified according to the outlined protocol. However, based upon the evidence present in an SDS PAGE gel (see **Figure 3.8**), there was a faint band present to represent the ScCM protein (~ 30 kDa), but there was also an additional band present at ~ 40 kDa. This band cannot be the TEV protease (it is ~ 27 kDa) or a multimer of ScCM protein, which means that this protein's origin and identity are unknown.

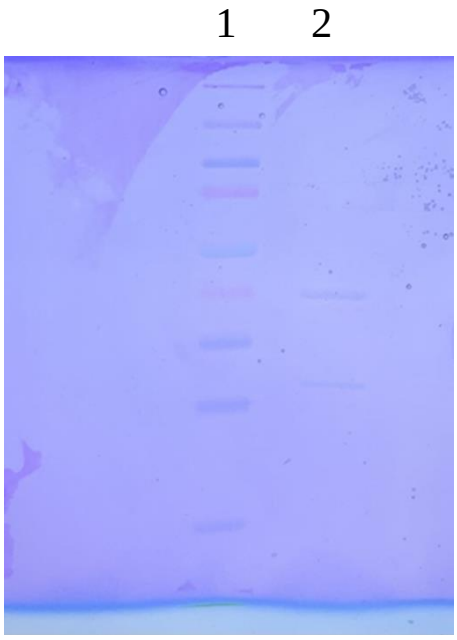


Figure 3.8: SDS PAGE for ScCM T226I/T257S indicates presence of an unidentifiable protein

Lane 1 – ladder; Lane 2 – sample

3.3.3 ScCM T226I also appears to misfold

I also expressed and purified the ScCM T226I variant, which has a Thr at the C-terminal residue instead of Ser. However, much like the ScCM T226I/T257S variant, upon analysis via SDS PAGE and protein NMR (^{1}H-^{13}C SOFAST-HMQC), there was evidence to suggest that the protein was once again misfolded, and that there may be an unidentifiable protein present. This was unexpected and continued to raise questions about why this variant would be misfolded or from where the protein contamination came.

The SDS PAGE gels for the T226I variant (see **Figures 3.9-3.10**) support the idea that an unidentifiable protein was present within the ScCM T226I sample. Although the gel from the first nickel column (see **Figure 3.9**) does not indicate that this extra protein was present, the gel used

to check the purity of the ScCM T226I NMR sample (see **Figure 3.10**) indicates presence of the extra protein following completion of purification and during NMR experimentation.

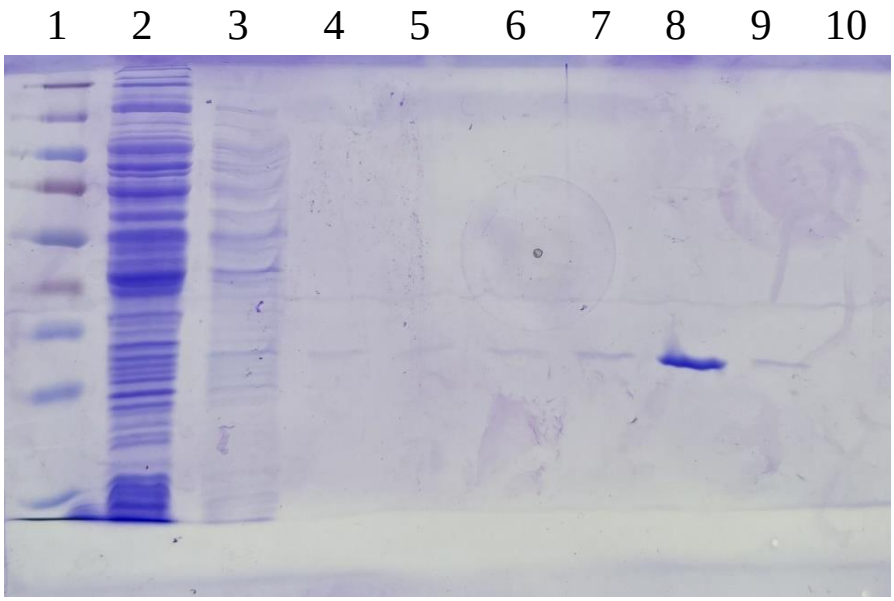


Figure 3.9: Purification of the T226I variant of ScCM across the first nickel affinity column as analyzed by SDS PAGE. Lanes 7-10 indicate a single protein that was retained by the nickel affinity column but then eluted using a higher imidazole concentration (>100 mM).

Lane 1 – ladder; Lane 2 – flowthrough from supernatant; Lane 3 – flowthrough from Nickel Column Buffer with 65 mM imidazole total and Triton detergent; Lane 4 – flowthrough from Nickel Column Buffer with 65 mM imidazole total; Lane 5 – flowthrough from Nickel Column Buffer with 75 mM imidazole total; Lane 6 – flowthrough from Nickel Column Buffer with 85 mM imidazole total; Lane 7 – flowthrough from Nickel Column Buffer with 100 mM imidazole total; Lane 8 – flowthrough from Nickel Column Buffer with 500 mM imidazole total; Lane 9 – flowthrough from Nickel Column Buffer with 500 mM imidazole total; Lane 10 – flowthrough from Nickel Column Buffer with 500 mM imidazole total

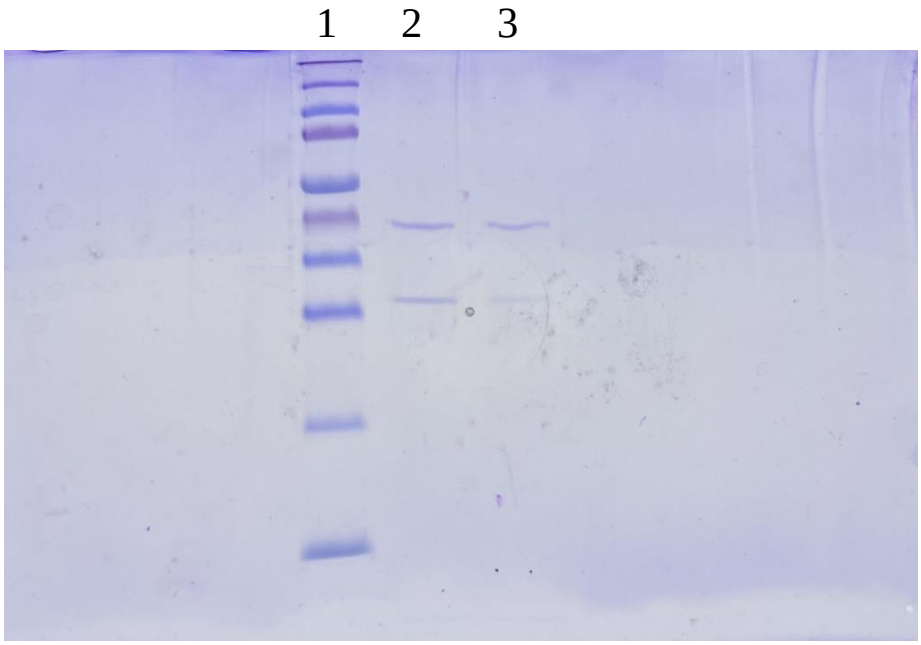


Figure 3.10: SDS PAGE for T226I variant – NMR purity check. Indicated presence of an unidentifiable protein.
Lane 1 – ladder; Lane 2 – ScCM T226I after purification; Lane 3 – ScCM T226I NMR sample

An NMR spectrum collected for this variant (see **Figure 3.11**) shows the substantial structural changes to ScCM apparently caused by the T226I substitution. The broader peaks for this sample did not appear with only 16 scans, which may indicate that these broad peaks are derived from a low-concentration protein, which may perhaps be folded ScCM T226I. The sharper peaks are likely representative of the unidentifiable protein. An overlay between the T226I spectrum and a T257S spectrum (see **Figure 3.12**) indicates chemical shift and peak intensity changes, suggesting that there are substantial structural and/or dynamic changes induced by the T226I substitution. While NMR spectra do not directly show structural changes, spectra do depend on three-dimensional protein structure. Thus, changes to a protein's structure appear as chemical shift changes in NMR spectra.

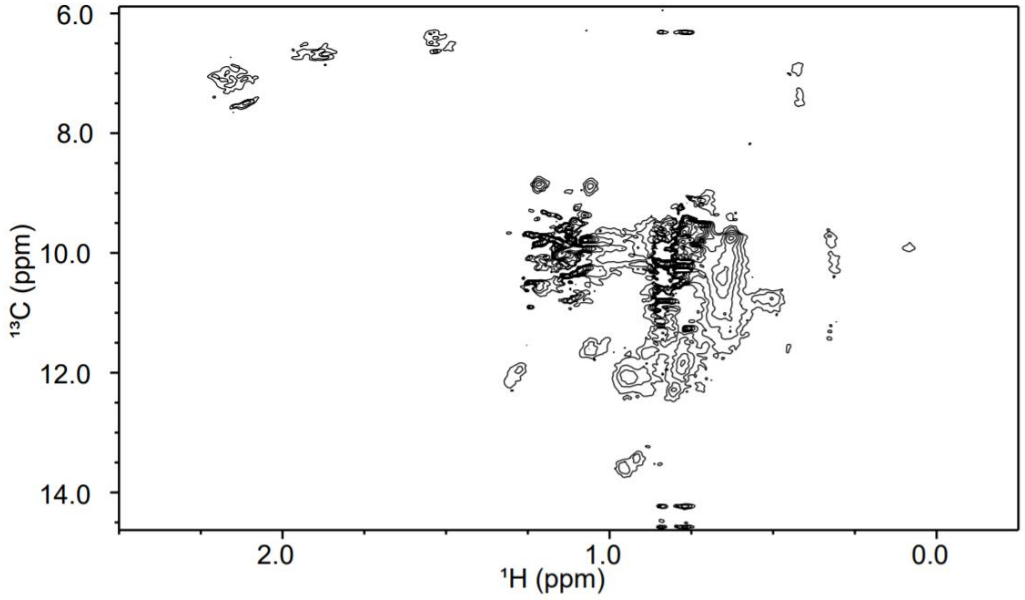


Figure 3.11: NMR spectrum for ScCM T226I shows misfolding

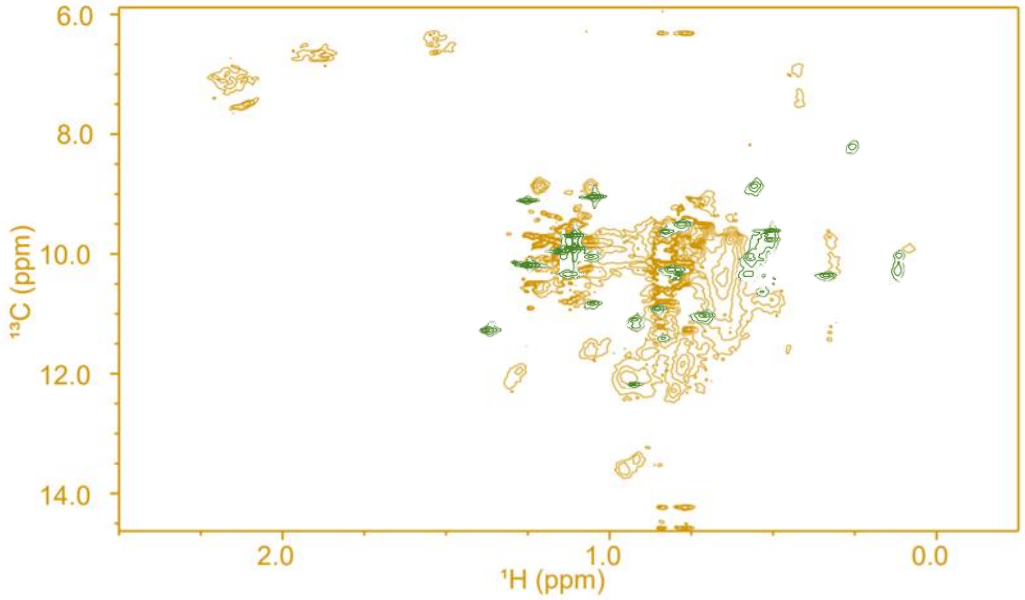


Figure 3.12: NMR Spectrum for ScCMT226I (in gold) overlaid with spectrum for T257S (in green). Noticeable differences in peaks and chemical shifts are readily observable.

3.4 Conclusions

Two of the ScCM variants that were expressed and purified do not appear to be typical of past results, which could suggest that the T226I mutation in the ScCM T226I/T257S and ScCM T226I variants has caused some issues with protein folding, and/or some protein contamination issues, that had not been observed in past experiments. To determine whether the apparent misfolding of these variants is due to human error instead of changes in protein folding due to the introduction of a nonsynonymous mutation, the ScCM T257S variant was expressed, purified, and analyzed to determine if the structure of the variant is typical. The ScCM T257S sample appears to be correctly folded, which indicates that it would be reasonable to believe that the misfolding of the ScCM T226I/T257S and ScCM T226I variants is – at least partially – due to the introduction of the T226I mutation.

Chapter 4 : Conclusions and Future Directions

Throughout this thesis, the importance of ScCM as a model system for the study of allostery has been sufficiently discussed. Due to the range of allosteric behaviors that ScCM displays, we were able to study sequence determinants of allosteric regulation using bioinformatic techniques, as well as begin to investigate the solution characteristics of the T226I variant.

As a result of the high levels of conservation and complexity of conservation patterns displayed in multiple sequence alignments in Chapter 2, not much could be concluded regarding species-specific CM behavior. While these results were less-than-satisfying, they do allow for the elimination of sequence as the main cause for differences in allosteric regulation. In order to further explore these species-specific differences in CM behavior, it may be worth engaging in amino acid covariance analysis as a means of further study. Amino acid covariance analysis is a method used to detect residues that coevolved in different species. These residues can then be used as constraints for three-dimensional protein structures, as we can assume that – at least partially – the force of coevolution for these choice residues is a product of their importance as direct contacts and steric constraints when in the folded protein state (23). More specifically, covariance analysis employs the use of multiple sequence alignments for computing couplings between pairs of residue positions in a protein sequence (23). The computed couplings represent frequencies of observing these residue pairs across all sequences in the multiple sequence alignment. Finally, couplings are used to calculate a range of covariance scores that can aid in analysis (23).

In Chapter 3, I explored the ScCM T226I variant, which effectively “locks” the enzyme into the active “R” conformation. Upon analysis of NMR spectra, CD spectra, and SDS PAGE gels, it appears that the substitution of isoleucine for threonine at position 226 causes misfolding

that had not been reported in previous studies. In order to further explore the apparently disruptive nature of the T226I mutation, there are a few options for future methods of investigation, including: 1) repetition of studies and 2) the use of mass spectrometry.

If the studies described in Chapter 3 were repeated to the exact same specifications, these novel results could serve as a means of supporting or refuting the findings reported here. Replicability is important, and it could provide further evidence for – or against – the apparently disruptive character of the T226I mutation.

Mass spectrometry (MS) may be useful in determining the identity of the unidentified protein present in SDS PAGE gels for the T226I/T257S and T226I variants in Chapter 3. Using MS for protein identification requires the digestion of proteins into peptides, which are then ionized and detected by the mass spectrometer (24). In order to identify the protein, peaks from the mass spectra are analyzed computationally, and each peak is representative of one peptide fragment ion (24). This, hopefully, would allow for identification of the unknown protein and would help us determine its origin.

I am hopeful that a future researcher will someday build upon the work I have described in this thesis and continue to explore the allosteric regulation of ScCM and other CM enzymes. Being able to contribute to developing research into allostery in ScCM has been an experience I shall not soon forget.

APPENDIX A – Supplemental Data for Multiple Sequence Alignments

Multiple Sequence Alignment for Yeasts Impacted by Phe/Tyr Binding

REFERENCES CITED

(1) Jiao, W., & Parker, E. J. (2012). Using a Combination of Computational and Experimental Techniques to Understand the Molecular Basis for Protein Allostery. Structural and Mechanistic Enzymology - Bringing Together Experiments and Computing Advances in Protein Chemistry and Structural Biology, 87, 391–413. doi: 10.1016/b978-0-12-398312-1.00013-5

(2) Monod, J., Wyman, J., & Changeux, J.-P. (1965). On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, *12*(1), 88–118. https://doi.org/10.1016/s0022-2836(65)80285-6

(3) Koshland, D. E., Némethy, G., & Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits*. *Biochemistry*, *5*(1), 365–385. https://doi.org/10.1021/bi00865a047

(4) Gorman, S. D. (2020). Insight into Mechanisms of Allosteric Regulation and Cooperativity from Yeast Chorismate Mutase (Unpublished doctoral dissertation). The Pennsylvania State University.

(5) Martínez, J. L., Liu, L., Petranovic, D., & Nielsen, J. (2012). Pharmaceutical protein production by yeast: Towards production of human blood proteins by microbial fermentation. *Current Opinion in Biotechnology*, *23*(6), 965–971. https://doi.org/10.1016/j.copbio.2012.03.011

(6) Goerisch, H. (1978). On the mechanism of the chorismate mutase reaction. *Biochemistry*, *17*(18), 3700–3705. https://doi.org/10.1021/bi00611a004

(7) Andrews, P. R., Smith, G. D., & Young, I. G. (1973). Transition-state stabilization and enzymic catalysis. kinetic and molecular orbital studies of the rearrangement of chorismate to prephenate. *Biochemistry*, *12*(18), 3492–3498. https://doi.org/10.1021/bi00742a022

(8) UniProt ConsortiumEuropean Bioinformatics InstituteProtein Information Resource SIB Swiss Institute of Bioinformatics. (n.d.). UniProt Consortium. Retrieved January 19, 2021, from https://www.uniprot.org/

(9) Protein BLAST: Search protein databases using a protein query. (n.d.). Retrieved February 02, 2021, from https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins

(10) Bank, R. (n.d.). 3CSM. Retrieved January 27, 2021, from https://www.rcsb.org/3d-view/ngl/3csm

(11) Bank, R. (n.d.). 4CSM. Retrieved January 27, 2021, from https://www.rcsb.org/3d-view/ngl/4csm

(12) Cobalt: phylogenetic tree widget - cobalt rid 37unysmx212. (n.d.). Retrieved February 23, 2021, from https://www.ncbi.nlm.nih.gov/blast/treeview/treeView.cgi?request=page&cobaltRID=37UNYSMX212&checkStatus=on&link_loc=alignPage&screenWidth=1280&screenHeight=720

(13) *MODEL*. SWISS. (n.d.). https://swissmodel.expasy.org/.

(14) Kroll, K., Holland, C. K., Starks, C. M., & Jez, J. M. (2017). Evolution of allosteric regulation in chorismate mutases from early plants. *Biochemical Journal, 474*(22), 3705-3717. doi:10.1042/bcj20170549

(15) UniProt ConsortiumEuropean Bioinformatics InstituteProtein Information Resource SIB Swiss Institute of Bioinformatics. (n.d.). UniProt Consortium. Retrieved March 11, 2021, from https://www.uniprot.org/

(16) Strater, N., Hakansson, K., Schnappauf, G., Braus, G., & Lipscomb, W. (1996). Crystal structure of the T state of allosteric yeast chorismate mutase and comparison with the R state. *Proceedings of the National Academy of Sciences of the United States of America, 93*(April), 3330-3334. doi:10.1073/pnas.93.8.3330

(17) Strater, N., Schnappauf, G., Braus, G., & Libscomb, W. N. (1997). Mechanisms of catalysis and allosteric regulation of yeast chorismate mutase from crystal structures. *Structure, 5*(11), 1437-1452. doi:10.1016/s0969-2126(97)00294-3

(18) Gorman, S. D., & Boehr, D. D. (n.d.). *Energy and enzyme activity landscapes of yeast chorismate mutase at cellular concentrations of allosteric effectors*. Biochemistry. Retrieved February 27, 2022, from https://pubmed.ncbi.nlm.nih.gov/31498992/

(19) Bode, R., & Birnbaum, D. (1991, January). *Regulation of chorismate mutase activity of various yeast species by aromatic amino acids*. Antonie van Leeuwenhoek. Retrieved February 27, 2022, from https://www.ncbi.nlm.nih.gov/pubmed/2059012

(20) G. Schnappauf, W.N. Lipscomb, G.H. Braus, Separation of inhibition and activation of the allosteric yeast chorismate mutase, Proc. Natl. Acad. Sci. 95 (1998) 2868–2873. https://doi.org/10.1073/pnas.95.6.2868.

(21) D.S. Winston, S.D. Gorman, D.D. Boehr, Conformational transitions in yeast chorismate mutase important for allosteric regulation as identified by nuclear magnetic resonance spectroscopy, *Journal of Molecular Biology* (2022), doi: https://doi.org/10.1016/j.jmb.2022.167531

(22) Roberts, Gordon C., and Bernhard Brutscher. "SOFAST HMQC." *Encyclopedia of Biophysics*, Springer Berlin, Berlin, 2013.

(23) Corcoran, D., Maltbie, N., Sudalairaj, S., Baker, F. N., Hirschfeld, J., & Porollo, A. (2021). CoeViz 2: Protein graphs derived from amino acid covariance. *Frontiers in Bioinformatics*, *1*. https://doi.org/10.3389/fbinf.2021.653681

(24) Wang, P., Wilson, S.R. Mass spectrometry-based protein identification by integrating de novo sequencing with database searching. *BMC Bioinformatics* **14,** S24 (2013). https://doi.org/10.1186/1471-2105-14-S2-S2

ACADEMIC VITA

Evan Nelson

efn5030@psu.edu | efn5030@gmail.com

EDUCATION

PENN STATE COLLEGE OF MEDICINE – Penn State Health Milton S. Hershey Medical Center
Doctor of Medicine pending 2026

PENN STATE UNIVERSITY, SCHREYER HONORS COLLEGE – University Park, PA
Bachelor of Science in Premedicine pending 2022
Major in Premedicine ♦ Smeal College Business Fundamentals Certificate

Leadership/ Activities:
- ♦Merchandise Chair, Schreyer Honors College Student Council
- ♦Springfield Club, a THON philanthropic organization
- ♦Blue & White Society
- ♦Penn State Fly Fishing Club

PETERS TOWNSHIP HIGH SCHOOL – McMurray, PA
2018 Graduate Summa Cum Laude ♦ PT Scholar ♦ Gifted Program

Leadership/ Activities:
Fundraiser Chair, National Honor Society
Treasurer, Math Club
Communications Chair, American Medical Student Association
Founding Member of the Academic Integrity Committee
Math League
Interact Club

UNDERGRADUATE EXPERIENCE

- **Clinical Research**
  UPMC Presbyterian Hospital, Pittsburgh, PA – Paula M. Novelli, MD; Philip D. Orons, DO, FAOCR, Chief Professor of Radiology; Cole Thompson, MD
  - Conducted clinical research via chart reviews while under the supervision of physicians from the interventional radiology division at UPMC Presbyterian Hospital
  - Research focused on how and when the use of pelvic artery embolization for postpartum hemorrhage patients can improve these patients' prognoses
  - Successfully navigated electronic medical records, specifically EpicCare and PowerChart

- Gained greater understanding of extra-clinical work and female health

- **Undergraduate Biochemistry Research**
  Penn State University Department of Chemistry, University Park, PA – David D. Boehr, Ph.D.
  - 10 hours/week during undergraduate studies
  - Engaged in undergraduate research investigating the biophysical chemistry of enzymes
  - Utilized NMR spectroscopy for determination of internal enzymatic motion and general enzyme function
  - Worked to overexpress amino acid using choice strains of *E. coli* bacteria
  - Research findings have implications on drug design and protein engineering

- **Physician Observation**
  UPMC Pittsburgh Internal Medicine Associates, Pittsburgh, PA – G. Richard Zimmerman, MD
  - Shadowed with Dr. Zimmerman, a primary care physician, for 22 hours
  - Gained experience with the roles and responsibilities of family practice, including: general health maintenance, chronic disease care, immunization recommendations, and analysis of basic mental health concerns
  - Learned proper way to establish bedside manner and think clinically while with patients

  UPMC Presbyterian Hospital, Pittsburgh, PA – Paula M. Novelli, MD; Philip D. Orons, DO, FAOCR, Chief Professor of Radiology; Nikhil B. Amesur, MD
  - Shadowed with UPMC interventional radiologists for 110 hours
  - Gained exposure to a wide range of minimally invasive procedures, including but not limited to: repair of splenic aneurysm, duodenal ulcer embolization, transjugular intrahepatic portosystemic shunt (TIPS), repair of spine hematoma, Y90 nuclear radiology procedure, and transcatheter arterial chemoembolization (TACE)
  - Rounded with various members of the IR team to all areas of care
  - Reviewed electronic medical records for relevant patients
  - Attended noon conference with residents

  The Washington Hospital, Washington, PA – Thomas Pirosko, DO
  - Shadowed with Washington Hospital emergency medicine physician Dr. Pirosko for 100 hours
  - Learned and practiced the process of assessing and diagnosing patients
  - Assisted with patient interviewing and communication of care options
  - Worked alongside all members of the emergency department
  - Gained a greater understanding of the role emergency medicine has in the progression of patient care

- **Freelance Writer**
  SE2.0 – A Search Engine Optimization and Content Writing Agency
  - Served as a freelance writer for over 9 months
  - Wrote and edited approximately 70 articles focusing on mental health and business topics

- Honed writing skills and gained experience as an effective communicator

♦ **Scheel Family Scholarship in Premedicine in the Eberly College of Science**

- Selected by the Eberly College of Science to receive the Scheel Family Scholarship as a reward for academic excellence

OTHER RELEVANT EXPERIENCE

Recipient of the United Way of Washington County M. Kathleen Ramsey Student Volunteer of the Year Award for Commitment and Dedication to Community Service

♦ **Patient Transport and Emergency Department Volunteer** (2014-2018, 256 hours)
St. Clair Hospital, Pittsburgh, PA

- Admitted and discharged patients; distributed and stocked supplies for patient care

♦ **Junior EMT** (2017-2018, 125 hours)
Peters Township Ambulance Service, McMurray, PA

- Assisted emergency personnel with rescues, provided basic first aid, and assessed patients

♦ **Volunteer Coordinator** (2017)
American Heart Association

- Researched partner corporations, assembled volunteer groups for Pittsburgh Heart Walk

♦ **Research Assistant** (2017)
Waynesburg University, Waynesburg, PA – Evonne Baldauff, Ph.D.

- Analyzed volatile coffee samples using SPME technique to aid local roaster in production

♦ **Physician Shadow** (2016-2018)

- Advanced Orthopaedics and Rehabilitation, Washington, PA – Vincent J. Ripepi, DO- Learned about the examination, diagnostic, and treatment processes for conditions of the musculoskeletal system from an orthopedic surgeon
- VA Pittsburgh Healthcare System, Pittsburgh, PA – Danny Chu, MD, FACS – Observed two full days of CABG surgeries in the operating room with a cardiac surgeon
- The Washington Hospital, Washington, PA – Thomas Pirosko, DO – Explored the duties of emergency medicine including: assessment of patient conditions, ordering diagnostic tests, stabilization of patients, treatment of injuries, and prescribing medication
- Monongahela Valley Hospital, Monongahela, PA – William Pendergast, MD – Experienced the roles and responsibilities of a hospitalist and learned about providing continued medical care to hospitalized patients