STRUCTURAL AND BIOCHEMICAL STUDIES ON ENZYMES INVOLVED IN NATURAL PRODUCT BIOSYNTHESIS AND CELLULAR RESPIRATION

BY

DAVID S. PARK

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biophysics and Computational Biology in the Graduate College of the University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Satish K. Nair, Chair, Director of Research
Professor Robert Gennis
Assistant Professor Hong Jin
Professor Claudio Grosman
ABSTRACT

Natural products have recently become an interest in modern research. This is due to their synthesis mainly by living organisms and their natural biosynthesis and abundance. Many of these compounds have been found to have pharmaceutical and industrial applications, and have been used as the basis for derivatives that are already being used for large scale production. Despite initial availability, there now is an urgent need to isolate and produce new and more effective natural products. The good news is we are still continuing to discover new sources. Natural products will still be an important source for industrial and pharmaceutical development. One focus improving them is to understand the biosynthetic pathways of some of these natural products. Elucidating these biosynthetic pathways will allow us to be improve efficacy and allow for rational design of new and better compounds. The studies that were conducted in this dissertation focuses on structural studies of three enzymes involved with natural product synthesis.

The most widely distributed component of brown algae is alginate, a cell wall polysaccharide composed of polymeric blocks of α-(1,4) O-linked β-D-mannuronate (M) and its C-5 epimer α-L-guluronate (G). Biochemical characterization of the *Saccharophagus degredans* polysaccharide lyase-17 family enzyme (Alg17c) demonstrates that the enzyme can depolymerize alginate di-, tri-, and tetra-saccharide into monosaccharides, providing the necessary precursor for ethanol fermentation. In chapter 1 we present several crystal structures of the PL-17 enzyme Alg7c from *S. degredans*, including the wild-type enzymes, respective mutants, and a co-crystal structure of one of the variants in complex with an alginate trisaccharide. Structure-guided
analysis of several active site variants, allow for the identification of residues that are critical for substrate recognition and for the bases of the exolytic reaction mechanism of PL-17 enzymes.

Plumbemycin is an antibacterial phosphonate peptide from Streptomyces plumbeus containing APPA (2-amino-5-phosphono-3-pentenoic acid), a compound that inhibits the threonine synthesis. The current pathway is unknown, but a proposed pathway has been formulated based on the pathway of a structurally similar APPA containing antifungal compound rhizocticin. There is a putative class II aldolase, PluG, in the biosynthetic pathway which is proposed to catalyze an aldol reaction between oxaloacetate and phosphonoacetaldehyde to form a reactive hydroxyl intermediate (2-keto-4-hydroxy-5-phosphonopentanoic acid), with a chance to form an irreversible, non-enzymatic dehydration reaction which forms an off pathway end product. In chapter 2, we obtained a high resolution structure of a PluG ortholog from Saccharothrix aerocolonigenes. The data from this structure shows an uncommon c-terminal domain, which we hypothesize to promote the formation of a complex with a subsequent enzyme from the plumbemycin biosynthetic pathway. We also propose that a hydrophobic tunnel forms between the two active sites of the complex and propose that this tunnel shuttles the reactive hydroxyl intermediate, thus preventing the formation of the non-enzymatic end product.

Cellular respiration is a critical components of a cellular organism’s ability to adapt and survive in its environment, as well as produce potentially useful natural products. It is not coincidence that many researchers are focusing on elucidating pathways involved with this process, especially that of the electron transport chain. \textit{B. subtilis} is also one of the most heavily studied
and utilized organisms, with its genome completely sequenced and easily found in soil, gastrointestinal tracts of humans, and other readily accessible sources. However chemistry involved with the electron transport chain still has many questions that need to be answered.

We carried out structural studies on the $aa_3$-menaquinone oxidase, which has high sequence similarity with ubiquinone oxidase from $E. coli$. We believe structural information on the $aa_3$ can elucidate the mystery of the quinone/menaquinone binding site, which is critical for the q cycle in the electron transport chain of bacteria.
ACKNOWLEDGEMENTS

I would like to thank all those who helped me through my PhD career. It would not have been possible without your support and encouragement.

I have deep gratitude for my PhD advisor, Professor Satish Nair, who willingly and generously took me into his lab and put me under his advisement to finish my degree. He gave me a place to continue my degree after the difficult passing of my late advisor Dr. Robert Clegg and helped me get back on my feet. During the past three years in the Nair Laboratory, I had the chance to work with some great people who helped me learn the techniques and recommendations to succeed. I had the opportunity to work on projects that were exciting and gave me much to experience in being a successful graduate student. It was thanks to his encouragement, excellent advise, great patience, and willingness to work with me that I am indebted to.

I would like to thank the Nair lab members that I have had the chance to work with, including Jon Chekan, Yue Hao, Shi-Hui Dong, Dillon Cogan, Phil Oliveres, Paola Estrada, Joesph Ly, Nektaria Petronikolou, and Jillian Chekan. I give great thanks for the wonderful advice I received from you as well as the patience to teach me the techniques in X-ray crystallography. For giving me support and encouragement during the difficult times of my PhD and all the times of fellowship we had. I would also like to thank my undergrad assistant Alex Bacalar for assisting me with my experiments and helping my achieve success on my research.

I would like to thank my committee members Professor Robert Gennis, Professor Hong Jin, and Professor Claudio Grosman for their advice and intellectual input on my graduate research. I
would also like to thank Professors Wilfred van der Donk, Professor Huimin Zhao, Professor William Metcalf, Dr. Kou-San Ju, Dr. Zedu Huang, and Dr. Despina Bougioukou for the opportunity to work with them on collaborative projects and for having helpful discussions with me.

I would also like to thank the late Professor Robert Clegg and the fellow Clegg Lab members John Eichorst and Kevin Teng, who were there to help me acclimate to graduate student life and advised me through my first three years as a PhD student. It was also through them that I had the excellent experience of teaching my first discussions and classes. This experience was invaluable to me and I am forever thankful.

I am truly grateful for support and guidance from Cindy Dodds, who tirelessly worked to keep not just me, but the other students and professors in the department together and organized. I would also like to thank Dr. Chester Brown, JP Swigart, Alejandra Stenger, Elizabeth Good, and Nicholas Kirchner for the opportunity to experience teaching and leading discussion with undergraduate students. Also the patience for the number questions I had to ask and for helping me find my love of teaching and instruction.

I would also like to thank the Center for Biophysics and Computational Biology for giving me place to pursue my graduate research and for the generous financial support I received in the Gregario Weber Fellowship my first year.

Most of all, I am indebted to my family, who were a constant source of support and encouragement through all the good and tough times through my graduate career. I sincerely
thank my parents for keeping up and checking up on me on a weekly basis. Their emotional support has been invaluable to my career and personal development.
# TABLE OF CONTENTS

CHAPTER 1: STRUCTURE OF A PL17 FAMILY ALGINATE LYASE DEMONSTRATES FUNCTIONAL SIMILARITIES AMONG DEPOLYMERASES .................................................................1

1.1: GENERAL INTRODUCTION ...........................................................................................................1

1.2: MATERIALS AND METHODS ........................................................................................................31

1.3: RESULTS AND DISCUSSION .......................................................................................................46

1.4: CONCLUDING REMARKS ..............................................................................................................53

CHAPTER 2: BIOCHEMICAL CHARACTERIZATION OF PLUMBEMYCIN BIOSYNTHESIS ..........65

2.1: GENERAL INTRODUCTION ..........................................................................................................65

2.2: MATERIALS AND METHODS ......................................................................................................72

2.3: RESULTS AND DISCUSSION ......................................................................................................78

2.4: CONCLUDING REMARKS ............................................................................................................82

CHAPTER 3: STRUCTURAL STUDIES ON AA$_3$-MENAQUINOL OXIDASE FROM BACILLUS SUBTILIS ......................................................................................................................103

3.1: GENERAL INTRODUCTION .........................................................................................................103

3.2: MATERIALS AND METHODS .....................................................................................................117

3.3: RESULTS AND DISCUSSION ......................................................................................................118

3.4: CONCLUDING REMARKS ............................................................................................................120

BIBLIOGRAPHY .................................................................................................................................123
CHAPTER 1: STRUCTURE OF A PL17 FAMILY ALGINATE LYASE DEMONSTRATES FUNCTIONAL SIMILARITIES AMONG DEPOLYMERASES

1.1 GENERAL INTRODUCTION

Introduction

The increasing cost and limited reserve of fossil fuels have resulted in massive research efforts aimed at the discovery of efficient sources of renewable energy. The identification of the dwindling sources of these fuels has been a hot topic in news lately, with the attention on alternative fuel sources being shifted to sources that have the potential for self-sustainability. Multiple sources have been identified as having that potential and can be categorized into four “generations” of biofuels. The first generation focuses directly from fermenting ethanol from already available food stocks and sources. The main conversion process involves the use of transesterification, where the isolation of fatty acid methyl esters (FAME) or biodiesel from vegetable oil (Xu, Qiao et al. 2003). Sources mainly are mainly sugar/starch containing crops, such as grain, sugar cane, wheat, fruits, rice, corn, etc., and cellulosic biomass, such as pine, wood, cedar, etc (Li, Zhang et al. 2003). The advantage of this generation is that the infrastructure for crop growth, harvesting, and processing are already available.

Second generation biofuels are primarily focused on lignocellulosic, hemicellulosic, or lignin materials produced from non-food crops such as the inedible portions of generation one biofuels (e.g. corn stover/husk), perennial crops (switchgrass, myscathus, Indian grass, etc.), and even components of woody biomass from wood (Li, Zhang et al. 2003). Two main process of processing second generation biofuels exists: one involving thermochemical processing, such as pyrolysis and gasification, and biochemical treatment, revolving around pretreatment too
speed up hydrolysis and separation of lignin, hemicellulose, and cellulose (Li, Zhang et al. 2003). The advantages for second generation biofuels are as follows. They are perennial, have relatively low fertilizer needs, and have a very high net yield while not impacting food processing. Third generation biofuels if focused on algae biomass, with focus on the production of ethanol from extraction of the lipids of the cell wall (Zhang and Chen 2003). Methods for conversion can be utilized from current conversion methods, such as transesterification and combustion. Advantages are the costs involved with space and operations, as algae can grow in open air ponds, bioreactors, or can grow anywhere it is warm. They can also help digest municipal waste if needed. Fourth Generation biofuels is the focus of specially engineered plants or biomass that are optimized for higher energy yields and easier conversion, while limiting the amount of CO\textsubscript{2} produced and released.

However, the biofuel sources are not without their problems. First generation biofuels, as already heard in the news and media, have the issue of “food vs fuel”, where the majority of biofuels directly from food crops result in the increasing cost of food prices in the last couple of years. Also the yields generated from these sources are generally not high enough to be counter the argument that it is a cost effective solution to self-sustainability. Second generation biofuels were initially developed to overcome the limitation of the first generations, with higher yields and competitive cost investment; however, the issues stems from the expensive process of ethanol conversion and the time required for the crops to reach a desired harvest density. Third generation biofuels have disadvantages involving expensive input costs needed, such as large amounts of water, nitrogen, and phosphorus to grow. Also the implementation of new production infrastructure would be necessary.
Over the past decade, massive resources have been invested into the optimization of methods for extracting ethanolic precursors from terrestrial biomass, such as already available food stocks to second generation bioenergy feed stocks perennial grasses (Martin 2010). Because of the already available infrastructure, much preference has been shifted in terms of research and funding toward optimizing generation one and two production. Comparatively, less effort has been directed towards the bioprocessing of marine algal biomass, which has shown potential as a replacement for land-based plant materials, but was shown to have a high startup cost (Hu, Ji et al. 2015). However, the potential untapped source of bio-ethanol cannot be denied. Red, brown, and green algae contain an abundance of polysaccharide-derived sugars and, owing to tolerance of changes in climate or soil conditions, represent the most sustainable source for fermentative biofuels (Markou, Angelidaki et al. 2012). Moreover, the use of large-scale of algae can be carried out in controlled environments, such as photo bioreactors, allows for the high yields without the concern for land requirements as those with energy crops. The unique carbohydrate composition of algae presents a formative challenge for bioprocessing (Blanch 2012). As a consequence, there has been an increasing interest in the discovery and development of enzymes that can depolymerize algal polysaccharides into fermentable monomeric sugars.

The most widely distributed component of brown algae is alginic acid or more commonly referred to as alginate, and is refined from brown algae or seaweed. The composition of alginate is a copolymer of β-D-mannuronic acid (M) and α-D-gluronic acid (G), which can be configured into homopolymeric G blocks and M blocks, as well as heteropolymeric MG blocks. Found abundantly in the cell walls of brown algae, alginate has found numerous industrial
applications in food, textile printing, and pharmaceuticals, as an additive, stabilizer, and gelling agent. It has even been used in the dental industry as a component of dental impression material, mainly as the widely used product sodium alginate. Research on alginates has shifted toward its potential as a biofuel source of ethanol as a way to combat the ever growing concerns of the limitation present in the current, popular feed stocks, such as corn and sugar cane (Wargacki, Leonard et al. 2012). Development of new techniques for improving yields to scalable quantities have yield promising results.

The process of lipid extraction from the walls is one that follows mostly the fermentation methods of the first generation biofuels. The following diagram shows the basic process of biodiesel from algae (figure 1) (Wei, Zhang et al. 2003). Lipid extraction starts with harvesting algae, dried, and spun down to ensure the highest dry weight. Then the lipids are extracted using mechanical separation with pretreatment of hexanes. This is followed by enzyme treatment with distillation and ethanol purification. Afterwards, there is a further removal of pretreatment components as well as hydrolysis, which is finally followed by biodiesel production. The goal of optimizing the current technology has become a key priority for researchers. One of the more expensive components of the process is the utilization of enzymes engineered to improve saccharification and fermentation of the product. To improve this, researchers have been looking into enzymes found naturally in the environment to breakdown complex polysaccharides. One group of enzymes that have garnered much interested is alginate lyases that target the abundant brown algae such as seaweed.
Alginate lyases are enzymes that catalyze the degradation of alginate heteropolymers through β-elimination mechanism. The result of this mechanism involves a product containing a double bond 4-deoxy-L-erythro-hex-4-enopyranosyluronic acid at the non-reducing end (Li, Dong et al. 2011). Alginate lyases can be classified into three categories based on substrate specificity (Li, Dong et al. 2011). First type is guluronate lyases, which target specifically G blocks of alginate. The second type is mannuronate lyases, which target specifically M blocks of alginate. And finally, the third type is bifunctional alginate lyases, which target both G and M blocks. Additionally, alginate lyases have been grouped into three categories based on molecular weight: Class 1 at around 20-35kDa, class 2 at around 40 kDa, and class 3 at >60 kDa (Osawa, Matsubara et al. 2005).

Alginate lyases can be isolated from a number of marine sources such as fungi, various marine algae, and bacteria. They also have a diverse range of uses, such as production of algal protoplasts and alginate structure studies. Structural studies of these alginate lysases not only give bioinformatical insight to conserved pathways amongst source organisms, but also potential roads toward rational design of these enzymes.

Alginate Lyase Sources
Alginate lyases can be isolated from a variety of sources, mainly marine based organisms and a variety of terrestrial microorganisms. These alginate lyases are categorized based on the preference of substrate, which is shown to be dependent on the environment that the organism is found, though it appears that the majority of lyases prefer poly (M) substrates or both poly (M) and poly (G); however, alginate lyases that prefer only poly (G) substrates have been studied and characterized recently.

Marine sources of alginate lyases can include those from various marine algae and invertebrates. Algae sources include various types of brown algae, such as *Colpomenia sinuosa*, *Endarachne binghamiae*, *Eisenia bicyclis*, and *Sargassum sagamianum* (Shiraiwa 1975). Invertebrates that contain alginate lyases include a sea hare *Alpysia depilans* (Boyen, Kloareg et al. 1990), pink abalone *Haliotis corrugate* (Nakada and Sweeny 1967), a type of sea snail *Haliotis tuberculata* (Boyen, 1990 (Boyen, Kloareg et al. 1990), and another type of sea snail *Turbo corunutus* (Tsuyoshi Muramatsua 1977). Alginate lyase sources from terrestrial organisms include a variety of gram-positive and gram-negative bacteria. Gram negative bacteria include *P. aeruginosa* (Boyd, Ghosh et al. 1993), *P. maltophilia* (Wong, Preston et al. 2000), *P. syringae* pv. *Syringae* (Wong, Preston et al. 2000), *Spingomonuas* sp (Yonemoto, Murata et al. 1991). Examples of gram-positive bacteria include *Bacillus* sp. ATB-1015 (Nakagawa, Ozaki et al. 1998), *B. circulans* (Hansen and Nakamura 1985), *Cl. Alginolyticum* (Kaiser, Le Borgne et al. 1968), and *Corynebacterium* sp. ALY-1 ((Matsubara, Iwasaki et al. 1998)

**Enzymatic Mechanism**
As mentioned previously, alginate is a polysaccharide consisting of blocks of polyM, polyG, or polyMG linked by a glycosidic 1→4 O-linkage (figure 2) (Agulhon, Markova et al. 2012).

Alginate lyases catalyzes the β-elimination reaction that targets the glycosidic 1→4 O-linkages of monomeric units of alginate. As a result, a double bond is formed between carbons 4 and 5 from the subsequently eliminated 4-O-glycosidic bond. This yields can yield either 4-deoxy-L-erythro-hex-4-enopyranosyluronic acid at the nonreducing terminal end of the product regardless of whether the by syn or anti elimination (Figure 3) (Lombard, Bernard et al. 2010).

The β-elimination reaction acted on alginate by the enzymes involves three main events (Wong, Preston et al. 2000). One is a charged residue of the active site catalyzing a general base removal of the C5 proton and/or another residue (or environmental solvent) acting as a proton donor. The usual amino acid residues are usually aspartic acid, histidine, lysine, glutamic acid, and cysteine. Another event is the neutralization of the negative charge on the carboxyl anion, usually involving a residue such as lysine. And finally, the transition of electrons from the carboxyl group to the bond between C4 and C5 results in a double bond formation and elimination of the 4-O-glycosidic bond.

**Substrate Preference: Endolytic vs Exolytic**

Substrate specificity for alginate lyases is shown to be influenced by the environment the source organism is living in. These lyases are structurally tailored to accommodate the various alginate sources that are readily available for use. As mentioned before, many of the initially discovered alginate lyases had a preference for polyM (poly(M) lyases [(1→4)-β-mannuronan
lyase}) substrate or even polyMG substrate, but there has been an increasing amount of enzymes preferring polyG (poly(G) lyases [(1→4)-α-L-guluronan lyase]) being discovered and studied. Out of these enzymes a majority of them appear to be endolytic or preference toward longer polysaccharides. There are several instances where the lyase has a preference for exolytic activity, forgoing longer chain oligosaccharides and catalyzing the removal of terminal alginate monomeric/dimeric units from the terminal ends.

Endolytic alginate lyases catalyze a β-elimination reaction of a 1-4 glycosidic linkage within a larger alginate polysaccharide. The result is an oligomer that contains a 4-deoxy-L-erythro-hex-4-enepyranosyluronate at the non-reducing end (Kim, Lee et al. 2011). Activity can be multifaceted, exhibiting polyM, polyG, or polyMG activity (Figure 4). Oligomeric alginate is taken up by exolytic alginate lyases, which catalyze a β-elimination reaction resulting in the formation of an unsaturated monosaccharide. This monosaccharide then is non-enzymatically converted to a 4-deoxy-L-erythro-hexoseulose uronic acid (DEH) (Takase, Ochiai et al. 2010) (Figure 4).

It is important that even though an alginate lyase is categorized as an M specific or G specific lyase, in many cases the enzyme is shown to have moderate to low activity against the other homopolymer. Investigation into the multiple substrate activity has led to the reclassification of enzymes characterized previous as having a single polymer preference. In most cases this was discovered from the fact that protocols responsible for generating “pure” polyM or polyG usually results in a small amount of the other homopolymer being present. However this is not
a bad thing as a lyase that equally targets both polyM and polyG substrates can help contribute the cost reduction of biodiesel production.

Alginate Lyase Structures

Alginate lyases fall under the superfamily of polysaccharide lyases, which are divided up into twenty families based on the hydrophobic cluster analysis and sequence similarity comparisons. More specifically, they belong to seven of these families: PL-5, PL-6, PL-7, PL-14, PL-15, PL-17, and PL-18 (Cantarel, Coutinho et al. 2009).

Polysaccharide Lyase Family 5 (PL5)

Alginate lyases from the polysaccharide lyase family 5, such as the known characterized A1-Ill from the *Sphingomonas* species A1, are shown to have endolytic activity, resulting in the production of di- and trisaccharides (Hye-Jin Yoon 1999). According to the CAZy database, these enzymes are structurally characterized by an (α/α)₆ barrel. PL5 alginate lyases demonstrate a stronger preference for polyM alginates, as in the case of algL from *Azotobacter chroococcum mcd 1* (Pecina, Pascual et al. 1999) and a homologous protein from *Azotobacter vinelandii* (Ertesvag, Hoidal et al. 1998). However, some of the PL5 lyases have been shown to act on poly alginate as well, such as the case of Smlt1473 from *Stenotrophomonas maltophilia (strain K279a)* (MacDonald and Berger 2014).
Structures of PL5 lyases can be best represented by that of Al-III as originally characterized by the Mikami group at Kyoto University (Hye-Jin Yoon 2001) (PDB ID 1HV6). 12 α-helices form a barrel containing a tunnel cleft which acts as the active site of the enzyme and the substrate (4-deoxy-L-erythro-hex-4-eneypyranosyluronate-mann-uronate-mannuronic acid) is show to bind in the middle of this cleft (Figure 5). Interaction with the substrate involves active residues Gly313, Arg88/306/312/342, Trp141, Try246/249, and His245. A complement of 8 water molecules acts as hydrogen bond donors for the bound substrate. A conformational change involving a loop of about 21 amino acids acts a cover to the active site upon substrate binding, with Tyr68 of the loop forming a hydrogen bond “lock” with Tyr245 in the active site (Mikami, Ban et al. 2012).

Polysaccharide Lyase 6 (PL6)

There are several alginate lyases found that correspond to the PL6 family. Structurally, enzymes are characterized by a parallel β-helix, which was found via the solved structure of the condroitinase B, another member of the PL6 family. There is a poly (β-D-mannuronate) lyase from the strain Sphingomonase sp. Strain A1 that acts exolytically on alginate and produces monosaccharide products (Hashimoto, Miyake et al. 2000). A rare polyMG-specific alginate lyase from Stenotrophomas maltophilia KJ-2 is found to have high specificity for polyMG blocks over polyM or polyG blocks and produces dimer, trimer, and tetramer oligosaccharides by endolytic cleavage (Lee, Choi et al. 2012).
Structural information can be deduced from a solved and characterized chondroitin B lyase from *Flavobacterium heparinum* (Huang, Matte et al. 1999) (PDB ID 1DBO) (Figure 6). The original structure was shown to adopt a right-handed, parallel β-helix which consists of three sets of antiparallel β-sheets, two of which run antiparallel to each other. Four α helices are also present, with three of them acting as active site stabilizers and one acting as an N-terminal cap. Chondroitinase B’s interaction with the substrate (a bound dermatan sulfate disaccharide) involves a unique and extensive internal/external asparagine ladder system found near the N-terminus. Also as large interior aromatic stack maximizing favorable interactions with the substrate. The main active site, though still being characterized, is shown to possibly involve residues Lys250, His272, Glu333, and Arg364, with a Ca$^{2+}$ ion being present as well (Michel, Pojasek et al. 2004) (PDB ID 1OFL) (Figure 7). According to the authors, the Ca$^{2+}$ ion played an important role in modulating activity, as enzymatic efficiency of the enzyme increased with increasing Ca$^{2+}$ concentration and no activity when no Ca$^{2+}$ was present (Michel, Pojasek et al. 2004). Residues Asn213, Glu243, and Glu245 are shown to be important for chelating the Ca$^{2+}$ ion.

**Polysaccharide Lyase Family 7 (PL7)**

PL7 alginate lyases represents one of the larger characterize families. These enzymes show diversity in terms of substrate preference and have both shown to act exolytically and endolytically on preferred substrate. Some polyM alginate lyases that can be found in this family are Alg7D from *Saccharophagus degradans* 2-40 with endolytic activity (Kim, Ko et al.
Lyases that prefer polyG from the family are alyPG from *Corynebacterium* sp. ALY-1 (Iwasaki and Matsubara 2000), AlyA from *Klebsiella pneumonia* subsp. arogenes (Baron, Wong et al. 1994), and AlyVGI from *Iibrio halioticoli IAM14596T* (Sugimura, Sawabe et al. 2000).

Lyases that prefer polyMG substrates are characterized in this family as well, with examples being AlyA1/A2/A3 from *Azotobacter vinelandii DJ ATCC BAA-1303* (Setubal, dos Santos et al. 2009) and PA1167 from *Pseudomonas aeruginosa PAO1* (Stover, Mayhew et al. 2000).

Based on several structurally characterized PL7 alginate lyases, the main fold of the structure appears to be that of a β-jelly roll. The crystal structure of ALY-1 (polyMG lyase) from *Corynebacterium* sp was solved in 2005 and the active site was deduced to through homology modelling by ClustalW, which found that absolutely conserved amino acid residues Arg72, Gln117, His119, Tyr189, Phe190, Lys191, Gly193, Tyr195, and Gln197 played a critical role in catalytic activity for folding of a tertiary structure (Figure 8) (PDB ID 1UAI) (Osawa, Matsubara et al. 2005). Also the very close proximity of these putative catalytic residues supported the belief of the active site locations. This is also indicated topologically as a cleft on the surface representation.

A recent structure in complex with Alginate Lyase A1-II with substrate yielded promising information on the active site (Figure 9) (PDB ID 2ZAA). The positioning of the GGMG substrate substantiated the position of the active site cleft that was suggested by previously solved
structures (PDB ID 1UAI/1VAV). However, the authors have not yet deduced the position of the active site residues and those involved with catalytic activity as of yet.

**Polysaccharide Lyase Family (PL14)**

PL14 is represented by the variety of sources that the alginate lyases are found, including a large amount from viruses. Majority of the enzymes shown prefer polyM as a substrate, though the structure that will be shown later in this section is a polyG lyase. Examples include those from eukaryote sources, such as AkAly30 from *Aplysia kurodai*, which is an endolytic, polyM alginate lyase (Rahman, Inoue et al. 2011). HdAly from *Haliotis discus hannai* is an endolytically cleaving alginate lyase that is only specific for polyM (Shimizu, Ojima et al. 2003). Unpublished studies revealed two alginate lyases: HdAlex from *Haliotis discus hannai* which is shown to be exolytically active (Suzuki 2005) and LbAly28, which is shown to be a poly(M) alginate lyase (Rahman, Inoue et al. 2011).

The only characterize structure of a PL-14 comes from the Chlorella virus ATCV-1, which is a poly(G) and interestingly has two modes of activity (either endolytically or exolytically) depending on the pH (either 7 or 10) (Ogura, Yamasaki et al. 2009). The crystal structure revealed the enzyme to adopt a β-jelly roll fold and contains a bound glucoronate residue (GlcA) attached to one of the catalytic, active site, similar to that of the PL7 lyases’ structure (Figure 10) (PDB ID 3A0N). The authors of the structure also showed, through mutation
experiments, the importance of a specific arrangement of residues in the active site that is critical for its pH dependent activity, a major difference from that of the other PL families.

The residues involved are Lys197, Ser209, His213, Ser219, Arg221, Tyr233, and /or Tyr235 (Ogura, Yamasaki et al. 2009). The substrate is bound to another site on the other side of the aforementioned active site, which validates the dual activity nature of the protein. Two separate structures show the bound substrate interacting with two sets of residues. One set, the β-D-Glucopyranuronic acid is interacting with residues Arg116, Gly105/223, Tyr106/128/130 and Lys92 (PDB ID 3IM0) (figure 11). The other site has the same substrate interacting with residues Gly150 and Gln120/153 (PDB ID 3A0N) (figure 11).

**Polysaccharide Lyase Family (PL15)**

The PL15 alginate lyases is a smaller family that is largely dominated by exotype alginate lyases. Some examples are the exotype alginate lyase Atu3025 and AGR_L_3558p from *Agrobacterium tumefaciens* str. C58 (Wood, Setubal et al. 2001). Also we have the A1-IV' and OalA from the *Sphingomonas sp* a1 (Hashimoto, Miyake et al. 2005). Structural information on the family can be found with exotype alginate lyase Atu3025 in complex with an alginate trisaccharide (polyG) (Ochiai, Yamasaki et al. 2010). The structure is shown to be comprised of three unique domains: N-terminal β-sheet domain (antiparallel), central α-domain (15 α-helices form a α₆/ α₆ barrel), and a C-terminal β-sheet domain (antiparallel) (Ochiai, Yamasaki et al. 2010) (PDB ID 3AD0).
According to the authors, there is a slight conformation change upon substrate binding, where the C-terminal β-sheet domain acts as a hinged cover (PDB ID 3AFL). When overlaying the PDB structures 3A0O and 3AFL (substrate bound structure), there is a conformational change in the active site with an RMSD of 1.502 (figure 12). The conformational change is best seen when zoomed up on the active as displacement of the loops and α6/α6 barrel is more obvious. Active site residues were concluded to responsible for production of monosaccharides involved H531, R314, Y365, W467, R199, Y555, and W198, with H311 and Tyr365 shown to function as the catalytic acid and base (PDB ID 3AFL). The authors concluded that these residues were important, as subsequence mutation experiments showed a marked decrease in active or a negative alteration to the kinetics of the enzyme.

**Polysaccharide Lyase Family 18 (PL18)**

PL18 alginate lyases feature a preference for different kind of alginate substrates. This is one of the more recently families to be discovered as there have been only a handful of discovered members. Some examples are AlyA from *Pseudoalteromonas atlantica* AR06, an endolytically active enzyme that targets polyM, polyG, and polyMG alginates (Matsushima, Danno et al. 2010). AlyPEEC from *Pseudoalteromonas* sp. IAM14594 is a polyMG lyase (Sawabe, Sugimura et al. 1998) and Aly-SJ02 from *Pseudoalteromonas* sp. SM0524, a bifunctional enzyme with a recently submitted native apo structure (Dong, Wei et al. 2014).
Two structures of PL18 alginate lyase have been submitted on the PDB. One is an alginate lyase from *Alteromonas* sp 272 (PDB ID 1J1T) and the other is the previously mentioned aly-SJ02 (4Q8L). Unfortunately, without substrate bound in complex with these enzymes, it hasn’t been deduced as to how similar these enzymes are to the other polysaccharide lyase families in a bioinformatical sense. However, in a structural sense, this enzyme also retains a β-jelly roll fold, similarly to PL7 and PL14 alginate lyases (figure 13). The RMSD values of the two structures aligned together is approximately 0.337, showing very similar structural features. A blast comparison between the two structures reveals a 99% sequence identity over a 97% query cover.

**Polysaccharide Lyase Family 17 (PL17)**

Characterization of this family of enzymes has been relatively sparse as shown on the CAZy database. There are only few characterized enzymes, those of which have been shown to be exolytically active enzymes. One is AlyII, an alginate lyase that preferably attacks polyM substrate and works in tandem from Aly (another alginate lyase, with both from *Pseudomonas* sp OS-ALG-9 (Kraiwattanapong, Motomura et al. 1999). A novel oligoalginate lyase from *Sphiogomonas* sp. MJ3 with a very high activity from polyM alginate, though it still have activity towards polyG and polyMG blocks (Park, Kam et al. 2012).

The work that we have focused on was an alginate lyase from a marine bacterium *Saccharophagus degredans* 2-40, which bioinformatically was classified as a putative PL17 enzyme and is shown to degrade complex polysaccharides such as alginate (Weiner, Taylor et
al. 2008). Also in the genome was the presence of nine putative alginate lyases from the families of 7, 14, and 18 which shows the possible ability for these enzymes to work in a cooperative manner. Biochemical characterization of the putative PL17 enzyme, or know more so as Alg17c from referenced literature, shows it is an exotype enzymes that is able to depolymerize alginate di-, tri-, and tetra-saccharides into smaller monosaccharides, which is important for the production of ethanol fermentation (Kim, Chung et al. 2012).

However, an interesting feature of the enzyme is compared to other alginate lyases, was limited sequence similarity to PL15, the closest enzyme in terms of identity, and the lack of conserved, critical active site residues leads to the. This is due to the fact there is a lack of structural information on PL17 enzymes available and a known mechanism for how activity is carried out. The structure of the PL17 enzyme that was discovered and will be described in detail provides an understanding for how alginate is digested in this enzyme. We present several crystal structures of the Alg17c from *S. degredans* including apo structures without the presence of substrate, two active site mutants, and an inactive mutant with a bound substrate. This allowed for identification of active site residues critical for substrate binding and the exolytic activity, which is representative of PL17 enzymes. This is the first structure of the PL17 enzymes and yields critical information of how the PL17 enzymes work.
Figure 1: The general outline biodiesel production. Algae cells are mechanically broken down followed by lipid extraction through pretreatment of enzymes and hexanes. This is followed by trans-esterification and separation of the target FAME products. Final steps to biodiesel production involve the complete removal of alcohols and acids used in the isolation process.
Figure 2: Structures of alginate polysaccharide blocks linked by glycosidic 1→4 O-linkage. A) Homopolymeric block of G. B) Homopolymeric block of M. C) Heteropolymeric block of GM.
Figure 3: General mechanism of alginate lyases whether by A) Syn elimination or B) anti elimination.
Figure 4: A) Endolytic Cleavage of a polyM substrate. B) Exolytic cleavage of a polyM oligomer and non-enzymatic conversion to 4-deoxy-l-erythro-5-hexose uronic acid.
Figure 5: Structure of AI-III PDB ID 1HV6. A) Structure with trisaccharide (4-deoxy-l-erythro-hex-4-ene.pyranosyluronate-mannuronate-mannuronic acid or MMM) substrate bound (yellow). B) Surface representation showing active site pocket. C) Active site residues show to interact with the trisaccharide (4-deoxy-l-erythro-hex-4-ene.pyranosyluronate-mannuronate-mannuronic acid).
Figure 6: A) Chondroitinase B (PDB ID 1DBO) with a bound dermatan sulfate disaccharide (yellow). B) Surface representation showing the binding pocket for the substrate. C) Putative active site residues involved with activity/substrate binding.
Figure 7: A) Chondroitinase B (PDB ID 1OFL) with several bound dermatan sulfate disaccharide (yellow) B) Surface representation showing bound Ca$^{2+}$ in active site as well as several bound dermatan sulfate disaccharide (yellow). C) Active site residues (green) and metal chelating residues (purple).
Figure 8: Aly-1 presented in A) cartoon representation and B) Surface representation, with active site cleft. C) active site residues are label and colored in silver.
Figure 9) Structure of A1-II (yellow) with bound tetrasaccharide GGMG (Orange), represented as A) cartoon representation and B) surface representation.
Figure 10: A) cartoon representation of Val-1 with bound β-D-Glucopyranuronic acid (salmon) and B) surface representation showing the binding pocket and the substrate bound PDB ID 3A0N).
Figure 11: Active sites of two solved structures of val-1. A) Active site of 3IM0 with bound β-D-Glucopyranuronic acid (salmon) and active site residues in blue. B) Active site residues of 3A0N with β-D-Glucopyranuronic acid (salmon) and residues in blue.
Figure 12: A) cartoon overlay with 3A0O (orange) and 3AFL (blue) with substrate (yellow). B) A close up of the active site of the same overlay with an RMSD of 1.502, where the conformational change is more noticeable. C) Surface representation of 3AFL (Blue) with substrate GGG (yellow) and D) zoom up of active site with labelled residues; however it is important to note since the structure was a H531A mutant, the A531 residue is not shown.
Figure 13: aly-SJ02 (PDB ID 4Q8L) in teal, represented as A) cartoon structure and B) surface representation, which shows the presence of the putative active site tunnel. Structural alignment of aly-SJ02 (teal) and alginate lyase from *Alteromonas* sp 272 (PDB ID 1J1T) (yellow).
1.2 MATERIALS AND METHODS

**Strains Used**

*Saccharophagus degradans* 2-40 genomic DNA was purchased from the American Tissue Culture Collection or ATCC. ATCC number is catalogue number is ATCC 43961. *S. degradans* is a gram-negative, marine-dwelling, rod-shaped bacteria that was originally found from salt marsh cord grass *Spartina alterniflora* which is found in the Chesapeake Bay (Taylor, Henrissat et al. 2006). The best growing conditions for this strain was found to be 5° to 40°C with an optimum temperature of 37°C with an optimum pH of 7.5 (Andrykovitch and Marx 1988). *S. degradans* is able to degrade at least 10 different complex polysaccharides including agar, alginate, cellulose, chitin, β-glucan, laminarin, pectin, pullulan, starch, and xylan, which are its sole carbon source (Ekborg, Taylor et al. 2006). This known factor makes *S. degradans* a versatile complex polysaccharide degrader which has potential in degrading wastes from agricultural wastes, as well as biofuel potential.

Engineered *Escherichia coli* (*E. Coli*) competent cells were used for cloning and protein expression purposes. Cells were purchased from the Cell Media Facility located on campus at the University of Illinois at Urbana-Champaign. One of the *E. coli* strains we used, DH5α, was engineered with multiple mutations that enable high-efficiency in transformation procedures. The mutations that are made to the DH5α strains are Δ lacZ M15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17(rK-mK+) supE44 thi-1 gyrA96 relA1 (Taylor 1993). The endA1 mutation inactivates an intracellular endonuclease that degrades plasmid DNA, thus allowing for lower rates endonuclease degradation and higher yields of plasmid transfer (Taylor 1993). The hsdR17
mutation eliminates the endonclease of the EcoKI restriction-modification system, so the DNA without methylation from EcoKI will not be degraded (Taylor 1993). The Δ lacZ M15 is the alpha acceptor allele required for blue-white screening with recombinant cells containing lacZ based vectors and the recA1 mutation reduces homologous recombination, in turn reducing deletion formation and plasmid multimerization (Taylor 1993).

*E. coli* strain BL21 (DE3) with an inserted Rosetta™ plasmid (pRARE) (all of which was engineered by Novagen) was the strain we used for protein expression. The pRARE plasmid allows for a high production rate of tRNA to circumnavigate the issue of expression yields from codon bias (Novy 2001). This allows for better cooperation with phage based, T7 expression platform, such as pET vectors, which are isopropyl-1-thio-β-D-galactopyranoside (IPTG) induced from a T7 polymerase (also on the pET vector), which allows for higher expression yields of desired protein uninterrupted.

**Cloning and Transformation**

Alg17c (GenBank Accession ADB82539.1) was amplified from the *S. degredans* genomic DNA. (Figure 14). It is important to note that the underlined and bolded sequence are the 24 residues that encode for the signal peptide which was removed during cloning, which allows for facilitation of heterologous expression (Kim, Chung et al. 2012). The primers used to clone out the gene were designed for the restriction sites Xhol and BamHI. Primers were designed in a 5’ to 3’ fashion (Figure 14).
The gene was cloned out and isolated by polymerase chain reaction or PCR protocol. The Tm used was 77°C with an elongation time of 45 seconds. Reagents used were from New England Biolabs using the Q5® High-Fidelity DNA Polymerase (M0491). The final mix solution included 10 µl of 5X Reaction Buffer, 200 µM of dNTPs, 0.4 µM of forward primer, 0.4 µM of reverse primer, approximately 200ng of template DNA, 0.02 U/µl of High-Fidelity DNA Polymerase, and filled to 50 µl with nuclease free water. The protocol for the thermocycler was an initial denaturation at 98°C followed by 20 cycles of the following: 10 second denaturation at 98°C, 30 second annealing at 77°C, and followed by extension at 72°C at 30 seconds. Then there is a final extension time at 72°C.

The gene is then checked using agarose gel electrophoresis. Casting of the gel is done in 50 mL of TAE buffer solution (tris base, acetic acid, and EDTA mix) which is mixed with 0.5 g Agarose (Fischer 21163). After heating it is poured into a castor mold with 0.4 µl of Ethidium Bromide (Fisher 15585011). After running the gel to separate out product, the desired fragment was then isolated using QIAquick Gel Extraction Kit from Qiagen (Cat. No. 28706). The DNA was excised and incubated in 250 µl of Buffer QG at a temperature of 50°C, where the agarose was dissolved. Then using spin prep columns provided, the sample was cleaned further in the provided PE buffer, which was then eluted of the prep column. The fragment was then digested over night with restriction enzymes BamHI (NEB RO126S) and XhoI (NEB R0146S) at 37°C.

After the overnight digest was cleaned using the QIAquick PCR Purification Kit (Cat. No. 28104), the purified gene was then ligated into a pre-double digested pET-28b vector with the same
restriction sites of BamHI and XhoI. The mixture for the ligation reaction contained 6 µl of purified Alg17c gene insert, 2 µl of double digested pET-28b, 1.5 µl of T4 DNA Ligase buffer from NEB (Cat. No. M0202S), and 0.5 µl of T4 DNA Ligase from NEB (Cat. No. M0202S). This was incubated at 16°C for 4 hours, where then transformed into DH5α E. coli.

Site-specific mutants were designed using a QuickChange Site-Directed Mutagenesis Kit from Agilent Technologies Inc (Cat. No. 200521) (Figure 15). All reagents for the reaction were provided with the kit. The final mix contains 5 µl of 10x reaction buffer, 10 ng of Alg17c in pET-28b, 125 ng of mutation primer forward, 125 ng of mutation primer reverse, 1 µl of dNTP mix, 3 µl of QuikSolution, and the filled the remaining volume to 50 µl using ddH2O. Then 1 µl of PfuUltra HF DNA polymerase is added (2.5 U/µl).

The initial denaturing cycles is at 95°C for 1 minute followed by a series of 18 cycles repeating the following: denaturing for 50 seconds at 95°C, annealing for 50 seconds at 60°C, then extension for 5 minutes at 68°C. Then a final elongation step at 68°C for 7 minutes.

Afterwards, the product undergoes a DpnI digestion, where 1 µl of DpnI restriction enzyme is added to the reaction and incubated for 1 hour at 37°C. DpnI works to cleave off methylated DNA that is formed previously.

After the ligation and mutation reactions, the genes were then transferred into DH5α E. coli using a heat shock transformation method. First step is incubation on ice for 30 minutes to allow for bacterial membrane to stabilize increasing interaction between the calcium cation and negatively charged membrane components. Then the cells plus DNA are heated for 2 minutes
in the 43˚C water bath to change the fluidity of the membrane to promote DNA uptake. Next the five minute incubation on ice reduces thermal motion of DNA to promote exogenous DNA to enter the competent cells. The final steps involve adding 250 µl of Luria Broth (LB) liquid media, shaking for 1 hr at 37°C, plating on pre-purchased LB plates infused with 50 µg/ mL Kanamycin, and finally left overnight in the 37°C incubator.

After colonies are formed on the plates, several colonies were picked and grown overnight at 37°C. Plasmid was then isolated using the QIAprep Spin Miniprep Kit from Qiagen (Cat. No. 27104). The clones were then sent for sequencing to check for positives using the ACGT, INC (Wheeling, IL) sequencing services. Positive hits were transformed into E. coli BL21 Rosetta DE3 using the same transformation method as DH5α.

**Culturing and Harvesting**

Transformed E. coli BL21 Rosetta DE3 containing the Alg17c, both WT and Mutants, were grown in 2 L cultures in LB media (BD Biosciences, Cat. No. 244610). The cultures were grown for 4 hours at 37°C at 200 rpm to OD 0.8. The cultures were then induced with 0.1 mM IPTG and grown overnight at 18°C. The next day the cells were harvested by centrifugation using the Sorvall RC-3B Plus with 1L bucket rotor from Thermo Scientific at 3,500 rpm for 25 minutes. The resulting pelleted was suspended in a buffer solution containing 500 mM NaCl, 20 mM Tris pH 8.0, and 10% glycerol. Depending on immediate or later use, the pellet was either immediately lysed or frozen in liquid nitrogen and stored at -80°C.
Selenomethionine (SeMet) cultures were grown in 2L of M9 minimal media made with 22.4 g Na₂HPO₄·7H₂O, 6 g KH₂PO₄, 2 g NH₄Cl, and 1 g NaCl. This was autoclaved at 121°C for 30 minutes and allowed to cool. The media was then further supplemented with 25% glucose, 0.1 mM MgSO₄, 4.2 mg/mL FeSO₄, and 0.05% thiamine, along with 20 mL of previously grown start up culture. Cultures were shaken vigorously at 37°C and grown to OD ~0.5, where then a 20 mL amino acid mix of 0.1 g/10 mL of Lysine, Phenylalanine, Threonine, and 0.05 g/10 mL of Leucine, Isoleucine, Valine are added and 15 minutes later, selenomethionine is added at 0.06 g/10 mL. After another 15 minutes, 0.1 mM IPTG is added and the cultures are shaken overnight at the same speed at 18°C.

**Protein Purification**

The resuspended pellet was lysed using an EmulsiFlex-C5 Homogenizer from AVESTIN (Ottawa, Canada). Pressure was approximately 50,000 kPa. This continued for 15 minutes or until the cells were fully lysed. The resulting lysate was then spun down at 16,000 rpm for 1 hours at 4°C using a Beckman J2-21M/E centrifuge from Beckman Coulter Life Sciences (Indianapolis, IN). The supernatant containing the soluble fraction was injected into a 5 mL Hi-Trap Ni-NTA, containing immobilized metal ion affinity resin (G.E. Healthcare), which was pre-equilibrated with binding buffer of 50 mM Tris-HCl at pH 7.5, 300 mM KCl, and 20 mM imidazole. After the sample was loaded, the protein was eluted by linear gradient from 20 mM imidazole to 200 mM imidazole over the course of 30 minutes using an AKTA UPC-900 FPLC (G.E. Healthcare Life Sciences, Pittsburgh, PA). Fractions were selected based on high UV signal intensity generated
by aromatic amino acids tryptophan, tyrosine, and phenylalanine at approximately 280 nm. Selected fractions were then running on a polyacrylamide, SDS-PAGE gel for 1 hour to observe sample purity. Fractions containing the purest sample were then incubated with thrombin overnight in a buffer containing 100 mM KCl, 20 mM Tris at pH 8.9 to remove imidazole and the N-Terminal polyhistidine tag.

After overnight dialysis, the sample was injected onto a Superdex HiLoad™ 75 16/60 size exclusion column (GE Healthcare, Pittsburgh, PA). The column was pre incubated in a buffer containing 20 mM HEPES pH 7.5 and 100 mM KCl. Injection of the sample on the column was done at using an AKTA UPC-900 FPLC (G.E. Healthcare Life Sciences, Pittsburgh, PA) at a flow rate of 1 mL/min. Slow elution was carried out over the course of 2 hours using the same buffer. Samples were then concentrated to 15 mg/mL using a 10,000 MWCO Amicon centrifugal filter (Merck Millipore, Carrigtwohill, CO).

Crystallization and Data Collection

Screening for crystals of the wild type, SeMet, and mutant variants of Alg17c was done using the Crystal Gryphon (Art Robbins Instruments, Sunnyvale, CA). Individual drops containing protein and screening solutions were set up at 0.3 ul on 96 well Intelli-PLATES (Art Robbins Instruments, Sunnyvale, CA) by sitting drop method. The initial hits were further optimized and the final condition for both wild type and SeMet crystals was 0.1 m Tris pH 8.0, 5% 2-methyl-2, 4-pentanediol (MPD), 10% Polyethylene glycol (PEG) 6000. These crystals were grown over the
course of 3 days at 16°C. Substrate soaks were carried out with digested alginate oligosaccharide was provided to us by the Huimin Zhao Lab at the University of Illinois at Urbana-Champaign. This substrate was then incubated with preformed apo-crystals at 2 mM for different time steps ranging from 15 minutes to 4 hours. Crystals were selected from all time points for data selection.

Next was the selection of cryopreservation agents for the crystals, which functions to reduce radiation damage to the crystal during irradiation. Cryo-protectant selection was based on the resolution produced from each, with the best resolution being selected at 1.85 Å. Cryo’s included the optimized condition, plus 20% PEG 2000/4000/8000/10000, as well as 25% rehalose/threitol/glycerol/PEG 400, The selected cry condition for flash freezing contained 20% PEG 6000, 12% ethylene glycol, 0.1 M Tris pH 8.0, and 5% MPD. After soaking in cryo, crystals were immediately flash frozen in liquid nitrogen.

**Structural Determination**

Crystals of the wild-type Alg17c diffracted to a resolution of 1.85 Å. Data was collected at the insertion device synchrotron beam line at Sector 21 ID-G LS-CAT (Advance Photon Source, Argonne, IL) using a Mar 300 CCD detector. Exposure was at 1 second per frame for 250 frames at a fixed wavelength of 0.97872 Å. Diffraction data was indexed and scaled using HKL2000 (Otwinowski, Borek et al. 2003) or XDS (G. Bricogne 2003). The data set was found to have an R\text{merge} = 9.2% and I/\sigma(I) =2, with a space group of P2\text{1}. SeMet – Alg17c crystals, were collected
on the same beamline and using the same parameters. The resolution of the data set was 2.2 Å and processed using single-wavelength anomalous diffraction or SAD to accommodate the phase problem found with the native data sets. The presence of heavy atom sites were determined used HySS and subsequently refined in SHARP to generate the required phases (G. Bricogne 2003). The model was then build using the resultant phase data via ARP/wARP (Langer, Cohen et al. 2008) and fitted finally in COOT, where residue bonds and angles were fitted properly (Emsley and Cowtan 2004). Final refinements were carried out using REFMAC (Murshudov, Vagin et al. 1997).

The data sets for the structures of the Alg17c mutants H202L and Y258A were collected at 1.7Å and 2.45Å respectively. The structure of Y258A was in complex with an alginate trisaccharide and diffracted to a resolution of 1.9Å. These variants’ structures were determined using Phaser-MR in the Phenix GUI using the wild-type Alg17c as the initial build model (Echols, Grosse-Kunstleve et al. 2012). The refined structures of all variants with statistics were then deposited in the protein data bank (Rose, Bi et al. 2013).

**Kinetic Analysis of Alg17c**

Measurements of the kinetics for Alg17c were carried out on the wild type as well as all generated mutants using low-viscosity alginate as the substrate purchased from MP Biomedicals (CAS 9005-38-3, Santa Barbara, CA). For each assay, 0.613 µM of enzymes was added to 1 mL of mixture which contained 50 mM Tris pH7.5 and varying concentrations of
alginate, from 0.1 mg/mL to 10 mg/mL). The reaction mixture was incubated at 30°C for one hour then boiled for 5 minutes to stop the reaction. The thiobarbituric acid (TBA) method was used to quantify the amount of unsaturated uronic acids produced in the reaction (Manabu Kitamikado 1992). 40 μl of the reaction mix was added to 0.25 mL of 5 mM H$_5$IO$_6$ (in 62.5 mM H$_2$SO$_4$) and incubated at 23°C for 20 minutes.

This is then followed by the addition of 0.50 mL of 2% sodium arsenite to the mixture, then vigorous shaking for 2 minutes. 2 mL of 0.3% thiobarbituric acid at a pH of 2 was added to the reaction mixture, followed by heating at 100°C for 10 minutes. After being allowed to cool for an extended amount of time, formation of β-formylpyruvic acid with thiobarbituric acid was measured at the absorbance wavelength at 548 nm. A basic mechanism for the formation of the resulting chromophore from the reaction with added TBA is shown (figure 16). The units of the resulting measurements were measured as one unit of enzyme activity required to free 1 μM of β-formylpyruvic acid per minute.

**Thin Layer Chromatography (TLC) and Isothermal Titration Calorimetry (ITC)**

Alginate oligosaccharides with degrees of polymerization of 2, 3, and 4 at the concentration of 0.5 mg/mL were incubated with 0.6 μM of Alg17c in a buffer solution containing 50 mM Tris pH 7.5. This mix was then incubated at 30°C for one hour. The reactions were then analyzed by TLC using a solvent containing a mixture of n-butanol: formic acid: water (4:6:1, v/v/v) respectively. The products were run on small silica based TLC plates, where migration occurs via capillary
action, and visualization was developed by spraying a solution of 10% sulfuric acid in ethanol. The standards used were D-glucose monosaccharide at 2mM and D-trehalose disaccharide at 2mM.

ITC measurements were carried out on the MicroCal VP-ITC system (GE Healthcare, Piscataway, NJ) located in the University of Illinois at Urbana-Champaign’s Microanalysis Laboratory. Measurements were used to calculate dissociation constant ($K_d$) for the binding of oligosaccharides to the Alg17c. Purified enzyme was concentrated to 15 mg/mL and diluted in a buffer solution containing 20 mM HEPES pH 7.5 and 100 mM KCl. Oligosaccharides with a degree of polymerization between 2 to 4 (DP 2-4) were diluted in identical buffer to approximately 1 mM. Both enzyme and substrate solutions were degassed and brought to room temperature. The reference cell contained ddH$_2$O and the sample chamber was washed with a ddH$_2$O and methanol. The sample was injected into a 1.44 mL reaction cell, followed by 26 successive injections of 10 µL of ligand at 300 second intervals. Data was fitted using a nonlinear regression curve on a single site model. This was done on the MicroCal Origin software.
Figure 14: The amino acid sequence for Alg17c. Signal sequence is bolded and underlined.

(GenBank Accession ADB82539.1)
Alg17c Forward Primer for restriction site XhoI:

5’ GCGCTCGAGCAAGTTTCTGGCAATGGTCATC 3’

Alg17c reverse Primer for restriction site BamHI:

5’ GCGGGATCTTTATACGTAACCACCGC 3’

Figure 15: Designed primers for cloning out Alg17c. Forward primer is designed for restriction site XhoI. Reverse primer is designed for restriction site BamHI.
H413A Mutation: Alg17C H415A Fw: 5'-GGTATGGGGCTGTGCTTTCGATAAG-3'
Alg17C H415A Rv: 5'-AAAGCTTAGTATGCAGACACAGG-3'

H415A Mutation: Alg17C H415A Fw: 5'-TATGGGGCTGTGCTTTCGATAAG-3'
Alg17C H415A Rv: 5'-TAAGCTTTATCGAAAGCACCATGCCC-3'

R438A Mutation: Alg17C R438A Fw: 5'-TACGGTGCGGCGGCATACCTAAAC-3'
Alg17C R438A Rv: 5'-AACGTTTAGTATGCCCGCCGACC-3'

R260A Mutation: Alg17c_R260A_Fw: 5'-ATACTACCAAGCATATGCACTTATGCC-3'
Alg17c_R260A_Rv: 5'-TAAGTGACATATCTTGATATGCCC-3'

Q149A Mutation: Alg17c_Q149A_Fw: 5'-GGCAAGCCTAGCAGAAGCTTGGTGG-3'
Alg17c_Q149A_Rv: 5'-CAACACATGTTTCTGCTAGGCTTTG-3'

N201A Mutation: Alg17c_N201A_Fw: 5'-ACAAGATTCAGCTACCAGTTGTACGT-3'
Alg17c_N201A_Rv: 5'-ACGTACCCTGAGCGTGAATCTTG-3'

H202L Mutation: Alg17C H200A Fw: 5'-TTCAACAAGATTGCAAATCACGGTACGTG-3'
Alg17C H200A Rv: 5'-CGTACCCTGAGCTTGGTGAATG-3'

Y450A Mutation: Alg17C Y450A Fw: 5'-GGCGGTCATGCTATTAGCTGAAAACAATAC-3'
Alg17C Y450A Rv: 5'-TTTAGCTAATGCATGACCGGCTTTTAG-3'

Y258A Mutation: Alg17C Y257A Fw: 5'-ACACAGAAGGCCCAGCATACCAACGC-3'
Alg17C Y257A Rv: 5'-GCATAGCGTTGCTATGTGCTTTCTTC-3'

Figure 16: Primer design for using a QuickChange Site-Directed Mutagenesis Kit from Agilent Technologies Inc (Cat. No. 200521)
Figure 17: General reaction leading to formation of the pink colored fluorophore. Unstaturated product reacts with periodic acid to for the β-Formyl Pyruvic Acid. The following reaction with 2 molecules of thiobarbituric acid yields the fluorophore.
1.3 RESULTS AND DISCUSSION

Kinetic Analysis

It is important to note that the enzymes used for these assays had the first 24 residues that encoded for the signal peptide removed to facilitate expression. Kinetic values for the wild-type Alg17c and subsequent mutants were acquired by using the TBA assay and the values were recorded (Table 1). The wild-type recorded a $k_{cat}$ of $56.9 \text{ s}^{-1}$, a $k_{cat}/K_m$ of $2.6 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$, $K_m$ of $21.3 \mu\text{M}$, and a $V_{max}$ of $34.9 \mu\text{M}$, and a $V_{max}$ of $34. \mu\text{M} \text{ S}^{-1}$. Comparison to other alginate and polysaccharide lyases show similarity in values, including that of AlgL from *Pseudomonas aeruginosa* (Farrell and Tipton 2012) and Atu3025 from *Agrobacterium tumefaciens* (Ochiai, Yamasaki et al. 2010).

Mutants to active site residues N201A, Y258A, and Y450F resulted in the abolishment of any detectable activity. R260A recorded $k_{cat}$ of $0.0149 \pm 0.002 \text{ s}^{-1}$, a $k_{cat}/K_m$ of $0.016 \pm 0.003 \text{ M}^{-1}\text{s}^{-1}$, $K_m$ of $12.1 \pm 3.1 \mu\text{M}$, and a $V_{max}$ of $0.0092 \mu\text{M}$. R438A recorded $k_{cat}$ of $3.63 \pm 0.29 \text{ s}^{-1}$, a $k_{cat}/K_m$ of $0.013 \pm 0.001 \text{ M}^{-1}\text{s}^{-1}$, $K_m$ of $31.1 \pm 5.92 \mu\text{M}$, and a $V_{max}$ of $10.9 \pm 0.88\mu\text{M}$. Q149A recorded $k_{cat}$ of $0.018 \pm 0.001 \text{ s}^{-1}$, a $k_{cat}/K_m$ of $0.0058 \text{ M}^{-1}\text{s}^{-1}$, $K_m$ of $31.1 \pm 5.9 \mu\text{M}$, and a $V_{max}$ of $0.011 \pm 0.0008 \mu\text{M}$. H202L recorded $k_{cat}$ of $10.9 \pm 0.3 \text{ s}^{-1}$, a $k_{cat}/K_m$ of $1.2 \pm 0.2 \text{ M}^{-1}\text{s}^{-1}$, $K_m$ of $91.6 \pm 18.6 \mu\text{M}$, and a $V_{max}$ of $6.7 \pm 0.2 \mu\text{M}$. H413A recorded $k_{cat}$ of $5.02 \pm 0.36\text{ s}^{-1}$, a $k_{cat}/K_m$ of $0.8 \pm 0.2 \text{ M}^{-1}\text{s}^{-1}$, $K_m$ of $85.1 \pm 9.51 \mu\text{M}$, and a $V_{max}$ of $15.1 \pm 1.08 \mu\text{M}$. H415A recorded $k_{cat}$ of $0.057 \pm 0.004$, a $k_{cat}/K_m$ of $0.026 \pm 0.002 \text{ M}^{-1}\text{s}^{-1}$, $K_m$ of $21.6 \pm 2.7 \mu\text{M}$, and a $V_{max}$ of $0.035 \pm 0.002 \mu\text{M}$. 
TLC Results and the Determination of Polymer Specificity

The preferred polymer length that can be digested by Alg17c was determined using TLC analysis of reaction products generated from reaction involving a mixture of alginate oligosaccharides with degrees of polymerization (DP) of 2, 3, and 4. The result TLC shows that Alg17c completely processes the tri- and tetra-saccharide substrates (Figure 18). The di-saccharide was shown to be only partially digested. The results show the minimum substrate DP preference of Alg17c to be that of a tri-saccharide. Kinetic measurements of the wild-type using tri-saccharide substrates yielded a $k_{cat}$ of 62.4 s$^{-1}$, $k_{cat}/K_m$ of $8.2 \times 10^6$ M$^{-1}$ s$^{-1}$, $K_m$ of 7.7 μM, and $V_{max}$ of 38.2 μM s$^{-1}$. When compared to other alginate lyases that are active against oligosaccharides with a similar DP, the kinetic values are consistent with those other enzymes. This represents the notion that an alginate tri-saccharide is the minimal length substrate for Alg17c.

Alg17c’s Overall Structure

Alg17c’s wild-type structure was determined to 1.85 Å. Phases used to solve the final structure were determined by SAD measurements generated from data collected on crystals of the SeMet labeled protein. The statics for wild-type, SeMet labeled Alg17c, H202L mutant, Y258A mutant, and Y258A mutant with bound substrate MMG are located in the following table (table 2). A monomer of Alg17c comprises of two domains: and N-terminal imperfect α-barrel and a C-terminal β-sheet domain (figure 19) (PDB 4OJZ). The N-terminal α-barrel is imperfect due to the presences of thirteen helices, an extra helix on the side of the barrel compared to the α6/α6-
barrel usual six helix-loop-helix motif which serves to increase rigidity of the barrel. The C-terminal β-sheet domain is comprised of three co-planar layers of β-sheets in an antiparallel arrangement, with four small helices interjecting in between the two domains (Figure 19). Because of the high sequence conservation between the other PL-17 enzymes at predicted key areas of the sequence, this structure is suggested to be representative of all enzymes in this family. Structurally, Alg17c is different from other alginate and polysaccharide lyases, which are shown to adopt either a α-barrel or a jelly roll fold (Lombard, Bernard et al. 2010).

A unique zinc binding domain was found on Alg17c. This was evident by the presence of electron density that resembled a metal ion near interface between the c-terminal and n-terminal (figure 20a). The residues His415, Asp433, and His464 are located around this density and their nitrogen and oxygen atoms position in a way to be interacting with it (Figure 20b). Based on coordination geometry, distance, an f” value of 2.55 with an enhanced data set collected at λ=1.0 which is similar to the wavelength of zinc, and comparison to other similar structures containing zinc such as heparinase II from P. heparinus, we can deduce that this electron density corresponds to zinc (Figure 21) (Shaya, Tocilj et al. 2006). To show the importance of the zinc ion on catalytic activity of Alg17c, a mutation to H415 to an alanine was made and the kinetics were characterized (table 1). The $k_{cat}/K_m$ was $2.6 \times 10^3$ M$^{-1}$s$^{-1}$, which shows compromised activity when compared to the wild-type (about 1000x fold lower). This shows the proposed importance of the metal ion as stabilizing binding to the substrate. This is a unique feature to the PL-17 enzymes as even the closest structural relative, Atu3205, does not contain a metal ion binding site in it.
Structure of Alg17c Bound to Alginate Trisaccharide

To deduce the residues responsible for binding to the oligosaccharide substrate, comparisons to the closest structurally related polysaccharide lyase Atu3205. Structural alignments show the residues believed to have an importance to catalysis and binding stabilization. Based on this alignment, H202L and Y258A mutations were generated as their positions were equivalent to the already characterized Atu30256 residues of His311 and Try365, which function as a catalytic acid and base. To determine if these residues were indeed involved with catalysis with Tyr258 and His202 as a catalytic acid and base respectively, mutations were made to these residues with kinetic analysis carried out. Kinetic characterization (as shown in table 1) show that Y258A had absolutely no activity, which the H202L mutant had a $k_{cat}$ of $10.9 \text{ s}^{-1}$ and a $K_m$ of $92 \mu\text{M}$, showing compromised activity. This shows that His202 may not necessarily participate in acid/base chemistry as originally predicted. Structurally, the two mutants did not show any major structural changes, with exception to very minor movements of loop regions in the β-sheet domain. This is also why we were probably unable to obtain a co-crystal structure with the H202L variant. However we did obtain a crystal of Y258A in complex with oligosaccharide trisaccharide, determined to a resolution of 2.45Å.

We then proceeded to determine the affinity for binding of Alg17c-Y258A using ITC analysis. What we found was that the mutant bound to the trisaccharide ligand with a 1:1 stoichiometry. The enzyme also had an affinity of $9.9 \times 10^3 \text{ M}^{-1}$ (Figure 22). Based on the curve that was generated from the ITC measurement as well as the calculated $K_D$, this reaction is found to be largely enthalpically driven.
To be able to further identify active site residues that participate in both substrate binding and the exolytic type mechanism, we generated alginate oligosaccharides with DP of 2 to 4 were generated from crude sodium alginate and soaked at varying concentrations into preformed crystals of Y258A. Crystals of Y258A in complex with alginate oligosaccharide diffracted to 1.9 Å and solved using molecular replacement (figure 23). The bound substrate that fit a density near the active site residue Tyr258 and corresponded to a tri-saccharide ΔMMG, where ΔM corresponds to the unsaturated D-mannuronate, and M and G correspond to saturated D-mannuronate and L-glucuronate. Structural statistics to the bound forms of the Alg17c variants can be found on table 2. Also another interesting feature was the presence of conformational changes between the bound form and unbound form of Alg17c (Figure 24a). The alignment of the α6/α6-barrel domain shows that c-terminal β-sheet domain undergoes a rotation upon ligand binding those results in the positioning of active site residues for interaction (Figure24b).

We can see a similar domain shift with PL-15 family Atu3205 in complex with ΔGGG (Ochiai, Yamasaki et al. 2010) and a series of computational modeling studies suggest the active site flexibility plays are a critical role in substrate engagement, as ween with PL-8 hyaluronate lyase (Jedrzejas, Mello et al. 2002). Despite having no sequence similarities, they share the common architecture, suggesting that the conformational flexibility may be a common feature of all enzymes from this particular (α/α)n toroid structural class.

With the information derived from the co-crystal structures, we were able to deduce a mechanism for how the tri-saccharide is processed by Alg17c (figure 25). The nomenclature for substrate binding in lyase active sites define subsites that interact with each unit of the
oligosaccharide, which cleavage primarily engaged at the glycosidic bond that links the sugars. These subsites are the part of the sugar that the catalytic residues interact, for this case denoted with numbers -1, +1, and +2 (in reference to the overall charge of that atom). In the usual exolytic mechanism, the negative charge on the C6 (for carbon 6) carboxylate of the oligosaccharide at subsite +1 is stabilized, which then promotes the removal of the C5 proton by a general base reaction. This is followed by the donation of a proton by the O4 atom via general acid reaction (Gacesa 1987).

The bound ΔMMG substrate occupies sites -1, +1, and +2, where the nonreducing mannuronate occupying the deep cleft in the binding pocket and the pyranose ring stacks with the Tyr261 side chain. Hydrogen bond interactions can be seen with Asn149, Tyr257, Arg260, His413, Arg438, and Glu667. At subsite +1, the saturated D-mannuronate interacts with residues Gln146 and Asn201, which is indicative of the flexibility in accommodating different bound saccharides at this position. The L-glucuronate at the +2 position interacts in a minimal manner with Lys136 located within hydrogen bonding distance.

Prediction of roles for the functional residues is based on their location in proximity of the uronic acid carboxylate, the glycosidic bond, and the acidic C5 proton. C6 carboxylate is located next to Asn201 and His202, which show that these residues likely function to stabilize the carboxylate charge of the C5 proton. Tyr450 is poised in a way to the C5 proton, where it likely functions as a general base and Tyr258 is positioned in line with the glycosidic bond between subsites -1 and +1, which means it is likely to be a general acid that protonates the O4 during cleavage (figure 25).
Kinetic Characterization of Active Site Mutants

To confirm the proposed functions of the active site residues in catalysis, kinetic studies were carried out on site specific mutants that targeted these residues. Consistent with the role of Tyr250 as the general base and Tyr258 as the general acid, mutations to these residues completely abolished activity. Although our initial assumption of the two tyrosine residues involvement in acid/base chemistry was correct, another mutant Y261A was also inactive, whose proximity to a non-reducing manuronate through stacking interactions is seen. This means that Tyr258 is more likely a general acid and not a base. His202 was presumed to be the general base; however, the H202L mutant was catalytically active and our substrate bound variant Y258A supports this. The function of H202 is believed to instead work with Asn201 to stabilize the substrate in position by stabilizing the charge on the C6 carboxylate. Residues that were initially believed to stabilize the oligosaccharide (Asn149, Arg260, His413) are negatively affected, with the $k_{cat}/K_m$ values magnitudes lower than that of the wild type. These mutants still show activity though, as it believed that the function of these residues is redundant amongst each other.

Mechanistic Basis for the Exolytic Mode of Alginate Depolymerization

Alg17c mechanisms follows that of an oligosaccharide lyase, which requires active site residues that participate in acid/base chemistry and stabilize the negative charge on the C6 carboxylate (figure 25). The mechanism uses basic residues to lower the pKa of the C5 proton to promote
proton abstraction by a general base and protonation of the O4 atom by a general acid during cleavage of the glycosidic bond. We can see this in many lyases, such as chondroitin AC lyase and hyaluronate lyase, where the C6 carboxylate is stabilized instead by an amide and/or histidine that forms a hydrogen bond with this moiety.

The co-crystal structure shows that Asn201 and His202 are near the C6 carboxylate and likely work together to neutralize the negative charge of the uronic acid, which is confirmed by mutational analysis. Tyr450 functions as a general base to remove C5 protons to promote glycosidic bond cleavage. Kinetic analysis further supports this assumption. For two tyrosine residues Tyr258/450 to act as a general acid/base respectively, is an unusual feature within a multi-domain (α/α)n toroid lyase superfamily. Giving that the phenolic side chain has a pKa of around 10, tyrosine is an odd choice to function as a general base. Compared to Atu3205, His311 functions as the catalytic base. The explanation for this is that the environment is having an impact on tuning the pKa of both tyrosine residues that can accommodate their unique function.

1.4 CONCLUDING REMARKS

Our studies provide the first crystallographic structure of the Pl-17 family of Alg17c from S. degredans. We also present the first detail structure-function characterization of active-site mutants, which was identified through our co-crystal structures that established the various residues involved alginate depolymerization. Our results also demonstrate a convergence both
in structure and function amongst the large class of lyases that fall in the multi-domain (α/α)_n toroid lyase superfamily. In spite of global similarities, there are a considerable number of differences that exist. This includes the identity of the catalytic residue and these can’t be identified by simple sequence or structure alignments. Additional studies of enzymes within this structural class will probably give new insight into the mechanism of oligosaccharide depolymerization.
Table 1: Steady state kinetic parameters for Alg17c wild-type and subsequent mutations.

<table>
<thead>
<tr>
<th>Active Site Residues</th>
<th>KM (µM)</th>
<th>Vmax (µM s⁻¹)</th>
<th>kcat (s⁻¹)</th>
<th>kcat/KM (x10² M⁻¹ s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>21.6 ± 2.73</td>
<td>34.9 ± 2.18</td>
<td>56.9 ± 3.6</td>
<td>25.7 ± 1.4</td>
</tr>
<tr>
<td>N201A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Y258A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Y450F</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R280A</td>
<td>12.1 ± 3.1</td>
<td>0.0092</td>
<td>0.0149 ± 0.002</td>
<td>0.016 ± 0.003</td>
</tr>
<tr>
<td>R438A</td>
<td>31.1 ± 5.9</td>
<td>10.9 ± 0.88</td>
<td>3.63 ± 0.29</td>
<td>0.013 ± 0.001</td>
</tr>
<tr>
<td>Q149A</td>
<td>31.1 ± 5.9</td>
<td>0.011 ± .00008</td>
<td>0.018 ± 0.001</td>
<td>0.0058</td>
</tr>
<tr>
<td>Catalytic Acid or Base</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H202L</td>
<td>91.6 ± 18.6</td>
<td>6.7 ± 0.2</td>
<td>10.9 ± 0.3</td>
<td>1.2 ± 0.2</td>
</tr>
<tr>
<td>H413A</td>
<td>85.1 ± 19.51</td>
<td>15.1 ± 1.08</td>
<td>5.02 ± 0.36</td>
<td>0.8 ± 0.2</td>
</tr>
<tr>
<td>Zinc Binding Domain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H415A</td>
<td>21.6 ± 2.7</td>
<td>0.035 ± 0.002</td>
<td>0.057 ± 0.004</td>
<td>0.026 ± 0.002</td>
</tr>
</tbody>
</table>
Figure 18: The TLC analysis of polymer specific for Alg17c. Stands for the monomer are D-glucose in lane 1 and dimer standard is D-Trehalose. A mixture of di-, tri-, and tetra-saccharides are in lane 3 (before enzyme incubation), and the processed mixture after digestion of mainly mono- and di-saccharides are located in lane 4 (After enzyme incubation).
Table 2: Data collection, phase, and refinement statistics. Highest resolution shell is shown in parenthesis. Figure of merit- probability weight average of the cosine of the phase error, before and after density modification. $R$-factor = $\sum (|F_{obs}| - k|F_{calc}|)/\sum |F_{obs}$ and the $R$-free is the $R$ value for a test set of reflections consisting of a random 5% of the diffraction data not sued in refinement.
Figure 19: Overall structure of an Alg17c monomer with the α-barrel domain (green) and antiparallel β-sheet domain (blue). A) is the cartoon representation and B) is the surface representation.
Figure 20: A) Structure with a bound zinc (red) and B) close up of the three residues (cyan) that are interacting with the bound zinc (red): His 415, Asp 433, and His 464. Notice the nitrogen and oxygen atoms rotated to accommodate the binding of the zinc atom.
Figure 21: Alg17c shares similar structural features as those of other polysaccharide lyases despite the diverse enzymatic activity. A) Alg17c colored in green and blue with red Zn$^{2+}$, B) Heparinase II colored in purple and pink with red Zn$^{2+}$ (PDB 2FUQ), and C) Atu3025 colored in yellow and green (PDB 3A0O).
Figure 22: ITC analysis, which characterize the binding interaction between a ΔMMG trisaccharide with a Y258A mutant of Alg17c. Data was fit to a single site model.
Figure 23: Alg17c with bound trissachride MMG. A) Cartoon representation to show the location of the active site with bound substrate (pink). B) Surface representation to show the binding pocket location of Alg17c. C) Active site containing all interacting residues (silver) with substrate (pink).
Figure 24: A) Comparison of the overall structure of Alg17c (green) and Y258A (silver) with bound tri-saccharide (pink). RMSD = 1.049. B) The active site overlay with Alg17c (green) and Y258A (silver). Note the large shift with some of the amino acids in the bound form, especially the large contortion of the Lys146.
Figure 25: This is the proposed mechanism of alginate depolymerization by Alg17c. Asn201 and His202 pair together to stabilize the negative charge at the carboxylate, which Tyr450 acts as a general base to abstract the proton at C5. Tyr258 serves as a general acid to donate a proton to the oxygen of the glycosidic bond.
CHAPTER 2: BIOCHEMICAL CHARACTERIZATION OF PLUMBEMYCIN BIOSYNTHESIS

2.1 GENERAL INTRODUCTION

There is a need for a new and cheaper source of antibiotics. Resistance to current antibiotics is continuing to be a major problem; therefore, there is a need to quickly deploy new antibiotics to combat the ever growing number of pathogens appearing. Natural product biosynthesis has shown have much potential in terms of readily accessible novel antibiotics. The key to fully taking advantage of this relatively newly discovered source is to be able to deduce the biosynthetic pathways for these compounds. One much underexploited group of bioactive compounds that show great potential in treating against pathogen are phosphonate based compounds.

These molecules are generally termed C-P compounds and have similar structure to phosphate esters and anhydrides, but contain C-P bonds in the place of the usual oxygen-phosphate bond. Because of the labile nature of the C-O-P bond, the phosphonate compounds have shown to have great resiliency in the presence of heat, acid/base exposure, and enzymatic digestion, as well as the ability to structurally mimic compounds in nature. (Metcalf and van der Donk 2009) Examples of these include FR-900098 which is an N-hydroxypropylphosphonic acid that functions as an anti-malarial compound which inhibits deoxyxylulose phosphate reductoisomerase (Shigi 1989) (figure 26A) and dehydrophos, which is a phosphonate isolated from *Streptomyces luridus* that is shown to inhibit pyruvate utilizing enzymes (Whitteck, Ni et al. 2007) (figure 26B). Other examples include Fosfomycin (1R,2S-Epoxypropylphosphonic acid) isolated from *Streptomyces fradiae, S. wedmorensis, S. viridochromogenes* (Hendlin, Stapley et
al. 1969), *P. syringae* (Shoji, Kato et al. 1986), and *P. viridiflava* (Katayama, Tsubotani et al. 1990), an FDA approved drug against uncomplicated acute cystitis (Bailey 1993) (figure 26C). Another is phosphinothricin-tripeptide is produced by *Streptomyces viridochromogenes* and is shown to have very high antibacterial activity *in vitro* (Thompson and Seto 1995) (figure 26D).

Plumbemycin A and B are oligopeptide, antibacterial phosphonates from *Streptomyces plumbeus* containing APPA (2-amino-5-phosphono-3-pentenoic acid), and acts as L-threonine antagonist that inhibits the synthesis of threonine. This unique mechanism is due to the structural similarity to phosphohomoserine (ester of serine and phosphoric acid), which is normally cleaved by threonine synthase to produce threonine (figure 27) (Gahungu, Arguelles-Arias et al. 2013). The key difference being is that there is a C-P bond in APPA instead of C-O-P moiety that is normally found in phosphohomoserine. Activity against bacteria was first noticed when crude supernatant of *Streptomyces plumbeus* showed activity against test strain of *E. coli* grown on minimal media (Park, Hirota et al. 1976). Its structure is similar to that of another APPA containing compound known as Rhizoctin, which exhibits antifungal activity through the same mechanism (Figure 28) (Borisova, Circello et al. 2010).

Both compounds contain proteogenic amino acids located on their N-terminus, which function as identification markers for uptake by target host’s oligopeptide transport system. From there the host peptidases will cleave the oligopeptide, releasing APPA. Rhizoctin contains a conserved arginine and a variable amino acid (denoted by single letter notation) and plumbemycin contains a conserved alanine with an asparagine/aspartate at its N-terminus. In our research, we are focusing on elucidating the early steps of Plumbemycin synthesis. The
complete biosynthetic pathway for Plumbemycin B is currently uncharacterized, though a proposed pathway has been formulated based on the pathway of Rhizoctin (figure 29).

We are making the assumption that the pathways are similar due to the APPA moiety being the same between plumbemycin and rhizoctin. Extensive work by Dr. Svetlana Borisova from the William Metcalf lab sequenced the gene clusters containing enzymes involved in the biosynthesis of both compounds. From that they were able to show homology between the proteins and functions of already characterized enzymes by bioinformatics. The enzyme we are focusing on is PluG, a putative class II aldolase which promotes an aldol condensation reaction between a phosphonoacetaldehyde (PnAA) and pyruvate generated from an oxaloacetic acid (OAA). PluG is shown to form a reactive hydroxyl intermediate (2-keto-4-hydroxy-5-phosphonopentanoic acid), which is an important precursor for the formation of the 3,4-cis double bond of APPA. However, the product of the PluG reaction can undergo an irreversible, non-enzymatic dehydration reaction which forms an off pathway end product. PluG was shown to be insoluble during initial purification steps; therefore, we attempted to crystallize orthologue of PluG isolated by Dr. Kou-San Ju from Metcalf lab. We obtained a high resolution structure (1.7 Å) of a PluG ortholog from *Saccharothrix aerocolonigenes* (P325), which shows an uncommon c-terminal domain, which we hypothesize to promote the formation of a complex with a subsequent enzyme from the plumbemycin biosynthetic pathway, which we propose to be either amino-transferase PluJ or dehydrogenase PluI.
Figure 26: Structures of phosphonates: A) FR900098 is an antimalarial compound. B) Dehydrophos inhibits pyruvate utilizing enzymes. C) Fosfomycin treats acute cystitis. D) Phosphinothricin is shown to have excellent in vitro activity against bacteria.
Figure 27: A) Threonine synthase catalyzing the removal of the phosphonate off HSerP to form Threonine. B) APPA Inhibiting Threonine synthase activity (Notice the C-P bond on APPA verses the C-O-P bond in (S)-HSerP.)
Figure 28: Structures of Rhizoctinin and Plumbemycin. Rhizoctinin contains a conserved arginine and variable amino acid on its n-terminal end. Plumbemycin contains a conserved alanine and variable amino acid. In circled is the common APPA “warhead” that is responsible for inhibiting threonine synthase.
Figure 29: Proposed pathway of plumbemycin biosynthesis based off the original proposed biosynthetic pathway of rhizoctin
2.2 MATERIALS AND METHODS

Cloning and Transformation

PluG, Plul, and PluJ from *Streptomyces plumbeus*, and P325 PluG a PluG ortholog from *Saccharothrix aerocolonigenes* already cloned into a Pet28 vector by Dr. Svetlana Borisova and was isolated by Dr. Kou-San Ju from William Metcalf Lab. Cloning and bioinformatical analysis were carried out by Dr. Svetlana Borisova and Dr. James Doroghazi. The PluG, Plul, and P325 clones transformed into rosetta DE3, while the PluJ construct was transformed into BL21DE3, were given to us by Dr. Vinayak Agarwal, a previous student in the lab. The cloned gene was inserted into a pET-28 vector.

RhiG, Rhil, and RhiJ were cloned from *Bacillus subtilis* into a pET-28b vector by Dr. Vinayak Agarwal after received in the genomic DNA from the Metcalf lab. RhiG was then transformed into rosetta DES for protein expression. Rhil and RhiJ were transformed in to BL21DE3 cells for protein expression.

Culturing and Harvesting

Transformed *E. coli* BL21 Rosetta DE3 containing the P325 PluG ortholog, were grown in 2 L cultures in LB media (BD Biosceinces, Cat. No. 244610). The cultures were grown for 4 hours at 37°C at 200 rpm to OD 0.8. The cultures were then induced with 0.1 mM IPTG and grown overnight at 18°C. The next day the cells were harvested by centrifugation using the Sorvall RC-3B Plus with 1L bucket rotor from Thermo Scientific at 3,500 rpm for 25 minutes. The resulting
pellet was suspended in a buffer solution containing 500 mM NaCl, 20 mM Tris pH 8.0, and 10% glycerol. Depending on immediate or later use, the pellet was either immediately lysed or frozen in liquid nitrogen and stored at -80°C.

Selenomethionine (SeMet) cultures were grown in 2L of M9 minimal media made with 22.4 g Na$_2$HPO$_4$·7H$_2$O, 6 g KH$_2$PO$_4$, 2 g NH$_4$Cl, and 1 g NaCl. This was autoclaved at 121°C for 30 minutes and allowed to cool. The media was then further supplemented with 25% glucose, 0.1 mM MgSO$_4$, 4.2 mg/mL FeSO$_4$, and 0.05% thiamine, along with 20 mL of previously grown start up culture. Cultures were shaken vigorously at 37°C and grown to OD ~0.5, where then a 20 mL amino acid mix of 0.1 g/10 mL of Lysine, Phenylalanine, Threonine, and 0.05 g/10 mL of Leucine, Isoleucine, Valine are added and 15 minutes later, selenomethionine is added at 0.06 g/10 mL. After another 15 minutes, 0.1 mM IPTG is added and the cultures are shaken overnight at the same speed at 18°C.

To promote complex formation of both the RhiGIJ and PluGIJ complexes, the individual proteins of PluG/RhiG, Plul/Rhil, and PluG/RhiG were grown separately in 1L cultures of TB (Terrific Broth) media (BD Biosciences, Cat No 243820). The individual cultures were grown for 8 hours at 37°C to an OD of >2.0 and induced with IPTD 1mM. Cells were then harvested and collected into the same conical with resuspension buffer containing 500 mM NaCl, 20 mM Tris pH 8.0, and 10% glycerol.
Protein Purification

The resuspended pellet was lysed using an EmulsiFlex-C5 Homogenizer from AVESTIN (Ottawa, Canada). Pressure was approximately 50,000 kPa. This continued for 15 minutes or until the cells were fully lysed. The resulting lysate was then spun down at 16,000 rpm for 1 hour at 4°C using a Beckman J2-21M/E centrifuge from Beckman Coulter Life Sciences (Indianapolis, IN). The supernatant containing the soluble fraction was injected into a 5 mL Hi-Trap Ni-NTA, containing immobilized metal ion affinity resin (G.E. Healthcare), which was pre-equilibrated with binding buffer of 50 mM Tris-HCl at pH 7.5, 300 mM KCl, and 20 mM imidazole. After the sample was loaded, the protein was eluted by linear gradient from 20 mM imidazole to 200 mM imidazole over the course of 30 minutes using an AKTA UPC-900 FPLC (G.E. Healthcare Life Sciences, Pittsburgh, PA). Fractions were selected based on high UV signal intensity generated by aromatic amino acids tryptophan, tyrosine, and phenylalanine at approximately 280 nm. Selected fractions were then run on a polyacrylamide, SDS-PAGE gel for 1 hour to observe sample purity. Fractions containing the most pure sample were then incubated with thrombin overnight in a buffer containing 100 mM KCl, 20 mM Tris at pH 8.9 to remove imidazole and the N-Terminal polyhistidine tag.

After overnight dialysis, the sample was injected onto a Superdex Hiload™ 75 16/60 size exclusion column (GE Healthcare, Pittsburgh, PA). The column was pre incubated in a buffer containing 20 mM HEPES pH 7.5 and 100 mM KCl. Injection of the sample on the column was done at using an AKTA UPC-900 FPLC (G.E. Healthcare Life Sciences, Pittsburgh, PA) at a flow rate of 1 mL/min. Slow elution was carried out over the course of 2 hours using the same
buffer. Samples were then concentrated to 10 mg/mL using a 10,000 MWCO Amicon centrifugal filter (Merck Millipore, Carrigtwohill, CO).

To promote complex formation of PluGJJ, the resuspended cells were collected and lysed in separate conicals. The lysates were purified separately through a column containing Ni resin and eluted using a buffer containing 20 mM Tris, 1M NaCl, 250 mM imidazole, pH 8.0. All fractions collected were run on a SDS-PAGE polyacrylamide gel (12% acrylamide). Fractions based on purity were then dialyzed overnight with thrombin to remove the polyhistidine tag of PluJ. The buffer used in dialysis was 100mM NaCl, 20mM Tris, pH 8.9. The following day the fractions were combined into one conical, then incubated with 5 mM MgCl$_2$ for 5 hours. The protein mixture was then purified by size exclusion chromatography, using the Superdex 200 10/300 GL (from GE Healthcare) and the buffer used was a low salt buffer (100 mM KCl, 20mM Hepes pH 7.5). 6 mls of dialyzed protein was loaded. The absorbance curve generated. Fractions that contained the target peak (4 fraction at 5mls each) were collected and ran on an SDS-PAGE polyacrylamide gel with the same standards of the previous runs.

To observe prospect in vivo factors that may constitute complex formation, the native host, *B. subtilis* ATCC 6633, which was shown to actively product APPA, was grown in 1L of LB media over the course of 72 hours at 30°C. The cells were then lysed, with the supernatant collected and incubated with 100 µM of PMSF (phenylmethanesulfonylfluoride) to inhibit host strain proteases. The supernatant was then collected and incubated with the RhiG, RhiI, and RhiJ-HisTag mixture for 18 hours and subsequent reactions were purified on a nickel column and eluted using a buffer containing 20 mM Tris pH 8.0, 1M NaCl, 250 mM Imidazole. All fractions
were collected and ran on a SDS-PAGE polyacrylamide gel. Afterwards, selected fractions were
diaiyzed overnight and incubated with thrombin to remove this HisTag off RhiJ. The following
day the fractions were further purified by size exclusion chromatography, using the Superdex
200 10/300 GL (from GE Healthcare) and the buffer used was a low salt buffer (100 mM KCl,
20mM Hepes pH 7.5). 6 mls of dialyzed protein was loaded.

**Crystallization and Structure Determination**

Crystallization of the native and SeMet P325 PluG was incubated at 9°C at 8mg/mL overnight.
The SeMet screening condition contained 0.1 M Tris pH 8.0, 0.2 M LiSO₄, and 10% Peg 5000
MME (Monomethyl either). The cryo used to flash freeze the crystals contained 10% glycerol
and PEG 5000 MME at 20% instead of 10%. The native crystals grew and were frozen in the
same conditions, but with 5 mM MgCl₂. Crystallization of the prospective PluGIJ complex was
incubated at 9°C at 15 mg/mL for two days. The condition of the screen is 0.2M NaCl, 0.1M
Sodium Cacodylate pH 6.5, and 2M Ammonium Sulfate. The cryo used to flash freeze the
crystals contained 30% trehalose in addition to the original screen condition.

Flash-cooled crystals of wild-type P325 PluG were diffracted X-rays to 1.7 Å respectively, using a
Mar 300 CCDetector at an insertion device synchrotron beam line (LS-CAT Sector 21 ID-G,
Advanced Photon Source, Argonne, IL). All data were indexed and scaled using either the
HKL2000 package or XDS. Crystallographic phases were determined using the single wavelength
anomalous dispersion (SAD) using data collected from a crystals of SeMet labeled P325 PluG, with
Diffraction data collected to and 1.9 Å. and processed as described above. Heavy atom sites were determined using HySS and imported into SHARP for maximum likelihood refinement. The resultant phases were further improved by two-fold non-crystallographic symmetry averaging, and subject to automated model building as implemented in ARP/wARP. An initial model containing most of the main chain and side chain atoms was used for refinement against the higher resolution native data set. The remainder of the model was fitted using Coot, and further improved by rounds of refinement with REFMAC. Cross-validation was routinely used throughout the course of model building and refinement using 5% of the data in the calculation of the free R factor.

**Activity analysis of P325 PluG.**

A reaction mixture containing 40 μM PnP decarboxylase, 80 μM RhiH, 45 μM of either RhiG (positive control) or P325 PluG, 12 mM of OAA (freshly prepared), 10mM PEP, 1mM TPP, 10mM MgCl₂, and diluted in 50mM HEPES, pH7.5. The reaction was carried out at 30°C for 1 hour. The mixture became precipitated during the incubation and was removed by a Micron Ym-30 column at 13,200 RPM. The filtrate (volume at 500ul) was then characterized by the presence of phosphonates using ³¹H decoupled ³¹P NMR in 20% D₂O (lock solvent). The reference was to 85% phosphoric acid standard at 0 ppm. The spectra were acquired on the Varian Unity Inova-600 spectrometer at room temperature at the Institute of Genomic Biology at the University of Illinois at Urbana-Champaign.
Samples of these same reactions were also given to Dr. Jae-Heon Lee to observe product formation by mass instead of the presence of phosphonate. The instrument used was the Agilent 5975B series Gas Chromatograph/Mass Spectrometer (GC/MS) (Agilent Technologies, Santa Clara, CA). Samples observed involved the reactions that were carried over the course of 24 hours (both the RhiG control and target reaction) and were desalted before being analyzed.

2.3 RESULTS AND DISCUSSION

Overall Structure of P325 PluG.

The structure of P325 PluG was determined to 1.7 Å resolution (Figure 30), using phases determined by single wavelength anomalous diffraction measurements collected using crystals of SeMet labeled protein. The final model for each of the structures consists of all residues spanning Ala4 through Cys330. The overall structure of P325 contains two domains: an imperfect TIM barrel (residues Ala4 through Lys 269) with a C-terminal domain (Pro 270 through Cys 330) (Figure 31). Nine alpha helices surround a β-barrel at the amino-terminus, showing a topological arrangement similar to that of already characterized TIM barrels such as triose phosphate isomerase (PDB code 8TIM) with the exception of the extra helix (figure 32). There is also the presence of Magnesium bound and coordinated by residues Asp 16, His 196, and His 198 (Figure 33).
Presence of a unique C-terminal Domain.

The C-terminal domain is comprised of an α-helical bundle, similar to that found on already characterized aldoase DmpG (PDB code 1NVM) (figure 34). DmpG is a 4-hydroxy-2-oxovalerate aldoase, which catalyzes the cleavage of 4-hydroxy-2-oxovalerate to acetaldehyde and pyruvate (Manjasetty, Powlowski et al. 2003) (Figure 35). The c-terminal domain is shown to be a communication domain, which promotes the formation of a tightly bound complex between DmpF (dehydrogenase) and DmpG (aldolase) Manjasetty, Powlowski et al. 2003). When the complex forms, a hydrophobic tunnel is created in the enzymes, which allows for the shuttling of the reactive acetaldehyde intermediate from one active site (DmpG) to the other (DmpF) (Figure 36). This prevents the acetaldehyde from escaping into the internal cellular environment, where it can cause oxidative damage.

Further proving this point, DmpG and DmpF by itself was shown to have little to no enzymatic activity (Manjasetty, Powlowski et al. 2003). Due to the structural similarities between P325 PluG and DmpG, we propose that P325 PluG forms a similar complex with either a PluJ ortholog (aminotransferase) or Plul ortholog (Dehydrogenase), or even both Plul/J from the same gene cluster (Figure36). This may be important to preventing the off pathway 3,4 trans double bond product that was shown to form non-enzymatically in the rhizoctin biosynthetic pathway. Individually P325 PluG was shown to have no activity (Figure 37), which compares P325 PluG with that of RhiG. As we can see with the figure, P325 PluG does not have activity by itself. The 24 hour reactions were also analyzed by mass spectrometry.
Analysis by Dr. Jae-Heon Lee showed by GC/MS shows the presence of startup product as well as converted OAA; however, there is a lack of the target trans-double bond product, which leads us to believe the turnover for this transformation is extremely slow (figure 38). Also important to note that the mass for the 4-hydroxyle intermediate was extremely low despite the fact there was the presence of a split peak. Due to lack of signal intensity, we cannot say with confidence that P325 had activity. We apply this theory to both the Rhizoctin and Plumbemycin biosynthetic pathways, where the enzymes involved with the APPA moiety formation individually have yet to be enzymatically characterized. Due to the energetically unfavorable formation of the cis-3,4 double bond from the 4 hydroxyl intermediate and the threat of the non-enzymatic inactive trans double bond product, it may be important for RhiG and PluG to respectively form complexes with RhiJ/I and PluJ/I. This will allow for the formation of a shuttling tunnel similar to that of DmpG and DmpF (figure 39).

With this assumption, we believe that there are three ways the PluGIJ and RhiGIJ complex can form. One is that first PluG/RhiG and PluI/RhiI form a complex to form the cis-3,4 double bond intermediate, which then the complex with PluJ and RhiJ forming APPA (figure 40). Another way is that PluG/RhiG and PluJ/RhiJ form a complex to form the tran-animated intermediate, which then the following complex with Rhil/PluI forms APPA (Figure 41). Or the third assumption is that all three proteins for a complex to make APPA (Figure 42).
**Putative Complex Formation**

Evidence for prospective PluGIJ complex formation was discovered after incubation of the three proteins together following overnight dialysis. After further purification of the protein mixture using size exclusion, we can see the presence of one peak that based on retention volume is possibly a two or three protein complex. This is also further validated when comparing this peak to the peaks generated by the individual proteins as a control (figure 43). The fractions within the retention volume of the peak between 40 mL and 60 mL were run on the SDS-PAGE electrophoresis gel and visualized using coomassie brilliant blue dye. It appears that PluG and PluI have the strongest association with each other, which PluJ appears co-elutes with the other two proteins to an extent (figure 44). Fraction 6 was then taken and concentrated down to 15 mg/mL and screened for protein crystals. Crystals grew after two days and flash frozen (figure 45). Flash frozen crystals were then taken to the insertion device synchrotron beam line at Sector 21 ID-G LS-CAT (Advance Photon Source, Argonne, IL) using a Mar 300 CCD detector. Unfortunately, diffraction was quite poor and the data appeared to be extremely split. Resolution was approximately 3.5 Å but was ultimately unable to be processed.

Evidence for a RhiGIJ complex also appeared. Initially, attempts to reconstitute the prospective complex using the same method as PluGIJ did not work, leading to a failure for the proteins co-eluting together. However, when the mixture was incubated with the lysate from ATCC 6633, there was evident of co-elution when the sample was co-eluted off the nickel column (Figure 46). The bands that eluted off the column match the standards migration distance of each individual protein. The gel filtration chromatograph also shows the presence of a strong affinity between
the three proteins as they elute together at a retention volume representative of the complex molecular weight of approximate 100 kDa (figure 47).

Fractions generated from the elution volume of the peak (between 40-60 mL) were run on a SDS-PAGE gel, with results once again co-elution showing that there is a small factor in the ATCC 6633 supernatent that promotes a strong affinity between RhiG, Rhil, and RhiJ (Figure 48). This may be due to a small factor that the native host alone produces. These factors have yet to be confirmed (other proteins, small compounds, etc.), but are considered vital for complex formation, as previous in vitro methods that were carried out without the native host lysate failed to generate the same results.

2.4 CONCLUDING REMARKS

The main challenge remaining is to be able to reconstitute activity with both the PluGIJ and RhiGIJ complexes. Though we can see some sort of association, we have yet to obtain clear activity and kinetics that demonstrates the enzyme’s abilities to carry out the proposed activities we have stated for them. Though RhiG/PluG has previously been shown to have the ability to carry out an aldol condensation reaction on its own, the other two enzymes Rhil/Plul and RhiJ/PluJ have yet to show detectable activity on their own. The next challenge is to be able to identify said components needed to generate activity for the two proposed protein complexes. The eventual goal is to generate a crystal structure of this complex with all co-
factors bound, which would allow us to better understand how the complex forms and to elucidate the APPA biosynthetic pathway of both Rhizoctin and Plumbemycin.
Figure 30: P325-PluG diffraction to 1.7 Å resolution.
Figure 31: Structure of P325-PluG. The structure comprises of two domains: an imperfect TIM barrel (green) and a C-terminal α-helical domain that is proposed to be a communication domain (orange). Bound magnesium is also present in blue.
Figure 32: Comparison of a normal TIM barrel of Triose Phosphate Isomerase (red) versus P325 PluG (imperfect TIM barrel in green). P325’s TIM barrel has an extra helix in addition to a unique C-terminal domain (orange)
Figure 33: The bound metal Mg$^{2+}$ (blue) is shown in blue coordinating with three residues: His196, His198, and Asp16. This metal ion is proposed to serve as an electron sink for the aldolase condensation reaction.
Figure 34: P325 and DmpG share a similar C-terminal domain and structure besides having no sequence similarity. The C-terminal domain of P325 (orange) is proposed to function as a communication domain as seen with DmpG’s C-terminal domain (purple). Activity is shown to be dependent on this domain working to form a complex with a subsequent enzyme.
Figure 35: A structure of the DmpGF Complex (PDB 1NVM). DmpG (blue and purple) forms a tight complex with DmpF (red). Interaction occurs at the C-terminal domain (purple). It is important to note that both only completely active when in complex.
Figure 36: Reaction mechanisms: cleavage catalyzed by DmpG to produce free acetaldehyde and DmpF catalyzes a dehydration to produce free acetyl-CoA.
Figure 37: Complex forms a hydrophobic tunnel that allows for movement of a reactive acetaldehyde between the DmpG and DmpF active sites. This is to prevent the release of the acetaldehyde intermediate into the cellular medium. Manganese ion is in red and the NAD and Oxalate are colored in yellow.
Figure 38: Activity Assays analyzed with P NMR, comparing the activity of RhiG with that of P325 PluG after 1 hr and 24 hr incubations at 30°C. No 4-hydroxyl intermediate is detected with P325 PluG, with RhiG acting as the positive control.
Figure 36: Mass Spectrometry measurements carried out by Jae-Heon Lee. There is a lack of the 3,4 trans-double bond product but this is due to the possibly very slow turn over of this product. Also though there is a split peak for the 4-hydroxyl intermediate in the PluG reaction, the signal was extremely low, leading to the possibly that PluG-P325.
Figure 40: Because of the structural similarities with DmpGF complex, we propose that PluG may form a similar complex with either Plul or PluJ due to the formation of a trans by-product that forms non-enzymatically over time. This functions to shuffle the condensation intermediate to a second active site.
Figure 41: Proposed complex formation. One is that first PluG/RhiG and PluI/Rhil form a complex to form the cis-3,4 double bond intermediate, which then the complex with PluJ and RhiJ forming APPA. Another way is that PluG/RhiG and PluJ/RhiJ form a complex to form the tran-animated intermediate, which then the following complex with Rhil/Plul forms APPA.
Figure 42: Proposed complex formation. The third assumption is that all three proteins for a complex to make APPA.
Figure 43: The gel filtration chromatographs for the fractions after dialysis. Top chromatograph is for the putative PluGIJ (black curve). Bottom chromatograph contains the three individual proteins of the putative complex. Each peak is represented by PluG (orange), PluI (grey), and PluJ (yellow).
Figure 44: SDS-PAGE gel following size exclusion chromatography. PluG and PluI have the strongest association with each other as evident on fraction 6, which PluJ co-elutes to an extent.
Figure 45: Crystals of the putative PluGIJ Complex.
Figure 46: Standards of separately purified RhiG, RhiI, and RhiJ are located on the left. Flow is the excess mixture from the Ni column and Wash is the excess protein and other aggregates that do not bind tightly. The SDS ladder is the SDS-PAGE Molecular Weight Standards, Broad Range from Bio-Rad (catalog number 161-0317). Fractions 1-6 are represented on the right.
Figure 47: Gel filtration chromatograph shows main peak appears at approximately 47-50 ml of elution volume.
Figure 48: Standards of separately purified RhiG, RhiI, and RhiJ are located on the left. The SDS ladder is the SDS-PAGE Molecular Wiegght Standards, Broad Range from Bio-Rad (catalog number 161-0317). Fractions 1-4 from the gel filtration are represented on the right.
CHAPTER 3: STRUCTURAL STUDIES ON AA3-MENAQUINOL OXIDASE FROM BACILLUS SUBTILIS

3.1 GENERAL INTRODUCTION

Bacteria exhibit an ability to be considerably flexible with their respiratory chains. They are able to vary the expression of the proteins involved with respiration to adjust the energy production based on the environmental needs of the host (Lauraeu 1993). This apparatus is called the electron transport chain or ETC, which is a network of proteins that couples the movement of electrons from donor to acceptor with the pumping of protons across the membrane. This electrochemical proton gradient is what drives the synthesis of adenosine triphosphate (ATP).

In eukaryotes, the ETC is located at inner membrane of the mitochondria, while in prokaryotes the ETC is found imbedded in the cell wall. In the case of plants, chloroplasts utilize photons from light to convert water and NADP+, to oxygen and NADPH respectively (Anraku 1988). In the case of mitochondria based ETC’s, it is oxygen and NADH being converted to water and NAD+ respectively, with the addition of the conversion of succinate to fumarate (Anraku 1988) (figure 49).

There are four membrane bound enzymes that are important to the ETC in both mitochondria and bacteria: Complex I, Complex II, Complex III, and Complex IV. In Complex I (known as the NADH dehydrogenase), two electrons are transferred from NADH to ubiquinone (Q) (Lauren 2010). It is here NADH is oxidized into NAD+ by reduction of Flavin mononucleotide to FMNH₂, where then FMNH₂ transfers two electrons to an Fe-S cluster (Lauren 2010). Then the Fe-S cluster transfers the electrons to Q, which results in the Q forming QH₂ or ubiquinol (Lauren...
The QH$_2$ diffuses across the membrane, while at the same time there are four protons translocated across the membrane, resulting in a proton gradient (Lauren 2010) (Figure 50).

Complex II or the succinate dehydrogenase delivers two additional electrons to the Q pool. The two main actions of complex II are the conversion of oxidized succinate to fumarate by actions of Flavin adenine dinucleotide (FAD) and to reduce coenzyme Q$_{10}$ (CoQ) to CoQH$_2$ (Grisham 2010). CoQH$_2$ is responsible for driving the action of complex III (Grisham 2010) (Figure 51).

Complex III (cytochrome bc$_1$ complex) is the location of the Q-cycle, where cytochrome (cyt) b binds to an ubiquinol and a ubiquinone, where 2Fe-S centers and the Bl heme pull an electron off the ubiquinol, which releases two protons into intermembrane space (Ferguson 2002) (figure 52). From there one electron is transferred cyt c$_1$ while another is transferred to the BH heme. Cyt C$_1$ then transfers its electron to cyt c, and BH heme transfers its electron to ubiquinone, resulting in the formation of a ubisemiquinone (Ferguson 2002). Here is where Cyt c diffuses across the intermembrane space and the ubiquinone is released while the semiquinone remains bound (Ferguson 2002). A second ubiquinol is bound by Cyt B, where 2 Fe-S centers and Bl heme again pull an electron off the ubiquinol, releasing two protons into the intermembrane space (Ferguson 2002). One electron is then transferred to cyt c$_1$ while another is transferred from Bl heme to BH heme (Ferguson 2002). Cyt c$_1$ then transfers its electron to cyt c, while the already present seminquinone picks up the electron from the BH heme and the two protons (Ferguson 2002). The second ubiquinone along with the new ubiquinol are released (Ferguson 2002).
Complex IV or the Cytochrome c oxidase contains two heme centers, \textit{cyt} \textit{a} and \textit{cyt} \textit{a}_3, and two copper centers \textit{Cu}_A and \textit{Cu}_B, with each of the copper centers associated with one of the cytochromes (\textit{Cu}_A associates with \textit{Cyt} \textit{a} and \textit{Cu}_B associates with \textit{cyt} \textit{a}_3) (Yoshikawa, Shimada et al. 2015). These copper sites function as electron carriers, transferring an electron at a time. \textit{Cyt c} transfers its electron to \textit{Cu}_A and gets oxidized in the process, which causes it to dissociate (Yoshikawa, Shimada et al. 2015). \textit{Cyt a} transfers an electron to \textit{Cu}_B, where then a second \textit{cyt c} binds an transfers its electron to \textit{Cu}_A (Yoshikawa, Shimada et al. 2015). This electron is then transferred to \textit{cyt a}, which is then transferred to \textit{cyt} \textit{a}_3, where the binuclear metal center of \textit{Cu}_B and \textit{cyt} \textit{a}_3 resulting in the possession of two electrons now and the allowed binding of \textit{O}_2 (Yoshikawa, Shimada et al. 2015). Two protons are then pushed through the membrane, with the addition of another electron from the same previously mentioned pathway. A fourth electron is then exchanged for a hydroxide at the heme center, which causes it to become protonated and dissociate as \textit{H}_2\textit{O} (figure 53).

There is one interesting factor that appears though. A number of prokaryotes also contain heme-copper oxygen reductases that utilize the quinol as a substrate electron (Pereira, Santana et al. 2001). These enzymes are known as quinol oxidases and are closely related to the cytochrome \textit{c} oxidase, with the key difference being that these quinol oxidases lack \textit{Cu}_A binding motif that is present in the usual cytochrome \textit{c} oxidase (Yi, Narasimhulu et al. 2010). The most heavily studied quinol oxidase currently is the cytochrome \textit{bo}_3 ubiquinol oxidase from \textit{E. coli} and approximately have over 400 sequences of homologues that are available (Matsumoto, Murai et al. 2006). One of these homologues of \textit{cyt bo}_3 is \textit{aa}_3-600, is from \textit{Bacillus subtilis} does not contain ubiquinone, but a similar menaquinone, thus making it a menaquinone oxidase (Yi, 2001).
Narasimhulu et al. 2010). What makes this interesting is *B. subtilis* only uses menaquinol as a substrate, in contrast with that of *E. coli* bo3 which only uses ubiquinol. Structurally menaquinol and ubiquinol are quite different as well (figure 54). We already know that ubiquinone acts an electron transporter from complex I/II to complex III, but it is still unsure how menaquinone works in *B. Subtilis*, though the similarities to the ubiquinone oxidase bo3, it does have a role in cellular respiration.

Cytochrome bo3 ubiquinol oxidase from *E. coli* is a four subunit, heme copper oxidase that catalyzes the reduction of oxygen to water, where the redox centers are located within subunit (Puustinen, Finel et al. 1991). The bound heme b acts as an electron donor to a binuclear center that contains an O-type heme (heme o3) and a copper ion (Cu6) (Abramson, Riistama et al. 2000). Subunits I, II, and III are homologous to corresponding subunits that can be found in the aa3-600 menaquinone oxidase (Lemieux, Calhoun et al. 1992). Unfortunately as of now Subunit IV has either little or no sequence homology with other oxidases and its function is still unknown. Also important to note is that subunit II of the ubiquinol oxidase has no CuA center or a cytochrome c binding site (Lemieux, Calhoun et al. 1992). Heme b instead receives electrons directly from the ubiquinol molecule and the protons from the result of this transfer are moved to the positive side of the membrane. The x-ray crystal structure of bo3 shows the presence of two ubiquinone binding sites: one high affinity site called QH and one low affinity site QL (Abramson, Riistama et al. 2000).
Structure of ubiquinol oxidase $bo_3$

The structures of ubiquinol oxidase from *Escherichia coli* and its ubiquinone binding site were solved by Jeff Abramson and associates and revealed important information on how ubiquinone interacts with its binding site in a novel manner (Abramson, Riistama et al. 2000). The overall structure of $bo_3$ is made up of 25 transmembrane helices (figure 52) (PDB ID 1FFT).

Subunit I is comprised of 15 transmembrane helices. Heme $b$ is coordinated by His106/421, heme $o_3$ is coordinate by His419, and the $Cu_b$ interacting with residues His284/333/334 (figure 54). Subunit II is comprised of two domains, the N-terminal transmembrane domain and the C-terminal extrinsic domain (figure 55). Subunit III is comprised of five helices and subunit IV is comprised of three helices, but function for this domain is unknown (figure 55).

It is proposed there are two ubiquinone binding sites on $bo_3$, one with high affinity ($Q_H$) and one with low affinity ($Q_L$) (Satowatanabe, Mogi et al. 1994). It is proposed that electrons are transferred from the $Q_L$ binding site to the heme $b$ by action of the ubiquinone at the $Q_H$ binding site, where it remains tightly bound (Satowatanabe, Mogi et al. 1994). The ubiquinone is not bound here do the crystallization conditions utilizing the detergent n-octyl-β-D-glucopyranoside, which has been shown in previous work to rip out the ubiquinone out of the active site. The authors propose prospective $Q_H$ ubiquinone binding sites at the interface of the C-terminal extrinsic domain of subunit II and subunit I, or in the membrane domain of subunit I. This is based on the proximity to the bound hemes. However, the $Q_L$ binding site location remains uncertain. Further work has elucidated this question, where it was shown that ubiquinone binds to the $Q_H$ site, where it accepts two elections from the quinol at the $Q_L$ site.
and transfers the electrons to heme $b$ (figure 56) (Matsuura, Yoshioka et al. 2004). It is then transferred to the heme $a_3$/Cu$_B$ site, where oxygen is reduced to two H$_2$O (Matsuura, Yoshioka et al. 2004). How exactly this process is carried out in the ubiquinone oxidase remains a question.

Perhaps the mystery of the Q$_L$ binding site as well as questions pertaining to the Q$_H$ binding site can be better elucidated with information derived from a structure of the homolog $aa_3$-600 menaquinone oxidase. Also there is little known about the structure of how $aa_3$-600 is compared to $bo_3$. In collaboration with the Robert Gennis lab, we hope to distinguish the differences of the protein-quinol interactions of the $bo_3$ type ubiquinol oxidases and the $aa_3$-600 menaquinol oxidase. We have obtained crystals of menaquinone oxidase with ubiquinone bound as evident by the color of the crystals containing a golden brown color, which is evident of the bound quinol.
Figure 49: The electron transport chain is the site of oxidative phosphorylation in eukaryotes, where the NADH and succinate generated in the citric acid cycle are oxidized and provide energy for the ATP synthase. In plants, the ETC is used to extract energy from sunlight in photosynthesis using redox reactions.
Overall Equation: \[ \text{NADH} + \text{H}^+ + \text{CoQ} + 4\text{H}^+\text{in} \rightarrow \text{NAD}^+ + \text{CoQH}_2 + 4\text{H}^+\text{out} \]

Figure 50: A diagram of complex I of the electron transport chain. Also called NADH-Coenzyme Q reductase since it transfers two electrons from NADH to coenzyme Q.
Figure 51: A diagram of complex II of the electron transport chain. Also known as succinate dehydrogenase, it is responsible for converting succinate to fumarate to drive the transfer of two electrons to coenzyme Q to produce CoQH₂.

Overall Equation: Succinate + CoQ $\rightarrow$ Fumarate + CoQH₂
Figure 52: A diagram of complex III of the electron transport chain. Also known as coenzymes Q-cytochrome c reductase because it passes the electrons for CoQH₂ to cyt c through the Q-cycle.

Overall Equation: \[ \text{CoQH}_2 + 2 \text{ cytochrome c}_1 (\text{Fe}^{3+}) + 2 \text{ H}^+ \text{ (matrix)} \rightarrow \text{CoQ} + 2 \text{ cytochrome c}_1 (\text{Fe}^{2+}) + 4 \text{ H}^+ \text{ (intermembrane)} \]
Figure 53: A diagram of complex IV of the electron transport chain. Complex IV is also known as cytochrome c oxidase. It accepts electrons from cytochrome c and reduces O₂ to form two molecules of H₂O.

Overall Reaction: $4\text{ cyt c (Fe}^{2+}\text{)} + 4\text{ H}^+ + \text{O}_2 \rightarrow 4\text{ cyt c (Fe}^{3+}\text{)} + 2\text{H}_2\text{O}$
Figure 54: Structures of Ubiquinone and menaquinone. Ubiquinone acts as electron carrier from complex I and complex II to complex III. Menaquinone plays a role in *B. subtilis* cellular respiration but as how is not clearly known.
Figure 55: Cartoon representation of $bo_3$ (PDB 1FFT). Subunit I is in blue and is where the heme $b$ and heme $o_3$ bind along with Cu$_B$. Subunit II is colored in green. Subunit III is colored in red. Subunit IV is in yellow.
Figure 56: Binding site of subunit I of $bo_3$. Heme b is in pink, Heme $O_3$ is in blue, and Cu$_B$ is in green. Residues interacting with them are in white, with His106/421 interacting with Heme b, His419 interacting with Heme $O_3$ and His 284/333/334 interacting with the Cu$_B$. 
3.2 MATERIALS AND METHODS

Protein Preparation

Purified $aa_3$-600 menaquinone oxidase with bound ubiquinone was purified and prepared by the Gennis lab. Purified protein was given to our lab in 10% glycerol to allow for flash freezing if needed. The protein mixture was then purified by size exclusion chromatography, using the Superdex 200 10/300 GL (from GE Healthcare) and the buffer used was a low salt buffer (100 mM Tris-Cl pH 8.0). This buffer also contained the detergent $n$-Dodecyl-$\beta$-D-maltopyranoside (DDM) at 0.05% (Anatrace, Cat. No. D310A). DDM was used as detergents allow for micelle formation which is beneficial to membrane proteins since it reduces contact of the hydrophobic portions of the protein with the sounding aqueous environment. Fractions were collected at a retention volume of approximately 55-75 mL. Injection of the sample on the column was done at using an AKTA UPC-900 FPLC (G.E. Healthcare Life Sciences, Pittsburgh, PA) at a flow rate of 1 mL/min. Slow elution was carried out over the course of 2 hours using the same buffer. Samples were then concentrated to 12 mg/mL using a 10,000 MWCO Amicon centrifugal filter (Merck Millipore, Carrigtwohill, CO).

*E. coli* lipids were also used as a way to alter the packing of the micelle that forms with the crystals. These were given to us by the Grosman lab upon recommendation to improve resolution. Lipid concentration at 1.5 mg/mL was incubated with $aa_3$-600 menaquinone oxidase before screening for 4 hours. Decyl Maltoside (DM) at 0.05%, Octyl Glucoside (OG) at 0.01%, Octyl Glucose Neopentyl Glycol (OGNG) at 0.075%, and N,N-Dimethyldecylamine N-oxide
(LDAO) at 1\% were used as well (all detergents from Anatrace chemicals). These are added as substitutes to DDM in the same buffer at 100 mM Tris-HCl pH 8.0.

**Crystallization Methods and Data Collection**

Crystallization of \(aa_3\)-600 was in the condition containing 0.1 M Sodium acetate pH 5.0, 0.2M MnCl\(_2\), and 18\% Peg400 with a protein concentration at 12 mg/mL. Incubation was a 4-5 days at the temperature of 4\degree C. The cryo used to flash freeze the crystals contained 25\% trehalose and 30\% Peg400. Crystals were flash frozen at the temperature of incubation due to the susceptibility of the crystals to temperature shifts. Data was collected at the insertion device synchrotron beam line at Sector 21 ID-G LS-CAT (Advance Photon Source, Argonne, IL) using a Mar 300 CCD detector. Exposure was at 2 second per frame for 300 frames at a fixed wavelength of 0.97872 Å. Diffraction data was indexed and scaled using HKL2000 (Otwinowski, Borek et al. 2003) or XDS (G. Bricogne 2003).

### 3.3 RESULTS AND DISCUSSION

The crystals of the \(aa_3\)-600 menaquinone oxidase have a golden brown color to them due to the bound menaquinone (figure 57). Diffraction of the crystals was resultant to 3.5 Å (figure 58); however, the data generate was not able to yield a conclusive space group or suitable solution to build a model. To improve resolution, utilizing of \(E.\ coli\) lipids were used with the idea that
packing of the crystal lattice would be altered. Upon incubation for 5 days at 4°C, crystals of similar shape were produced; however, these diffracted poorly.

Our next solution was to utilize different detergents in the place of DDM to alter protein packing of crystal formation. OG failed to form crystals and there was a color loss of the resultant fractions after desalting. DM failed to produce crystals when using the same crystallization conditions that previously generated crystals. LDAO caused protein precipitation upon concentration, leading to the belief that LDAO cause the protein to become stable during desalting/ size-exclusion chromatography. OGNG failed to reproduce crystals using the same crystallization conditions. Broad crystallization screening using DM and OGNG were also carried out at 10-15 mg/mL, however these screens failed to produce hits as well.

We are continuing to alter detergent concentrations for DDM, DM, OGNG, LDAO, and OG. One consideration is to alter the concentration to either higher or lower than the critical micelle concentration (CMC). CMC is the concentration of detergents that is required for micelles to form (Ana Domínguez 1997). Each detergent has a specific CMC: DM is 1.8 mM, DDM is 0.2 mM, OGNG is 1.02 mM, LDAO is 2mM, OG is 20 mM. Adding the detergent above the recommended CMC usually increases the amount of micelles that are formed. Another method we are planning is the screening of various low buffer conditions other than Tris-HCl during the size exclusion purification step. This will alter the stability of the target protein in respects to pH and the actual buffering compound. Other candidates include HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), citric acid-NaOH, Sodium citrate, sodium acetate, MES (2-(N-morpholino)ethanesulfonic acid)-NaOH, and imidazole-HCl.
Our goal for this continuing project is to utilize structural studies to distinguish the differences of the protein-quinol interactions of the $bo_3$ type ubiquinol oxidases and the $aa_3$-600 menaquinol oxidase. It is still unclear as to how menaquinone is utilized by $B.~Subtilis$ in either a similar or different manner then that of ubiquinol by $E.~coli$. Additional studies will be required to understand the true nature of this interaction as a bound quinone/menaquinone crystal structure will be significant discovery. A high degree of sequence identity between the $Q_H$ sites means that they are structurally similar and evolutionarily conserved. Therefore understanding the structural changes and interactions of one $Q_H$ or $Q_L$ will give a great understanding to the other.
Figure 57: Pictures of the crystals of $aas_5$-600 menaquinone oxidase.
Figure 58: Diffraction image of $aa3$-600 menaquinone oxidase crystals. Resolution achieved was 3.5 Å.
BIBLIOGRAPHY


Park, B. K., A. Hirot and H. Sakai (1976). "Studies on New Antimetabolite Produced by Microorganism .2. 2-Amino-5-Phosphono-3-Pentenoic Acid, a New Amino-Acid from N-1409 Substance, an Antagonist of Threonine." Agricultural and Biological Chemistry 40(9): 1905-1906.


