2017

Structural and Functional Studies of Proteins from the Agricultural Pests Tetranychus Urticae and Aspergillus Fumigatus

Caleb Schlachter
University of South Carolina

STRUCTURAL AND FUNCTIONAL STUDIES OF PROTEINS FROM THE AGRICULTURAL PESTS *TETRANYCHUS URTICAE* AND *ASPERGILLUS FUMIGATUS*

by

Caleb Schlachter

Bachelor of Science
University of Tennessee, 2011

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Chemistry

College of Arts and Sciences

University of South Carolina

2017

Accepted by:

Maksymilian Chruszcz, Major Professor

Thomas Makris, Committee Member

Linda Shimizu, Committee Member

Gregorio Gomez, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge and thank Dr. Maksymilian Chruszcz for accepting me into his lab and providing wonderful guidance and wisdom throughout my graduate career. I would like to acknowledge Dr. Thomas Makris and Jose Amaya for all the insight they provided on some projects not mentioned in this publication. I would like to acknowledge Dr. Leslie Lovelace and Dr. Amy Taylor-Perry for giving me my preferred Teaching Assistantships. I would like to acknowledge my former and current labmates: Dr. Lesa Offermann, Dr. Nick Mank, Dr. William Booth, Nikita Ussin, Swanandi Pote, Brenda Kapingidza and Leily Daneshian. I would like to thank my former and current undergraduates for their patience and help with crystallization: Vincent Klapper, Taylor Radford, Josh Burgess and Anna Bagnell. I would like to acknowledge my friends and family for their continued support. I would like to acknowledge the University of South Carolina for providing me with several opportunities to present my research at conferences, and for awarding me the SPARC Grant. Lastly, I would like to acknowledge and thank the members of my committee: Dr. Thomas Makris, Dr. Linda Shimizu, Dr. Gregorio Gomez and Dr. Cory Robinson (former committee member).

ABSTRACT

Agricultural pests are a worldwide problem and cause billions of dollars in crop loss. In the United States alone, an estimated $40 billion USD is lost per year due to insecticide resistance [1]. Studied here are proteins (potentially new pesticide targets) from the agricultural pests *Tetranychus urticae* and *Aspergillus fumigatus*. *T. urticae*, or two-spotted spidermite, is a polyphagous pest, and three proteins from this pest, a cyanase, a glutathione S-transferase and an intradiol ring-cleavage dioxygenase are described [2]. Cyanase is involved in the conversion of bicarbonate and cyanate, a toxic self-defense metabolite produced by plants, into ammonia and carbon dioxide. Glutathione S-transferase conjugates reduced glutathione to xenobiotics for detoxification and have been associated with insecticide resistance [3]. Intradiol ring-cleavage dioxygenase is involved in the breakdown and metabolism of toxic aromatic compounds.

Furthermore, *A. fumigatus* is a ubiquitous fungus that is not only a problem in agriculture, but also in healthcare. *A. fumigatus* drug resistance is becoming more prominent which is mainly attributed to the widespread use of fungicides in agriculture [4]. The 2-methylcitrate cycle, which is only present in fungi, is responsible for detoxifying propionyl-CoA, a toxic metabolite produced as the fungus breaks down proteins [5]. The enzyme responsible for this detoxification is 2-methylcitrate synthase (mcsA) and is a potential candidate for the design of new anti-fungals. However, mcsA shares a similar reaction to human citrate synthase (hCS), and both the structure and function of each enzyme is studied in parallel to find mcsA-specific inhibitors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

4,4'-DDT ..... 1,1,1-trichloro-2,2-bis(4-chlorophenyl) Ethane

4HNE 4-hydroxynonenal

4NBC ..... 4-nitrobenzyl Chloride

4NPA ..... 4-nitrophenethyl Acetate

4NPB ..... 4-nitrophenethyl Bromide

A-CoA ..... Acetyl-CoA

AMP ..... Ampicillin

APS ..... Advanced Photon Source

AU ..... Asymmetric Unit

β-ME 2-mercaptoethanol

BmGST ..... *Bombyx mori* Glutathione S-transferase

B. mori ..... *Bombyx mori*

cCS ..... Chicken Citrate Synthase

CMM ..... CheckMyMetal Server

CDNB 1-chloro-2,4-dinitrobenzene

CoA ..... Coenzyme A

CoA-SH ..... Coenzyme A (free thiol)

CP ..... Chlorfenapyr

dATP ..... Deoxyadenosine Triphosphate

dCTP ..... Deoxycytidine Triphosphate

DEAE ..... Diethylaminoethanol

dGTP ..... Deoxyguanosine Triphosphate

DNDGIC ..... Dinitrosyl-diglutathionyl Iron Complex

DSF ..... Differential Scanning Fluorimetry

DTNB ..... 5,5-dithio-bis-(2-nitrobenzoic Acid)

dTTP ..... Deoxythymidine Triphosphate

DTT ..... Dithiothreitol

EcCyanase ..... *Escherichia coli* Cyanase

E. coli ..... *Escherichia coli*

EDTA ..... Ethylenediaminetetraacetic Acid

EPNP ..... 1,2-epoxy-3-(4-nitrophenoxy)-propane

FPLC ..... Fast-Protein Liquid Chromatography

GFP ..... Green Fluorescent Protein

GMQE ..... Global Model Quality Estimation

GSH ..... Reduced Glutathione

GST ..... Glutathione S-transferase

HCN ..... Hydrogen Cyanide

hCS ..... Human Citrate Synthase

HPLC-MS ..... High-Performance Liquid Chromatography-Mass Spectrometry

ID-RCDs ..... Intradiol Ring-Cleavage Dioxygenases

IPTG β-D-1-thiogalactopyranoside

ITC ..... Isothermal Titration Calorimetry

KAN ..... Kanamycin

k_{cat} ..... Turnover Number

kDa ..... kiloDaltons

K_m ..... Michael-Menten Constant

LB ..... Luria-Broth

LB-KAN ..... Luria-Broth Kanamycin

LIC ..... Ligation Independent Cloning

Mbs ..... Mega bases

MBP ..... Maltose-binding Protein

mcsA ..... 2-methylcitrate Synthase

MW ..... Molecular Weight

NaCl ..... Sodium Chloride

NiNTA ..... Nickel-nitrilo-triacetic agarose resin

OAA ..... Oxaloacetate

OCN^- ..... Cyanate

P-CoA ..... Propionyl-CoA

PCR ..... Polymerase Chain Reaction

pCS ..... Pig Citrate Synthase

PDB ..... Protein Data Bank

Pro-CoA ..... Propionyl-CoA

RMS ..... Root Mean-Square

RMSD ..... Root Mean-Square Deviation

SERp ..... Surface Entropy Reduction

SDS ..... Sodium Dodecyl Sulfate

SDS-PAGE ..... Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SER-CAT ..... Southeast Regional Collaborative Access Team

SOC ..... Super Optimal Broth with Catabolite Repression

S. proteamaculans ..... *Serratia proteamaculans*

SpCyanase ..... *Serratia proteamaculans* Cyanase

SPR ..... Surface Plasmon Resonance

TEV ..... Tobacco Etch Virus

TLS ..... Translational/Libration/Screw

T_m ..... Melting Temperature

TNB 2-nitro-5-thiobenzoate

TSSM ..... Two-spotted spidermite

T. urticae ..... *Tetranychus urticae*

TuCyanase ..... *Tetranychus urticae* Cyanase

TuDioxygenase ..... *Tetranychus urticae* Dioxygenase

TuGST ..... *Tetranychus urticae* Glutathione S-transferase

V_{max} ..... Velocity Max

TVMV ..... Tobacco Vein Mottling Virus

CHAPTER 1

STRUCTURAL CHARACTERIZATION OF A EUKARYOTIC CYANASE FROM *TETRANYCHUS URTICAE*

1.1 BACKGROUND AND INTRODUCTION

Chelicerates are a basal lineage within the Arthropoda phylum and represent the second largest group of terrestrial animals after insects [6]. They comprise of horseshoe crabs, scorpions, spiders, mites and ticks and include various economically important species for human health and agriculture. Mites exhibit a diverse range of lifestyles including herbivory, predation, parasitism, detritivory and symbiosis [7, 8]. With 48,000 species described by the turn of the century [9] and a total estimate of 0.5 and 1 million species, mites are probably one of the most diverse animal groups [10]. The two-spotted spider mite (TSSM), *Tetranychus urticae,* is an agricultural herbivorous pest and is the first chelicerate for which the complete genome was sequenced [11]. Due to its rapid developmental rate (can reach adulthood from egg in 7 days at 30°C), easy laboratory rearing and strong research community, *T. urticae* is presently a versatile chelicerate model organism [12]. Moreover, its compact genome of 90 Mbp (54% of the genomic sequence encodes for proteins) makes it an attractive experimental system to investigate gene and protein function.

TSSM is one of the most polyphagous arthropods, feeding on more than 1,100 plant species including more than 150 agricultural crops [13]. As plants produce a vast number of secondary metabolites, whose composition varies qualitatively and quantitatively with

and between species [14], TSSM is able to overcome a plethora of plant defense systems [15]. TSSM infestations are controlled with acaricides of which the global market value is estimated at $1.6 billion [16]. The TSSM has the highest incidence of pesticide resistance, resulting in extreme difficulty of controlling populations by conventional chemical pest management [17]. Analyses of the TSSM genome have previously revealed an unprecedented proliferation of multi-gene families that are commonly associated with the xenobiotic metabolism. These lineage-specific expansions in families such as cytochrome P450 monooxygenases and ABC transporters are considered to strongly contribute to the mite's ability to quickly develop pesticide resistance and to adapt to a wide host plant range [11, 18, 19]. In addition, spider mite outbreaks and crop damage are strongly facilitated by high temperatures and drought stress, both of which are intensified by climate change [20]. With increased reproductive potential under conditions of global warming [21], combined with an outstanding ability to adapt to new crops and develop resistance to pesticides, the TSSM is becoming a high-risk pest threatening global crop security.

Many plant species, including a number of food plants, are cyanogenic and release toxic hydrogen cyanide (HCN) from non-toxic glycoside or lipid precursors upon herbivore attack [22, 23]. Previous studies have gathered a body of evidence showing that the TSSM overcomes cyanide toxicity by expressing genes of foreign origin within its genome acquired through the process of horizontal gene transfer [24, 25].

Two mutually non-exclusive cyanide detoxification pathways, catalyzed by two horizontally transferred genes, have been proposed for TSSM. Cyanide might be first oxidized to cyanate (OCN^-) which is then catabolized by the horizontally acquired *T. urticae* cyanase enzyme into ammonia and carbon dioxide [24]. It was shown that the

cyanase gene was transcribed in all mite feeding stages (larvae, nymphs and female adults) and that gene-expression was host plant dependent. However, a later study revealed the presence of yet another gene from bacterial origin that codes for a β-cyanoalanine synthase enzyme that directly detoxifies cyanide through the formation of the amino acid derivative β-cyanoalanine [25]. This raises the question whether cyanide detoxification is the main function of *T. urticae* cyanase (in this manuscript further referred to as TuCyanase). The genomes of some non-phytophagous mites also possess cyanase genes [15]. Cyanate is also formed by the dissociation of carbamoyl phosphate and inhibits further synthesis of the compound by interacting with carbamoyl phosphate synthase. As carbamoyl phosphate is a main substrate for arginine and pyrimidines biosynthesis, cyanase might be indirectly involved in the regulation of their biosynthesis by changing cyanate concentrations [22, 23, 26–28]. To better understand the biological function of the horizontally transferred cyanase genes and their protein products, it is of crucial importance to understand the quaternary structures and how these evolved after a horizontal transfer event.

Cyanases in nematodes, plants and fungi are not very well characterized and the enzyme from *E. coli* is the most studied representative of this group of enzymes [26, 29–34]. It was shown that the *E. coli* cyanase (here referred to as EcCyanase) is a decameric protein that uses bicarbonate and cyanate as substrates (Figure 1.1) [35]. The enzyme is inhibited by various anions such as chloride, azide, nitrate, oxalate, oxaloacetate and malonate. Moreover, bicarbonate at high concentrations is responsible for substrate inhibition of the enzyme. Several structures of the *E. coli* enzyme have been determined [34]. At the time of this publication, the only other cyanase structure present in the Protein Data Bank (PDB) originates from the bacteria *Serratia proteamaculans* (SpCyanase) [36].

The TuCyanase presented here is the first mite cyanase that has had its structure determined, and is also the very first structure of any eukaryotic cyanase. This manuscript discusses in detail the differences between bacterial and TuCyanase which have a similar overall fold, despite relatively low sequence identity. Our structural and functional data provide a deeper understanding on the role of cyanase in *T. urticae* and other mites.

1.2 Purification of recombinant *T. urticae* Cyanase

Originally, tetur28g02430 (TuCyanase) was ordered from DNA 2.0 (Menlo Park, CA) in the expression vector pJExpress411 (Kanamycin resistance (KAN)) that has an isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible T7 promoter. The gene was synthesized to contain an N-terminal cleavable 6xHis-tag for ease of purification with the Tobacco Etch Virus (TEV) cut site MHHHHHHSSGVDLGTENLYFQ/SGSG where the cut site is shown with a slash. DNA was transformed into BL-21 (DE3) *E. coli* via heat shock and cells were plated on Luria-broth (LB)-KAN (50 µg/mL) plates at 37ºC for 16 hours. For inoculation, 10 mL of starter culture was used in 1.0 L of LB containing 50 µg/mL KAN. Cultures were shaken at 37ºC until an OD of 0.8 was reached and then cooled down to 16ºC for protein expression with 0.4 mM IPTG for 16 hours. Cell pellets were harvested by spinning cultures in a Beckman Coulter Ultracentrifuge (Indianapolis, IN) at 4ºC which were immediately frozen at -80ºC until needed further.

Pellets were resuspended in lysis buffer (50 mM Tris pH 7.4, 500 mM NaCl, 2% glycerol, 20 mM β-mercaptoethanol (β-ME), 10 mM imidazole) using 5.0 mL buffer per gram of pellet and lysed by sonication using a Branson Sonifier 450 (ThermoFisher, Grand Island, NY). Cell lysate was spun down in a Beckman Coulter Ultracentrifuge (Indianapolis, IN) at 4ºC and supernatant was loaded onto a 5.0 mL resin bed of HisPur NiNTA Resin (ThermoScientific, Grand Island, NY) equilibrated in wash buffer (50 mM

Tris pH 7.4, 500 mM NaCl, 2% glycerol, 20 mM β-ME, 30 mM imidazole). The column was washed several times with wash buffer and protein was eluted with elution buffer (50 mM Tris pH 7.4, 500 mM NaCl, 20 mM β-ME, 250 mM imidazole). Elutions were collected in 1.5 mL fractions and run on a 12% SDS-PAGE gel to determine fractions that contain protein. Elutions containing protein were put into SnakeSkin Dialysis Tubing with a 3,000 molecular weight (MW) cutoff (ThermoScientific, Grand Island, NY) and dialyzed in dialysis buffer (10 mM Tris pH 7.4, 150 mM NaCl, 5 mM β-ME) at 4ºC for 16 hours. TuCyanase was concentrated using an EMD Millipore Amicon Ultra-15 concentrator (Billerica, MA) with a 3,000 MW cutoff, and concentration was determined by A_{280} using the molar extinction coefficient 18,910 $M^{-1}cm^{-1}$ and MW 18,540 as determined by the ExPASy ProtParam tool [37, 38].

Purification of TuCyanase with N-terminal his-tag resulted in very poor yields of protein (around 1.0 mg per liter culture). Figure 1.2A shows a 12% SDS-PAGE gel for Ni-NTA purification results of this protein. To increase yield, the fusion protein maltose-binding protein (MBP) was used in pMBPcs1.

1.3 Construction of pMBPcs1 and Cloning of *T. urticae* Cyanase

Construction of pMBPcs1

The base vector for making pMBPcs1 was pMCSG71 [39] from DNASU Plasmid Repository (Tempe, AZ). This pMCSG71 vector expresses fusion protein as follows: MBP, Tobacco Vein Mottling Virus (TVMV) cut site, Ligation Independent Cloning (LIC) site, TEV cut site, 6xHis-tag where the LIC site is the location of gene insertion. The end product pMBPcs1 was designed to express the fusion protein as follows: MBP, AlaAlaSer-linker, TuCyanase (LIC site), TEV cut site, 6xHis-tag. The megaprimers labeled as pCSAA-F and pCSAA-R in Table 1.1 were used to knockout the TVMV cut site in

pMCSG71 and replace it with two alanine codons (AGCAGC); the serine of the AAS linker was added to the 5' end of the TuCyanase insert later in the process. Primers were designed using the program on rf-cloning.org [40]. First, megaprimers were synthesized in a 50 µL primary reaction by mixing 10 µM of each primer (pCSAA-F and pCSAA-R) with phusion polymerase following the manufacturer's setup protocol (NEB, Ipswich, MA) and thermalcycling as follows: initial 98ºC for 30 seconds, 15 cycles of 98ºC for 10 seconds, 58ºC for 15 seconds, 72ºC for 10 seconds, a final extension at 72ºC for 30 seconds and an infinite hold at 12ºC.

After the primary polymerase chain reaction (PCR), secondary PCR was performed by adding 1.0 µL of 100 ng pMCSG71 to the primary PCR and thermalcycling as follows: initial 98ºC for 30 seconds, 35 cycles of 98ºC for 10 seconds, 64ºC for 30 seconds, 72ºC for 6 minutes, a final extension at 72ºC for 10 minutes and an infinite hold at 12ºC. Twenty units of DpnI (NEB, Ipswich, MA) were added to the secondary PCR and incubated at 37ºC for 2 hours. DpnI was then inactivated at 80ºC for 20 minutes. For transformation, 20 µL of secondary PCR was added to chemically competent DH5-α *E. coli* cells and heat shocked at 42ºC for 45 seconds. Cells were allowed to recover by adding 950 µL of super optimal broth (SOC) media and incubated at 37ºC for 1.0 hour with shaking. Cells were plated on LB-ampicillin (AMP) plates (50 µg/mL) and grown at 37ºC for 16 hours. Clones were miniprepped with a GeneJET Plasmid Miniprep Kit (ThermoFisher, Grand Island, NY) and sent to EtonBioscience (Research Triangle Park, NC) to confirm the correct sequence with T7 forward and T7 reverse primers.

CLONING OF TUCYANASE INTO PMBPCS1

After the correct sequence was confirmed for pMBPcs1 (Figure 1.3), TuCyanase

(Uniprot: T1KZQ3) from DNA 2.0 (Menlo Park, CA) was cloned into it using Ligation Independent Cloning. For LIC, pMBPcs1 was amplified at the LIC site to generate blunt ends for T4 DNA Polymerase 3' → 5' exonuclease activity using the primers pMBPcs1-LIC-F and pMBPcs1-LIC-R in Table 1.1. The PCR was performed following the standard protocol for KOD Polymerase (EMD Millipore, Billerica, MA) with the addition of 1.0 M betaine monohydrate, 2.0 mM magnesium sulfate and 130 ng of pMBPcs1 in a 50 µL reaction with thermalcycling as follows: initial 95ºC for 2 minutes, 30 cycles of 95ºC for 20 seconds, 54ºC for 10 seconds, 70ºC for 6 minutes, a final extension at 70ºC for 10 minutes and an infinite hold at 12ºC. The reaction was run on a 1.0% agarose gel and the band corresponding to pMBPcs1 was gel excised using a GeneJET Gel Extraction Kit (ThermoFisher, Grand Island, NY) and eluted with sterile water.

TuCyanase was prepared for LIC into pMBPcs1 by using pMBP-Cyan-F and pMBP-Cyan-R primers listed in Table 1.1; the forward primer for TuCyanase was designed to introduce a serine residue at the N-terminal of cyanase so the fusion protein linker would be three residues consisting of AlaAlaSer. For PCR, TuCyanase in pJExpress411 from DNA 2.0 (Menlo Park, CA) was used as the template (100 ng) and the protocol for Phusion polymerase was used (NEB, Ipswich, MA). The conditions for thermalcycling were as follows: initial 98ºC for 30 seconds, 35 cycles of 98ºC for 10 seconds, 70ºC for 30 seconds, 72ºC for 25 seconds, a final extension at 72ºC for 1.0 minute and an infinite hold at 12ºC. PCR product was run on a 1.0% agarose gel and gel excised using a GeneJET Gel Extraction Kit (ThermoFisher, Grand Island, NY) and eluted with sterile water.

Sticky ends for LIC were made by incubating gel excised PCR products with T4 DNA polymerase (NEB, Ipswich, MA). A 40 µL reaction containing 600 fmoles of

pMBPcs1, 5.0 mM dithiothreitol (DTT), 2.5 mM deoxyadenosine triphosphate (dATP), 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase, and a 40 µL reaction containing 1500 fmoles of tetur28g02430, 5.0 mM DTT, 2.5 mM deoxythymidine triphosphate (dTTP), 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase were incubated at room temperature for 30 minutes. The T4 DNA polymerase was then inactivated at 75ºC for 20 minutes and the reactions were mixed 10:10 at room temperature for 5.0 minutes. Next, 1.0 µL of 25 mM ethylenediaminetetraacetic acid (EDTA) was added and the reaction was incubated another 5.0 minutes at room temperature. Transformation was performed with heat shock as mentioned previously after adding the 20 µL LIC reaction to the cells, and cells were plated on LB-AMP (50 µg/mL) plates at 37ºC for 16 hours. Clones were miniprepped with a GeneJET Plasmid Miniprep Kit (ThermoFisher, Grand Island, NY) and insertion of TuCyanase was first confirmed by restriction digest with HindIII and XbaI following the manufacturer's instructions (NEB, Ipswich, MA). After restriction digest, clones with correctly sized fragments were sent to EtonBioscience (Research Triangle Park, NC) for sequencing with T7 forward and T7 reverse primers. Correct clones were transformed via heat shock into BL-21 (DE3) *E. coli* cells.

1.4 Purification of MBP-Cyanase

Purification of MBP-Cyanase was the same as previously mentioned for cyanase except it yielded much higher quantities of protein (around 30 mg per liter culture, Figure 2B) and dialysis buffer contained 5.0 mM maltose. MBP-Cyanase concentration was determined by A_{280} using the molar extinction coefficient 85,260 $M^{-1}cm^{-1}$ and MW 60,834 as determined by the ExPASy ProtParam tool [37, 38]. MBP-Cyanase was further purified with a Superdex 200PG column attached to an ÄKTA Pure FPLC system (GE Healthcare, Marlborough, MA) equilibrated in FPLC buffer (10 mM Tris pH 7.4, 150 mM NaCl, 5.0

mM maltose). Gel filtration results are shown in Figure 1.4. Interestingly, MBP-Cyanase eluted off the column as a single peak around a size of 670 kiloDaltons (kDa) (based on Bio-Rad (Hercules, CA) Thyroglobulin 670 kDa standard), which would indicate the fusion protein is in a decameric state.

1.5 Confirmation of Cyanase Activity of MBP-Cyanase

To determine if the MBP-cyanase fusion protein had enzymatic activity, MBP-Cyanase was incubated in a 500 µL reaction of 100 mM Tris pH 7.5 with 10 mM sodium cyanate and 10 mM sodium bicarbonate; all chemicals used were purchased from Sigma-Aldrich. To detect the presence of ammonia, one product of the cyanase reaction, a piece of Whatman pH paper dampened with diH_2O was placed above the opening of the 1.5 mL tube containing the 500 µL reaction mix. After ~20 seconds, the pH paper turned green which would indicate the presence of base i.e. ammonia (data not shown). As negative controls, reactions were made as follows: only sodium bicarbonate, only sodium cyanate, both sodium bicarbonate and sodium cyanate, sodium bicarbonate with Tris pH 7.5 and MBP-Cyanase, sodium cyanate with Tris pH 7.5 and MBP-Cyanase. These controls were all tested with pH paper in the same manner. All of the controls failed to change the color of the pH paper, except the Tris reaction mix containing sodium cyanate and MBP-Cyanase. It is possible that residual bicarbonate (from atmospheric CO_2) was present after the purification of MBP-Cyanase, even though bicarbonate was not used during purification.

1.6 Crystallization of MBP-Cyanase

MBP-Cyanase crystallization experiments were performed at room temperature using the sitting-drop vapor diffusion method and MRC 2-drop 96-well crystallization plates (Hampton Research, Aliso Viejo, CA). Recombinant protein (~13 mg/mL) was

mixed with mother liquor in a 1:1 ratio. A condition of 0.1 M Tris pH 7.5, 15% w/v PEG6000 produced crystals after about 2 months. Crystals were cryo-cooled in liquid N_2.

1.7 Data Collection, Structure Determination and Refinement of TuCyanase

Data was collected using remote access to Southeast Regional Collaborative Access Team (SER-CAT) 22ID at the Advanced Photon Source (APS), Argonne National Lab (Argonne, IL). The diffraction images were processed with the HKL-2000 software package [41]. Data collection statistics are reported in Table 2. Molecular replacement was performed using MOLREP [42] integrated with HKL-3000 [43]. The crystal structure *E. coli* cyanase (PDB code: 2IV1) was used as the starting model for molecular replacement. The initial model was rebuilt using BUCCANEER [44] and CCP4 package [45]. Refinement was performed using REFMAC [46] and HKL-3000. Non-crystallographic symmetry was used during the whole process of refinement. TLS refinement was used during the last stages of the refinement and the TLS Motion Determination server was used for partitioning protein chains into the rigid bodies undergoing vibrational motions [47]. The model was updated and validated with COOT [48]. MOLPROBITY was used in the final steps of the model validation [49]. The final model together with structure factors were deposited to the Protein Data Bank [50] with the accession number 5UK3. Dali [51] and PDBeFold [52] were used to identify structural homologs of the *T. urticae* cyanase. PDBePISA [53] was used to analyze oligomeric assembly formed by the enzyme. Figures were created with PyMOL [54].

1.8 Oligomeric Assembly of TuCyanase

The recombinant MBP-cyanase protein was shown to be enzymatically active by detecting the accumulation of the reaction product ammonia in the reaction mix (data not

shown). Although the recombinant protein was unstable at room temperature after 4-5 days (data not shown), we were still able to crystallize TuCyanase. The model presented here is the first mite cyanase structure determined and the first structure of any eukaryotic cyanase (Figure 1.5). The comparison of *T. urticae* cyanase to the two known bacterial cyanase structures revealed that a homo-decameric state and active site residues were conserved, but there were some variabilities in the dimer interfaces and protein domains. Consequently, TuCyanase appears to be less compact than the prokaryotic cyanases, EcCyanase and SpCyanase.

Structural analysis revealed that only the cyanase part of MBP-Cyanase is present in the crystal. The crystal form contained 10 protein chains in the asymmetric unit (AU). The quaternary structure was determined to be decameric which is in agreement with gel filtration results (Figure 1.4), as well as the oligomeric forms observed for EcCyanase [34] and SpCyanase [36]. The decamer may be treated as a pentamer of dimers (Figure 1.6A). A single protein chain is composed of two segments: an N-terminal domain (helices α1-α5) and the cyanate lyase domain (all β-strands and helix α6). The role of the N-terminal domain is not well understood and has a similar structure to some proteins involved in DNA binding. For example, a structural domain part of cytolysin repressor 2 from *Enterococcus faecalis* (PDB code: 2XJ3) overlaps with an RMSD of 1.8 Å (over 58 Cα atoms). The cyanate lyase domain of TuCyanase contained all amino acids necessary to form the active site.

The determined cyanase structure from *T. urticae* included mostly residues 22-156 with a small variability of the sequence span between the 10 protein chains in the asymmetric unit. Comparison of the TuCyanase structure with the bacterial homologs from

E. coli and *S. proteamaculans* revealed a significant degree of structure similarity despite a relatively low sequence identity (Figure 1.5, 1.7). The sequence identity/similarity between *T. urticae* and *E. coli* or *S. proteamaculans* proteins is 35%/62% or 34%/56% respectively, with the N-terminal domains being less conserved in comparison with the cyanate lyase domains. Superposition of TuCyanase and EcCyanase or SpCyanase structures resulted in RMSD values of 1.8Å (over 129 Cα atoms; PDB code 1DW9) or 1.7Å (over 128 Cα atoms; PDB code 4Y42), respectively.

1.9 Dimer Interface of TuCyanase

As mentioned before, functionally active cyanase enzymes of *T. urticae* and bacteria form a decamer that may be described as a pentamer of dimers. The dimer interface (interface 1, Figure 1.6B) has a very large area (~3050 $Å^2$ for *T. urticae* and ~3250 for *E. coli*). In addition, each protein chain forming the dimer interacts with four other protein chains from neighboring dimers (Figure 1.6B). In TuCyanase, interface 2 (blue and yellow chains) had a relatively large area (~1200 $Å^2$) compared to interfaces 3 (blue and green), 4 (blue and brown) and 5 (blue and red) which were relatively small (~260 $Å^2$ each); a significant degree of interface 2 was formed by the C-terminal end of a neighboring unit. The decamer is barrel-shaped with a height of approximately 70 Å and diameter of 75 Å. Compared to TuCyanase, EcCyanase is more compact with larger interface areas between the monomers and neighboring dimers: interface 1 is ~3250 $Å^2$, interface 2 is ~1400 $Å^2$, interface 3 and 4, are ~800 $Å^2$, interface 5 is ~300 $Å^2$). Despite differences in compactness and interface areas, the decameric barrel of TuCyanase and EcCyanase have a similar height and diameter.

The major difference in the oligomeric forms observed for proteins from *T. urticae* and *E. coli* (as well as *S. proteamaculans*) is related to the fact that the spider mite protein structure does not possess the first 20 amino acids. These missing amino acids correspond to a fragment of helix α1 in the *E. coli* structure and N-terminal part that does not have a defined secondary structure; it is placed on the "edge" of the barrel mediating contact between dimers. The presence of the N-terminal amino acids in the *E. coli* structure increases the surface area interface between dimers by almost 40% in comparison with the corresponding interactions in the structure for *T. urticae.* In addition, in *E. coli* cyanase the N-terminal part of the protein is packed in the barrel, thus limiting the size of the substrate channel through the decamer. The comparison of the TuCyanase and EcCyanase also shows that the helices α1 and α5 for EcCyanase are shifted significantly towards the center of the barrel.

1.10 Active Site of TuCyanase

Structural and sequence analyses of TuCyanase, EcCyanase and SpCyanase (Figure 1.8) show that the most conserved amino acids are located in the areas mediating interactions between the dimeric assemblies. Two dimers are necessary to create an active site and there are five active sites in one decamer. Mapping of the sequence conservation onto the structure of TuCyanase clearly showed that the cyanate lyase domain is significantly more conserved than the N-terminal domain. The active site is composed of amino acids provided by four protein chains (Figure 1.9A). The studies of cyanase from *E. coli* identified the following residues as being critical for the protein activity: R96, E99 and S122 [34]. These residues correspond to R100, E103 and S126 from TSSM and are

conserved in all studied cyanases (Figure 1.9B) [24, 28, 55], including the recently discovered silverleaf whitefly homolog (R119, E122 and S145).

The cyanase active site is quite unusual as not only is it formed by residues originating from four protein chains, it also exhibits a pseudo-two-fold symmetry (Figure 1.9C). The number and composition of the active sites are in agreement with the kinetic and binding studies that show 0.5 stoichiometry in the case of substrate and inhibitor binding [30]. At the entrance to the active site, K149 may help to attract negatively charged substrates. In addition, the entrance to the active site was aligned with hydrophobic residues like I124, I128 and L151 that limit the size of the molecules/ions that can access the active site. It is not clear what the role of S126 is during the reaction, but it is believed that R100 is responsible for substrate binding, while E103 positions the arginine in proper orientation. There are also two proline residues that were conserved in the TuCyanase, EcCyanase and SpCyanase. The first proline (P96) changes the direction of polypeptide chain and initiates helix α6 that contains critical catalysis residues R100 and E103. The second proline (P152) ends a short β-strand (β2b). It is also worth mentioning that the regions between β1a and β1b (S126-A127-I128), as well as β2a and β2b (K149), contained residues that are highly conserved among cyanases.

1.11 Discussion

Although the results presented here provide a better understanding of the horizontally acquired TuCyanase, it remains unclear how the substrates and products enter and leave the active site, respectively. It was proposed that solvent exposed residues L151 (Figure 1.9C) control the access to the active site [34, 36], and the reaction occurs through a random sequential mechanism [32]. The active sites have an overall net positive charge which provides some explanation as to how the anionic substrates bind. Furthermore, the

active site residues consist of arginine, serine and glutamic acid residues and are conserved among all cyanases including those from bacteria, fungi and *T. urticae* presented here [22, 24, 28, 34, 36, 55]. While there is an agreement on the identity of major cyanase substrates (Figure 1.1), it is not clear what source(s) of protons is necessary for the reaction. EcCyanase and TuCyanase display a bell-shaped pH dependence of activity with maxima at 7.4 and 7.6 respectively [24, 29]. It seems that serines 126 (from two chains) are the only amino acids in the vicinity of the active site (Figure 1.9C) that may participate in a proton transfer. However, it cannot be excluded that these residues play a different role during the reaction.

Decomposition of cyanate and bicarbonate by cyanase (Figure 1.1) generates two molecules of carbon dioxide. The *E. coli* cyanase operon also includes a *cynT* gene that codes for a carbonic anhydrase which aids in the recycling of the produced carbon dioxide to bicarbonate [26, 35, 56]. The role of the anhydrase is not only important for carbon dioxide recycling, but more importantly, it prevents CO_2 from escaping from the cell. In the absence of the anhydrase, the carbon dioxide diffuses faster from the cell than it would being hydrated to bicarbonate [56]. The escape of CO_2 would lead to a depletion of cellular bicarbonate that in turn would affect various metabolic processes using HCO_3^-/CO_2 as substrates [57]. Furthermore, in *E. coli* it was shown that deletion of the *cynT* gene results in growth inhibition and an inability to degrade cyanate [57]. The importance of the cyanase/anhydrase tandem in the detoxification of cyanate was also suggested in fungal *Fusarium oxysporum* species [58]. Therefore, we speculate that a similar situation takes place in *T. urticae.* It is also likely that the protons generated during CO_2 recycling are used during the reaction catalyzed by cyanase (Figure 1.1). Carbonic anhydrases have been

identified within the animal kingdom and belong to the α- or β-carbonic anhydrase family [59, 60]. However, β-carbonic anhydrases with InterPro protein domains IPR001765 could not be detected, although 9 α-carbonic anhydrases with InterPro domain IPR001148 were found [11, 61].

1.12 TABLES

Table 1.1: Primers used for cloning TuCyanase. Lowercase letters show the two alanine insertions for pMBPcs1-F/R. Primers pMBPcs1-LIC-F/R were used to amplify pMBPcs1 at the LIC region to generate linear DNA for cloning. In pMBP-Cyan-F, lowercase letters show the introduction of a serine (codon AGT) at the N-terminal of tetur28g02430. Furthermore, the lowercase letters in pMBP-Cyan-R shows the C-terminal end of tetur28g02430 with the addition of an alanine (codon GCT) which was needed for LIC.

Primer	Sequence
pMBPcs1-F	5' CCTGAAAGACGCGCAGACTAATTCGgctgctGGGGA 3'
pMBPcs1-R	5' CGGATTGGAAGTACAGGTTCTCCCCagcagcCGAAT 3'
pMBPcs1-LIC-F	5' GGGGAGAACCTGTACTTCCAATCCGCCGGG 3'
pMBPcs1-LIC-R	5' AGCAGCCGAATTAGTCTGCGCGTCTTTCAG 3'
pMBP-Cyan-F	5' TTCGGCTGCTagtcgtatctacagccgcttgtttcaaa 3'
pMBP-Cyan-R	5' GGTTCTCCCCagcgtccagctggcctttgtaatacgg 3'

Table 1.2: Summary of data collection and structure refinement statistics for TuCyanase. Values in parentheses refer to the highest resolution shell. Abbreviations: AU – Asymmetric Unit.

PDB Accession code	5UK3
Data collection	
Diffraction source	Synchrotron (APS 22ID)
Wavelength (Å)	1.000
a, *b*, *c* (Å)	71.1, 81.6, 136.5
α, β, γ (°)	90, 100.4, 90
Space group	$P2_1$
Solvent content (%)	50
Protein chains in AU	10
Resolution range (Å)	40.00-2.80 (2.85-2.80)
Unique reflections	38010 (1906)
Redundancy	4.1 (4.2)
Completeness (%)	99.9 (100)
R_{merge}	0.069 (0.546)
R_{pim}	0.050 (0.350)
R_{rim}	0.102 (0.721)
CC half	(0.800)
Average $I/\sigma(I)$	22.5 (2.0)
Refinement	
R_{work}	0.221 (0.341)
R_{free}	0.240 (0.346)
Mean *B* value (Å^2)	95.8
B from Wilson plot (Å^2)	93.3
RMSD bond lengths (Å)	0.012
RMSD bond angles (°)	1.7
No. of amino acid residues	A = 135; B,E,H = 137; C,D,F,G,J = 138; I = 141
No. of water molecules	20
Ramachandran plot	
Most favored regions (%)	98.9
Additional allowed regions (%)	100.0

1.13 FIGURES

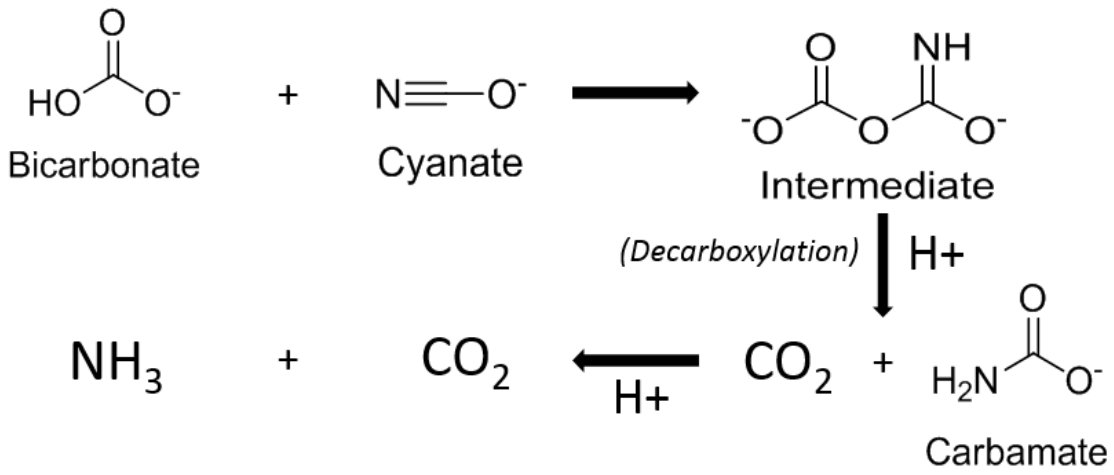

Figure 1.1: Reaction scheme of cyanase. Cyanase catalyzes the reaction of bicarbonate with cyanate to form an intermediate, which is decarboxylated to give carbamate and carbon dioxide. The end products of the reaction are ammonia and carbon dioxide (carbamate decomposes nonenzymatically) [35].

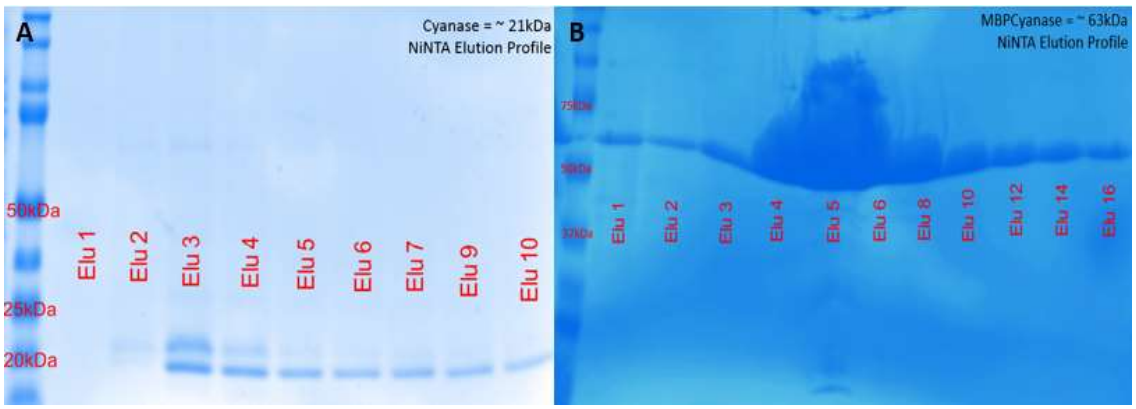

Figure 1.2: SDS-PAGE NiNTA purification of TuCyanase and MBP-Cyanase. A) SDS-PAGE NiNTA TuCyanase purification. **B)** SDS-PAGE NiNTA MBP-Cyanase purification.

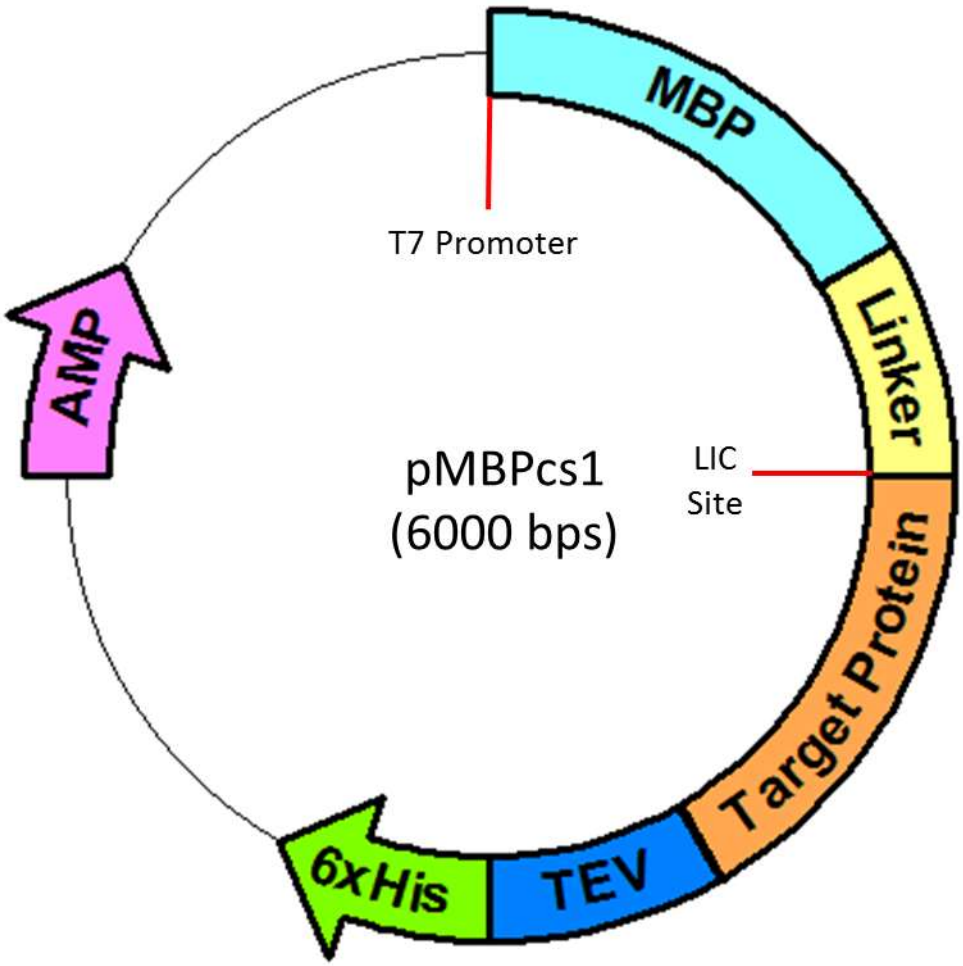

Figure 1.3: Plasmid map for pMBPcs1. The pMBPcs1 vector expresses fusion protein as MBP, AlaAlaSer Linker, Target Protein, TEV cut site, 6xHis-tag and operates under a T7 promoter. MBP – Maltose Binding Protein, TEV – Tobacco Etch Virus, AMP – Ampicillin resistance, LIC – Ligation Independent Cloning Site, bps – basepairs. Plasmid map was made with the program pDRAW32.

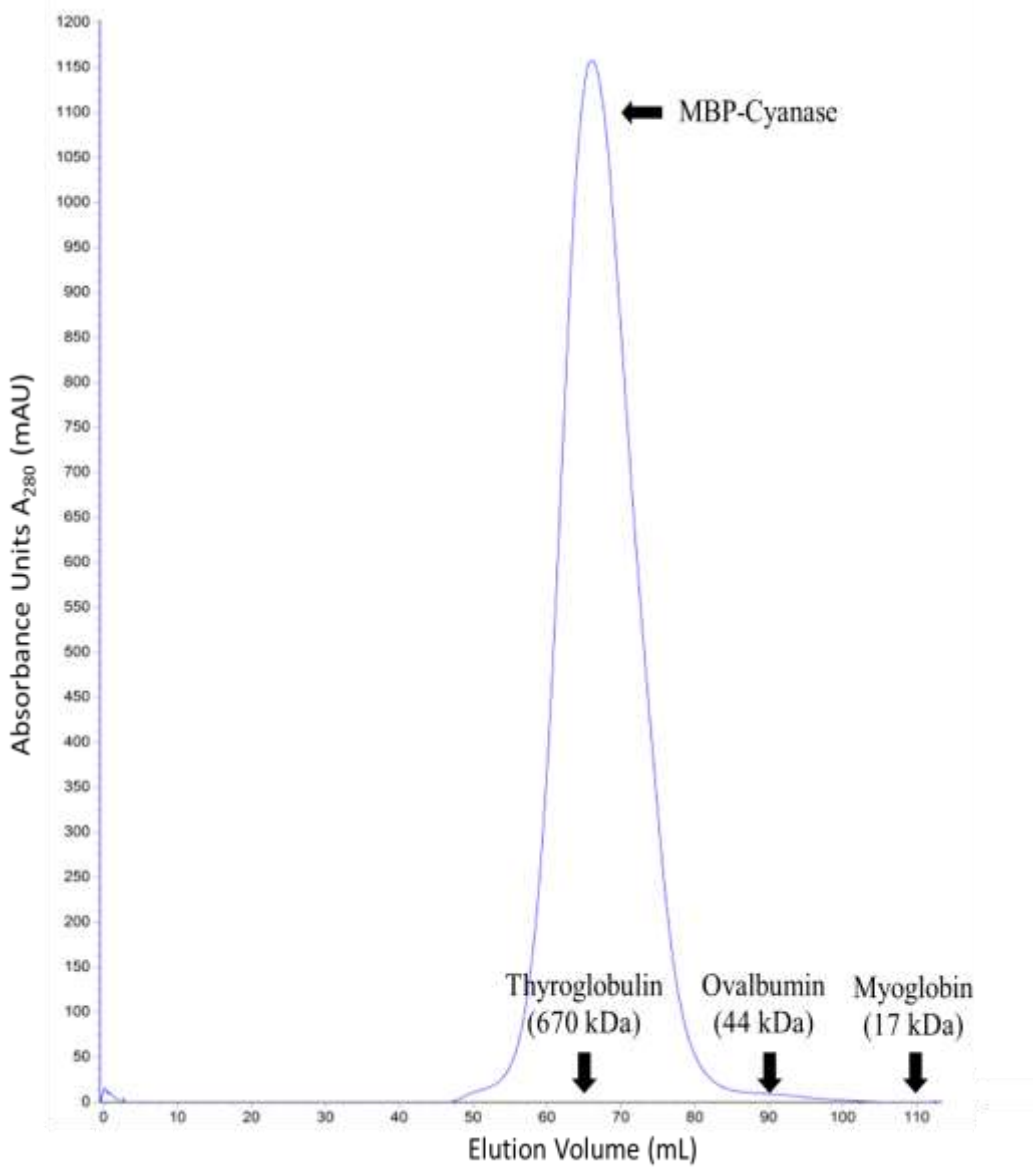

Figure 1.4: Gel filtration results for MBP-Cyanase. One molecule of MBP-cyanase has a molecular weight of ~60 kDa. Standards used were thyroglobulin, ovalbumin, and myoglobin.

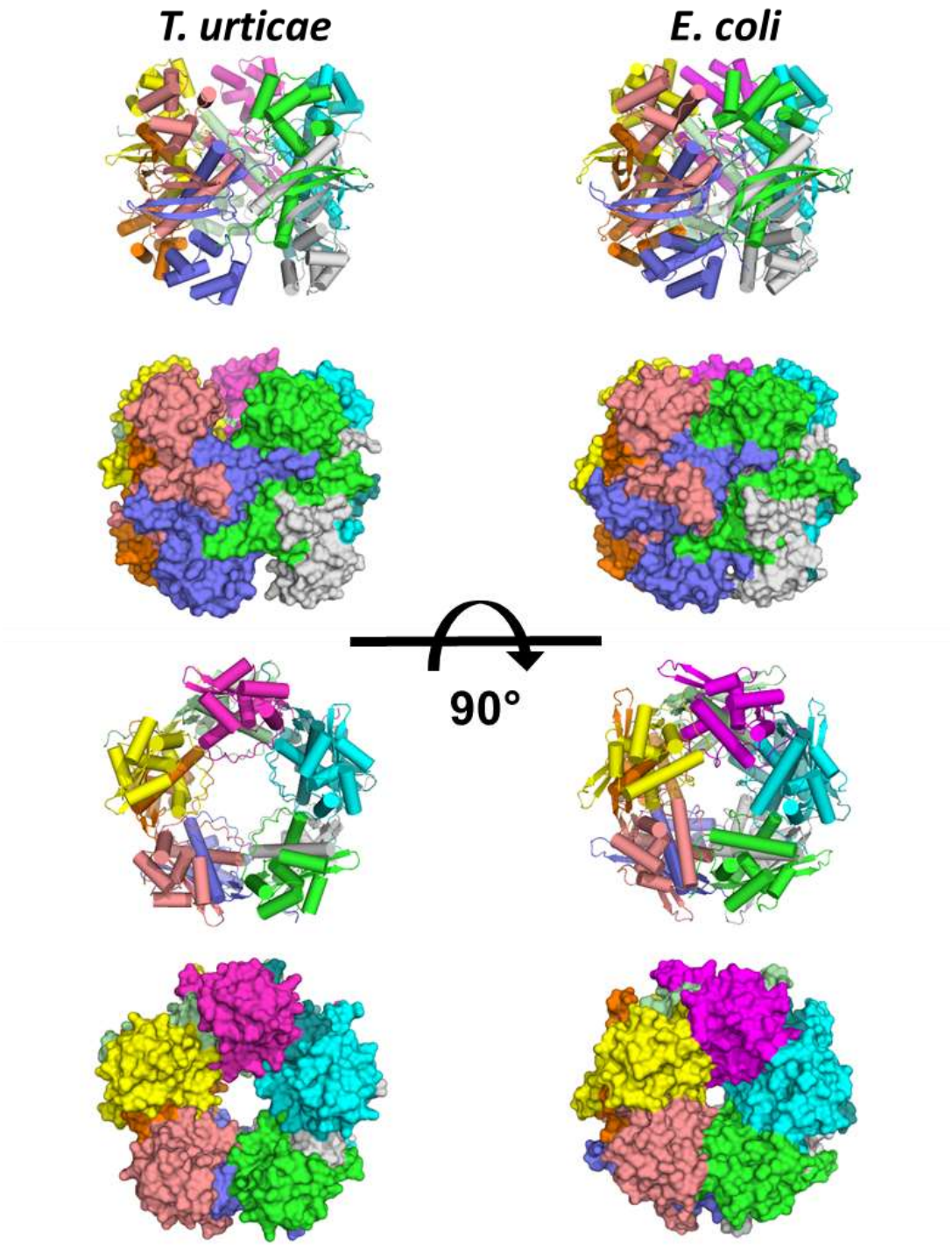

Figure 1.5: Oligomeric state of *T. urticae* and *E. coli* cyanases. Both form a pentamer of dimers and are shown in cartoon and surface representations with each monomer in a different color. Structures flipped 90° around the X-axis are also shown.

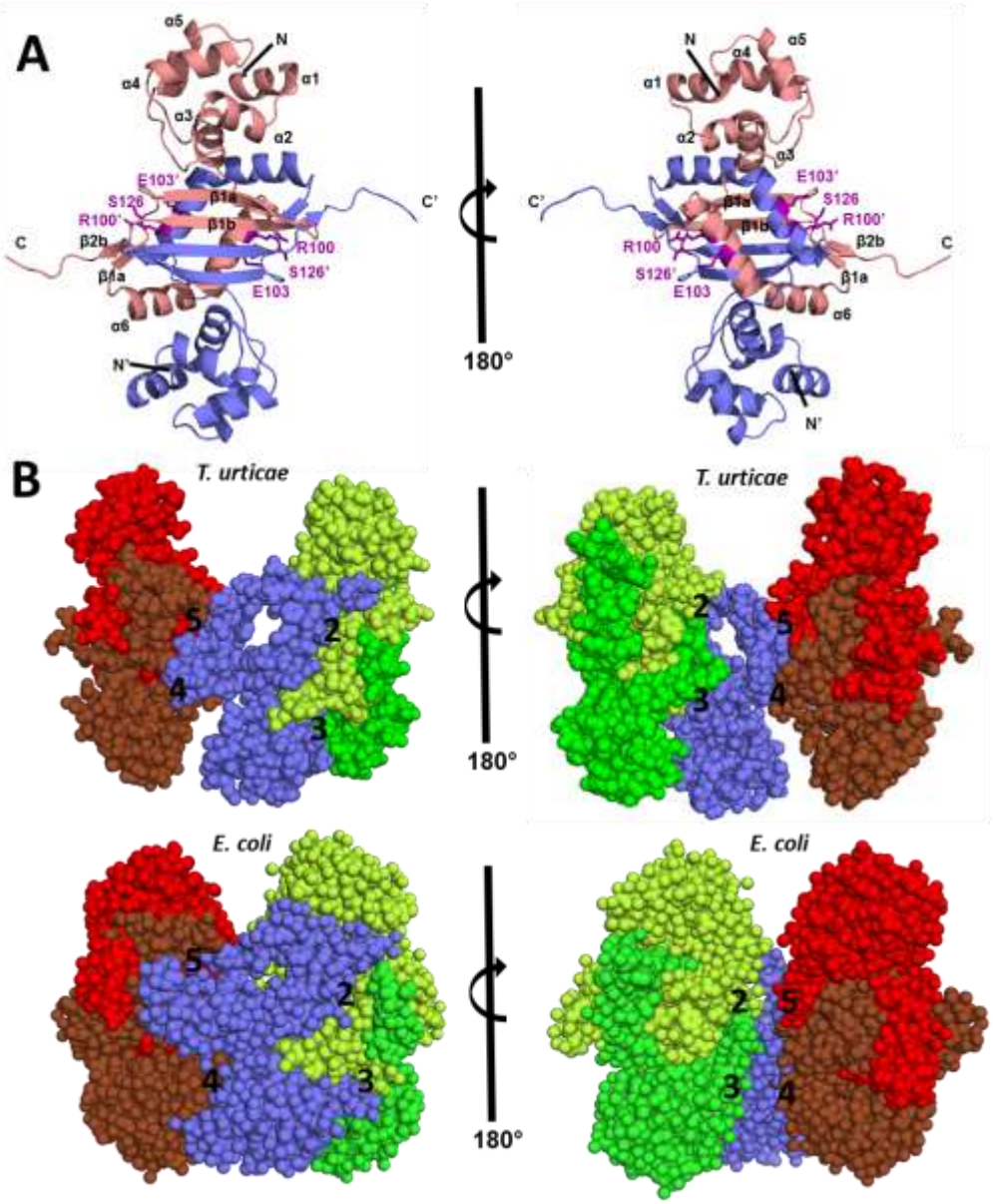

Figure 1.6: Cyanase dimer and surface interfaces. A) Structure of the TuCyanase dimer shown in cartoon representation. The dimer is shown in two orientations related by rotation of 180° along the Y-axis and each chain is depicted in a different color. Secondary structure elements are labeled on one of the chains forming the dimer. Residues forming active site are shown in stick representation (magenta) and are labeled. The dimer interface of two monomers for TuCyanase is ~3060 Å^2 and for EcCyanase is ~3250 Å^2. **B)** Space-filling models of TuCyanase and EcCyanase. The interfaces formed with neighboring dimers by one monomer (blue) are shown with the numbers 2, 3, 4 or 5. The area for surface interfaces 2, 3, 4 and 5 are ~1200 Å^2, ~270 Å^2, 270 Å^2 and 270 Å^2 for TuCyanase and ~1410 Å^2, 800 Å^2, 800 Å^2 and 295 Å^2 for EcCyanase, respectively. Surface interface areas were calculated using PDBePISA.

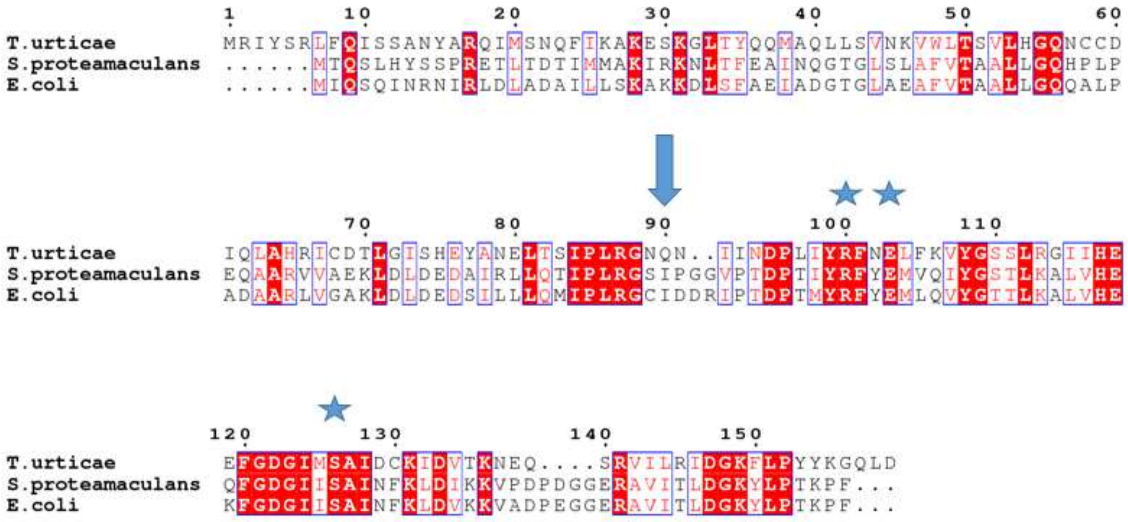

Figure 1.7: Amino acid sequence alignment of *T. urticae, S. proteamaculans and E. coli* cyanases. Conserved active sites residues are marked with blue stars, while the arrow indicates the end of the N-terminal domain and beginning of the cyanate lyase domain. Red boxes indicate identical residues whereas blue boxes indicate similar residues. The sequence identity/similarity between *T. urticae* and *E. coli* or *S. proteamaculans* proteins is 35%/62% or 34%/56% respectively.

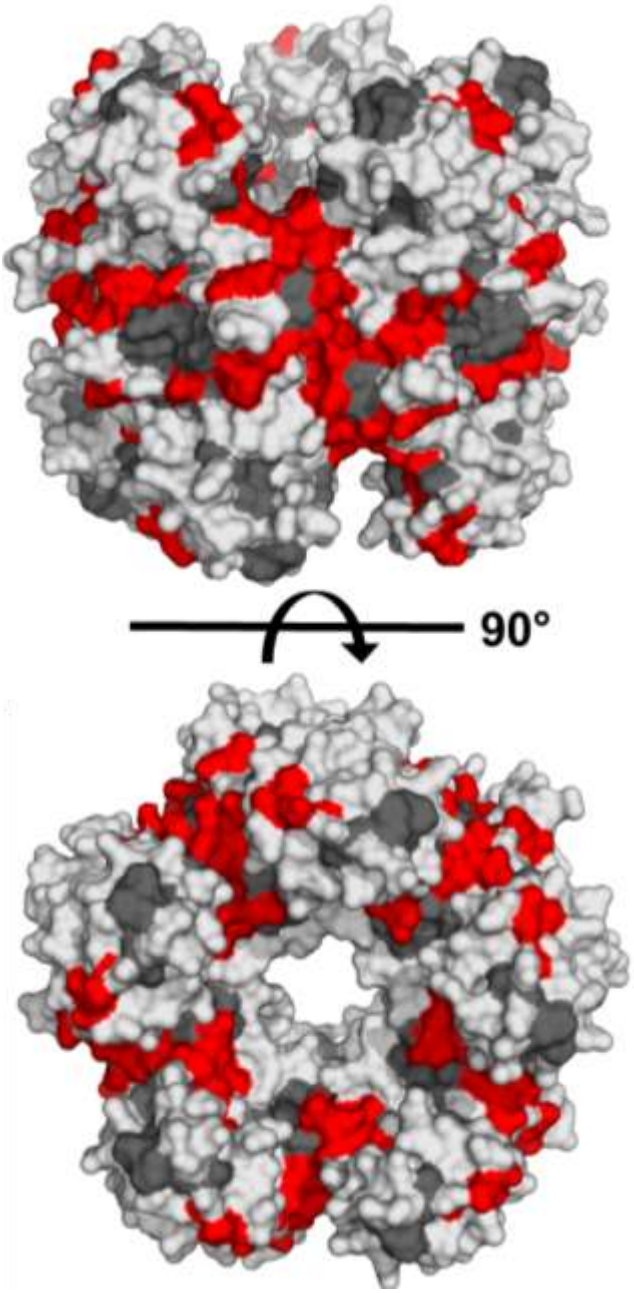

Figure 1.8: Amino acid sequence conservation between TuCyanase, EcCyanase and SpCyanase. Residues that are identical in all structures are colored in red and residues that are similar are colored in dark gray. Residues that are not identical or similar are colored white.

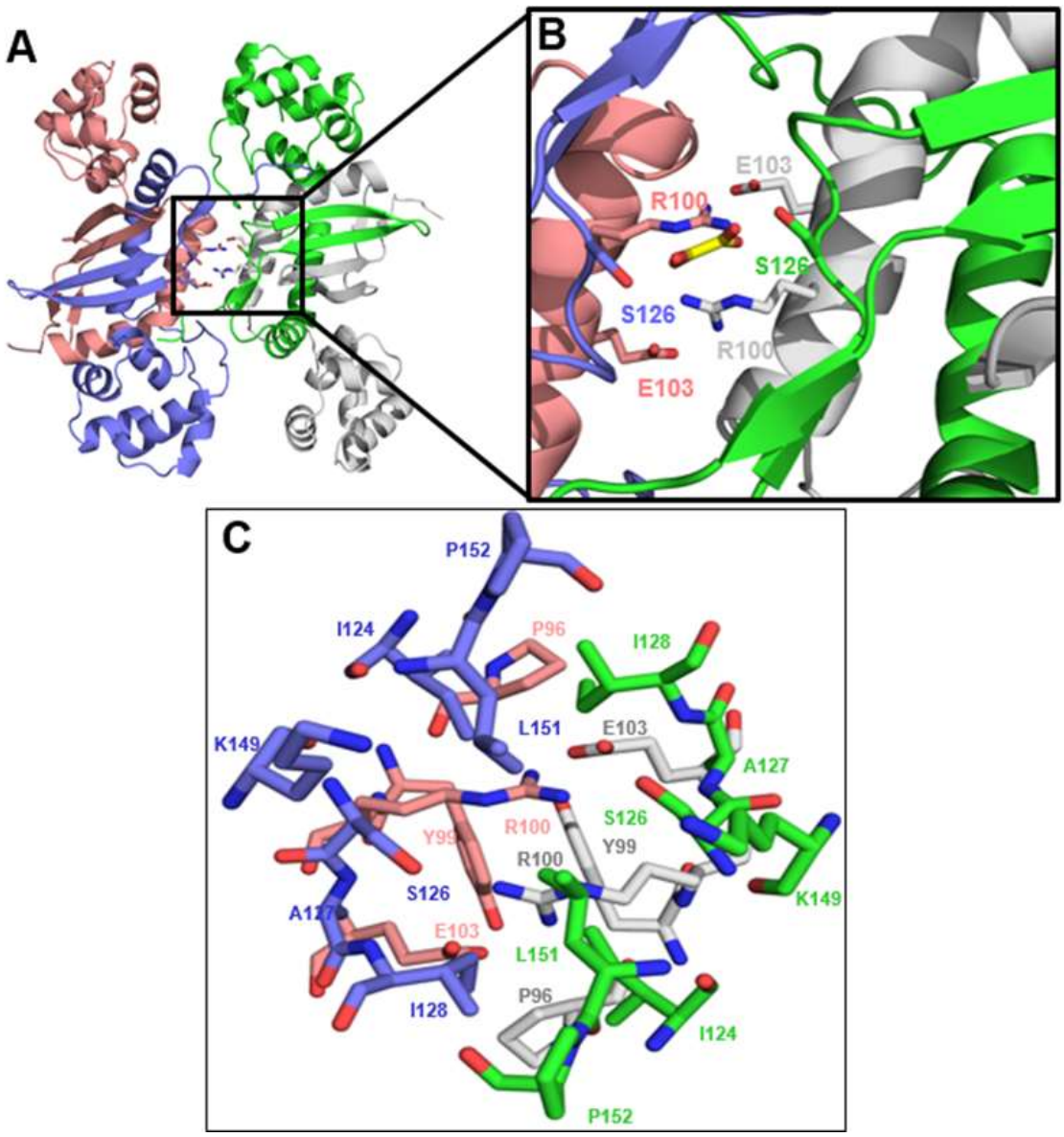

Figure 1.9 The active site of cyanase from *T. urticae*. A) Cartoon representation of active site cyanase formed from 4 monomers. **B)** The active site is formed by residues R100, S126 and E103 originating from four different protein chains (shown in four different colors). The residues forming the active site are shown in stick representation and are labeled. Oxalate (yellow) is modeled in the active site. The position of oxalate was derived from EcCyanase structure (PDB code: 2IU7). **C)** The active site of the TuCyanase is formed by residues from four adjacent protein chains (each colored differently) and all residues near the vicinity of the active site that were conserved in the enzymes from *T. urticae*, *E. coli* and *S. proteamaculans* are shown.

CHAPTER 2

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF A GLUTATHIONE S-TRANSFERASE FROM *TETRANYCHUS URTICAE*

2.1 BACKGROUND AND INTRODUCTION

Glutathione S-transferases (GSTs) are present in several trees of life (both prokaryotic and eukaryotic) and serve to protect cells from oxidative stress, endogenous toxins, organophosphates, organotins, carbamates, pyrethoids and other xenobiotics [2, 62, 63]. The thiol group of reduced glutathione (GSH) is conjugated to an electrophilic compound to make an excretable, water soluble product; alternatively, glutathione can act as a co-factor such as in a dehydrochlorination reaction [62]. There are several classifications of GSTs which are based on their primary amino acid sequence [62]. In insects, the current classifications of GSTs known are delta, epsilon, mu, omega, sigma, theta and zeta [64, 65]. GST diversity is very common in insects which may be attributed to alternative gene splicing, genetic rearrangements and local gene duplications that can result in altered substrate specificity [62, 63]. Insect GSTs are responsible for the detoxification of several insecticides used in agriculture such as 1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane (4,4'-DDT), chlorfenapyr and permethrin [2, 62]. Furthermore, in addition to gene modifications, some insect GSTs can be upregulated in the presence of insecticides and display slow turnover of the insecticide(s), effectively causing sequestration and broad-spectrum resistance [63].

Almost all GSTs form homodimers or heterodimers with two active sites [62, 66]. In one special case, however, a GST in *Plasmodium falciparum* has been shown to form a low GSH affinity tetramer but a high affinity homodimer in the presence of increased GSH concentration [67]. The active sites of GSTs consist of two subsites: G-site (formed by N-terminal residues) and H-site (formed by C-terminal residues). The G-site consists of residues near the N-terminus and is responsible for binding GSH [66]. The main G-site residue involved with activating the reduced thiol of GSH can either be a cysteine, serine, threonine or tyrosine [66]. The H-site consists of residues near the C-terminus and binds hydrophobic compounds that are targets for GSH conjugation.

T. urticae is a polyphagous, worldwide agricultural pest that is rapidly developing insecticide resistance due to its rapid reproductive cycle and the widespread use of insecticides [2]. The growing insecticide resistance in *T. urticae* can be attributed to several cytochrome P450 and GST gene families [3]. Four GSTs from *T. urticae* have been characterized to date: TuGSTd10, TuGSTd14, TuGSTm09 [3] and TuGSTd05 [65]. The TuGSTd10, TuGSTd14 and TuGSTm09 enzymes exhibited affinity towards abamectin, an insecticide, and also peroxidase activity towards Cumene hydroperoxide [3]. TuGSTd05 exhibited activity towards the de-esterified metabolite of cyflumetofen, a more recently introduced insecticide [65]. A delta-class GST (tetur01g02230) is studied here and was recombinantly expressed in *E. coli*. Several compounds were tested by UV-Vis spectroscopy to determine *T. urticae* GST (TuGST) specificity and some binding cooperativity was observed. Furthermore, structural elements of a predicted structure for TuGST and implications for cooperativity and substrate binding are explored.

2.2 Expression and Purification of TuGST in pJExpress411

Initially, TuGST (tetur01g02230) was ordered from DNA2.0 in the IPTG-inducible expression vector pJExpress411 with a T7 promoter. The gene was synthesized to contain an N-terminal cleavable 6xHis-tag for ease of purification with the TEV cut site MHHHHHHSSGVDLGTENLYFQ/SGSG where the cut site is shown with a slash. DNA was transformed into BL-21 (DE3) *E. coli* via heat shock and cells were plated on KAN (50 µg/mL) plates at 37ºC for 16 hours. For inoculation, 10 mL of starter culture was used in 1.0 L of LB containing 50 µg/mL KAN. Cultures were shaken at 37ºC until an OD of 0.8 was reached and then cooled down to 16ºC for protein expression with 0.4 mM IPTG for 16 hours. Cell pellets were harvested by spinning cultures in a Beckman Coulter Ultracentrifuge (Indianapolis, IN) at 4ºC which were immediately frozen at -80ºC until needed further.

Pellets were resuspended in lysis buffer (50 mM Tris pH 7.4, 500 mM NaCl, 2% glycerol, 20 mM β-ME, 10 mM imidazole) using 5.0 mL buffer per gram of pellet and lysed by sonication using a Branson Sonifier 450. Cell lysate was spun down in a Beckman Coulter Ultracentrifuge (Indianapolis, IN) at 4ºC and supernatant was loaded onto a 5.0 mL resin bed of ThermoFisher HisPur NiNTA Resin (Grand Island, NY) equilibrated in wash buffer (50 mM Tris pH 7.4, 500 mM NaCl, 2% glycerol, 20 mM β-ME, 30 mM imidazole). The column was washed several times with wash buffer and protein was eluted with elution buffer (50 mM Tris pH 7.4, 500 mM NaCl, 20 mM β-ME, 250 mM imidazole). Elutions were collected in 1.5 mL fractions and run on a 12% SDS-PAGE gel to determine fractions that contain protein. Elutions containing protein were put into ThermoFisher SnakeSkin Dialysis Tubing (Grand Island, NY) with a 3,000 MW cutoff and dialyzed in

dialysis buffer (10 mM Tris pH 7.4, 150 mM NaCl, 5.0 mM β-ME) at 4ºC for 16 hours.

TuGST was concentrated using an EMD Millipore Amicon Ultra-15 concentrator (Billerica, MA) with a 3,000 MW cutoff, and concentration was determined by A_{280} using the molar extinction coefficient 50,880 M^{-1} cm^{-1} and MW 27,163 as determined by the ExPASy ProtParam tool [37]. Concentrated protein was put on a Superdex 200 SEC column attached to an ÄKTA Pure FPLC system (GE Healthcare, Marlborough, MA) equilibrated in FPLC buffer (10 mM Tris pH 7.4, 150 mM NaCl). Peaks corresponding to TuGST were pooled and the protein concentration was determined using the Bradford method [68]. To remove the N-terminal his-tag from TuGST, it was subjected to TEV protease cleavage. Briefly, pure TuGST (1-2 mg/mL) was incubated with TEV protease in a 1:100 (w/w) protease/protein ratio and dialyzed for 12 hours in dialysis buffer at 4˚C. After cleavage, TuGST was loaded onto an NiNTA column equilibrated in FPLC buffer and TuGST was collected in the flow through and concentrated. Due to poor yields and instability of TuGST, fusion partners for protein expression and a tagless construct were pursued.

2.3 Cloning TuGST into pMCSG28, pMCSG29 and pMBPcs1

pMCSG28 and pMCSG29 Cloning

Purification of TuGST with N-terminal his-tag resulted in very poor yields of protein (around 3.0 mg per liter culture). Furthermore, TEV his-tag cleavage resulted in most of the protein denaturing and the final yield of TuGST was less than one milligram (data not shown). To increase TuGST yield, expression options with fusion partner maltose-binding protein (MBP) were explored. Two plasmids, pMCSG29 (C-terminal MBP, C-terminal 6xHis-tag) and pMBPcs1 (N-terminal MBP, C-terminal 6xHis-tag) were

tested for their abilities to increase TuGST expression. An additional plasmid, pMCSG28, was also used to generate a tagless TuGST construct that contains no terminal His-tag or fusion partner.

The pMCSG28 and pMCSG29 plasmids were purchased from DNASU Plasmid Repository (Tempe, AZ). These vectors are designed to express protein with a C-terminal TEV-cleavable 6xHis-tag or N-terminal MBP, TVMV cut site, LIC site (gene insert), TEV cut site and 6xHis-tag, respectively; however, the reverse insertion primer for pMCSG28 (pMCSG28-GST-R) was designed to include a stop codon on the insert so this vector could be used for tagless TuGST expression. The vectors were amplified at the LIC site to generate blunt ends for LIC using the primers pMCSG28/29-AMP-F and pMCSG28/29-R listed in Table 2.1; both pMCSG28 and pMCSG29 contain identical LIC sites so only one set of primers was needed to amplify both vectors. KOD polymerase (EMD Millipore, Billerica, MA) was used following the manufacturer's protocol with the addition of 1.0 M betaine monohydrate. Thermalcycling was as follows: initial denaturation 95˚C for 2 minutes, 30 cycles of 95˚C for 20 seconds, 61˚C for 10 seconds and 70˚C for 2 minutes 30 seconds, followed by incubation at 4˚C. After PCR, the pMCSG28/29 products were purified by gel excision.

TuGST in pJExpress411 was used as base template for all reactions that involve amplifying the TuGST gene. The main sequence for TuGST was amplified with the primers pMCSG28/29-GST-F and pMCSG28-GST-R for pMCSG28 insertion, and primers pMCSG28/29-GST-F and pMCSG29-GST-R for pMCSG29 insertion listed in Table 2.1. Primers were designed to make the insert compatible with ligation independent cloning (LIC) into the target vectors. For pMCSG28 insert amplification, Phusion

polymerase (NEB) was used following the manufacturer's protocol. Thermalcycling was as follows: initial 98ºC for 30 seconds, 35 cycles of 98ºC for 10 seconds, 61ºC for 15 seconds, 72ºC for 10 seconds, a final extension at 72ºC for 30 seconds and an infinite hold at 12ºC. For pMCSG29 insert amplification, KOD polymerase (EMD Millipore, Billerica, MA) was used following the manufacturer's protocol with the addition of 1.0 M betaine monohydrate. Thermalcycling was as follows: initial denaturation of 95˚C for 2 minutes, 20 cycles of 95˚C for 20 seconds and 71˚C for 20 seconds, followed by incubation at 4˚C. All PCR products were purified by gel excision.

Sticky ends for LIC were made by incubating gel excised PCR products with T4 DNA polymerase (NEB, Ipswich, MA). A 40 µL reaction containing about 300 fmoles of vector (pMCSG28/29), 5.0 mM DTT, 2.5 mM dATP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase, and a 40 µL reaction containing about 1000 fmoles of insert, 5.0 mM DTT, 2.5 mM dTTP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase were incubated at room temperature for 30 minutes. The T4 DNA polymerase was then inactivated at 75ºC for 20 minutes and the reactions were mixed 10:10 at room temperature for 5 minutes. Next, 1.0 µL of 25 mM EDTA was added and the reaction was incubated another 5 minutes at room temperature. Transformation was performed with heat shock after adding the 20 µL LIC reaction to chemically competent DH5-α *E. coli* cells and plated on LB-AMP (50 µg/mL) plates at 37ºC for 16 hours. Clones were miniprepped with a GeneJET Plasmid Miniprep Kit (ThermoFisher, Grand Island, NY) and insertion of TuGST was first confirmed by restriction digest with XhoI following the manufacturer's instructions (NEB, Ipswich, MA). After restriction digest, clones with correctly sized fragments were sent to EtonBioscience (Research Triangle Park, NC) for sequencing with

T7 forward and T7 reverse primers. Correct clones were transformed via heat shock into chemically competent BL-21 (DE3) *E. coli* cells for protein expression.

pMBPcs1 Cloning

The construction of pMBPcs1 is described in Chapter 1.3 of this manuscript. TuGST was prepared for LIC into pMBPcs1 by using primers pMBPcs1-GST-F and pMBPcs1-GST-R listed in Table 2.1. For PCR, TuGST in pJExpress411 was used as template and the protocol for Phusion polymerase was followed (NEB, Ipswich, MA). The conditions for thermalcycling were as follows: initial 98ºC for 30 seconds, 35 cycles of 98ºC for 10 seconds, 70ºC for 30 seconds, 72ºC for 25 seconds, a final extension at 72ºC for 1 minute and an infinite hold at 12ºC. PCR product was run on a 1.0% agarose gel and gel excised using a ThermoFisher GeneJET Gel Extraction Kit (Grand Island, NY) and eluted with sterile water.

Sticky ends for LIC were made by incubating gel excised PCR products with T4 DNA polymerase (NEB, Ipswich, MA). A 40 µL reaction containing about 400 fmoles of pMBPcs1, 5.0 mM DTT, 2.5 mM dATP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase, and a 40 µL reaction containing about 1500 fmoles of TuGST, 5.0 mM DTT, 2.5 mM dTTP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase were incubated at room temperature for 30 minutes. The T4 DNA polymerase was then inactivated at 75ºC for 20 minutes and the reactions were mixed 10:10 at room temperature for 5 minutes. Next, 1.0 µL of 25 mM EDTA was added and the reaction was incubated another 5 minutes at room temperature. Transformation was performed with heat shock after adding the 20 µL LIC reaction to chemically competent DH5-α *E. coli* cells and plated on LB-AMP (50 µg/mL) plates at 37ºC for 16 hours. Clones were miniprepped with a ThermoFisher

GeneJET Plasmid Miniprep Kit (Grand Island, NY) and insertion of TuGST was first confirmed by restriction digest with HindIII and XbaI following the manufacturer's instructions (NEB, Ipswich, MA). After restriction digest, clones with correctly sized fragments were sent to EtonBioscience (Research Triangle Park, NC) for sequencing with T7 forward and T7 reverse primers. Correct clones were transformed via heat shock into chemically competent BL-21 (DE3) *E. coli* cells for protein expression.

2.4 TuGST Expression with pMCSG28, pMCSG29 and pMBPcs1

Both fusion partner constructs (pMCSG29 and pMBPcs1) produced appreciable amounts of protein (around 20 mg per liter culture) after NiNTA and FPLC purifications, but using TEV protease to cleave 6xHis-tags would also result in most of the protein denaturing and yield less than one milligram of TuGST, similar to pJExpress411 expression and purification (data not shown). However, purifying tagless TuGST (pMCSG28) resulted in considerable yields of active protein (about 15 mg per liter culture).

Purification and expression was the same as pJExpress411 up to the sonication. After sonication, lysate was dialyzed in 4.0 L of 25 mM Tris pH 7.5 at 4˚C for 12 hours. Next, the lysate loaded onto a HiPrep DEAE FF anion exchange column attached to an ÄKTA Pure FPLC system (GE Healthcare, Marlborough, MA) equilibrated in 25 mM Tris, pH 7.5. A stepwise elution gradient with sodium chloride was used in increments of 100 mM up to 500 mM (data not shown). TuGST eluted off the column at 100 mM NaCl, was concentrated, then loaded onto a Superdex 200 SEC column attached to an ÄKTA Pure FPLC system (GE healthcare, Marlborough, MA) (Figure 2.1) equilibrated in 50 mM HEPES pH 8.0 and 150 mM NaCl. Final yield of protein was around 14 mg per liter culture. The FPLC results showed the correct size for monomeric TuGST, but the predicted

structure and kinetic results herein show that the protein is most likely homodimeric, as is most common for GSTs [3, 69–71]. Furthermore, purified TuGST was run both on SDS-PAGE and Native-PAGE; SDS-PAGE showed one band at 24 kDa as expected, but the Native-PAGE showed the possibility of several oligomeric states (data not shown).

2.5 TuGST Kinetics

All reagents mentioned here were purchased from Sigma-Aldrich (St. Louis, MO), Santa Cruz Biotechnology (Dallas, TX) or FisherScientific (Grand Island, NY). Data was fit using OriginPro software (Northampton, MA). Several compounds that are known targets of GSTs were tested which include GSH, 1-chloro-2,4,-dinitrobenzene (CDNB), 1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane (4,4'-DDT), chlorfenapyr (CP), permethrin (PM), hydrogen peroxide, 1,2-epoxy-3-(4-nitrophenoxy)-propane (EPNP), 4-nitrophenethyl bromide (4NPB), 4-nitrobenzyl chloride (4NBC), 4-hydroxynonenal (4HNE), ethacrynic acid (ECA) and 4-nitrophenyl acetate (4NPA)(Figure 2.2) by UV-Vis spectroscopy in a ThermoFisher NanoDrop 2000 and processed with NanoDrop 2000 software (Grand Island, NY). The only compounds that seemed to exhibit any activity were GSH and CDNB as measured at 340 nm with a molar extinction coefficient of 9600 M^{-1} cm^{-1} [64]. Reactions consisted of 10 mM GSH or 1.5 mM CDNB in 1.0 mL of 100 mM HEPES pH 8.0 with a working TuGST concentration of about 47 nM; TuGST showed the highest thermal stability in pH 8.0 buffer (data not shown). Reactions were performed in triplicate, and background GSH reactivity with target compounds was subtracted from reactions containing enzyme to determine the concentration of enzyme-conjugated GSH products. Interestingly, TuGST showed positive cooperativity towards GSH binding (Figure 2.3A) and negative cooperativity towards CDNB binding (Figure 2.3B). Kinetic

parameters are listed in Table 2.2. TuGST had a higher affinity for GSH than CDNB but a lower V_{max}. The specific activity for both substrates was very low (around 15 nmoles/min/mg of protein) relative to other GSTs [3, 64–66, 70, 72–75], and the k_{cat} and catalytic efficiency for CDNB were about 2-fold higher than those for GSH.

2.6 PREDICTED STRUCTURE OF TUGST

The structure for TuGST was predicted using the amino acid sequence for TuGST (Uniprot: T1JQ77) in the SWISS-MODEL online program [76–78]. The template model chosen was a silkworm delta-class GST from *Bombyx mori* (PDB code: 3VK9) with a 43.7% sequence identity (Figure 2.4). The Global Model Quality Estimation (GMQE) was 0.75 (range is from zero to one where one is more reliable predictions based on template used and sequence alignment). The residue range was 1 to 214 with 0.99 coverage. Measurements were made in COOT and structures made in PyMOL [79, 80].

The predicted structure of TuGST is shown in Figure 2.5A. Similar to other GSTs, it is shown as a homodimer with two active sites. To observe potential substrate binding, the predicted TuGST model was aligned with a human theta-class GST, hGSTT2-2 (PDB code: 3LJR) (Figure 2.5A). Despite relatively low sequence identity (about 25%), aligning the structures of predicted TuGST and hGSTT2-2 had an RMSD value of 1.2 Å (performed in PyMOL [80]). Based on amino acid sequence TuGST appears to be a delta-class GST. Theta-class hGSTT2-2 was chosen for comparison for two reasons: 1) hGSTT2-2 has the ligand 1-menaphthyl glutathione conjugate bound which can be used as a homologous model for substrate binding, 2) Cooperativity has been observed in hGSTT2-2, and residues that may be the basis for this cooperativity are found in the predicted TuGST structure [66].

In hGSTT2-2, the residues Tyr73, Leu89 and Ala93 correspond to Tyr73, Val89 and Ala93 in TuGST (Figure 2.5B) and may be responsible for cooperativity [66]. The 1-menaphthyl glutathione conjugate is modeled in the active sites of TuGST. Residues in the G-site of TuGST include Ser11 (interacts with reduced thiol of GSH), His52, Ile54 (main chain backbone salt bridge), Glu66, Ser67 (main chain backbone salt bridge) and Arg68 (Figure 2.5D) similar to those found in BmGST (3VK9 - Ser11, Gln51, His52, Glu66, Ser67 and Arg68 with a difference in only one residue (Gln51 in BmGST corresponds to Phe51 in TuGST) [81]; the Ile54 residue of BmGST could potentially form a salt bridge with 1-menaphthyl glutathione conjugate similar to that of TuGST. Furthermore, residues in the H-site potentially include Leu8, Tyr107, Phe115, Ile118 and Phe209 (Figure 2.5D) similar to Tyr107, Tyr115, Phe119, Phe206 and Ser212 found in BmGST [81]. The hydrophobic H-site residues in BmGST appear to be much bulkier than those found in TuGST which may contribute to their difference in specificities.

2.7 DISCUSSION

The delta-class TuGST investigated here only had reactivity towards GSH and CDNB. Initially, expression and purification of an N-terminal 6xHis-tagged TuGST in *E. coli* lead to poor yields of protein, especially after cleaving the tag with TEV protease. The fusion partner MBP was also used and resulted in poor yields of TuGST. Only using a construct with no tags or fusion partners resulted in appreciable yields of TuGST. All constructs were tested for GST activity, and all of them except non-tagged TuGST displayed extremely deficient GST activity (relative to the non-tagged construct; data not shown). This observation may be explained by observing the predicted structure of TuGST.

For most GSTs, the N-terminal region contains residues responsible for binding

GSH, and the C-terminal region contains residues responsible for binding hydrophobic substrates. Adding a 6xHis-tag or fusion partner to either of the terminal ends may disrupt the protein active sites and overall protein folding, which is a possible explanation as to why these constructs had very poor GST activity or did not express well. Only after expressing TuGST with no tags or fusion partners was activity observed by UV-Vis spectroscopy.

Positive and negative cooperativity was observed for GSH and CDNB, respectively. For most GSTs, a lock-and-key motif (example hGSTP1-1, PDB code: 6GSS) is observed upon substrate binding, but an alternative "clasp" motif has been described in the middle of the subunit interface for some GSTs [66]. It has been speculated that the clasp motif behaves like two hands interlocking that may facilitate cooperativity [66]. Observing the predicted TuGST structure, the residues to form a clasp motif are present, but site-directed mutagenesis experiments would need to be performed to confirm this type of motif and mechanism. Furthermore, negative cooperativity was observed in several GSTs upon binding the toxic nitric oxide adduct dinitrosyl-diglutathionyl iron complex (DNDGIC) [66]. The authors describe DNDGIC binding as a perturbation of normal enzymatic activity, and that cooperativity is displayed as a means of possible cooperative self-preservation [66].

The negative cooperativity may help to maintain residual conjugating activity of GST against other endogenous toxins so complete GST activity is not lost [66]. A similar situation may be true for the TuGST studied here, although it did not show reactivity with almost any of the compounds tested. In this case, it is possible this TuGST may serve only sequester xenobiotics, which would effectively detoxify them, but more sensitive binding

studies and instrumentation such as Surface Plasmon Resonance (SPR) or Isothermal Titration Calorimetry (ITC) would provide more information on substrate binding than UV-Vis spectroscopy.

2.8 TABLES

Table 2.1: Primers used for TuGST cloning. Nucleotides in bold are complementary to TuGST sequence. Lowercase letters are nucleotides insertions.

Primer Name	Primer Sequence
pMCSG28/29-GST-F	5' GTCTCTCCCATG**GTCCTGGAGTTGTACCAACTGC** 3'
pMCSG28-GST-R	5' GGTTCTCCCCAGCGGGGCTCGAG**TTACGCCAGA** 3'
pMCSG29-GST-R	5' GGTTCTCCCCAG**CCGCCAGACGTGCGCGCA** 3'
pMCSG28/29-F	5' GGGGAGAACCTGTACTTCCAATCCGC 3'
pMCSG28/29-R	5' GGGAGAGACTCCTTCTTAAAGTTAAACAAA 3'
pMBPcs1-GST-F	5' TTCGGCTGCTATG**GTCCTGGAGTTGTACCAACTGCCGATGAGCG** 3'
pMBPcs1-GST-R	5' GGTTCTCCCCAGC**CGCCAGACGTGCGCGCAGGAA** 3'
pMBPcs1-F	5' CCTGAAAGACGCGCAGACTAATTCGgctgctGGGGA 3'
pMBPcs1-R	5' CGGATTGGAAGTACAGGTTCTCCCCagcagcCGAAT 3'
pMBPcs1-LIC-F	5' GGGGAGAACCTGTACTTCCAATCCGCCGGG 3'
pMBPcs1-LIC-R	5' agcagcCGAATTAGTCTGCGCGTCTTTCAG 3'

Table 2.2: Kinetic parameters for GSH and CDNB. All reactions were performed in triplicate. The measurement for specific activity was performed taking the average of three substrate concentrations performed in triplicate. Abbreviations: GSH – Reduced glutathione, CDNB – 2,4-dinitrochlorobenzene.

Parameter	GSH	CDNB
Wavelength (nm)	340	340
K_m (mM)	1.48 ± 0.05	1.74 ± 1.22
Vmax (μM/sec)	1.28 ± 0.10	2.82 ± 0.62
Hill Coefficient (n)	1.40 ± 0.10	0.58 ± 0.05
Specific Activity (nmoles/min/mg)	13.8 ± 6.6	15.5 ± 3.8
k_{cat} (sec^{-1})	27.2	60.0
Catalytic efficiency (M^{-1} sec^{-1})	18.4 x 10^3	34.5 x 10^3

2.9 Figures

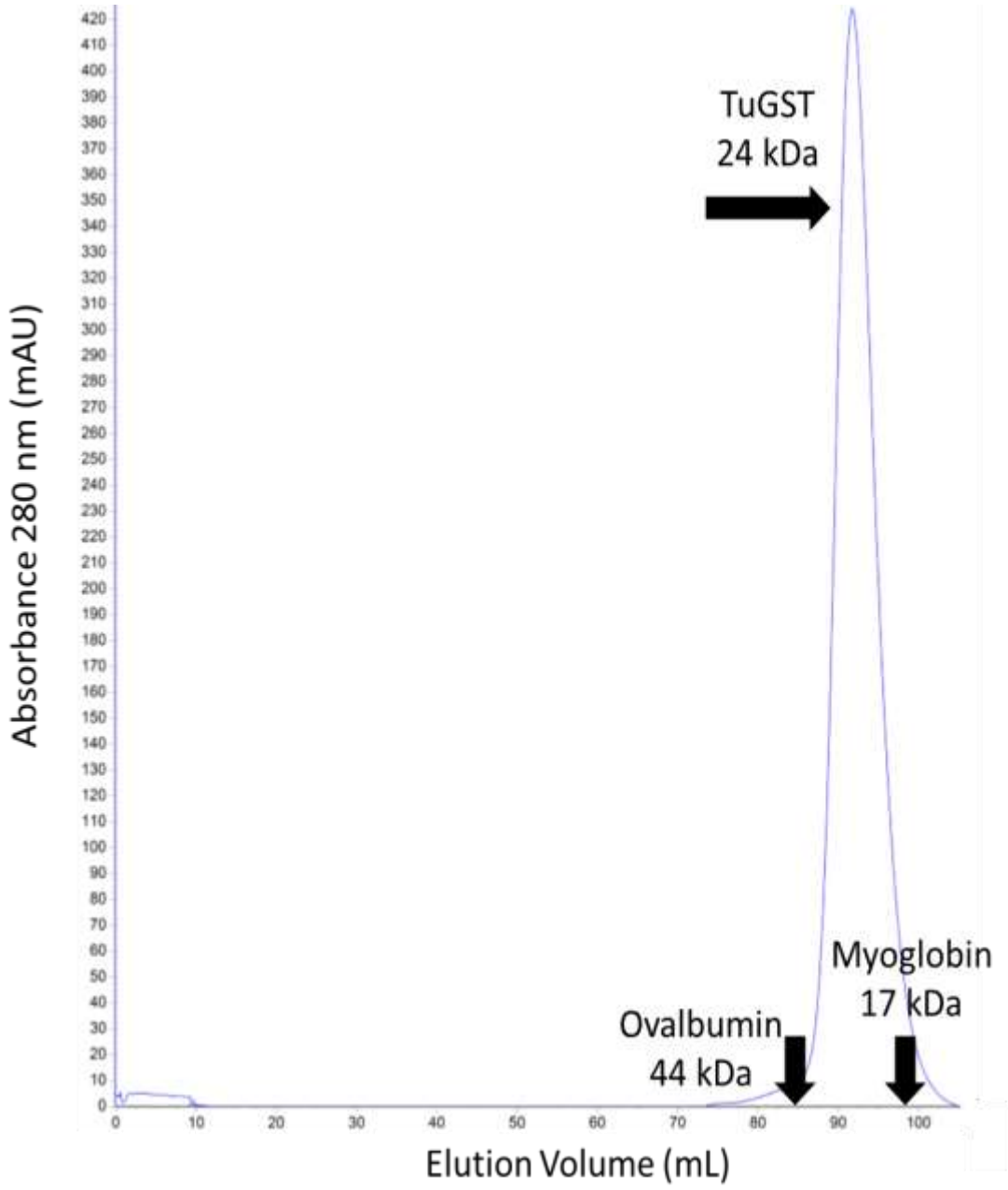

Figure 2.1: Size exclusion results for purification of TuGST. Bio-Rad molecular weight standards (Hercules, CA) are shown (Ovalbumin and Myoglobin).

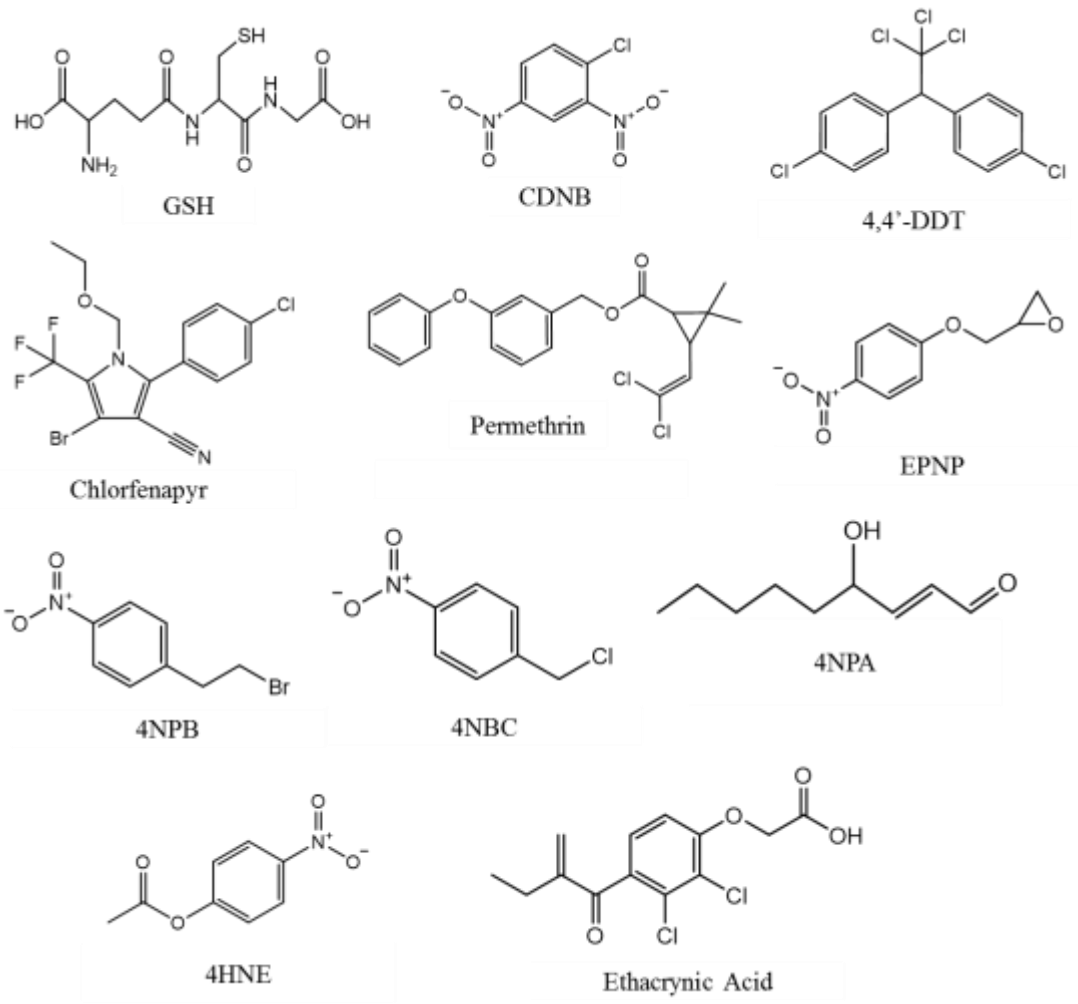

Figure 2.2: Compounds tested with TuGST. GSH – reduced glutathione; CDNB - 1-chloro-2,4-dinitrobenzene; 4,4'-DDT - 1,1,1-trichloro-2,2-bis(4-chlorophenyl) ethane; EPNP - 1,2-epoxy-3-(4-nitrophenoxy)-propane; 4NPB - 4-nitrophenethyl bromide; 4NBC - 4-nitrobenzyl chloride; 4NPA - 4-nitrophenyl acetate; 4HNE - 4-hydroxynonenal. Not shown: hydrogen peroxide.

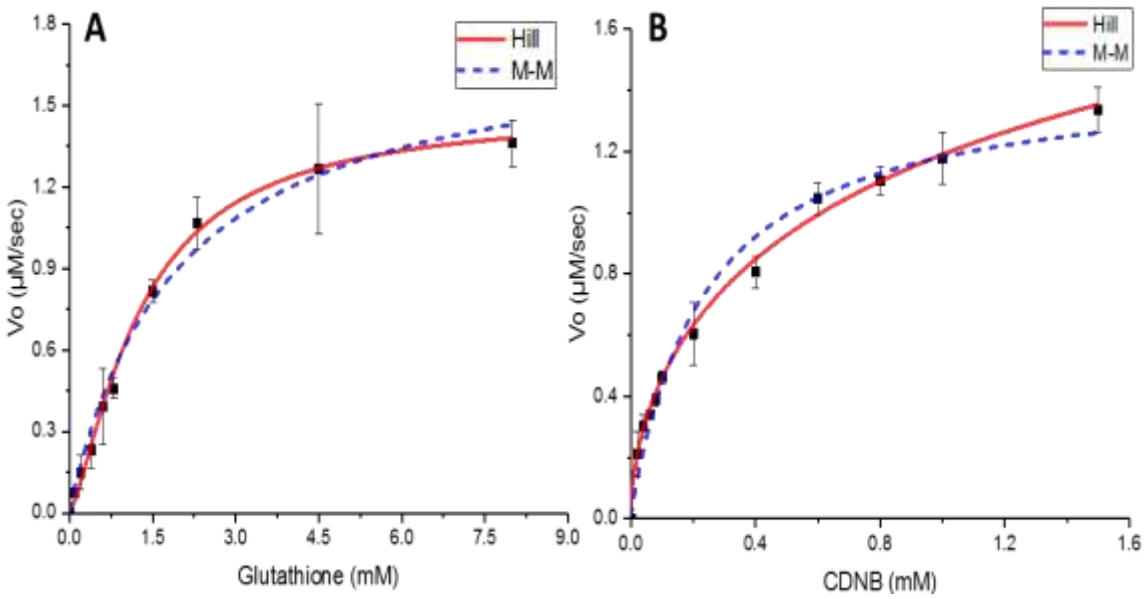

Figure 2.3: Kinetic data for GSH and CDNB. Both data sets show Hill (red line) and Michaelis-Menten (M-M, blue dash) fits. **A)** UV-Vis spectroscopy with varied GSH displayed positive cooperativity. **B)** UV-Vis spectroscopy with varied CDNB displayed negative cooperativity. Both were fit using the Hill and Michaelis-Menten equations in OriginPro software (Northampton, MA).

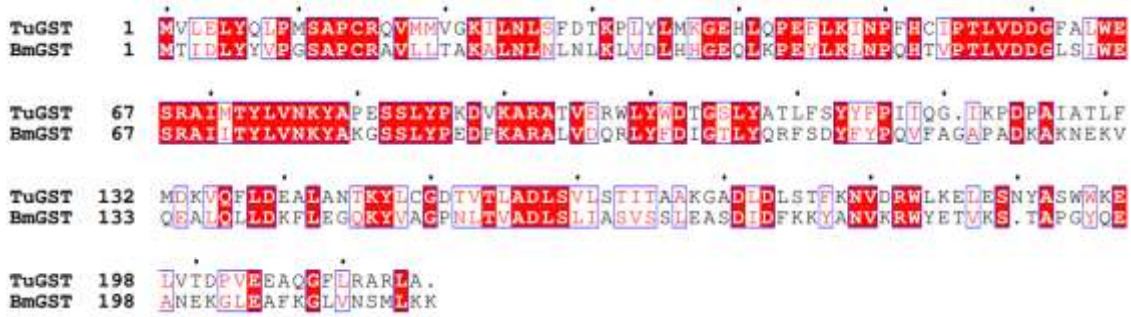

Figure 2.4: Sequence alignment of TuGST and BmGST. TuGST and BmGST share a sequence identity of about 44%. Residues highlighted in red are identical. Residues in red font are similar. Residues surrounded by a blue box represents similarity across groups of residues. Sequence alignment was performed with ESPript 3.0 [82].

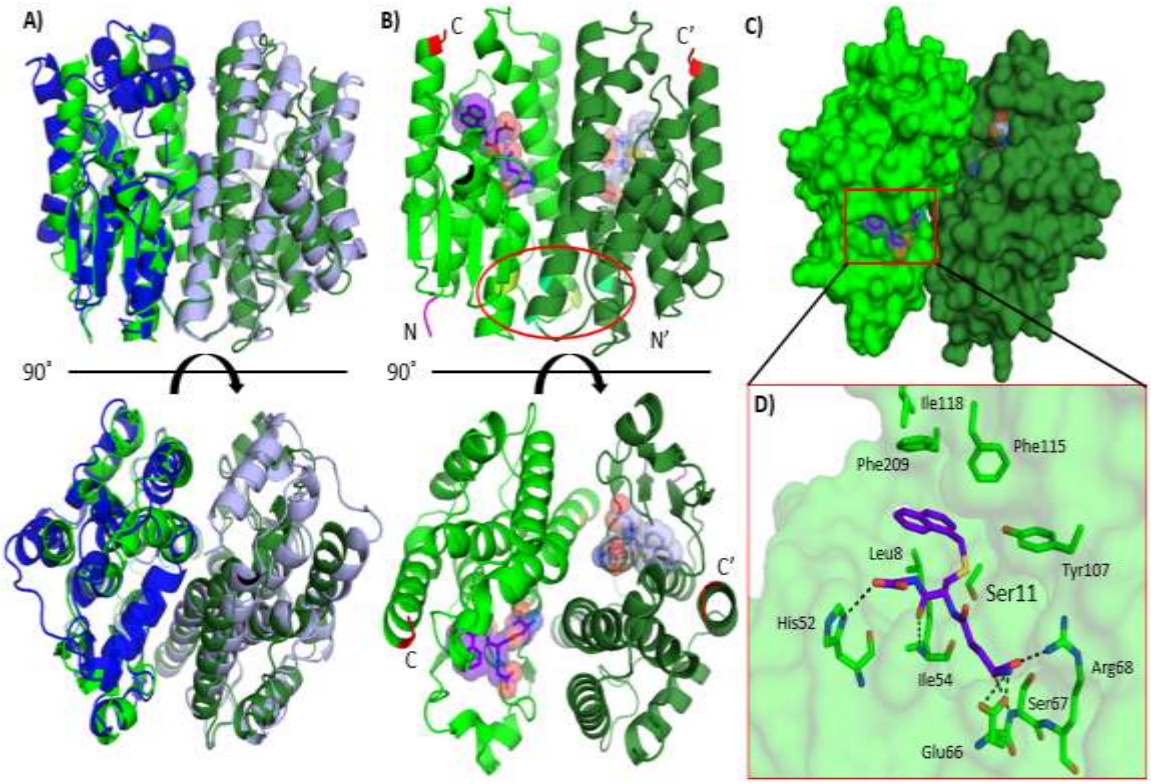

Figure 2.5: Predicted structure of TuGST. **A)** Structure alignment of predicted TuGST with hGSTT2-2 (PDB code: 3LJR). HGSTT2-2 is a homodimer and TuGST is predicted to be a homodimer. TuGST monomers are colored in shades of green and hGSTT2-2 are colored in shades of blue. The RMSD value from alignment was about 1.2 Å. **B)** Homologous model of TuGST (monomer 1 – green, monomer 2 – forest) with substrate analogue 1-menaphthyl glutathione conjugate (from 3LJR) modeled in the active sites. The N-terminal and C-terminal ends are colored in magenta and red, respectively. The red circle shows residues potentially involved in cooperative movement: Tyr73, Val89, Ala93 (colored yellow for monomer 1 and cyan for monomer 2). **C)** Surface representation of TuGST that shows 1-menaphthyl glutathione conjugate buried in the active site. **D)** Different residues in the H and G sites of TuGST. Based on this predicted structure and the positioning of Ser11, TuGST is most likely a Ser-GST.

CHAPTER 3:

STRUCTURAL CHARACTERIZATION OF AN INTRADIOL RING-CLEAVAGE DIOXYGENASE FROM *TETRANYCHUS URTICAE*

3.1 BACKGROUND AND INTRODUCTION

Aromatic compounds are present throughout the environment, whether they are natural (such as the plant biopolymer lignin) or man-made contaminations (such as phthalates in papers, insecticides, cosmetics) [83, 84]. They are stabilized by their high resonance energy, thus making them very stable, and consequently, permit their persistence in the environment and sometimes contain electron-withdrawing groups that can further prevent their biodegradation [83, 84]. Their resistance to degradation is problematic as they can be stored in animal and plant tissues and accumulate to toxic levels [83].

To combat some of these toxic aromatic compounds, many organisms (both prokaryotic and eukaryotic) are equipped with metalloproteins called dioxygenases that are capable of catabolizing aromatic compounds into non-toxic metabolites by utilizing a non-heme iron in the active site [85, 86]. The genome of *T. urticae* includes a set of 17 genes encoding for secreted proteins that belong to the "intradiol dioxygenase-like" subgroup that may promote the *T. urticae's* polyphagous lifestyle in detoxifying allelochemicals produced by plants [86]. Catechol, or 1,2-dihydroxy benzene, is a common metabolite found in plants which can be involved with many metabolic pathways [87] and act as substrate for intradiol ring-cleavage dioxygenases (ID-RCDs). Typically, ID-RCDs detoxify catechols (and catechol derivatives, i.e. substituted with halides, other functional

groups) by cleaving the aromatic ring between C1 and C2, resulting in the formation of cis,cis-muconic acid (Figure 3.1) that can be further metabolized into succinate and acetyl-CoA [84, 85]. In addition to allelochemical resistance, ID-RCDs may play a role in insecticide resistance as several insecticides are catechol-based such as carbofuran and propoxur (Figure 3.1) [88]. Tetur07g02040, a gene coding for a putative ID-RCD, is the only ID-RCD in the spidermite genome that contains introns [86] and is the focus of this research. Studying this protein has the potential to not only provide more information on the structural properties of ID-RCDs and their substrate specificities, but also could serve as a target for the development of new insecticides as the importance of this TuDioxygenase is not known.

3.2 Protein expression and purification of TuDioxygenase

The gene for tetur07g02040 was ordered from DNA2.0 (Newark, California), codon optimized for *E. coli* and inserted into pJExpress411 with a T7 promoter, terminator and KAN resistance. Initially, the TuDioxygenase gene was ordered with the first 47 amino acids truncated for two reasons: 1) residues 1-22 were predicted to be a signal peptide (Uniprot reference number T1K8P1 [61], 2) residues 28-49 were predicted to be disordered [89]. Furthermore**,** the gene was synthesized to contain an N-terminal cleavable 6xHis-tag for ease of purification with the Tobacco Etch Virus (TEV) cut site MHHHHHHSSGVDLGTENLYFQ/SGSG where the cut site is shown with a slash**.**

Plasmid was transformed into BL-21 (DE3) cells by heat shock and grown in 1.0 L cultures of Luria-Broth (LB) with 50 µg/mL KAN or 100 µg/mL AMP shaking at 37˚C until an O.D. of 0.8 was reached. Cultures were cooled down to 16˚C and protein

expression was induced with 0.4 mM IPTG for 16 hours while shaking. Cells were pelleted and frozen at -80˚C until needed further.

For purification of TuDioxygenase, cell pellets were resuspended in lysis buffer (50 mM Tris, 500 mM NaCl, 10 mM imidazole, 20 mM β-ME, 2% glycerol). After resuspension, cells were lysed by sonication with a Branson 45 Sonifier (ThermoFisher, Grand Island, NY). Crude extract was separated by spinning the lysate in a Beckman Coulter centrifuge with a JA-17 rotor at 16000 RPMs (Indianapolis, IN) for 25 minutes at 4˚C. The clear, yellowish supernatant was poured into a 12 x 1.5 cm Bio-Rad column (Hercules, CA) filled with 5.0 mL of NiNTA resin (ThermoFisher, Grand Island, NY) previously equilibrated in wash buffer (50 mM Tris pH 7.4, 150 mM NaCl, 30 mM Imidazole, 2% glycerol, 20 mM β-ME) and then washed with wash buffer. Protein was eluted using elution buffer (50 mM Tris pH 7.4, 150 mM NaCl, 250 mM imidazole, 2% glycerol, 20 mM β-ME) and immediately put into dialysis buffer (10 mM Tris pH 7.4, 150 mM NaCl) in ThermoFisher SnakeSkin Dialysis Tubing (Grand Island, NY) with a 10,000 MW cutoff. Protein was dialyzed in dialysis buffer for 12 hours at 4˚C.

After dialysis, protein was concentrated with Amicon Ultra concentrators (EMD Millipore, Billerica, MA) with a 10,000 MW cutoff. Concentrated protein was put on a Superdex 200 column attached to an ÄKTA Pure FPLC system (GE Healthcare, Marlborough, MA) equilibrated in FPLC buffer (10 mM Tris pH 7.4, 150 mM NaCl). Peaks corresponding to TuDioxygenase were pooled and the protein concentration was determined using A_{280} with the MW 27068.39 Da and molar extinction coefficient 25,900 $M^{-1}cm^{-1}$. Attempts at removing the his-tag resulted in heavy protein denaturation which is probably attributed to the instability of dioxygenase that was observed during purification.

Furthermore, purified TuDioxygenase was red in color, indicating the presence of non-heme ferric iron, but the protein was very unstable and frequent red precipitate was observed in purified protein samples. Due to instability, the fusion partner MBP was pursued.

3.3 Molecular cloning of TuDioxygenase into pMCSG29 and pMBPcs1

The dioxygenase gene was cloned into pMCSG29 using ligation independent cloning (LIC). This plasmid is designed to generate a fusion protein as follows: protein of interest, TEV cut site, 6xHis-tag, TVMV cut site, MBP). Primers used for pMCSG29 cloning are listed in Table 3.1. The pJExpress411 vector containing dioxygenase was used as template for the first PCR with the primers p29-Dioxy-F and p29-Dioxy-R. PCR was performed using phusion polymerase following the manufacturer's protocol (NEB, Ipswich, MA) and thermalcycling as follows: initial denaturation 98˚C 30 seconds, 30 cycles of 98˚C for 10 seconds, 68˚C for 25 seconds and 72˚C for 25 seconds with a final extension at 72˚C for 5.0 minutes followed by an infinite hold at 12˚C. PCR product was purified using a GeneJet Gel Extraction Kit (ThermoFisher, Grand Island, NY). The pMCSG29 was amplified at the LIC site to generate blunt-ended vector with the primers pMCSG28/29-F and pMCSG28/29-R as mentioned in Chapter 2 with TuGST cloning. KOD polymerase (Millipore, Billerica, MA) was used following the manufacturer's protocol with the addition of 1.0 M betaine monohydrate to amplify pMCSG29 by PCR. Thermalcycling was as follows: initial denaturation 95˚C for 2 minutes, 30 cycles of 95˚C for 20 seconds, 61˚C for 10 seconds and 70˚C for 2 minutes 30 seconds followed by an infinite hold at 12˚C. After PCR, product was purified by gel excision and cleaned up using a GeneJet Gel Extraction Kit (ThermoFisher, Grand Island, NY).

Sticky ends for LIC were made by incubating gel excised PCR products with T4 DNA polymerase (NEB, Ipswich, MA). A 40 µL reaction containing 170 fmoles of pMCSG29, 5.0 mM DTT, 2.5 mM dATP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase, and a 40 µL reaction containing 670 fmoles of dioxygenase insert, 5.0 mM DTT, 2.5 mM dTTP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase were incubated at room temperature for 30 minutes. The T4 DNA polymerase was then inactivated at 75ºC for 20 minutes and the reactions were mixed 10:10 at room temperature for 5.0 minutes. Next, 1.0 µL of 25 mM EDTA was added and the reaction was incubated another 5.0 minutes at room temperature. Transformation was performed with heat shock as mentioned previously after adding the 20 µL LIC reaction to the cells, and cells were plated on LB-AMP (50 µg/mL) plates at 37ºC for 16 hours. Clones were miniprepped with a GeneJET Plasmid Miniprep Kit (ThermoFisher, Grand Island, NY) and insertion of dioxygenase was first confirmed by restriction digest with XhoI following the manufacturer's instructions (NEB, Ipswich, MA). After restriction digest, clones with correctly sized fragments were sent to EtonBioscience (Research Triangle Park, NC) for sequencing with T7 forward and T7 reverse primers. Correct clones were transformed via heat shock into BL-21 (DE3) *E. coli* cells.

Cloning TuDioxygenase into pMBPcs1.

The pMBPcs1 plasmid was generated and amplified as previously mentioned in Chapter 1 (Page 5). Dioxygenase insert was amplified using the primers pMBPcs1-Dioxy-F and pMBPcs1-Dioxy-R shown in Table 3.1 and PCR performed using phusion polymerase following the manufacturer's protocol (NEB, Ipswich, MA) with thermalcycling as follows: initial denaturation at 98˚C for 30 seconds, 35 cycles of 98˚C

for 10 seconds, 70˚C for 30 seconds and 72˚C for 25 seconds and a final extension of 72˚C for 5.0 minutes followed by an infinite hold at 12˚C. PCR product was purified using a GeneJET Gel Extraction Kit (ThermoFisher, Grand Island, NY). Sticky ends for LIC were made by incubating gel excised PCR products with T4 DNA polymerase (NEB, Ipswich, MA). A 40 µL reaction containing 600 fmoles of pMBPcs1, 5.0 mM DTT, 2.5 mM dATP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase, and a 40 µL reaction containing 1000 fmoles of dioxygenase insert, 5.0 mM DTT, 2.5 mM dTTP, 1X NEB2.1 buffer and 3.0 units of T4 DNA polymerase were incubated at room temperature for 30 minutes. The T4 DNA polymerase was then inactivated at 75ºC for 20 minutes and the reactions were mixed 10:10 at room temperature for 5.0 minutes. Next, 1.0 µL of 25 mM EDTA was added and the reaction was incubated another 5.0 minutes at room temperature.

Transformation was performed with heat shock as mentioned previously after adding the 20 µL LIC reaction to the cells, and cells were plated on LB-AMP (50 µg/mL) plates at 37ºC for 16 hours. Clones were miniprepped with a GeneJET Plasmid Miniprep Kit (ThermoFisher, Grand Island, NY) and insertion of dioxygenase was first confirmed by restriction digest with HindIII and XbaI following the manufacturer's instructions (NEB, Ipswich, MA). After restriction digest, clones with correctly sized fragments were sent to EtonBioscience (Research Triangle Park, NC) for sequencing with T7 forward and T7 reverse primers. Correct clones were transformed via heat shock into BL-21 (DE3) *E. coli* cells.

3.4 Expression and Purification of Dioxygenase-MBP and MBP-Dioxygenase

Dioxygenase-MBP and MBP-dioxygenase were transformed, expressed and purified in the same manner as TuDioxygenase but with 5.0 mM maltose in all purification

buffers. Tris and imidazole (50 mM) were used in dialysis and FPLC buffers. Dioxygenase-MBP yielded decent amounts of protein (~10 mg/L culture); however, the protein did not have the red hue observed from TuDioxygenase purification that would indicate the presence of the non-heme iron (and most likely properly folded protein) so this construct was not pursued further. MBP-dioxygenase yielded decent quantities of protein (~12 mg/L culture) and had the desired red hue after purification. Protein yield was slightly increased by adding 3.0 mg of ferrous sulfate heptahydrate to the 1.0 L LB culture pre-induction and to the lysis buffer. Protein concentration of MBP-dioxygenase was determined using A_{280} with the MW 67,027.82 Daltons and molar extinction coefficient 92,375 $M^{-1}cm^{-1}$ as determined by using ExPASy ProtParam tool [37]. Attempts were not made to remove the his-tag from MBP-dioxygenase due to instability observed from purification of TuDioxygenase

3.5 Crystallization of MBP-Dioxygenase

All chemicals were purchased from Hampton Research (Aliso Viejo, CA), ThermoFisher (Grand Island, NY) or Sigma-Aldrich (St. Louis, MO). Crystallization experiments were performed at room temperature using the sitting-drop vapor diffusion method and MRC 2-drop 96-well crystallization plates (Hampton Research, Aliso Viejo, CA). Initially, MBP-dioxygenase crystals were grown in 0.1 M sodium cacodylate trihydrate pH 6.5, 0.2 M magnesium acetate tetrahydrate, 20% w/v PEG8000 (3:1 of protein:mother liquor where protein ~13 mg/mL). These crystals were microseeded due to their small size and red color. For microseeding, crystals were crushed and resuspended in a 1.5 mL tube filled with 150 µL of the cacodylate crystallization solution and a glass bead then vortexed. An equal volume of MBP-dioxygenase (~13 mg/mL) was added to the seed

stock and drops were set with 1:1 ratio of protein to mother liquor as written above. Using the seed stock, red crystals grew after about one month in 0.1 M Tris pH 7.5, 15% w/v PEG6000.

3.6 Data Collection, Structure Determination and Refinement of TuDioxygenase.

Table 3.2 shows the data collections statistics for the TuDioxygenase crystal structure mentioned here. The crystal structure for TuDioxygenase was deposited in the Protein Data Bank (PDB) with the accession code 5VG2. Crystals were cryo-cooled in liquid nitrogen and data was collected using Southeast Regional Collaborative Access Team (SER-CAT) 22ID beamline at the Advanced Photon Source (APS), Argonne National Lab (Argonne, IL). Data were processed with the HKL-2000 software package [41]. Molecular replacement was performed using MOLREP [42] integrated with HKL-3000 [43] using the PDB model 4ILT for molecular replacement. BUCCANEER [44] and HKL-3000 were used to rebuild the initial model. Refinement was performed using REFMAC [46] and HKL-3000 [43]. Non-crystallographic symmetry was used during the whole process of refinement. TLS refinement was used during the last stages of refinement and the TLS Motion Determination server was used for partitioning protein chains into the rigid bodies undergoing vibrational motions [90, 47]. Model was updated and validated with COOT [48]. MOLPROBITY was used in the final steps of the model validation [49]. The final model together with structure factors were deposited to the Protein Data Bank.

The programs ProFunc [91] and PDBePISA [53] were used to analyze oligomeric assembly and secondary structure, bond measurements were performed in COOT [48] and structures were made using PyMOL [54] and UCSF-Chimera [92].

3.7 CRYSTAL STRUCTURE OF MBP-DIOXYGENASE

Similar to the crystallization of MBP-cyanase (chapter 1), MBP was not present in the crystal structure of MBP-dioxygenase. TuDioxygenase crystallized in monoclinic system with four protein chains present in the asymmetric. In crystal TuDioxygenase appears to be monomeric which is consistent with size exclusion results (data not shown). The interface area between neighboring monomers is around 600 $Å^2$ and 360 $Å^2$. Each chain consists of residues 56-259 (with respect to complete dioxygenase gene sequence) where the first 8 amino acids (residues 48-55) of the truncated recombinant dioxygenase studied here are not visible in the electron density, nor is the TEV cut site and 6xHis-tag on the C-terminus. β-ME was used in some purification buffers, but each chain starts at Cys56 which forms a disulfide bridge with Cys99 on the same chain. There are six cacodylate ions present in the crystal structure: four cacodylate ions are coordinated by Lys106 and His175 for each chain, one cacodylate ion coordinated by Asn148 in chain C and one cacodylate ion that mediates some minor crystal contact between chains with Arg208 of chain A and Arg126 of chain C.

TuDioxygenase contains 7 β-sheets that form a β-sandwich and 6 α-helices. The β-sandwich core is conserved with most intradiol-ring cleavage dioxygenases [93]. However, the conserved β-sandwich core usually consists of 8 β-sheets instead of 7, but the "missing" additional β-sheet is most likely in the truncated part of the dioxygenase (residues 23-47) as the N-terminal of truncated TuDioxygenase is located next to the β-sandwich (Figure 3.2). The overall fold of dioxygenase is most similar to that of SACTE_2871 (PDB code: 4ILT) from *Streptomyces* sp. SirexAA-E which was used as the starting model for molecular replacement [93]. There is only a 17% sequence identity but a 34% sequence

similarity between TuDioxygenase and SirexAA-E (Figure 3.3). Furthermore, TuDioxygenase has a similar ID-RCD domain to that of SACTE_2871 such that an entrance to the active site is a narrow cleft that is more solvent accessible than most other structurally similar ID-RCDs [93]. This cleft is mainly aligned with several ionizable amino acids. The intradiol dioxygenase domain consists of residues 64-189.

3.8 Coordination of Active Site Iron

Metal binding sites were validated by the CheckMyMetal (CMM) server [94, 95]. The non-heme Fe^{+3} is coordinated by two tyrosines and two histidines in each active site: Tyr118, Tyr163, His169, His171 (Figure 3.4A). The ferric center is coordinated in a trigonal bipyramidal manner with the four active site residues and a water molecule which is common to ID-RCDs [96]. The water molecules (or OH^- ions) are not shown as the corresponding electron density for them is of poor quality to allow for unambiguous modeling. Figure 3.4B demonstrates bidentate binding of catechol to the ferric center via the structural homologue PDB code: 3HHY, a 1,2-dioxygenase from the Gram-positive bacteria *Rhodococcus opacus* (shares 24% sequence identity with TuDioxygenase) [97]. Upon binding catechol, Tyr196 (Tyr163 in TuDioxygenase) is displaced from being equatorial to axial, and the trigonal bipyramidal binding is retained [85].

3.9 Putative Mechanism of TuDioxygenase

Based on the mechanism of ID-RCDs and homologues in the PDB [84, 85, 93, 98], catechol binding displaces Tyr163 that permits bidentate binding to the active site iron [99] (Figure 3.4B). Cleaving an oxygen-oxygen (O_2) bond generates an iron-bound oxide or hydroxide that acts as a nucleophile to hydrolyze the anhydride which is a result of Criegee rearrangement, ultimately generating the cis, cis-muconic acid product [99].

3.10 DISCUSSION

Structural homologs of TuDioxygenase identified by the Dali server [51] were dioxygenases such as: SACTE_2871 (PDB code 4ILT, 153/155 residues aligned, RMSD 1.8 Å); catechol 1,2-dioxygenase from *Burkholderia multivorans* (PDB code 5UMH, 167/307 residues aligned, RMSD 2.4 Å); 3-chlorocatechol 1,2-dioxygenase from *Rhodococcus Opacus 1Cp* (PDB code 2BOY, 157/250 residues aligned, RMSD 2.1 Å); catechol 1,2-dioxygenase from *Rhodococcus opacus 1Cp* (PDB code 3HHY, 158/256 residues aligned, RMSD 2.2 Å). No other structural homologues SACTE_2871 has a carbohydrate binding motif domain located on the C-terminal that appears to not be shared by the TuDioxygenase studied here [93]. However, similar to SACTE_2871, TuDioxygenase does not have an extensive N-terminal dimerization domain, an attribute common to ID-RCDs, as observed in the TuDioxygenase crystal structure [93].

Although the first 47 residues were truncated for this TuDioxygenase construct, the first 22 residues are predicted to be a signal peptide and most likely are not present in a mature form of the protein. The remaining 25 residues that are missing may facilitate oligomerization of TuDioxygenase, but further studies will need to be performed with the full-length protein. Most likely, the missing β-strand of the 7-strand β-sandwich is formed by these residues.

3.11 TABLES

Table 3.1: Primers used for TuDioxygenase cloning.

Primer	Sequence
p29-Dioxy-F	5' GTCTCTCCCATGTCGTTTGTTACCCGTTTCACCGAGT
p29-Dioxy-R	5' GGTTCTCCCCAGCGGCCCACATCAGATTGCT
pMCSG28/29-F	5' GGGGAGAACCTGTACTTCCAATCCGC
pMCSG28/29-R	5' GGGAGAGACTCCTTCTTAAAGTTAAACAAA
pMBPcs1-Dioxy-F	5' TTCGGCTGCTAGTTCGTTTGTTACCCGTTTCACCGAG
pMBPcs1-Dioxy-R	5' GGTTCTCCCCAGCGGCCCACATCAGATTGCTGC

Table 3.2: Data collection statistics for TuDioxygenase. Values in parentheses are for the highest resolution shell.

PDB Accession code	5VG2
Data collection	
Diffraction source	Synchrotron
Wavelength (Å)	1.000
a, *b*, *c* (Å)	60.48, 43.07, 165.72
α, β, γ (degrees)	90.0, 95.2, 90.0
Space group	P 1 2 1
Solvent content (%)	39.89
Protein chains in AU	4
Resolution range (Å)	40.0-2.70
Highest resolution shell (Å)	2.75-2.70
Unique reflections	38010 (1906)
Redundancy	3.4 (3.5)
Completeness (%)	99.4 (100.0)
R_{merge}	0.049 (0.281)
R_{pim}	0.038 (0.194)
R_{rim}	0.070 (0.366)
CC half	0.895
Average $I/\sigma(I)$	18.3 (3.9)
Refinement	
R_{work} (%)	18.7 (26.5)
R_{free} (%)	21.9 (27.9)
Mean *B* value (Å^2)	43.48
B from Wilson plot (Å^2)	42.1
RMSD bond lengths (Å)	0.016
RMSD bond angles (degrees)	1.643
No. of amino acid residues	A, B, C, U = 204
No. of water molecules	74
Ramachandran plot	
Most favored regions (%)	97.10
Additional allowed regions (%)	100.00

3.12 FIGURES

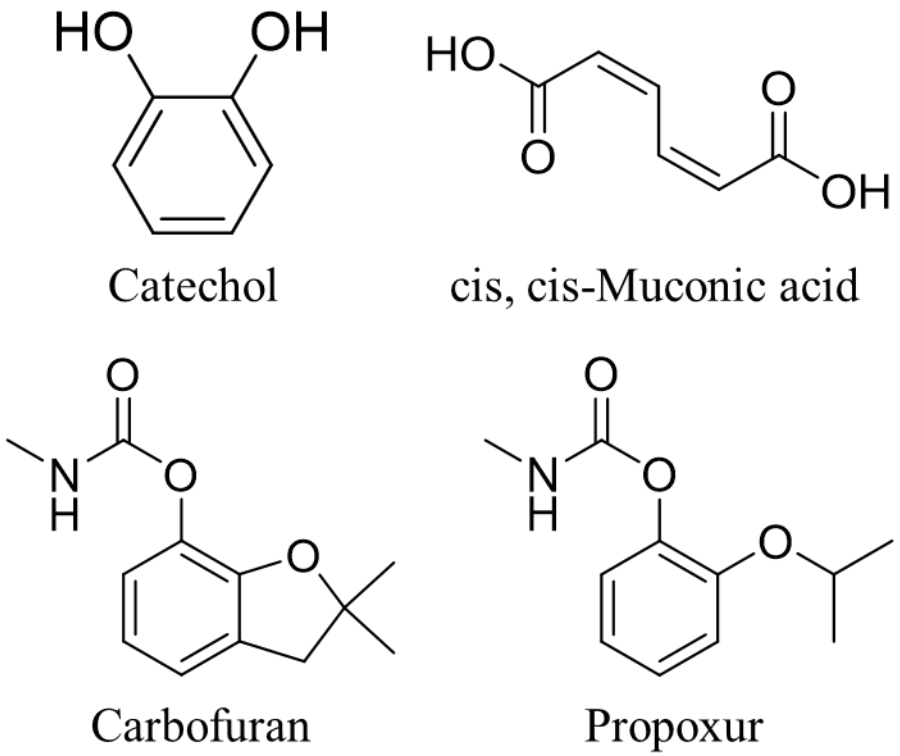

Figure 3.1: The structure of catechol and catechol derivatives. The structure of catechol, the product of cleaving catechol via an intradiol ring-cleavage dioxygenase (cis, cis-Muconic acid) and two catechol-derived insecticides.

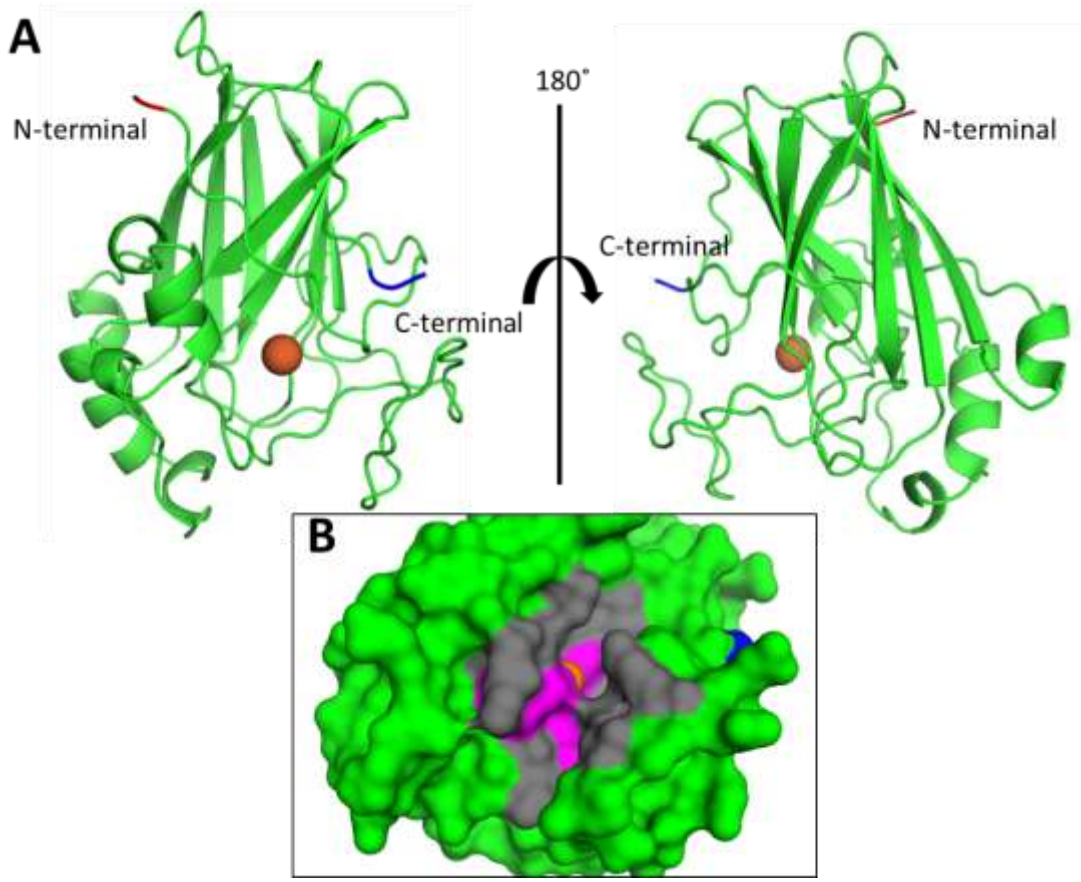

Figure 3.2: Crystal structure of TuDioxygenase. A) Cartoon representation of TuDioxygenase (PDB code: 5VG2). The β-sandwich core is conserved with most ID-CRDs. The β-sandwich only consists of 7 β-strands which is different from the typical 8 β-strands in ID-CRDs, but it is most likely attributed to the truncation of residues 1-47 of the recombinant TuDioxygenase studied here. The ferric center is shown by an orange sphere. The N-terminal is colored red and the C-terminal is colored blue. **B)** Surface representation of TuDioxygenase. Active site residues are colored in magenta and residues aligning the active site which form a narrow cleft are colored grey.

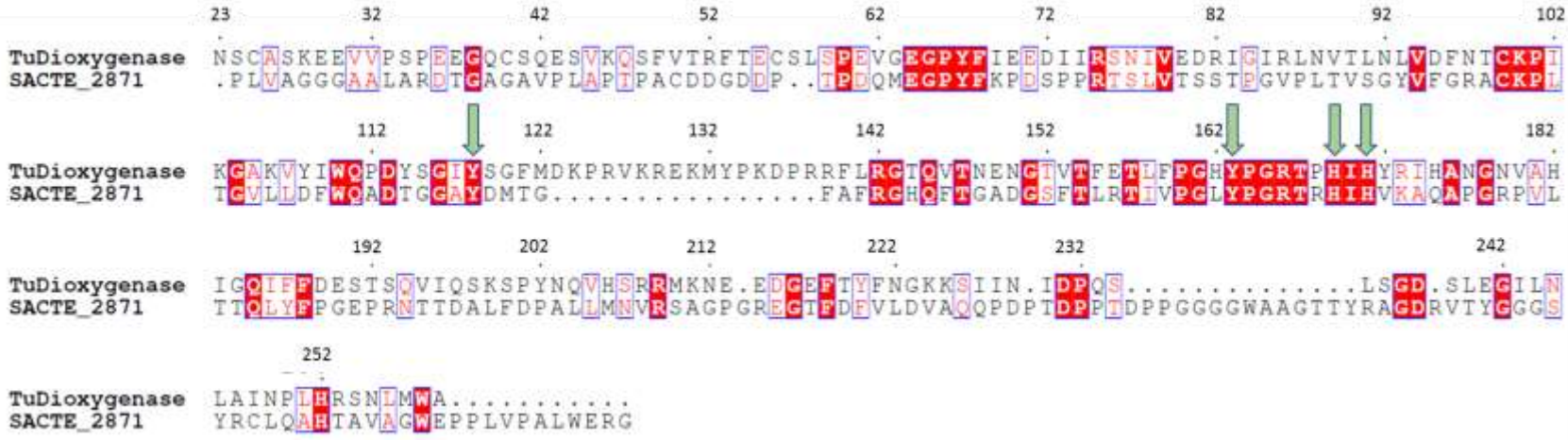

Figure 3.3: Sequence alignment of TuDioxygenase with SACTE_2871 homologue. Sequences are aligned without their respective signal peptides. Numbered residues are with respect to the sequence for TuDioxygenase. TuDioxygenase and SACTE_2871 (PDB code: 4ILT) share a low sequence identity of 17% but a decent sequence similarity of 34%. Residues highlighted in red are identical and residues colored red are similar. Conserved secondary structure is shown by blue boxes. Active site residues are shown with greens arrows. Figure was made using ESPript 3.0 [82].

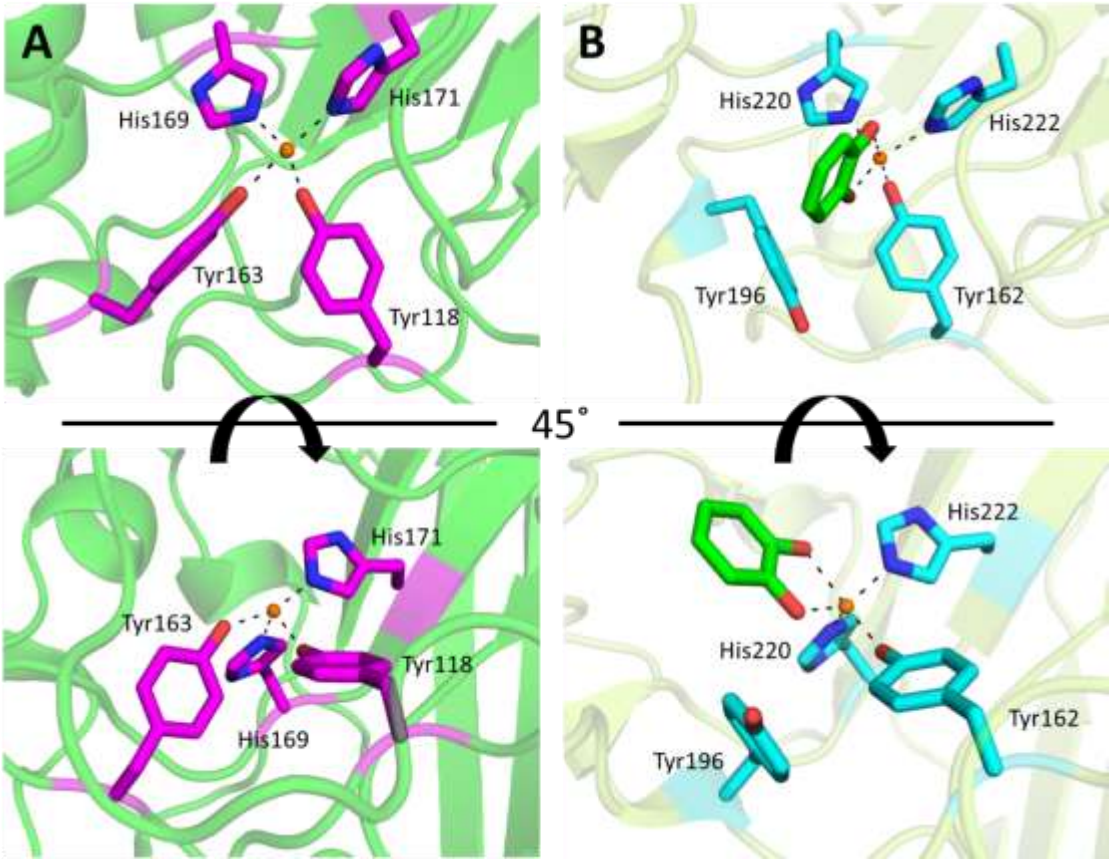

Figure 3.4: Active site of TuDioxygenase and a bacterial 1,2- dioxygenase homologue. **A)** Active site of TuDioxygenase. Two tyrosines (Tyr118, Tyr163), two histidines (His169, His171) and a water molecule (not shown) coordinate the active site iron in a trigonal bipyramidal manner. Both tyrosines are in equatorial positions when no substrate is bound. **B)** Active site of 1,2-dioxygenase from *Rhodococcus opacus* (PDB code: 3HHY) which has a similar intradiol dioxygenase domain to TuDioxygenase. Upon binding catechol, Tyr196 (corresponds to Tyr163 of TuDioxygenase) is repositioned form equatorial to axial and the catechol binds in a bidentate manner. The water molecule bound to the ferric center is also displaced and the trigonal bipyramidal binding on the iron is retained [85].

CHAPTER 4

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF 2-METHYLCITRATE SYNTHASE FROM *ASPERGILLUS FUMIGATUS* AND CITRATE SYNTHASE FROM HUMANS

4.1 BACKGROUND AND INTRODUCTION

Aspergillus fumigatus is a fungus that typically grows on compost and agricultural crops worldwide [4]. As part of its life cycle, *A. fumigatus* will form stress-tolerant, readily-airborne spores in high capacity [100]. Consequently, this permits *A. fumigatus* to be one of the most prevalent airborne pathogens in the world [5, 101]. *A. fumigatus* is an opportunistic pathogen and spores are inhaled by humans on a frequent basis but normally eradicated from the body by the innate immune system neutrophils and macrophages [100, 102]. However, inhalation of *A. fumigatus* spores by immunocompromised individuals can lead to invasive aspergillosis which has a mortality rate of 50-95% [4]. Individuals undergoing chemotherapy, organ transplants or that have chronic diseases such as tuberculosis, asthma or cystic fibrosis are especially more susceptible to invasive aspergillosis [4, 100]. Drugs currently used to treat *A. fumigatus* infections include azole compounds, polyenes and echinocandins, but resistance to these drugs is becoming more prevalent and new classes of anti-fungals are needed [100, 103–105].

A new potential target for treating *A. fumigatus* infections is the metabolic enzyme 2-methylcitrate synthase (mcsA) [5, 106, 107]. The mcsA enzyme catalyzes the primary step in the 2-methylcitrate synthase pathway which is specific to fungi (Figure 4.1) [5]. McsA performs a condensation reaction with oxaloacetate and propionyl-Coenzyme

(CoA), a toxic metabolite that is produced by leucine, methionine and valine breakdown, into 2-methylcitrate and non-toxic CoA-SH (Figure 4.2B)[5]. It has been shown that knocking out this enzyme greatly impairs growth and dissemination of *A. fumigatus* [5, 107]. Furthermore, mcsA also has citrate synthase activity in which acetyl-CoA is used as co-substrate instead of propionyl-CoA to generate citrate from oxaloacetate [5].

In humans, propionyl-CoA is detoxified by converting it to methylmalonyl-CoA (via a carboxylation and isomerization reactions) which is then fed into the citric acid cycle (Figure 4.1) [5]. Although fungi and humans incorporate different metabolic cycles to detoxify propionyl-CoA, mcsA performs the same reaction as human citrate synthase (hCS), an enzyme part of the citric acid cycle (Figure 4.1, 4.2A). Moreover, mcsA and hCS share a 51% amino acid sequence identity which makes finding mcsA-specific inhibitors more challenging. Because of their similarities, mcsA and hCS were studied in parallel. Both enzymes had their crystal structures determined to observe any structural differences, and inhibitor efficacy was studied for both enzymes for the purpose of finding mcsA-specific inhibitors.

4.2 Cloning of hCS into pMCSG53

The plasmid HsCD00434245 containing the gene for hCS was purchased from DNASU [108–110], and the hCS gene (residues 29-466) was cloned into pMCSG53 (DNASU) using ligation independent cloning (LIC). Briefly, hCS was amplified using the primers in Table 1 and the protocol for Phusion polymerase was followed (NEB, Ipswich, MA); for thermalcycling, an extension time of 30 seconds for 35 cycles was used with 100 ng of template. Primers were designed to create complimentary 5' and 3' overhangs to pMCSG53 for LIC. The hCS insert was run on a 1.0% agarose gel and gel excised and purified using a ThermoFisher GeneJET Gel Excision Kit (Grand Island, NY).

Instead of cutting pMCSG53 with SmaI for LIC, pMCSG53 was amplified at the LIC site to generate blunt-ended 5' and 3' ends. To amplify pMCSG53, the primers pMCSG53-F and pMCSG53-R in Table 4.1 were used and the protocol for KOD polymerase was followed. For the KOD polymerase reaction, 4.0 µL of 25 mM magnesium sulfate and 10 µL of betaine monohydrate (5.0 M) with 10 ng of template were added in addition to other reagents listed in the manufacturer's protocol (EMD Millipore, Billerica, MA). Thermalcycling was as follows: 95˚C for 2 minutes, 95˚C for 20 seconds, 10 second annealing at 61˚C, 2 minute extension at 70˚C for 25 cycles and a final extension at 70˚C for 10 minutes. After PCR, reactions were run on a 1.0% agarose gel and amplified pMCSG53 was gel excised and purified using a ThermoFisher GeneJET Gel Excision Kit (Grand Island, NY).

After amplification and purification of PCR products, the hCS insert and PCR-amplified vector were prepared for LIC. For the insert, a 40 µL reaction of 8.5 µL insert (2,500 fmoles), 4.0 µL 10X T4 DNA polymerase buffer, 1.0 µL deoxycytidine triphosphate (dCTP; 100 uM), 2.0 µL DTT (100 mM), 23.5 µL diH_2O and 1.0 µL T4 DNA polymerase (3.0 units). Vector was prepared in a 40 µL reaction containing 28 µL vector (620 fmoles), 4.0 µL 10X T4 DNA polymerase buffer, 1.0 µL deoxyguanosine triphosphate (dGTP; 100 uM), 2.0 µL DTT (100 mM), 4.0 µL diH_2O and 1.0 µL of T4 DNA polymerase (3.0 units). Reactions were incubated at room temperature for 30 minutes and then T4 DNA polymerase was inactivated by heating at 75˚C for 20 minutes. After inactivation, 10 µL and insert and vector reactions were mixed together for 5 minutes at room temperature. Post-incubation, 1.0 µL of EDTA (25 mM) was added to the reaction and incubated for 5 more minutes at room temperature. The entire reaction volume was transformed into

chemically competent DH5-α *E. coli* cells using heat shock at 42˚C for 45 seconds. Cells were plated on LB-Agar plates containing AMP (50 µg/mL) and grown at 37˚C overnight. Clones were inoculated into 5.0 mL LB containing AMP and grown at 37˚C overnight (16 hours) with shaking. Minipreps were performed with a ThermoFisher GeneJET Plasmid Miniprep Kit (Grand Island, NY) following the manufacturer's protocol and eluted with 50 µL of nuclease-free water. Clones containing the hCS insert were first verified by setting up a restriction digest with NdeI and XhoI follow the manufacturer's protocol (NEB, Ipswich, MA), and then sent to EtonBioscience (Research Triangle Park, NC) to confirm the correct sequence with T7 forward and T7 reverse primers.

4.3 Expression and Purification of hCS and mcsA

Recombinant hCS and mcsA were truncated to residues 30-466 and 29-465, respectively, due to the presence of signal peptides at the N-terminus of these proteins; pig heart citrate synthase (pCS) was purchased as an ammonium sulfate suspension from Sigma-Aldrich (St. Louis, MO). The hCS-pMCSG53 plasmid was transformed into BL-21 (DE3) pLysS cells due to the presence of rare *E. coli* codons in the hCS gene. The mcsA plasmid was ordered from DNA 2.0 (Newark, CA) in KAN-resistant vector pJExpress411 and codon optimized for *E. coli*. Both plasmids operate under an IPTG inducible T7 promoter and contain a cleavable N-terminal 6xhis-tag with TEV protease. Cells were grown to an O.D. of 0.8 at 37˚C with shaking and protein expression was induced with 0.4 mM IPTG; cells were then cooled down to 16˚C, shaken overnight (12 hours), pelleted in a Beckman Coulter centrifuge (Indianapolis, IN) for 10 minutes at 4˚C and frozen at -80˚C until needed further.

For purification, the hCS pellet was resuspended in lysis buffer (50 mM Tris, 500 mM NaCl, 10 mM imidazole, 20 mM β-ME, 2% glycerol) and mcsA pellet was resuspended in mcsA buffer (50 mM Tris, 500 mM NaCl, 20 mM β-ME, 20% glycerol). The same buffers were used for both proteins for the remainder of the purification. After resuspension, cells were lysed by sonication with a Branson 45 Sonifier (ThermoFisher, Grand Island, NY). Crude extract was separated by spinning the lysate in a Beckman Coulter centrifuge (Indianapolis, IN) for 25 minutes at 4˚C. The clear, yellowish supernatant was poured into a 12 x 1.5 cm Bio-Rad column (Hercules, CA) filled with 5.0 mL of NiNTA resin (ThermoFisher, Grand Island, NY) previously equilibrated in wash buffer (50 mM Tris pH 7.4, 150 mM NaCl, 30 mM Imidazole, 2% glycerol, 20 mM β-ME) and then washed with wash buffer. Protein was eluted using elution buffer (50 mM Tris pH 7.4, 150 mM NaCl, 250 mM imidazole, 2% glycerol, 20 mM β-ME) and immediately put into dialysis buffer (10 mM Tris pH 7.4, 150 mM NaCl, 5 mM β-ME) in ThermoFisher SnakeSkin Dialysis Tubing (Grand Island, NY) with a 10,000 MW cutoff. Protein was dialyzed in dialysis buffer for 12 hours at 4˚C.

After dialysis, protein was concentrated with Amicon Ultra concentrators (EMD Millipore, Billerica, MA) with a 10,000 MW cutoff. Concentrated protein was put on a Superdex 200 column attached to an ÄKTA Pure FPLC system (GE Healthcare, Marlborough, MA) equilibrated in FPLC buffer (10 mM Tris pH 7.4, 150 mM NaCl). Peaks corresponding to either hCS or mcsA were pooled and the protein concentration was determined using the Bradford method [68]. To remove the N-terminal his-tag from mcsA, it was subjected to TEV protease cleavage. The hCS protein also contained a cleavable N-terminal his-tag; however, TEV cleaves hCS at Ser166 so the his-tag could not be removed

from hCS. Briefly, pure mcsA (1-2 mg/mL) was incubated with TEV protease in a 1:100 (w/w) protease/protein ratio and dialyzed for 12 hours in dialysis buffer. After cleavage, mcsA was loaded onto an NiNTA column equilibrated in FPLC buffer and mcsA was collected in the flow through and concentrated. The mutant constructs, hCSA348G and mcsAG352A, were purified in the same manner as their wild type counterparts.

4.4 Data Collection, Structure Determination and Refinement

Table 4.2 shows the data collections statistics for crystal structures mentioned here. All chemicals mentioned here were purchased as a screen from Hampton Research (Aliso Viejo, CA) or single chemicals from ThermoFisher (Grand Island, NY) or Sigma-Aldrich (St. Louis, MO). Ethyl-CoA was purchased from Jena Bioscience (Jena, Germany). Crystallization experiments for pCS, hCS, hCSA348G, mcsA and mcsAG352A were all performed at room temperature using the sitting-drop vapor diffusion method and MRC 2-drop 96-well crystallization plates (Hampton Research, Aliso Viejo). Recombinant protein (around 4.0 mg/mL; 0.5 mg/mL for pCS) was mixed with mother liquor in a 1:1 ratio. Crystallization conditions for each structure in Table 4.2 are as follows: PDB code: **5UQO** – 0.1 M Tris pH 8.5, 25% w/v PEG3350; PDB code: **5UQQ** – 0.2 M sodium tartrate dibasic dihydrate pH 7.85, 20% w/v PEG3350; PDB code: **5UQR** – 0.15 M DL-malic acid pH 7.0, 20% w/v PEG3350 (co-crystallized with 1.0 mM ethyl-CoA and oxaloacetate); PDB code: **5UQU** – 0.2 M ammonium sulfate, 0.1 M Tris pH 8.5, 25% w/v PEG3350 (co-crystallized with 1.0 mM propionyl-CoA and oxaloacetate); PDB cod: **5UQ5** – 0.2 sodium tartrate pH 7.85, 20% w/v PEG3350; PDB code: **5UZR** – 0.2 M ammonium acetate pH 7.1, 20% w/v PEG3350; PDB code: **5UZQ** – 0.1 M succinic acid pH 7.0, 12% w/v PEG3350 (co-crystallized with 1.0 mM ethyl-CoA and oxaloacetate); PDB code: **5UZP** – 0.2 M

ammonium acetate, 0.1 M sodium citrate tribasic dihydrate pH 5.6, 15% w/v PEG4000 (co-crystallized with 1.0 mM propionyl-CoA and oxaloacetate). Cryo-protectant of 50% mineral oil with 50% paratone-N was only used for crystal structures 5UQQ, 5UQR, 5UQ5 and 5UZR.

Crystals were cryo-cooled in liquid nitrogen and data was collected using either Southeast Regional Collaborative Access Team (SER-CAT) 22BM, SER-CAT 22ID, Structural Biology Center (SBC) 19BM, SBC 19ID [111] or the Life Sciences Collaborative Access Team (LS-CAT) 21-ID-F or 21-ID-G beamlines at the Advanced Photon Source (APS), Argonne National Lab (Argonne, IL). Data were processed with the HKL-2000 software package [112]. Molecular replacement was performed using MOLREP [113, 114] integrated with HKL-3000 [43] using the PDB model 3ENJ as a starting model for 5UQO and 5UZR. The crystal structure of pCS with citrate and CoA-SH bound (PDB code: 2CTS) was used as the starting model for holo-mcsA (5UQR). Refinement was performed using REFMAC [46] and HKL-3000 [43]. Non-crystallographic symmetry was used during the whole process of refinement. TLS refinement was used during the last stages of refinement and the TLS Motion Determination server was used for partitioning protein chains into the rigid bodies undergoing vibrational motions [90, 115, 116]. Model was updated and validated with COOT [79]. MOLPROBITY was used in the final steps of the model validation [117]. The final models together with structure factors were deposited to the Protein Data Bank with accession numbers reported in Table 4.2.

The programs ProFunc [91] and PDBePISA [53] were used to analyze oligomeric assembly of the enzymes, bond measurements were performed in COOT and structures were made using PyMOL [80].

4.5 Structural Comparison of pCS and cCS to Recombinant hCS

Citrate synthase from pig, chicken (cCS) and human all share very high amino acid sequence identity as follows: pCS and cCS – 93%, pCS and hCS – 96% and cCS and hCS – 92%. Figure 4.3 shows a sequence alignment between pCS, cCS and hCS, and Figure 4.4 shows secondary structure conservation. Overall structure of hCS is similar to that of pCS and cCS with 19 alpha helices and four beta sheets (two parallel and two anti-parallel); helices 7, 13 and 17 are kinked at residues Ser179, Leu303 and His404 respectively, in these helices. Furthermore, pCS, cCS and hCS all form non-cooperative homodimers that contain two active sites where both monomers (~49 kDa each without signal peptide) contribute residues to both active sites (Figure 4.5A).

At the dimer interface are helices 6, 7, 11 and 12 from both chains that are anti-parallel to one another; several electrostatic, van der Waals and hydrogen bonding interactions keep these helices sandwiched together with an interface area of about 6000 $Å^2$ (Figure 4.5B). In the active sites of hCS are two histidines (His265, His347 from chain A) and three arginines (Arg356, Arg428 from chain A and Arg448 from chain B) that are involved in binding oxaloacetate (Figure 4.5C). Several amine and carbonyl groups from peptide backbone form salt bridges and hydrogen bonds with the adenosine moiety of acetyl-CoA. A conserved Asn400 binds the carbonyl adjacent to the thiol group of acetyl-CoA. Arg73 from chain A binds to P2 of CoA and Arg191 from chain B binds to the ribose sugar of CoA based on superimposition of hCS (PDB code: 5UZR) and cCS (PDB code:

4CSC) structure that contains bound D-malate and acetyl-CoA. Based on whether ligand is present, citrate synthases mentioned here can adopt "open," "partially closed" or "closed" conformations which has been well-studied for pCS and cCS [118]; a hinge region for these conformational changes is located at His301 and Gly302 in hCS.

4.6 Mechanism for pCS, cCS and hCS

The mechanism of the condensation reaction for pCS and cCS has been well studied and characterized [118–122]. Substrates bind in an ordered mechanism where oxaloacetate binds first and increases enzyme affinity for acetyl-CoA by a factor of at least 20 [118]. Binding of oxaloacetate results in enzyme conformational change to "partially closed" and a binding site for acetyl-CoA is formed. It has been proposed that acid-base catalysis with His274 and Asp402 in the active site, respectively, promotes enolization of the thioester group on acetyl-CoA which is the rate-limiting step [118, 119]. A citryl-thioester intermediate is then formed with oxaloacetate via Claisen Condensation that is hydrolyzed most likely by water molecules trapped in the active site to generate the products citrate and CoA [118, 119]. Kinetic values obtained for recombinant his-tagged hCS purified here is comparable to that of pCS, cCS and hCS in literature [120, 121, 123] and are discussed later for mcsA comparison.

4.7 Comparison of hCS and mcsA Crystal Structures

Secondary Structure Conservation

The sequence identity shared between hCS (466 residues) and mcsA (465 residues) is 51% (Figure 4.3), and structure conservation is mapped between hCS and mcsA in Figure 4.6. The apo structures of hCS (PDB code: 5UZR) and mcsA (PDB code: 5UQO) were superimposed in PyMOL with an RMSD value of 0.7 Å. Similar to other citrate synthases

mentioned here, mcsA forms a homodimer that adopts “open,” “partially closed” or “closed” conformations upon binding ligand; however, positive cooperativity was observed between active sites of mcsA based on kinetic results. Figure 4.7 shows space-filling models of apo-hCS (PDB code: 5UZR), apo-mcsA (PDB code: 5UQO), holo-cCS (PDB code: 4CSC - homologue) and holo-mcsA (PDB code: 5UQR). The conformational changes upon binding substrate is comparable for both hCS and mcsA.

McsA has a total of four beta sheets (two parallel and two anti-parallel), but 21 alpha helices compared to the 19 alpha helices found in hCS. At the C-terminal of helix 8 in mcsA is Pro220 (corresponds to Asn219 in hCS helix 8) which causes the formation of additional helix 9. The other additional helix in mcsA, helix 17, is one turn and starts at Pro378. This residue corresponds to Pro372 in hCS; however, the residues of helix 17 in mcsA are more rigid and capable of forming peptide backbone hydrogen bonds whereas the corresponding residues in hCS form a partially flexible loop. Several helices are kinked in mcsA: helix 7 at Ser180, helix 8 at Tyr195 and helix 14 at Leu307. The main dimer interface of mcsA is an 8 anti-parallel alpha helix sandwich formed by helices 6, 7, 12 and 13. Each active site of mcsA is similar to that of hCS and contains two histidines (His269 and His351 from chain A) and three arginines (Arg360, Arg434 from chain A, Arg454 from chain B) that contribute to oxaloacetate binding (Figure 4.8). The binding of CoA for hCS and mcsA appears to be similar except for a difference of one residue: Arg191 in hCS (chain B) and Lys192 in mcsA (chain B). Lys192 forms a single salt bridge with the ribose sugar of CoA. Furthermore, upon adopting the “closed” conformation, trapped water molecules can be observed in the active site for both hCS and mcsA.

The main dimer interface of mcsA is an 8 anti-parallel alpha helix sandwich formed by helices 6, 7, 12 and 13. The dimer interface area of mcsA was determined to be around 5400 $Å^2$ which is about 600 $Å^2$ less than the dimer interface area hCS. Interestingly, when substrate is the bound, the dimer interface area increases to about 6200 $Å^2$ for both hCS and mcsA. This significant increase of about 600 $Å^2$ may contribute to the positive cooperativity observed for mcsA. The hinge region in mcsA is at His305 and Gly306. Although hCS and mcsA have similar "closed" conformations, several residues in hCS have greater movement with respect to corresponding residues in mcsA (Figure 4.9). Most of these residues are located near the active site or hinge region and move a greater distance (based on Cα's) in hCS than they do in corresponding residues in mcsA. Based on these measurements, hCS specificity for only acetyl-CoA may be attributed to larger residue movements, which consequently may be forming a tighter "closed" conformation that cannot accommodate propionyl-CoA due to a more compact active site with respect to mcsA.

4.8 Comparison of hCS to hCSA348G and mcsA to mcsAG352A

Measurements mentioned here were all performed in COOT (from Cα's unless otherwise stated) [79]. Initially, apo-hCS (PDB code: 5UZR) and apo-mcsA (PDB code: 5UQO) were compared to determine any differences in the active site which may explain why hCS has no 2-methylcitrate synthase activity with propionyl-CoA. The only difference observed in the active site was the presence of Ala348 in hCS and Gly352 in mcsA which are located near the CoA binding site (Figure 4.10) (for hCSA348G and mcsAG352A site-directed mutagenesis, see section 4.14).

One crystal structure was obtained (PDB code: 5UZP) for hCSA348G in which oxaloacetate was bound in both active sites (Figure 4.11). Interestingly, the crystal structure does not show a "partially closed" conformation as is normal upon oxaloacetate binding (compared to pCS with oxaloacetate bound, PDB code: 4CTS) and formation of the acetyl-CoA binding sight is not observed. In the hCSA348G crystal structure, active site residues Arg448 (chain B) and His347 (chain A) are not in their normal positions for oxaloacetate binding and resemble that of apo-hCS (Figure 4.5A). Upon oxaloacetate binding, His347 moves about 3.7 Å towards oxaloacetate but it is not observed in the crystal structure for hCSA348G. Furthermore, based on 4CTS, hCS should form a salt bridge with both Glu266 and oxaloacetate, but the positioning of Arg448 in hCSA348G only allows it to bind to oxaloacetate.

No holo structures of hCS or hCSA348G were obtained from these studies; therefore, cCS with acetyl-CoA and D-malate bound (PDB code: 4CSC) will be used as a structural homologue for holo-hCS and conformational change discussion. In 4CSC, the methyl side chain of Ala321 (Ala348 in hCS) is about 3.4 Å from O4 of P2 in the adenosine moiety and the Cα is 3.3 Å. Although this may cause some steric clash, it does not explain how the A348G mutation hinders the "partially closed" conformational change. Comparing hCSA348G to apo-hCS and apo-pCS (PDB code: 4CTS), A348G displays some distortion in neighboring residues such as Val349 and Leu350 that could be inhibiting enzyme conformational changes. The hCSA348G mutant still displayed citrate synthase activity which is discussed further in the kinetics results.

One crystal structure was obtained for mcsAG352A (PDB code: 5UQU) with oxaloacetate bound in one active site and a sulfate ion with CoA in the other (Figure 4.12).

The sulfate ion came from the crystallization condition and two other sulfates were bound on the surface of the protein. The G352A mutation did not seem to have an impact on positioning of active site residues and substrate binding which was observed by comparing apo-mcsA (PDB code: 5UQU) to holo-mcsA (PDB code: 5UQR). All active site residues involved in binding oxaloacetate and acetyl/propionyl-CoA substrates were bound to the sulfate ion and CoA, respectively, in their normal substrate binding orientations. Ala352 shows little steric clash with CoA. The side chain methyl group of Ala352 is about 3.8 Å away from O4 of P2 in the adenosine moiety and the Cα is 3.4 Å which is slightly greater than wild type hCS. In apo-mcsA, Gly352 Cα is 3.3 Å from P2 and little difference in residue positioning is observed between Gly352 and G352A. Binding of either oxaloacetate only or the sulfate ion and CoA induced the "partially closed" or "closed" confirmations, respectively, that are observed for holo-mcsA. Comparing the structures of hCSA348G and mcsAG352A, only a disruption in the conformational change of hCSA348G was observed. Both mutants retained their native activities, and hCSA348G was not able to utilize propionyl-CoA as co-substrate.

4.9 Other Crystal Structures of hCS and mcsA

The apo-hCS 5UZQ crystal structure only contained one chain in the asymmetric unit. Two cysteines in the structure, Cys211 and Cys359, were modified to S, S-2 (hydroxyethyl) thiocysteines (CME) molecules. Both apo-pCS (PDB code: 5UQ5) and apo-hCS (PDB code: 5UZR) have a chlorine ion bound in each active site. Apo-mcsA 5UQQ crystal structure contains 6 chains in the asymmetric unit and has one tartrate molecule bound to Arg74 (chain B) and Lys456 (chain C). The tartrate molecule did not

cause significant conformational change except to Gly352 in which the Cα moves about 5.3 Å away from the tartrate molecule.

4.10 hCS, hCSA348G, mcsA and mcsAG352A Substrate Kinetics

All chemicals mentioned here were purchased from ThermoFisher (Grand Island, NY) or Sigma-Aldrich. Kinetic parameters for hCS, hCSA348G, mcsA and mcsAG352A are shown in Table 4.3 and were measured under pseudo-first order conditions by monitoring the reaction of the product CoA-SH with 5-5'-dinitrobenzoic acid (DTNB) [5]; the reaction of CoA-SH with DTNB produces 2-nitro-5-thiobenzoate (TNB) that absorbs at 412 nm with a molar extinction coefficient of 14,150 M^{-1} cm^{-1}. Measurements were performed by using a ThermoScientific Nanodrop 2000 Spectrophotometer (Grand Island, NY) using a 1.0 cm path length. Reaction volumes were a total of 1.0 mL and contained 100 mM Tris pH 8.0, 1.0 mM DTNB and 40 µM oxaloacetate and 40 µM acetyl-CoA for hCS; the same reaction volumes were used for mcsA except with 30 µM oxaloacetate and 30 µM acetyl or propionyl-CoA. Enzyme was added last to each reaction. The enzyme working concentrations were 2.5 nM hCS, 6.15 nM hCSA348G, 22.6 nM mcsA or 50.8 nM mcsAG352A; for inhibitor studies, concentrations of substrates were increased to 50 µM and 40 µM for hCS and mcsA, respectively. After addition of enzyme to the reaction, the cuvette was mixed by inversion and immediately the A_{412} was measured. All reactions were performed in triplicate. Slope values were calculated using ThermoFisher Nanodrop 2000 software (Grand Island, NY). Data was further processed using OriginPro software (Northampton, MA). The Michaelis-Menten equation was used to fit hCS and hCSA348G data, and the Hill coefficient was calculated using the Hill equation; the Hill equation was used for mcsA and mcsAG352A data (Figure 4.13).

OXALOACETATE BINDING

hCS had higher K_m for oxaloacetate than hCSA348G but a lower V_{max}. McsA had lower K_m for oxaloacetate than mcsAG352A but a higher V_{max}. The cooperativity displayed by hCS towards oxaloacetate was n = 1 (Hill coefficient) which indicates no cooperativity, and n values for mcsA and mcsAG352A were above one indicating positive cooperativity. The turnover number for hCS (208 sec^{-1}) and mcsA (29.2) was greater than hCSA348G (111 sec^{-1}) and mcsAG352A (11.2 sec^{-1}), respectively. The catalytic efficiency of hCS (5.9 x 10^7) and mcsA (6.0 x 10^6 sec^{-1} M^{-1}) was greater than hCSA348G (1.9 x 10^7) and mcsAG352A (3.2 x 10^6), respectively. Comparing hCS and mcsA, hCS had higher K_m, lower V_{max}, higher k_{cat} and higher catalytic efficiency than mcsA for oxaloacetate.

ACETYL-COA BINDING

hCS had higher K_m for acetyl-CoA than hCSA348G but a lower V_{max}. McsA had lower K_m for acetyl-CoA than mcsAG352A but a higher V_{max}. The Hill coefficient of hCS towards acetyl-CoA was n = 1, but hCSA348G had a value of n = 0.85 ± 0.08 which indicates potential negative cooperativity. The Hill coefficient for mcsA and mcsAG352A were above one, where mcsAG352A seemed to display greater positive cooperativity with a value of n = 2.9. The turnover number for hCS (260 sec^{-1}) and mcsA (32.7 sec^{-1}) were greater than hCSA348G (117 sec^{-1}) and mcsAG352A (10.4 sec^{-1}), respectively. The catalytic efficiency of hCS (3.4 x 10^7 sec^{-1} M^{-1}) and mcsA (6.6 x 10^6 sec^{-1} M^{-1}) were greater than hCSA348G (1.1 x 10^7 sec^{-1} M^{-1}) and mcsAG352A (2.9 x 10^6 sec^{-1} M^{-1}), respectively. Comparing hCS and mcsA, had higher K_m and V_{max} for acetyl-CoA, but hCS had greater turnover and catalytic efficiency.

PROPIONYL-COA BINDING

2-methylcitrate synthase activity was not observed with using propionyl-CoA as co-substrate for hCS and hCSA348G. McsA had lower K_m but higher V_{max} for propionyl-CoA than mcsAG352A. The Hill coefficient was above one for both mcsA and mcsAG352A. The turnover number (30.1 sec^{-1}) and catalytic efficiency (6.5 x 10^6 sec^{-1} M^{-1}) of mcsA were both greater than mcsAG352A (12.2 sec^{-1} and 3.2 x 10^6 sec^{-1} M^{-1}).

4.11 SCREENING FOR INHIBITORS USING DIFFERENTIAL SCANNING FLUORIMETRY (DSF)

DSF is a high-throughput technique that measures protein stability via melting temperature (T_m) in the presence of different ligands [124, 125]. Briefly, protein (about 1.0 mg/mL) was mixed 1000:1 with SYPRO-Orange dye (ThermoFisher Scientific, Grand Island, NY) in 50 mM Tris pH 8.0. Protein was set in a Bio-Rad Hardshell 96-well PCR plate (Hercules, CA) then covered with a Bio-Rad adhesive PCR plate seal. A Bio-Rad CFX96 Real-Time PCR machine (Hercules, CA) was used to measure SYRPO-Orange excitation at 488 nm and fluorescence resonance energy transfer (FRET). Reactions were started at 30˚C and increased to 90˚C in 2˚C increments every minute. Protein T_m in the presence of ligands was measured and compared to protein T_m without any ligands present. Data was processed with Bio-Rad CFX Manager software (Hercules, CA). Reactions with substrates and products were performed in duplicate. Reactions with other ligands were performed in triplicate.

Initially, inhibitor screening using this technique was validated by using 1.0 mM substrates and products as ligands (Table 4.4). As expected, all substrates and products increased thermal stability of hCS, hCSA348G, mcsA and mcsAG352A with the most dramatic increase in stability arising from incubation with oxaloacetate and oxaloacetate

with CoA-SH for all proteins. To find mcsA-specific inhibitors, ligands that only stabilize mcsA were tested further for their inhibitor efficacy via UV-Vis. Non-substrate/product compounds (1.0 mM) that were tested include 3-phosphonopropionic acid, 3,3-thiodipropionic acid, dihydroxyfumaric acid, trimethylcitrate, DL-malic acid, maleic acid and hydroxycitrate (Figure 4.14). With the exception of trimethylcitrate, all of these ligands increased the stability of mcsA only. Trimethylcitrate did not increase the stability of either mcsA or hCS, but was still tested for inhibitory properties, along with the other ligands mentioned, with UV-Vis spectroscopy because it is a substrate analogue.

4.12 Testing Inhibitors by UV-Vis Spectroscopy

For inhibitor studies, inhibitors were used in three different concentrations and Lineweaver-Burk plots were used to calculate inhibitor constants (K_i). Ligands from DSF were tested for their efficacy against both substrates for each enzyme: oxaloacetate and acetyl-CoA for hCS; oxaloacetate and propionyl-CoA for mcsA. Inhibitor reactions were performed in 50 mM Tris pH 8.0, 1.0 mM DTNB in 1.0 mL total. The working concentration of enzyme for reactions was about 5.0 nM for hCS and about 36 nM for mcsA. Saturating concentrations of 50 µM oxaloacetate and acetyl-CoA were used for hCS reactions, and 40 µM oxaloacetate and propionyl-CoA used for mcsA reactions. Most inhibitors tested had their pH adjusted to around 8.0 and the pH of every reaction with inhibitor was checked to make sure inhibition was not due to pH change. The ligands 3-phosphonopropionic acid, 3,3-thiodipropionic acid, dihydroxyfumaric acid, trimethylcitrate, DL-malic acid and maleic acid were used in concentrations of 20 mM, 40 mM and 70 mM; trimethylcitrate did not increase thermal stability of hCS or mcsA, but it was still tested because it is a substrate analogue. Hydroxycitrate was tested in

concentrations of 8.0 mM, 16 mM and 24 mM. Lineweaver-Burk plots for each ligand are shown in Figure 4.15. A summary of inhibitor constants and type of inhibition displayed towards each substrate is shown in Table 4.5.

All the inhibitors tested had K_i values in the low millimolar range except trimethylcitrate. Trimethylcitrate had a K_i around 300 mM or higher for both hCS and mcsA (data not shown). The most mcsA-specific inhibitor was dihydroxyfumaric acid (a derivative of fumarate – a precursor in both the citric acid cycle and 2-methylcitrate cycle (Figure 4.1) in which variable oxaloacetate showed mixed inhibition (K_{ia} = 23.6 mM, K_{ib} = 90.6 mM) and variable propionyl-CoA showed competitive inhibition (K_i = 17.1 mM) (Figure 4.15 N, P); for hCS, dihydroxyfumaric acid had competitive inhibition with oxaloacetate (K_i = 90.5 mM) and acetyl-CoA (K_i = 73.0 mM). The best inhibitor of hCS was malic acid, another precursor in both the citric acid cycle and 2-methylcitrate cycle. Malic acid had mixed inhibition with both oxaloacetate (K_{ia} = 14.2 mM, K_{ib} = 169.3 mM) and acetyl-CoA (K_{ia} = 6.5 mM, K_{ib} = 19.3 mM) for hCS (Figure 4.15 Q, S); malic acid was noncompetitive with both oxaloacetate (K_i = 67.5 mM) and propionyl-CoA (K_i = 62.1 mM) for mcsA. Hydroxycitrate, a product analogue, inhibited both hCS and mcsA almost equally for both of their respective substrates (Figure 4.15 I - L). Interestingly, competitive inhibition with oxaloacetate (Ki = 8.0 mM) and acetyl-CoA (K_i = 14.6 mM) was observed for hCS, and noncompetitive inhibition with oxaloacetate (K_i = 16.6 mM) and competitive inhibition with propionyl-CoA (K_i = 6.9 mM) for mcsA.

Maleic acid, a substrate analogue of fumarate, had stronger inhibition of hCS than mcsA. It showed mixed inhibition with both oxaloacetate (K_{ia} = 20.0 mM, K_{ib} = 116.4 mM) and acetyl-CoA (K_{ia} = 17.5 mM, K_{ib} = 75.2 mM) for hCS which contrasts to the competitive

inhibition with oxaloacetate (K_i = 30.9 mM) and noncompetitive inhibition with propionyl CoA (K_i = 97.3 mM) for mcsA. The 3,3-thiodipropionic acid showed competitive inhibition for the substrates of both enzymes, but the inhibition of hCS (K_i = ~22 mM) was better than that of mcsA (K_i = ~50 mM) (Figure 4.15 E - H). Lastly, 3-phosphonodipropionic acid was a better inhibitor with acetyl-CoA for hCS than propionyl-CoA for mcsA, but a better inhibitor of oxaloacetate for mcsA than hCS (Figure 4.15 A - D). It had competitive inhibition with oxaloacetate (K_i = 71.7 mM) and acetyl-CoA (K_i = 21.2 mM) for hCS, and mixed inhibition with oxaloacetate (K_{ia} = 34.4 mM, K_{ib} = 193.4 mM and competitive inhibition with propionyl-CoA (K_i = 23.1 mM) for mcsA.

Propionyl-CoA was also used to determine the inhibitor constant with oxaloacetate and acetyl-CoA for hCS. Concentrations of 60 µM, 80 µM and 120 µM propionyl-CoA were used. With these concentrations, mixed inhibition was seen with acetyl-CoA (K_{ia} = ~47 µM, K_{ib} = ~170 µM) (data not shown). The inhibition of propionyl-CoA for oxaloacetate showed almost no inhibition at propionyl-CoA concentrations of 60 µM and 80 µM (data not shown); 120 µM propionyl-CoA displayed weak inhibition with oxaloacetate (data not shown).

4.13 Discussion

A. fumigatus is a problem worldwide in both agriculture and healthcare, and new protein targets are needed. The protein characterized here, mcsA, has had its crystal structure determined along with hCS, an important enzyme in the citric acid cycle, to note any differences in these 51% sequence identical enzymes. Although their sequence identity is high, it is not known how mcsA can utilize both acetyl-CoA and propionyl-CoA as co-

substrates whereas hCS is limited to only acetyl-CoA. Both hCS and mcsA form homodimers with two active sites and share nearly identical active site residues.

Upon binding oxaloacetate, both hCS and mcsA undergo a conformational change from an "open" state to a "partially closed" state in which the binding site for either acetyl-CoA or propionyl-CoA is formed. With both substrates bound, the enzymes form a "closed" state in which the active site becomes inaccessible to solvent as observed by our structures here and current structures for citrate synthase in the PDB. The residues Ala348 in hCS and Gly352 in mcsA were one main difference observed in the active site, and these residues were mutated and had their crystal structure determined with and without substrates. Although the G352A mutation in mcsA did not appear to have a significant impact on substrate binding and conformational change, the A348G mutation in hCS was much more pronounced. This mutation appeared to hinder the "partially closed" conformational state as observed by the crystal structure for hCSA348G with oxaloacetate bound in both active sites (PDB code: 5UZP). Furthermore, investigating individual residue movements from apo-hCS and apo-mcsA to holo-hCS and holo-mcsA structures showed that residues in hCS move a greater distance (based on Cα's) than those of mcsA, which may imply that hCS forms a tighter "closed" conformation that may permit substrate specificity for acetyl-CoA only.

The mechanism of the condensation reaction with acetyl-CoA and propionyl-CoA was not elucidated for mcsA. However, being that the active site of mcsA contains identical residues to that of citrate synthase, it is a good assumption that it would be very similar to the mechanism of citrate synthase [118, 119, 122, 126]. Most likely, either a citryl-CoA

intermediate or 2-methylcitryl-CoA intermediate is formed based on the presence of either acetyl-CoA or propionyl-CoA, respectively.

Kinetic parameters of hCS, hCSA348G, mcsA and mcsAG352A were determined with DTNB and UV-Vis spectroscopy. Both mutants had lower turnover numbers and catalytic efficiency than their respective wild type counterparts. Overall, hCS had a higher k_{cat} by a factor of about 8 and higher catalytic efficiency by about 10 for substrates than mcsA. Furthermore, mcsA displayed positive cooperativity whereas hCS did not display any type of cooperativity.

To screen for potential inhibitors, the high throughput technique DSF was used. Seven compounds were tested for their inhibitory properties: 3-phosphonopropionic acid, 3,3-thiodipropionic acid, potassium hydroxycitrate, dihydroxyfumaric acid, DL-malic acid, maleic acid and trimethylcitrate. Trimethylcitrate showed very poor inhibition of both hCS and mcsA at concentrations of 70 mM. None of these compounds showed an increase in protein stability for hCS based on DSF results, but they were still capable of inhibiting hCS, and in some cases, inhibited hCS better than mcsA (i.e. DL-malic acid and maleic acid). Although DSF allows the rapid screening of hundreds of compounds, it is apparent that the ligands used do not necessarily need to show an increase in protein thermal stability to serve as an inhibitor. Furthermore, the increase in protein thermal stability does not necessarily reflect how potent an inhibitor will be. For example, 3-phosphonopropionic acid showed a T_m increase of 6˚C degrees whereas dihydroxyfumaric acid displayed a 1˚C increase. However, based on inhibitor results, dihydroxyfumaric acid turned out to be a better inhibitor of mcsA than 3-phosphonopropionic acid.

All the inhibitors screened here showed either competitive, noncompetitive or mixed inhibition with respect to hCS and mcsA substrates. In most cases, hCS and mcsA were not subject to the same type of inhibition by the same inhibitor. For example, hCS displayed mixed inhibition with maleic acid for both oxaloacetate and acetyl-CoA, whereas competitive and noncompetitive inhibition was observed for oxaloacetate and propionyl-CoA, respectively, for mcsA. The best inhibitor of mcsA was dihydroxyfumaric acid which is an analogue of fumarate. This compound inhibited hCS poorly in comparison to mcsA. Based on this result, it would be logical to try derivatives of compounds found in the 2-methylcitrate cycle and the citric acid cycle to determine their inhibitor efficacy of mcsA and hCS. Currently, it is unclear why different types of inhibition are observed between hCS and mcsA for the same ligand, but it may be attributed to the tightness of the clamping motion from "open" to "closed" conformations of these enzymes. Co-crystallization of these inhibitors with hCS and mcsA may provide further insight as to why different types of inhibition are seen and potentially the development of mcsA-specific inhibitors based on these differences.

4.14 Site-directed Mutagenesis of hCSA348G and mcsAG352A

Primers used for mutagenesis are listed in Table 3.1. The hCSA348G mutation was made using complimentary primers and the thermalcycling protocol for Phusion polymerase was followed (NEB, Ipswich, MA). Briefly, a 4 minute extension time was used for 25 cycles with 50 ng of template. For mcsAG352A, non-overlapping primers were designed with NEBquickchange using Q5 polymerase with the mutation in the forward primer (NEB, Ipswich, MA). Reaction conditions and thermalcycling used were based on the NEB Q5 polymerase protocol. Briefly, a 2 minute 45 second extension time was used

for 25 cycles with 1.0 ng of template. After hCS and mcsA PCR reactions, DpnI was added per the manufacturer's protocol (NEB, Ipswich, MA) and incubated at 37˚C for 2 hours. Post-DpnI digest, reaction mixtures were transformed into chemically competent DH5-α *E. coli* cells and spread onto LB-Agar plates containing the appropriate antibiotic.

Plates were incubated at 37˚C overnight. Clones were inoculated into 5.0 mL LB containing the appropriate antibiotic and grown at 37˚C overnight (about 16 hours) with shaking. Minipreps were performed with a ThermoFisher GeneJET Plasmid Miniprep Kit (Grand Island, NY) following the manufacturer's protocol and eluted with 50 µL of nuclease-free water. Clones were sent to EtonBioscience (Research Triangle Park, NC) for sequencing with T7 forward and T7 reverse primers. When the point mutations were confirmed, the correct mutant plasmids were transformed into BL-21 (DE3) pLysS *E. coli* cells for hCSA348G and BL-21 (DE3) *E. coli* cells for mcsAG352A via heat shock at 42˚C for 45 seconds. Mutants were purified in the same manner as wild type.

4.15 TABLES

Table 4.1: Primers used for hCS and mcsA cloning and mutagenesis.

Primer	Sequence
pMCSG53-hCS-F	5' TACTTCCAATCCAATGCCTCCTCCACGAATTTGAAAGACATATTGG 3’
pMCSG53-hCS-R	5’ TTATCCACTTCCAATGTTACCCTGACTTAGAGTCCACAAACTTCAT 3’
pMCSG53-F	5’ATTGGAAGTGGATAACGGATCCGAATTCGA 3'
pMCSG53-R	5’ ATTGGATTGGAAGTACAGGTTCTCGGTAC 3'
hCSA348G-F	5’ CCAGGCTATGGCCATGGAGTACTAAGGAAGACT 3’
hCSA348G-R	5’ AGTCTTCCTTAGTACTCCATGGCCATAGCCTGG 3’
mcsAG352A-F	5’ TACGGCCACGCTGTTCTGCGCAAAC 3’
mcsAG352A-R	5’ ACCCGGCACAACGCGACC 3’

Table 4.2: Data collection statistics for hCS and mcsA crystal structures. Numbers in parentheses represent statistics from the highest resolution shell. Abbreviations: AU – Asymmetric Unit.

Protein	mcsA	mcsA	mcsA	mcsAG352A	pCS	hCS Dimer	hCS monomer	hCSA348G
PDB Accession code	5UQO	5UQQ	5UQR	5UQU	5UQ5	5UZR	5UZQ	5UZP
Data collection								
a, *b*, *c* (Å)	60, 116, 153	82., 130, 261	69., 94, 124	70, 94, 123	58, 59, 74	58, 74, 60	76, 76, 198	58, 111, 74
α, β, γ (degrees)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	99, 98, 117	62, 93, 80	90, 90, 90	90, 99, 90
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	P1	P1	$P42_12$	$P12_11$
Solvent content (%)	56.25	49.24	41.93	42.38	45.73	47.29	58.98	49.60
Protein chains in AU	2	6	2	2	2	2	1	2
Resolution range (Å)	50.01-2.50	31.41-2.30	40.00-1.75	50.01-1.70	50.01-1.60	40.00-2.30	50.01-2.16	45.06-2.29
Highest resolution shell (Å)	2.54-2.50	2.33-2.30	1.78-1.75	1.73-1.70	1.63-1.60	2.34-2.30	2.19-2.15	2.34-2.30
Unique reflections	37351 (1814)	125981 (4770)	82446 (3801)	85824 (4534)	109748 (5428)	37732 (1924)	30789 (1500)	42042 (2123)
Redundancy	4.9 (4.7)	4.9 (4.8)	5.7 (3.9)	7.8 (7.9)	2.2 (2.2)	2.1 (2.1)	8.5 (8.5)	3.7 (3.7)
Completeness (%)	94.8 (92.4)	99.7 (99.7)	98.0 (91.6)	94.2 (100.0)	96.9 (94.9)	96.5 (97.7)	94.1 (93.6)	100 (100)
R_{merge}	0.066 (0.590)	0.081 (0.528)	0.078 (0.529)	0.062 (0.799)	0.061 (0.344)	0.078 (0.375)	0.068 (0.323)	0.123 (0.628)
R_{pim}	0.044 (0.318)	0.052 (0.354)	0.041 (0.303)	0.029 (0.278)	0.046 (0.359)	0.083 (0.372)	0.029 (0.148)	0.082 (0.403)
R_{rim}	0.104 (0.724)	0.117 (0.777)	0.104 (0.634)	0.081 (0.785)	0.068 (0.530)	0.123 (0.548)	0.094 (0.483)	0.159 (0.780)
CC half	0.806	0.809	0.722	0.857	0.733	0.776	0.968	0.750
Average $I/\sigma(I)$	21.1 (2.04)	18.4 (2.35)	16.4 (2.33)	35.2 (3.28)	19.95 (2.17)	10.3 (2.18)	25.87 (3.29)	11.57 (2.5)
Refinement								
R_{work} (%)	19.1 (30.3)	18.5 (25.8)	13.48 (20.9)	15.84 (20.8)	15.25 (24.0)	17.02 (22.8)	19.60 (27.4)	15.59 (21.3)
R_{free} (%)	23.3 (36.6)	21.4 (28.1)	17.48 (27.0)	19.21 (24.3)	17.84 (24.8)	21.45 (30.8)	23.23 (33.6)	20.55 (29.1)
Mean *B* value (Å^2)	57.96	42.98	16.92	27.60	20.65	26.66	53.62	28.05
B from Wilson plot (Å^2)	56.90	42.20	17.90	27.70	21.29	26.32	52.37	27.63
RMSD bond lengths (Å)	0.012	0.012	0.018	0.015	0.013	0.013	0.011	0.014
RMSD bond angles (degrees)	1.6	1.5	1.8	1.7	1.6	1.6	1.5	1.6
No. of amino acid residues	A, B = 434	A,F = 435; B = 436; C = 437; D,E = 434	A = 439; B = 435	A = 438; B = 436	A,C = 434	A = 434; D = 435	A = 434	A,B = 436
No. of water molecules	100	1449	1288	749	1058	400	159	576
Ramachandran plot								
Most favored regions (%)	98.30	98	98.9	98.5	99	98.3	98.4	99
Additional allowed regions (%)	100.00	99.8	100	100	100	100	100	100

Table 4.3: Kinetic parameters of hCS, hCSA348G, mcsA and mcsAG352A determined by UV-Vis spectroscopy. Abbreviations: OAA – Oxaloacetate, A-CoA – Acetyl-CoA, P-CoA – Propionyl-CoA.

Kinetic Parameter	hCS			hCSA348G			mcsA			mcsAG352A		
	OAA	A-CoA	P-CoA	OAA	A-CoA	P-CoA	OAA	A-CoA	P-CoA	OAA	A-CoA	P-CoA
K_m (µM)	3.6 ± 0.2	7.7 ± 0.6	n/a	5.7 ± 0.5	10.4 ± 1.5	n/a	4.9 ± 0.2	5.0 ± 1.3	4.7 ± 0.3	3.5 ± 0.3	3.6 ± 0.3	3.9 ± 0.1
V_{max} (µM/sec)	0.52 ± 0.01	0.65 ± 0.02	n/a	0.68 ± 0.02	0.7 ± 0.1	n/a	0.66 ± 0.02	0.7 ± 0.1	0.68 ± 0.03	0.57 ± 0.03	0.53 ± 0.04	0.62 ± 0.01
Hill Coefficient (n)	0.9 ± 0.1	0.9 ± 0.1	n/a	1.0 ± 0.1	0.9 ± 0.1	n/a	1.5 ± 0.1	1.6 ± 0.3	1.5 ± 0.1	1.8 ± 0.3	2.9 ± 0.5	1.7 ± 0.1
k_{cat} (sec^{-1})	208	260	n/a	111	117	n/a	29.2	32.7	30.1	11.2	10.4	12.2
K_{cat}/K_m ($sec^{-1}M^{-1}$)	5.9×10^7	3.4×10^7	n/a	1.9×10^7	1.1×10^7	n/a	6.0×10^6	6.6×10^6	6.5×10^6	3.2×10^6	2.9×10^6	3.2×10^6

Table 4.4: Differential Scanning Fluorimetry (DSF) results using substrates as ligands. All substrates and products increased the thermal stability of hCS, hCSA348G, mcsA and mcsAG352A.

Protein	Increase in Melting Temperature T_m (˚C)						
	Oxaloacetate	**Acetyl-CoA**	**Propionyl-CoA**	**Citrate**	**2-Methylcitrate**	**CoA-SH**	**Oxaloacetate + CoA-SH**
hCS	22	8	8	7	5	10	26
hCSA348G	11	4	5	1	1	5	13
mcsA	12	8	8	4	4	8	14
mcsAG352A	10	2	4	2	2	4	12

Table 4.5: Summary of calculated inhibitor constants and inhibition types for hCS and mcsA.

Inhibitor	hCS		mcsA	
	Oxaloacetate K_i (mM)	Acetyl-CoA K_i (mM)	Oxaloacetate K_i (mM)	Propionyl-CoA K_i (mM)
3-phosphonopropionic acid	71.7 ± 20.0	21.2 ± 0.6	34.4 ± 1.1, 193.4 ± 57.1	23.1 ± 2.1
	Competitive	Competitive	Mixed	Competitive
3,3-thiodipropionic acid	24.2 ± 2.9	22.3 ± 1.9	52.1 ± 19.0	45.1 ± 5.6
	Competitive	Competitive	Competitive	Competitive
Potassium Hydroxycitrate	8.0 ± 1.3	14.6 ± 4.9	16.6 ± 1.1	6.9 ± 1.2
	Competitive	Competitive	Noncompetitive	Competitive
Dihydroxyfumaric Acid	90.5 ± 29.7	73.0 ± 26.2	23.6 ± 9.6, 90.6 ± 36.6	17.1 ± 6.2
	Competitive	Competitive	Mixed	Competitive
Malic Acid	14.2 ± 0.7, 169.3 ± 46.1	6.5 ± 1.6, 19.3 ± 7.3	67.5 ± 17.7	62.1 ± 11.9
	Mixed	Mixed	Noncompetitive	Noncompetitive
Maleic Acid	20.0 ± 5.6, 116.4 ± 19.5	17.5 ± 5.2, 75.2 ± 3.3	30.9 ± 4.9	97.3 ± 13.2
	Mixed	Mixed	Competitive	Noncompetitive

4.16 FIGURES

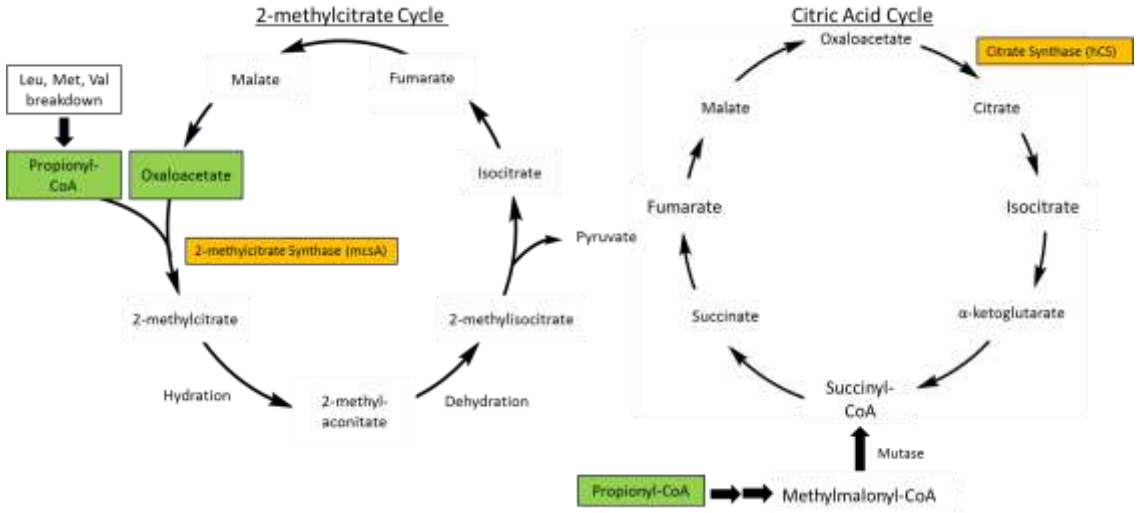

Figure 4.1: The 2-methylcitrate Cycle and Citric Acid Cycle. The 2-methylcitrate cycle (left) is used by fungi to detoxify propionyl-CoA via the enzyme 2-methylcitrate synthase (mcsA). In humans, propionyl-CoA is converted to methylmalonyl-CoA (right) and fed into the citric acid cycle. McsA also can perform the same reaction as citrate synthase (hCS) in this cycle (both colored in orange).

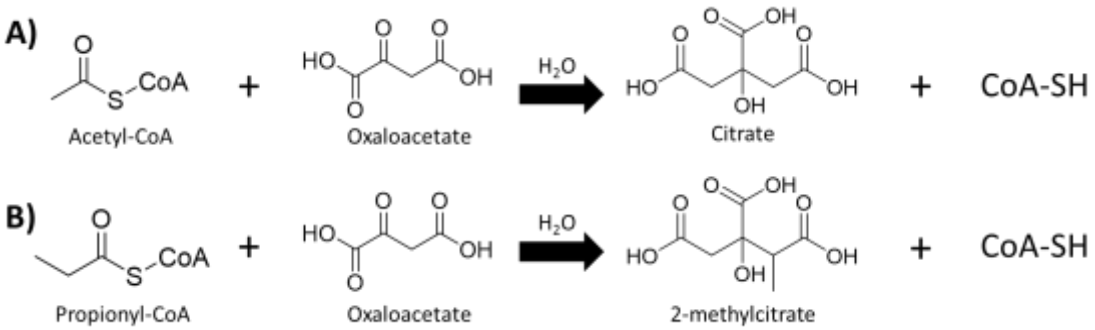

Figure 4.2: Condensation reactions catalyzed by mcsA and hCS. **A)** Condensation reactions catalyzed by both hCS and mcsA. **B)** Reaction catalyzed by only mcsA. hCS cannot use propionyl-CoA as co-substrate.

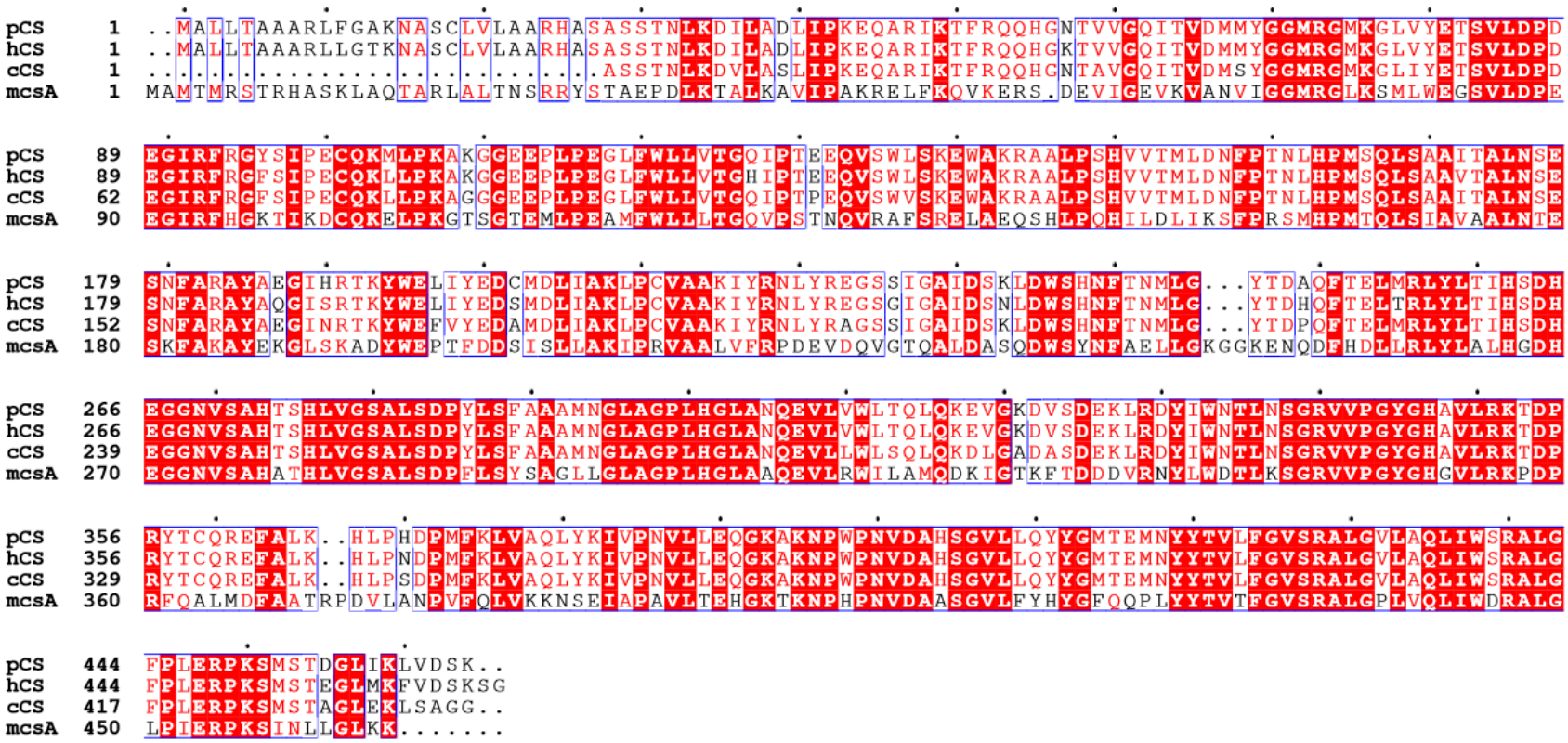

Figure 4.3: Sequence alignment of pCS, hCS, cCS and mcsA. Enzymes pCS, hCS and cCS all share over 91% sequence identity. HCS and mcsA share a 51% sequence identity. Conserved residues are highlighted in red. Conserved secondary structure is shown by blue boxes. Sequence alignment was performed with ESPript 3.0 [82]

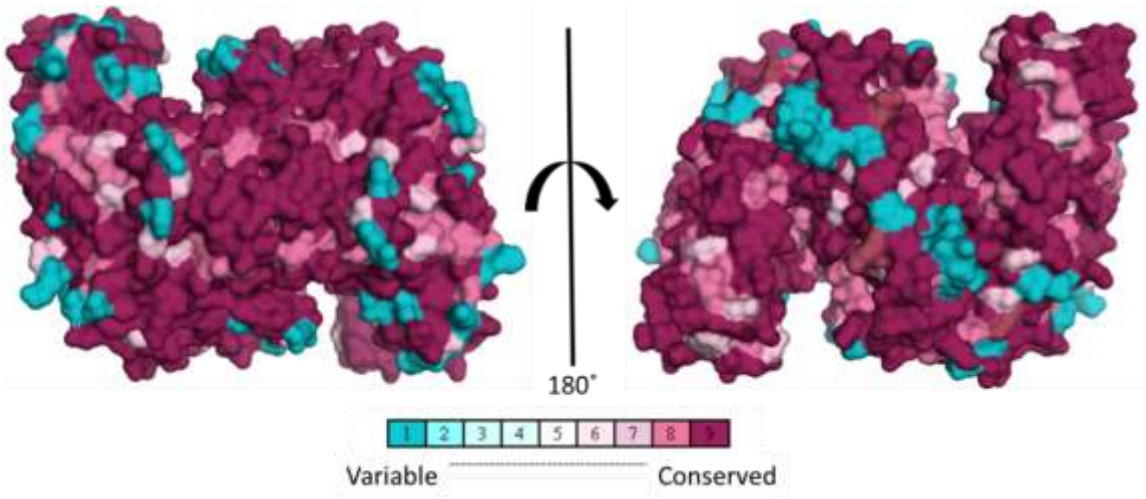

Figure 4.4: Secondary sequence conservation of pCS, hCS and cCS. Conserved residues are shown in dark purple and variable residues are shown in teal between these citrate synthases; figure was made using Consurf [127].

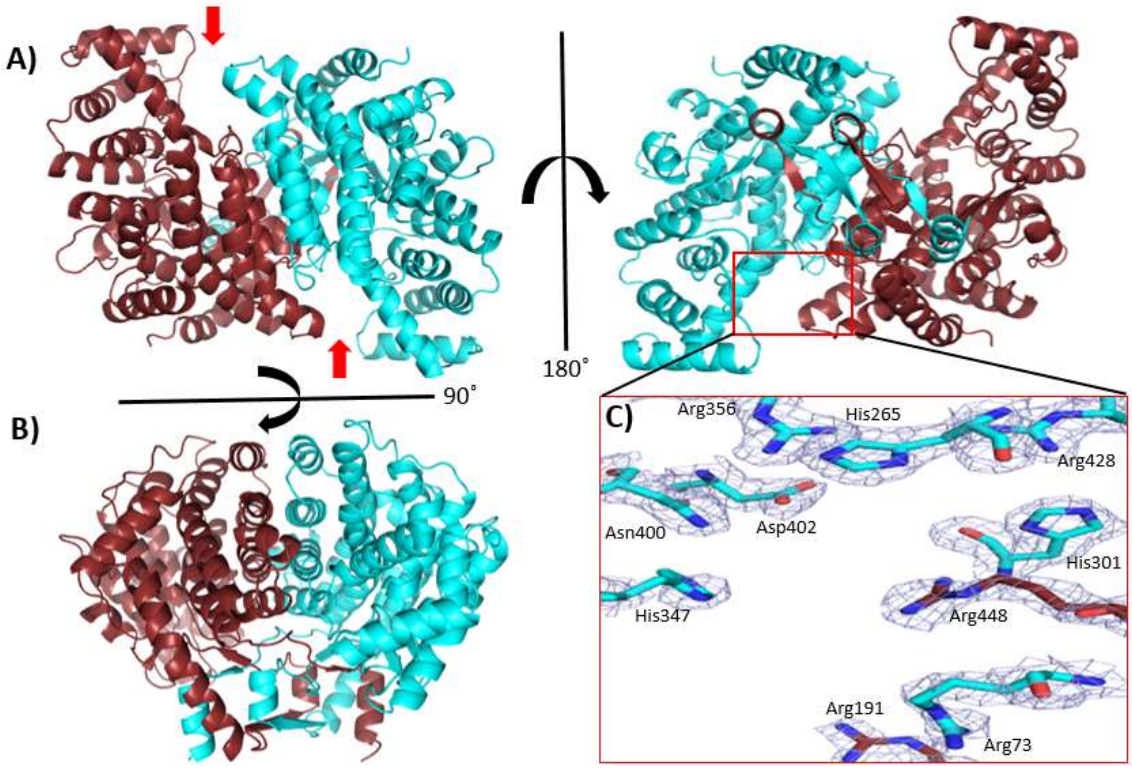

Figure 4.5: Crystal structure of hCS. A) Crystal structure of hCS (PDB code: 5UZR). Citrate synthases are homodimeric and form two active sites. Red arrows indicate the active sites. Each monomer is colored in cyan or brown and both contribute residues to each active site. A 180˚ rotation about the Y-Axis is shown. **B)** HCS rotation of 90˚ about the X-axis to show the dimer interface between monomers. The dimer interface area is about 6000 $Å^2$. **C)** Zoomed in view of one active site of hCS. Several basic residues contribute to substrate binding. Residues from chain A are shown in cyan and residues from chain B in brown.

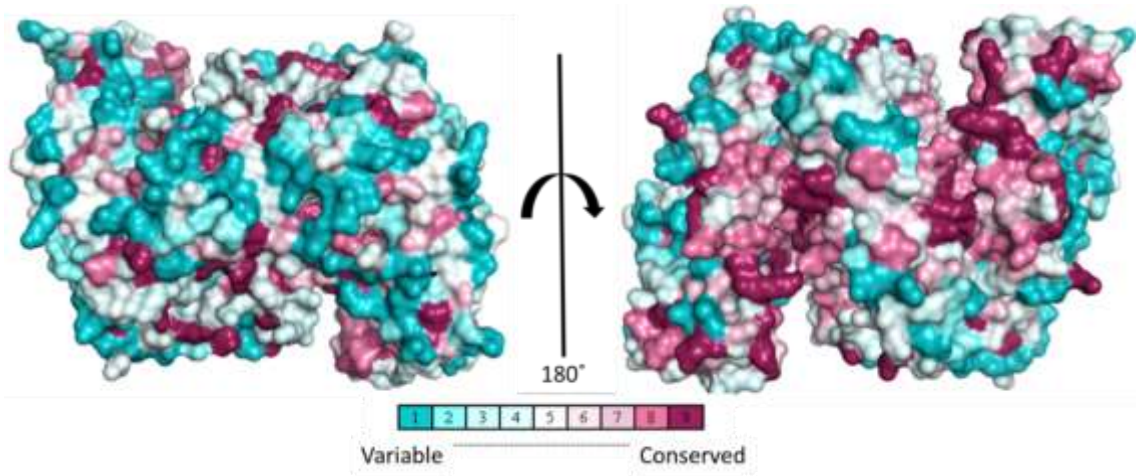

Figure 4.6: Secondary sequence conservation of hCS and mcsA. Highly conserved residues are shown in purple and variable residues shown in teal. Figure was made using Consurf [127].

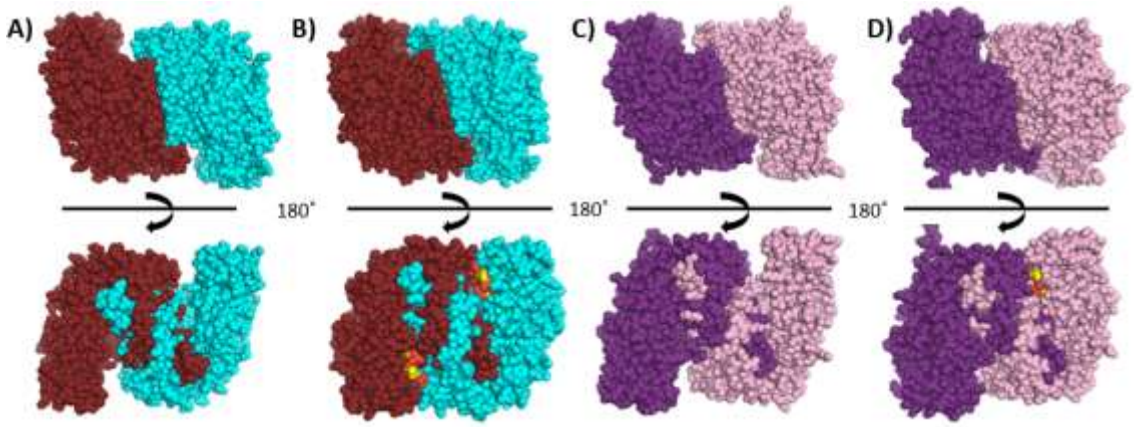

Figure 4.7: Comparison of conformational changes between hCS and mcsA. Both hCS and mcsA adopt "open," "partially closed" or "closed" conformations upon binding substrate. All structures are shown as space-filling models with each monomer of the homodimer colored differently. Substrates are colored with red and yellow spheres. **A)** Apo-hCS (PDB code: 5UZR) "open" conformation. **B)** Holo-cCS (PDB code: 4CSC - homologue) in "closed" conformation with D-malate and acetyl-CoA bound in the active sites. **C)** Apo-mcsA (PDB code: 5UQO) in "open" conformation. **D)** Holo-mcsA (PDB code: 5UQR) with oxaloacetate bound in one active site ("partially closed" conformation) and oxaloacetate and ethyl-CoA bound in the other ("closed" conformation).

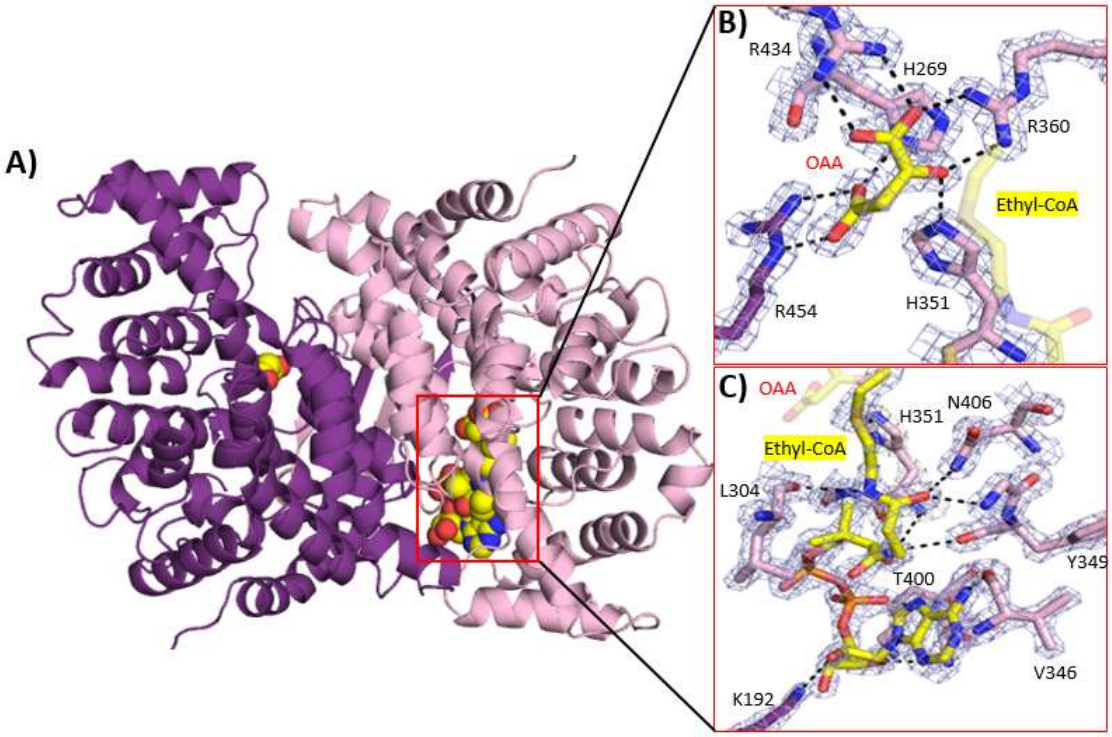

Figure 4.8: Crystal structure of mcsA with oxaloacetate and ethyl-CoA bound. **A)** McsA (PDB code: 5UQR) was co-crystallized with oxaloacetate (OAA) and ethyl-CoA. Oxaloacetate was present in both active sites whereas ethyl-CoA was only in one of them. **B)** Zoomed in view of the oxaloacetate binding. Residues from both chains involved in binding oxaloacetate are shown. **C)** Zoomed in view of ethyl-CoA binding. Several peptide bonds form salt bridges with the adenosine moiety of ethyl-CoA. Only K192 from the other chain seems to be involved in binding ethyl-CoA.

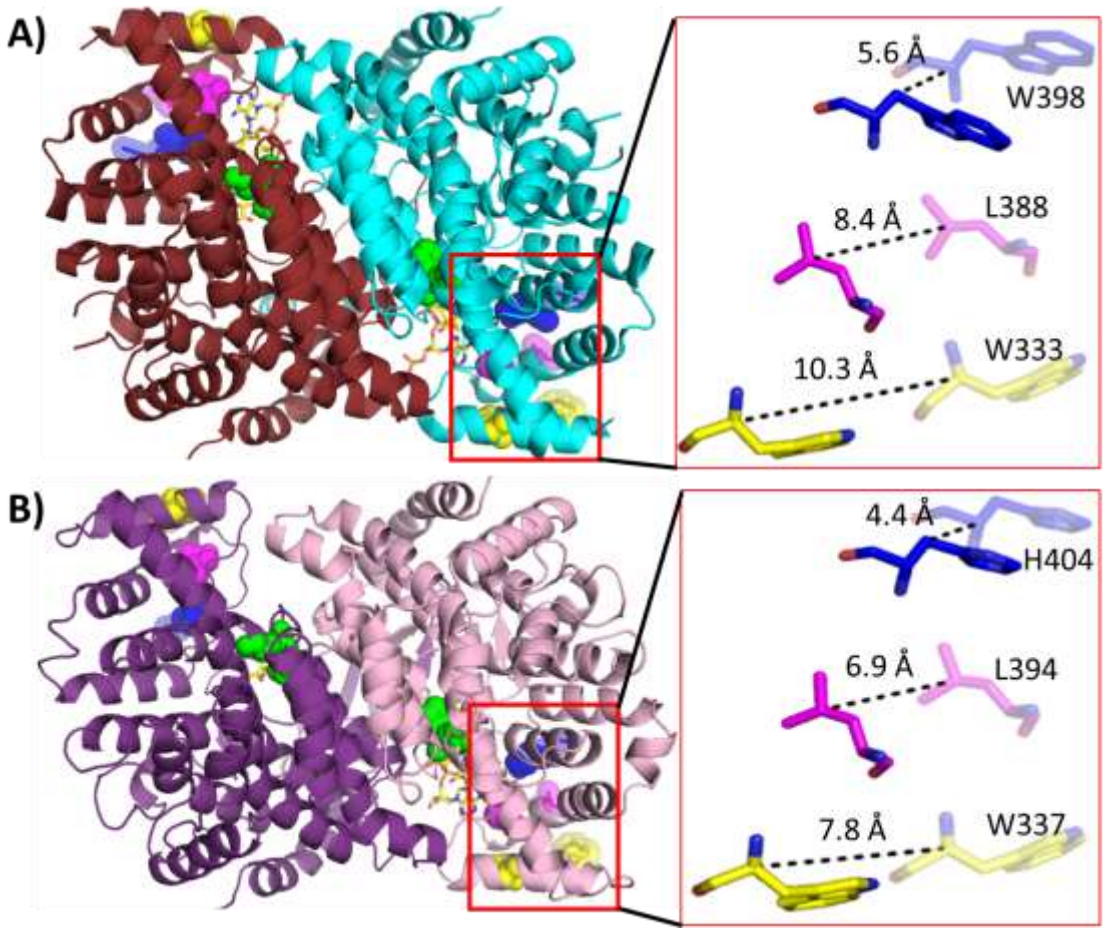

Figure 4.9: Residue movement comparison upon substrate binding by hCS and mcsA. Cartoon representations are shown for holo-cCS (PDB code: 4CSC – homologue) and holo-mcsA (PDB code: 5UQR). Overall residue movement in each chain between hCS and mcsA was compared. Comparison of only three residues are shown as spheres: W398 (hCS) and H404 (mcsA) in blue, L388 (hCS) and L394 (mcsA) in magenta, W333 (hCS) and W337 (mcsA) in yellow. Apo-enzyme residue positions are shown with transparent spheres and stick representation. The hinge region for each active site is shown by green spheres. **A)** Residue movement observed in apo-hCS (PDB code: 5UZR) and holo-cCS (PDB code: 4CSC – homologue). **B)** Residue movement observed in apo-mcsA (5UQO) and holo-mcsA (PDB code: 5UQR). Overall, the residues in hCS showed greater movement than the corresponding residues in mcsA which could imply a tighter "closed" conformation. Moreover, this may help explain substrate specificity between hCS and mcsA.

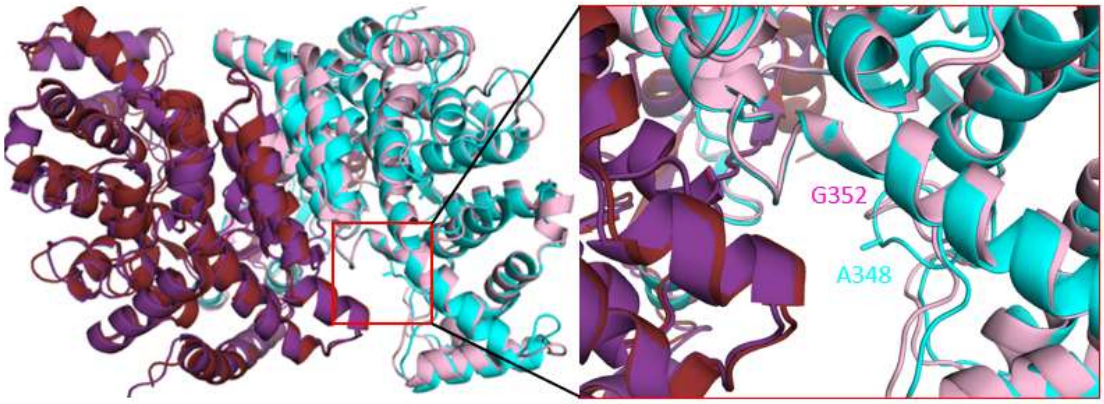

Figure 4.10: Ala348 in hCS and Gly352 in mcsA. Pictured are aligned cartoon representations of apo-hCS (PDB code: 5UZR) and apo-mcsA (PDB code: 5UQO). Apo-hCS is colored in cyan and brown. Apo-mcsA is colored in pink and purple. One minor difference in the active sites are A348 in hCS and G352 in mcsA. The alanine in hCS was thought to possibly cause some steric clash with CoA substrate and was investigated further.

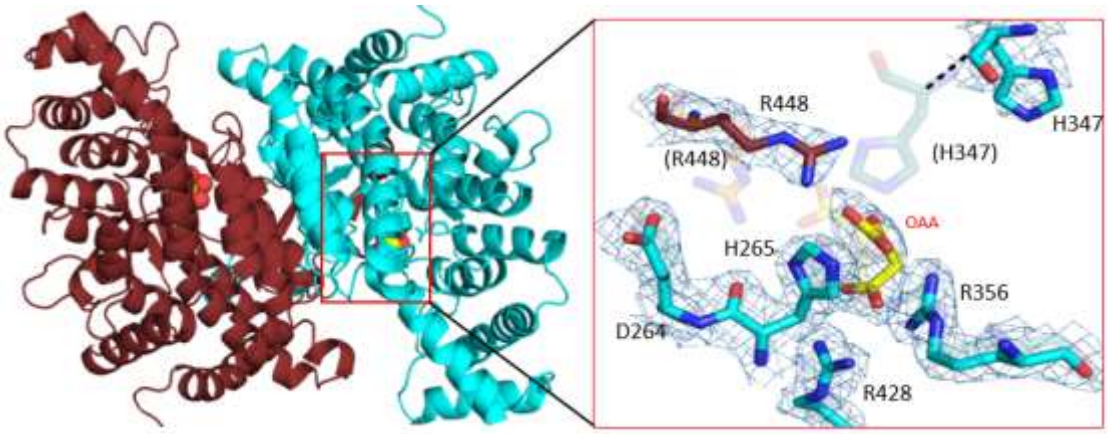

Figure 4.11: Crystal structure of hCSA348G. The hCSA348G crystal structure (5UZP) was aligned with pCS (PDB code: 4CTS – homologue with only OAA bound) to show difference in positioning of active site residues. OAA is bound in both active sites (red and yellow spheres) for both structures. Most of the active site residues are in their normal positions when OAA is bound except H347 and R448. In the stick representations, 4CTS residues H347, R448 and OAA are shown with transparency and parentheses. Furthermore, the A348G mutation seemed to hinder the formation of the acetyl-CoA binding site by preventing the "partially closed" conformation.

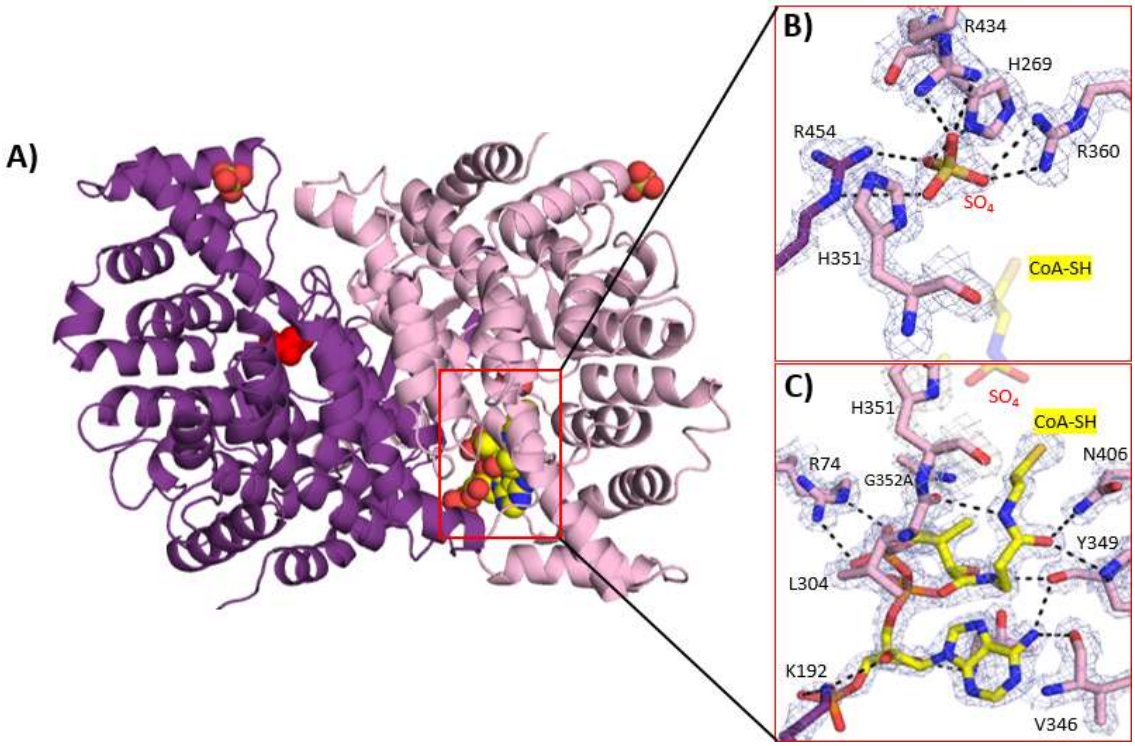

Figure 4.12: Crystal structure of mcsAG352A. A) mcsAG352A (PDB code: 5UQU) was co-crystallized with oxaloacetate and CoA-SH. Oxaloacetate (shown as red spheres) was bound in one active site, and the other contained a sulfate ion (SO_4^{2-}) and CoA-SH. The sulfate came from the crystallization condition. Two other sulfate ions were also bound on the surface of the protein (shown as red and yellow spheres). **B)** Active site coordination of the sulfate ion. **C)** Active site coordination of CoA-SH.

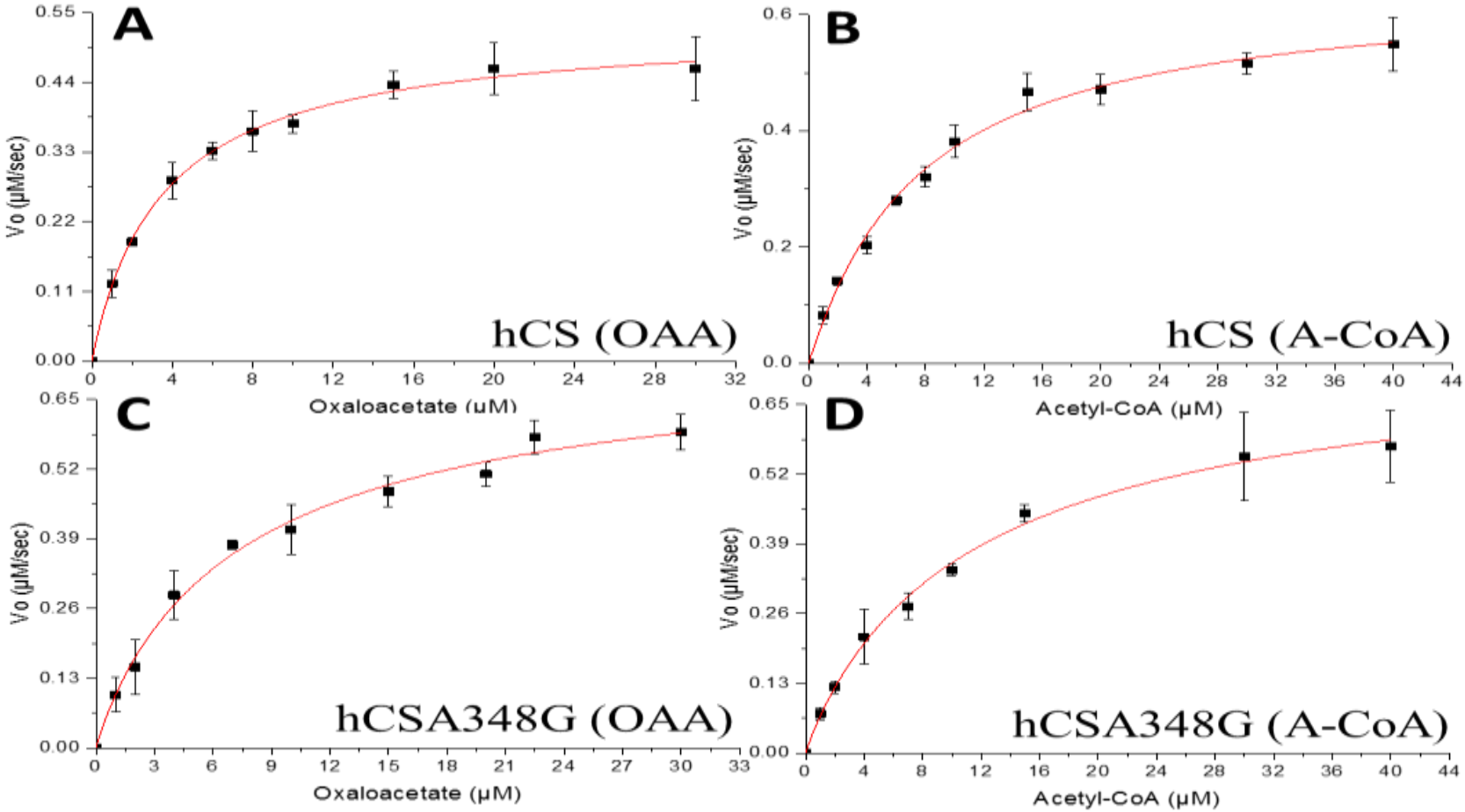

Figure 4.13: Kinetic graphs of hCS, hCSA348G, mcsA and mcsAG352A (A-D).

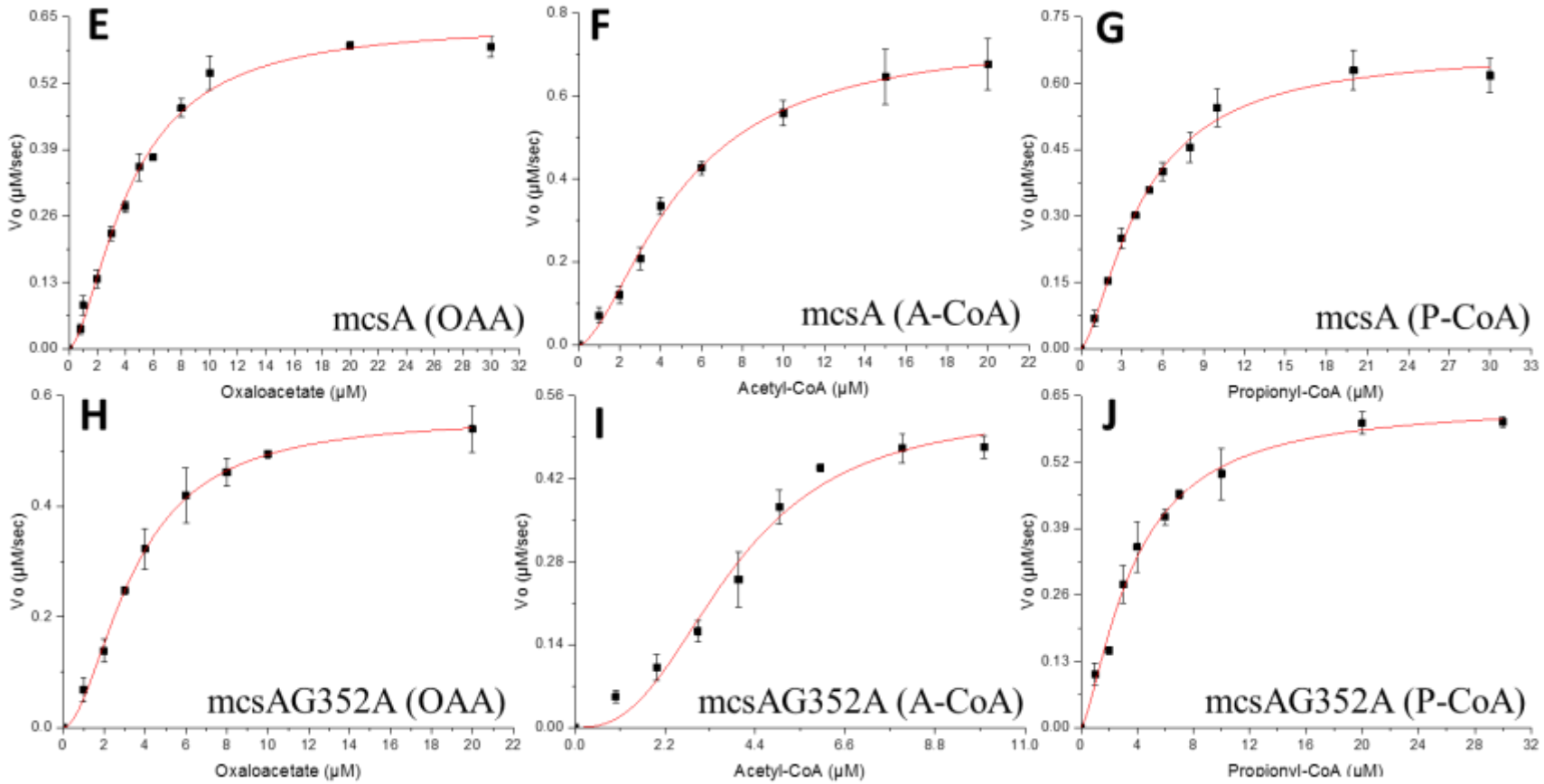

Figure 4.13: Kinetic graphs of hCS, hCSA348G, mcsA and mcsAG352A (E-J). For hCS fitting, the Michaelis-Menten equation was used. For mcsA fitting, the Hill equation was used (displayed positive cooperativity). All x-axes (substrate concentration) are in micromolar (µM) and y-axes (initial velocity) are in micromolar per second (µM/sec). **A, C, E, H)** Varied oxaloacetate (OAA). **B, D, F, I)** Varied acetyl-CoA (A-CoA). **G, J)** Varied propionyl-CoA (P-CoA). Propionyl-CoA could not be used as co-substrate for hCS or hCSA348G. Abbreviations: OAA – Oxaloacetate, A-CoA – Acetyl-CoA, P-CoA – Propionyl-CoA.

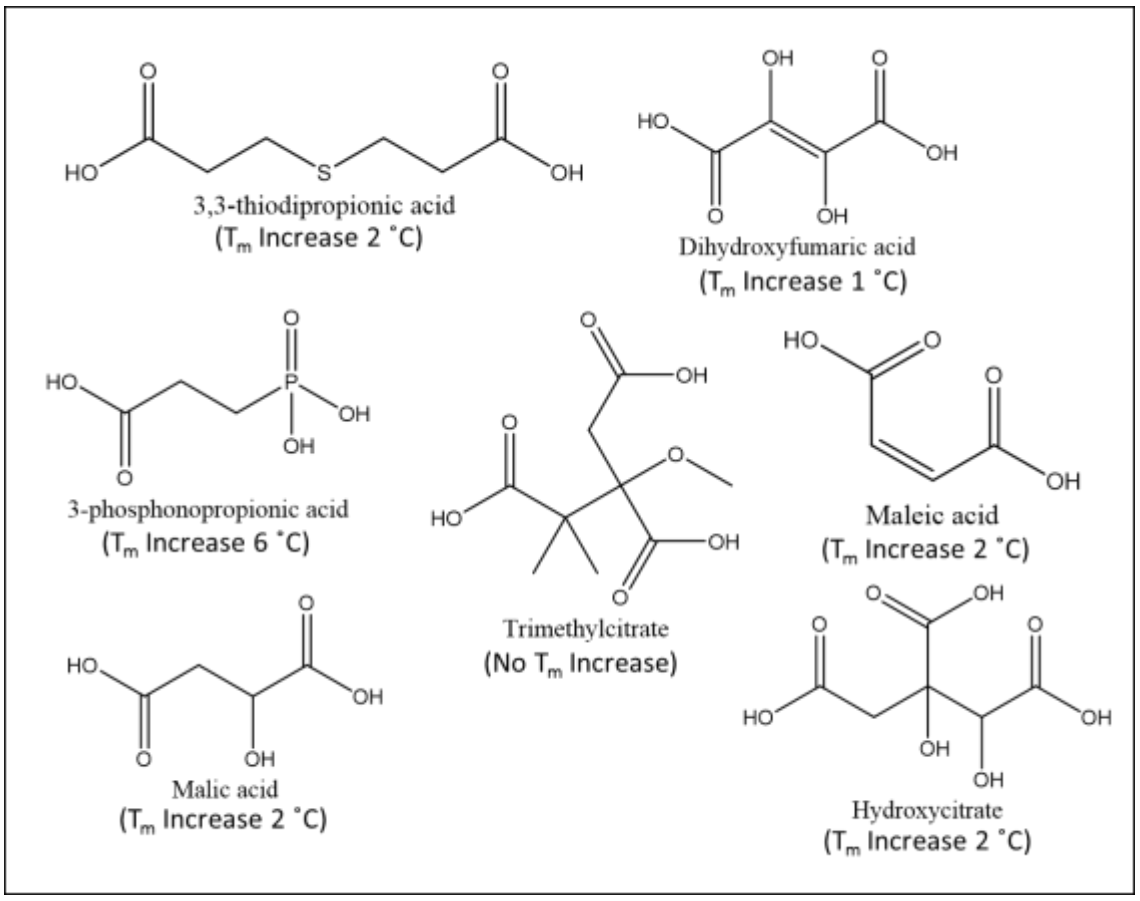

Figure 4.14: Ligands used for DSF. All ligands pictured, except trimethylcitrate, increased the thermal stability of mcsA and did not increase the thermal stability of hCS.

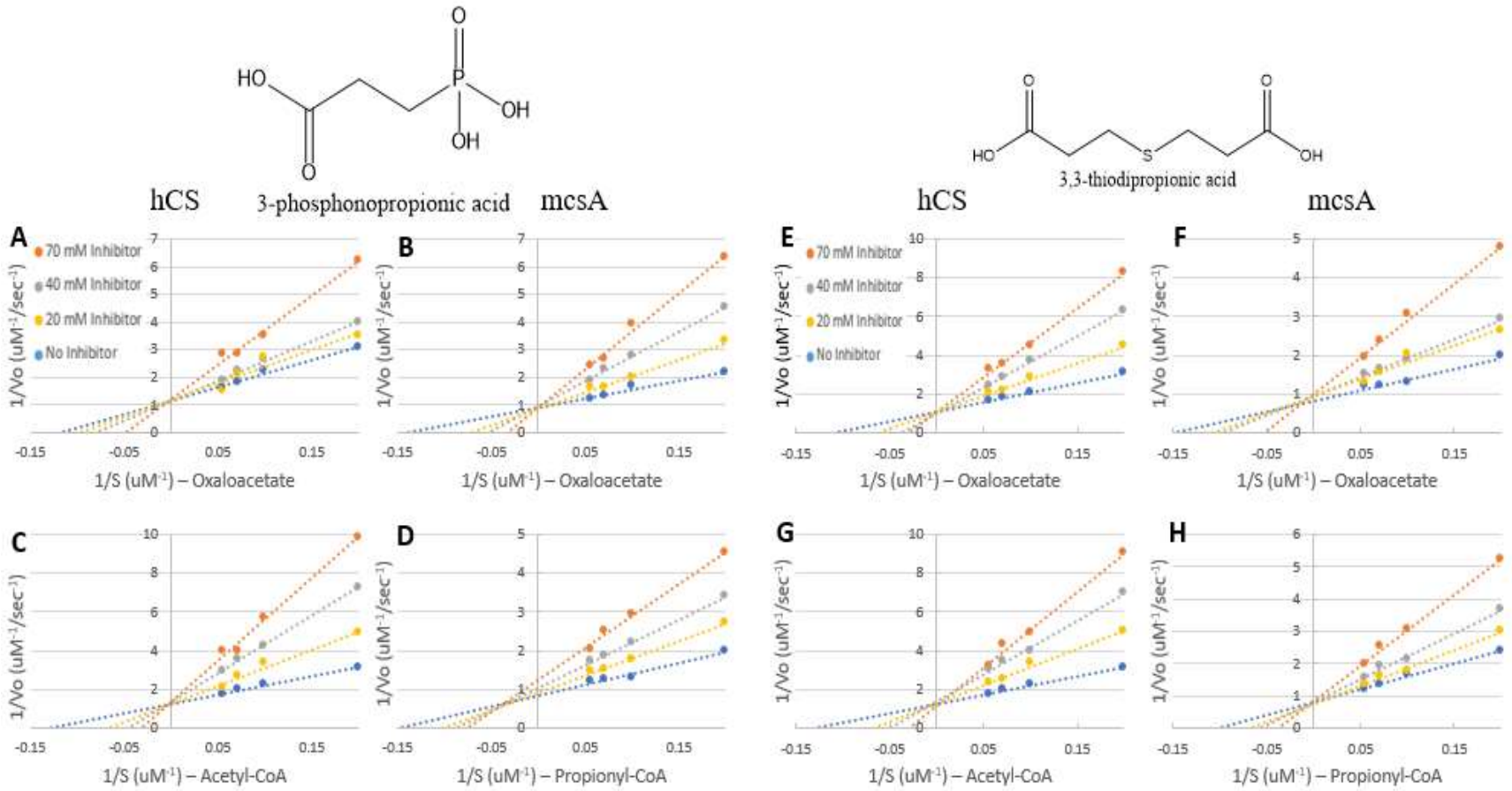

Figure 4.15: Lineweaver-Burk plots and inhibitor kinetics (A-H).

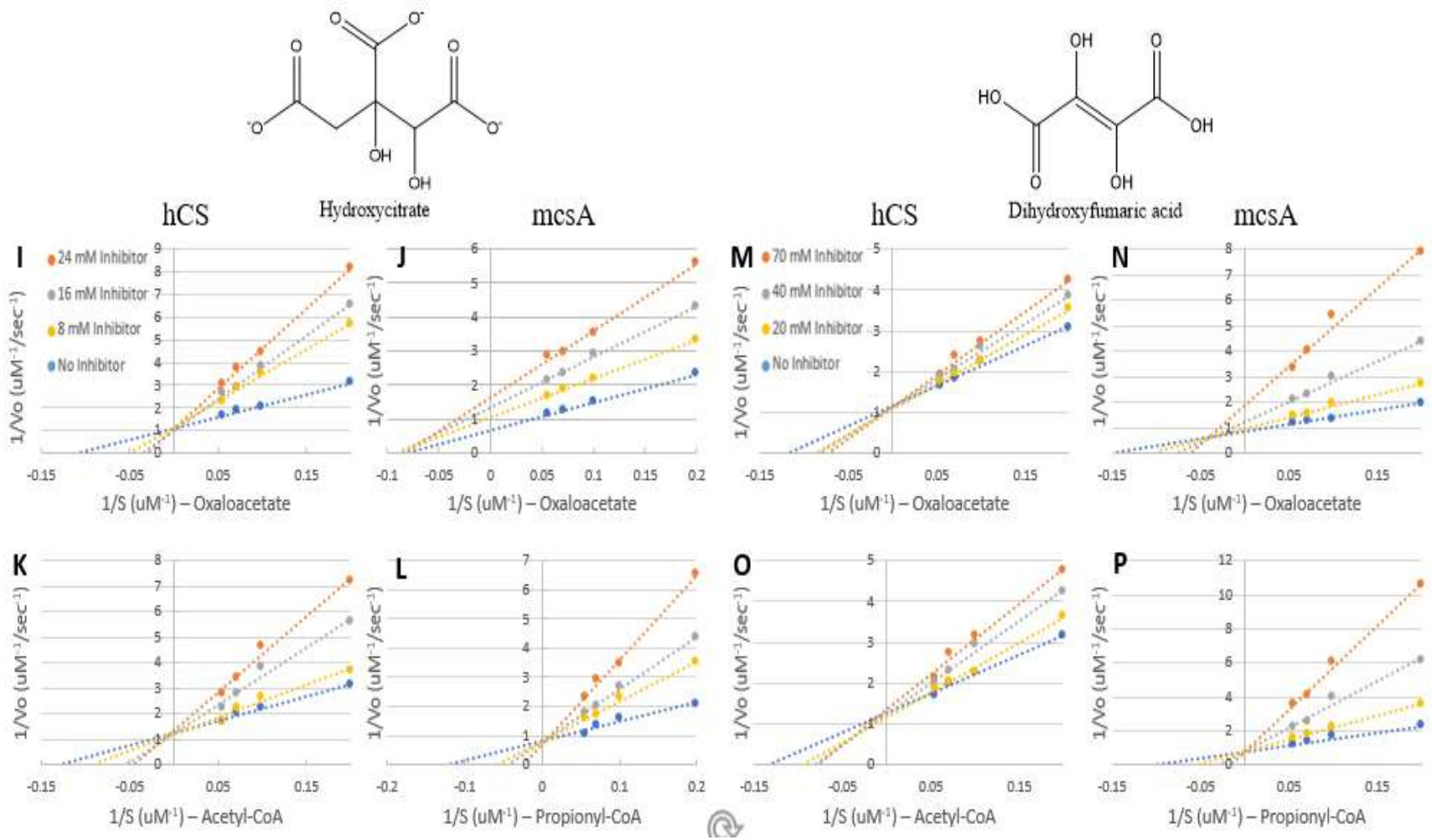

Figure 4.15: Lineweaver-Burk plots and inhibitor kinetics (I-P).

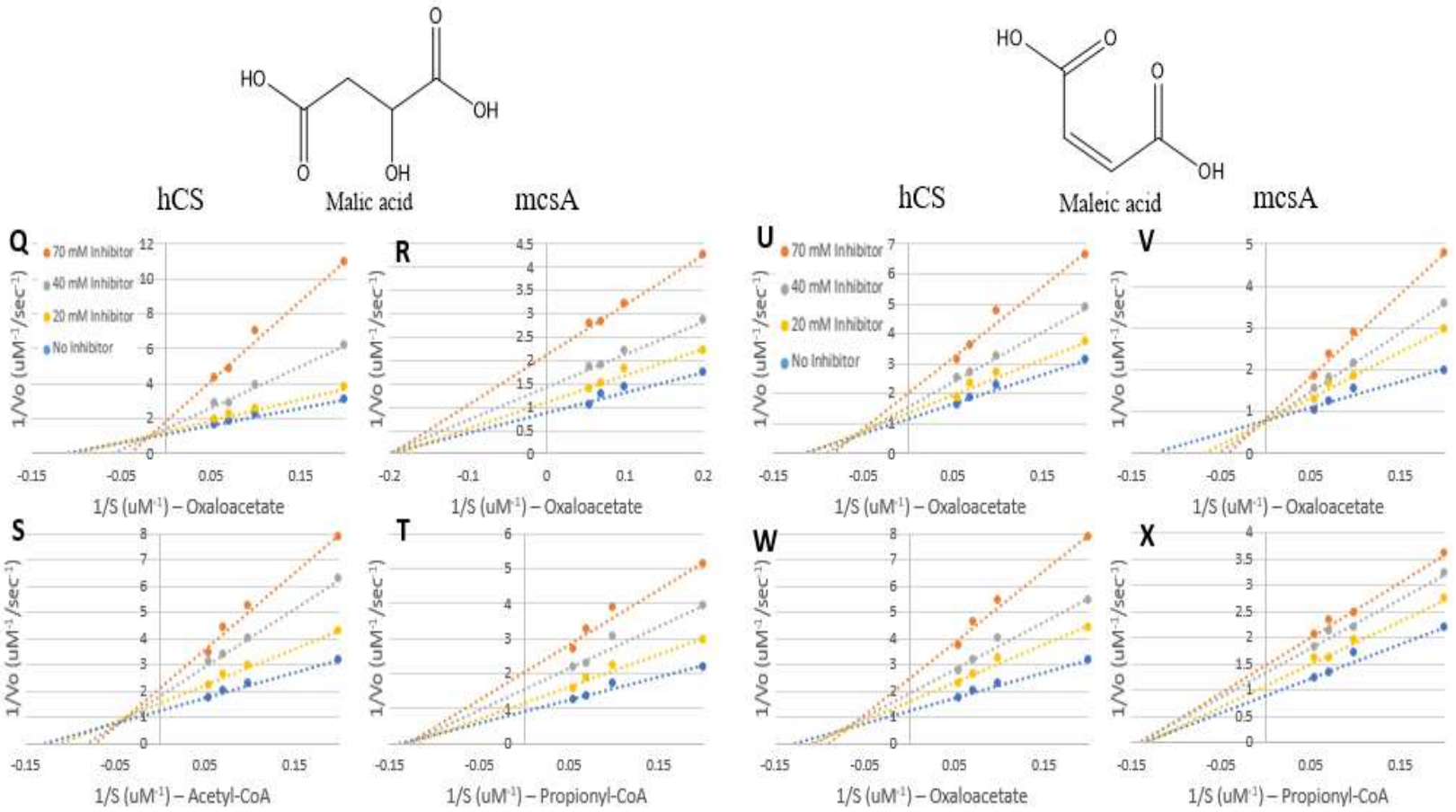

Figure 4.15: Lineweaver-Burk plots and inhibitor kinetics (Q-X). All inhibitors (except hydroxycitrate) were tested with concentrations of 70 mM (orange), 40 mM (grey), 20 mM (yellow) and 0 mM inhibitor (blue, no inhibitor). Hydroxycitrate was tested with concentrations of 24 mM (orange), 16 mM (grey), 8.0 mM (yellow) and 0 mM inhibitor (blue, no inhibitor). Trimethylcitrate showed poor inhibition for both enzymes and is not shown. **A-D)** 3-phosphonopropionic acid. **E-H)** 3,3-thiodipropionic acid. **I-L)** Hydroxycitrate. **M-P)** Dihydroxyfumaric acid. **Q-T)** Malic acid. **U-X)** Maleic acid.

CHAPTER 5

SUMMARY, CONCLUSIONS AND FUTURE DIRECTIONS

The work presented here focuses on proteins from two agricultural pests, *Tetranychus urticae* and *Aspergillus fumigatus*, and characterizing these proteins may set the foundation for the discovery of new pesticides. Chapter 1 focused on TuCyanase, a protein important in the detoxification of the toxic metabolite produced by plants, cyanate. TuCyanase was crystallized and had its structure determined then compared to the bacterial cyanases. Future work with this project would include testing different cyanase inhibitors, and possibly trying to express and crystallize TuCyanase without fusion partner MBP. This may also provide insight as to the purpose of the N-terminal domain of TuCyanase, and how it is related to the bacterial cyanases and what purpose it may serve.

Co-crystallization and soaking experiments with potential ligands may help explain how substrates enter and exit the active sites. The cellular localization of TuCyanase is not known, but could be determined using fluorescent probes or markers such as green fluorescent protein (GFP). Knowing the localization TuCyanase might explain the purpose of the negatively charged core of the decamer, and the significance of the TuCyanase decamer being less compact than the bacterial cyanases mentioned here. Furthermore, studying the interaction between *T. urticae* carbonic anhydrases and TuCyanase may display how important both of these enzymes are for the survival of *T. urticae* and potentially lead to the development of new inhibitors towards these enzymes. Lastly, this is the first eukaryotic cyanase deposited to the PDB.

Chapter 2 focused on a Glutathione S-Transferase (TuGST) from *T. urticae*. GSTs are important to almost all forms of life, including both prokaryotic and eukaryotic organisms. For GSTs, reduced glutathione is normally used as substrate to detoxify a target xenobiotic and in turn, makes it more soluble and permits it to be excreted from cells as waste. A few GSTs have been studied in *T. urticae* already, but here a new GST was investigated. TuGST did not seem to have any obvious affinity for all the substrates tested except CDNB, in which it bound with negative cooperativity. The future directions of this project include trying crystallization optimization by using techniques such as Surface Entropy Reduction (SERp) to determine the actual crystal structure of this TuGST, although GST secondary structure is well-conserved and the predicted structure shown here is probably a fair representation.

Active site residues could be mutated to alanine, especially Ser11 (interacts with reduced thiol of GSH), to provide further validity to the predicted structure if a crystal structure for TuGST cannot be obtained. Although attempting to use fusion partners to promote the crystallization of TuGST could be pursued (such a thioredoxin, rubredoxin and lysozyme), it would most likely not work based on protein expression and kinetic results in which the fusion partners seemed to completely hinder TuGST activity. Furthermore, the compounds all tested by UV-Vis spectroscopy could be used with more sensitive techniques such as SPR or ITC to determine dissociation constants and could provide very informative results on the purpose of this TuGST in *T. urticae.* Another option is to use High Performance Liquid Chromatography-Mass Spec (HPLC-MS) to measure the amount of substrate consumed or product formed during a given time period.

Chapter 3 described the function and crystal structure of an intradiol ring-cleavage dioxygenase from *T. urticae.* This type of enzyme is metalloprotein that utilizes an active site non-heme iron to degrade aromatic compounds. The typical target aromatic compounds of ID-RCDs is catechol, a compound found in plants that not only can serve as a self-defense chemical, but can also be a precursor molecule in some plant metabolic pathways. The ferric center of TuDioxygenase is coordinated by two tyrosines and two histidines which is common to most ID-RCDs. Based on the use of several bioinformatic programs, TuDioxygenase appears to only have one main domain which is the intradiol domain, despite the fact that SACTE_2871 (PDB code: 4ILT), the model used for molecular replacement of TuDioxygenase, contains a lignin-binding domain [93]. Both TuDioxygenase and SACTE_2871 contain signal peptides on the N-terminus, although the signal peptide for TuDioxygenase is predicted to be around 20 amino acids whereas the signal peptide for SACTE_2871 is around 40 amino acids [93].

Co-crystallization and crystal soaking with catechol and catechol derivatives could be performed to acquire holo-TuDioxygenase structures in which bidentate binding should occur of catechol to the active site iron. Kinetic assays need to be performed by UV-Vis spectroscopy to determine substrate specificity of TuDioxygenase for catechol and catechol derivatives (if any) or by using HPLC-MS as described for TuGST. Electroparamagnetic Resonance (EPR) can be performed to determine free radical formation during the reaction of catechol with TuDioxygenase which may provide more insight into interaction between substrate and the active site iron, and also provide information on the coordination spheres around the ferric center. Lastly, assays (such as DSF, EPR) or co-crystallization experiments could be performed with catechol-based

insecticides to observe if this TuDioxygenase may have implications in insecticide resistance.

Chapter 4 focused on 2-methylcitrate synthase (mcsA) from *A. fumigatus* which was compared and contrasted to a similar enzyme, human citrate synthase (hCS). Structurally, there were very few differences between mcsA and hCS, but it seems from the crystal structures that hCS may be able to form a tighter "closed" conformation than mcsA, which may attribute to its substrate specificity for only acetyl-CoA and not propionyl-CoA. Computational studies could be done to investigate the clamping motions between them, and ultimately may explain why mcsA displays positive cooperativity and hCS does not. The inhibitor studies performed here had interesting results, such that the type of inhibition towards hCS and mcsA seemed to be mostly different for the same inhibitor. Co-crystallizing the inhibitors tested here with either hCS or mcsA would provide a better explanation as to why different inhibition types are seen for each enzyme. Crystal soaking would probably not work as this has been tried in the past and will destroy the crystals, most likely due to the significant conformational changes that occur when the proteins bind ligands.

More DSF experiments can be performed to find any new potential mcsA-specific inhibitors. Alternatively, a 96-well plate reader may yield better (and more informative results), but difficulties may arise due to the fast catalysis performed by these enzymes. Organically synthesizing 2-methylcitryl-CoA analogue (or other similar CoA analogues) could be useful and informative if inhibitors that target the mcsA active site are still pursued. Allosteric inhibitors would be more ideal, but may prove challenging as mcsA and hCS are structurally similar and have similar surface charge distributions. Lastly, the

very first crystal structure of mcsA and hCS are reported here. There are several crystal structures for pCS and cCS in the PDB already; however, the hCS structures presented here are the first citrate synthase crystal structures from humans deposited in the PDB.

REFERENCES

1. Paini DR, Sheppard AW, Cook DC, et al (2016) Global threat to agriculture from invasive species. Proc Natl Acad Sci U S A 113:7575–7579. doi: 10.1073/pnas.1602205113
2. Van Leeuwen T, Van Pottelberge S, Tirry L (2006) Biochemical analysis of a chlorfenapyr-selected resistant strain of Tetranychus urticae Koch. Pest Manag Sci 62:425–433. doi: 10.1002/ps.1183
3. Pavlidi N, Tseliou V, Riga M, et al (2015) Functional characterization of glutathione S-transferases associated with insecticide resistance in Tetranychus urticae. Pestic Biochem Physiol 121:53–60. doi: 10.1016/j.pestbp.2015.01.009
4. Binder U, Lass-Florl C (2013) New insights into invasive aspergillosis--from the pathogen to the disease. Curr Pharm Des 19:3679–3688.
5. Ibrahim-Granet O, Dubourdeau M, Latge J-P, et al (2008) Methylcitrate synthase from Aspergillus fumigatus is essential for manifestation of invasive aspergillosis. Cell Microbiol 10:134–148. doi: 10.1111/j.1462-5822.2007.1025.x
6. Siddall ME (2004) Invertebrates.—R.C. Brusca and G. J. Brusca. 2003. Sinauer Associates, Sunderland, Massachusetts. xix + 936 pp. ISBN 0–87893–097–3. $109.95(cloth). Syst Biol 53:664–666.
7. Dunlop JA (2010) Geological history and phylogeny of Chelicerata. Arthropod Struct Dev 39:124–142. doi: 10.1016/j.asd.2010.01.003
8. Dunlop JA, Alberti G (2007) The affinities of mites and ticks: a review. J Zool Syst Evol Res 0:070907105857006–??? doi: 10.1111/j.1439-0469.2007.00429.x
9. Halliday RB, Oconnor BM, Baker AS (2000) Global diversity of mites. National Academy Press
10. Holt DC, Burgess ST, Reynolds SL, et al (2013) Intestinal proteases of free-living and parasitic astigmatid mites. Cell Tissue Res 351:339–352. doi: 10.1007/s00441-012-1369-9
11. Grbic M, Van Leeuwen T, Clark RM, et al (2011) The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature 479:487–492.
12. Grbic M, Khila A, Lee KZ, et al (2007) Mity model: Tetranychus urticae, a candidate for chelicerate model organism. Bioessays 29:489–496.
13. Jeppson LR, Keifer HH, Baker EW (1975) Mites injurious to economic plants. Univ of California Press
14. Verpoorte R, Choi YH, Kim HK (2010) Metabolomics: will it stay? Phytochem Anal 21:2–3. doi: 10.1002/pca.1191
15. Van Leeuwen T, Dermauw W (2016) The Molecular Evolution of Xenobiotic Metabolism and Resistance in Chelicerate Mites. Annu Rev Entomol 61:475–498. doi: 10.1146/annurev-ento-010715-023907
16. Van Leeuwen T, Tirry L, Yamamoto A, et al (2015) The economic importance of acaricides in the control of phytophagous mites and an update on recent acaricide

mode of action research. Pestic Biochem Physiol 121:12–21. doi: 10.1016/j.pestbp.2014.12.009
17. Van Leeuwen T, Vontas J, Tsagkarakou A, et al (2010) Acaricide resistance mechanisms in the two-spotted spider mite Tetranychus urticae and other important Acari: a review. Insect Biochem Mol Biol 40:563–572.
18. Ahn SJ, Dermauw W, Wybouw N, et al (2014) Bacterial origin of a diverse family of UDP-glycosyltransferase genes in the Tetranychus urticae genome. Insect Biochem Mol Biol 50:43–57. doi: 10.1016/j.ibmb.2014.04.003
19. Dermauw W, Osborne EJ, Clark RM, et al (2013) A burst of ABC genes in the genome of the polyphagous spider mite Tetranychus urticae. BMC Genomics 14:317. doi: 10.1186/1471-2164-14-317
20. English-Loeb GM (1990) Plant drought stress and outbreaks of spider mites: a field test. Ecology 71:1401–1411.
21. Maxmen A (2013) Crop pests: Under attack. Nature 501:S15-7. doi: 10.1038/501S15a
22. Kamennaya NA, Post AF (2011) Characterization of cyanate metabolism in marine Synechococcus and Prochlorococcus spp. Appl Env Microbiol 77:291–301. doi: 10.1128/AEM.01272-10
23. Luque-Almagro VM, Huertas MJ, Saez LP, et al (2008) Characterization of the Pseudomonas pseudoalcaligenes CECT5344 Cyanase, an enzyme that is not essential for cyanide assimilation. Appl Env Microbiol 74:6280–6288. doi: 10.1128/AEM.00916-08
24. Wybouw N, Balabanidou V, Ballhorn DJ, et al (2012) A horizontally transferred cyanase gene in the spider mite Tetranychus urticae is involved in cyanate metabolism and is differentially expressed upon host plant change. Insect Biochem Mol Biol 42:881–889.
25. Wybouw N, Dermauw W, Tirry L, et al (2014) A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. Elife 3:e02365. doi: 10.7554/eLife.02365
26. Anderson PM, Sung YC, Fuchs JA (1990) The cyanase operon and cyanate metabolism. FEMS Microbiol Rev 7:247–252.
27. Palatinszky M, Herbold C, Jehmlich N, et al (2015) Cyanate as an energy source for nitrifiers. Nature 524:105–108. doi: 10.1038/nature14856
28. Elleuche S, Poggeler S (2008) A cyanase is transcriptionally regulated by arginine and involved in cyanate decomposition in Sordaria macrospora. Fungal Genet Biol 45:1458–1469. doi: 10.1016/j.fgb.2008.08.005
29. Anderson PM (1980) Purification and properties of the inducible enzyme cyanase. Biochemistry 19:2882–2888.
30. Anderson PM, Johnson W V, Endrizzi JA, et al (1987) Interaction of mono- and dianions with cyanase: evidence for apparent half-site binding. Biochemistry 26:3938–3943.
31. Anderson PM, Korte JJ, Holcomb TA, et al (1994) Formation of intersubunit disulfide bonds and properties of the single histidine and cysteine residues in each subunit relative to the decameric structure of cyanase. J Biol Chem 269:15036–15045.
32. Anderson PM, Little RM (1986) Kinetic properties of cyanase. Biochemistry

25:1621–1626.
33. Little RM, Anderson PM (1987) Structural properties of cyanase. Denaturation, renaturation, and role of sulfhydryls and oligomeric structure in catalytic activity. J Biol Chem 262:10120–10126.
34. Walsh MA, Otwinowski Z, Perrakis A, et al (2000) Structure of cyanase reveals that a novel dimeric and decameric arrangement of subunits is required for formation of the enzyme active site. Structure 8:505–514.
35. Johnson W V, Anderson PM (1987) Bicarbonate is a recycling substrate for cyanase. J Biol Chem 262:9021–9025.
36. Butryn A, Stoehr G, Linke-Winnebeck C, Hopfner KP (2015) Serendipitous crystallization and structure determination of cyanase (CynS) from Serratia proteamaculans. Acta Crystallogr F Struct Biol Commun 71:471–476. doi: 10.1107/S2053230X15004902
37. Wilkins MR, Gasteiger E, Bairoch A, et al (1999) Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 112:531–552.
38. Gasteiger E, Gattiker A, Hoogland C, et al (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788.
39. Eschenfeldt WH, Makowska-Grzyska M, Stols L, et al (2013) New LIC vectors for production of proteins from genes containing rare codons. J Struct Funct Genomics 14:135–144. doi: 10.1007/s10969-013-9163-9
40. Bond SR, Naus CC (2012) RF-Cloning.org: an online tool for the design of restriction-free cloning projects. Nucleic Acids Res 40:W209-13. doi: 10.1093/nar/gks396
41. Otwinowski Minor W. Z (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol Macromol Crystallogr , part A 276:307–326.
42. Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. J Appl Crystallogr 30:1022–1025.
43. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M (2006) HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. Acta Crystallogr D Biol Crystallogr 62:859–866. doi: 10.1107/S0907444906019949
44. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr 62:1002–1011. doi: 10.1107/S0907444906022116
45. Winn MD, Ballard CC, Cowtan KD, et al (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr Sect D-Biological Crystallogr 67:235–242. doi: Doi 10.1107/S0907444910045749
46. Murshudov GN, Skubak P, Lebedev AA, et al (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67:355–367. doi: 10.1107/S0907444911001314
47. Painter J, Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. J Appl Crystallogr 39:109–111.
48. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126–2132.
49. Davis IW, Leaver-Fay A, Chen VB, et al (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35:W375-83.

50. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242.
51. Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38:W545–W549. doi: Doi 10.1093/Nar/Gkq366
52. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 60:2256–2268.
53. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372:774–797. doi: 10.1016/j.jmb.2007.05.022
54. DeLano W. S (2002) The PyMOL Molecular Graphics System.
55. Qian D, Jiang L, Lu L, et al (2011) Biochemical and structural properties of cyanases from Arabidopsis thaliana and Oryza sativa. PLoS One 6:e18300. doi: 10.1371/journal.pone.0018300
56. Guilloton MB, Korte JJ, Lamblin AF, et al (1992) Carbonic anhydrase in Escherichia coli. A product of the cyn operon. J Biol Chem 267:3731–3734.
57. Kozliak EI, Guilloton MB, Gerami-Nejad M, et al (1994) Expression of proteins encoded by the Escherichia coli cyn operon: carbon dioxide-enhanced degradation of carbonic anhydrase. J Bacteriol 176:5711–5717.
58. Elmore MH, McGary KL, Wisecaver JH, et al (2015) Clustering of two genes putatively involved in cyanate detoxification evolved recently and independently in multiple fungal lineages. Genome Biol Evol 7:789–800. doi: 10.1093/gbe/evv025
59. Hewett-Emmett D, Tashian RE (1996) Functional diversity, conservation, and convergence in the evolution of the alpha-, beta-, and gamma-carbonic anhydrase gene families. Mol Phylogenet Evol 5:50–77. doi: 10.1006/mpev.1996.0006
60. Syrjanen L, Tolvanen M, Hilvo M, et al (2010) Characterization of the first beta-class carbonic anhydrase from an arthropod (Drosophila melanogaster) and phylogenetic analysis of beta-class carbonic anhydrases in invertebrates. BMC Biochem 11:28. doi: 10.1186/1471-2091-11-28
61. (2015) UniProt: a hub for protein information. Nucleic Acids Res 43:D204-12. doi: 10.1093/nar/gku989
62. Enayati AA, Ranson H, Hemingway J (2005) Insect glutathione transferases and insecticide resistance. Insect Mol Biol 14:3–8. doi: 10.1111/j.1365-2583.2004.00529.x
63. Bass C, Field LM (2011) Gene amplification and insecticide resistance. Pest Manag Sci 67:886–890. doi: 10.1002/ps.2189
64. Hossain MDT, Yamada N, Yamamoto K (2014) Glutathione-binding site of a bombyx mori theta-class glutathione transferase. PLoS One 9:e97740. doi: 10.1371/journal.pone.0097740
65. Pavlidi N, Khalighi M, Myridakis A, et al (2017) A glutathione-S-transferase (TuGSTd05) associated with acaricide resistance in Tetranychus urticae directly metabolizes the complex II inhibitor cyflumetofen. Insect Biochem Mol Biol 80:101–115. doi: 10.1016/j.ibmb.2016.12.003
66. Bocedi A, Fabrini R, Lo Bello M, et al (2016) Evolution of Negative Cooperativity in Glutathione Transferase Enabled Preservation of Enzyme Function. J Biol Chem 291:26739–26749. doi: 10.1074/jbc.M116.749507
67. Tripathi T, Rahlfs S, Becker K, Bhakuni V (2007) Glutathione mediated regulation

of oligomeric structure and functional activity of Plasmodium falciparum glutathione S-transferase. BMC Struct Biol 7:67. doi: 10.1186/1472-6807-7-67
68. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254.
69. Kakuta Y, Usuda K, Nakashima T, et al (2011) Crystallographic survey of active sites of an unclassified glutathione transferase from Bombyx mori. Biochim Biophys Acta 1810:1355–1360. doi: 10.1016/j.bbagen.2011.06.022
70. Chen L, Hall PR, Zhou XE, et al (2003) Structure of an insect delta-class glutathione S-transferase from a DDT-resistant strain of the malaria vector Anopheles gambiae. Acta Crystallogr D Biol Crystallogr 59:2211–2217.
71. Yamamoto K, Higashiura A, Hossain MT, et al (2015) Structural characterization of the catalytic site of a Nilaparvata lugens delta-class glutathione transferase. Arch Biochem Biophys 566:36–42. doi: 10.1016/j.abb.2014.12.001
72. Cui J, Li LG, Jiang P, et al (2015) Biochemical and functional characterization of the glutathione S-transferase from Trichinella spiralis. Parasitol Res 114:2007–2013. doi: 10.1007/s00436-015-4410-6
73. Dirr H, Reinemer P, Huber R (1994) X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. Eur J Biochem 220:645–661.
74. Lien S, Gustafsson A, Andersson AK, Mannervik B (2001) Human glutathione transferase A1-1 demonstrates both half-of-the-sites and all-of-the-sites reactivity. J Biol Chem 276:35599–35605. doi: 10.1074/jbc.M103789200
75. Qin G, Jia M, Liu T, et al (2013) Characterization and functional analysis of four glutathione S-transferases from the migratory locust, Locusta migratoria. PLoS One 8:e58410. doi: 10.1371/journal.pone.0058410
76. Biasini M, Bienert S, Waterhouse A, et al (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252-8. doi: 10.1093/nar/gku340
77. Guex N, Peitsch MC, Schwede T (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. Electrophoresis 30 Suppl 1:S162-73. doi: 10.1002/elps.200900140
78. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201. doi: 10.1093/bioinformatics/bti770
79. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66:486–501. doi: 10.1107/S0907444910007493
80. Schrodinger LLC (2015) The PyMOL Molecular Graphics System, Version 1.8.
81. Yamamoto K, Usuda K, Kakuta Y, et al (2012) Structural basis for catalytic activity of a silkworm Delta-class glutathione transferase. Biochim Biophys Acta 1820:1469–1474. doi: 10.1016/j.bbagen.2012.04.022
82. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:W320-4. doi: 10.1093/nar/gku316
83. Brown CK, Vetting MW, Earhart CA, Ohlendorf DH (2004) Biophysical analyses of designed and selected mutants of protocatechuate 3,4-dioxygenase1. Annu Rev

Microbiol 58:555–585. doi: 10.1146/annurev.micro.57.030502.090927
84. Urszula Guzik, Katarzyna Hupert-Kocurek DW (2013) Intradiol Dioxygenases — The Key Enzymes in Xenobiotics Degradation. Biodegrad Hazard Spec Prod. doi: DOI: 10.5772/56205
85. Broderick JB (1999) Catechol dioxygenases. Essays Biochem 34:173–189.
86. Dermauw W, Wybouw N, Rombauts S, et al (2013) A link between host plant adaptation and pesticide resistance in the polyphagous spider mite Tetranychus urticae. Proc Natl Acad Sci U S A 110:E113-22. doi: 10.1073/pnas.1213214110
87. Weng J-K (2014) The evolutionary paths towards complexity: a metabolic perspective. New Phytol 201:1141–1149. doi: 10.1111/nph.12416
88. Heinz-Gerhard Franck JWS (1988) Industrial Aromatic Chemistry, 1st ed. doi: 10.1007/978-3-642-73432-8
89. Slabinski L, Jaroszewski L, Rychlewski L, et al (2007) XtalPred: a web server for prediction of protein crystallizability. Bioinformatics 23:3403–3405. doi: 10.1093/bioinformatics/btm477
90. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 62:439–450. doi: 10.1107/S0907444906005270
91. Laskowski RA, Watson JD, Thornton JM (2005) ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res 33:W89-93. doi: 10.1093/nar/gki414
92. Pettersen EF, Goddard TD, Huang CC, et al (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. doi: 10.1002/jcc.20084
93. Bianchetti CM, Harmann CH, Takasuka TE, et al (2013) Fusion of dioxygenase and lignin-binding domains in a novel secreted enzyme from cellulolytic Streptomyces sp. SirexAA-E. J Biol Chem 288:18574–18587. doi: 10.1074/jbc.M113.475848
94. Zheng H, Chordia MD, Cooper DR, et al (2014) Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. Nat Protoc 9:156–170. doi: 10.1038/nprot.2013.172
95. Zheng H, Cooper DR, Porebski PJ, et al (2017) CheckMyMetal: a macromolecular metal-binding validation tool. Acta Crystallogr Sect D, Struct Biol 73:223–233. doi: 10.1107/S2059798317001061
96. Shaik S, Munro AW, Sen S, et al (2011) Iron-Containing Enzymes. doi: 10.1039/9781849732987
97. Matera I, Ferraroni M, Kolomytseva M, et al (2010) Catechol 1,2-dioxygenase from the Gram-positive Rhodococcus opacus 1CP: Quantitative structure/activity relationship and the crystal structures of native enzyme and catechols adducts. J Struct Biol 170:548–564. doi: http://doi.org/10.1016/j.jsb.2009.12.023
98. Guzik U, Hupert-Kocurek K, Sitnik M, Wojcieszynska D (2013) High activity catechol 1,2-dioxygenase from Stenotrophomonas maltophilia strain KB2 as a useful tool in cis,cis-muconic acid production. Antonie Van Leeuwenhoek 103:1297–1307. doi: 10.1007/s10482-013-9910-8
99. Vaillancourt FH, Bolin JT, Eltis LD (2006) The ins and outs of ring-cleaving dioxygenases. Crit Rev Biochem Mol Biol 41:241–267. doi: 10.1080/10409230600817422

100. Paulussen C, Hallsworth JE, Alvarez-Perez S, et al (2017) Ecology of aspergillosis: insights into the pathogenic potency of Aspergillus fumigatus and some other Aspergillus species. Microb Biotechnol 10:296–322. doi: 10.1111/1751-7915.12367
101. Latge JP (1999) Aspergillus fumigatus and aspergillosis. Clin Microbiol Rev 12:310–350.
102. Askew DS (2008) Aspergillus fumigatus: virulence genes in a street-smart mold. Curr Opin Microbiol 11:331–337. doi: 10.1016/j.mib.2008.05.009
103. Segal BH (2009) Aspergillosis. N Engl J Med 360:1870–1884. doi: 10.1056/NEJMra0808853
104. Fraczek MG, Bromley M, Bowyer P (2011) An improved model of the Aspergillus fumigatus CYP51A protein. Antimicrob Agents Chemother 55:2483–2486. doi: 10.1128/AAC.01651-10
105. Howard SJ, Arendrup MC (2011) Acquired antifungal drug resistance in Aspergillus fumigatus: epidemiology and detection. Med Mycol 49 Suppl 1:S90-5. doi: 10.3109/13693786.2010.508469
106. Brock M, Fischer R, Linder D, Buckel W (2000) Methylcitrate synthase from Aspergillus nidulans: implications for propionate as an antifungal agent. Mol Microbiol 35:961–973.
107. Maerker C, Rohde M, Brakhage AA, Brock M (2005) Methylcitrate synthase from Aspergillus fumigatus. Propionyl-CoA affects polyketide synthesis, growth and morphology of conidia. FEBS J 272:3615–3630. doi: 10.1111/j.1742-4658.2005.04784.x
108. Seiler CY, Park JG, Sharma A, et al (2014) DNASU plasmid and PSI:Biology-Materials repositories: resources to accelerate biological research. Nucleic Acids Res 42:D1253-60. doi: 10.1093/nar/gkt1060
109. Cormier CY, Mohr SE, Zuo D, et al (2010) Protein Structure Initiative Material Repository: an open shared public resource of structural genomics plasmids for the biological community. Nucleic Acids Res 38:D743-9. doi: 10.1093/nar/gkp999
110. Cormier CY, Park JG, Fiacco M, et al (2011) PSI:Biology-materials repository: a biologist's resource for protein expression plasmids. J Struct Funct Genomics 12:55–62. doi: 10.1007/s10969-011-9100-8
111. Rosenbaum G, Alkire RW, Evans G, et al (2006) The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. J Synchrotron Radiat 13:30–45. doi: 10.1107/S0909049505036721
112. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276:307–326.
113. Murshudov GN, Vagin AA, Lebedev A, et al (1999) Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr D Biol Crystallogr 55:247–255. doi: 10.1107/S090744499801405X
114. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53:240–255. doi: 10.1107/S0907444996012255
115. Winn MD, Isupov MN, Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr D Biol Crystallogr 57:122–133.

116. Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol 374:300–321. doi: 10.1016/S0076-6879(03)74014-2
117. Chen VB, Arendall WB 3rd, Headd JJ, et al (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12–21. doi: 10.1107/S0907444909042073
118. Wiegand G, Remington SJ (1986) Citrate synthase: structure, control, and mechanism. Annu Rev Biophys Biophys Chem 15:97–117. doi: 10.1146/annurev.bb.15.060186.000525
119. Karpusas M, Branchaud B, Remington SJ (1990) Proposed mechanism for the condensation reaction of citrate synthase: 1.9-A structure of the ternary complex with oxaloacetate and carboxymethyl coenzyme A. Biochemistry 29:2213–2219.
120. Karpusas M, Holland D, Remington SJ (1991) 1.9-A structures of ternary complexes of citrate synthase with D- and L-malate: mechanistic implications. Biochemistry 30:6024–6031.
121. Alter GM, Casazza JP, Zhi W, et al (1990) Mutation of essential catalytic residues in pig citrate synthase. Biochemistry 29:7557–7563.
122. Kurz LC, Shah S, Frieden C, et al (1995) Catalytic strategy of citrate synthase: subunit interactions revealed as a consequence of a single amino acid change in the oxaloacetate binding site. Biochemistry 34:13278–13288.
123. Smitherman TC, Mukherjee A, Robinson JBJ, et al (1979) Human heart citrate synthase: purification, properties, kinetic and immunologic studies. J Mol Cell Cardiol 11:149–160.
124. Pantoliano MW, Petrella EC, Kwasnoski JD, et al (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. J Biomol Screen 6:429–440. doi: 10.1177/108705710100600609
125. Lavinder JJ, Hari SB, Sullivan BJ, Magliery TJ (2009) High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. J Am Chem Soc 131:3794–3795. doi: 10.1021/ja8049063
126. Bayer E, Bauer B, Eggerer H (1981) Evidence from inhibitor studies for conformational changes of citrate synthase. Eur J Biochem 120:155–160.
127. Ashkenazy H, Erez E, Martz E, et al (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res 38:W529-33. doi: 10.1093/nar/gkq399