Biochemical and structural characterization of a novel enzyme involved in uronic acid metabolism

by

Seung Hyae Lee
BSc., University of Victoria, 2012

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

Supervisory Committee

Biochemical and structural characterization of a novel enzyme involved in uronic acid metabolism

by

Seung Hyae Lee
BSc, University of Victoria, 2012

Supervisory Committee

Dr. Alisdair B. Boraston (Department of Biochemistry and Microbiology)
Supervisor

Dr. Douglas Briant (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Jürgen Ehlting (Department of Biology)
Outside Member

Abstract

Supervisory Committee

Dr. Alisdair B. Boraston (Department of Biochemistry and Microbiology)
Supervisor
Dr. Douglas Briant (Department of Biochemistry and Microbiology)
Departmental Member
Dr. Jürgen Ehlting (Department of Biology)
Outside Member

Polyuronic acids are an important constituent of seaweed and plants, and therefore represent a significant part of global biomass, providing an abundant carbon source for both terrestrial and marine heterotrophic bacteria. Through the action of polysaccharide lyases, polyuronic acids are degraded into unsaturated monouronic acid units, which are fed into the Entner-Doudoroff pathway where they are converted into pyruvate and glyceraldehyde-3-phosphate. The first step of this pathway was thought to occur non-enzymatically. A highly conserved sequence, *kdgF* was found in alginate and pectin utilization loci in a diverse range of prokaryotes, in proximity to well established enzymes catalyzing steps downstream in the Entner-Doudoroff pathway and I hypothesized that KdgF was involved in the catalysis of the first step of this pathway. The *kdgF* genes from both *Yersinia enterocolitica* and a locally acquired *Halomonas sp.* were expressed in *Escherichia coli* and their activity was examined using unsaturated galacturonic acid depletion activity assays. To gain perspective on the general structure of KdgF, x-ray crystallography was used to obtain a crystal structure of both *Ha*KdgF and *Ye*KdgF. These crystal structures provided insight into the molecular details of catalysis by the KdgF proteins, including their putative catalytic residues and a

coordinated metal binding site for substrate recognition. To elucidate amino acids that may be involved in binding and/or catalysis, mutants were created in *Ha*KdgF, and lack of activity was observed in four mutants (Asp102A, Phe104A, Arg108A, and Gln55A). The research done in this study suggests that KdgF proteins use a metal binding site coordinated by three histidines and several additional residues to cause a change in monouronic acid, thereby, affecting the unsaturated double bond. This suggests that KdgF is involved in the first step in the Entner-Doudoroff pathway, which is the linearization of unsaturated monouronic acids.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to first thank my supervisor, Alistair Boraston, for taking me on as his student in his lab. His patience and guidance have been a source of motivation to complete this degree. I would also like to thank my committee members, Dr. Jürgen Ehlting and Dr. Douglas Briant for being so approachable and being such great teachers and mentors during my time here at UVic. I would also like to thank everyone in the lab for his or her encouragement in this process and made my time in the Boraston lab a fun and enjoyable place to be every day during my three years here. A special thanks to Dr. Joanne Hobbes for reading and editing my (what was probably dreadful) initial drafts. I would also like to thank Cuong Le for helping me get through my final drafts.

Lastly, an immense thank you to Alexander Fillo and Kaleigh Giles for taking time to help me through this daunting task that would have been insurmountable without their help.

Dedication

I would like to dedicate this to my parents, Miree and John Lee. I would not be as privileged to be where I am today without their dedication, sacrifice and unconditional love and support. I would also dedicate this to my older brother, Kevin Lee. His enthusiasm and encouragement for higher education has been an important part of my motivation to learn and thrive.

엄마, 아빠 & 오빠

감사하고 사랑해요

Chapter 1. Introduction

1.1 Carbohydrate diversity

Carbohydrates are one of the most abundant organic molecules on the planet due to their widespread roles in all domains of life. They have an important place in hereditary information as they make up the structural framework, in the form of ribose and deoxyribose sugars, of RNA and DNA respectively. Carbohydrates also act as a form of energy storage in all forms of life. When linked to proteins and lipids, carbohydrates play a role in intercellular communication and interactions. Additionally, carbohydrates impart mechanical stability to cell walls in bacteria, fungi and plants.

The diverse roles in which carbohydrates function is achieved by having tremendous chemical and structural diversity. Carbohydrates are composed of basic structural units called monosaccharides, which are aldehydes or ketones with the empirical formula $(C\text{-}H_2O)_n$, where n ranges from 3-9. One or more carbon centers and their respective hydroxyl groups can vary in the stereochemical configuration, which can generate molecules with distinct chemical structures. For example, glucose and mannose are distinct monosaccharides with different biological roles, yet they differ only in the stereochemistry of their hydroxyl groups at C2. The diversity of monosaccharides, when linked by glycosidic bonds to form polysaccharides, contributes to the versatility and functionality of the overall carbohydrate structure.

1.1.1 Polyuronic acids: Unique negatively charged polysaccharides

Polyuronic acids are polysaccharides consisting of hexuronic acid units, and are linked by a variety of glycosidic linkages (Figure 1). Compared to glucose or other hexamer sugars, hexuronic acids, such as mannuronic acid, guluronic acid and glucuronic acid, are unique as they have a carboxylic group on the C6 position. This negative charge can alter the solubility of the respective polymer through the coordination of either monovalent or divalent cations (Davis *et al*., 2003). Different modifications to uronic acids exist in a diverse range of organisms and have a variety of different roles. Galacturonic acids are the main component of pectin in plant cell walls, which is important for plant growth and structure (Caffall and Mohnen, 2009). Guluronic acid and mannuronic acid make up a significant proportion of the brown algae cell wall and extra cellular matrix of some biofilm forming bacteria (Fourest and Volesky, 1997; Boyd and Chakrabarty, 1995). Glucuronic acid and iduronic acid are present in glycosaminoglycans, which are highly polar molecules that have numerous biological roles in cell biology (Anower-E-Khuda and Kimata, 2015)(Figure 2).

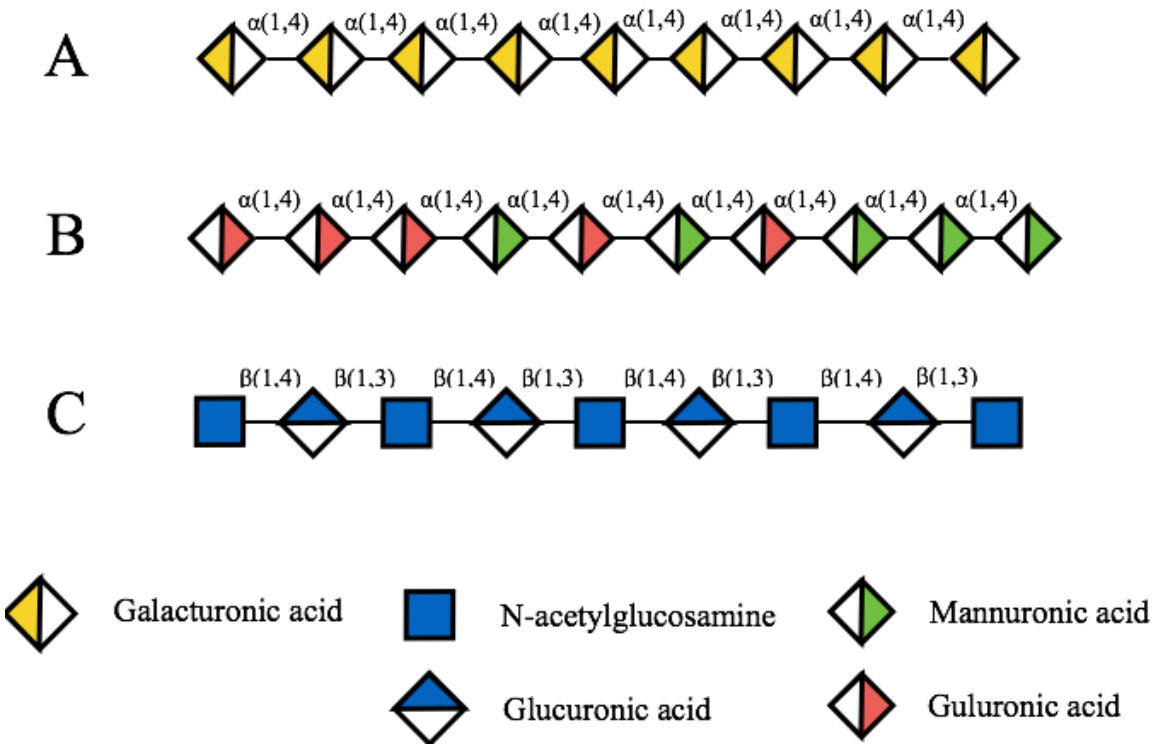


Figure 1. Different polysaccharide containing hexuronic acids.

A) Homogalacturonic acid in pectin joined by α(1,4) linkages. B) alginate composed of mannuronic acid and guluronic acid joined by α(1,4) linkages. C) Hyaluronan composed of glucuronic acid and N-acetylglucosamine and sugars are joined by alternating β(1,3) and β(1,4) linkages.

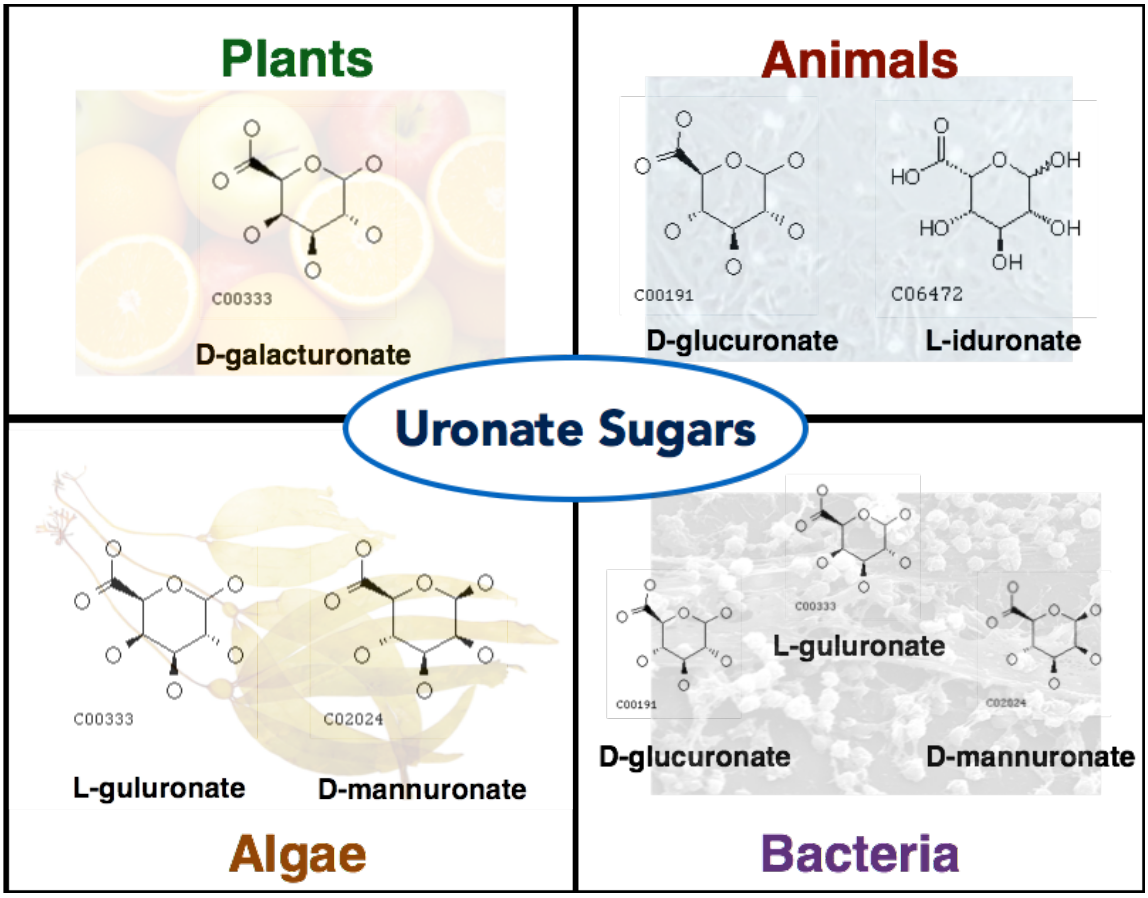


Figure 2. Different hexose uronate sugars from various sources.

D-galacturonate, which makes up the backbone of pectin polysaccharides in plants (top left); D-glucuronate and L-iduronate, which make up proteoglycans and hyaluronan in mammalian cells (top right); D-mannuronate and L-guluronate, which make up the alginate polysaccharide in brown algae (bottom left); D-glucuronate is secreted as a exoheteropolysaccharide (gellan) by *Sphingomonas* and D-mannuronate and L-guluronate is secreted as alginate co-polymer by *Pseudomonas aeruginosa* and *Azotobacter vinelandii* (Bottom right). Monosaccharide images were taken from KEGG database.

1.1.2 Alginate

Alginate is a linear polysaccharide that is made up of varying proportions of L-guluronic acid (G) and D-mannuronic acid (M) joined by β-(1,4) linkages. The residues can be arranged in either homomeric (GG/MM) or heteromeric (MG) sequences and the proportion of these constituents vary depending on species. The relative proportion of GG blocks contributes to the gelling capacity of the alginate polymer, due to the sugar's ability to sequester divalent cations (Ca^{2+}, Zn^{2+} and Cu^{2+}) (Ertesvåg and Valla, 1998; Draget *et al*., 1996). For this reason, alginates have various industrial uses as stabilizers and gel-forming agents (Ertesvåg and Valla, 1998).

While commercial alginate can be collected through the growth of select Gram-negative bacteria, which produce alginate in biofilm form, alginate is traditionally obtained from processing of Phaeophyceae (brown algae) (Hay *et al.*, 2010; Ertesvåg and Valla, 1998). Brown algae are complex multicellular photosynthetic organisms that share some features with terrestrial plants, such as the presence of cellulose in the cell wall. What sets these algae apart is the utilization of alginate, which composes a significant proportion of the cell wall, and contributes to more than 40% of brown algae's dry weight (Andriamanantoanina and Rinaudo, 2010; Percival and McDowell, 1967). The alginate is thought to contribute to the flexibility and strength of the algae (Smidsrød and Draget, 1996).

1.1.3 Pectin

Pectins are complex heteropolysaccharides that are found in the primary cell walls and the middle lamella of terrestrial plants and green algae. They provide many functions within the plant, which can include contributing to structural integrity, cell adhesion, or mediation of defense responses (Caffall and Mohnen, 2009). There are four structural classes of pectic

polysaccharides: homogalacturonan, xylogalacturonan, rhamnogalacturonan I, and rhamnogalacturonan II **(**Caffall and Mohnen, 2009**)**. Homogalacturonan (HG) is the most abundant, accounting for over 60% of pectin in the cell wall (Ridley *et al*., 2001).

1.2 Carbohydrate metabolism

Polysaccharides constitute a major source of energy. The depolymerisation of this energy source requires two distinct classes of enzymes. The first and most common class is the Glycoside Hydrolases (GHs), which are responsible for the depolymerisation of carbohydrates by hydrolysis of glycosidic bonds (Figure 3A). The second class that depolymerises carbohydrates is the Polysaccharide Lyases (PLs). PLs cleave glycosidic linkages via a β-elimination mechanism and targets polysaccharides that contain uronic acids (Yip *et al*., 2006; Figure 3B). The release of sugar monomers through the application of these carbohydrate active enzymes enables organisms to then import and metabolize them.

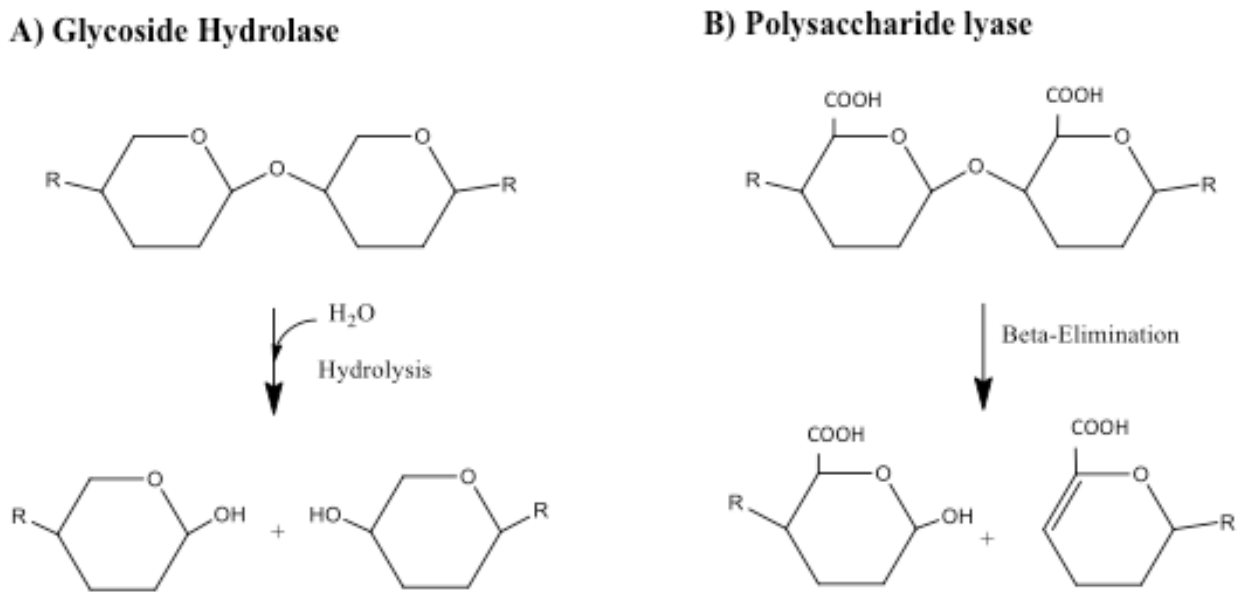


Figure 3. Glycoside hydrolases versus Polysaccharide lyase
A) Glycoside hydrolases use hydrolysis mechanism (addition of water) to break glycosidic bonds in polysaccharides. B) Polysaccharide lyases use β-elimination to cleave glycosidic bonds in polyuronic acids and introduce an unsaturated bond between C4 and C5 of the sugar on the reducing end.

Carbohydrate metabolism is composed of highly integrated networks of chemical reactions that occur in a cell. Monosaccharides undergo several sequential transformations to provide chemical energy in the form of ATP and reducing equivalents, such as NADH. These sequential transformations are made possible by enzyme catalysts. Enzymes ensure that chemical reactions reach equilibrium faster by decreasing the activation energy required. The main pathway that produces energy from these converted monosaccharides is the citric acid cycle. Three main pathways exist that convert simple monosaccharides into substrates for the citric acid cycle: the Embden-Myerhoff (glycolytic) pathway, the pentose phosphate pathway, and the Entner-Doudoroff (ED) pathway (Romano and Conway, 1996). Glycolysis is the pathway by which glucose is sequentially converted into pyruvate and is also the most studied pathway that

converts monosaccharides into substrates for the citric acid cycle (Figure 4A). An alternative pathway that runs parallel to glycolysis is the pentose phosphate pathway where glucose is converted to pyruvate by generating 5-carbon sugars (Kruger and von Schaewen, 2003) (Figure 4A). Another alternative pathway is the ED pathway, which involves the catabolism of hexuronic sugars that possess carboxylic functional group at C6 to generate pyruvate and glyceraldehyde-3-phosphate (G3P) (Romano and Conway, 1996) (Figure 4B). Despite their differences, the pattern and scheme of the glycolytic and ED pathways are similar. They both involve six-carbon sugars that are converted into phosphorylated intermediates, which are cleaved by aldolases into two three-carbon intermediates that will eventually be converted to pyruvate (Conway, 1992).

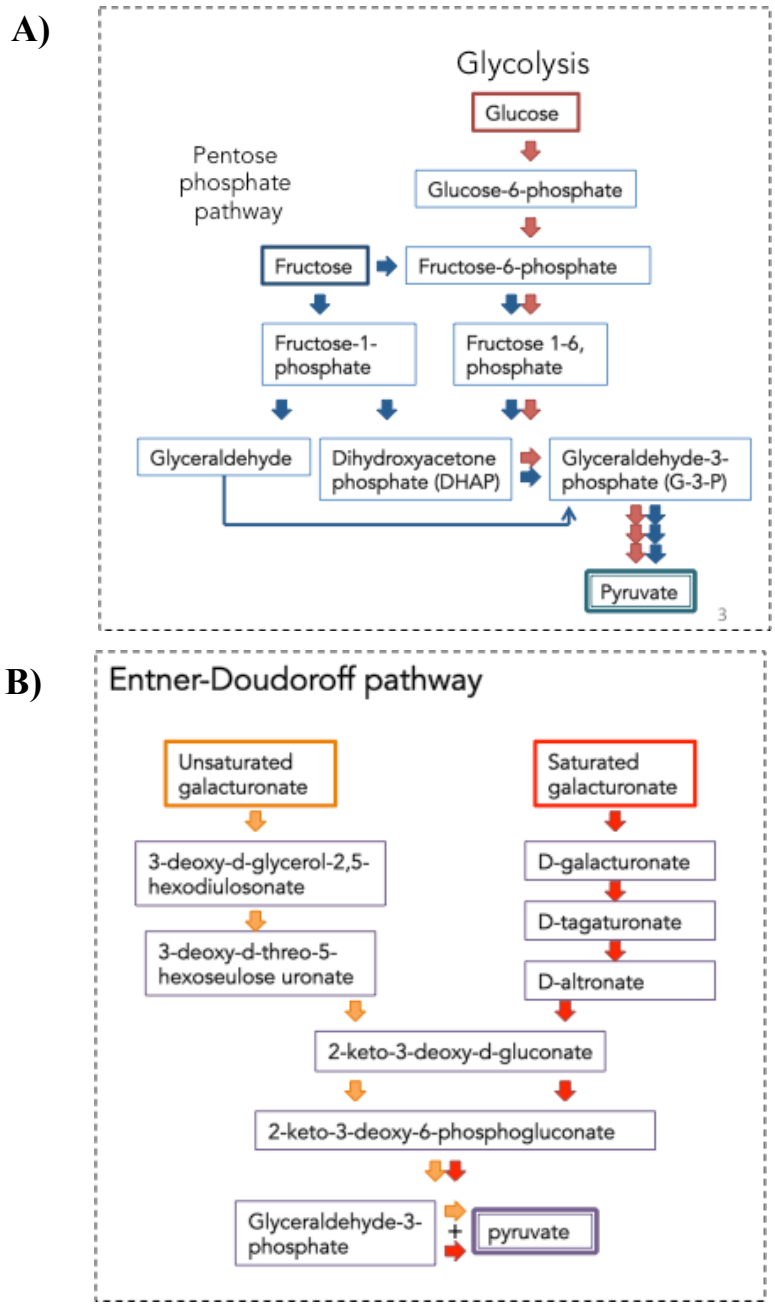


Figure 4. Carbohydrate metabolic pathways

A) Glycolysis and pentose phosphate pathway versus. B) Entner-Doudoroff (ED) pathway. These two pathways are very distinct from each other in substrate and reactions but in the end, they produce pyruvate that feeds into the citric acid cycle.

1.2.1 Entner Doudoroff pathway

The ED pathway was first characterized in *Pseudomonas saccharophila* by Entner and Doudoroff, and since then has been identified in a variety of other organisms such as *Escherichia coli*, *Thermoproteus tenax* (hyperthermophilic archaea), *Phaeodactylum tricornutum* (unicellular eukaryotic algae) (Entner and Doudoroff, 1952; Eisenberg and Dobrogosz, 1967; Ahmed *et al*., 2005; Fabris *et al*., 2012). The presence of this pathway in the three major domains is thought to stem from its utilization in primitive life as a precursor to glycolysis (Romano and Conway, 1996). The ED pathway accomplishes the same goal as glycolysis, producing pyruvate and G3P; however, it does so using different mechanics. The primary distinction lies in the nature of the central substrate, as the ED pathway consumes hexuronic acids while glycolysis primarily uses glucose.

Hexuronic acid sugars are converted through a series of enzymatic reactions, eventually into 2-keto-3-deoxygluconate (KDG). KDG is then phosphorylated by a kinase to produce 2-keto-3-deoxy-6-phosphogluconate (KDPG), which is the substrate for cleavage by aldolases, producing pyruvate and G3P (Figure 5).

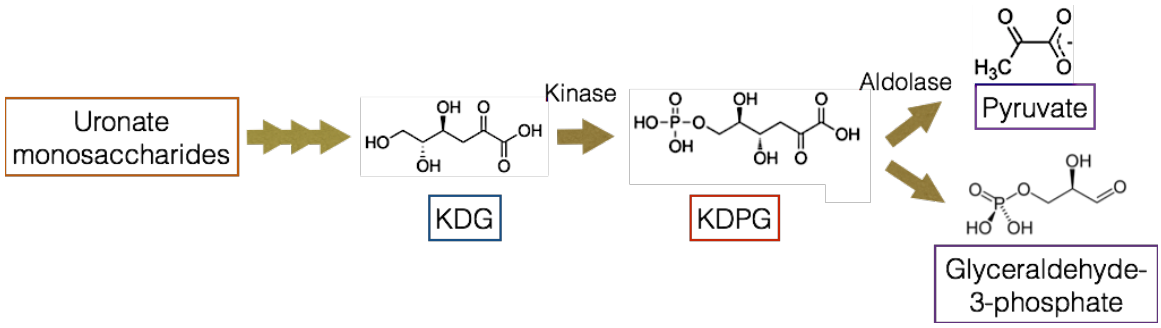


Figure 5. Central steps in the Entner-Doudoroff pathway.
KDG, 2-keto-3-deoxygluconate, KDPG: 2-keto-3-deoxy-6-phosphogluconate.

Although the ED pathway is similar to other pathways that metabolize monosaccharides, its ability to metabolize hexuronic acids has been shown to have significant effects in adaptation and pathogenesis. For instance, the bacterium *Chromohalobacter salexigens* preferentially uses the ED pathway over the glycolysis for glucose catabolism to better cope with a high salt environment. Using the ED pathway allows better control of the production of redox equivalents and fine-tuning of its redox states to maximize growth and the production of ectoine (osmolyte that allows protection from extreme osmotic stress) (Pastor *et al*., 2013). In the pathogenic bacterium that causes cholera, *Vibrio cholera*, the ED pathway plays a significant role in its pathogenic cycle by up-regulating virulence genes that are required for colonization and toxin production (Patra *et al*., 2012).

1.3 Model alginate degrading organism, *Zobellia galactanivorans*

To date, the study of this pathway has mainly focused on a small number of terrestrial and pathogenic species that degrade bacterial alginates in the environment. However, bacterial alginate is not very different from marine alginate produced by brown algae, the only difference being bacterial alginate can be secreted in an acetylated form depending on the bacterial species (May and Chakrabarty, 1994). Therefore, we would expect the central assimilation of KDG into the ED pathway to be the same in terrestrial and marine microorganisms.

The wide distribution of brown algae in the marine ecosystem makes alginate a plentiful potential energy source for heterotrophic organisms. The flavobacterium *Z. galactanivorans* is a model marine organism for studying degradation of algal biomass (Thomas *et al*., 2012) because it has been shown to utilize a wide variety of algal polysaccharides as sole carbon sources and to

possess a complete pathway for alginate degradation. This pathway is overseen by two alginolytic operons, collectively called the Alginate Utilization System (AUS) (Thomas *et al*., 2011).

Within the AUS, there are many genes that encode for lyases that generate unsaturated (Δ) monosaccharides from alginate poly- and oligosaccharides (Figure 6B). The ΔMannuronic acid and ΔGuluronic acid undergo spontaneous tautomerization into 4-deoxy-erythro-5-hexoseulose (DEH). DEH undergoes a single step reduction by 2-dehydro-3-deoxy-D-gluconate 6-dehydrogenase (Sdr) to form KDG (Figure 6A) (Thomas *et al*., 2012). KDG then progresses further through the ED pathway and is finally converted to pyruvate and G3P (Takase *et al*., 2012).

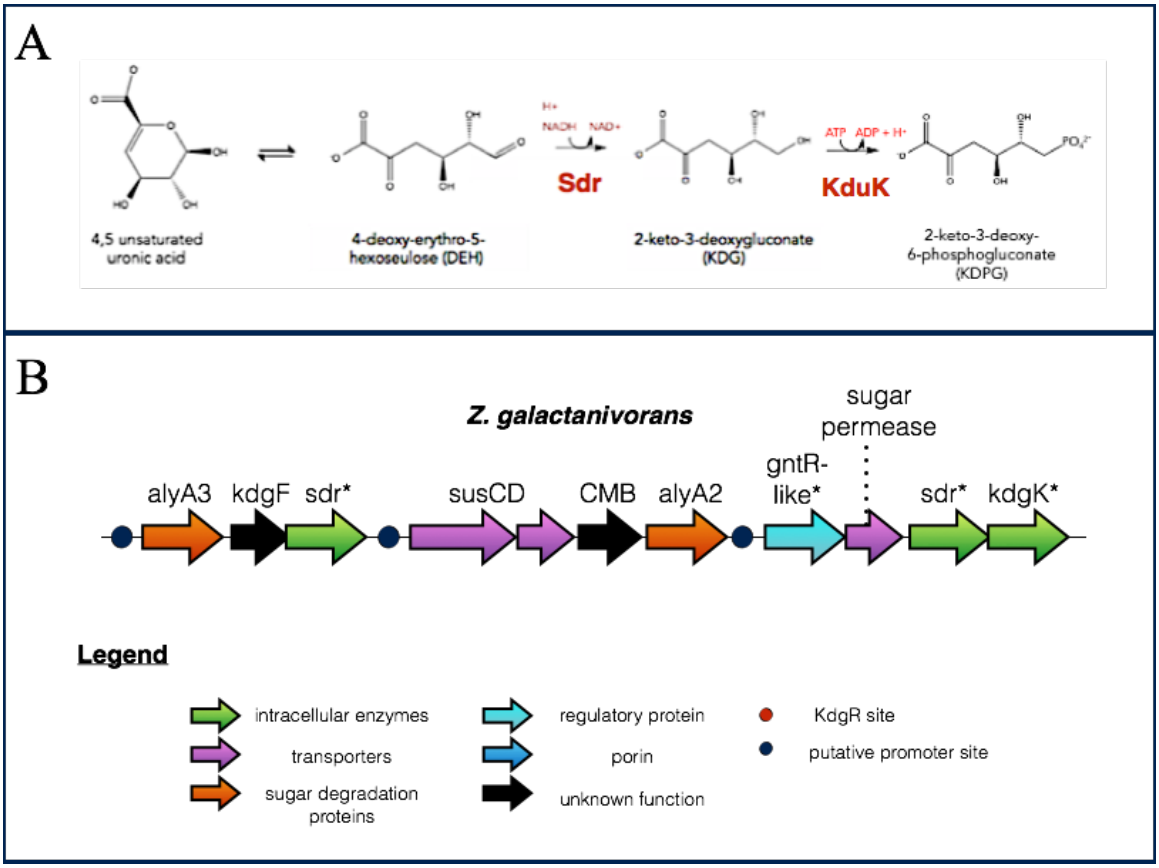


Figure 6. Central steps in the ED pathway for metabolizing Δmonosaccharides derived from alginate and the genetic organization of alginate utilization loci in *Z. galactanivorans.*

A) Alginate metabolism in *Z. galactanivorans.* Δ uronic acid is converted to DEH by spontaneous linearization followed by reduction (by Sdr) and is phosphorylated by KdgK to produce KDPG. B) Genetic organization alginolytic loci found in *Z. galactanivorans* genome and the function of genes and promoter sites are color coded and indicated in the legend (Thomas *et al*., 2012).

1.3.1 Alginate from brown algae as feedstock for biotechnological applications

Brown macroalgae is a desirable source of feedstock for biorefining and there are many advantages of using macroalgae as a feedstock versus land crops. Firstly, macroalgae are much more productive because they do not require arable land, irrigation or fertilizer (Georgianna and Mayfield, 2012; Adams *et al.,* 2009, Yoon *et al.,* 2010). Secondly, macroalgae are easier to process than plants because they contain no lignin (John *et al*., 2011). It has recently been shown that *Sphingomonas* sp. Strain A1 and *Escherichia coli* can be used to produce ethanol from brown algae alginate on a large scale (Takeda *et al.,* 2011; Wargacki *et al*, 2012). The combination of *Sphingomonas* sp. strain A1 and brown algae alginate has also been used for the production of pyruvate, a widely used starting material in the pharmaceutical industry (Kawai *et al*., 2013). All of these systems were bioengineered to take advantage of the conversion of Δuronic acid, the final product in the depolymerisation stage of polyuronic acids degradation before entering the ED pathway. Therefore, it is important to characterize all the components that are involved in this pathway, as it will further contribute towards optimization this the utilization of these feedstock for biorefining.

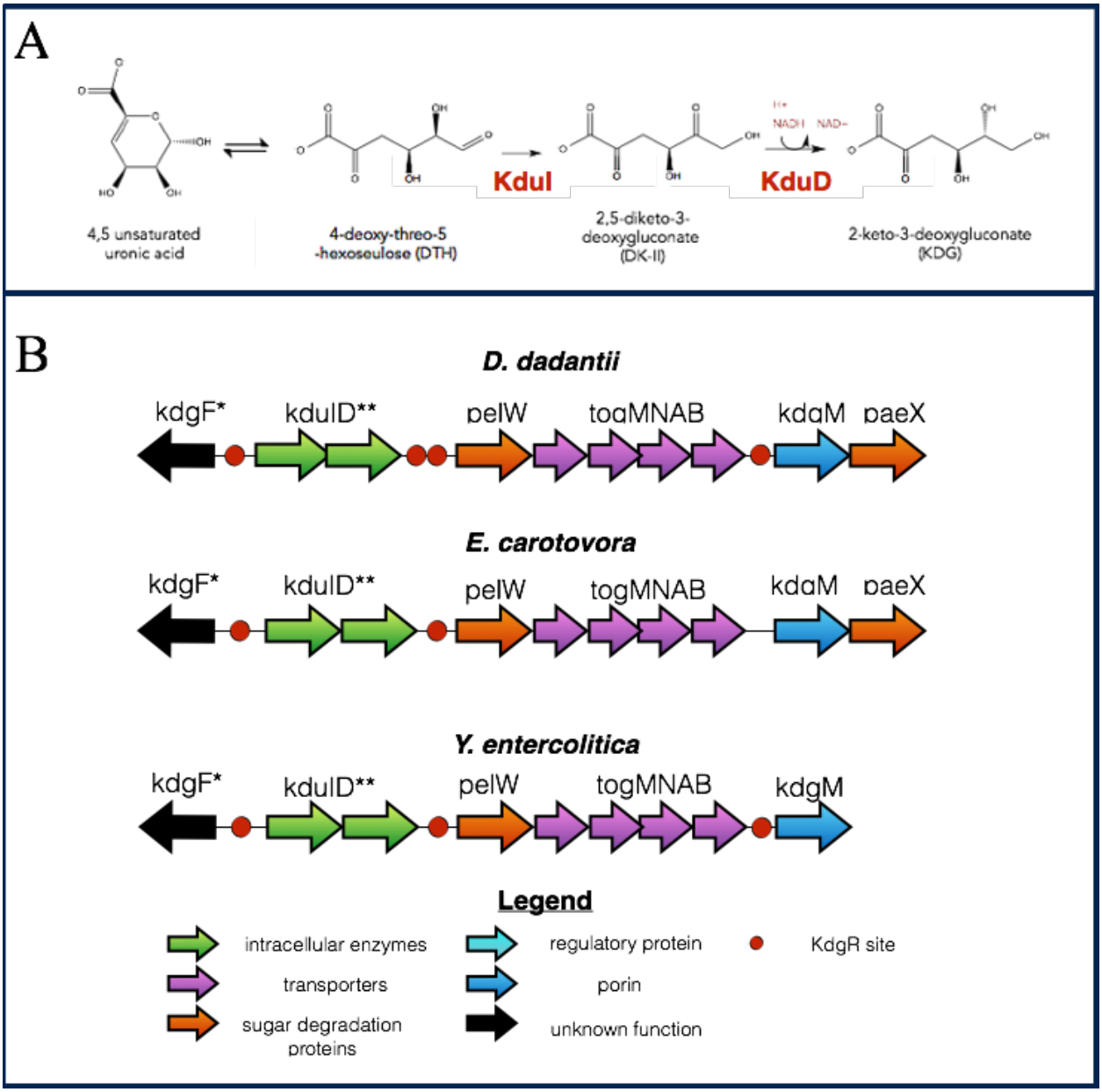
1.4 Model pectin degrading organism, *Dickeya dadantii*

Due to pectin's abundance and diversity in biological roles, it is a source of energy-rich carbohydrates, in particular galacturonic acids, and is targeted by pectinolytic microorganisms. Pectin polysaccharides are pliable substrates compared to rigid crystalline structural polysaccharides due to their hydration level, flexibility, and polarity (Abbott *et al*., 2010). Because of their branching and the diversity of pectin monomers, pectin requires a plethora of

PLs, GHs and other CAZymes in order to be fully degraded. In this way, it requires a much more complex pathway than crystalline cellulose.

The cytoplasmic pathway of pectin utilizations has been extensively studied in *D. dadantii* and through comparative genomic studies (Rodionov *et al*., 2004; Reverchon *et al*., 2006; Condemine, *et al*., 1986). The symptoms of *D. dadantii* infection (soft rot) arise from the depolymerisation and disorganization of the plant cell wall due to the variety of pectinases and oligogalacuronate transporters secreted by the bacterium (Hugouvieux-Cotte-Pattat *et al*., 1996). *D. dadantii* is equipped with many genes encoding for pectin-active enzymes that are dedicated to pectin metabolism. *D. dadantii* represents a model system because the gene organization dedicated for the depolymerisation and metabolism is well conserved in the enterobacteriaceae family (Figure 8B)(Rodionov *et al*., 2004).

Pectin is first deconstructed by a series of lyases into tri- and digalacturonic acids, both extracellularly and periplasmically, and then transported into the cytoplasm. There, exo-polysaccharide lyases and oligogalacturonate lyases (OGLs) produce 4,5 unsaturated monogalacturonic acid (ΔGalA) and saturated monogalacturonate (to a lesser extent)(Shevchik *et al*., 1999). The final depolymerized product (ΔGalA) catabolism begins with the spontaneous formation of 4-deoxy-5-threo-hexoseulose (DTH)(Preiss and Ashwell, 1963). In a two-step process, DTH is first isomerized by KduI, then subsequently reduced by KduD, in order to form the central metabolite of the ED pathway, KDG (Figure 7A). This contrasts the single step conversion observed in alginate metabolism.



;

Figure 7. Central steps in the ED pathway for metabolizing Δmonosaccharides derived from pectin and the genetic organization of pectin utilization loci in enterobacteriaceae family.

A) Pectin metabolism in *Dickeya dadantii.* Δuronic acid requires three reactions and two enzymes (KduI and KduD) for the production of the central ED metabolite, KDG. Genetic organization pectinolytic clusters found in *Dickneya dadantii*, *Erwinia carotovora* and *Yersinia enterocolitica* genomes. The functions and regulatory sites are color coded and indicated in the legend (Rodionov *et al.*, 2004).

1.4.1 Plant pectin as a feedstock for biotechnological applications

Pectin polysaccharides represent a source of underutilized and relatively abundant sugars. Because pectin is most abundant in primary cell walls of soft and growing tissues, the wastes from processed fruits and vegetables are particularly pectin rich. Such examples are citrus peels, apple pomace and beet-pulp (with pectin content between 20-40%)(Kennedy *et al*., 1999; Edwards and Doran-Peterson, 2012). Currently, these by-products are disposed of in landfills or dried and processed in an inefficient way for cattle feed (Benz *et al*., 2014). These by-products have little or no lignin content due to the pre-treatment procedures during sugar and juice extractions, making it attractive potential as feedstock for biofuel production (Edwards and Doran-Peterson, 2012).

1.5 Objectives and Hypotheses

The ED pathway, which almost exclusively catabolizes monouronic acids such as those derived from pectin and alginate, has been well studied and it is believed that the enzymes responsible for all catalytic steps have been identified. The majority of these enzymes are localized within pectin and alginate utilization loci. The only gene that is widely conserved within these loci but has yet to be assigned a role is *kdgF* (Rodionov *et al*., 2004).

kdgF was first identified as part of the KdgR regulon of *D. dadantii* based on the presence of a KdgR binding box in its regulatory region (Condemine and Robert-Baudouy, 1991). When the *kdgF* was knocked out, there was no change in the *D. dadantii* growth phenotype on polygalacturonic acid. However, the phenotype did negatively impact the expression of pectate lyases, which was thought to be due to reduced production of regulon inducer, KDG. This

suggests that KdgF is somehow involved in KDG metabolism and thus pectin degradation but the specific function is unknown (Codemine and Robert-Baudouy, 1991).

In pectin degradation, the first step of the ED pathway for ΔGalA, is their tautomerization into DTH. This reaction is widely thought to occur spontaneously (Preiss and Ashwell, 1962; Takase *et al*., 2010; Hugouvieux-Cotte-Pattat *et al*., 1996). As the only presumed non-enzymatic step, this conversion is rate limiting for the entire ED pathway. If an enzyme were to facilitate this tautomerization, it would improve the overall efficiency of the pathway by preventing the accumulation of starting substrate for the pathway. Furthermore, the parallels between this reaction and the tautomerization of ΔM and ΔG to DEH in alginate degradation may be enzymatically catalyzed in a similar fashion.

Since DTH and DEH are both precursors of KDG, the production of KDG would dependent on their tautomerization. Therefore, *I hypothesize that Δmonouronic acid tautomerization does not only occur spontaneously, but is expedited by the catalytic activity of KdgF*. I planned to test this hypothesis by studying two bacterial KdgF proteins – *Ye*KdgF from the pectinolytic pathogenic bacterium *Yersinia enterocolitica* and *Ha*KdgF from the alginolytic marine bacterium *Halomonas.* This study of these proteins will provide new details about an otherwise well-researched pathway. The specific objectives are:

1. Investigate the ability of *Ye*KdgF and *Ha*KdgF to catalyze the conversion of cyclic Δmonouronic acid.
2. Determine the three-dimensional structure of *Ye*KdgF and *Ha*KdgF using X-ray crystallography

3. Investigate the effects of residues in the active site pocket of KdgF through mutagenesis of *Ha*KdgF.

The ED pathway is arguably one of the most ancient metabolic pathways and to date, and no enzyme has been assigned to the first step of this pathway due to it occurring spontaneously. From an evolutionary perspective, it would be unusual that the ED pathway's first step is the major rate-limiting step due to the lack of a designated enzyme. This study provides evidence that the first committed step is enzymatically facilitated and provides better understanding of interactions between proteins and uronic acids. In addition, many organisms utilized in the biotechnological industry take advantage of this pathway (Takeda *et al*., 2011; Wargacki *et al.*, 2012; Kawai *et al*., 2013). Therefore, in addition to expanding the bounds of scientific knowledge, characterizing this molecular step in the ED pathway may facilitate the exploitation of organism containing this pathway by the biotechnological industry.

Chapter 2: Materials and Methods

2.1 Materials

Y. enterocolitica subsp. *Enterocolitica* 8081 genomic DNA (ATCC 23715™) and *Halomonas spp.* genomic DNA (provided by Dr. Jan Hehemann) were used to clone respective constructs. Digalacturonic acid was purchased from Sigma Aldrich.

2.2 Bioinformatics

All amino acid sequence similarity searchers were performed using the NCBI BLASTp and PSI-BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.3 Cloning of *HakdgF* and *YekdgF*

PCR amplification

The *kdgF* genes were amplified from genomic DNA in a 50 µL PCR reaction consisting of 500 nM of forward and reverse primer (Table 1), 1x Phusion HF buffer, 200 µM dNTPs, 1.0 U Phusion DNA Polymerase, 100 ng template DNA, and nuclease free water. Template DNA was initially denatured at 94 °C for 3 minutes (min). The denaturation, annealing, and elongation steps for each gene were 30 cycles of 94 °C for 45 seconds (sec), 58 °C for 30 sec, and 72 °C for 12 min, respectively. Five microliters of the PCR product was visualized on a 1 % agarose gel supplemented with EtBr using the EagleEye II system. PCR products were purified using the PCR purification kits from BioBasic.

Table 1. Primers used for amplification of *kdgF* gene in respective bacterium

	Forward primer (5'-3') (NheI site underlined)	Reverse primer (5'-3) (XhoI site underlined)
Halomonas spp.	CATATCGCTAGCAACACAGGCAGTTTCTTC	GGTGGTCTCGAGTTATGATTTCAGCATGTCATCGCG
Yersinia enterocolitica	CATATCGCTAGCAAGATGTTCTTTATTAATGATGAAACG	GGTGGTCTCGAGTTACAAGAAATCATCCCGTTTG

Restriction digestion

The two purified gene inserts and previously purified pET28a(+) vector were digested with NheI and XhoI (1 U enzyme per 1 µg of DNA). The digestion mixture was incubated at 37 ˚C for 1 hour and inactivated by heat at 65 ˚C for 20 min. The digested PCR products were PCR purified to remove unwanted DNA fragments, enzymes, and reagents using the QIAquick® PCR Purification Kit. Further purification of the digested vector was done by running it on a 1% agarose gel for 30 min at 100 V, excising the band, and purifying the DNA from the agarose gel using the QIAquick® Gel Extraction Kit.

Ligation

Ligation of the digested gene and pET28a plasmid was performed according to the T4 DNA Ligase kit protocol by Invitrogen. The following formula was used to determine the required amounts of insert and vector DNA required for optimal ligation:

$$ng\ insert = \frac{ng\ vector \times basepairs\ insert \times 3}{basepairs\ vector}$$

The appropriate amount of DNA for each gene, 1 U T4 DNA ligase, and 1X T4 DNA ligase buffer were used in the 20 μL reaction and incubated at room temperature for 1 hour. Both genes were ligated into the pET28a multiple cloning site in-frame with the N-terminal hexahistidine tag.

Transformation

Chemically competent *Escherichia coli* BL21 Star (DE3) cells were thawed on ice for 10 min and incubated with 1 μL of the ligation reaction for 15 min. Cells were heat shocked at 42 °C for 30 sec, incubated on ice for 2 min, suspended in 250 μL LB media, incubated at 37 °C with shaking (180 rpm) for 1 hour, and plated on LB agar supplemented with 50 mg/L kanamycin. Colony PCR was conducted on each *kdgF* construct using Taq polymerase with T7 forward and reverse primers. Colony PCR products were visualized on a 1 % agarose gel with EtBr and the EagleEye II system. Positive colonies were grown in 10 mL of LB media overnight at 37°C with shaking at 180 rpm. Plasmid DNA was extracted from cultures using QIAprep® Spin Miniprep Kit and constructs were confirmed by sequencing (Sequetech, Mountain View, CA).

2.4 Protein expression and purification

2.4.1 *Ha*KdgF

BL21 (DE3) cells containing the *HakdgF* pET28a construct were grown to an OD_{600} of ~0.6 in 2 L of LB media supplemented with 50 mg/L of kanamycin at 37 ˚C with shaking (180 rpm). Cultures were induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside and incubated at 16 ˚C overnight with shaking. Cells were harvested by centrifugation at 8000 g for 10 min and lysed

via chemical lysis. The cells were subjected to chemical lysis using the following procedure: The cell pellet was resuspended in sucrose solution (25 % sucrose and 20 mM Tris-HCl, pH 7.5) and lysozyme was added and the solution was stirred for 15 min, deoxycholate solution (1 % sodium deoxycholate, 1 % Triton X-100; 20 mM Tris-HCl, pH 7.5, and 100 mM NaCl) was added and the solution stirred for 10 min, two-hundred micrograms of DNase and 5 mM $MgCl_2$ was added and the solution stirred for 10 min. Cell free lysate was obtained via centrifugation at 4 °C for 30 min at 17000 rpm in a Beckman centrifuge.

The cell-free lysate was loaded onto a nickel-charged immobilized metal ion affinity chromatography (IMAC) column, washed with two column volumes of binding buffer (20 mM Tris-HCl, pH 8.0, and 500 mM NaCl) and eluted with a stepwise 5 – 500 mM imidazole gradient. Each fraction as analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) for confirmation of protein purity and approximate size (11kDa). Fractions containing pure *Ha*KdgF were pooled and dialyzed in 20 mM Tris-HCl, pH 8.0. No further purification was required for activity assays.

*Ha*KdgF was dialyzed in 150 mM NaCl, 2 mM $CaCl_2$, and 20 mM Tris-HCl, pH 8.0. To remove the N-terminal hexahistidine purification tag, thrombin (1.0 U per 2 mg of protein) was added to this solution and digested at room temperature for 32 hours. The protein was subjected to two purification methods set up in tandem: the protein was purified by a nickel IMAC column connected to an anion exchange chromatography system. The protein sample eluted through the nickel IMAC column first so that any uncut protein with the histidine tag and free histidine tags were captured. Any cut protein of interest was caught in the anion exchange and eluted with salt buffer gradient (500 mM NaCl, 20 mM Tris-HCl, pH 8.0). Fractions from the major peaks of the anion exchange chromatogram were analyzed by SDS-PAGE to ensure purity. Fractions

containing sufficiently pure protein were pooled and concentrated down to 20 mg/mL for crystallization trials using the stirred ultrafiltration unit (Amicon, Beverly, MA) on a 10 kilodalton (KDa) molecular weight cut-off membrane.

2.4.2 *Ye*KdgF

Expression and Nickel IMAC purification of *Ye*KdgF was carried out as described for *Ha*KdgF except the additional thrombin cleavage step was not required for crystallization. The SDS-PAGE confirmed fractions from IMAC purification were pooled and concentrated down to 2 mL. The protein was subjected to a secondary purification using size exclusion chromatography (SEC). The 2 mL sample was injected into a HiPrep 16/60 Sephacryl S-100 HR column (GE Healthcare) and was eluted with 20mM Tris-HCl, pH 8.0. Fractions constituting the major peaks on the chromatogram produced by the SEC purification were run on an SDS-PAGE gel to confirm *Ye*KdgF presence and purity. Fractions with pure *Ye*KdgF were pooled and concentrated to 25 mg/mL using the stirred ultrafiltration unit (Amicon, Beverly, MA) on a 10 KDa molecular weight cut-off membrane

2.4.3 *Ye*Ogl

YeOGL was produced and purified as described previously for *Ye*KdgF.

2.5 Determining activity of *Ye*KdgF and *Ha*KdgF

*Ye*Ogl was used to produce ΔGalA from digalacturonic acid ($GalA_2$). Cleavage of digalacturonate by lyases generates monosaccharides with 4,5-unsaturated bonds at the non-reducing ends (Abbott et al., 2010). The reaction was done in UVStar 96 well plates (that do not absorb UV at 232 nm) with 1 μM *Ye*Ogl, 1 mM digalacturonic acid, and 100 mM Tris, pH 7.5. The production of ΔGalA was monitored by the absorbance at 232 nm. The total volume of the reaction was 100 μL and was allowed to proceed at 30 °C until it plateaued (7 minutes). Subsequently, 25nM of *Ha*KdgF or *Ye*KdgF was added to the reaction and the absorbance at 232 nm was measured for additional 2 minutes. The negative control for this experiment was no KdgF protein added.

2.6 Crystallization, Data collection and Structure solution.

2.6.1 *Ha*KdgF

Concentrated *Ha*KdgF (20 mg/mL) was used for crystallization screening via hanging drop vapor diffusion methods. Three drop ratios were tested in the initial screening with 1:1, 1:2 and 1:3 ratios (volume ratio of respective protein:mother liquor) in the hanging drop. The screenings were left incubating at 18 °C for approximately 2 days.

The crystals were optimized using 1.2 M sodium citrate and 20 mM Tris-HCl, pH 7.5. The best crystals from optimizations were cryoprotected with 20% ethylene glycol and diffraction data was collected at University of Victoria. Diffraction data set was integrated with

iMosfilm (Leslie, 2006), scaled and merged to 1.9 Å in Scala (Evans, 2011). Five percent of reflections were set-aside for the calculation of R_{free}. Four molecules of *Ha*KdgF in the AU were searched for using PHASER (McCoy et al., 2007) and a CHAINSAW generated model (Stein, 2008) of Cupin 2 conserved barrel domain protein from *Shewanella frigidimarina* (PDB: 2PFW chain a; 37.5% identity). PHASER produced one solution and was used for automated building using ARP/WARP Crystallographic Macromolecular Model Building program in CCP4 (Langer et al., 2008; Winn et al., 2011; and Murshudov et al., 2011). The unmodeled portions of the structure were manually built using COOT as well as the metal and two waters were modeled to fit the density within the active site (Emsley and Cowtan, 2004). All of the remaining water molecules were automatically modeled by COOT. The *Ha*KdgF structure was refined multiple times using REFMAC (Murshudov et al., 2011) to an R_{work} of 21% and R_{free} of 25%. Stereochemical analysis of the *Ha*KdgF refined structure was completed with PROCHECK and SFCHECK in CCP4 (Vaguine *et al.*, 1999; Laskowski, MacArthur, Moss, & Thornton, 1993). The Ramachandran plot showed adequate stereochemistry with 91.3 % of the residues in the favoured conformations and no residues modeled in disallowed regions (Data collection statistics are presented in Table 5).

2.6.2 *Ye*KdgF

Concentrated *Ye*KdgF were screened for crystallization conditions via the hanging drop vapor diffusion method. Three drop ratios were tested in the initial screening with 1:1, 1:2 and 1:3 ratios (volume of respective protein: volume of mother liquor) was used for crystallization. The screenings were left incubating at 18 ºC for 2 days.

*Ye*KdgF crystals in the condition containing 75% tacsimate, pH 7.0, were observed and all three drops contained crystals. The most intact and singular crystals were found in 1:2 ratio drops. The crystal was cryoprotected with 75% tacsimate and diffraction data was collected on beam line 9-2 at the Stanford Synchrotron Radiation Lightsource (SSRL) at 0.9792 Å. Diffraction data was processed using SSRL data analysis software and images collected were scaled and merged to 1.45 Å (Gonzalez et al., 2008) (Data collection statistics are presented in Table 6). Five percent of reflections were set aside for R_{free}. The previously solved model of *Ha*KdgF was processed through CHAINSAW (Stein, 2008) in CCP4 and was used as a search model to solve the *Ye*KdgF structure. One molecule of *Ye*KdgF in the AU was searched in PHASER (McCoy et al., 2007). PHASER produced one solution and was used for automated building using ARP/WARP Crystallographic Macromolecular Model Building program in CCP4 (Langer et al., 2008; Winn et al., 2011; and Murshudov et al., 2011). Some water molecules were modeled automatically, whereas others were modelled in manually using COOT. The unmodeled portions of the structure were manually built using COOT (Emsley and Cowtan, 2004). The metal and malonic acid (from the crystallization condition) was modeled to fit the density within the active site. Compounds from the crystallizing solutions were manually modeled into densities as well. The *Ye*KdgF structure was refined multiple times using REFMAC (Murshudov et al., 2011) to an R_{crys} of 18% and R_{free} of 21%. The refined structure of *Ye*KdgF PROCHECK was used in CCP4 (Laskowski et al., 1993), which showed 95.9% of the residues in favoured conformations and no residues in disallowed orientations.

2.7 Mutagenesis and activity

2.7.1 Quick-change site directed mutagenesis

Site directed mutagenesis primers for *Ha*KdgF were designed based on the guidelines indicated in previously described methods (Zheng et al., 2004). Using the primers in Table 2, QuikChange II Site-Directed Mutagenesis reactions were set up using Phusion polymerase as described in General methods. The first cycle was at 94°C for 3 minutes and the next 16 cycles were as follows: 94°C for 1 minute, 52°C for 1 minute and 68°C for 8 minutes. These 16 cycles were followed by a final extension at 68°C for 60 minutes. The products (5uL) were visualized on 1 % (w/v) agarose gel supplemented with EtBr using the EagleEye II system. The remaining (45uL) of PCR product was purified using PCR purification kit (BIOBASIC) as instructed by the manufacturers. The reaction was treated with DpnI restriction enzyme (NEB) for 1 hour at 37°C. Afterwards, 5uL of digested sample was used to transform competent BL21 (DE3) cells. Colony PCR was used to find successfully transformed clones and the mutation was confirmed by digestion of colony PCR gene product by restriction enzymes that cut the added restriction site during mutagenesis (Table 2). Plasmids of selected colonies were sequenced using the T7 forward primer at Sequetech (Mountain View, CA).

Expression and purification of *Ha*KdgF mutants was carried out as described for *Ha*KdgF wild-type and no secondary purification was required. The protein was dialyzed and buffer exchanged with 20 mM Tris-HCl, pH 7.5.

Table 2. Respective primers used QuikChange Site-directed Mutagenesis of *Ha*KdgF. The restriction sites are shown in bold and the mutated base pairs are capitalized.

Mutant	Forward (5'-3')	Reverse (5'-3)	Restriction site
Q55A	gagcgtcatgacGCg**atCg**gc	gacatag**cGat**cGCgtcatga	DpnII
D102A	agtctggt**gatC**gCcctgttc	cgaggcgagaacaggGc**Gatc**a	DpnII
F104A	gtgattgacct**AGCT**tcgc	gcggcgaggcga**AGCT**agg	AluI
R108A	ctgttctcgcctcg**AGCT**gatgac	ttcagcatgtcatc**AGCT**cgaggc	AluI

2.7.2 Activity of *Ha*KdgF mutants

Reactions were done in UVStar 96 well plates with 1.2 µM *Ye*Ogl, 2.5 mM digalacturonic acid and 100 mM Tris-HCl, pH 7.5. The total volume of the reaction was 100 µL and was allowed to proceed at 30 °C until it plateaued (7 minutes). 25 nM *Ha*KdgF and respective mutants were added to separate reactions and the absorbance was measured for an additional 4 minutes.

Chapter 3. Results

3.1 Bioinformatics Analysis of KdgF

Initial BLAST analysis of KdgF from both *Halomonas* sp. and *Yersinia enterocolitica* revealed significant shared sequence identity with members of the cupin superfamily. However, because of the diverse functionality of this superfamily, this provided no additional insight as to the function of either protein. Thus, the search was expanded to incorporate the genomic context of *kdgF*.

In the alginate utilization locus, *HakdgF* is located between *kdgR* and *sdr,* which are homologs to those genes involved in pectin metabolism (Figure 9A and Table 3). *Ha*kdgF is also situated in close proximity to two putative poly(β-D-mannuronate) lyase genes and one putative broad substrate alginate lyase, homologous to *alyPI* from *Pseudoalteromonas* sp. CY24.

In the pectin utilization locus, *YekdgF* is present downstream of *kduD* and *kdgI* homologs from *D. dadantii* (Figure 9B and Table 4). Also present in the locus, are the *togMNAB* transporter, polysaccharide lyase 2 and a *kdgM* homolog from *D. dadantii* (Blot *et al*., 2002; Hutter *et al*., 2014 Abbott and Boraston, 2007).

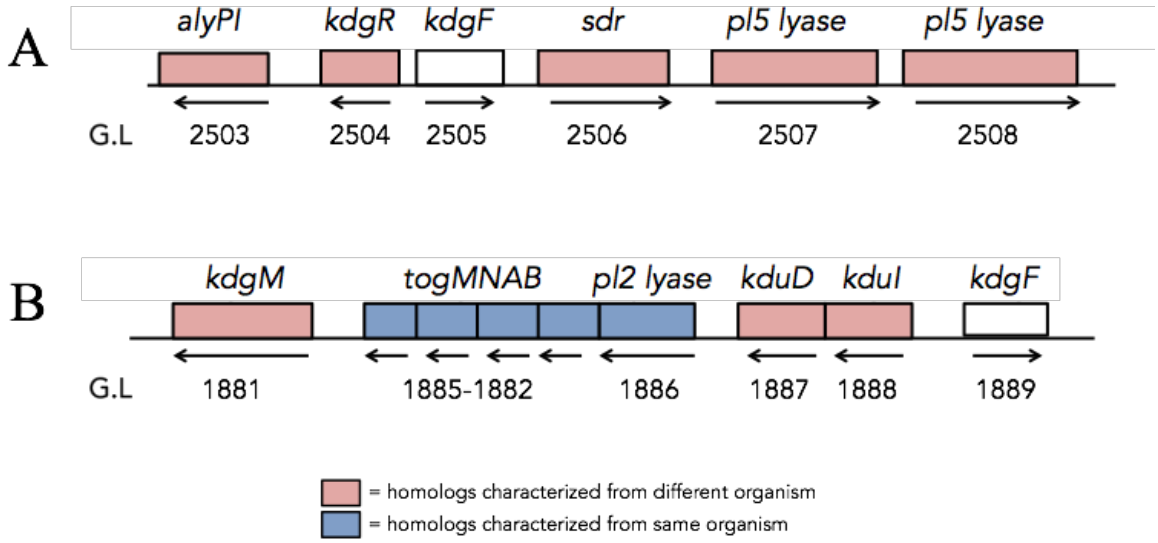


Figure 8. Organization of Pectin and Alginate utilization loci in bacteria.

A) Alginate utilization locus from *Halomonas* sp. and B) Pectin utilization locus from *Yersinia enterocolitica.* Legend: G.L: gene locus. Annotations of genes in loci can be found in Tables 3 and 4.

Table 3. Genomic context of genes surrounding *kdgF* in the *Halomonas* sp. alginate utilization locus from Figure 8A.

Gene locus	Homolog/ Family	Enzyme	Organism	Identity (%)	Coverage (%)	Reference
2503	Polysaccharide lyase family 7	AlyPI	*Pseudoalteromonas* sp. CY24	34	93	Duan, Han, and Yu. (2009)
2504	KdgR**	Transcriptional regulator;, KDG operon repressor	*Dickeya dadantii*	48	98	Rodionov *et al*. (2004)
2505	KdgF	Unknown	*Halomonas*	N/A	N/A	NA
2506	Sdr**	2-dehydro-3-deoxy-D-gluconate 6-dehydrogenase	*Zobellia galactanivorans*	35	97	Thomas *et al*. (2012)
2507	Poly(β-D-mannuronate) lyase PL family 5	poly(β-D-mannuronate) lyase	*Agarivorans albus*	46	96	Uchimura *et al*. (2009)
2508	poly(beta-D-mannuronate) lyase PL family 5	poly(beta-D-mannuronate) lyase	*Agarivorans albus*	44	94	Uchimura *et al*. (2009)

**Directly associated with, or integrated in, the Entner Doudoroff pathway
N/A – not applicable
PL = polysaccharide lyase

Table 4. Genomic context of genes surrounding *kdgF* in the *Y. enterocolitica* pectin utilization locus from Figure 8B.

Gene locus	Homolog/ Family	Enzyme	Organism	Identity (%)	Coverage (%)	Reference
1881	KdgM porin	Oligogalacturonate porin	*Dickeya dadantii*	66	100	Blot *et al*. (2002), Hutter et al. (2014)
1882-1885	TogMNAB ABC transporter	Oligogalacturonate transporter	*Yersinia enterocolitica*	N/A	N/A	Abbott and Boraston (2007)
1886	Polysaccharide lyase family 2	Galacturonic acid lyase	*Yersinia enterocolitica*	N/A	N/A	Abbott and Boraston (2007)
1887	KduD**	2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase	*Dickeya dadantii*	85	100	Condemine and Robert-Baudouy (1991)
1888	KduI**	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase	*Dickeya dadantii*	82	100	Condemine and Robert-Baudouy (1991)
1889	KdgF	Unknown	*Yersinia enterocolitica*	N/A	N/A	N/A

**Directly associated with, or integrated in, the Entner Doudoroff pathway
N/A – not applicable

3.2 Production and purification of *Ha*kdgF and *Ye*KdgF

*Ha*KdgF and *Ye*KdgF were recombinantly expressed in an *E. coli* system. Each protein was purified via nickel immobilized metal affinity chromatography (IMAC) and all samples from purification were run on SDS-PAGE to assess the size and purity of the proteins (Figure 9). *Ha*KdgF and *Ye*KdgF ran between 7 to 17 kDa, which is consistent with their predicted mass (molecular weight of *Ha*kdgF and *Ye*KdgF: 12.5kDa and 12.6kDa respectively) (Figure9A and 9B). The purity after IMAC was sufficient for activity assays.

For crystallization trials, an additional purification step was required to achieve sufficient homogeneity. For *Ha*KdgF, prior to the secondary purification, this protein required the removal of the histidine tag by treatment with thrombin. After digestion, *Ha*Kdgf was purified via tandem

nickel exchange and anion exchange chromatography. *Ha*Kdgf began to elute with 25mM NaCl, 20mM Tris, pH 8.0, (5% Buffer B) and the samples from elution had excellent purity (Figure 10A). *Ye*KdgF was purified via size exclusion chromatography (SEC) column. The first peak of the elution profile showed little to no *Ye*KdgF and slight contamination by what presumably are *E. coli* proteins. The second peak however, had a higher absorbance and thus contained significantly more *Ye*KdgF than the first peak (Figure 10B). Elutions containing pure protein were pooled for crystallization.

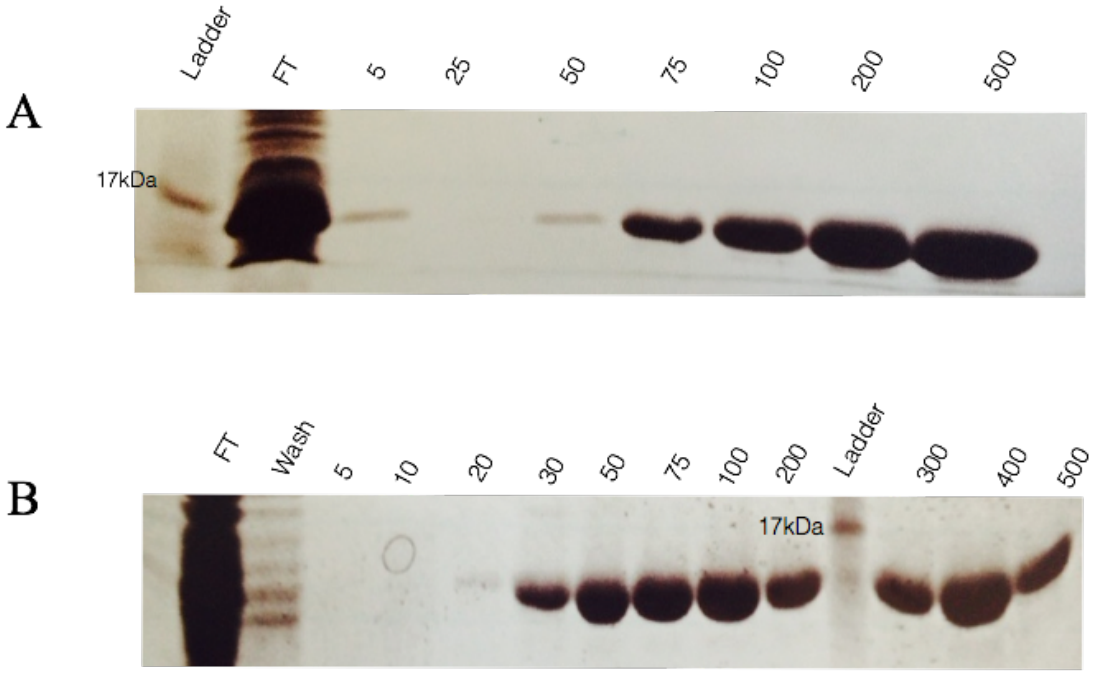


Figure 9. Purification of HaKdgF and YeKdgF. SDS-PAGE gel images of samples eluted from various wash steps during IMAC. The gel was run at 200V for 50 minutes and stained with Coomassie blue stain. FT: flow through 5-500: concentration of imidazole in mM A) The distance travelled by the band correlates with the theoretical weight of HaKdgf (12.5kDa). B) The distance travelled by the band correlates with the theoretical weight of YeKdgF (12.6kDa).

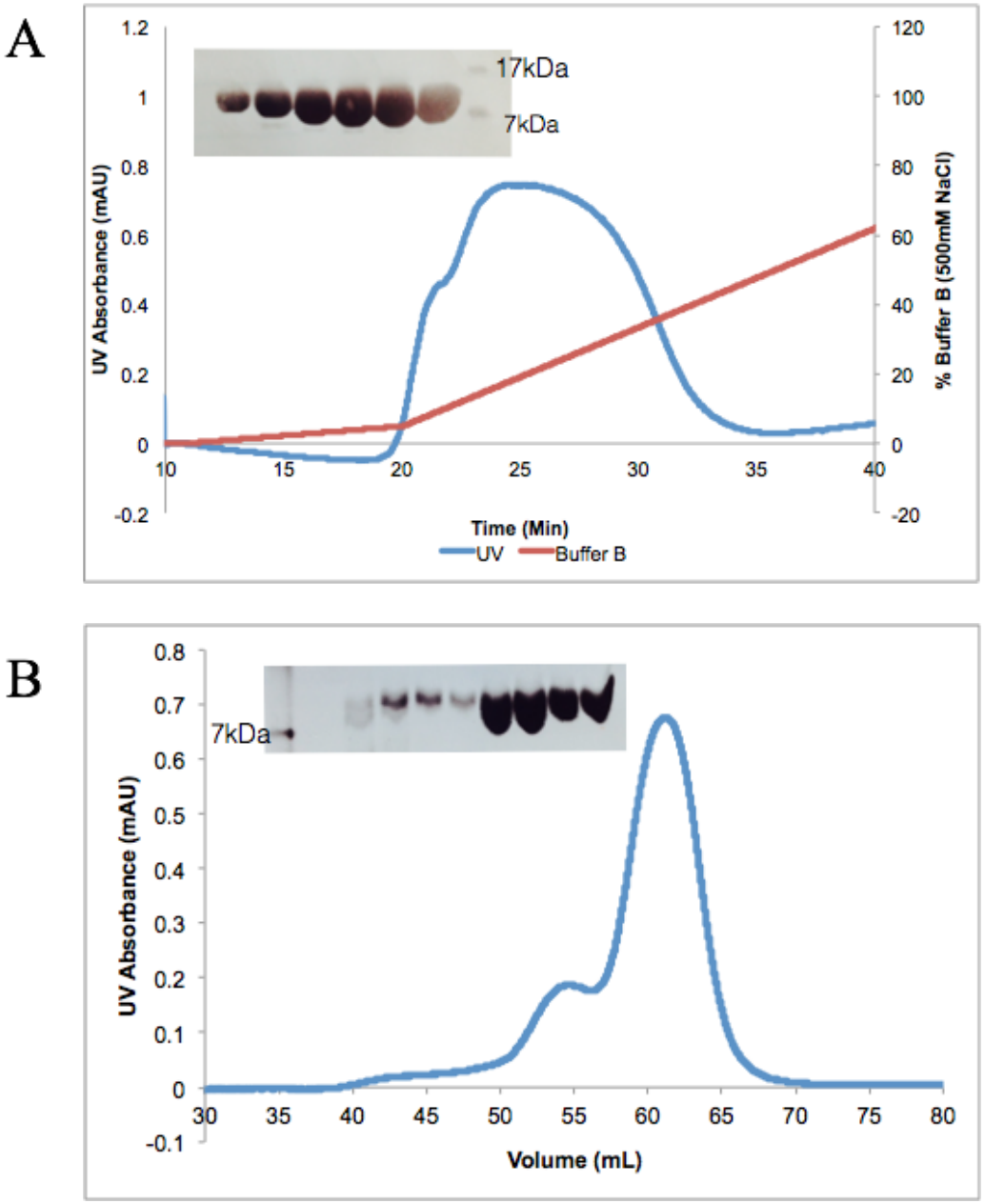


Figure 10. Secondary purification of HaKdgF and YeKdgF for crystallization trials.
A) *Ha*KdgF anion exchange profile, eluted at 5% Elution Buffer (500mM NaCl, 20mM Tris-HCl, pH 8.0) with accompanying SDS-PAGE image B) *Ye*KdgF SEC profile, elution beginning at 58mL of 20mM Tris-HCl (pH 8.0), with accompanying SDS-PAGE image.

3.3 Determining KdgF activity using *Ye*Ogl and digalacturonic acid

To determine whether KdgF has an effect on the Δmonouronic acid from alginate or pectin degradation, a qualitative assay was developed which measures the depletion of Δuronic acid by measuring the absorbance of the unsaturated bond at the C4 and C5 at 232 nm. However, because of the tendency of Δmonouronic acids to spontaneously tautomerize, the substrate needed to be produced immediately before its interaction with KdgF. For this purpose, *Ye*Ogl was utilized to produce the ΔGalA from $GalA_2$ immediately prior to *Ye*KdgF activity experiments (Figure 11A). As expected, when *Ye*Ogl was added to $GalA_2$, the relative absorbance at 232nm increased until a plateau was reached (Figure 11B). Upon reaching the absorption plateau, *Ye*KdgF was added and immediately resulted in a sharp decrease in absorption compared to the negative control (no KdgF), which observed no decrease (Figure 12A).

While we were unable to provide an alginate specific assay for the *Ha*KdgF, its natural substrates (mannuronic acid and guluronic acid) are very similar to the chemical structure of galacturonic acid from pectin (Figure 12B). Thus, *Ha*KdgF activity was tested using the same ΔGalA depletion assay. The addition of *Ha*KdgF caused a drop in absorbance, similar to what was observed in the *Ye*KdgF experiment (Figure 12C).

To ensure that the depletion in ΔGalA observed with *Ye*KdgF and *Ha*KdgF was not an artifact of the assay, the relationship between enzyme concentration and the rate of depletion was investigated using *Ha*KdgF. It was determined that as the concentration of *Ha*KdgF increased, the rate of depletion in absorbance proportionally increased as well (Figure 13). This indicates that the rate of depletion was dependent on the concentration of *Ha*KdgF.

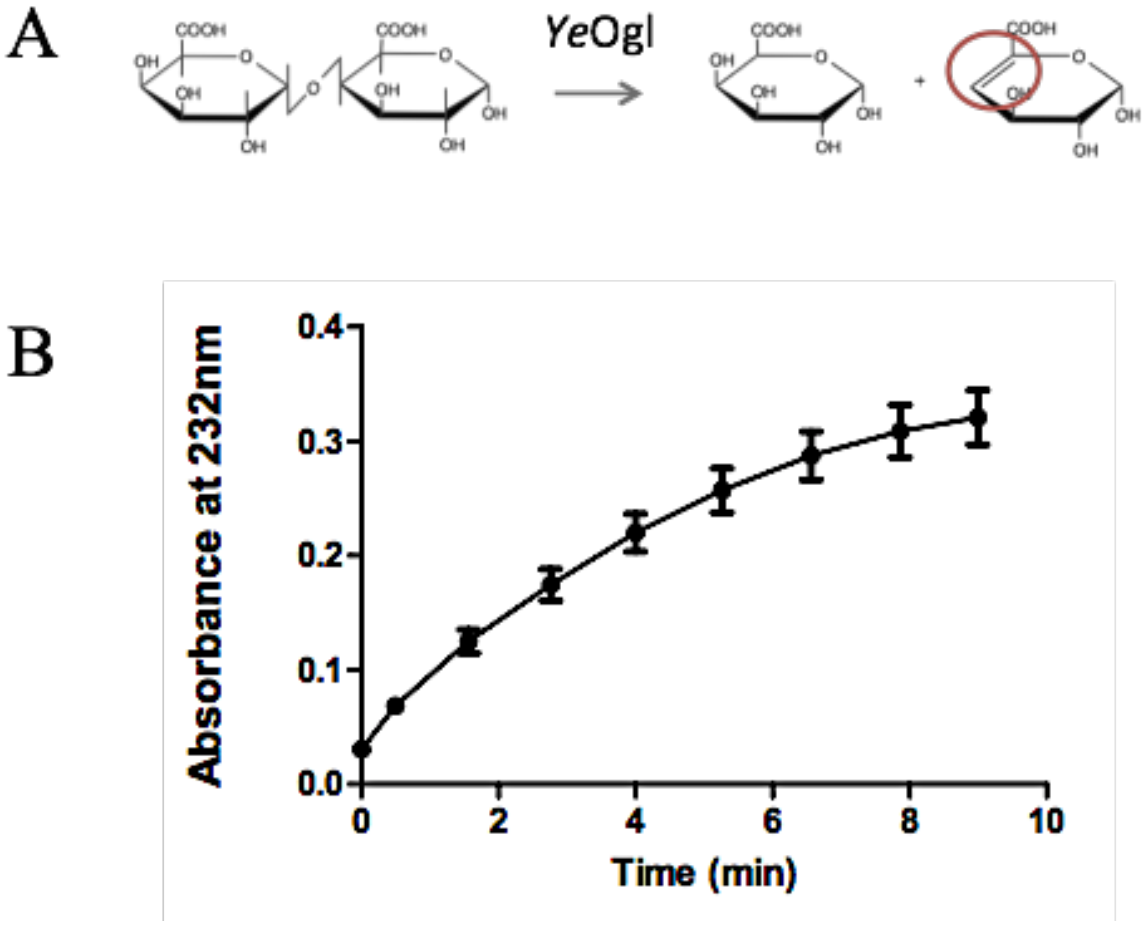


Figure 11. YeOgl activity using digalacturonic acid.

A) *Ye*Ogl catalyzed reaction of digalacturonic acid into monogalacturonic acid and unsaturation of the C4-C5 bond of the monogalacturonic acid. Unsaturated bond is shown with red circle. B) *Ye*Ogl activity assays showing increase in absorbance at 232 nm due to the unsaturated bond formation from the OGL activity.

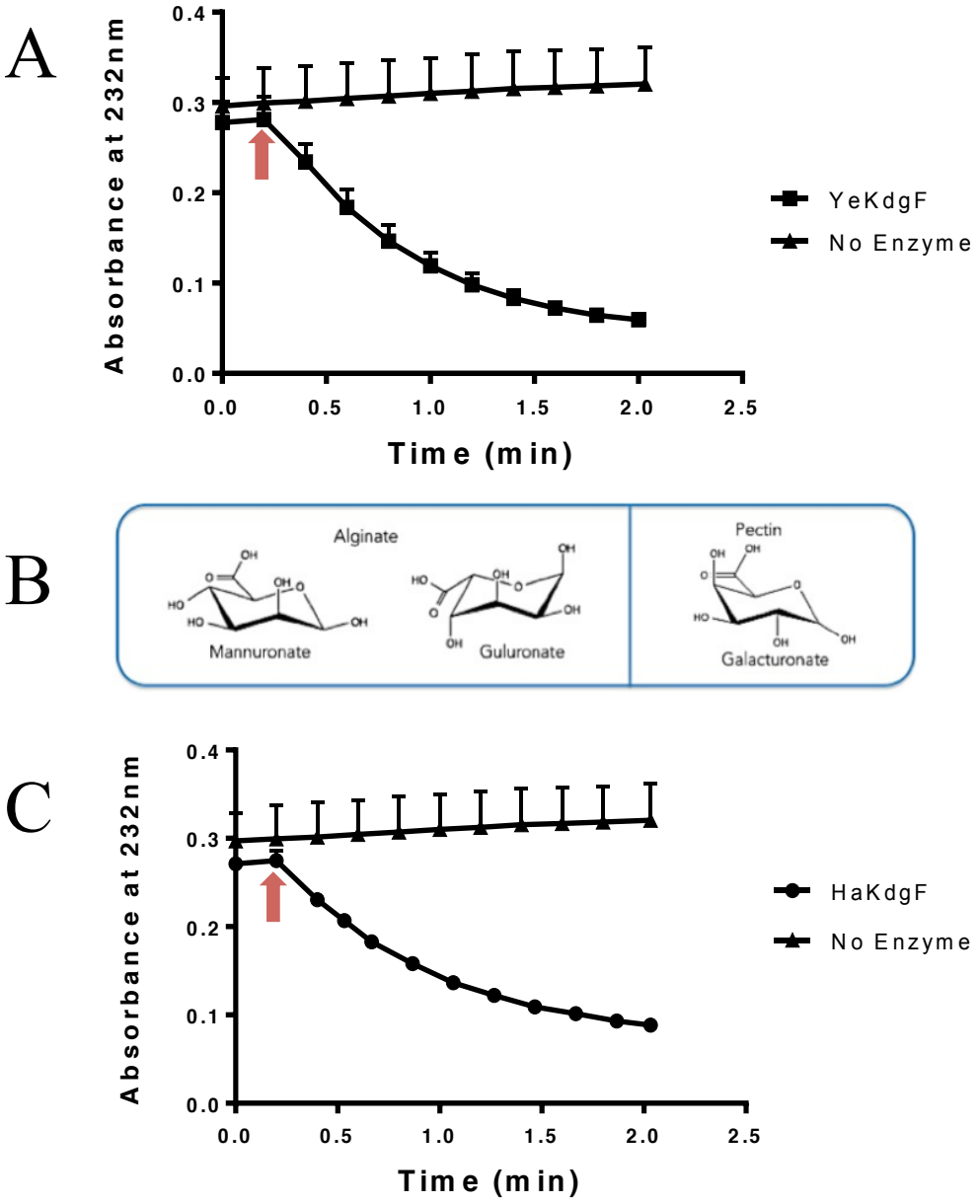


Figure 12. Testing YeKdgF and HaKdgf using ΔGalA depletion assay

A) Assay measuring absorbance at 232 nm from ΔGalA upon addition of *Ye*KdgF at time zero. Error bars, where visible, represent the standard deviation of triplicate experiments. Red arrow indicates when *Ye*KdgF was added B) Monosaccharide units of alginate (left and middle) and pectin (right). C) Assay measuring absorbance at 232 nm from ΔGalA upon addition of *Ha*KdgF at time zero. Error bars, where visible, represent the standard deviation of triplicate experiments. Red arrow indicates when *Ha*KdgF was added.

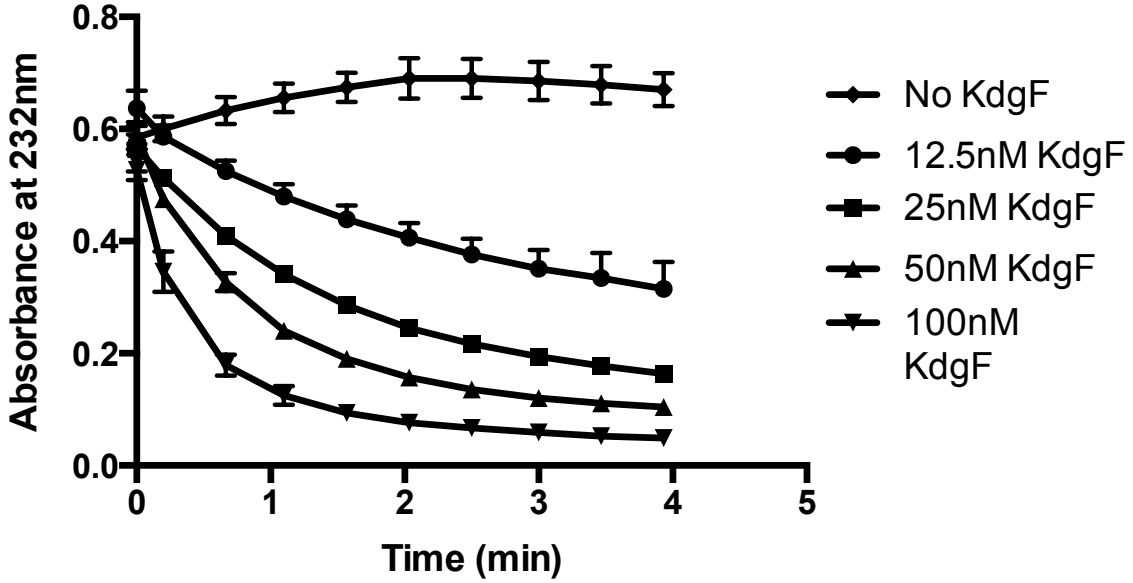


Figure 13. Enzyme concentration dependency of *Ha*KdgF.
Assay measuring absorbance at 232 nm from ΔGalA upon addition of increasing concentrations of *Ha*KdgF. YeOgl and digalacturonic acid was incubated for 7 minutes and *Ha*KdgF was added and absorbance was measured for 4 minutes. Error bars, where visible, represent the standard deviation of triplicate experiments.

3.4 Crystal structure of HaKdgF

The x-ray crystal structure of *Ha*KdgF was solved by molecular replacement using the structure of a Cupin 2 conserved barrel domain, protein chain A, from *Shewanella frigidimarina* with waters and side-chains removed (37.5% sequence identity; PDB: 2PFW). Molecular replacement produced one solution (with 4 molecules in the asymmetric unit) and the majority of the model was automatically built using ARP-WARP in CCP4 suite of programs. The model was refined to 2.0 Å in space group C2 with R and R_{free} values of 21% and 25%. There are four

molecules of *Ha*KdgF in the asymmetric unit (AU), chains A, B, C and D, with A and B forming one dimer and C and D forming the second dimer. The final refined structure begins at Gly4 and extended until Leu112 for chain A and chain B. For chain C and D, the structures start from Phe6 (chain C) and Ser5 (chain D) and continued until Met111 and Leu112, respectively. The significant portion of residues (94.1%) lies in the favorable regions of the Ramachandran plot and no residues are in the disallowed regions (Table 5).

*Ha*KdgF possess a cupin-type β-barrel with two antiparallel β-sheets that form a barrel in the center: one containing five β-strands and the other containing four β-strands (Figure 14A). The center of the β-barrel forms a pocket that is predominantly hydrophobic (Figure 14B) and this is where the metal-binding site is located. Electron density shows the metal ion and coordinating residues: His48, His50, His89 and Gln55, and two water molecules (Figure 14C).

Table 5. Data collection and structure statistics for *Ha*KdgF

Data collection	*Ha*KdgF native
Space group	C2
Cell dimensions	
a, *b*, *c* (Å)	68.00, 78.68, 103.99
α, β, γ (°)	90.0000 102.80, 90.0000
Resolution (Å)	33.20-1.9
R_{sym}	0.128
$I / \sigma I$	4.0
Completeness (%)	100.0
Redundancy	3.6
Refinement	
Resolution (Å)	33.2-1.9
No. reflections	36147
No. Free reflections	1830
R_{work} / R_{free}	0.208/0.246
No. atoms	
Protein Chain A	862
Protein Chain B	862
Protein Chain C	844
Protein Chain D	858
Ligand/ion	12
Water	171
B-factors	
Protein Chain A	31.901
Protein Chain B	30.850
Protein Chain C	32.572
Protein Chain D	35.310
Metal	38.04
Water	63.960
R.m.s. deviations	
Bond lengths (Å)	0.0187
Bond angles (°)	2.0775
Ramachandran	
Preferred (%)	91.3%
Allowed (%)	8.7%
Disallowed (%)	0.0%

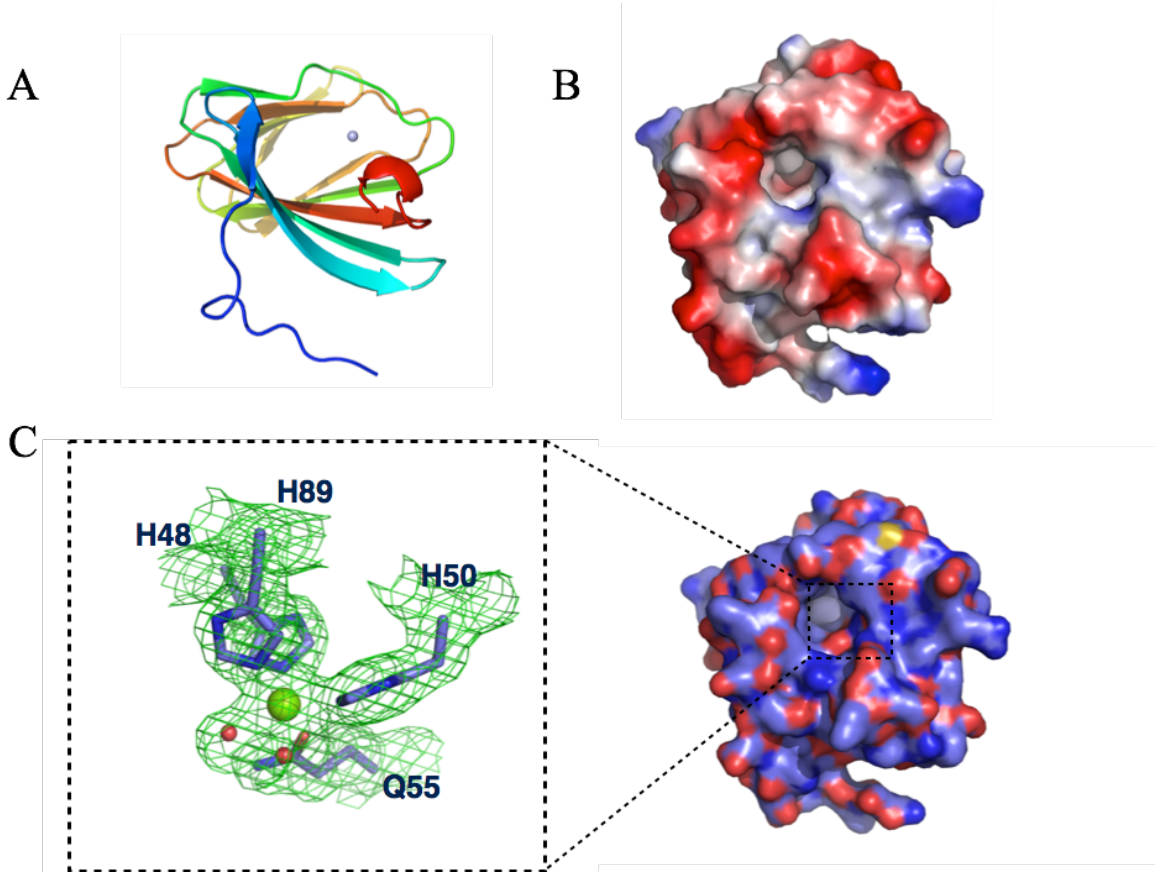


Figure 14. Overall structure of *Ha*KdgF

A) Cartoon representation of *Ha*KdgF showing the β-barrel characteristic of the cupin fold, as well as the metal present in the center of the barrel. The N-terminus begins with blue and C-terminus ends in red. B) Representation of the electrostatic surface of *Ha*KdgF showing the putative active site as a hydrophobic pocket (generated by PyMOL). The protein is centered to display the entrance to the active site. C) His48, His 89, His50 and Gln55 coordinating the metal in the putative active site pocket. There are also two coordinated waters (shown as red spheres). The three histidines and glutamine are shown as blue sticks, and the metal ion is modeled as a green sphere. Green mesh represents the 2Fo-Fc map contoured to 2.0σ. Overall structure of *Ha*KdgF is depicted on the right. Nitrogen and Oxygen in side chains are depicted as red and blue respectively.

3.5 Crystal structure of *Ye*KdgF

The crystal structure of *Ye*KdgF was solved by molecular replacement using *Ha*KdgF chain A with waters and side-chains removed (54 % sequence identity). The model was refined to 1.45 Å in space group P3121 with R and R_{free} values of 18 % and 21 % and had one molecule in the asymmetric unit. 95.9% of the residues lie in the favourable regions of the Ramachandran plot and no residues are in the disallowed regions (Table 6). The overall fold of *Ye*KdgF is a cupin-type β-barrel type, similar to *Ha*KdgF (Figure 15A). The cupin domain is comprised of two antiparallel β-sheets: one containing five strands and the other consisting of four strands. The center of the β-barrel forms a pocket where the entrance is basic but the interior that is predominantly hydrophobic (Figure 15B). The pocket contains a metal that is coordinated by four residues: His46, His48, His87 and Gln53. The electron density in the shape of malonic acid (MLA), a component from the crystallization condition, was also present in this pocket and was coordinated by the metal (Table 6 and Figure 15C).

Table 6. Data collection and structure statistics for *Ye*KdgF

Data collection	*Ye*KdgF native
Space group	$P3_12_1$
Cell dimensions	
a, *b*, *c* (Å)	60.190, 60.190, 67.830
α, β, γ (°)	90, 90, 120
Resolution (Å)	20.74-1.45 (1.52-1.45)*
R_{sym} or R_{merge}	0.049 (0.458)
$I / \sigma I$	15.6
Completeness (%)	99.9 (100.0)
Redundancy	4.8 (3.8)
Refinement	
Resolution (Å)	19.7-1.45
No. reflections	23184
No. free reflections	1171
R_{work} / R_{free}	0.1757/0.2134
No. atoms	1184
Protein	860
Malonic acid	7
Zinc	1
Water	314
B-factors	
Protein	13.2
Zinc	16.9
Malonic acid	12.8
Water	40.8
R.m.s. deviations	
Bond lengths (Å)	0.0178
Bond angles (°)	2.0773
Ramachandran	
Preferred (%)	95.9
Allowed (%)	4.1
Disallowed (%)	0.0

***Values for highest resolution shell are shown in parentheses**

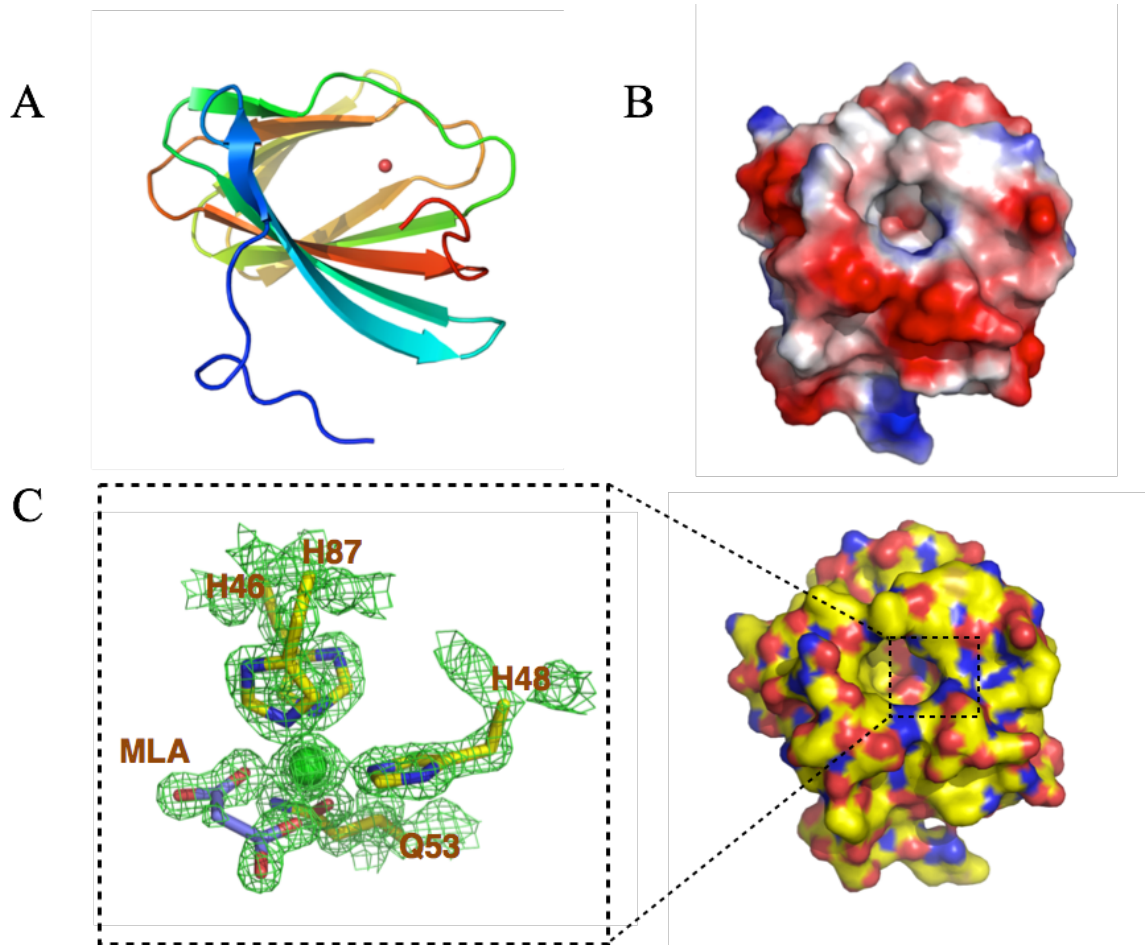


Figure 15. Overall structure of YeKdgF

A) Cartoon representation of *Ye*KdgF showing the β-barrel fold characteristic of the cupin fold, as well as the metal present in the center of the barrel. The N-terminus begins with blue and the C-terminus ends in red. B) Representation of the electrostatic surface of *Ye*KdgF showing the putative active site as a slightly hydrophobic pocket (generated by PyMOL). The protein is centered to display the entrance to the active site. C) His46, His 87, His48 and Gln53 coordinating the metal in the putative active site pocket. There is also a coordinated malonic acid (MLA; shown as blue sticks). The three histidines and glutamine are shown as yellow sticks, and the metal ion as a green sphere. Green mesh represents the 2Fo-Fc map contoured to 2.0σ. Overall structure of *Ye*KdgF is depicted on the right. Nitrogen and Oxygen in side chains are depicted as red and blue respectively

3.6 *Ha*KdgF and *Ye*KdgF are part of the Cupin superfamily

Despite *Ha*KdgF and *Ye*KdgF originating from different uronate degradation systems, the two proteins have high percent identity (54%) in their amino acid sequence (Figure 16A). Structural overlay revealed that the two proteins aligned with RMSD of 0.515 Å (90 of 95 Cαs aligned)(Figure 16B). The structure of *Ha*KdgF and *Ye*KdgF confirmed that these enzymes are a part of the cupin superfamily; however due to the lack of sequence conservation and diverse functionality of this superfamily, it does not give us insight of the function or specificity of these enzymes. The catalytic amino acids of the cupin superfamily have generally been located at the center of the barrel where the metal lies (Figure 17C-E). Figure 17A-B shows the metal binding site of *Ye*Kdgf and *Ha*Kdgf respectively, and Figure 17C-E shows the metal binding site of three representative cupin proteins with substrates or inhibitors bound. When the binding residues of both KdgF and the chosen cupin proteins were aligned, the residues involved in metal binding were spatially conserved and their substrates localized in the same area close to the metal, suggesting this area is the active site (Figure 17F).

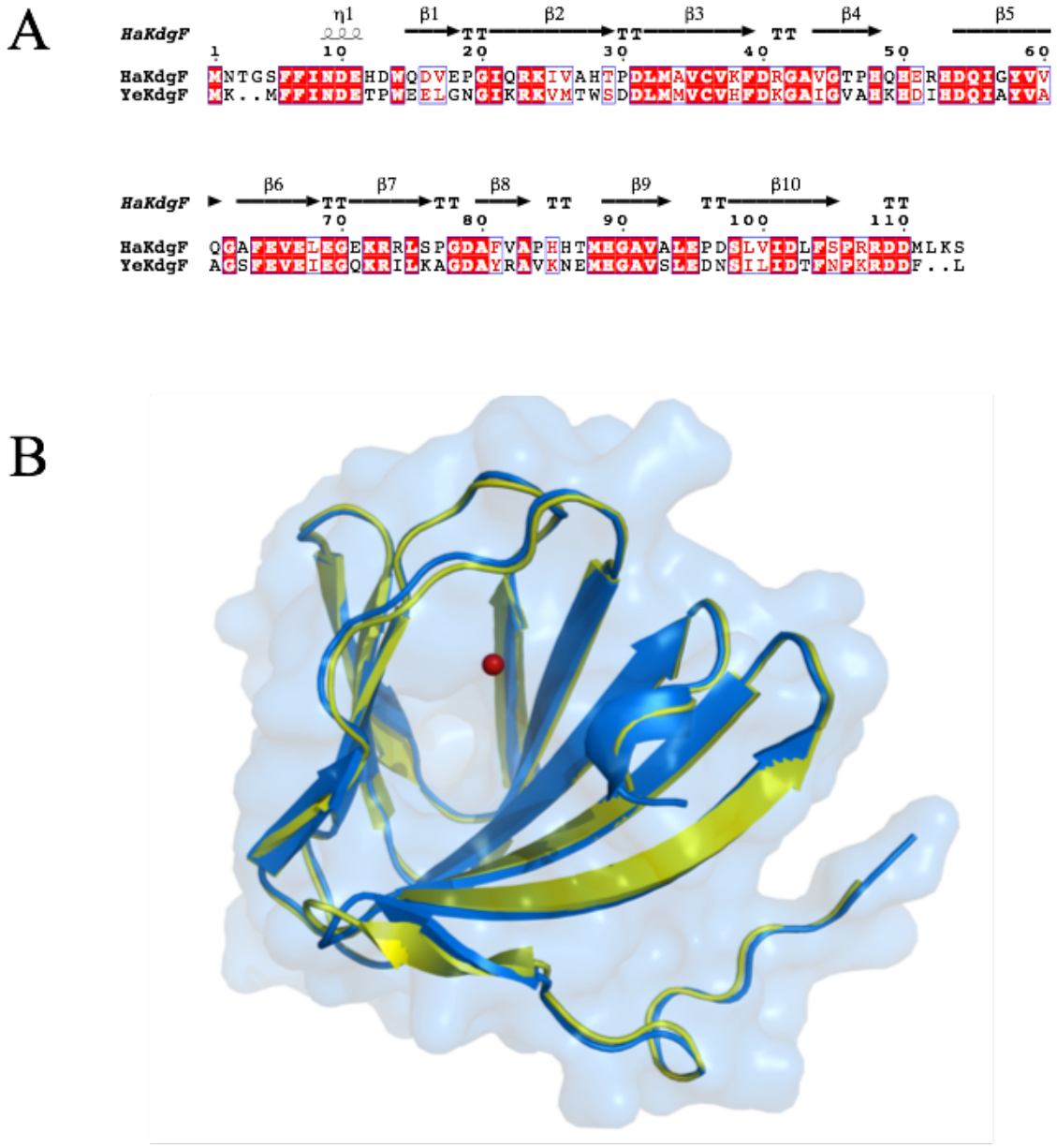


Figure 16. *Ha*Kdgf and *Ye*Kdgf alignments show high sequence and structure conservation.
A) Amino acid sequence alignment of *Ha*KdgF and *Ye*KdgF using Clustal Omega. This figure was generated using ESPript with secondary structure elements indicated from structure file of *Ha*KdgF. B) Structural overlay of *Ha*KdgF (Yellow) and *Ye*KdgF (Blue) with RMSD= 0.515 Å. Red sphere represents the coordinated metal.

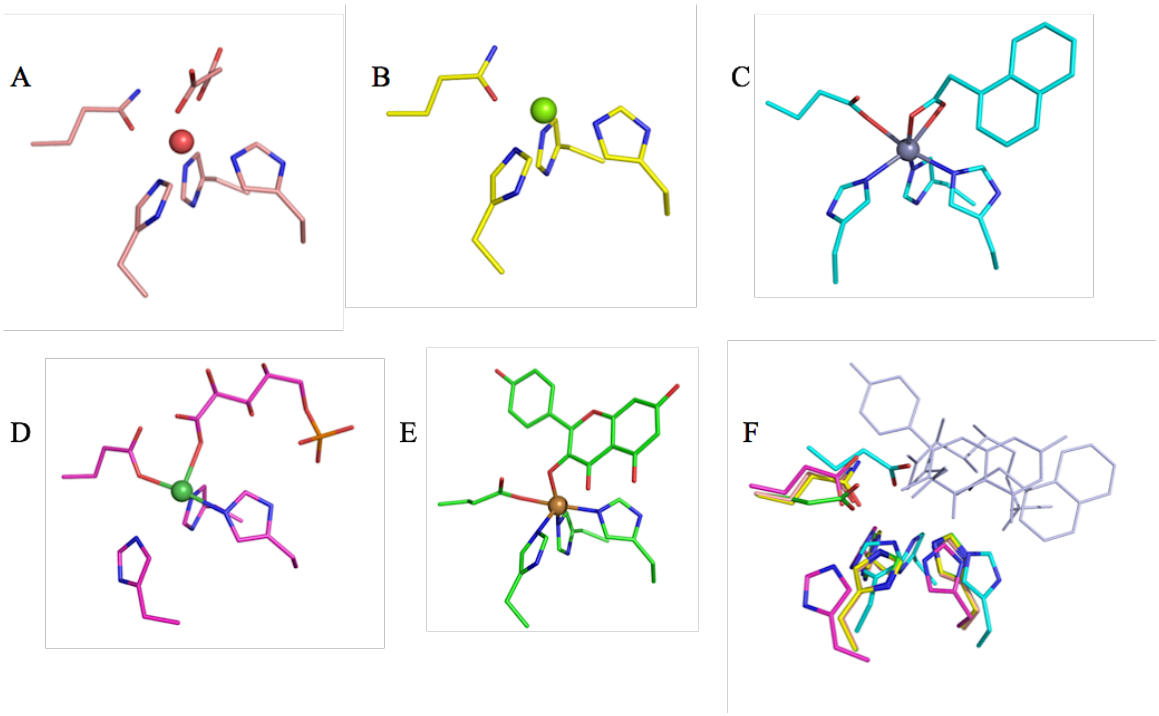


Figure 17. Conserved metal coordination among Cupin superfamily enzymes.
A) *Ye*KdgF with malonic acid, B) *Ha*KdgF, C) auxin-binding protein 1 complex with naphthalene-1-YL-acetic acid (NPL or auxin) from *Zea mays*. Coordinated metal is Zn^{2+} (PDB: 1LRH Woo et al., 2002), D) Phosphoglucose isomerase enzyme complex with 5-phosphoarabinonate from *Pyrococcus furiosus* (coordinated metal is Fe^{2+})(PDB: 1QXR Swan et al., 2003), E) Quercetin 2,3-dioxygenase complex with 3,5,7-Trihydroxy-2-(4-Hydroxyphenyl)-4H-Chromen- 4-one from *Aspergillus japonicus* (coordinated metal is Cu^{2+})(PDB: 1H1M Steiner et al., 2002), F) alignment of the metal coordination sites from A-E and respective substrates shown in C-E.

3.7 Active site mutagenesis

Putative catalytic residues were selected based on their biochemical properties and their location near the coordinated metal in the putative active site (Figure 18). The candidate residues to be involved in binding and catalysis in the active site were Gln55, Asp102, Phe104 and Arg108. These residues were mutated to alanine in *Ha*KdgF using the Quik-change method, which was confirmed by sequencing. These mutants were tested for activity similar to the testing of the wild-type and all mutants' ability to deplete ΔGalA was abrogated and no depletion of absorbance at 232nm was observed, compared to the wild-type (Figure 19).

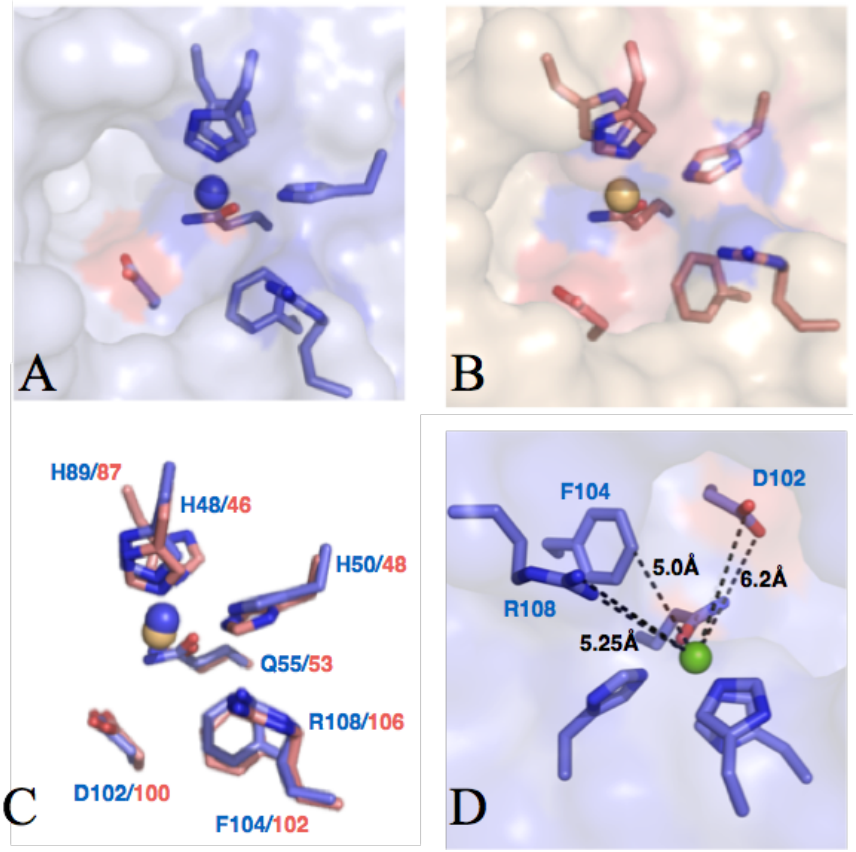


Figure 18. Close up stick representation of the metal binding pocket coordinating and putative catalytic residues.

A) *Ha*KdgF, B) *Ye*KdgF, C) Alignment of the residues shown in A and B, and D) distances from putative catalytic residues to metal. Distances were generated by PyMOL.

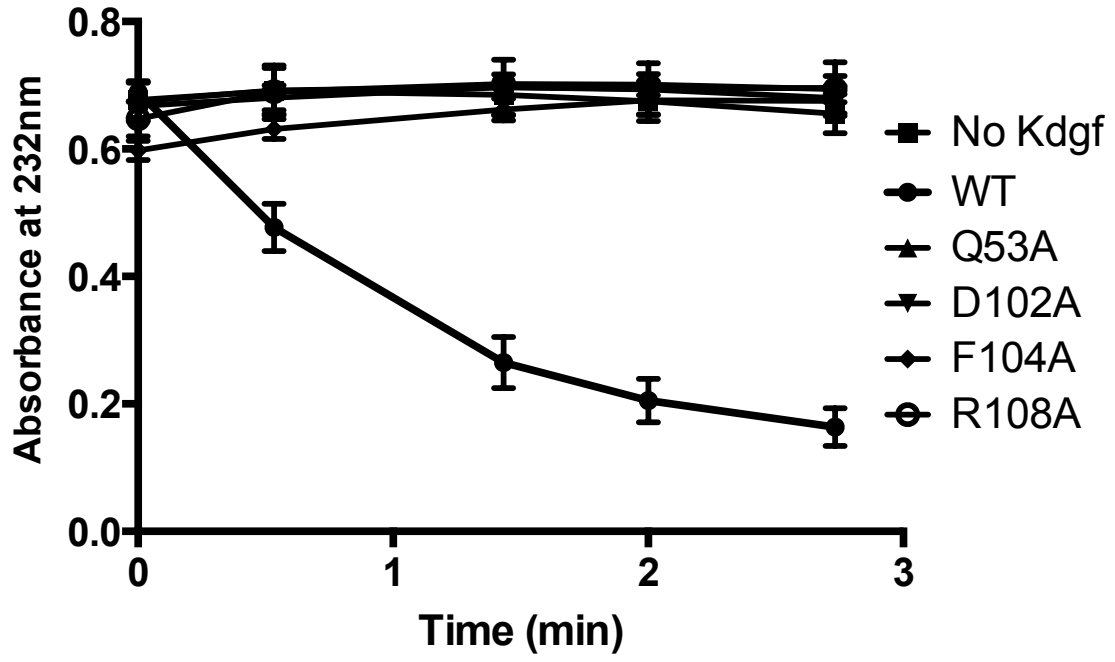


Figure 19. Assay measuring the depletion of ΔGalA upon addition of KdgF mutants and wild type. Assay measuring absorbance at 232 nm from ΔGalA upon addition of *Ye*KdgF at time zero. The reaction was started with 1.2μM *Ye*Ogl, 2.5mM digalacturonic acid and 100mM Tris-HCl pH 7.5. Once the reaction plateaued, 25nM of *Ha*KdgF wild type and respective mutants were added. Error bars, where visible, represent the standard deviation of triplicate experiments

Chapter 4. Discussion

4.1 Bioinformatics analysis suggests Kdgf is involved in uronic acid metabolism.

An uncharacterized gene called *kdgF* was identified in an alginolytic locus from *Halomonas sp*. and a pectinolytic locus from *Yersinia enterocolitica*. Initial amino acid sequence analysis identified KdgF as a member of the cupin superfamily, a family of structurally conserved but functionally divergent proteins (Table 3 and 4) (Dunwell *et al.,* 2001). The amino acid sequence of KdgF is highly conserved among a broad range of other polyuronic acid-utilizing prokaryotes. Due to the functional diversity within this family, it was difficult to predict the function of KdgF. The genomic context was analyzed and the genes that catalyzed steps in the ED pathway were close in proximity to *kdgf* in both alginolyic and pectinolytic loci (Figure 8). All steps in the ED pathway except one in the *D. dadantii* pectinolytic system had been studied in detail and the enzymes responsible for catalysis have been identified (Hugouvieux-Cotte-Pattat *et al*., 1996). Prior to this study, the only gene in the *D. dadantii* pectinolytic locus whose function has not determined is *kdgF* (Rodionov *et al.,* 2004). The genomic context of *YekdgF* and *HakdgF* was concurrent with previous analysis, which has shown that genes involved in the ED pathway are also closely associated with *kdgF* and are often immediately adjacent to it. Additionally, genes involved in the ED pathway can be controlled by the same regulatory elements as *kdgF* in a wide range of Enterobacteriaceae families and marine bacteria (Rodionov*, et al.* 2004; Thomas *et al.,* 2012). This bioinformatics analysis revealing the conservation and genomic context of *kdgF* suggested that it has a role in uronic acid metabolism through the ED pathway.

4.2 *Ye*KdgF and *Ha*KdgF deplete ΔGalA

Given that the bioinformatics analysis of *YekdgF* and *HakdgF* suggested that it is involved in the first step of the ED pathway, the ability of *Ye*KdfgF and *Ha*KdgF to act on ΔGalA were tested. ΔGalA was the only substrate tested due to limitation of substrate availability. *Ye*KdgF caused depletion in absorbance at 232 nm, which indicates that the concentration of ΔGalA was decreasing. This suggested that *Ye*KdgF causes a change to the double bond of ΔGalA. The monomeric units that make up polygalacturonic acid and alginate polymers have the same chemical composition and differ only in the orientation of hydroxyl groups on the sugar (Figure 12B). For this reason, *Ha*Kdgf was tested under the same experimental conditions as *Ye*KdgF. When tested with ΔGalA, *Ha*Kdgf induced a depletion of absorbance at 232 nm, suggesting that *Ha*Kdgf can affect the double bond of ΔGalA and it is not specific for substrates with certain stereochemistry. However, further testing with other uronic acids must be tested to confirm stereochemical specificity. This is an interesting finding as the source of KdgF was from an alginolytic system, however it was active on a pectin-derived substrate. Conversely, whether *Ye*KdgF can act on alginate derived substrate is unknown.

Attempts on determining the function of KdgF has been made, but was never fully elucidated previous to this study. The *kdgf* gene has been knocked out in *D. dadantii* in parallel with *kduD* and *kduI* and two additional enzymes in the pectin degradation pathway (Condemine and Robert-Baudouy, 1991). The level of induction of pectate lyases (controlled by KdgR) by polygalacturonic acid was reduced in the knockout of *kdgf* compared to the wild-type strain, however no growth phenotype could be associated with the knockout. If the conversion of ΔGalA to DTH is spontaneous and KdgF catalyzes this reaction, the rate of degradation would decrease but no noticeable phenotype would be observed since the reaction can occur

spontaneously. It is well established that KDG, a downstream product of the ED pathway, interacts with the repressor KdgR, allowing the expression of pectinolytic genes (Hugouvieux-Cotte-Pattat et al., 1996). Because the conversion of ΔGalA to DTH is rate limiting, the removal of KdgF would resulting in a build up in substrate concentration upstream of the pathway and subsequently result in decrease in the rate of KDG production, resulting in the decrease in induction of the pectinolytic genes in *kdgF* knockouts. This is concurrent with what was observed in the *kdgF* knockouts used in the Condemine and Robert-Baudouy studies.

4.3 Crystal structure of KdgF

Since the activity of *Ye*KdgF and *Ha*KdgF on ΔGalA was confirmed, the structure of both these enzymes were solved by x-ray crystallography in attempt to elucidate how they recognize and act on ΔGalA. Initial primary amino acid sequence analysis identified both *Ye*KdgF and *Ha*KdgF as members of the cupin superfamily. The crystal structures confirmed that both *Ye*KdgF and *Ha*KdgF adopt a β-barrel-like structure that is characteristic of the cupin fold (Figure 14A and 15A). In the center of the β-barrel, a metal binding site with chemical similarity was observed in a pocket of both the structures that is likely an active site (Figure 14B and 15B). Taking a closer look at the metal binding site, the electron density revealed a bound metal that is coordinated with octahedral geometry with three histidines and a glutamine in both *Ye*KdgF and *Ha*KdgF. The cupin superfamily has a bound metal that is coordinated with typically three histidines and a glutamine in the substrate-binding pocket (Dunwell *et al.,* 2004). The alignment of *Ye*KdgF or *Ha*KdgF crystal structures produced an rmsd of 0.515 (over 90 out of 95 Cαs) with a percent identity of 54% (Figure 10). This structural comparison suggested that *Ye*KdgF and *Ha*KdgF are related to members of the cupin superfamily.

4.4 Structural insight on putative catalytic residues

Given the structure of *Ye*KdgF and *Ha*KdgF suggested that they are in the cupin superfamily with a metal that is coordinated by three histidines and a glutamine, putative key residues in the metal binding pocket were selected. Figure 17 C-E shows various proteins from the cupin superfamily with substrates interacting directly with the metal of the active site. Additionally, those same members of the cupin superfamily have been shown to have surface exposed residues near the metal involved in binding and catalysis (Swan *et al.,* 2003, Steiner *et al.,* 2002). Therefore, it can be speculated that *Ye*KdgF and *Ha*KdgF have residues near the metal that are involved in binding or catalysis. Taking a closer look at the metal binding pocket, in addition to the metal binding residues, there are three amino acids that are present and conserved in both *Ye*KdgF and *Ha*KdgF: Asp100/102, Phe 102/104, Arg106/108 (Figure 18A and B). When the *Ye*KdgF and *Ha*KdgF crystal structures were aligned, these residues as well as the metal binding residues aligned perfectly in three-dimensional space (Figure 18C). The Arg106/108 is positioned at the entrance to the pocket, Asp100/102 is positioned in the interior, and Phe102/104 is positioned in between the two. These residues lie very closely to the metal binding site (approximately 5-6 Å away) and therefore, it is likely that these residues contribute to the stabilization and/or catalytic conversion of the substrate (Figure 18D).

4.5 Mutants and their activity

To confirm whether these amino acids are involved in the activity of KdgF, they were mutagenised in *Ha*KdgF. Four residues Gln55, Asp102, Phe104 and Arg108 in *Ha*KdgF were mutated to alanine, expressed separately, and depletion assay were performed on each of the mutant proteins. The absorbance at 232 nm over time was similar to the negative control (no

KdgF) for all mutants, indicating that KdgF with any one of these residues mutated cannot act on the substrate (ΔGalA). This lack of enzyme activity indicates that the mutants are catalytically inactive; this could be due to a lack of affinity for the substrate or an inability to attack the substrate. The underlying molecular mechanism cannot be fully elucidated without a structure of *Ye*KdgF or *Ha*KdgF in complex with the substrate. Prior to this study, there were no known enzymes that act upon Δmonouronic acids; however, there have been studies on a class of enzymes called glucuronyl hydrolases (UGL) that hydrolyze disaccharides consisting of Δglucuronic acid (ΔGlcA) and N-acetylgalactosamine (components of glycosaminoglycans) at the non-reducing terminus (Hashimoto *et al.,* 1999; Itoh *et al.,* 2006; Johnkees and Withers, 2011). In the process of cleaving the glycosidic bond, the unsaturated sugar on the non-reducing end collapses because of its hemiketal transition state; this results in the formation of DTH before it forms a ring again (Itoh *et al*., 2006; Jongkees *et al*., 2011). Itoh suggested that the specificity for the non-reducing terminus is due to the enrichment of electrons at the C4=C5 double bond; therefore, the catalytic reaction mechanism involves a water addition reaction of the vinyl ether group (C4=C5-O5) (Itoh *et al*., 2006). There are two aspartates (Asp149 and Asp88) in the active site of UGL, both of which interact intimately with ΔGlcA of the disaccharide and have been shown to be important for catalytic activity. Asp149 was hypothesized to be a general acid/base catalyst to protonate the vinyl group (C4=C5-C6) and deprotonate the water molecule, respectively. This results in the formation of the hemiketal state, which then spontaneously converts ΔGlcA to an α-keto-acid form (DTH). This triggers the hydrolysis of the glycosidic bond. The other aspartate (Asp88) stabilizes the oxocarbenium ion state of C5 during the transition state, setting it up for attack by the endocyclic oxygen. In the mutant activity assays, Gln55A, Asp102A, Phe104A, and Arg108A also displayed a complete

loss of activity, but their roles in catalysis are also unclear without a substrate complex structure. In the active site of UGL, a glutamine residue Gln211 also interacts closely with ΔGlcA. The authors of this study suggested that the Gln211 plays a crucial role in coordinating a water molecule at a suitable position for catalysis. We can therefore speculate that the loss of activity in the Gln55A mutant of *Ha*KdgF is due to a failure to coordinate the metal ion and/or water properly. The Phe91, due to its hydrophobic nature, was suggested to be important for the stabilization of the hydrophobic ring of the sugar in UGL. Since the mutant with Phe104A displayed complete loss in activity, it is possible that it is also important for the stabilizations of the hydrophobic ring of the ΔGlcA. The arginine residue in UGL (Arg221) interacts with the carboxylic group of ΔGlcA, but activity was still comparable to the wild-type when Arg221 was mutated. In *Ha*KdgF, there was a complete loss of activity when Arg108 was mutated, suggesting that it may have a more important role in catalysis than Arg221 from UGL. Together, the assays with the mutants of Asp102, Gln55A, Phe104A and Arg108A suggest that these residues are essential for catalysis by *Ha*KdgF and provided insight into the molecular details of how all KdgF proteins function.

4.6 Conclusions

Bioinformatics analysis suggested that *Ye*KdgF and *Ha*KdgF were involved in the ED pathway and act on monouronic acids. More specifically, the genomic context and conservation of *kdgF* suggests that it catalyzes the first step of the ED pathway. Taking advantage of the spectrophotometric properties of the ΔGalA, we were able to qualitatively measure the depletion of ΔGalA upon addition of both *Ha*KdgF and *Ye*KdgF. Together, the data presented in this thesis suggest that KdgF acts on Δmonouronic acid in the first step of the pathway, which was

previously thought to occur strictly spontaneously uronic acid metabolism in terrestrial and marine bacteria. Since the first study of alginate and pectin degradation in the 1950's, this spontaneous first step of the degradation of mono-uronate sugars produced by lyases was assumed to be a keto-enol tautomerization reaction. The linearization of Δuronate does occur spontaneously through initial water assisted ring opening followed by tautomerization, however in this study we have shown that KdgF is often present in pectinolytic and alginolytic loci and is expressed to catalyze a change affecting the double bond in Δuronate sugars. To understand how the protein functions, three-dimensional structure of *Ye*KdgF and *Ha*Kdgf was determined using X-ray crystallography. *Ye*KdgF and *Ha*Kdgf were both characterized to be in the cupin superfamily due to the highly conserved structure and metal binding motif. From deducing where the putative active site is, the surrounding residues were examined and the most conserved and exposed residues were chosen for site-directed mutagenesis studies of *Ha*KdgF. All the selected amino acids were shown to be important in the activity of *Ha*KdgF. This gave us a better understanding of the residues that may be involved in binding and catalysis of KdgF proteins. However, the exact underlying mechanism still remains to be elucidated. One way that could determine the mechanism and molecular details of how both *Ye*KdgF and *Ha*KdgF function is by using x-ray crystallography to obtain crystal structures complexed with their respective substrates. Nevertheless, the results obtained in this study give a greater understanding KdgF functions in the ED pathway.

Although further research is required, these data provide evidence to apprise the standard ED pathway for degradation of Δuronic acids with the addition of an initial enzymatically catalyzed step. This step, although exergonic under standard conditions, would have its reaction rate increased with addition of an enzyme. The ED pathway is argued to be one of the earliest

sugar catabolism pathways and an increase in rate of degradation in an energy-limited environment may have been essential to the survival of these species. By studying the enzymes in the ED pathway in their current state we can learn more about how catabolism of simple sugars and has evolved.

Bibliography

Abbott, D. W., & Boraston, A. B. (2007). A Family 2 Pectate Lyase Displays a Rare Fold and Transition Metal-assisted -Elimination. *Journal of Biological Chemistry*, *282*(48), 35328–35336. doi:10.1074/jbc.M705511200

Abbott, D. W., & Boraston, A. B. (2007). Specific Recognition of Saturated and 4,5-Unsaturated Hexuronate Sugars by a Periplasmic Binding Protein Involved in Pectin Catabolism. *Journal of Molecular Biology*, *369*(3), 759–770. doi:10.1016/j.jmb.2007.03.045

Abbott, D. W., Gilbert, H. J., & Boraston, A. B. (2010). The Active Site of Oligogalacturonate Lyase Provides Unique Insights into Cytoplasmic Oligogalacturonate -Elimination. *Journal of Biological Chemistry*, *285*(50), 39029–39038. doi:10.1074/jbc.M110.153981

Adams, J. M., Gallagher, J. A., & Donnison, I. S. (2009). Fermentation study on Saccharina latissima for bioethanol production considering variable pre-treatments. *Journal of Applied Phycology*, *21*(5), 569–574. doi:10.1007/s10811-008-9384-7

Ahmed, H., Ettema, T. J. G., Tjaden, B., Geerling, A. C. M., van der Oost, J., & Siebers, B. (2005). The semi-phosphorylative Entner–Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochemical Journal*, *390*(2), 529. doi:10.1042/BJ20041711

Andriamanantoanina, H., & Rinaudo, M. (2010). Characterization of the alginates from five madagascan brown algae. *Carbohydrate Polymers*, *82*(3), 555–560. doi:10.1016/j.carbpol.2010.05.002

Anower-E-Khuda, M. F., & Kimata, K. (2015). Human Blood Glycosaminoglycans: Isolation and Analysis. In K. Balagurunathan, H. Nakato, & U. R. Desai (Eds.), *Glycosaminoglycans*

(Vol. 1229, pp. 95–103). New York, NY: Springer New York. Retrieved from http://link.springer.com/10.1007/978-1-4939-1714-3_10

Blot, N. (2002). The Oligogalacturonate-specific Porin KdgM of Erwinia chrysanthemi Belongs to a New Porin Family. *Journal of Biological Chemistry*, *277*(10), 7936–7944. doi:10.1074/jbc.M109193200

Boyd, A., & Chakrabarty, A. M. (1995). Pseudomonas aeruginosa biofilms: role of the alginate exopolysaccharide. *Journal of Industrial Microbiology*, *15*(3), 162–168. doi:10.1007/BF01569821

Caffall, K. H., & Mohnen, D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate Research*, *344*(14), 1879–1900. doi:10.1016/j.carres.2009.05.021

Condemine, G., & Robert-Baudouy, J. (1987). Tn *5* insertion in *kdgR* , a regulatory gene of the polygalacturonate pathway in *Erwinia chrysanthemi*. *FEMS Microbiology Letters*, *42*(1), 39–46. doi:10.1111/j.1574-6968.1987.tb02296.x

Conway, T. (1992). The Entner-Doudoroff pathway: history, physiology and molecular biology. *FEMS Microbiology Reviews*, *9*(1), 1–27.

Davis, T. A., Volesky, B., & Mucci, A. (2003). A review of the biochemistry of heavy metal biosorption by brown algae. *Water Research*, *37*(18), 4311–4330. doi:10.1016/S0043-1354(03)00293-8

Draget, K. I., Skjåk-Bræk, G., Christensen, B. E., Gåserød, O., & Smidsrød, O. (1996). Swelling and partial solubilization of alginic acid gel beads in acidic buffer. *Carbohydrate Polymers*, *29*(3), 209–215. doi:10.1016/0144-8617(96)00029-X

Draget, K. I., Smidsrød, O., & Skjåk‐Bræk, G. (2005). Alginates from algae. Biopolymers Online.

Duan, G., Han, F., & Yu, W. (2009). Cloning, sequence analysis, and expression of gene alyPI encoding an alginate lyase from marine bacterium Pseudoalteromonas sp. CY24. *Canadian Journal of Microbiology*, *55*(9), 1113–1118. doi:10.1139/w09-051

Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R., & Goodenough, P. W. (2001). Evolution of functional diversity in the cupin superfamily. *Trends in Biochemical Sciences*, *26*(12), 740–746.

Edwards, M. C., & Doran-Peterson, J. (2012). Pectin-rich biomass as feedstock for fuel ethanol production. *Applied Microbiology and Biotechnology*, *95*(3), 565–575. doi:10.1007/s00253-012-4173-2

Eisenberg, R. C., & Dobrogosz, W. J. (1967). Gluconate metabolism in Escherichia coli. *Journal of Bacteriology*, *93*(3), 941–949.

Emsley, P., & Cowtan, K. (2004). *Coot* : model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography*, *60*(12), 2126–2132. doi:10.1107/S0907444904019158

ENTNER, N., & DOUDOROFF, M. (1952). Glucose and gluconic acid oxidation of Pseudomonas saccharophila. *The Journal of Biological Chemistry*, *196*(2), 853–862.

Ertesvåg, H., & Valla, S. (1998). Biosynthesis and applications of alginates. *Polymer Degradation and Stability*, *59*(1-3), 85–91. doi:10.1016/S0141-3910(97)00179-1

Fabris, M., Matthijs, M., Rombauts, S., Vyverman, W., Goossens, A., & Baart, G. J. E. (2012). The metabolic blueprint of Phaeodactylum tricornutum reveals a eukaryotic Entner-

Doudoroff glycolytic pathway: The Phaeodactylum tricornutum metabolic blueprint. *The Plant Journal*, *70*(6), 1004–1014. doi:10.1111/j.1365-313X.2012.04941.x

Fourest, E., & Volesky, B. (1997). Alginate Properties and Heavy Metal Biosorption by Marine Algae. *Applied Biochemistry and Biotechnology*, *67*(3), 215–226. doi:10.1007/BF02788799

Georgianna, D. R., & Mayfield, S. P. (2012). Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature*, *488*(7411), 329–335. doi:10.1038/nature11479

Hashimoto, W., Kobayashi, E., Nankai, H., Sato, N., Miya, T., Kawai, S., & Murata, K. (1999). Unsaturated Glucuronyl Hydrolase of Bacillus sp. GL1: Novel Enzyme Prerequisite for Metabolism of Unsaturated Oligosaccharides Produced by Polysaccharide Lyases. *Archives of Biochemistry and Biophysics*, *368*(2), 367–374. doi:10.1006/abbi.1999.1305

Hay, I. D., Ur Rehman, Z., Ghafoor, A., & Rehm, B. H. A. (2010). Bacterial biosynthesis of alginates. *Journal of Chemical Technology & Biotechnology*, *85*(6), 752–759. doi:10.1002/jctb.2372

Hugouvieux-Cotte-Pattat, N., Blot, N., & Reverchon, S. (2008). Identification of TogMNAB, an ABC transporter which mediates the uptake of pectic oligomers in Erwinia chrysanthemi 3937: TogMNAB oligogalacturonide transport system. *Molecular Microbiology*, *41*(5), 1113–1123. doi:10.1046/j.1365-2958.2001.02564.x

Hugouvieux-Cotte-Pattat, N., Condemine, G., Nasser, W., & Reverchon, S. (1996). REGULATION OF PECTINOLYSIS IN *ERWINIA CHRYSANTHEMI*. *Annual Review of Microbiology*, *50*(1), 213–257. doi:10.1146/annurev.micro.50.1.213

Hugouvieux-Cotte-Pattat, N., & Robert-Baudouy, J. (1985). Isolation of kdgK-lac and kdgA-lac Gene Fusions in the Phytopathogenic Bacterium Erwinia chrysanthemi. *Microbiology*, *131*(5), 1205–1211. doi:10.1099/00221287-131-5-1205

Hugouvieux-Cotte-Pattat, N., & Robert-Baudouy, J. (1989). Isolation of Erwinia chrysanthemi mutants altered in pectinolytic enzyme production. *Molecular Microbiology*, *3*(11), 1587–1597.

Hutter, C. A. J., Lehner, R., Wirth, C., Condemine, G., Peneff, C., & Schirmer, T. (2014). Structure of the oligogalacturonate-specific KdgM porin. *Acta Crystallographica. Section D, Biological Crystallography*, *70*(Pt 6), 1770–1778. doi:10.1107/S1399004714007147

Itoh, T., Hashimoto, W., Mikami, B., & Murata, K. (2006). Crystal Structure of Unsaturated Glucuronyl Hydrolase Complexed with Substrate: MOLECULAR INSIGHTS INTO ITS CATALYTIC REACTION MECHANISM. *Journal of Biological Chemistry*, *281*(40), 29807–29816. doi:10.1074/jbc.M604975200

John, R. P., Anisha, G. S., Nampoothiri, K. M., & Pandey, A. (2011). Micro and macroalgal biomass: A renewable source for bioethanol. *Bioresource Technology*, *102*(1), 186–193. doi:10.1016/j.biortech.2010.06.139

Jongkees, S. A. K., & Withers, S. G. (2011). Glycoside Cleavage by a New Mechanism in Unsaturated Glucuronyl Hydrolases. *Journal of the American Chemical Society*, *133*(48), 19334–19337. doi:10.1021/ja209067v

Kawai, S., Ohashi, K., Yoshida, S., Fujii, M., Mikami, S., Sato, N., & Murata, K. (2014). Bacterial pyruvate production from alginate, a promising carbon source from marine brown macroalgae. *Journal of Bioscience and Bioengineering*, *117*(3), 269–274. doi:10.1016/j.jbiosc.2013.08.016

Kruger, N. J., & von Schaewen, A. (2003). The oxidative pentose phosphate pathway: structure and organisation. *Current Opinion in Plant Biology*, *6*(3), 236–246. doi:10.1016/S1369-5266(03)00039-6

Langer, G., Cohen, S. X., Lamzin, V. S., & Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols*, *3*(7), 1171–1179. doi:10.1038/nprot.2008.91

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291. doi:10.1107/S0021889892009944

Leslie, A. G. W. (2006). The integration of macromolecular diffraction data. *Acta Crystallographica Section D Biological Crystallography*, *62*(1), 48–57. doi:10.1107/S0907444905039107

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). *Phaser* crystallographic software. *Journal of Applied Crystallography*, *40*(4), 658–674. doi:10.1107/S0021889807021206

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., … Vagin, A. A. (2011). *REFMAC* 5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D Biological Crystallography*, *67*(4), 355–367. doi:10.1107/S0907444911001314

Pastor, J. M., Bernal, V., Salvador, M., Argandoña, M., Vargas, C., Csonka, L., … Cánovas, M. (2013). Role of central metabolism in the osmoadaptation of the halophilic bacterium Chromohalobacter salexigens. *The Journal of Biological Chemistry*, *288*(24), 17769–17781. doi:10.1074/jbc.M113.470567

Patra, T., Koley, H., Ramamurthy, T., Ghose, A. C., & Nandy, R. K. (2012). The Entner-Doudoroff Pathway Is Obligatory for Gluconate Utilization and Contributes to the Pathogenicity of Vibrio cholerae. *Journal of Bacteriology*, *194*(13), 3377–3385. doi:10.1128/JB.06379-11

Percival, E., & McDowell, R. H. (1967). Chemistry and enzymology of marine algal polysaccharides.

PREISS, J., & ASHWELL, G. (1962a). Alginic acid metabolism in bacteria. I. Enzymatic formation of unsaturated oligosac-charides and 4-deoxy-L-erythro-5-hexoseulose uronic acid. *The Journal of Biological Chemistry*, *237*, 309–316.

PREISS, J., & ASHWELL, G. (1962b). Alginic acid metabolism in bacteria. II. The enzymatic reduction of 4-deoxy-L-erythro-5-hexoseulose uronic acid to 2-keto-3-deoxy-D-gluconic acid. *The Journal of Biological Chemistry*, *237*, 317–321.

PREISS, J., & ASHWELL, G. (1963). Polygalacturonic acid metabolism in bacteria. I. Enzymatic formation of 4-deoxy-L-threo-5-hexoseulose uronic acid. *The Journal of Biological Chemistry*, *238*, 1571–1583.

Reverchon, S., Nasser, W., & Robert-Baudouy, J. (1991). Characterization of *kdgR* , a gene of *Erwinia chrysanthemi* that regulates pectin degradation. *Molecular Microbiology*, *5*(9), 2203–2216. doi:10.1111/j.1365-2958.1991.tb02150.x

Ridley, B. L., O'Neill, M. A., & Mohnen, D. (2001). Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*, *57*(6), 929–967.

Rodionov, D. A. (2004). Comparative genomics of the KdgR regulon in Erwinia chrysanthemi 3937 and other gamma-proteobacteria. *Microbiology*, *150*(11), 3571–3590. doi:10.1099/mic.0.27041-0

Rodionov, D. A., Mironov, A. A., Rakhmaninova, A. B., & Gelfand, M. S. (2000). Transcriptional regulation of transport and utilization systems for hexuronides, hexuronates and hexonates in gamma purple bacteria. *Molecular Microbiology*, *38*(4), 673–683.

Romano, A. H., & Conway, T. (1996). Evolution of carbohydrate metabolic pathways. *Research in Microbiology*, *147*(6-7), 448–455.

Selig, M., Xavier, K. B., Santos, H., & Schönheit, P. (1997). Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium Thermotoga. *Archives of Microbiology*, *167*(4), 217–232.

Shevchik, V. E., Condemine, G., Robert-Baudouy, J., & Hugouvieux-Cotte-Pattat, N. (1999). The exopolygalacturonate lyase PelW and the oligogalacturonate lyase Ogl, two cytoplasmic enzymes of pectin catabolism in Erwinia chrysanthemi 3937. *Journal of Bacteriology*, *181*(13), 3912–3919.

Smidsrød, O., & Draget, K. I. (1996). Chemistry and physical properties of alginates. Carbohydr. Eur, 14(6).

Stein, N. (2008). *CHAINSAW* : a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography*, *41*(3), 641–643. doi:10.1107/S0021889808006985

Steiner, R. A., Kalk, K. H., & Dijkstra, B. W. (2002). Anaerobic enzyme*substrate structures provide insight into the reaction mechanism of the copper-dependent quercetin 2,3-dioxygenase. *Proceedings of the National Academy of Sciences*, *99*(26), 16625–16630. doi:10.1073/pnas.262506299

Swan, M. K., Solomons, J. T. G., Beeson, C. C., Hansen, T., Schonheit, P., & Davies, C. (2003). Structural Evidence for a Hydride Transfer Mechanism of Catalysis in Phosphoglucose

Isomerase from Pyrococcus furiosus. *Journal of Biological Chemistry*, *278*(47), 47261–47268. doi:10.1074/jbc.M308603200

Takase, R., Ochiai, A., Mikami, B., Hashimoto, W., & Murata, K. (2010). Molecular identification of unsaturated uronate reductase prerequisite for alginate metabolism in Sphingomonas sp. A1. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1804*(9), 1925–1936. doi:10.1016/j.bbapap.2010.05.010

Takeda, H., Yoneyama, F., Kawai, S., Hashimoto, W., & Murata, K. (2011). Bioethanol production from marine biomass alginate by metabolically engineered bacteria. *Energy & Environmental Science*, *4*(7), 2575. doi:10.1039/c1ee01236c

Thomas, F., Barbeyron, T., & Michel, G. (2011). Evaluation of reference genes for real-time quantitative PCR in the marine flavobacterium Zobellia galactanivorans. *Journal of Microbiological Methods*, *84*(1), 61–66. doi:10.1016/j.mimet.2010.10.016

Thomas, F., Barbeyron, T., Tonon, T., Génicot, S., Czjzek, M., & Michel, G. (2012). Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine Flavobacteriia to their independent transfers to marine Proteobacteria and human gut Bacteroides: Emergence and transfer of alginolytic operons. *Environmental Microbiology*, *14*(9), 2379–2394. doi:10.1111/j.1462-2920.2012.02751.x

Uchimura, K., Miyazaki, M., Nogi, Y., Kobayashi, T., & Horikoshi, K. (2010). Cloning and Sequencing of Alginate Lyase Genes from Deep-Sea Strains of Vibrio and Agarivorans and Characterization of a New Vibrio Enzyme. *Marine Biotechnology*, *12*(5), 526–533. doi:10.1007/s10126-009-9237-7

Vaguine, A. A., Richelle, J., & Wodak, S. J. (1999). *SFCHECK* : a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the

atomic model. *Acta Crystallographica Section D Biological Crystallography*, *55*(1), 191–205. doi:10.1107/S0907444998006684

Wargacki, A. J., Leonard, E., Win, M. N., Regitsky, D. D., Santos, C. N. S., Kim, P. B., … Yoshikuni, Y. (2012). An Engineered Microbial Platform for Direct Biofuel Production from Brown Macroalgae. *Science*, *335*(6066), 308–313. doi:10.1126/science.1214547

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., … Wilson, K. S. (2011). Overview of the *CCP* 4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography*, *67*(4), 235–242. doi:10.1107/S0907444910045749

Yip, V. L., & Withers, S. G. (2006). Breakdown of oligosaccharides by the process of elimination. *Current Opinion in Chemical Biology*, *10*(2), 147–155. doi:10.1016/j.cbpa.2006.02.005

Yoon, J. J., Kim, Y. J., Kim, S. H., Ryu, H. J., Choi, J. Y., Kim, G. S., & Shin, M. K. (2010). Production of Polysaccharides and Corresponding Sugars from Red Seaweed. *Advanced Materials Research*, *93-94*, 463–466. doi:10.4028/www.scientific.net/AMR.93-94.463

Zheng, L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*, *32*(14), e115–e115. doi:10.1093/nar/gnh110