2019-08-23

Structural determination of the carboxy-terminal portion of ATP-citrate lyase

Nguyen, Vinh Huynh

Nguyen, V. H. (2019). Structural determination of the carboxy-terminal portion of ATP-citrate lyase (Unpublished master's thesis). University of Calgary, Calgary, AB.

<http://hdl.handle.net/1880/110825>

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Structural determination of the carboxy-terminal portion of ATP-citrate lyase

by

Vinh Huynh Nguyen

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGICAL SCIENCES

CALGARY, ALBERTA

AUGUST, 2019

© Vinh Huynh Nguyen 2019

Abstract

ATP-citrate lyase (ACLY) is a key enzyme that links carbohydrate and lipid metabolism. Understanding the catalytic mechanism of ACLY can aid in drug development to decrease lipid production as a possible treatment for obesity and cancer. The goal of this study was to determine the structure of the carboxy(C)-terminal portion of *Chlorobium limicola* ACLY (CIACLY). A tobacco etch virus protease site was introduced into CIACLY so that the N- and C-terminal portions could be separated after cleavage. A 1.9 Å dataset was obtained from a crystal of the C-terminal portion of CIACLY co-crystallized with acetyl-CoA. The structure revealed that the C-terminal portion was the oligomerization site for ACLY and contained the CoA-binding site. The results of this study further advance the knowledge of how the C-terminal portion of ACLY participates in catalysis.

Acknowledgements

I would like to thank my supervisor Dr. Marie Fraser for giving me the opportunity to work in her lab to learn about X-ray crystallography. I initially joined the lab as an undergraduate project student and learned a lot about the field of X-ray crystallography. She also gave advice about professionalism and how to present myself in the scientific community. The experience I gained is invaluable and I greatly appreciate the opportunity I had to work in her lab.

I am also thankful for my committee members Dr. Sergei Noskov and Dr. Vanina Zaremberg. Both were thoughtful and offered useful advice for my project and I appreciate their inquiries about my future as a scientist. My lab members, Ji Huang and Kyle McDade, gave substantial advice about molecular biology, protein purification and crystallization that greatly helped my project. I also acknowledge Dr. Isabel Usón and Ana Medina for their work determining the initial phases that were used to solve the structure.

Finally, I would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Queen Elizabeth II scholarship for financially supporting my research. I also appreciate The Canadian Light Source and Advanced Photon Source for allowing me to collect diffraction data at their facilities.

Table of Contents

Abstract.....	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vi
List of Figures and Illustrations	vii
List of Symbols, Abbreviations and Nomenclature	x
 Chapter 1: Introduction	 1
ATP-citrate lyase.....	1
Prevalence of ACLY in eukaryotes	3
Eukaryotic versus prokaryotic ACLY	4
Acyl-CoA synthetase superfamily (NDP-forming)	6
Proposed catalytic mechanism of ACLY	7
Citrate synthase and the C-terminal portion of ACLY.....	8
Previous work on the C-terminal portion of <i>Cl</i> ACLY	12
Phase problem of X-ray crystallography	14
Goals and objectives	15
 Chapter 2: Materials and methods.....	 17
Introduction of the recognition site for the TEV protease.....	17
Purification of TEV protease	22
Production and purification of the C-terminal portion of <i>Cl</i> ACLY	23
Purification of selenomethionine labelled protein for phase determination	25
Crystallization and data collection	26
Production of the C-terminal portion of <i>Cl</i> ACLY.....	28
 Chapter 3: Results.....	 30
Insertion of the recognition site for TEV protease into <i>Cl</i> ACLY.....	30
Purification of TEV protease	32
Purification of the C-terminal portion of <i>Cl</i> ACLY	33
Crystallization of the C-terminal portion of <i>Cl</i> ACLY	39
Purification and crystallization of selenomethionine labelled <i>Cl</i> ACLY	41
Obtaining phase estimates for solving the structure	47
Production and purification of the C-terminal portion of <i>Cl</i> ACLY	50
 Chapter 4: Discussion	 56
Purification of the C-terminal portion of <i>Cl</i> ACLY	56
Structural determination of the C-terminal portion of <i>Cl</i> ACLY	56
CoA-binding site of <i>Cl</i> ACLY	57
<i>Comparison of the CoA-binding sites of ClACLY, Nter hACLY, and SCS....</i>	62
Oligomerization of the C-terminal portion of <i>Cl</i> ACLY	64
Catalytic mechanism of ACLY	68
Evolution of ACLY and CS.....	69
Production of the C-terminal portion of <i>Cl</i> ACLY.....	70

Chapter 5: Conclusion.....	72
Chapter 6: Future studies	74
References	75

List of Tables

Table 2.1. Summary of the four constructs that were made to test the insertion of the recognition site for the TEV protease at the two candidate sites	20
Table 2.2. The sets of forward and reverse primers utilized.....	21
Table 3.0. Crystallographic statistics for the datasets obtained from crystals of the C-terminal portion of <i>CIACLY</i>	47
Table 3.1. Statistics for the model of the C-terminal portion of <i>CIACLY</i>	49

List of Figures and Illustrations

Figure 1.1. ACLY catalyzes a key reaction that links carbohydrate and lipid metabolism.....	2
Figure 1.2. The RTCA cycle that allows autotrophic CO ₂ fixation in some prokaryotes.	5
Figure 1.3. Gene architecture of human ACLY versus <i>Chlorobium limicola</i> ACLY.....	6
Figure 1.4. Diagram of the domain architectures of SCSs and ACLYs.	7
Figure 1.5. Catalytic mechanism of CS to form citryl-CoA from acetyl-CoA and oxaloacetate .	10
Figure 1.6. Sequence alignment between multiple ACLYs and CSs	11
Figure 1.7. SDS-PAGE gel stained with Coomassie blue R-250 of dissolved <i>Cl</i> ACLY crystals produced from <i>in situ</i> chymotrypsin cleavage.....	13
Figure 1.8. Mass spectrometry results from chymotrypsin cleaved <i>Cl</i> ACLY crystals.	13
Figure 1.9. Plasmid map for <i>Cl</i> ACLY in the pET42b(+) vector.	16
Figure 2.1. Predicted chymotrypsin cleavage sites in the large (A) subunit of <i>Cl</i> ACLY.....	17
Figure 3.1. Sequence alignment of <i>aclA</i> and the sequencing result for TEV347G <i>Cl</i> ACLY.....	30
Figure 3.2. Sequence alignment of <i>aclA</i> and the sequencing results for XhoCanTEV354S <i>Cl</i> ACLY.....	31
Figure 3.3. Sequence alignment of <i>aclA</i> and the sequencing result for ConTEV354G <i>Cl</i> ACLY.....	31
Figure 3.4. Sequence alignment of <i>aclA</i> and the sequencing result for CanTEV354S <i>Cl</i> ACLY..	32
Figure 3.5. 14% SDS-PAGE gel for the purification of TEV protease.	33
Figure 3.6. 14% SDS-PAGE gel of sample from Ni-NTA for TEV347G <i>Cl</i> ACLY.....	34
Figure 3.7. 14% SDS-PAGE gel for cation exchange for the C-terminal portion of <i>Cl</i> ACLY....	35
Figure 3.8. Cation exchange chromatogram for TEV347G <i>Cl</i> ACLY after TEV proteolysis..	35
Figure 3.9. Anion exchange chromatogram for the C-terminal portion of <i>Cl</i> ACLY.	36
Figure 3.10. Superposition of the gel filtration chromatograms of wild-type <i>Cl</i> ACLY (red) and XhoCanTEV354S <i>Cl</i> ACLY (blue) <i>Cl</i> ACLY.....	37
Figure 3.11. Western blot showing that the C-terminal portion of <i>Cl</i> ACLY was purified.	38
Figure 3.12. Crystals of the C-terminal portion of <i>Cl</i> ACLY.....	39

Figure 3.13. Diffraction pattern from a crystal of the C-terminal portion of <i>ClACLY</i>	40
Figure 3.14. 14 % SDS-PAGE for the purification of selenomethionine labelled <i>ClACLY</i>	41
Figure 3.15. Cation exchange chromatogram for the C-terminal portion of <i>ClACLY</i> labelled with selenomethionine.	42
Figure 3.16. Anion exchange chromatogram for the C-terminal portion of <i>ClACLY</i> labelled with selenomethionine.	42
Figure 3.17. Crystals of the C-terminal portion of <i>ClACLY</i> labelled with selenomethionine.	43
Figure 3.18. Diffraction pattern obtained from a crystal of selenomethionine C-terminal portion of <i>ClACLY</i>	44
Figure 3.19. Crystals of the C-terminal portion of <i>ClACLY</i> produced from streak seeding of: a) native protein and b) selenomethionine labelled protein.	45
Figure 3.20. Diffraction pattern obtained from a crystal of the C-terminal portion of <i>ClACLY</i> grown from streak seeding.....	46
Figure 3.21. $CC_{1/2}$ as a function of resolution for the selenomethionine data.	48
Figure 3.22. Ribbon diagram of the C-terminal portion of <i>ClACLY</i>	49
Figure 3.23. Electron density and modelled portion of acetyl-CoA bound.	50
Figure 3.24. The sequencing result for V351 <i>ClACLY</i> showing the 5' and 3' ends.....	51
Figure 3.25. Gel filtration chromatogram of V351 <i>ClACLY</i>	52
Figure 3.26. Cation exchange chromatogram of V351 <i>ClACLY</i>	52
Figure 3.27. Anion exchange chromatogram for V351 <i>ClACLY</i>	53
Figure 3.28. 14% SDS-PAGE gel of samples from each purification step for V351 <i>ClACLY</i>	53
Figure 3.29. Crystal of V351 <i>ClACLY</i> grown with a well solution of 200 mM lithium citrate tribasic pH 8.4 and 20% w/v PEG 3350.	54
Figure 3.30 Crystals of V351 <i>ClACLY</i> grown with a well solution of 200 mM ammonium tartrate dibasic pH 6.6 and 20% w/v PEG 3350.	55
Figure 3.31. Crystals of V351 <i>ClACLY</i> grown with a well solution of 200 mM ammonium tartrate dibasic pH 6.6 and 20% w/v PEG 3350.	55
Figure 4.1. The adenine group of acetyl-CoA bound into a hydrophobic pocket composed of residues 485 to 489 in the C-terminal portion of <i>ClACLY</i>	58

Figure 4.2. The long-distance hydrogen bonding interaction between N1 of the adenine ring and the amide nitrogen of V486.....	59
Figure 4.3. The ionic interaction between 3' phosphate of acetyl-CoA and K534.	59
Figure 4.4. Superposition of three structures of the C-terminal portion of <i>Cl</i> ACLY (6NZY (blue), 6HXN (red), and 6QCL (green)) to show the degree of openness.....	61
Figure 4.5. Superposition of two structures of the C-terminal portion of <i>Cl</i> ACLY (6NZY, gold and 6QCL, green) showing the molecules bound at the active sites of both structures.	62
Figure 4.6. Full-length <i>h</i> ACLY showing CoA-binding between N- and C-terminal portions.	63
Figure 4.7. The interactions of the α -helices at the central interface of the tetramer.	65
Figure 4.8. Structural superposition of chicken citrate synthase (yellow) with two protomers of the C-terminal portion of <i>Cl</i> ACLY (blue and red).	66
Figure 4.9. Structural alignment of the C-terminal portion of <i>Cl</i> ACLY and chicken CS.	67
Figure 4.10. Superposition of chicken CS (green) and the C-terminal portion of <i>Cl</i> ACLY.	68
Figure 4.11. Structure of full-length <i>Cl</i> ACLY shows that the first candidate site (Grey, EVMKQG) is exposed.....	71

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
2-ME	2-mercaptoethanol
A ₂₈₀	absorbance at 280 nm
Akt	protein kinase B
AD	anomalous dispersion
ADP	adenosine diphosphate
AMP	adenosine monophosphate
ATP	adenosine triphosphate
AS	ammonium sulfate
ACLY	ATP-citrate lyase
CoA	coenzyme A
CRAT	carnitine acetyltransferase
CLS	Canadian Light Source
<i>ClACLY</i>	<i>Chlorobium limicola</i> ACLY
CMCF	Canadian Macromolecular Crystallography Facility
CS	citrate synthase
COOT	crystallographic object-oriented toolkit
Da	Dalton
DNA	deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
GSK-3	glycogen synthase kinase-3

g	gram
x <i>g</i>	times gravity
<i>h</i> ACLY	human ACLY
IMCA-CAT	Industrial Macromolecular Crystallography Association Collaborative Access Team
IPTG	isopropyl β-D-1-thiogalactopyranoside
IR	isomorphous replacement
ITC	isothermal calorimetry
LB	lysogeny broth
mL	milliliter
mS/cm	milliSiemens per centimeter
M	mole per liter
MES	2-(N-morpholino)ethanesulfonic acid
MMS	microseed matrix screening
MR	molecular replacement
NEB	New England Biolabs
Ni-NTA	nickel-nitrilotriacetic acid
Phenix	python-based hierarchical environment for integrated Xtallography
rpm	rotations per minute
RTCA cycle	reductive tricarboxylic acid cycle
PCAF	p300/CBP-associated factor
PCR	polymerase chain reaction

PEG 3350	polyethylene glycol 3350
PDB	Protein Data Bank
SIRT2	Protein deacetylase sirtuin 2
PKA	cAMP-dependent protein kinase
rmsd	root-mean-square deviation
<i>S. cerevisiae</i>	<i>Sacchromyces cerevisiae</i>
SAD	single-wavelength anomalous dispersion
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SCS	succinyl-CoA synthetase
TB	terrific broth
TCA cycle	tricarboxylic acid cycle
TEV protease	tobacco etch virus protease
Tris	tris(hydroxymethyl)aminomethane
TCEP	tris(2-carboxyethyl)phosphine

Chapter 1: Introduction

ATP-citrate lyase

ATP-citrate lyase (ACLY, EC 2.3.3.8) is localized in the cytoplasm, catalyzing the following reaction in the presence of magnesium ions (Srere, 1959):

ACLY is an important enzyme linking carbohydrate and lipid metabolism (Figure 1.1). ACLY takes citrate produced by the tricarboxylic acid (TCA) cycle and makes acetyl-CoA in the cytosol (Kaplan, Mayor, & Wood, 1993; Wiegand & Remington, 1986). Although acetyl-CoA is produced in the mitochondria by the pyruvate dehydrogenase complex, it cannot pass through the mitochondrial membranes to enter the cytosol. Mitochondrial citrate synthase (CS) converts acetyl-CoA into citrate, which can enter the cytosol through mitochondrial tricarboxylate transport proteins. ACLY can convert the exported citrate back into acetyl-CoA in the cytosol. The oxaloacetate produced by ACLY can be reduced to malate, which can travel back into the mitochondria through the tricarboxylate transport proteins to re-enter the TCA cycle (Bauer, Hatzivassiliou, Zhao, Andreadis, & Thompson, 2005; Hatzivassiliou et al., 2005). Thus, ACLY serves as a link between carbohydrate and lipid metabolism by converting citrate back into acetyl-CoA for cells to synthesize lipids. Due to its involvement in lipid metabolism, ACLY is a proposed target for drug therapy to treat obesity and cancer.

Figure 1.1. ACLY catalyzes a key reaction that links carbohydrate and lipid metabolism.

ACLY is involved in gene regulation (Wellen et al., 2009). The acetyl-CoA produced by ACLY can be utilized by histone acetyltransferases to selectively regulate genes involved in carbohydrate metabolism. However, global cellular acetylation still occurs when ACLY is silenced with siRNA. It is unclear how the acetyl-CoA produced by ACLY can selectively regulate genes involved in carbohydrate metabolism.

Human ACLY (*hACLY*) can undergo post-translational modifications that affect its stability and activity. A disordered region of the protein (residues 425 – 485) that has not been seen in structures of ACLY (PDB ID: 3MWD, (Sun, Hayakawa, Bateman, & Fraser, 2010)) contains T and S residues that can be phosphorylated (Berwick, Hers, Heesom, Moule, & Tavaré, 2002; Potapova, El-Maghrabi, Doronin, & Benjamin, 2000). T446 and S450 are phosphorylated by glycogen synthase kinase-3 (GSK-3), whereas S454 is phosphorylated by

cAMP-dependent protein kinase (PKA) or insulin-stimulated kinase; the phosphorylation of these residues increases the activity of *hACLY*. Hormones can indirectly regulate the activity of *hACLY*. The presence of insulin inhibits GSK-3 and PKA, decreasing the phosphorylation of T446, S450, and S454 in the linker region, thereby decreasing the activity of *hACLY*. This disordered region appears to regulate *hACLY* through phosphorylation of T446, S450, and/or S454.

The amount of *hACLY* present in cells is partly regulated by acetylation and ubiquitination (Tao et al., 2014). K540, 546, and 554 are ubiquitinated on the ϵ nitrogen by UBR4 to target *hACLY* for degradation; however, acetylation of the same residues prevents ubiquitination. p300/CBP-associated factor (PCAF) has been shown to acetylate K540, 546, and 554, preventing *ACLY* from being targeted to a proteasome through ubiquitination. Protein deacetylase sirtuin 2 (SIRT2) removes the acetyl group from K540, 546, and 554. Ubiquitination is partly regulated by glucose concentrations because the interaction between UBR4 and *hACLY* decreases as glucose concentrations increase. Ubiquitination or acetylation of the three K residues regulate the *in vivo* abundance of *hACLY*. There have been post-translational modifications found on other residues of *ACLY* from mass spectrometry experiments, but the biological relevance of those modifications has not been examined.

Prevalence of *ACLY* in eukaryotes

Although *ACLY* is found in many eukaryotic organisms, it is not the only enzyme that produces cytosolic acetyl-CoA. A large survey of yeasts has identified three different ways for eukaryotes to synthesize cytosolic acetyl-CoA *in vivo* (Nissen, Schulze, Nielsen, & Villadsen, 1997). The first pathway involves *ACLY* producing acetyl-CoA from citrate using CoA and

ATP. The second approach involves acetyl-CoA synthetase, which catalyzes the following reaction (Tucek, 1967):

The third process involves converting acetyl-CoA into acetylcarnitine through carnitine acetyltransferase (CRAT). Acetylcarnitine can travel through carnitine transporters from the mitochondria to reach the cytosol, where CRAT converts it back to acetyl-CoA. *S. cerevisiae* only utilizes acetyl-CoA synthetase to produce cytosolic acetyl-CoA. However, other yeast strains such as *Rhodotorula gracilis* and *Lipomyces lipofers* have ACLY (Boulton & Ratledge, 1981). Thus, ACLY is one of three enzymes that can create cytosolic acetyl-CoA.

Eukaryotic versus prokaryotic ACLY

Eukaryotic ACLY has been studied predominantly in humans (Bauer et al., 2005; Berwick et al., 2002; Boulton & Ratledge, 1981; Hatzivassiliou et al., 2005; Wellen et al., 2009). Prokaryotic ACLY has not been studied extensively because it has only been found in prokaryotes that utilize the reductive tricarboxylic acid (RTCA) cycle (Fuchs, Stupperich, & Gerolf, 1980) (Figure 1.2). The RTCA cycle enables autotrophic CO₂ fixation in prokaryotes such as green sulfur bacteria (Wahlund & Tabita, 1997). Within the family of green sulfur bacteria, *Chlorobium* was the genus where the RTCA cycle was discovered. The RTCA cycle starts with by converting oxaloacetate to malate via malate dehydrogenase (Evans, Buchanan, & Arnon, 1966). The incorporation of two carbon atoms from two CO₂ molecules eventually produces citrate, which is utilized by ACLY to produce acetyl-CoA and oxaloacetate. ACLY has not been found in prokaryotes that do not utilize the RTCA cycle, so it is not an essential enzyme for prokaryotes.

Figure 1.2. The RTCA cycle that allows autotrophic CO₂ fixation in some prokaryotes. Enzymes that differ from those in the the TCA cycle are boxed.

A comparison of human and prokaryotic ACLY shows differing gene architecture (Fuchs et al., 1980). In humans, there are several isoforms of ACLY that result from alternative splicing, with isoform 1 (1101 residues) representing the canonical form. *Chlorobium limicola* ACLY (*ClACLY*) is encoded by two genes (Kanao, Fukui, Atomi, & Imanaka, 2001). The two genes, *AclB* (Accession no. AB054669) and *AclA* (Accession no. AB054670), code for the B and A subunits respectively. The quaternary structure of *hACLY* is a homotetramer, whereas *ClACLY* forms a heterooctamer (A₄B₄). *ClACLY* does not have phosphorylation sites that are present in *hACLY* (Figure 1.3) (Berwick et al., 2002).

Figure 1.3. Gene architecture of human ACLY versus *Chlorobium limicola* ACLY. The B and A subunits are coloured blue and red, respectively. The regions in hACLY that show sequence similarity between the B and A subunits are coloured. The yellow region in the human ACLY are regulatory sites that are not present in ClACLY. Residue numbers of hACLY are labelled.

Acyl-CoA synthetase superfamily (NDP-forming)

The acyl-CoA synthetase superfamily (NDP-forming) catalyzes the general reaction:

where the reverse reaction would classify enzymes as acyl-CoA synthetases (Sánchez, Galperin, & Müller, 2000). Some members can only catalyze the above chemical reaction in one direction, whereas other members can catalyze the forward and reverse reactions. Succinyl-CoA synthetase (SCS) is the prototype enzyme for this family and can catalyze the forward and reverse directions. ACLY is a member of this family, catalyzing the reverse reaction. Each of the members of this superfamily has five domains that are common amongst them (Sun et al., 2010). The domain architecture of the superfamily was deduced from the structural analysis of *Escherichia coli* SCS because it was the first enzyme in the superfamily that had its structure solved. The domain architecture could be deduced for ACLY by aligning its sequence with the domains found in *E. coli* SCS (Figure 1.4). The sequence of the C-terminal portion of ACLY aligns with CS. Based on the proposed domain architecture, there are three catalytic sites in ACLY.

Figure 1.4. Diagram of the domain architectures of SCSs and ACLYs.

Domains 3 and 4 come together to form the nucleotide binding site. Domain 2 contains a histidine that can be phosphorylated. Domain 5 contains the binding site for the organic acid. The C-terminal portion of ACLY has sequence similarity to citrate synthase (CS). There are three catalytic sites in ACLY from the sequence similarity with SCS and CS.

Proposed catalytic mechanism of ACLY

The proposed four step catalytic mechanism of ACLY is,

where E represents ACLY, E-P is the phosphorylated enzyme, and $\bullet$ indicates that E interacts non-covalently with either citryl-P or citryl-CoA (Sun, Hayakawa, & Fraser, 2011). Multiple structures of the N-terminal portion of *h*ACLY have been solved with the substrates citrate (Sun et al., 2010) and ADP-Mg²⁺ bound (Sun et al., 2011) as well as with inhibitors tartrate (Hu, Komakula, & Fraser, 2017; Sun et al., 2010) and hydroxycitrate (Hu et al., 2017). These structures support the hypothesis that the nucleotide- and citrate-binding domains are in the N-

terminal portion. The phosphohistidine loop in domain two has been seen in some structures of the N-terminal portion of ACLY that was hypothesized to shuttle the phosphoryl group between the nucleotide- and citrate-binding sites to facilitate steps one and two in the reaction mechanism (Hu et al., 2017). The N-terminal portion of *h*ACLY cannot bind CoA, even though domain one shares sequence similarity with the CoA-binding domain of SCS (Sun et al., 2010). Therefore, structures of the N-terminal portion only provide data to explain the first two steps of the catalytic mechanism. The C-terminal portion was predicted to contain the CoA-binding site. Solving the structure of the C-terminal portion of *Cl*ACLY could determine its role in steps three and four of the catalytic mechanism.

Citrate synthase and the C-terminal portion of ACLY

One hint to the function and structure of the C-terminal portion of ACLY is that *h*ACLY and human CS have 40% sequence similarity. CS is a mitochondrial enzyme that catalyzes a Claisen condensation of acetyl-CoA and oxaloacetate to form a citryl-CoA intermediate, which is hydrolyzed to produce citrate and CoA (Srere, 1969; Wiegand & Remington, 1986):

An important consideration is that CS catalyzes the formation of citrate and CoA from acetyl-CoA and oxaloacetate, whereas ACLY catalyzes the formation of acetyl-CoA and oxaloacetate from CoA, citrate and ATP. Thus, the similarity of their sequences and catalytic reactions suggest that the two could have a similar structure that facilitates their activity.

Previous structures of chicken CS revealed that there are three key residues (H274, D375, and H320) that are involved in catalysis (Alter, Zhi, Nemeth, Srere, & Evans, 1990; Karpusas, Branchaud, & Remington, 1990). H274 and D375 participate in enolization of acetyl-CoA,

where D375 deprotonates acetyl-CoA, while H274 donates a proton to stabilize the enol (Figure 1.2). Oxaloacetate is protonated by H320, making it susceptible to nucleophilic attack by the double bond of the enol to form citryl-CoA (Wiegand & Remington, 1986). Computational studies also supported the role of these three key residues (Mulholland & Richards, 1997; van der Kamp, Zurek, Manby, Harvey, & Mulholland, 2010; Yang & Druekhammer, 2003).

From sequence alignments, the catalytic residues H274, D375, and H320 of CS align with F450, D543, and H491, respectively, in the A subunit of *Cl*ACY (Figures 1.3). The conservation of D375 and H320 in *Cl*ACY suggest that these residues are essential for cleavage of citryl-CoA as citryl-CoA is the intermediate for both CS and ACY (Alter et al., 1990; Karpusas et al., 1990). Previously, the activity of H491A or D543A mutant proteins were measured and the H491A mutation decreased specific activity by 64%, whereas the D543A mutation completely abolished activity (Singh, 2016). The mutation of H491 did not abolish activity, but the significant decrease in activity with the loss of the side chain would explain its conservation in ACYs. Residue H274 found in chicken CS is conserved across multiple CS; however, multiple ACYs have a conserved phenylalanine instead of a histidine in the equivalent position (Figure 1.3). The role of the conserved phenylalanine is not known, but a structure of the C-terminal portion of ACY could help explain its role in catalysis.

Figure 1.5. Catalytic mechanism of CS to form citryl-CoA from acetyl-CoA and oxaloacetate. Catalytic residues are numbered based on chicken CS

Figure 1.6. Sequence alignment between multiple ACLYs and CSs. The key catalytic residues in CS and their equivalent residues in ACLY have been boxed.

There are two types of CS that differ in their structure and regulation. Type II CS are found exclusively in gram-negative bacteria, forming a hexamer and some are allosterically regulated by NADH (Nguyen et al., 2001). Type I CS are found in other organisms, forming a dimer that is unregulated. Both type I and II CS need to oligomerize to function properly and are mostly α -helical. The sequence similarity between the C-terminal portion of ACLY and CS suggests that the full-length ACLY could also oligomerize in this portion of the protein.

Structures of Type I CS from pig and chicken show that it can adopt open and closed conformations (Remington, Wiegand, & Huber, 1982). In the closed conformation, residues from another subunit form interactions with the pantothenate group of acetyl-CoA/CoA to facilitate binding. Acetyl-CoA/CoA cannot bind when CS is in an open conformation because residues from the other subunit are not within distance to interact with the pantothenate group. A closed CS also protects the substrates in the active site from hydrolysis by free water molecules. Therefore, the closing of CS is necessary for acetyl-CoA binding. The presence of oxaloacetate

in the active site appears to induce a closed conformation in CS. Crystals of CS in an open conformation cracked when soaked with oxaloacetate, suggesting that the enzyme underwent a conformational change. Oxaloacetate induced closure of CS can explain why the binding affinity of acetyl-CoA increases by 20x when CS is bound with oxaloacetate (Johannsson & Pettersson, 1977). The open and closed conformation that CS adopts may also be present in ACLY and can influence catalysis.

Previous work on the C-terminal portion of *CI*ACLY

Previous work utilized *in situ* chymotrypsin cleavage to determine the first structure of the N-terminal portion of *h*ACLY (Sun et al., 2010). A similar approach was done with *CI*ACLY to see if crystals would grow after proteolysis. Crystals formed when *CI*ACLY was cleaved with chymotrypsin. Dissolved crystals were analyzed on SDS-PAGE and a ~28 kDa protein was found (Figure 1.4). Analysis of trypsin cleaved peptides from the crystals through mass spectrometry found peptides that matched to the sequence of the C-terminal portion of *CI*ACLY (Figure 1.5). High-resolution data ($< 3 \text{ \AA}$) were obtained from these crystals, but the structure could not be solved by molecular replacement. The multiple different cleavage products from chymotrypsin could crystallize differently. To improve the protein sample, a protease, such as tobacco etch virus (TEV) protease, that cleaves at a specific site would be necessary to produce only one cleavage product.

Figure 1.7. SDS-PAGE gel stained with Coomassie blue R-250 of dissolved *CIACLY* crystals produced from *in situ* chymotrypsin cleavage.

Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications: [Carbamidomethyl \(C\)](#)
Variable modifications: [Oxidation \(M\)](#)

Protein sequence coverage: 25%

Matched peptides shown in **bold red**.

```

1  MSILANKDTR AVIIGGVAGV NAAKRMAQFD FLVNRPLTVQ AFVYPPEAGQ
51  QKEIFRGGEL KNTVTVYDSL PALEEHDPIN TALIYLGASR AAQAAKEALE
101 SPNIQLVSMI TEGVPEKDAK RLKKLAQKLG KMLNGPSSIG IMSAGECRLG
151 VIGGEFKNLK LCNLYRQGSF GVLTKSGGLS NEAMWLCAQN GDGITSVAI
201 GGDAYPGTDF VTYLEMFEDK PATKAVVMIG EVGGNLEEEA AEWLAAEPRR
251 IKLIAAIGGT CQEVLPQGMK FGHAGAKEGK KGAGSARSKM NALRDAGAYV
301 PDTFGGLSKE IKKVYEELIA AGEISTEIDE AVLPELPPRV QEVMKQGEVI
351 VEPLIRTTIS DDRGEEPRYA GYAASELCSK GYGIEDVIGL LWNKKLP TRE
401 ESEIIKRIVM ISADHGPAVS GAFGSILAAC AGIDMPQAVS AGMTMIGPRF
451 GGAVTNAGKY FKMAVEDYPN DIPGFLSWMK KNVGPVPGIG HRVKS VKNPD
501 QRVKYLVSII KNETSLHTPC LDYALEVEKV TTAKKGNLIL NVDGTIGCIL
551 MDLDFPVHSL NGFFVLARTI GMIGHWIDQN NQNSRLIRLY DYLINYAVKP
601 EQEVPEKK

```

Figure 1.8. Mass spectrometry results from chymotrypsin cleaved *CIACLY* crystals. The protein band from Figure 1.4 was excised and sent to the Southern Alberta Mass Spectrometry Facility. Peptide detected (red) were matched to the sequence of the C-terminal portion of *CIACLY*.

Phase problem of X-ray crystallography

There is no lens that can focus X-rays to reconstruct the image of protein molecules from diffracted X-rays. The use of detectors allows measurement of X-ray diffraction, but the phase problem occurs because detectors can only measure the intensity of the diffracted X-rays. Both amplitudes and phases are required to calculate electron density. Therefore, phases must be obtained to solve the structure of a protein. There are three techniques that are commonly utilized to obtain phase information for determining protein structures: molecular replacement (MR), anomalous dispersion (AD), and isomorphous replacement (IR).

MR relies on the availability of a protein structure, called the search model, that is similar to the protein crystallized. The search model is translated and rotated in the crystal until it is correctly placed. The phases from the search model are used as the initial phases once the search model has been placed.

AD and IR involves the use of marker atoms, such as heavy atoms, to calculate a difference map that shows the contribution of only the marker atoms on the diffraction pattern. The difference map decreases the complexity of phase determination by looking at a smaller number of marker atoms, instead of 1000s of atoms in proteins. IR involves soaking a crystal with a heavy atom to produce a derivative crystal. The native and derivative crystal diffraction patterns are subtracted to obtain a difference map that is used to determine the location of the heavy atoms. The phases and amplitudes of the heavy atoms are calculated and used to estimate the phases of the protein. The native and derivative crystals must have the same unit cell dimensions, hence isomorphous. AD involves measurement of anomalous signal from heavy atoms either introduced during protein production like selenomethionine, ones that are already present in the protein, ones that are soaked into protein crystals. A difference map can be

calculated in AD based on the anomalous signal detected. AD has advantages over IR in that it requires a good dataset from only one crystal, thus no isomorphous requirements. Thus, AD and MR are more commonly used to obtain phases than IR.

Goals and objectives

The current work focuses on determining the structure of the C-terminal portion of *Cl*ACYL. The structure of the C-terminal portion can reveal how the residues equivalent to the catalytic residues of CS are arranged in *Cl*ACYL to facilitate catalysis. It was predicted that the CoA-binding site resides in the C-terminal portion of ACYL. The oligomerization site for ACYL was also predicted to be the C-terminal portion.

Proteolysis was used to isolate the C-terminal portion of *Cl*ACYL from the rest of the protein. TEV protease was used instead of chymotrypsin due to its specificity, cleaving at a consensus sequence of EXXYXQ[^]S/G, where [^] indicates where the peptide bond would be cleaved (Carrington & Dougherty, 1988). The specificity of proteolysis would result in only one cleavage product so that a homogenous protein solution would be used for crystallization experiments.

The current work can be split into five major steps. The first step is to determine where to introduce the TEV proteolysis site in the protein to separate the N- and C-terminal portions. *E. coli* codon-optimized genes coding for the A and B subunit cloned in the pET42b(+) vector were used as the template for mutagenesis (Figure 1.6). The second step is to see if TEV protease could cleave *Cl*ACYL after the recognition site had been introduced. For proteolysis to be successful, the recognition site needs to be exposed so that TEV protease has access to the site. The third step is to purify the C-terminal portion from the N-terminal portion, the B subunit, and

TEV protease. Nickel-nitrilotriacetic acid (Ni-NTA) will be the first purification step because the B subunit contains a His-tag, followed by ion exchange chromatography. The fourth step is to obtain crystals of the C-terminal portion that diffract to high-resolution. The final step is to obtain phases to solve the structure using either MR or AD.

Figure 1.9. Plasmid map for *CIACLY* in the pET42b(+) vector. Codons coding for the His-tag are located at the 3' end of the gene that encodes the B subunit.

Chapter 2: Materials and methods

Introduction of the recognition site for the TEV protease

PeptideCutterTM (Gasteiger et al., 2005) was used to predict the possible chymotrypsin cleavage sites that would produce a ~28 kDa protein. There were two candidate sites determined from the prediction, and four constructs were made to test the insertion of the TEV protease site at these two locations (Figure 2.1). No off-target cleavage should occur because no TEV protease sites were found in the wild-type *ClACLY*.

Figure 2.1. Predicted chymotrypsin cleavage sites in the large (A) subunit of *ClACLY*. The blue region represents the portion of the sequence of the large (A) subunit that aligns with the primary sequence of the structure of the N-terminal portion of hACLY (PDB ID: 3PFF). The green region represents the N-terminal peptide detected by mass spectrometry in the crystals grown from chymotrypsin-cleaved *ClACLY*. The most probable sites that produced a ~28 kDa protein were highlighted grey. These two sites would produce a C-terminal portion with theoretical molecular weights 28.5 and 27.5 kDa after TEV proteolysis.

The first candidate site has the sequence EVM^KQG, where ^ represents the predicted chymotrypsin cut site. Due to the similarity of the sequence of the first site and the consensus sequence for TEV protease, (EXXYXQ^G/S, (Carrington & Dougherty, 1988)), the first construct was made by inserting a tyrosine between methionine and lysine and changing the lysine to a phenylalanine (Table 2.1). The tyrosine was introduced because it is essential for TEV proteolysis based off the consensus sequence. The lysine was mutated to phenylalanine because the structure of TEV protease (PDB ID: 1LVB) indicated that the binding pocket for this residue is hydrophobic. Phenylalanine was chosen because it is found in the canonical sequence for TEV protease (ENLYFQ^S).

The first construct, called TEV347GC $\textit{Cl}$ ACY, was created following the Agilent Quikchange LightningTM mutagenesis protocol using BesTaqTM polymerase (Applied Biological Materials). Primers containing the mutations were elongated using the plasmid with codon optimized genes for $\textit{Cl}$ ACY (Genscript) as the template. The protocol was modified so that the primer elongation occurs initially in two 20 μL samples, each containing 1x BesTaqTM PCR buffer, 1x BesTaqTM GC enhancer, 25 ng template DNA, 1 mM dNTPs, and 1.25 U BestaqTM DNA polymerase. The forward and reverse primers (VN2.1 and 2.2, Table 2.2) were added separately to a concentration of 1 μM into the two 20 μL samples so they can be elongated, rather than annealing together (Wang & Malcolm, 1999). The two samples were placed in a thermocycler for 12 cycles and mixed together afterwards, where they were elongated together for 18 cycles. 20 U DpnI (New England Biolabs (NEB)) was added and the mixture was incubated for one hour at 37°C to remove template DNA. The digested sample was transformed into Inoue competent *E. coli* DH5 α (Sambrook & Green, 2012). Colonies were selected and grown overnight at 37°C in 5 mL LB containing 30 $\mu\text{g}/\text{mL}$ kanamycin. The plasmid was purified

using the Monarch Plasmid Miniprep Kit (NEB) and sent to the University of Calgary DNA sequencing centre to confirm successful mutation and to ensure no other mutations were introduced into the two genes of *CIACLY*.

Constructs two to four were made to test the second candidate site, EV[^]IVEPL. The second construct, called XhoCanTEV354SC[^]IA[^]CLY, was made by replacing the second candidate site with the canonical sequence. The mutagenesis protocol used to generate the first construct was not utilized for the second construct because of the large number of nucleotide changes required to replace the second candidate site with the canonical sequence. Primer VN3.1 contained codons for the canonical sequence and the XhoI site in its 5' overhang, whereas the reverse primer VN3.2 had a 5' overhang coding for the XhoI site (Table 2.1). The two primers were designed to introduce XhoI sites on the 5' and 3' ends after elongation so that the amplified product could be circularized through ligation. An XhoI site was chosen because the restriction site codes for residues LE: only a L would be introduced after amplification of the primers. Primers VN3.1 and 3.2 will change the sequence from EVIVEPL to LENLYFQS at the second candidate site (Table 2.1). The reaction mixture contained 1x BesTaq[™] PCR buffer, 1x BesTaq[™] GC enhancer, 50 ng template DNA, 1 mM dNTPs, 1 µM each of VN3.1 and 3.2, and 2.5 U Bestaq[™] DNA polymerase. 20 U DpnI (NEB) were added after 25 cycles of primer elongation, and the digestion mixture was incubated for one hour at 37°C. 20 U XhoI (Invitrogen) was added to the digested sample and the mixture was incubated overnight at 37°C. The Monarch PCR and DNA Cleanup Kit was used to remove excess nucleotides and enzymes from the overnight digestion sample. 5 U T4 DNA ligase was added, and the ligation mixture was incubated overnight at 4°C. The ligation mixture was transformed into Inoue competent *E. coli* DH5α (Sambrook & Green, 2012). Overnight cultures were grown in 5 mL LB with 30

µg/mL kanamycin, and the purified plasmid was sent to the University of Calgary DNA sequencing centre to confirm the introduction of the canonical TEV protease site and to ensure that no other mutations were introduced in the two genes.

The third construct, called ConTEV354GCIACLY was created by introducing the consensus recognition site at the second candidate site. The last four residues at the second site were mutated to match the consensus sequence (EVIVEPL → EVIYVQG) using primers VN4.1 to 4.4 (Table 2.1 & 2.2). Primers VN4.1 and VN4.2.2 were used to mutate residues PL to QG. The mutant product generated using primers VN 4.1 and 4.2.2 was used as the template for a second cycle of mutagenesis, where primers VN4.3 and 4.4 were elongated to mutate residues VE to YV. The protocol for mutagenesis was the same as the one used to generate TEV347GCIACLY, except that different primers were used, and a different template was used for the second round of mutagenesis. The fourth construct, called CanTEV354SCIACLY, was made by removing the leucine introduced before the glutamic acid in the construct XhoCanTEV354SCIACLY (Table 2.1). The protocol for mutagenesis follows the one used for TEV347GCIACLY, except primers VN5.1 and 5.2 were used instead.

Table 2.1. Summary of the four constructs that were made to test the insertion of the recognition site for the TEV protease at the two candidate sites

Construct	Candidate Site Tested	Primary Sequence Changes
TEV347GCIACLY	1	EVMKQG ↓ EVMYFQG
XhoCanTEV354SCIACLY	2	EVIVEPL ↓ LENLYFQS
ConTEV354GCIACLY	2	EVIVEPL ↓ EVIYVQG
CanTEV354SCIACLY	2	EVIVEPL ↓ LENLYFQS ↓ ENLYFQS

Table 2.2. The sets of forward and reverse primers utilized for mutagenesis.

	Sequence
TEV347GCIACLY Forward Mutagenic Primer VN2.1	GCC GCG TGT GCA GGA AGT TAT <u>GTA TTT</u> <u>T</u> ¹ CA AGG TGA AGT TAT TGT CG
TEV347GCIACLY Reverse Mutagenic Primer VN2.2	CGA CAA TAA CTT CAC CTT <u>GAA AAT ACA</u> TAA CTT CCT GCA CAC GCG GC
XhoCanTEV354SCIACLY PCR Forward Primer VN3.1	TTT TAT <u>CTC GAG AAC CTG TAT TTT CAG AGC</u> ATC CGC ACC ACG ATT AGT GAT
XhoCanTEV354SCIACLY PCR Reverse Primer VN3.2	ACA TCC <u>CTC GAG</u> ACC TTG TTT CAT AAC TTC CTG CA
ConTEV354GCIACLY Mutagenesis Forward Primer Set #1 VN4.1	GAA GTT ATT GTC GAA <u>CAA GGT</u> ATC CGC ACC ACG ATT AGT GAT G
ConTEV354GCIACLY Mutagenesis Reverse Primer Set #1 VN4.2.2	CAT CAC TAA TCG TGG TGC GGA <u>TAC CTT GTT</u> CGA CAA TAA CTT C
ConTEV354GCIACLY Mutagenesis Forward Primer Set #2 VN4.3	GAA GTT ATT <u>TAT GTT CAA GGT</u> ATC CGC ACC ACG ATT AGT GAT G
ConTEV354GCIACLY Mutagenesis Reverse Primer Set #2 VN4.4	CAT CAC TAA TCG TGG TGC GGA <u>TAC CTT GAA</u> <u>CAT AAA</u> TAA CTT C
CanTEV354SCIACLY Mutagenesis Forward Primer VN5.1	GGA AGT TAT GAA ACA AGG TGA GAA CCT GTA TTT TCA GAG CAT CCG C
CanTEV354SCIACLY Mutagenesis Reverse Primer VN5.2	GCG GAT GCT CTG AAA ATA CAG GTT CTC ACC TTG TTT CAT AAC TTC C
G347CIACLY Forward Primer VN 6.1	GTC TTA <u>CAT ATG</u> GGT GAA GTT ATT GTC G
G347CIACLY & V351CIACLY Reverse Primer VN6.2	GAC TCT <u>CTC GAG</u> TTT TTT TTC CGG CAC
V351CIACLY Forward Primer VN8.1	GTC TTA <u>CAT ATG</u> GTC GAA CCG CTG

¹ Underlined nucleotides indicate the mutations in the primers

Purification of TEV protease

The availability of the TEV protease gene in the pRK793 plasmid allowed for protein production and purification for proteolysis experiments (Tropea, Cherry, & Waugh, 2009). pRIL was used to increase the yield of TEV protease by increasing the supply of *argU* tRNA. pRK793 containing the TEV protease gene and pRIL were transformed into *E. coli* Inoue competent BL21(DE3) and grown overnight at 37°C on LB agar plates containing 100 µg/mL ampicillin and 30 µg/mL chloramphenicol. A colony was introduced into 100 mL LB containing the same antibiotics. The culture was incubated at 37°C overnight with shaking at 225 rpm. 25 mL of the overnight culture was transferred into 500 mL terrific broth (TB; 6 g tryptone, 12 g yeast extract, 0.4% v/v glycerol, 1.16 g KH₂PO₄, 6.27 g K₂HPO₄) with the same antibiotics. The TB culture was incubated at 37°C with shaking at 225 rpm until the optical density at 600 nm (A₆₀₀) reached 0.6. The temperature was lowered to 30°C and IPTG was added to a final concentration of 1 mM. The culture was shaken at 250 rpm for four hours and the cells were collected at 2500 x g (Sorvall RC 5B Plus centrifuge with a SLA3000 rotor) for 30 minutes. The supernatant was discarded, and the cell pellet was stored at -80°C.

Prior to nickel affinity chromatography, cells were thawed on ice. The wet cell pellet was resuspended in lysis buffer (50 mM sodium phosphate pH 8.0, 100 mM NaCl, 10% glycerol, and 25 mM imidazole), and lysed using a Branson Sonicator at medium setting with 30 seconds of sonication and 30 second pauses for five minutes. The soluble cell lysate was separated from cell debris by centrifugation at 11000 x g (SS34 rotor) for 30 minutes and applied onto a nickel-nitrilotriacetic acid (Ni-NTA) column (Qiagen) previously equilibrated with lysis buffer. Five column volumes of lysis buffer were used to remove proteins that did not bind tightly to the column. The proteins that bound to the Ni-NTA column were eluted using elution buffer (50 mM

sodium phosphate pH 8.0, 100 mM NaCl, 10% glycerol, and 400 mM imidazole). The eluted protein was precipitated using an equivalent volume of saturated ammonium sulfate (AS). The precipitated protein was stored at 4°C.

The protein precipitate was collected by centrifugation at 8000 x g (SS34 rotor) for 30 minutes. The precipitate was dissolved with desalting buffer (25 mM sodium phosphate pH 8.0, 200 mM NaCl, 10% glycerol, 2 mM EDTA, and 10 mM DTT), and loaded onto a Sephadex G-25 column (GE Healthcare) previously equilibrated with desalting buffer. Fractions that contained protein were combined and the absorbance at 280 nm (A_{280}) was measured. The protein concentration was calculated using the Beer-Lambert Law with a molar absorption coefficient of 1.19 mg⁻¹mL for TEV protease (Gasteiger et al., 2005). The protein was diluted to 1 mg/mL and flash frozen in liquid nitrogen for storage at -80°C.

Production and purification of the C-terminal portion of *CIACLY*

Inoue competent BL21(DE3) STARTM pLysS was transformed with one of the four mutant constructs and grown overnight on LB agar plates containing 30 µg/mL kanamycin and 35 µg/mL chloramphenicol. Single colonies were introduced into 100 mL LB containing the same antibiotics and incubated overnight at 37°C, with shaking at 225 rpm. 25 mL of the overnight culture was transferred into 500 mL TB containing the same antibiotics and the culture was shaken at 225 rpm and 37°C until the A_{600} reached 1.6. IPTG was added to a final concentration of 0.1 mM and the culture was incubated at 18°C overnight, with shaking at 250 rpm. The cells were collected by centrifugation at 2500 x g for 30 minutes and stored at -80°C.

Prior to Ni-NTA purification, the frozen wet pellet was thawed and resuspended in lysis buffer (50 mM imidazole, 50 mM sodium phosphate pH 8.00, 300 mM NaCl, 5% glycerol, and

10 mM 2-ME). Cells were lysed following the same procedure used during purification of TEV protease. Soluble cell lysate was isolated using centrifugation at 11000 x g and loaded onto a Ni-NTA column previously equilibrated with lysis buffer. The column was washed with five column volumes of lysis buffer to remove proteins that did not bind tightly. Bound proteins were removed with five column volumes of elution buffer (200 mM imidazole, 50 mM sodium phosphate pH 8.00, 300 mM NaCl, 5% glycerol and 10 mM 2-ME). The eluted proteins were precipitated using an equivalent volume of saturated AS. The protein precipitate was stored at 4°C.

The precipitated protein was collected at 8000 x g for 30 minutes. The precipitated protein samples from constructs 1 – 4 were dissolved with TEV protease buffer (50 mM Tris pH 8.00, 0.5 mM EDTA, and 1 mM DTT). TEV protease was added in a 1:100 w/w ratio (TEV protease:*ClACLY*). *ClACLY* was digested for 24 hours at 4°C.

The digestion mixture was diluted with low salt buffer for cation exchange (20 mM MES pH 6.3, 15% glycerol, and 10 mM 2-ME) to decrease the conductivity (mS/cm) to less than 7. The diluted sample was loaded onto a HiTrap SP HP column (GE Healthcare). A salt gradient was made by mixing the low salt and high salt (1 M KCl, 20 mM MES pH 6.3, 15% glycerol, and 10 mM 2-ME) buffers. Fractions with the highest A_{280} were diluted with low salt buffer for anion exchange (20 mM Tris pH 8.0, 15% glycerol, and 10 mM 2-ME) and the diluted sample was loaded onto a HiTrap Q HP column. A gradient was made by mixing low and high salt (1 M KCl, 20 mM Tris pH 8.0, 15% glycerol, and 10 mM 2-ME) buffers, and the eluted protein was dialyzed into 20 mM potassium phosphate pH 7.0 and 10 mM 2-ME using an Amicon® Ultra-30 centrifuge filter at 3082 x g (Allegra X-22 centrifuge). The A_{280} was measured and protein concentration was determined using a molar absorption coefficient of 1.14 mg⁻¹mL for the C-

terminal portion of *CIACLY* (Gasteiger et al., 2005). Protein samples with a concentration of 5 mg/mL were flash frozen in liquid nitrogen and stored at -80°C.

The C-terminal portion of *CIACLY* and TEV protease have similar molecular weights, 28.5 and 26 kDa, respectively, making it difficult to discern the two on an SDS-PAGE gel. A western blot was made to prove that the purified protein was not TEV protease. Monoclonal mouse anti-His antibody (Qiagen, Cat No. 34698) was used because TEV protease has a His-tag, whereas the C-terminal portion of *CIACLY* does not. The anti-His antibody should also bind to the small subunit of *CIACLY* because it has a His-tag. Samples from each purification step were run on a 14% SDS-PAGE gel and transferred onto a 0.45 µm nitrocellulose membrane (Bio-Rad) at 4°C for 16 hours. To visualize protein content, the membrane was stained with Ponceau S and washed with distilled water until bands were visible. The chemiluminescent detection protocol from the QIAexpress Detection and Assay Handbook was followed. The secondary antibody was goat anti-mouse HRP-conjugate (Thermofisher, Cat No. G-21040) and ECLTM prime western blotting detection reagent (AmershamTM) was used as the chemiluminescent reagent. The membrane was visualized using an AmershamTM Imager 600.

Purification of selenomethionine labelled protein for phase determination

Selenomethionine labelled protein was produced and purified to perform single-wavelength anomalous dispersion (SAD) experiments to obtain initial phase estimates. A single colony of *E. coli* BL21 DE3(STAR)TM pLysS containing TEV347G*CIACLY* was introduced into 100 mL LB containing 30 µg/mL kanamycin and 35 µg/mL chloramphenicol. The LB culture was grown overnight at 37°C, with shaking at 225 rpm. The cells from the LB culture were collected at 2500 x g for 30 minutes, discarding the supernatant. The pellet was

resuspended in 10 mL M9 minimal media (1x M9 salts (12.8 g Na_2HPO_4 , 3 g KH_2PO_4 , 0.5 g NaCl , 1 g NH_4Cl), 2 mM MgSO_4 , 0.4% glucose, 0.1 mM CaCl_2 ; (Sambrook & Green, 2012)) and transferred into 1 L M9 media with the same antibiotics as the overnight culture. The culture was incubated at 37°C, with shaking at 225 rpm, until the OD_{600} reached 0.6. To inhibit *in vivo* methionine production, six amino acids (100 mg each lysine, phenylalanine, threonine and 50 mg each isoleucine, leucine, valine) were added along with 60 mg selenomethionine (Sylvie, 1997). The culture was shaken for 15 minutes at 37°C, then for 15 minutes at room temperature, and induced with IPTG at a final concentration of 0.1 mM and incubated overnight at 18°C, with shaking at 250 rpm. The cells were collected using at 2500 x g for 30 minutes and stored at -80°C under argon.

The purification of the selenomethionine protein followed the procedure used for the native *ClACLY* with some modifications. To decrease oxidation of selenomethionine, all buffers were degassed, and containers were filled with argon to displace oxygen. EDTA was added to all buffers to prevent metal induced oxidation, except during nickel-affinity purification. A 1 mL HisTrapFF column (GE Healthcare) was used for the first step of purification instead of Ni-NTA due to lower yield of the selenomethionine labelled protein. Instead of precipitating the protein with AS, the protein that eluted was diluted with TEV protease cleavage buffer and digested for 24 hours at 4°C. Cation and anion exchange chromatography were completed within a day to minimize oxidation.

Crystallization and data collection

Vapour diffusion from hanging drops was used to crystallize the native and selenomethionine labelled C-terminal portion of *ClACLY*. Initial crystallization screens were conducted using the PEG/Ion kits from Hampton Research. Conditions were optimized to obtain

large, single crystals, which were mounted in Hampton CryoLoopsTM and submerged into a cryo-solution with the same components as the mother liquor along with 20% glycerol. The cryo-protected crystals were vitrified using a Cryostream (Oxford Cryosystems) then stored in liquid nitrogen. Crystals were sent to the Canadian Light Source (CLS) for remote data collection. The data were collected at the Canadian Macromolecular Crystallography Facility ID (CMCF-ID) beamline using a Pilatus3 S 6M detector and processed using XIA2 with DIALS (Winter, 2010).

Streak seeding was attempted to promote crystal formation in the crystallization conditions (Stura & Wilson, 1991). Sacrificial crystals were poked with a strand of hair and streaked across a crystallization drop containing a 1:1 v/v ratio of protein and well solution. The hair strand would have microscopic crystal fragments that are deposited into the drop during streaking. Solutions of native and selenomethionine protein were streak seeded and the crystals were sent to the Advanced Photon Source (APS) for remote data collection. Data were collected on the Industrial Macromolecular Crystallography Association Collaborative Access Team (IMCA-CAT) 17-ID beamline using a Pilatus 6M detector and processed using XIA2 with DIALS (Winter, 2010).

There were two methods employed to determine initial phases: SAD and Arcimboldo Shredder. The data from the selenomethionine protein crystals were used in the first attempt to obtain phase information. The C-terminal portion of ACLY was a good candidate for Arcimboldo because a high-resolution dataset was available, and ACLY was predicted to be mostly α -helical based on the structures of CS. The data from the unlabelled protein crystals were subsequently sent to the developers of Arcimboldo Shredder in Spain. A search model based on *Acetobacter aceti* CS (PDB ID: 2H12) was used to obtain initial phase estimates.

Fitting of the model to the electron density maps and refinement were performed using *Coot* (Emsley & Lohkamp, 2010) and PHENIX refine (Adams et al., 2010), respectively.

Production of the C-terminal portion of *CIACLY*

Two additional constructs, called G347*CIACLY* and V351*CIACLY*, were made to test if the C-terminal portion of *CIACLY* could be produced in the absence of the N-terminal portion. PCR was used to amplify fragments containing the gene to produce either residue 347 to 608 or residues 351 to 608 of the C-terminal portion of *CIACLY*. NdeI and XhoI sites would be introduced at the 5' and 3' ends, respectively, so that the amplified fragment could be ligated into pET42b(+). Once ligated, the product would have the start codon before residue 347 or 351 and contain a LEHHHHHHHH tag at the C-terminus. Primers VN11.1 and 11.3 were used to amplify a fragment containing codons for residues 347 to 608, whereas VN 11.2 and 11.3 would amplify a fragment containing codons for residues 351 to 608 (Table 2.2). The fragments were amplified following the protocol used to generate XhoCanTEV354*SCIACLY*. 10 U each of XhoI, DpnI, and NdeI were added after primer elongation, and the digestion mixture was incubated overnight at 37°C. The Monarch PCR & DNA Cleanup Kit (NEB) was used to purify the digested product prior to ligation. pET42b(+) was digested with 10 U each of XhoI and NdeI overnight at 37°C. The digested vector and fragment were separated on a 0.9% agarose gel, and the digested vector was extracted using the Monarch Gel Extraction Kit (NEB). The digested PCR product and vector were incubated with 5 U T4 DNA ligase overnight at 4°C and transformed into Inoue competent DH5α. To determine if the ligation was successful, purified plasmid was digested overnight at 37°C with 10 U each of XhoI and NdeI and the digestion products were analyzed on

a 0.9% agarose gel. Plasmids were sequenced at the University of Calgary DNA sequencing lab to determine if the ligation was successful and that there were no mutations.

G347*CI*ACY and V351*CI*ACY were transformed into BL21(DE3) and protein production followed the same protocol used for constructs one to four, with slight modifications. Following precipitation of the protein with saturated AS, the precipitates were dissolved in gel filtration buffer (50 mM Tris pH 8, 150 mM NaCl, 15% glycerol, and 10 mM 2-ME) instead of TEV protease buffer. The dissolved protein was loaded onto a Superdex 200 column (GE Healthcare) previously equilibrated with gel filtration buffer to determine if the protein was properly folded. Fractions that were expected to contain the C-terminal portion were pooled and purified through cation and anion exchange chromatography and concentrated using Amicon® Ultra-30 centrifuge filters.

Chapter 3: Results

Insertion of the recognition site for TEV protease into *ClACLY*

The first goal was to determine where to insert the recognition site for TEV protease in *ClACLY*. The potential locations for the recognition site were determined using PeptideCutter. Two candidate sites were found that would produce a ~28 kDa protein after proteolysis (Figure 2.1). Four constructs were made to test insertion of the TEV protease recognition site at the two sites.

TEV347G*ClACLY* tested the first candidate site (EVMKQG). There was a three-nucleotide insertion (TAT) to introduce a tyrosine between the methionine and lysine. The codon for the lysine residue was changed to phenylalanine (AAA → TTT). The primary sequence at the first candidate site was changed from EVMKQG to EVMYFQG, matching the consensus sequence EXXYXQ^{G/S}. The changes were confirmed by aligning the result from sequencing with the codon optimized genes for *ClACLY* (Figure 3.1). No other mutations were found in the two genes.

Figure 3.1. Sequence alignment of *aclA* and the sequencing result for TEV347G*ClACLY*. The chromatogram shows the quality of the sequencing read. The boxed nucleotides indicate where the mutation was introduced. The nucleotides shown correspond to the region of the gene that primers VN 2.1 and 2.2 annealed to *aclA*.

The second candidate site was replaced with the canonical recognition site in the XhoCanTEV354SClACLY. The primary sequence at the second candidate site was changed from EVIVEPL to LENLYFQS. The results from sequencing confirmed that the second candidate site was replaced with the canonical site (Figure 3.2). No other mutations were found in the two genes.

Figure 3.2. Sequence alignment of *aclA* and the sequencing results for XhoCanTEV354SClACLY. The chromatogram shows the quality of the sequencing read. The boxed nucleotides indicate where the mutation was introduced. Nucleotides shown correspond to the region of the gene to which the primers VN 3.1 and 3.2 annealed to *aclA*.

The ConTEV354GClACLY was generated to determine the effects of introducing the consensus sequence at the second candidate site, rather than the canonical sequence. The last four residues were mutated to match the consensus sequence (EVIVEPL → EVIYVQG) and the changes were confirmed from the results from sequencing (Figure 3.3).

Figure 3.3. Sequence alignment of *aclA* and the sequencing result for ConTEV354GClACLY. The chromatogram shows the quality of the sequencing read. The boxed nucleotides indicate where the mutation was introduced. Nucleotides shown correspond to the region of the gene to which the primers VN 4.1 – 4.4 annealed to *aclA*.

The CanTEV354*SCI*ACLY was generated to remove the leucine introduced from the XhoI site in the second construct. The primary sequence was changed from LENLYFQS to ENLYFQS. The removal of the leucine was confirmed from the results from sequencing (Figure 3.4).

Figure 3.4. Sequence alignment of *aclA* and the sequencing result for CanTEV354*SCI*ACLY. The chromatogram shows the quality of the sequencing read. The boxed nucleotides indicate where the mutation was introduced. Nucleotides shown corresponded to the region of the gene that the primers VN 5.1 and 5.2 annealed to *aclA*.

Purification of TEV protease

A two-step protocol was used to purify the TEV protease: Ni-NTA and Sephadex G-25 chromatography. The protein sample that eluted from the Ni-NTA column shows four bands on the SDS-PAGE gel (Figure 3.5). The band just above the 25 kDa marker is expected to be the TEV protease because its calculated molecular weight is 26 kDa (Gasteiger et al., 2005). The band below the 45 kDa marker could be a contaminate protein, whereas the bands below the 60 and 52 kDa could be a dimer and trimer, respectively, of TEV protease. Only one band just above the 25 kDa marker is visible after size exclusion chromatography (Figure 3.5).

Figure 3.5. 14% SDS-PAGE gel for the purification of TEV protease. Coomassie blue R-250 was used to visualize proteins. Lane one contained the BlueAQUA prestained protein ladder with select molecular weights shown. Lane two contained the proteins that eluted from the Ni-NTA column. Lane three contained the sample collected from Sephadex G-25 column.

Purification of the C-terminal portion of *ClACLY*

The purification of TEV347G*ClACLY* utilized three columns: Ni-NTA, cation exchange, and anion exchange. The theoretical molecular weights of the small and large subunits of *ClACLY* are 45 and 66 kDa (Gasteiger et al., 2005). TEV proteolysis will separate the large subunit into its N- and C-terminal portions that will have molecular weights of 37.5 and 28.5 kDa, respectively.

Ni-NTA was the first purification step used to purify the full-length *ClACLY* before TEV proteolysis. The protein sample that eluted from the Ni-NTA column showed two bands just above the 35 kDa and 60 kDa markers (Figure 3.6). These two bands were expected to be the small and large subunits of *ClACLY*.

Figure 3.6. 14% SDS-PAGE gel of sample from Ni-NTA for TEV347GCIACLY. Coomassie blue R-250 was used to visualize proteins. Lane one contained the BlueAQUA prestained protein ladder with select molecular weights shown. Lane two contained the protein that eluted from the Ni-NTA column.

Proteolysis for 24 hours using TEV protease was successful because the protein above the 60 kDa marker was cleaved to form two new proteins just above the 25 and 35 kDa protein markers (Figure 3.7). The proteins above the 25 and 35 kDa markers were expected to be the N- and C-terminal portions of the large subunit, respectively. The digested sample was loaded onto a HiTrap SP HP column (GE Healthcare) and multiple peaks were visible on the chromatogram (Figure 3.8). A sample from the first peak that eluted between 10 – 15 mS/cm showed one band just above the 25 kDa marker. The other peaks appeared to be different proportions of the small subunit interacting with the N-terminal portion of the large subunit.

Figure 3.7. 14% SDS-PAGE gel for cation exchange for the C-terminal portion of *CIACLY*. Coomassie blue R-250 was used to visualize proteins. Lane one contained the BlueAQUA prestained protein ladder with select molecular weights shown. Lane two was contained the digested protein sample after TEV proteolysis for 24 hours. Lane three was a sample from the first eluted peak from cation exchange (fraction A12, Figure 3.8). Lanes 3 to 7 were samples from the next 5 peaks of the chromatogram.

Figure 3.8. Cation exchange chromatogram for TEV347GCIACLY after TEV proteolysis. The chromatogram shows absorbance at 280 nm (mAU) as a function of elution volume (mL). The right y-axis shows conductivity (mS/cm).

Anion exchange was the final step for purification of the C-terminal portion of *CIACLY*. The fractions A11 to B1 were pooled from cation exchange because the band just above the 25 kDa marker was predicted to be the C-terminal portion of *CIACLY*. The pooled sample from cation exchange showed two peaks on the anion exchange chromatogram (Figure 3.9). The first peak showed no visible bands on the SDS-PAGE gel. The second peak showed one band just above the 25 kDa protein marker that corresponds to expected molecular weight of the C-terminal portion of *CIACLY*.

Figure 3.9. Anion exchange chromatogram for the C-terminal portion of *CIACLY*. The first peak from cation exchange (Figure 3.8) was loaded onto a HiTrap Q HP column. The left y-axis shows absorbance at 280 nm (mAU) as a function of elution volume (mL). The second y-axis shows conductivity (mS/cm). Fraction F3 shows one band just above the 25 kDa marker in lane two of the SDS-PAGE gel stained with Coomassie blue R-250.

Purification of XhoCanTEV354SClACLY was not as successful as TEV347GC lACLY.

The gel filtration chromatogram indicated that the protein produced from XhoCanTEV354SClACLY eluted at the void volume (Figure 3.10). Superposing the gel filtration chromatogram of the wild-type *ClACLY* onto XhoCanTEV354SClACLY shows that the elution profile is different for the two proteins. The difference between the wild-type and mutant *ClACLY* suggests that the latter aggregated. The aggregation could be due to the mutations at the second candidate causing the protein to fold incorrectly or could produce a new region on the protein that could form interactions with other proteins. The purification of the mutant proteins produced from constructs 2 to 4 were not pursued because the second candidate site were mutated in these constructs.

Figure 3.10. Superposition of the gel filtration chromatograms of wild-type *ClACLY* (red) and XhoCanTEV354SClACLY (blue) *ClACLY*.

A western blot was utilized to prove that the TEV protease was not purified. The proteins from each purification step for the C-terminal portion (lanes 2 to 5) and samples of purified TEV protease (lanes 6 to 10) were visible after staining with Ponceau S (Figure 3.11). Luminescence was detected from all the lanes containing TEV protease. Luminescence was also detected in lane 2 just above the 35 kDa marker as expected for the small subunit. Several bands were also detected in lane two above and below the band representing the small subunit. The higher molecular weight bands could be different oligomeric forms and the lower molecular weight bands could be degradation products of the small subunit. The two bands just above the 35 and 25 kDa markers in lane three correspond to the small subunit and TEV protease, respectively. The sample from cation exchange (lane four) has the same two bands as in lane three, with a lower concentration of TEV protease. TEV protease and the small subunit were not detected after anion exchange chromatography (lane five).

Figure 3.11. Western blot showing that the C-terminal portion of *CIACLY* was purified. Ponceau S was used to visualize protein content from each purification step. Monoclonal mouse tetra-His antibody (Qiagen) was used to bind proteins that had a His-tag. Goat antimouse HRP conjugate antibodies were used for chemiluminescent detection. Lane one contained the blueAQUA prestained protein ladder. Lane two contained protein sample after Ni-NTA purification. The third lane contained the Ni-NTA sample after 24-hour TEV proteolysis. Lanes four and five were samples from cation and anion exchange, respectively. Lane six contained 2.5 μ g TEV protease purified in May 2018. Lanes 7 to 10 contained 10, 5, 2.5 and 1.25 μ g of TEV protease purified in September 2018.

Crystallization of the C-terminal portion of *CIACLY*

TEV347G*CIACLY* was used in crystallization experiments. The protein solution contained 5 mg/mL, 20 mM potassium phosphate pH 7.0, 5 mM acetyl-CoA, 10 mM 2-ME, and 5 mM TCEP (tris(2-carboxyethyl)phosphine). Optimization of crystallization conditions led to a well solution containing 200 mM sodium acetate, 100 mM MES pH 6.3, and 20% w/v PEG 3350. A hanging drop made from a 1:1 v/v ratio of protein and well solution produced large, needle crystals that grew from one nucleation site (Figure 3.12). These crystals were sent to the CLS and a 1.9 Å resolution data set was obtained from one crystal (Figure 3.13 & Table 3.1).

Figure 3.12. Crystals of the C-terminal portion of *CIACLY*.

Figure 3.13. Diffraction pattern from a crystal of the C-terminal portion of CIACLY. Data was collected at the CMCF-ID beamline in the CLS.

Purification and crystallization of selenomethionine labelled *CIACLY*

To perform SAD experiments, TEV347G*CIACLY* was labelled with selenomethionine. Samples from each purification step were analyzed on a 14% SDS-PAGE gel (Figure 3.14). The sample that eluted from the Histrap FF column showed two bands whose molecular weights corresponded to the large and small subunit of *CIACLY*. Although there is a methionine in the recognition site, labelling with selenomethionine did not prevent TEV proteolysis. Both the cation and anion exchange chromatograms for the selenomethionine labelled protein resembled those from the unlabelled C-terminal portion of *CIACLY* (Figure 3.15 and 3.16).

Figure 3.14. 14 % SDS-PAGE for the purification of selenomethionine labelled *CIACLY*. Coomassie blue R-250 was used to visualize proteins. Lane 2 and 3 contained samples of the flow through and from the wash steps from Histrap FF, respectively. Lane 4 contained the protein sample that was eluted from the Histrap FF column. Lane 5 contained the digested protein sample after TEV proteolysis for 24 hours. Lane 6 shows the flow through of the HiTrap SP HP column, while lane 7 contained sample from fraction C11 from HiTrap SP HP (Figure 3.14). Lane 8 is a sample from the largest peak from HiTrap Q HP (Figure 3.15). Lane 9 contained the concentrated the selenomethionine labelled C-terminal portion of *CIACLY*.

Figure 3.15. Cation exchange chromatogram for the C-terminal portion of *ClACLY* labelled with selenomethionine. The chromatogram shows absorbance at 280 nm on the left y-axis as a function of elution volume (mL). The right y-axis shows conductivity (mS/cm).

Figure 3.16. Anion exchange chromatogram for the C-terminal portion of *ClACLY* labelled with selenomethionine. The chromatogram shows absorbance at 280 nm on the left y-axis as a function of elution volume (mL). The right y-axis shows conductivity (mS/cm).

Obtaining high-resolution data from crystals of the C-terminal portion of *CIACLY* labelled with selenomethionine would allow for initial phases to be determined using SAD. Crystal growth was not observed in the same conditions used to grow native crystals of the C-terminal portion of *CIACLY* that diffracted to high-resolution. The condition that produced the native crystal that diffracted to high-resolution (1.9 Å) contained MES and sodium acetate. Due to the presence of two buffers in the original condition, one was removed to observe the effects on crystal growth. Upon removal of MES, crystals readily grew using sodium acetate as the buffer. Long, needle crystals formed from more than one nucleation site (Figure 3.17). Crystals were grown with a well solution of 200 mM sodium acetate pH 5.2, and 17% w/v PEG 3350. These crystals were sent to the CLS and a 3.5 Å resolution data set was obtained (Figure 3.18 & Table 3.1).

Figure 3.17. Crystals of the C-terminal portion of *CIACLY* labelled with selenomethionine.

Figure 3.18. Diffraction pattern obtained from a crystal of selenomethionine C-terminal portion of *ClACLY*. Data was collected on the CMCF-ID beamline in the CLS.

In order to obtain higher resolution data for SAD experiments, streak seeding was used to induce nucleation of the selenomethionine labelled *CIACLY* in the original conditions that used MES buffer (Stura & Wilson, 1991). Streak seeding was tested using both the native and selenomethionine labelled C-terminal portion of *CIACLY* and both produced long, needle crystals from the same conditions (200 mM sodium acetate, 20% PEG 3350, and 100 mM MES pH 6.3) that produced the crystal that diffracted to high-resolution (Figure 3.19). The crystals grew from one nucleation site or from other crystals, instead of from multiple sites when sodium acetate buffer was used. These crystals were sent to Advanced Photon Source (APS) for data collection and high-resolution data (2 Å) were obtained only from a native crystal (Figure 3.20 & Table 3.0).

Figure 3.19. Crystals of the C-terminal portion of *CIACLY* produced from streak seeding of: a) native protein and b) selenomethionine labelled protein.

Figure 3.20. Diffraction pattern obtained from a crystal of the C-terminal portion of *ClACLY* grown from streak seeding. Data was collected in the 17-ID beamline of APS.

Table 3.0. Crystallographic statistics for the datasets obtained from crystals of the C-terminal portion of *CIACLY*.

Structure	VNC10	VNC15	VNCSe23
X-ray source	CLS-ID	APS 17-ID	CLS-ID
Wavelength (Å)	0.97949	1.00001	0.97949
Space group	P 3 ₂	P 1	P 3 ₂ 2 ₁ 1
Cell dimensions	82.19 Å, 82.19 Å, 151.200 Å 90°, 90°, 120°	81.76 Å, 82.07 Å, 157.19 Å 75.22°, 78.18°, 60.40°	124.184 Å, 124.184 Å, 199.209 Å 90°, 90°, 120°
Resolution range (High resolution)	75.59 – 1.89 (1.92 – 1.89)	61.92 – 2.06 (2.09 – 2.06)	107.55 – 3.51 (3.57 – 3.51)
R _{merge}	0.060 (0.581)	0.044 (0.651)	0.256 (0.895)
I/σ	13.5 (1.80)	5.40 (1.10)	6.30 (2.00)
CC _{1/2}	0.99 9(0.545)	0.998 (0.583)	0.997 (0.670)
Number of unique reflections	91455 (4474)	205109 (9688)	22879 (1134)
Multiplicity (High resolution)	4.90 (3.60)	1.70 (1.70)	9.50 (9.40)
Anomalous Multiplicity (High resolution)	2.3 (1.8)	0.6 (1.0)	4.9 (4.6)
Completeness % (High resolution)	99.97 (98.33)	96.82 (91.61)	100.00 (100.00)

Obtaining phase estimates for solving the structure

SAD and Arcimboldo Shredder were attempted to obtain initial phases. The 3.5 Å data from the selenomethionine crystals were utilized for SAD, however the anomalous signal from selenium was significant to only ~ 5 Å (Figure 3.21). Due to the low resolution, the locations of the selenium atoms could not be determined, and phasing was not possible with SAD.

Arcimboldo Shredder was used to determine initial phases for the 1.9 Å resolution data obtained from the native crystals. A modified model of *A. acetii* CS (PDB ID: 2H12) was used as the model to search for four molecules in the asymmetric unit.

Figure 3.21. $CC_{1/2}$ as a function of resolution for the selenomethionine data. The blue curve corresponds to the $CC_{1/2}$ of the non-anomalous signal, whereas the orange corresponds to the $CC_{1/2}$ of the anomalous signal. The dotted line indicates the significant $CC_{1/2}$ values based on $p = 0.01$.

The initial model of the C-terminal portion had 639 residues of the expected 1048 residues, in 26 fragments. Autobuild (Terwilliger et al., 2008) was used for automatic fitting, which was able to model an additional 174 residues and combined all the fragments into four chains. Another 207 residues were manually fitted using *Coot* for a total of 1020 residues modelled in four chains, covering 97% of the expected residues (Figure 3.22). Other molecules (e.g., phosphate and glycerol) were modelled along with a portion of acetyl-CoA in two of the protomers (Figure 3.23). Final statistics for the model are presented in Table 3.1. The protein coordinates were deposited into the PDB and given the designation 6NZY.

Table 3.1. Statistics for the model of the C-terminal portion of *CIACLY*

Resolution range (High resolution) (Å)	71.17 – 1.90 (1.92 – 1.89)
Number of reflections in working set	80395 (2215)
Number of reflections in test set	4287 (117)
R _{free}	0.1896
R _{work}	0.1666
Coordinate error	0.19 Å
Number of non-hydrogen atoms	Protein: 8074 Water: 308 Other: 182
Average B-factors	Protein: 51 Å ² Water: 45 Å ² Other: 78 Å ²
Ramachandran plot	Favoured regions: 98.54% Allowed regions: 1.46% Outliers: 0%
rmsd from ideal values	Bonds: 0.006 Å Angles: 0.667°

Figure 3.22. Ribbon diagram of the C-terminal portion of *CIACLY*.
Termini are labelled, and the portion of acetyl-CoA is shown as a stick model. This image was generated using UCSF Chimera (Pettersen et al., 2004).

Figure 3.23. Electron density and modelled portion of acetyl-CoA bound. The phosphoadenosine group was the only portion of acetyl-CoA that could be modelled into the electron density. The $2F_o-F_c$ electron density map is coloured blue and was contoured at 1 rmsd. The F_o-F_c electron density map is coloured green or red and was contoured at 3 rmsd. This image was generated using *Coot* (Emsley & Lohkamp, 2010).

Production and purification of the C-terminal portion of *CIACLY*

The structure of the C-terminal portion revealed potential sites that could serve as the N-terminus for the truncated C-terminal portion of *CIACLY*. A G347*CIACLY* and V351*CIACLY* were made to test if *CIACLY* could be truncated to residues 347 to 608 or residues 351 to 608 to isolate the C-terminal portion. Only V351*CIACLY* was confirmed by sequencing (Figure 3.24). The deletion of 2289 nucleotides correspond to the N-terminal portion that was removed so that the start codon is before V351.

The gel filtration chromatograms for V351C/ACLY showed a large peak at ~80 mL, with a small shoulder before the large peak (Figures 3.25). Fractions C3 – C5 were loaded onto a cation exchange column and the elution showed one major peak and a larger shoulder (Figures 3.26). Fractions F6 – 9 were loaded onto an anion exchange column and one peak was present in the chromatogram. Amicon® Ultra-30 centrifuge filter was used to concentrate the protein to 8.9 mg/mL in 200 mM potassium chloride, 20 mM potassium phosphate pH 7.0, and 10 mM 2-ME.

Figure 3.25. Gel filtration chromatogram of V351CIACLY.

Figure 3.26. Cation exchange chromatogram of V351CIACLY.

Figure 3.27. Anion exchange chromatogram for V351CIACLY. Fractions F6 - 9 (Figure 3.26) from cation exchange chromatography were loaded onto the anion exchange column.

Figure 3.28. 14% SDS-PAGE gel of samples from each purification step for V351CIACLY. Coomassie blue R-250 was used to visualize proteins. Lane one contains the blueAQUA prestained protein ladder with select molecular weights labelled. Lane two is a sample of protein that eluted from Ni-NTA. Lane three contains protein sample from the dissolved AS precipitate. Lane four contains sample of the major peak (C4) from gel filtration. Lanes five and six contain fractions F8 and G6 from cation exchange, respectively. Lane seven contains a sample from the major peak from anion exchange.

Crystallization screens were made to see if V351C/ACLY could be crystallized. 0.28 mM D- and L-malic acid were added along with 5 mM acetyl-CoA into the protein solution to see if the protein could adopt a closed conformation to trap acetyl-CoA. V351C/ACLY was also crystallized in the absence of malic acid and acetyl-CoA. Crystal growth was observed in several conditions with similar long needle crystal morphology compared to TEV347G/ACLY, although the crystals do not appear to form from one nucleation site like TEV347G/ACLY (Figures 3.29 – 3.31). However, there were tetragonal, bipyramidal crystals grown in some conditions with D-malate and acetyl-CoA that were not seen during the crystallization of TEV347G/ACLY. No diffraction data were collected from these crystals yet.

Figure 3.29. Crystal of V351C/ACLY grown with a well solution of 200 mM lithium citrate tribasic pH 8.4 and 20% w/v PEG 3350. No acetyl-CoA or malate were added.

Figure 3.30 Crystals of V351C/ACLY grown with a well solution of 200 mM ammonium tartrate dibasic pH 6.6 and 20% w/v PEG 3350. D-malate was added along with acetyl-CoA.

Figure 3.31. Crystals of V351C/ACLY grown with a well solution of 200 mM ammonium tartrate dibasic pH 6.6 and 20% w/v PEG 3350. L-malate was added along with acetyl-CoA.

Chapter 4: Discussion

Purification of the C-terminal portion of *ClACLY*

TEV347G*ClACLY* was the only one out of the four constructs that produced protein that did not aggregate. One challenge during purification was separation of the N- and C-terminal portions after TEV proteolysis due to the interactions that they would have with each other. This issue was not present when chymotrypsin was used to cleave the protein because it may have digested the N-terminal portion. The N- and C-terminal portions were separated during cation exchange, possibly because the interactions between the two were disrupted at pH 6.1. However, the western blot revealed that there were trace amounts of the N-terminal portion and TEV protease in the first peak that eluted from the cation exchange column (Figure 3.11). Anion exchange chromatography was able to separate the residual N-terminal portion and TEV protease from the C-terminal portion. The protein sample after anion exchange chromatography showed no luminescence for the N-terminal portion or TEV protease on the western blot. The C-terminal portion of *ClACLY* was successfully purified after being cleaved by TEV protease.

Structural determination of the C-terminal portion of *ClACLY*

A 1.9 Å dataset was obtained from a crystal of the C-terminal portion. However, crystal nucleation was rarely seen upon replicating the condition. Streak seeding was attempted for the native and selenomethionine proteins, but only the native crystal grown from streak seeding diffracted to high resolution. The space group (P 1) and unit cell dimensions of the native crystal grown from streak seeding were different from the P 3₂ crystal grown without streak seeding (Table 3.0). Microseed matrix screening (MMS) could explain why streak seeding produced a different crystal form. In MMS, crystal seeds from an initial crystallization experiment are

transferred into conditions that differ from the initial experiment. The goal of MMS is to improve the quality of the crystals by growing new crystals in different conditions. Both MMS and streak seeding share similarity due to their introduction of seeds into conditions to promote growth of crystals. The use of MMS could lead to new crystals that possess a different space group than the original seeds, but may also produce the same crystal form (D'Arcy, Bergfors, Cowan-Jacob, & Marsh, 2014). Therefore, it is possible that the crystals grown from streak seeding were influenced by the initial seeds.

Initial attempts at obtaining phases using molecular replacement in *Phenix* did not produce a clear solution using several CS's as the search model. SAD was also unsuccessful, due to the inability to grow selenomethionine crystals that diffract to high-resolution. Arcimboldo Shredder was able to obtain initial phase estimates. The C-terminal portion of *CIACLY* was a good candidate for Arcimboldo Shredder because of the high-resolution data, and because most of the structure was predicted to be α -helical based on the structures of CS. The periodicity of α -helices leads to strong signal, aiding in determining their locations (Millán et al., 2018). Initial phases were obtained by Dr. Isabel Usón's group in Spain. A tetrameric protein model was built with these initial phases, showing that the C-terminal portion of *CIACLY* contains the CoA-binding site and is the oligomerization site for *ACLY*.

CoA-binding site of *CIACLY*

Two of the four predicted CoA-binding sites had visible electron density for the phosphoadenosine group of acetyl-CoA. Residues 485 – 489 form a hydrophobic pocket, where the adenine group of acetyl-CoA is inserted (Figure 4.1). L538 also forms hydrophobic interactions with the adenine ring. Only one long-distance hydrogen bond (3.2 Å) is present

between N1 of the adenine ring and the amide nitrogen of V486 (Figure 4.2). There is one ionic interaction between K534 and the 3' phosphate group (2.8 Å) (Figure 4.3). The presence of only one hydrogen bond suggests that hydrophobic interactions are the primary force in the binding of the adenine ring of acetyl-CoA/CoA. There is no electron density for the pantothenate, β -mercaptoethylamine, and acetyl groups.

Figure 4.1. The adenine group of acetyl-CoA bound into a hydrophobic pocket composed of residues 485 to 489 in the C-terminal portion of *ClACLY*.

Figure 4.2. The long-distance hydrogen bonding interaction between N1 of the adenine ring and the amide nitrogen of V486.

Figure 4.3. The ionic interaction between 3' phosphate of acetyl-CoA and K534. The blue skeleton represents the $2F_o - F_c$ electron density map contoured at 1 rmsd.

One explanation for the lack of electron density for the pantothenate, β -mercaptoethylamine, and acetyl groups is that the C-terminal portion of *Cl*ACYL is in an open conformation, preventing residues from another protomer to form interactions with the pantothenate group. The lack of interactions allows the other groups of acetyl-CoA to adopt multiple conformations in the crystals. In CS, the presence of oxaloacetate (or similar organic acids) induces a closed conformation, allowing residues from another subunit to interact with the pantothenate group of CoA (Remington et al., 1982). The lack of an organic acid bound in the active site prevents the C-terminal portion of *Cl*ACYL from adopting a closed conformation.

Recently, structures of the C-terminal portions of *Cl*ACYL and *h*ACYL were published (Verschueren et al., 2019). There were two structures (PDB IDs: 6QCL & 6HXN) that showed different degrees of openness of the C-terminal portion of *Cl*ACYL. Superposition of my structure of the C-terminal portion onto 6QCL and 6HXN confirms that my structure adopts an open conformation (Figure 4.4). 6QCL is in a closed conformation, where acetyl-CoA and malate were modelled in the active site. There is a 15° rotation of the CoA-binding domain (DynDom3D, (Girdlestone & Hayward, 2016)) between my structure and 6QCL. The shift in the CoA-binding domain moves the phosphoadenosine group 5.2 Å measured from N6 of the adenine rings of the two acetyl-CoA molecules. 6HXN is in an open conformation compared to 6QCL. My structure is like 6HXN because only the phosphoadenosine group of acetyl-CoA was modelled, whereas 6QCL has the entire acetyl-CoA and malate modelled in the active site. 6HXN and my structure cannot adopt a closed conformation because there is no organic acid in the active site. The three structures support the idea that the mechanism for inducing a closed conformation is similar between the C-terminal portion of ACYL and CS and requires the presence of an organic acid.

Figure 4.4. Superposition of three structures of the C-terminal portion of *CIACLY* (6NZY (blue), 6HXN (red), and 6QCL (green)) to show the degree of openness. 6QCL represents the closed conformation, whereas 6HXN and 6NZY show different degrees of openness of the CoA-binding site. a) View of the CoA-binding domain that shows the degree of openness of the three structures. There is a 15° rotational difference between the most open *CIACLY* and the closed *CIACLY*. b) 90° rotation revealing the CoA-binding site. The acetyl-CoA molecule is shifted 5.2 Å between the closed, and most open form of *CIACLY*.

Although there was no evidence for bound acetyl-CoA in the active site, glycerol and phosphate were modelled into the active sites of the four protomers. The superposition of the different structures of the C-terminal portion revealed that there is a bound glycerol in the active site of my structure where the acetyl group of acetyl-CoA binds in 6QCL (Figure 4.5). The amide nitrogen of G451 forms hydrogen bonds with the glycerol in the structure I solved, whereas it interacts with the oxygen of the thioester group of acetyl-CoA in 6QCL. The phosphate modelled in the active site of my structure superposes onto malate found in 6QCL. There is no evidence that glycerol or phosphate have a biological role in the active site of the C-terminal portion.

Figure 4.5. Superposition of two structures of the C-terminal portion of *ClACLY* (6NZY, gold and 6QCL, green) showing the molecules bound at the active sites of both structures.

Comparison of the CoA-binding sites of ClACLY, Nter hACLY, and SCS

A region of the N-terminal portion of ACLY shares sequence similarity to the CoA-binding site of SCS. However, there was no evidence that CoA could bind in the N-terminal portion of *hACLY* in the absence of the C-terminal portion (Bazilevsky et al., 2019; Sun et al., 2010). The structure of the C-terminal portion shows that the phosphoadenosine group of CoA binds in the C-terminal portion of ACLY. The flexibility of the CoA molecule could allow the pantothenate, and β -mercaptoethylamine groups to swing to the N-terminal portion to form citryl-CoA, which swings back to the C-terminal portion to be cleaved into acetyl-CoA and oxaloacetate. A structure of full-length *hACLY* (PDB ID: 6HXX) provides evidence that CoA can swing into the N-terminal portion. In 6HXX, the phosphoadenosine group is bound in the C-terminal portion and the thiol end is bound in the N-terminal portion (Figure 4.6). The structure of full-length *hACLY* suggests that the C-terminal portion is required for the N-terminal portion to bind the pantothenate and β -mercaptoethylamine groups, which explains why CoA does not bind to the N-terminal portion in the absence of the C-terminal portion. Therefore, the CoA-binding site was found in the C-terminal portion of structure, and the evidence for CoA swinging to the N-terminal portion is seen in the full-length structure of *hACLY*.

Figure 4.6. Full-length *hACLY* showing CoA-binding between N- and C-terminal portions. The phosphoadenosine group of CoA is bound in the C-terminal portion, whereas the thiol end is bound in the N-terminal portion near the citrate-binding site.

The published structures of the C-terminal portion of ACLY from human, and *C. limicola* show that there are four CoA-binding sites (Verschuere et al., 2019). However, a published structure of full-length *ClACLY* (PDB ID: 6HXJ) found eight CoA molecules bound. Four molecules of CoA were bound in the C-terminal portions, but an additional four molecules were bound in the N-terminal portion. The hypothesis presented was that *ClACLY* is an ancestral ACLY that retains the ability to bind the phosphoadenosine group in the N-terminal portion that aligns with the CoA-binding site of SCS. The presence of four CoA-binding sites in the N-terminal portion differs from a previous result showing that CoA does not bind onto the N-terminal portion of *hACLY*. I would argue against this hypothesis because the four CoA molecules bound in the N-terminal portion could also be artifacts of crystallization. Previous isothermal calorimetry (ITC) experiments with full-length *ClACLY* found a binding stoichiometry of two for CoA, which is odd because four CoA-binding sites were expected (Singh, 2016). Future work is required to determine how many CoA-binding sites *ClACLY* has.

Oligomerization of the C-terminal portion of *Cl*ACLY

The structure of the C-terminal portion of *Cl*ACLY revealed a tetrameric protein, proving that ACLY oligomerizes in this portion. The quaternary structure of the C-terminal portion of *Cl*ACLY has not been seen before in CS, which exists as either a dimer or a hexamer formed from three dimers (Nguyen et al., 2001).

There are four central α -helices (residues 435 – 445) that form the major interface between all four protomers of the C-terminal portion of *Cl*ACLY (Figure 4.7). Three of the 10 residues (Q437, S440, T444) are hydrophilic and form hydrogen bonding interactions with ordered water molecules within the central core. Each of the six ordered waters can form hydrogen-bonding interactions between two residues from different α -helices. The ordered waters within the central core are seen in other structures of the C-terminal portion of ACLY. The four central helices also form hydrogen bonding interactions with glycerol. There are four other α -helices (residues 420 – 430) that are adjacent to the central core that have hydrophobic residues at their interfaces. Both hydrogen-bonding and hydrophobic interactions are present at the central interface between the four protomers.

Figure 4.7. The interactions of the α -helices at the central interface of the tetramer. The four central α -helices are primarily hydrophilic residues and are enclosed in the black box. Two of the α -helices adjacent to the central four core α -helices shows hydrophobic residue interaction are enclosed in the red box. The red circles represent water molecules.

The protomers also interact with each other at residues near their termini. A two stranded β -sheet forms between C-terminal residues (594 to 596) of one protomer and N-terminal residues (355 to 357) of another protomer (Figure 3.22). The interaction of the N- and C-terminal residues between protomers has only been seen in structures of CS from thermophiles, suggesting that these interactions help stabilize the protein (Russell, Ferguson, Hough, Danson, & Taylor, 1997).

Superposition of CS and ACLY reveals structural similarities between the two. When one protomer of the C-terminal portion of *Cl*ACLY and CS are superposed, some α -helices are

missing in the C-terminal portion that are present in CS (Figure 4.8). However, the missing α -helices superpose onto α -helices of another protomer of *ClACLY*. The donation of α -helices from the adjacent protomer shows that the protomers may require another protomer to fold properly.

A sequence alignment can be made based on the structural superposition of the C-terminal portion of *ClACLY* and CS. This type of alignment can be used to infer evolutionary relationships between proteins based on their structures, rather than solely on their primary sequences. The structural alignment shows that one protomer of *ClACLY* aligns with the N-terminal portion of CS, whereas another protomer aligns with the C-terminal portion of CS (Figure 4.9). The structural alignment of *ClACLY* with CS suggests that there is an evolutionary relationship between the two. The structural alignment supports the hypothesis that CS is related to the C-terminal portion of *ACLY*.

Figure 4.8. Structural superposition of chicken citrate synthase (yellow) with two protomers of the C-terminal portion of *ClACLY* (blue and red). A) Superposition of one protomer of *ClACLY* with CS. B) Superposition of the two protomers of the C-terminal portion of *ClACLY* shows that the α -helices from the red protomer superpose onto α -helices that are missing in the blue protomer.

Both the initial sequence alignment and the updated structural alignment show that D543 and H491 of *ClACLY* have equivalent conserved residues in CS (Figure 1.2 & 4.9). H274 of chicken CS aligns with F450 of *ClACLY*. Studies of chicken CS found that H274 forms a hydrogen bonding interaction with the oxygen of the thioester group of acetyl-CoA (Karpusas et al., 1990). The sidechain of F450 in *ClACLY* cannot form hydrogen bonds, thus the interaction with the thioester group could be lost. However, superposition of *ClACLY* and CS shows a peptide flip in *ClACLY* between F450 and G451 compared to the equivalent residues in CS (Figure 4.10). The flipped peptide in *ClACLY* allows for hydrogen bonding interactions to form from the amide nitrogen. A published structure of the C-terminal portion of *ClACLY* (PDB ID: 6QCL) shows amide nitrogen of G451 forming a hydrogen bond with acetyl-CoA

Figure 4.10. Superposition of chicken CS (green) and the C-terminal portion of *ClACLY*. There is a flipped peptide bond in *ClACLY* between residues F450 and G451 compared to the equivalent residues in CS.

(PDB ID: 6QCL). Although the histidine in CS is changed to phenylalanine in ACLY, the interaction with the thioester group of acetyl-CoA is not lost in ACLY.

Catalytic mechanism of ACLY

The structures of the C-terminal portion of *ClACLY* and full-length *hACLY* reveal new information that supports the proposed catalytic mechanism. Previous structures of the N-terminal portion of *hACLY* showed the location of the nucleotide- and citrate-binding sites, along with the phosphohistidine loop (Hu et al., 2017; Sun et al., 2010, 2011). ATP binds in site II, where the γ -phosphate transfers onto a histidine residue (H760 in *hACLY*) in step I of the proposed catalytic mechanism. The phosphohistidine residue swings approximately 40 Å to the citrate-binding site, where the phosphoryl group gets transferred onto citrate during step II of the reaction. The next two steps of the reaction mechanism were elucidated from the structures of the C-terminal portion, along with the full-length ACLY. The structure of the full-length *hACLY* supports step III, where the thiol end of CoA binds in the N-terminal portion, while the

phosphoadenosine group is bound in the C-terminal portion (Figure 4.6). Once in the N-terminal portion, the thiol group receives the citryl group to form citryl-CoA. During step III, the C-terminal portion of ACLY is in an open conformation, allowing the citryl end of citryl-CoA to bind into the C-terminal portion for step IV. During the final reaction step, citryl-CoA is cleaved in the C-terminal portion to form acetyl-CoA and oxaloacetate.

Evolution of ACLY and CS

Previous studies have presented phylogenetic relationships between SCS, CS, and ACLY (Fatland et al., 2002). It was hypothesized that the small subunit of *Cl*ACLY shares a common ancestor with the β -subunit of SCS. The N- and C-terminal portions of the large subunit of *Cl*ACLY were hypothesized to have originated from the α -subunit of SCS and CS, respectively. However, a new hypothesis was recently presented based on the structures of citryl-CoA synthetase and citryl-CoA lyase (Verschueren et al., 2019). The current hypothesis is that ACLY resulted from a fusion of a citryl-CoA synthetase and a citryl-CoA lyase, which correspond to the N- and C-terminal portions of ACLY, respectively. A structure of a citryl-CoA lyase shows that it forms a similar tetrameric structure to the C-terminal portion of ACLY. CS was hypothesized to have originated from ACLY through deletion of domains, where the structural superposition of the two proteins shows that CS is missing domains that are present in ACLY (Figure 4.8). Instead of ACLY originating from CS, it was proposed that CS originates from ACLY. Therefore, ACLY emerged from fusion of citryl-CoA synthetase and lyase, whereas CS formed through deletion of domains from ACLY.

Production of the C-terminal portion of *CIACLY*

One goal was to insert the recognition site for TEV protease in an exposed region on *CIACLY*. TEVG347*CIACLY* was successfully cleaved at the first candidate site, indicating that the first candidate site is accessible to TEV protease. Assuming that the conformation seen in the structure is similar to conformation that is cleaved by TEV protease, the published structure of full-length *CIACLY* shows that the first candidate site (residues 342 to 347) is exposed, giving TEV protease access to this site to cleave *CIACLY* (Figure 4.11). The second candidate site (residues 348 to 354) is buried in the full-length structure of *CIACLY*, preventing TEV protease from accessing the site. However, mutations at the second candidate caused the protein to aggregate, either because the protein cannot properly fold, or the mutations introduced a new interaction site where proteins can cluster. The first candidate site was the better location for introducing the TEV protease site compared to the second candidate site because the site was accessible, and the mutant protein did not aggregate.

Previous attempts were made to produce just the C-terminal portion composed of residues 355 to 608 along with an N-terminal His-tag. The truncation of the C-terminal portion to residues 355 to 608 was not successful because the protein did not behave well in crystallization experiments. The initial thought was that the C-terminal portion needed to be produced with N-terminal portion to properly fold. However, one other possibility is that the N-terminus was chosen incorrectly. Residue 355 is part of a β -sheet formed between N-terminal residues 355 to 357 and C-terminal residues 594 to 596 of another protomer. The C-terminal portion of *CIACLY* may need to be produced as a longer fragment than residues 355 to 608 to be properly folded. To test if the C-terminal portion of *CIACLY* could be produced alone, two constructs were made that comprised residues 347 to 608 or 351 to 608.

Figure 4.11. Structure of full-length *ClACLY* shows that the first candidate site (Grey, EVMKQG) is exposed. The second candidate site mutated in constructs two to four is coloured orange. The residues of the first candidate are represented as sticks.

V351*ClACLY* was successfully produced and purified, showing that the C-terminal portion can be produced alone without the N-terminal portion. The purified protein was crystallized and there were multiple conditions that produced crystals. Diffraction data have not been measured yet for these crystals, but the ability to produce the C-terminal portion of *ACLY* simplifies future experiments by not having to utilize TEV protease.

Chapter 5: Conclusion

The overall goal was to determine the structure of the C-terminal portion of *Cl*ACYL. The structure would allow us to interpret the reaction mechanism, find the location of the CoA-binding site, and to determine if the C-terminal portion of ACYL was the oligomerization site. Proteolytic cleavage of *Cl*ACYL using TEV protease was able to separate the N- and C-terminal portions from each other. The C-terminal portion of *Cl*ACYL was purified through Ni-NTA, cation exchange, and anion exchange chromatography. SDS-PAGE and western blot were used to verify that the C-terminal portion of *Cl*ACYL had been purified. The C-terminal portion of *Cl*ACYL was co-crystallized with acetyl-CoA. A crystal of the C-terminal portion was grown in sodium acetate with MES buffer and 20% w/v PEG 3350 that diffracted to 1.9 Å. SAD experiments were attempted to obtain phase information to solve the structure, but no high-resolution data could be obtained. Streak seeding was utilized to introduce crystal seeds of the selenomethionine protein into the conditions that produced the native crystal that diffracted to high-resolution. However, no high-resolution dataset was obtained from the selenomethionine crystals produced from streak seeding. Arcimboldo Shredder was utilized to obtain phases after unsuccessful SAD experiments. The phases were obtained from a modified structure of *A.aceti* CS. The structure of the C-terminal portion of *Cl*ACYL revealed a tetrameric protein, as well as revealing the CoA-binding site.

Electron density for the phosphoadenosine group of acetyl-CoA was seen in C-terminal portion. However, the pantothenate, β -mercaptoethylamine and acetyl groups had no visible electron density. The flexibility of the groups can allow the pantothenate, and β -mercaptoethylamine groups to swing between the N- and C-terminal portions to facilitate catalysis. The groups may not bind in the active site because the structure is in an open

conformation since there is no organic acid bound in the active site to induce a closed conformation.

Comparisons of current structures of ACLY and CS reveals details about the catalytic mechanism and possible explanations as to why the C-terminal portion of ACLY forms a tetramer. A structural superposition of CS and the C-terminal portion of *Cl*ACLY shows that the missing α -helices of one protomer from *Cl*ACLY are donated from another protomer. The structure of the C-terminal portion also shows that the C-terminal residues of each protomer forms a two-stranded β -sheet with the N-terminal residues of another protomer. The intimate contacts that the protomers have with each other are reasons as to why the C-terminal portion forms a tetramer.

A structural alignment was generated for chicken CS and *Cl*ACLY. One of the catalytic residues, H274, in chicken CS aligns with F450 in *Cl*ACLY. The structure of the C-terminal portion of *Cl*ACLY reveals that there is a peptide bond flip between F450 and G451, exposing the backbone nitrogen to the active site. The exposed backbone nitrogen is vital because a published structure of the C-terminal portion of *Cl*ACLY bound to acetyl-CoA and malate shows that the backbone nitrogen interacts with the oxygen of the thioester of acetyl-CoA. The flipped peptide functions similarly to H274 in chicken CS, assuming an acid-base reaction is not required.

The final part of my project was to determine if the C-terminal portion of *Cl*ACLY can be produced without the N-terminal portion. Preliminary results indicate that the C-terminal portion of *Cl*ACLY can be produced, purified, and crystallized. However, no diffraction data has been collected yet for the crystals of 351V*Cl*ACLY.

Chapter 6: Future studies

The presence of glycerol and phosphate in the active site would not be ideal if the experimental goals were to see substrate or products bound. Future purification protocols should avoid using glycerol or phosphates if possible. Since glycerol was used as a cryo-protectant, other molecules should be used instead, such as PEG 400. By removing the possibility of other molecules binding in the active site, it may allow us to co-crystallize the C-terminal portion with substrates, analogues, or drugs more readily. The lack of organic acid could be one reason why the structure appears to be in an open conformation. Organic acids such as citrate or malate should be included in the crystallization experiments if the goal is to obtain the closed structure of ACLY. 351VC/ACLY was crystallized in the presence of D- or L-malate to see if a close structure could be obtained.

TEV347GC/ACLY can be utilized in further research for drug design targeting cancer and/or obesity. The production of a protein that can be cleaved with TEV protease can allow researchers to look at specifically the N- or C-terminal portions. The purification protocol can be optimized to isolate the N- or C-terminal portions depending on which portion researchers are interested in. Further work needs to be done to determine if C/ACLY contains eight CoA-binding sites as suggested in the literature. Different crystal forms could be pursued into order to see if the four CoA molecules bound in the N-terminal portion of C/ACLY are artifacts of crystallization. Biophysical techniques could also be used to determine the binding stoichiometry of CoA. ITC should be conducted again to see if the binding stoichiometry of two for CoA is reproducible.

References

- Adams, P. D., Pavel, V., Chen, V. B., Ian, W., Echols, N., Moriarty, N. W., ... Thomas, C. (2010). research papers PHENIX : a comprehensive Python-based system for macromolecular structure solution research papers, *D66*, 213–221.
<https://doi.org/10.1107/S0907444909052925>
- Alter, G. M., Zhi, W., Nemeth, P., Srere, P. A., & Evans, C. T. (1990). Mutation of essential catalytic residues in pig citrate synthase. *Biochemistry*, *29*(33), 7557–7563.
<https://doi.org/10.1021/bi00485a003>
- Bauer, D. E., Hatzivassiliou, G., Zhao, F., Andreadis, C., & Thompson, C. B. (2005). ATP citrate lyase is an important component of cell growth and transformation. *Oncogene*, *24*(41), 6314–6322. <https://doi.org/10.1038/sj.onc.1208773>
- Bazilevsky, G. A., Affronti, H. C., Wei, X., Campbell, S. L., Wellen, K. E., & Marmorstein, R. (2019). ATP-citrate lyase multimerization is required for coenzyme-A substrate binding and catalysis. *Journal of Biological Chemistry*, *4*, jbc.RA118.006685.
<https://doi.org/10.1074/jbc.RA118.006685>
- Berwick, D. C., Hers, I., Heesom, K. J., Moule, K., & Tavaré, J. M. (2002). The Identification of ATP-citrate Lyase as a Protein Kinase B (Akt) Substrate in Primary Adipocytes. *Journal of Biological Chemistry*, *277*(37), 33895–33900. <https://doi.org/10.1074/jbc.m204681200>
- Boulton, C. A., & Ratledge, C. (1981). Correlation of Lipid Accumulation in Yeasts with Possession of ATP: Citrate Lyase. *Microbiology*, *127*(1), 169–176.
<https://doi.org/10.1099/00221287-127-1-169>
- Carrington, J. C., & Dougherty, W. G. (1988). A Viral Cleavage Site Cassette : Identification of Amino Acid Sequences Required for Tobacco Etch Virus Polyprotein. *Proceedings of the*

National Academy of Sciences, 85(10), 3391–3395.

D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W., & Marsh, M. (2014). Microseed matrix screening for optimization in protein crystallization: what have we learned? *Acta Crystallographica Section F Structural Biology Communications*, 70(9), 1117–1126.

<https://doi.org/10.1107/s2053230x14015507>

Emsley, P., & Lohkamp, B. (2010). research papers Features and development of Coot research papers, 486–501. <https://doi.org/10.1107/S0907444910007493>

Evans, M. C. W., Buchanan, B. B., & Arnon, D. I. (1966). A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Biochemistry*, 55(4), 928–934.

<https://doi.org/10.1073/pnas.55.4.928>

Fatland, B. L., Ke, J., Anderson, M. D., Mentzen, W. I., Cui, L. W., Allred, C. C., ... Wurtele, E. S. (2002). Molecular Characterization of a Heteromeric ATP-Citrate Lyase That Generates Cytosolic. *Plant Physiology*, 130, 740–756. <https://doi.org/10.1104/pp.008110>

Fuchs, G., Stupperich, E., & Gerolf, E. (1980). Autotrophic CO₂ fixation in *Chlorobium limicola*. Evidence against the operation of the Calvin cycle in growing cells. *Archives of Microbiology*, 128(1), 56–63. <https://doi.org/10.1007/BF00422306>

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (pp. 571–607). Humana Press. <https://doi.org/10.1385/1592598900>

Girdlestone, C., & Hayward, S. (2016). The DynDom3D Webserver for the Analysis of Domain Movements in Multimeric Proteins. *Journal of Computational Biology*, 23(1), 21–26. <https://doi.org/10.1089/cmb.2015.0143>

Hatzivassiliou, G., Zhao, F., Bauer, D. E., Andreadis, C., Shaw, A. N., Dhanak, D., ...

- Thompson, C. B. (2005). ATP citrate lyase inhibition can suppress tumor cell growth. *Cancer Cell*, 8(4), 311–321. <https://doi.org/10.1016/j.ccr.2005.09.008>
- Hu, J., Komakula, A., & Fraser, M. E. (2017). Binding of hydroxycitrate to human ATP-citrate lyase. *Acta Crystallographica Section D: Structural Biology*, 73(8), 660–671. <https://doi.org/10.1107/S2059798317009871>
- Johannsson, C.-J., & Pettersson, G. (1977). Substrate-Inhibition by Acetyl-CoA in the Condensation Reaction between Oxaloacetate and Acetyl-CoA Catalyzed by Citrate Synthase from Pig Heart. *Biochimica et Biophysica Acta*, 484, 208–215.
- Kanao, T., Fukui, T., Atomi, H., & Imanaka, T. (2001). ATP-citrate lyase from the green sulfur bacterium *Chlorobium limicola* is a heteromeric enzyme composed of two distinct gene products. *European Journal of Biochemistry*, 268(6), 1670–1678. <https://doi.org/10.1046/j.1432-1327.2001.02034.x>
- Kaplan, R. S., Mayor, J. A., & Wood, D. O. (1993). The mitochondrial tricarboxylate transport protein. cDNA cloning, primary structure, and comparison with other mitochondrial transport proteins. *Journal of Biological Chemistry*, 268(18), 13682–13690.
- Karpusas, M., Branchaud, B., & Remington, S. J. (1990). Proposed Mechanism for the Condensation Reaction of Citrate Synthase: 1.9-Å Structure of the Ternary Complex with Oxaloacetate and Carboxymethyl Coenzyme A. *Biochemistry*, 29(9), 2213–2219. <https://doi.org/10.1021/bi00461a002>
- Millán, C., Sammito, M. D., McCoy, A. J., Nascimento, A. F. Z., Petrillo, G., Oeffner, R. D., ... Usón, I. (2018). Exploiting distant homologues for phasing through the generation of compact fragments, local fold refinement and partial solution combination. *Acta Crystallographica Section D Structural Biology*, 74(4), 290–304.

<https://doi.org/10.1107/s2059798318001365>

- Mulholland, A. J., & Richards, W. G. (1997). Acetyl-CoA enolization in citrate synthase: A quantum mechanical/molecular mechanical (QM/MM) study. *Proteins: Structure, Function and Genetics*, 27(1), 9–25.
- Nguyen, N. T., Maurus, R., Stokell, D. J., Ayed, A., Duckworth, H. W., & Brayer, G. D. (2001). Comparative analysis of folding and substrate binding sites between regulated hexameric type II citrate synthases and unregulated dimeric type I enzymes. *Biochemistry*, 40(44), 13177–13187. <https://doi.org/10.1021/bi010408o>
- Nissen, T. L., Schulze, U., Nielsen, J., & Villadsen, J. (1997). Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*. *Microbiology*, 143(1), 203–218. <https://doi.org/10.1099/00221287-143-1-203>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera — A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Potapova, I. ., El-Maghrabi, E., Doronin, S. ., & Benjamin, W. . (2000). Phosphorylation of Recombinant Human ATP:Citrate Lyase by cAMP-Dependent Protein Kinase Abolishes Homotropic Allosteric Regulation of the Enzyme by Citrate and Increases the Enzyme Activity. Allosteric Activation of ATP:Citrate Lyase by Phosphorylated Sug. *Biochemistry*.
- Remington, S., Wiegand, G., & Huber, R. (1982). Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 Å resolution. *Journal of Molecular Biology*, 158(1), 111–152. [https://doi.org/10.1016/0022-2836\(82\)90452-1](https://doi.org/10.1016/0022-2836(82)90452-1)
- Russell, R. J. M., Ferguson, J. M. C., Hough, D. W., Danson, M. J., & Taylor, G. L. (1997). The

crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution. *Biochemistry*, 36(33), 9983–9994.

<https://doi.org/10.1021/bi9705321>

Sambrook, J., & Green, M. R. (2012). *Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Protocols* (4th ed.). Cold Harbor Spring Laboratory Press.

<https://doi.org/10.3724/SP.J.1141.2012.01075>

Sánchez, L. B., Galperin, M. Y., & Müller, M. (2000). Acetyl-CoA synthetase from the amitochondriate eukaryote *Giardia lamblia* belongs to the newly recognized superfamily of acyl-CoA synthetases (nucleoside diphosphate-forming). *Journal of Biological Chemistry*, 275(8), 5794–5803. <https://doi.org/10.1074/jbc.275.8.5794>

Singh, N. (2016). *Examining the Biological Significance of Histidine and Aspartic Acid Residues in the Reaction Mechanism of ATP-Citrate Lyase by DEGREE OF Master of Science*.

Srere, P. A. (1959). The Citrate Cleavage Enzyme. *The Journal of Biological Chemistry*, 234(10), 2544–2547.

Srere, P. A. (1969). Citrate Synthase. *Methods in Enzymology*, 13, 3–11.

<https://doi.org/10.1107/S1600536810014716>

Stura, E. A., & Wilson, I. A. (1991). Applications of the streak seeding technique in protein crystallization. *Journal of Crystal Growth*, 110, 270–282.

Sun, T., Hayakawa, K., Bateman, K. S., & Fraser, M. E. (2010). Identification of the citrate-binding site of human ATP-citrate lyase using X-ray crystallography. *Journal of Biological Chemistry*, 285(35), 27418–27428. <https://doi.org/10.1074/jbc.M109.078667>

Sun, T., Hayakawa, K., & Fraser, M. E. (2011). ADP-Mg²⁺ bound to the ATP-grasp domain of ATP - Citrate lyase. *Acta Crystallographica Section F: Structural Biology and*

Crystallization Communications, 67(10), 1168–1172.

<https://doi.org/10.1107/S1744309111028363>

Sylvie, D. (1997). Preparation of Selenomethionyl Proteins for Phase Determination. *Methods in Enzymology*, 276(1989), 523–530. <https://doi.org/10.1557/PROC-1165-M09-01>

Tao, R., Li, T., Zhou, X., Guan, K.-L., Xiong, Y., & Lei, Q.-Y. (2014). Acetylation Stabilizes ATP-Citrate Lyase to Promote Lipid Biosynthesis and Tumor Growth. *Molecular Cell*, 51(4), 506–518. <https://doi.org/10.1016/j.molcel.2013.07.002>.Acetylation

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V, Moriarty, N. W., Zwart, P. H., Hung, L.-W., ... Adams, P. D. (2008). Iterative model building , structure refinement and density modification with the PHENIX AutoBuild wizard research papers. *Acta Crystallographica Section D: Structural Biology*, D64, 61–69. <https://doi.org/10.1107/S090744490705024X>

Tropea, J. E., Cherry, S., & Waugh, D. S. (2009). *Expression and Purification of Soluble His6-Tagged TEV Protease*. Humana Press.

Tucek, S. (1967). Subcellular Distribution of Acetyl-CoA Synthetase, ATP Citrate Lyase, Citrate Synthase, Choline Acetyltrasnferase, Fumarate Hydrate and Lactate Dehydrogenase in Mammalian Brain Tissue. *Journal of Neurochemistry*, 14, 531–545.

van der Kamp, M. W., Zurek, J., Manby, F. R., Harvey, J. N., & Mulholland, A. J. (2010). Testing High-Level QM / MM Methods for Modeling Enzyme Reactions : Acetyl-CoA Deprotonation in Citrate Synthase. *The Journal of Physical Chemistry B*, 114, 11303–11314.

Verschueren, K. H. G., Blanchet, C., Felix, J., Dansercoer, A., De Vos, D., Bloch, Y., ...

Verstraete, K. (2019). Structure of ATP citrate lyase and the origin of citrate synthase in the Krebs cycle. *Nature*, 568, 571–575. <https://doi.org/10.1038/s41586-019-1095-5>

- Wahlund, T. M., & Tabita, F. R. (1997). The reductive tricarboxylic acid cycle of carbon dioxide assimilation: Initial studies and purification of ATP-citrate lyase from the green sulfur bacterium *Chlorobium tepidum*. *Journal of Bacteriology*, 179(15), 4859–4867.
<https://doi.org/10.1128/jb.179.15.4859-4867.1997>
- Wang, W., & Malcolm, B. A. (1999). Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. *BioTechniques*, 26(4), 680–682.
- Wellen, K. E., Hatzivassiliou, G., Bui, T. V., Sachdeva, U. M., Cross, J. R., & Thompson, C. B. (2009). ATP-Citrate Lyase Links Cellular Metabolism to Histone Acetylation. *Science*, 324(5930), 1076–1080. <https://doi.org/10.1126/science.1164097>
- Wiegand, G., & Remington, S. J. (1986). Citrate Synthase: Structure, Control, and Mechanism. *Annual Review of Biophysics*, 15, 97–117.
- Winter, G. (2010). xia2 : an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, 2000, 186–190.
<https://doi.org/10.1107/S0021889809045701>
- Yang, W., & Druekhammer, D. G. (2003). Computational Study of the Citrate Synthase Catalyzed Deprotonation of Acetyl-Coenzyme A and Fluoroacetyl-Coenzyme A Demonstration of a Layered Quantum Mechanical Approach. *Journal of Physical Chemistry B*, 107, 5986–5994.