For the Love of Lignin: Six Catabolic Enzymes from S. paucimobilis SYK-6

By

Ann-Marie Illsley

Faculty Advisor: Professor Erika A. Taylor

A dissertation submitted to the faculty of Wesleyan University in partial fulfillment of the requirements for the degree of Master of Arts.

Middletown, Connecticut       May 2013
Acknowledgements

I would very much like to express my deep gratitude to all of those who helped make this thesis a reality. First and foremost of those: Dr. Erika Taylor - thank you for guiding me in research and elsewhere but most of all for sticking with me (and bearing with me) through this process. Your patience, enthusiasm, calm and passion have helped make me the scientist and person that I am today. I would also like to thank the members of the Taylor lab, especially Kevin, Dan, and Ling. We’ve all toiled, celebrated, and spent the early morning hours in Hall-Atwater together.

I would also like to thank the members of my committee for their support and constructive suggestions.

During my tenure in Taylor Lab, I’ve been fortunate to have a number of wonderful roommates from Home Avenue to Fountain Ave. to Court St. Thank you all for eating my baking experiments, laughing and crying with me (and not minding my bizarre hours too much!).

Finally, I want to thank my family, who have been a constant and strong source of love and support. I love you and I certainly couldn’t have ever completed this without you!
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Energy Issues</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Not All Biofuels Are Created Equally</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Lignin: An Untapped Resource</td>
<td>9</td>
</tr>
<tr>
<td>1.4 Lignin Depolymerization</td>
<td>12</td>
</tr>
<tr>
<td>1.5 Catabolism of Lignin Derived Aromatic Compounds</td>
<td>16</td>
</tr>
<tr>
<td>1.7 Amidohydrolases</td>
<td>18</td>
</tr>
<tr>
<td>1.7 Dioxygenases</td>
<td>25</td>
</tr>
<tr>
<td>1.8 References</td>
<td>29</td>
</tr>
<tr>
<td>II. EXPERIMENTARY</td>
<td>30</td>
</tr>
<tr>
<td>2.1 Amidohydrolase Enzymes</td>
<td>31</td>
</tr>
<tr>
<td>2.2 Dioxygenase Enzymes</td>
<td>35</td>
</tr>
<tr>
<td>III. RESULTS and DISCUSSION</td>
<td>44</td>
</tr>
<tr>
<td>3.1 Amidohydrolase Enzymes</td>
<td>45</td>
</tr>
<tr>
<td>3.2 Dioxygenase Enzymes</td>
<td>74</td>
</tr>
<tr>
<td>3.3 Conclusions and Future Directions</td>
<td>107</td>
</tr>
<tr>
<td>3.4 References</td>
<td>109</td>
</tr>
</tbody>
</table>
IV. CONCLUSIONS and FUTURE DIRECTIONS ................. 111

V. APPENDIX ..................................................................... 113
ABSTRACT

Lignin is a complex, aromatic bio-polymer found in virtually all plant life. It forms an irregular matrix that functions as a protective barrier and confers rigidity onto terrestrial plants. With the decreasing availability of fossil fuels and the abundance and carbon-richness of lignin molecules, a strong case for the utilization of lignin catabolism towards biofuel precursor purposes is made.

Lignin is catabolized in nature by both fungi and bacteria. *Pseudomonas paucimobilis* SYK-6 is one such bacterial species that possesses a number of enzymes that catabolize lignin monomers and dimer into small molecule participants of the Kreb’s cycle.

The Taylor lab has been working to characterize a number of these enzymes. In this dissertation, six of these enzymes will be discussed: Lig I, Lig J, Lig W, Lig Y, Des B and Des Z. The expression of soluble protein for each of these is investigated over a number of conditions and some basic characterization is performed. Large scale expression conditions for all four amidohydrolase enzymes were determined and their multimeric state studied. Soluble expression conditions of Des B and Des Z tests were attempted in two separate vectors over a number of conditions. Unfortunately, no large amounts of soluble protein were produced.
I. Introduction

1.1 Energy Issues

1.2 Not All Biofuels Are Created Equally

1.3 Lignin: An Untapped Resource

1.4 Lignin Depolymerization

1.5 Catabolism of Lignin Derived Aromatic Compounds

1.6 Amidohydrolases

1.7 Dioxygenases

1.8 References
1.1 Energy Issues

The majority of the world’s energy needs are currently being met primarily by the burning of fossil fuel products such as petroleum, natural gas, and coal. There is a continual increase in demand from major consumers such as the United States of America as well as a burgeoning need from the developing third world. Given the finite amount of fossil fuels available, the current energy system is quite unfeasible and wholly unsustainable moving into the future.\(^1\)

In 2008, the total fuels consumed were 85,255.339 thousand barrels per day (171.050 quadrillion Btu, QBtu) of petroleum, 111 billion cubic ft. (111 QBtu) of natural gas and 7,345,641 short tons (139 QBtu) of coal.\(^1\) The United States of America alone consumed 21.8 % of global petroleum, 21.4 % of natural gas, and 16.1 % of coal.\(^1\) Consumption has been increasing consistently and can be expected to continue in the absence of major energy use reform.\(^1\)

It is estimated that the current obtainable reserves for fossil fuels are as follows: petroleum - 1,354 billion barrels; natural gas - 6,609 trillion cubic feet; and coal - 930 million short tons.\(^1\) These reserves have been formed in a process that uses intense heat and pressure over millions of years and thus cannot be replenished in any significant way in the
near future. Considering the current consumption levels, it is likely that all obtainable fossil fuel will be used within a lifetime, much faster in fact if the rising consumption trends are considered (Figure 1). However, it is important to note that a more likely scenario would be that reserves will be tapped until the cost rises so high as to be economically unfeasible. Either way, continuing use of fossil fuels at current levels is completely unsustainable looking into the future. Given the combined factors of shrinking stores, rising consumption, and little to no sustainability, it is likely that the U.S.A. and indeed the rest of the world will need adopt to new primary energy sources and soon.
Figure 1 – Projected depletion of major fossil fuels based on total world proven reserves and consumption from 2010. It is likely a conservative estimate given that consumption continues to increase, suggesting that the rate of depletion would not be linear.¹
1.2 Not All Biofuels Are Created Equally

The limited supply and increasing cost of fossil fuels have spurred the development of fuel from new sources. Many alternative energy efforts are focused on biologically derived fuels, or biofuels which have the benefit of renewability on a much smaller timescale. A number of biological sources have been identified and are currently being researched. These include plant sources such as switch grass, canola, sugarcane, corn and soy for the production of ethanol, as well as human and animal derived sources such as manure and biowaste, all of which has gained increasing popularity in the United States of late.³

However, while biofuels may improve on fossil fuel in terms of renewability, the same cannot always be said for their energetic efficiency and the potential environmental effects from large scale implementation would have. With this in mind, recent studies have focused on taking such factors into account. Scharlemann et al. produced a Figure that took this one step further by developing a system that compared biofuels (taking into account biofuel precursors and production) and fossil fuels on a basis of both green-house gas emissions and total environmental impact (Figure 2).³ This comparison was achieved by quantifying multiple factors in the growth, processing and production and usage of a variety of biofuels. For example, green-house gas emission values consider not only biofuel combustion, but
also emissions from things like nitrogen based fertilizers. Environmental impact considers a number of factors such as land and water usage as well as the impact to human and animal ecosystems. This analysis reveals that while many potential biofuels show an improvement in emissions, they often show no change or even an increase in environmental impact.
Figure 2- Comparison of several biofuel precursors based on both greenhouse gas emissions and total environmental impact, standardized to fossil fuels. (modified from Ref. 3).
It is also important to consider the potential negative impact of using food crops as a biofuel feed stock. This is especially true of biofuels created using corn. Corn ethanol has, at best, a small net energy gain, though a study by Ulgiati, S (2001) has suggested that the energetic cost of corn ethanol production exceeds the amount of energy produced by the fuel. Moreover, even if all U.S. corn and soy farming products were used to produce fuel it would only cover a fraction of the nation’s needs. Diverting crops away from consumption as food can also easily drive up food prices as well as have unforeseen secondary effects. United States corn farming, fueled by government subsidies, has been increasing. As corn farming replaces soy (of which the U.S. is lead world producer), soy prices have subsequently been driven higher. In response, deforestation in the Amazon is occurring in Brazil to allow for greater farming of soy (a negative environmental impact).

In the near future, biofuels will likely become a vital part of the energy landscape. As that happens, policy and practice must be established with care. An ideal biofuel must combine a number of positive factors: easy and quick renewability (i.e. short growth), decreased green-house gas emissions and environmental impact, high net energetic gain, limited ancillary negative effects (driving up food prices, or negative secondary environmental effects) and logistical feasibility.
1.3 Lignin: An Untapped Resource

Lignin is a carbon rich, aromatic bio-polymer present in virtually all plant life. It is highly abundant, second only to cellulose, and its degradation is a vital part of carbon recycling. Found primarily in the middle lamella, lignin establishes the physical rigidity of plants by acting as a sort of bio-cement to hold individual cells together. It also functions as a physical barrier against the hydrolysis of layered cellulose (which forms the cell walls of plants) by creating a barrier (Figure 3).

Lignin possesses a highly interconnected and complex multidimensional structure. It is primarily composed of three of hydroxycinnamyl alcohols: p-coumaryl alcohol (no methoxy-substituents), coniferyl alcohol (mono-methoxy substituted) and sinapyl alcohol (di-methoxy substituted). These monolignols undergo a free-radical coupling process to form lignin (Figure 4). Approximately half of the sixteen different bond motifs between the phenylpropanoid units (referred to after incorporation into lignin as p-hydroxyphenyl, guaiacyl, and syringyl units, respectively) (Figure 4) are β–O-4 aryl ethers. Other key linkages are phenyl coumarans, biphenyl ethers, biphenyls, diarylpropane diols and resinols (Figure 5). The nature of the
Figure 3 – An example of the complex, irregular lignin polymer (modified from Ref. 11).
Corresponding phenyl-propanoid units in lignin polymer

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>p-coumaryl alcohol.</td>
<td>p-hydroxyphenyl lignin unit formed from p-coumaryl alcohol.</td>
<td>coniferyl alcohol.</td>
<td>guiacyl lignin unit formed from coniferyl alcohol.</td>
</tr>
<tr>
<td>sinapyl alcohol.</td>
<td>syringyl lignin unit formed from sinapyl alcohol.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4 – Common Monolignols found in Lignin. A.1) p-coumaryl alcohol. A.2.) p-hydroxyphenyl lignin unit formed from p-coumaryl alcohol. B.1.) coniferyl alcohol. B.2.) guiacyl lignin unit formed from coniferyl alcohol. C.1.) sinapyl alcohol. C.2.) syringyl lignin unit formed from sinapyl alcohol.
biosynthesis of lignin reaction is such that, not only are there a multitude of linkages, the a number of stereoisomers are formed, creating the highly irregular, un-patterned structure. Those features also make it very resistant to degradation as well as able to adapt well to varying levels metabolic stress.\textsuperscript{12}

1.4 Lignin Depolymerization

The highly complex and un-patterned structure of lignin makes it fairly resistant to degradation from both chemical and biological sources. However,
there are a number of fungi and bacteria that specifically breakdown lignin.\textsuperscript{7,13-14} White-rot basidiomycetes are one of the most common and well characterized examples of a fungal delignifier and is additionally noteworthy because of its ability to degrade lignin, cellulose and hemicellulose concurrently (called simultaneous rot) or lignin only (called selective delignification).\textsuperscript{8} This is achieved in fungi through a series of extracellular peroxidases and small molecule accessory metabolites. These most frequently include laccases and three distinct, high-redox potential, heme peroxidases: Lignin Peroxidase (LiP), Manganese Peroxidase (MnP), and Versatile Peroxidase (VP).

The first of these, the monomeric LiP has catalytic activity that resemble standard peroxidases.\textsuperscript{9} The heme iron is activated ($\text{Fe}^{\text{III}} \rightarrow \text{Fe}^{\text{IV}}$) by hydrogen peroxide in the active site, which in turn oxidizes two substrates by removing a single electron from each. Though it is secreted, LiP is not able to directly oxidize lignin substrates because it is too large to pass through plant cell walls. It has been suggested that veratryl alcohol may act as an oxometabolite intermediate, with LiP conveying its redox ability on to the smaller molecule which can pass through the plant cell wall and interact directly with the lignin molecule.\textsuperscript{9}

MnP is another ferrous heme-peroxidase responsible for depolymerization of phenolic units in lignin. It also operates similarly to
standard peroxidases but does not possess a catalytic tryptophan residue and thus is incapable of direct substrate interaction (Figure 6). Instead MnP produces a diffusible oxidizer in the form of Mn$^{3+}$, which can more easily permeate rigid lignin structures. VP, the third heme peroxidase, possesses hybrid-like qualities, sharing a structure similar to LiP and a catalytic process similar to both MnP and LiP (Figure 6).
Figure 6 – Lignin Depolymerase Structures. Blue arrows mark heme access channels, green spheres mark Ca\(^{2+}\), Mn ions are shown as van der Waals spheres (lilac in left column, gray in right). A) Full LiP structure (pdb 1LLP). B) Cut away view of LiP heme access channels (modified from Ref. 15). C) Full MnP structure (pdb 1YYD). D) Cut away view of MnP access channels (modified from Ref 15). E) Full VP structure (pdb 2BOQ). F) Cut away view of VP (modified from Ref 15).
1.5 Catabolism of Lignin Derived Aromatic Compounds

Certain bacteria continue the mineralization of lignin by further breaking down depolymerized lignin units into small molecules that the organisms use as a carbon source. Masai et al., have elucidated such a catabolic pathway in the bacterial species Sphingomonas paucimobilis SYK-6, a gram-negative bacteria found in the waste of a kraft pulp mill (Figure 7). They have described separate pathways for the break-down of β-aryl ethers, ferulate and biphenyl units, all converging on vanillate. They also suggest that a number of other lignin components (phenyl coumarane, pinoresinol, diarylpropane and guaiacyl lignin) are also broken down into vanillate during decomposition, thought they have not suggested a specific pathway as yet. All pathways converging on vanillate continue through what has been termed by Masai et al., the Protocatechuate (PCA) 4,5-cleavage pathway, resulting in two participants in the Krebs cycle (pyruvate and oxaloacetate). Syringyl lignin moieties are degraded via multiple pathways that converge within the PCA cleavage pathway directly before or after the decarboxylation of 2-pyrone-4,6-dicarboxylate (PDC). Of the twenty-three or so enzymes identified as relevant to the catabolism of lignin in Sphingomonas paucimobilis SYK-6, eight have been, or are
Figure 7– Lignin degradation pathway in *Sphingomonas paucimobilis* SYK-6 as proposed by Masai et al.\textsuperscript{17a, 19} Amidohydrolases studied by Taylor lab are shown in green, dioxygenases in red, all others are shown in purple.
currently being researched in the Taylor Lab. Of those, my research focuses on six. Four are members of the amidohydrolase superfamily: 2-pyrone-4,6-dicarboxylate hydrolase (Lig I), 4-oxalomesaconate hydratase (Lig J), 5-carboxyvanillate decarboxylase (Lig W), and 2,2′,3-trihydroxy-3′-methoxy-5,5′-dicarboxybiphenyl meta-cleavage product hydrolase (Lig Y). The other two are dioxygenases: gallate dioxygenase (Des B) and 3-O-methylgallate dioxygenase (Des Z).

1.6 Amidohydrolases

The amidohydrolase superfamily consists of over 1000 potential enzyme members and was first characterized based on conserved structural features as described by Holm. Chief among these are an eight strand β-barrel and divalent metal ion(s) complexed in the active site. Amidohydrolases generally have a conserved structure resembling the outer ring of β-sheets in the alternating (β/α)₈ structural motif of TIM-barrels (although some exhibit the inner α-helix part of the ring as well) (Figure 8). This is somewhat unsurprising given that a large portion of enzymes with TIM-barrel structures are metabolic hydrolases. The loops extending over the active site provide a measure of substrate specificity. One of most interesting features of the amidohydrolase superfamily lies in the catalytically relevant active site metal ions coordinated by the C-terminal region of the barrel core. Raushel et al. have published a review describing seven distinct family subtypes based on the diversity of metal ions and conserved loop.
Figure 8 – The amidohydrolase superfamily shows a conserved β-barrel structural motif similar to TIM (α/β)_n barrels. A) Top view of Triosephosphate isomerase (TIM) (pdb 8TIM), a strong example of TIM barrel motif. B) Side view of TIM (pdb 8TIM). C) Top view of Phosphotriesterase (PTE) with dinuclear metal ions displayed (pdb 1ez2). D) Side view of PTE with a substrate analogue, showing only β-sheets and metal ions.
These subtypes include mononuclear and both homo- and hetero- dinuclear metal sites involving divalent Zn, Ni and Fe (and possibly more). The Raushel group has also suggested that there are potentially more subtypes which have yet to be characterized in the literature. These enzymes catalyze a wide range of reactions but primarily hydrolyze substrates with amide and ester groups. Despite the wide variety of substrates, several amidohydrolase superfamily members share basic catalytic similarities. Specifically, for many with dinuclear metal active sites, the β-metal ion is responsible for electrophilically activating carbonyl or phosphoryl groups for attack by a water molecule that has been nucleophilically activated by the α-metal ion (Figure 9). In several cases with mononuclear metal active sites, the metal ion is also responsible for activating the nucleophilic water by general base catalysis. Proton transfer from an active site residue assists in the activation of the moiety that water targets (Figure 10).
Figure 9 – Proposed amide cleavage mechanism of dihydroorotase (DHO), an example of two catalytic metal ions in amidohydrolase. Recreated from the mechanism proposed by Raushel et al.\textsuperscript{21} Starting at the top left and moving clockwise. The entrance of dihydroorotase displaces the water molecule coordinated to the β-metal ion. Interaction between the β-metal ion and the carbonyl of the amide bond is polarizing and makes the carbonyl carbon electrophilic. Abstraction of the proton from the hydroxyl bridged between the two metal ions by the aspartate, activates the hydroxyl for nucleophilic attack of the carbonyl carbon. The amide bond is cleaved when the amide nitrogen abstracts the proton from aspartic acid following reformation of the carbonyl complexed with the α-metal ion. Newly formed carbamoyl aspartate is released and the enzyme regenerated upon the entrance of water molecules.
Figure 10 – Proposed mechanism of adenosine deaminase, an example of a single catalytic metal ion in amidohydrolases. Recreated from the mechanism present by Raushel et al.\textsuperscript{21} Starting at the top left and moving clockwise. The catalytic metal ion is complexed with a water molecule (the protons of which exhibit hydrogen bonding with a histidine residue and aspartate residue that is also complexed with the metal ion). The complexed water molecule is activated for nucleophilic attack by abstraction of a proton by the histidine, causing the adjacent nitrogen to abstract a proton from the nearby glutamatic acid residue. This creates a tetrahedral intermediate that collapses when the aspartate abstracts a proton from the carbinol and the amino-substituent abstracts a proton from the histidine residue and leaves as ammonia. The enzyme is regenerated when the glutamate abstracts a proton from the aspartic acid followed by the entrance of a water molecule.
Lig Y is an amidohydrolase enzyme originally identified by Peng and coworkers in 1999 and is found in the biphenyl catabolic pathway of *Sphingomonas paucimobilis* SYK-6.\textsuperscript{17a} The enzyme immediately hydrolyzes the unstable meta-cleavage product of 2,2′,3-trihydroxy-3′-methoxy-5,5′-dicarboxybiphenyl (OH-DDVA) catalyzed by Lig Z to produce 5-carboxyvanillate (5-CVA) (Figure 7). The hydrolase activity of Lig Y was confirmed by monitoring the incorporation of $\text{H}_2^{18}\text{O}$ into Lig Y produced 5-CVA by GC-MS.

In 2002, Peng and co-workers reported the identification of another amidohydrolase they termed 5-carboxyvanillate decarboxylase (Lig W). Directly following Lig Y in the biphenyl catabolic pathway, Lig W shares a 20 % sequence identity to both Lig J and Lig Y. Lig W catalyzes the decarboxylation of 5-CVA to vanillate. (Figure 7) The decarboxylase activity was confirmed by the observation, by GC-MS, of deuterium uptake from solution. Because this was observed without the presence of cofactors, Lig W was determined to be a *non-oxidative* aromatic decarboxylase.\textsuperscript{24} However, its putative amino acid sequence did not share significant sequence identity to any of the known non-oxidative aromatic decarboxylase families.\textsuperscript{17a} Ishii, Narimatsu et al. noted that Lig W does share 30 % sequence identity to γ-resorcylic acid decarboxylase, which suggests a potential new family of non-oxidative decarboxylases.\textsuperscript{25}

Lig I, or PDC hydrolase was first identified by Kersten et al. in *Pseudomonas testosteroni* and later identified in *Sphingomonas paucimobilis*
SYK-6 by Masai et al.\textsuperscript{26-27} It is a monomeric amidohydrolase that lies on the (PCA) 4,5-cleavage pathway of \textit{Sphingomonas paucimobilis} SYK-6 and catalyzes the reversible conversion of 2-pyrone-4,6-dicarboxylate (PDC) to 4-carboxy-2-hydroxymuconate (CHM), which interconverts with the keto-form tautomer of the product, 4-oxamesaconate (OMA) (Figure 7).\textsuperscript{17a, 28} While the Masai lab had determined basic kinetic properties of both the forward and reverse reactions by spectrophotomically observing the appearance and disappearance of PDC and OMA ($K_{\text{m-PDC}} = 74 \ \mu\text{M}$, $K_{\text{m-OMA}} = 49 \ \mu\text{M}$, $V_{\text{max hydrolysis}} = 506 \ \text{U/mg}$, and $V_{\text{max synthesis}} = 283 \ \text{U/mg}$), a more extensive kinetic analysis was recently performed by the Raushel group. By monitoring the disappearance of PDC at 312 nm, a $k_{\text{cat}}$ value of \textit{342} ± \textit{25} \text{s}^{-1}, a $K_{\text{m}}$ value of \textit{48} ± \textit{11} \mu\text{M} and a $k_{\text{cat}}/K_{\text{m}}$ value of \textit{7.5} ± \textit{1.0} \times 10^{6} \ M^{-1} \text{s}^{-1} were determined for the forward, Lig I catalyzed, hydrolysis of PDC to CHM and OMA.\textsuperscript{28} This was unaffected by the addition of several divalent metals (Mn\textsuperscript{2+}, Zn\textsuperscript{2+}, Co\textsuperscript{2+}, Cu\textsuperscript{2+}, or Ni\textsuperscript{2+}), which the Raushel group have suggested indicates that Lig I is the first member of the amidohydrolase superfamily to not require the metal ions for catalysis. This theory was given further support when the crystal structure of Lig I (and several mutants in which potential divalent metal ion associated residues were altered) with product and without metal ion were solved.\textsuperscript{28} The active site of Lig I was also explored by monitoring enzyme function with thiol reagent incubation. Having determined that this severely limited activity, the Masai group hypothesized the presence of cysteine in the active site. Based on this observation and the $\alpha/\beta$ fold hydrolase active site
characterized by Schrag (amino acid residue motif, Small-X-Nucleophile-X-Small), the Masai group have also suggested the active site of Lig I to be Ala$_{74}$-Ser$_{75}$-Cys$_{76}$-His$_{77}$-Gly$_{78}$.\textsuperscript{29}

4-Oxalomesaconate hydratase, also called Lig J, is a dimeric amidohydrolase that directly follows Lig I in the protocatechuate (PCA) 4,5-cleavage pathway of \textit{Sphingomonas paucimobilis} SYK-6. First described by Kersten et al., Lig J catalyzes the conversion of 4-oxalomesaconate (OMA) to 4-carboxy-4-hydroxy-2-oxoadipate (CHA) (Figure 7).\textsuperscript{17a, 27} OMA exists in equilibrium between its keto- and two enol- forms, but based on the structure of CHA, Masai et al. (who first isolated Lig J in \textit{Sphingomonas paucimobilis} SYK-6) have suggested that the keto-OMA is the likely substrate. Similarly to experiments performed with Lig I, Masai et al. determined basic kinetic values for Lig J by spectrophotometrically observing the disappearance of a peak at 265 nm ($K_m$-OMA of 138 µM, $V_{max}$ of 440 U/mg). They also investigated enzyme function in the presence of thiol reagents and due to Lig J's inactivity hypothesized that a cysteine is involved in catalysis. Lig J shares a 20 % sequence identity with Lig W and an even greater 37 % sequence identity with Lig Y. Masai et al. have suggested that this may indicate a common ancestor between Lig Y and Lig J.\textsuperscript{17b}

\subsection*{1.7 Dioxygenases}

Dioxygenases play a vital role in the bacterial degradation of catecholic compounds. This has made them important research targets for a wide range of topics from catabolic pathways to the dismantling of soil pollutants.\textsuperscript{17a, 30}
While there exists many types of dioxygenase enzymes from multiple superfamilies, all catalyze the addition of $O_2$ to a variety of substrates.

There are two main classes of ring-cleaving dioxygenases: intradiol and extradiol, which can be distinguished by which aromatic bond (in relationship to the two alcohol substituents) is cleaved. Intradiol dioxygenases cleave the bond between two alcohol substituents (ortho), whereas extradiol dioxygenases cleave the bond adjacent to the diol (meta) (Figure 11). Extradiol dioxygenases have a large diversity of substrates and are generally considered to be more versatile than intradiol dioxygenases.

![Figure 11 – Overview of Intradiol and Extradiol Dioxygenase Reactions (Ref. 29)](image-url)
The dioxygenases that participate in the degradation of lignin in *Sphingomonas paucimobilis* SYK-6 are extradiol dioxygenases and, as such, focus will be placed on that class.

Extradiol dioxygenases are found in three different superfamilies. Type I is composed of members of the Vicinal oxygen chelate (VOC) superfamily. Type II dioxygenases are monomeric, homo-multimeric or hetero-multimeric and contain 4,5-protocatechuate dioxygenase (Lig AB), gallate dioxygenase (Des B), and 3-O-methylgallate dioxygenase (Des Z), but has not yet been classified as a single superfamily. Type III is made up of members of the Cupin superfamily.31

Kasai et al. have determined by mutational analysis that the degradation of syringate, via 3-O-methylgallate (3MGA), occurs through more than one path. Identified originally in *Sphingomonas paucimobilis* SYK-6 and then produced in *e. coli* by Kasai et al., Des B is a type II extradiol dioxygenase involved in one of these pathways.32 3MGA is converted to gallate by Lig M after which Des B opens the aromatic ring producing the enol form of OMA which is also the product of the Lig I reaction. Ferric iron was the only metal to fully reinstate activity. The $K_{M\text{-gallate}}$ has been determined to be 67 µM and the $V_{max}$ 43 U/mg by monitoring oxygen consumption.32 In addition to the reaction pathway, Kasai and colleagues also investigated the native state and metal dependency of Des B. Characterization of the native state was performed by gel filtration chromatography and suggested a homodimer. The $Fe^{2+}$ metal dependency of Des B was determined by
incubation with various metals with Des B after it had been stripped by EDTA and monitoring of the oxygen consumption (indicating dioxygenase activity).\textsuperscript{32} A crystallographic study of Des B has been performed by Sugimoto, Yamamoto et al. who have reported obtaining crystals of Des B but have not yet published a structure.\textsuperscript{33}

3-O-Methylgallate dioxygenase (Des Z) is a type II extradiol dioxygenase involved in the other two branches of the syringate degradation part of the pathway.\textsuperscript{17a, 34} Des Z is believed to be responsible for the conversion of 3MGA to both 4-carboxy-2-hydroxy-6-methoxy-6-oxohexa-2,4-dienoate (CHMOD) and PDC (Figure 7). Kasai et al. have defined $K_{M}-{3MGA}$ to be 210 µM and $V_{MAX}$ to be 3.6 U/mg.\textsuperscript{17a, 35}

All six of these enzymes play an important role in the catabolic breakdown of lignin into TCA cycle precursors. Here we show initial characterization Lig I, Lig J, Lig W, Lig Y, Des B and Des Z.
1.8 References


II. Experimental

2.1 Amidohydrolase Enzymes

A. Materials
B. Small Scale Expression Screening of Lig I, Lig J, Lig W and Lig Y
C. Isolation and Purification of Enzymes
D. Size Exclusion Chromatography
E. Sequence Alignments

2.2 Dioxygenase Enzymes

A. Materials
B. Small Scale Expression Screening of Des B and Des Z
C. Cloning of desB and desZ into pET-21b
D. Site-Directed Mutagenesis
E. Small Scale Expression Screening of Des Z with a C-terminal (his)_6-tag
I. Amidohydrolases

A. Materials

Genes of the four amidohydrolase enzymes were synthesized and provided to the Taylor lab by Stephen Almo, Ph.D of Albert Einstein College of Medicine of Yeshiva University (Lig Y (9453a) clone: 9453a2BCt6p1, Lig W (9453b) clone: 9453b1BCt3p1, Lig I (10053d) clone: 10053d1BCt7p1, Lig J (9453d) clone: 9453d2BCt9p1). E. coli cell lines BL-21AI and DH5α were obtained from Invitrogen. Miniprep kits were purchased from Qiagen. Bacterial protein extraction reagent (B-PER) was obtained from Fisher Scientific. All media growths contained 50 µg/L kanamycin.

B. Small Scale Expression Screening of Lig I, Lig J, Lig W, and Lig Y

pTOM-15b vectors containing ligI, ligJ, ligW, ligY were transformed into BL-21AI cells via heat shock and grown on LB-KAN agar plates. Colonies were picked and grown over night at 37 °C with 200 rpm shaking. The cultures were then used to make frozen stocks and inoculate a 5 mL aliquot of LB-KAN which was grown overnight at 37° C. The overnight growth was used to inoculate 25 mL of LB-KAN for expression screening. Inoculated cultures were allowed to grow at 37 °C with shaking until the absorbance at 600 nm was 0.4-0.8. Once confluency had been reached, cultures were split into six aliquots of 4 mL each and induced with IPTG to a final concentration of either 1.0 mM or 0.1 mM IPTG and L-arabinose at 0.2 %, 0.002 %, or 0.0002 % and allowed to continue growing at 37 °C. At 0, 5, and 24 hours,
1 mL timepoints were removed, centrifuged and the media decanted. After all timepoints had been collected, cell pellets were re-suspended in 100 µL of B-PER and then centrifuged for 10 minutes at 13,000 rpm. The supernatant containing the soluble protein fraction was then removed and the pellet (containing the insoluble protein fraction) was re-suspended in another 100 µL of B-PER. An equal amount of loading dye was then added to both soluble and insoluble fractions and all fractions were incubated in a warm water bath for 5 minutes before a final 10 minute centrifugation at 13,000 rpm. Soluble and insoluble fractions were visualized via SDS-PAGE run at 100 V with Bio-Rad’s Precision Plus Protein Standard. Gels were stained in Coomassie Blue Protein Stain (400 mL methanol, 400 mL acetic acid, 500 mL ddH2O, and 50 mg Coomassie Blue Dye), and destained in Coomassie Destain (800 mL methanol, 400 mL acetic acid, 2800 mL ddH2O).

C. Isolation and Purification of Lig I, Lig J, Lig W, and Lig Y
A scratch of frozen stock of each enzyme (Lig I, Lig J, Lig W, and Lig Y) in DH5α competent cells was individually used to inoculate separate 5 mL aliquots of LB-Kan and was grown overnight at 37 °C with shaking. The overnight growth was used to inoculate 2 x 1 L of LB-KAN which was grown until A600 reached between 0.4-0.8. Cultures were then induced and grown under conditions determined by the previous expression screen. After growth had completed, cells were harvested by centrifugation for 10 minutes at 6,000 rpm at 4 °C and re-suspended in approximately 40 mL of Binding Buffer (20 mM Tris-HCl, pH 7.9, containing 5 mM imidazole and 0.5 M NaCl). Cells were broken open by three passes through a French Pressure Cell Press at
16,000 psig and then recentrifuged for 45 minutes at 11,500 rpm to pellet cell debris. The supernatant was then applied to a nickel-charged HisPur Ni-NTA Resin (Thermo Scientific, 40 mL), primed with Binding buffer and charged with 1 column volume (CV) of 0.2 M nickel sulfate. After allowing the protein to enter the column, an additional two CV of Binding Buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 7.5) were added. The column was then washed by 2 CV of Wash Buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 7.5), eluted with 1 CV of Elution Buffer (50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0). Column fractions were collected in 20 mL portions (excepting the elution fractions, which were collected in 4 x 10 mL portions) and used to visualize protein presence via SDS-PAGE as previously mentioned.

**Mass spectrometry** Samples of Lig I, Lig J, Lig W, and Lig Y were sent to the Keck Lab of Yale University for analysis by mass spectrometry. MALDI-TOF was used for Lig I and ESMS was used for the other enzymes. Lig I was the only enzyme for which a successful spectrum was obtained due to protein fall out of solution. The Lig I spectrum is found in the appendix.

**D. Size Exclusion Chromatography**
Fractions showing primarily pure target protein were combined and concentrated using spin concentrators (Millipore 0.5 mL, MWCO 10,000 kDa) and buffer exchanged 3 times into a 20 mM Heps, 150 mM NaCl buffer. A small portion of concentrated protein (Lig I 1st run – 0.087 mg/mL, Lig I 2nd run
- 0.88 mg/mL, Lig J 1st run - 0.087 mg/mL, Lig J 2nd run – 72 mg/mL, Lig W – 12.7 mg/mL, Lig Y 1st run – 0.421 mg/mL, Lig Y 2nd run – 0.21 mg/mL, Lig Y 2nd run reinjection – 0.05 mg/mL), diluted to 0.5 mL was added to an equilibrated Hi Load™ Superdex 200 10/300 column (from GE Life Sciences) connected to an AKTA FPLC system (GE Amersham Pharmacia) with an Amersham GE FRAC-950 Frac950 AKTA Fraction Collector. After the protein had been loaded onto the column, approximately 35 mL of buffer was run through the column (0.5 mL/min).

E. Sequence Alignment

Sequence alignments of Lig I, Lig J, Lig W, Lig Y with other amidohydrolases (subtype samples recommended in Raushel et al. from 2005) were performed using the ClustalW alignment tool on the Biology Workbench of the San Diego Supercomputer Center.

II. Dioxygenases

A. Materials

Because genomic DNA from Sphingomonas paucimobilis SYK-6 or the plasmids containing the genes for DesB and DesZ were unavailable, DNA sequences for each was found on the NCBI databank. An thrombin cleavable (his)_6 tag, and an NdeI endonucleation site were added to the DNA sequence upstream and stop codons and a BamHI endonucleation site were added
downstream of the gene. The modified sequences were submitted to DNA2.0 to be synthesized.

Once the DNA sequences had been synthesized (in a provided pJ241 vector), they were transformed into DH5α cells using the heat shock method and grown on LB-Kan agar plates. An additional vector stock was obtained by miniprepping overnight growths prepared in LB-Kan media. Sam Berman and Julie Huang of Taylor Lab subcloned the genes into pET-15b and pTOM-15b after using a restriction enzyme digest with NdeI and BamHI to liberate the gene. The vectors were then transformed into DH5α and BL-21 competent cells. Miniprep and maxiprep kits were purchased from Qiagen. Bacterial protein extraction reagent (B-PER) was obtained from Fisher Scientific. All media growths contained 100 µg/L ampicillin.

The two plasmids used were originally purchased (or modified from plasmids purchased) from Novagen:

1. pTOM-15b, which encodes ampicillin resistance, an NdeI/BamHI cloning site, T7 promoter and terminator regions, and a N-terminal 10-histidine tag. It is a modified version of pET-15b in which the two serine residues found immediately following and prior to the histidine-tag were mutated to histidine residues by Dr. Toomas Haller of the Gerlt Lab at U. Illinois.
2. pET-21b, which encodes ampicillin resistance, an NdeI/BamHI cloning site, T7 promoter and terminator regions, and a C-terminal 6-histidine tag.
NdeI, BamHI, SAPI, and Quick Ligase enzymes as well as 100 bp and 1 kb DNA ladders were purchased from New England Biosytems (NEB). Supplies for colony PCR (including Taq polymerase) were purchased from Invitrogen. These included T7 promoter and terminator primers.

\[
\text{T7 pro: } \text{taatacgactcactatagg} \\
\text{T7 term: } \text{gctagttattgctcagcgg}
\]

QuikChange™ Site-Directed Mutagenesis Kits were obtained from Agilent Technologies (formerly Stratagene). Primers were designed (using the QuikChange Primer Design program by Agilent) and synthesized by Agilent Technologies.

**B. Small Scale Expression Screening of Des B and Des Z**

Two pTOM-15b vectors, one containing \(\text{desB}\) and one containing \(\text{desZ}\), were transformed into both BL-21 and BL-21AI cells via heat shock and grown on LB-AMP agar plates. Colonies were picked and grown over night at 37 °C with 200 rpm shaking. The cultures were then used to make frozen stocks and inoculate LB-AMP for expression screening.

A scratch of BL-21 frozen stock was used to inoculate 5 mL of LB-AMP which was grown overnight at 37° C. The overnight growth was then used to inoculate 10 mL of LB-AMP for expression screening. Inoculated cultures were allowed to grow at 37 °C with shaking until the absorbance at 600 nm was 0.4-0.8. Once confluency had been reached, cultures were split into two aliquots of 4 mL each and induced with IPTG to a final concentration of either
1.0 mM or 0.1 mM and allowed to continue growing at 37 °C. At 0, 5, and 24 hours, 1 mL timepoints were removed, centrifuged and the media decanted.

A scratch of BL-21Al frozen stock was used to inoculate 5 mL of LB-AMP which was grown overnight at 37° C. The overnight growth was then used to inoculate 3 x 35 mL of LB-AMP with 0.2 mM Fe^{2+} for expression screening. Inoculated cultures were allowed to grow at 37 °C with shaking until the absorbance at 600 nm was 0.4-0.8. Once confluency had been reached, cultures were split into eight aliquots of 4 mL each and induced with IPTG to a final concentration of either 1.0 mM or 0.1 mM IPTG and L-arabinose at 0.2 %, 0.002 %, or 0.0002 % and allowed to continue growing at either 37 °C, 30 °C, or 16 °C. Three 1 mL timepoints were taken from each induction condition (0, 5, 24 hours for 37 °C and 30 °C; 0, 8, 24 hours for 16 °C), centrifuged and the media decanted.

After all timepoints had been collected, cell pellets were re-suspended in 100 µL of B-PER and then centrifuged for 10 minutes at 13,000 rpm. The supernatant containing the soluble protein fraction was then removed and the pellet (containing the insoluble protein fraction) was re-suspended in another 100 µL of B-PER. An equal amount of loading dye was then added to both soluble and insoluble fractions and all fractions were incubated in a warm water bath for 5 minutes before a final 10 minute centrifugation at 13,000 rpm. Soluble and insoluble fractions were visualized via SDS-PAGE run at 100 V with Bio-Rad’s Precision Plus Protein Standard.
C. Large scale purification and isolation

Conditions from the above expression screen that appeared to produce soluble protein were tested on 1L growths of Des Z and Des B (with N-terminal His-tag). Cells were harvested and purified as previously discussed for large scale purification of amidohydrolase enzymes. Additionally, purification was attempted over columns charged with cobalt and zinc.

D. Cloning of desB and desZ into pET-21b

Qiagen Maxiprep and miniprep kits were used to obtains stocks of desB in pTOM-15b. pET-21b vector containing hepl (cloned by Daniel Czyzyk) and des B in pTOM-15b were subjected to digest at their Ndel/BamHI cloning site. Plasmids were incubated with Ndel (~400 ng plasmid/1 µL enzyme) in Buffer 3 at 37 °C overnight. BamHI (~600 ng plasmid/1 µL enzyme) was then added and incubation continued for an additional 3-4 hours. During the last hour 1 µL of SapI was added. DNA fragments from both digests were extracted with 25:24:1 Phenol/Chloroform/Isoamyl alcohol and ethanol precipitated.

Cloning of desB and desZ Into pET-21b individually. Qiagen Maxiprep and miniprep kits were used to obtain separate stocks of desZ and desB in pTOM-15b vectors. Pure (unaltered) pET-21b vector and desB in pET-15b were subjected to digest at their Ndel/BamHI cloning site. Plasmids were incubated with Ndel (~400 ng plasmid/1 µL enzyme) in Buffer 4 at 37 °C overnight. BamHI (~600 ng plasmid/1 µL enzyme) was then added and
incubation continued for an additional 3-4 hours. Digestion products were separated by size with an EtBr agarose gel run at 100 V and imaged with a U:Genius gel imager from Syngene. Following confirmation of expected fragment sizes another digestion was performed under the same conditions. Digestions of the target genes were performed at 10x excess to target vector digestions. DNA fragments from both pairs of digest (desB and desZ focused) were extracted with 25:24:1 Phenol/Chloroform/Isoamyl alcohol and ethanol precipitated.

**Ligation into pET-21b.** Extracted fragments from both pairs of digestions were ligated using the Quick Ligase protocol from NEB with the following modifications. Two ligation reactions were set up for each. In the first, 50 µL (~200 ng/mL) of target vector DNA (fragments from pET-21b digestions) were incubated at room temperature with 100 µL (~2000 µg/mL) of target gene DNA (fragments from desB and desZ digestions at a 10x excess), 10 µL Quick ligase 2x buffer, and 1 µL of Quick Ligase enzyme. The second reaction incubated 40 µL of both vector and gene DNA with 10 µL of Quick ligase at room temperature. Ligation reaction I product (10 µL) was transformed via heat shock into DH5α cells twice, at 3 hours and at 24 hours. Ligation reaction II product (30 µL) was transformed via heat shock into DH5α cells three times, at either 1 hour, 3 hours, and at 24 hours. Transformation reactions were plated on LB-AMP agar plates and grown overnight at 37 °C. Colonies were chosen, grown in LB-AMP and plasmids harvested by miniprep.
Because multiple ligation products were possible, colony PCR was performed on harvested plasmids to screen for products of the correct size. Reactions were performed in triplicate and contained 1x standard buffer, 200 µM dNTP, 0.2 µM of both T7 pro and term primers, 0.75 U of Taq DNA polymerase, between 1 ng and 1 µg of ligation product and ddH2O to a final size of 25 µL. An Applied Biosystems thermal cycler was used. PCR products were separated by size with an EtBr agarose gel run at 100 V and imaged by a U:Genius gel imager from Syngene.

E. Site-Directed Mutagenesis

Mutation of two adjacent stop codons to serine codons was performed using Agilent Technologies QuikChange Lightning Site-Directed Mutagenesis Kit. Three sets of primers were designed using Agilent’s online QuikChange Primer Design Program (desired mutations are shown in capitals).

Des B – Forward Primer #1:
attcggtagccgcaacagtCTtCTgatccgaattcgagctccgtc

Des B – Forward Primer #2:
gtagccgcaacagtCCtCCgatccgaattcgagctccgtc

Des Z – Forward Primer:
sgggttcgtctgctggcgatCTtCTgatccgaattccgagctccgtc

The QuikChange mutagenesis procedure recommended by Agilent was used with the following modifications. In addition to the recommended 60 °C annealing temperature, reactions were attempted with annealing...
temperatures at 50 °C and 55 °C. The annealing time was also increased to 30 s, and the extension time was increased to 4.5 minutes. Furthermore, due to the higher presence of parent DNA in early attempts, the amount of DpnI added was quadrupled and the digest time increased from 5 minutes to an hour. Mutation reaction products were transformed into XL10-Gold ultracompetent cells according the procedure provided and plated on LB-AMP plates. Colonies were chosen and grown in LB-AMP overnight at 37 °C. Plasmids were harvested from the cultures using a Qiagen Miniprep kit and sent for sequencing to ACGT, Inc.

F. Small Scale Expression Screening of Des Z with a C-terminal (his)₆-tag

Plasmids containing Des Z in the mutated pET-21b vector were transformed via heat shock into DH5α cells (for future stock of plasmid) as well as both BL-21 and BL-21AI (for expression screens) and frozen stocks made of each. The expression screen was performed as previously described for N-terminal His-tag Expression Screen, but with the following expansion of conditions. Protein expression in BL-21 cells was investigated at three temperatures (37 °C, 30 °C, and 16 °C), with three metal conditions in the growth media (no metal, 0.2 mM Fe²⁺, and 0.2 mM Mn²⁺), and at two concentrations of IPTG (1 mM and 0.1 mM). Protein expression in BL-21AI cells was investigated with all of the above conditions as well as L-arabinose concentrations of 0.2 %, 0.02 %, 0.002 %, 0.0002 %. Timepoints were taken at 0, 5, and 24 hours for screens at 37 °C and 30 °C, and at 0, 8, 24 hours for
screens at 16 °C. The soluble and insoluble protein fractions were separated and visualized as previously described.

*Large Scale Purification of Des Z with C-terminal His-tag.* Conditions from the above expression screen that appeared to produce soluble protein were tested on 1 L growths of Des Z (with C-terminal His-tag). Cells were harvested and purified as previously discussed for large scale purification of amidohydrolase enzymes.
IV. Results, Discussion, and Conclusions

4.1 Amidohydrolase Enzymes

A. Lig I
   i. Expression Screenings
   ii. Isolation and Purification
   iii. Characterization
   iv. Conclusion

B. Lig J
   i. Expression Screenings
   ii. Isolation and Purification
   iii. Characterization
   iv. Conclusion

C. Lig W
   i. Expression Screenings
   ii. Isolation and Purification
   iii. Characterization
   iv. Conclusion

D. Lig Y
   i. Expression Screenings
   ii. Isolation and Purification
   iii. Characterization
   iv. Conclusion

4.2 Dioxygenase Enzymes

A. Des B
   i. Expression Screenings with N-terminal His-tag
   ii. Creating a C-terminal His-tag through Genetic Manipulations
   iii. Conclusion

B. Des Z
   i. Expression Screenings with N-terminal His-tag
   ii. Creating a C-terminal His-tag through Genetic Manipulations
   iii. Expression Screening with C-terminal His-tag
   iv. Conclusion
4.1 Amidohydrolase Enzymes

The genes *ligI*, *ligJ*, *ligW*, and *ligY* which produce the amidohydrolase enzymes Lig I, Lig J, Lig W, and Lig Y, respectively, were given to Taylor lab by the Almo lab of Albert Einstein College of Medicine. These enzymes had been minimally characterized previously by the Masai lab\textsuperscript{23-24, 32, 34} and we sought to both confirm those previous discoveries and add to our own knowledge of these amidohydrolases both as individual enzymes and as members of the larger lignin catabolizing pathway.

A. Lig I

i. Expression Screenings

The first challenge undertaken with these enzymes was to determine conditions for soluble protein expression which could be used to obtain large amounts of protein for chemical and structural analysis. This process begins on the small scale to allow multiple conditions to be tested simultaneously. Growths of *E. coli* cells, optimized for protein expression, into which the desired gene has been cloned are aliquoted into 4 mL portions after reaching confluency. Each aliquot is induced with a different set of conditions and timepoints are taken. Depending on the type of competent cell used, chemicals are added to induce high transcription levels of the target gene. Each timepoint is processed to separate the soluble and insoluble protein and visualized using SDS-PAGE. The resulting gels can be analyzed to determine the best conditions by the appearance of soluble, time-dependent growth at the expected molecular weight.
Figure 12 displays the SDS-PAGE visualized results of Lig I induced at 37 °C in a metal-less LB-media solution and varied IPTG and L-arabinose concentrations. The top row of gels shows growths induced with 1.0 mM IPTG and decreasing concentrations of L-arabinose (0.2 %, 0.02 %, and 0.002 % moving left to right). The soluble and insoluble fractions for each condition are placed next to each other for easy comparison. Molecular weight of protein bands in each lane can be ascertained by comparing them to the ladder (L), on which the standards have been labeled (in kDa). Optimal growth conditions are determined by looking for time-dependent growth within soluble lanes. Growth progression through time is shown by the increase in protein across time points.

Figure 12 – Gel results of the small expression screening of Lig I (33,320 Da) from BL-21AI cells at 37 °C with varied IPTG and L-arabinose. Optimal condition circled in orange.
Lig I has an expected molecular weight of 33,320 Da. Therefore, soluble, fully denatured Lig I should appear in a band just below the 37 kDa band of the ladder. Lig I showed especially high amounts of soluble protein expression in a number of different conditions. Of those, inducing with 0.1 mM IPTG and 0.2 % L-arabinose followed by 24 hours of growth at 37 °C was chosen for future growths because of the significant increase of soluble protein from the 0 hour time point to the 24 hour time point, compared with minimal increase in the insoluble fraction (Figure 12 circled).

**ii. Isolation and Purification**

Once potential expression screening conditions have been determined on the small scale, expression is scaled up to growths of 1 or 2 liters (in order to obtain greater amounts of protein). Because the soluble-insoluble extraction process used in small scale expression screens denatures protein (thus making it unusable for many future tests), an alternative process is used. Cells are broken open and a soluble protein fraction obtained using a French Press Pressure Cell and centrifugation. Lig I can then be isolated from the other soluble proteins using affinity column chromatography.

Lig I has a C-terminal (his)$_6$-tag that allows for its isolation on an affinity column charged with nickel ions. When the total soluble protein fraction is run over the column, much of the cellular protein passes straight through the column but a large portion of the his-tagged Lig I protein (and some of the cellular protein) are attracted to the nickel ions and remain on the column. A wash buffer is then applied to the column to wash away and any untagged
cellular proteins. Finally an elution buffer with a high concentration of imidazole is applied to the column which out-competes the histidine-nickel interaction, leading to the elution of Lig I off of the column. Column fractions are collected along with samples from the total protein fraction and the insoluble protein pellet and visualized by SDS-PAGE. Just as with expression screening gels, a band just below the 37 kDa band on the ladder is identified as the target protein. From the fractions containing Lig I, the purest are selected.

Fractions collected from an affinity column charged with nickel show that Lig I maintained its high level of expression in a larger growth (Figure 13). As previously observed in the small scale condition screening, there does seem to be some insoluble Lig I that is produced with these conditions, as is observed in examination of the lane containing the pellet. As demonstrated by the appearance large bands near the 37 kDa band of the ladder (L), Lig I is present in strong concentration in all wash fractions and the first three elution fraction (Figure 13). Of those soluble fractions containing Lig I, four were selected as sufficiently pure (wash fraction 4 and elution fractions 1-3), as determined by the lack of bands above or below the Lig I band (indicating proteins of other molecular weights).
Figure 13 – Nickel affinity column fractions showing purification of Lig I enzyme from cell lysate. Pure fractions are enclosed in black.

Collected fractions are then concentrated by centrifugation of concentrators with a size based filter, buffer exchanged into a HEPES solution and stored for future characterization.

iii. Characterization
Initially, a kinetic analysis of Lig I was planned. However, as Lig I substrates were not commercially available, focus turned towards characterization of the protein through other means. Towards that end, the multimeric state of Lig I was investigated using size exclusion chromatography (SEC). This method separates proteins by size and does not denature them (as SDS-PAGE does), and therefore it is possible to detect the natural quaternary state of a protein. The expected molecular weight of a single Lig I protein is 33,320 Da and thus if Lig I exists as a monomer it would appear between the ovalbumin (44 kDa) and myoglobin (17 kDa) peaks, between 15 and 17 mL. If Lig I exists as a dimer, its peak would fall near or over the BSA peak (66 kDa) at 14 mL. Two injections were of Lig I were performed. The first (Figure 14, green line) showed a single peak between 15 and 16 mL. When Figure 14 – Size exclusion column results for Lig I, showing UV absorption (mAU, y-axis) of Lig I solution as it exits the column (mL, x-axis). The blue line is a standard
comprising several protein markers (thyroglobulin 670 kDa – 9-11 mL, γ-globulin 158 kDa – 13 mL, ovalbumin 44 kDa – 15 mL, myoglobin 17 kDa – approx. 17 mL, and vitamin B₁₂ 1.35 kDa -20 mL). The red line is a bovine serum albumin (BSA standard) 66 kDa – 14 mL. The green line shows the first injection of Lig I (16 mL and 25-27 mL) and the purple line is the second injection of Lig I (15-16 mL, 20 mL, 25-26 mL).

overlaid with protein standards, it falls between myoglobin and ovalbumin (17 kDa and 44 kDa, respectively). This suggests a Lig I monomer (33.32 kDa), which supports the results reported by the Masai lab. However, because of the low concentration of the first run (Lig I will not stay in wash or elution buffers at high concentrations for more than a day or two, and must be stored or used quickly), another run was performed on Lig I purified just beforehand. As a result, the second column run had a higher concentration of protein and showed larger peaks. However, it also presented a more complex spectrum. (Figure 14, purple line) It showed a larger peak that stretched roughly between 15 and 16 mL and given the uneven top, could possibly be two conflated peaks. The right side partial peak resembles the monomer peak from the first run. The left portion of the dual peak lies closer to ovalbumin than BSA, indicating that the monomer and dimer do not exist independently in the same solution. What could be possible, however, is a concentration based equilibrium between the two states. If Lig I proteins were interchanging between monomers and dimers, the rate of travel through the
column would be altered and a peak between the two expected states could appear.

The characterization of Lig I was continued by examining its possible amidohydrolase subtype. Raushel et al. describe a series of subtypes within the amidohydrolase superfamily.21 These were defined by conserved residues within the loop regions and determine the type of divalent cation(s) present in the catalytic core. (Figure 8) To investigate the possible subtype of Lig I, it was aligned with each subtype sample and examined for conserved residues. Lig I shared the most similarity of highlighted residues with phosphotriesterase (PTE), which is part of Subtype I (Figure 15). The residues indicated by Raushel et al. were highlighted as well as residues in Lig I thought to be related.
iv. Conclusions

Lig I, which hydrolyses PDC to create 4-OMA in the lignin degradation pathway, was found to solubly express in large amount under several conditions. Out of these, Lig I grown in BL-21AI cells at 37 °C in LB-media with no metal and induced with 0.1 mM IPTG and 0.2 % L-arabinose followed...
by 24 hours of growth was determined to be the optimal expression conditions. Lig I was then successfully purified with affinity column chromatography. When characterized by size exclusion chromatography, Lig I displayed two conflated peaks, one of which suggests the presence of a monomer. The other peak falls in between the expected locations of a monomer and dimer, suggesting that Lig I may interconvert between the two. The amidohydrolase subtype of Lig I was also examined by aligning its sequence with known members of the other subtypes. Lig I showed the greatest similarity to Subtype I. However, given the recent publication by the Raushel group which suggests that Lig I does not have divalent metal ions in the active site, the enzymes subtype classification may change.

B. Lig J

i. Expression Screenings

Lig J was screened for an expression condition that would provide soluble protein using the same process and conditions as described for Lig I in the previous section. Figure 15 shows the SDS-PAGE visualized results of Lig J induced in a LB-media solution not supplemented metal ion and with varied IPTG and L-arabinose concentrations. Gels from growths induced with 1.0 mM IPTG are shown on the left side under the dark green column while growths induced with 0.1 mM IPTG are shown on the right below the light green column. Within each IPTG concentrations, L-arabinose concentration is decreased from left to right (0.2 %, 0.002 %, and 0.0002 %). The soluble fractions of each condition are shown in the top row with the corresponding insoluble fractions directly below. Just as with Lig I, the condition resulting in the greatest time-dependent growth determines the optimal expression condition for Lig J.

Lig J has an expected molecular weight of 38,171 Da and thus should appear as a band at the same height as the 37 kDa band on the ladder. Lig J does show increasing time-dependent growth in several conditions, between in the 24 hour timepoints. However, inducing with 0.1 mM IPTG and 0.2 % L-arabinose followed by growth for 24 hours at 37 °C was determined to be the optimal condition (Figure 16 circled). This was because increasing
Figure 16 – Gel results of the small scale screening of Lig J from BL-21Al cells at 37 °C with varied IPTG and L-arabinose. Optimal condition circled in black.

growth of Lig J could be seen over time (as evidenced by a band in the 5 hour timepoint) greater than that of the other *E. coli* proteins in the same lanes.

**ii. Isolation and Purification**

Once expression conditions have been determined on a small scale for Lig J, growth size and expression is performed on a larger scale in order to obtain protein for characterization. Cells are broken open using a French Press Pressure Cell and centrifuged. Because Lig J also has a (his)$_6$-tag, it can be isolated from the cellular extract using affinity chromatography on a column charged with nickel ions, the same alternative process used to isolate and purify Lig I (described in section 4.1.A.ii). The fractions collected from the column are then visualized using SDS-PAGE. As with the small-scale expression screening, fractions containing Lig J should display a band near
the 37 kDa band of the ladder. Using the gel as a guide, only pure Lig J contacting fractions are selected. Figure 17 shows the SDS-PAGE visualized column fractions along with the crude cell lysate and the pellet. The presence of a distinct band near the 37 kDa band of the ladder indicates that Lig J is present in several fractions (B4, W1-W4, and E1-E2). However, while the fourth binding buffer fraction and first and second wash fractions have Lig J in greater amounts (larger bands), the presence of additional bands above and below indicate that other proteins with different molecular weights are present.

![Figure 17– Nickel affinity column fractions for showing purification of the Lig I enzyme from cell lysate. Pure fractions are circled in black.](image)

Therefore, only wash fractions 3 and 4 and elution fractions 1 and 2 were collected and concentrated to use for further study of Lig J. Pure fractions were concentrated using centrifugation and stored for further characterization.

**iii. Characterization**

The multimeric state of Lig J was investigated using size exclusion chromatography (SEC). SEC separates proteins (and other macromolecules) by size but doesn’t denature proteins (a necessary part of SDS-PAGE), which
allows the multimeric state of the enzyme to be examined. Because the expected molecular weight of Lig J is 38,171 Da, a monomer would show a peak between myoglobin (17 kDa) and ovalbumin (44 kDa) peaks, between 17 mL and 15 mL.

Figure 18 compares two injections of Lig J on the column to several standards (ranging from 1.35 – 670,000 Da) by plotting the UV absorption of the eluent as it exits the column against the volume of the eluent. Due to low concentrations of pure Lig J in column fractions and Lig J’s instability in wash and elution buffers (another similarity to Lig I), both column runs displayed only small peaks (in comparison to the standards). However, magnification of the spectrum demonstrates similarities of the two runs. In both, there is a peak at 8 mL, which is approximately the void volume, suggesting the protein may be aggregating (and thus falling out of solution). Both runs also show a peak at 12 mL, falling on the left leg of the γ-globulin peak (158 kDa). This suggests a possible pentamer protein with an expected molecular weight of 190,855 Da. A smaller peak is visible in both the first and second runs between 15 and 16 mL, (falling between myoglobin and ovalbumin standards - 17 kDa and 44 kDa, respectively), suggesting that
Figure 18 – A. Size exclusion column results for Lig J, showing UV absorption (mAU, y-axis) of Lig J solution as it exits the column (mL, x-axis). The blue line is a standard comprising several protein markers (thyroglobulin 670 kDa – 9-11 mL, γ-globulin 158 kDa – 13 mL, ovalbumin 44 kDa – 15 mL, myoglobin 17 kDa – approx. 17 mL, and vitamin B$_{12}$ 1.35 kDa -20 mL). The red line is a bovine serum albumin (BSA standard) 66 kDa – 14 mL. The green line shows the first injection of Lig J (8 mL, 12 mL and 19-20 mL) and the purple line is the second injection of Lig J (8.5 mL, 11 mL, 12 mL, 15.5-16.5 mL, 22 mL). B. A magnification of plot shown in A. focused on y-values between 0 and 100 to better show the peaks of Lig J.
some Lig J is present as a monomer. The first run of Lig J also shows a large peak between 19 and 20 mL, which falls in between myoglobin (17 kDa) and Vitamin B\textsubscript{12} (1.35 kDa), suggesting a possible contaminant as it is smaller than the expected molecular weight of Lig J. It is possible that the sample of Lig J used for both runs was falling out of solution and thus showing a greater insoluble aggregate population, leading to the presence of peaks at larger molecular weights. Given that Lig J has been previously characterized as a dimer, further study of its multimeric state would be needed to make a definitive characterization.\textsuperscript{17b}

The subtype of Lig J within the amidohydrolase family was also investigated. Following attempted alignments with the sample enzymes provided by Raushel et al, Phosphotriesterase Homology Protein (PHP), a member of subtype II appears to have the closest relation (Figure 19).
iv. Conclusion

Lig J, which catalyzes the conversion of 4-OMA into 4-carboxy-4-hydroxy-2-oxoadipate (CHA) in the lignin catabolism pathway was found to solubly express when grown in BL-21AI cells at 37 °C in LB-media with no metal and induced with 0.1 mM IPTG and 0.2 % L-arabinose followed by 24 hours of
growth was determined to be the optimal expression conditions. Using an affinity chromatography column, Lig J was purified. Size exclusion chromatography was used to characterize the multimeric state of the enzyme. When characterized by size exclusion chromatography, Lig J displayed two major peaks, one indicating a monomer and the other a possible pentamer. However, low and decreasing concentrations of the protein and the appearance of several peaks indicating larger groupings of protein suggests that the protein may have been aggregating and falling out of solution. The amidohydrolase subtype of Lig J was also examined by aligning its sequence with known members of the other subtypes. Lig I showed the greatest similarity to Subtype II.
C. Lig W

i. Expression Screenings

Lig W was also screened for soluble expression using the same conditions and process as described for Lig I. Figure 20 shows the results of each growth condition, separated into soluble and insoluble fractions (top and bottom rows respectively) and visualized with SDS-PAGE. Conditions for the small scale expression of Lig W in Figure 20 are organized similarly to Lig J in Figure 16. Growths induced with 1.0 mM IPTG are shown on the left below the dark green rectangle and growths induced with 0.1 mM IPTG are shown on the right below the light green rectangle. Within each concentration variation of IPTG, L-arabinose is decreased from left right (0.2 %, 0.002 %, and 0.0002 %).

Figure 20 – Gel results of the small expression screening of Lig W from BL-21Al cells at 37 °C with varied IPTG and L-arabinose. Optimal conditions are circled in orange.
The ladder used has a band at 37 kDa. Because Lig W has an expected molecular weight of 39,072.5 Da, it would be expected to appear as a band close to the 37 kDa band in the ladder. In the gel visualized results, the band appearing in line to slightly below the 37 kDa band of the ladder was determined to be Lig W because of it showed time-dependent growth distinctive of other *E. coli* proteins shown. Lig W displayed very little to no insoluble growth. However, slightly greater growth (increasing sizably between 0 and 5 hours after induction) can be seen in some of the soluble fractions (1.0 mM IPTG and 0.2 % L-arabinose, 0.1 mM IPTG and 0.2 % L-arabinose, as well as 0.1 mM IPTG and 0.002 % L-arabinose). Of the conditions tested, 0.1 mM IPTG and 0.2 % L-arabinose followed by 24 hours of growth were chosen. (Figure 20 circled).

**ii. Isolation and Purification**

Expression is scaled up from a smaller scale once expression conditions are determined, so that larger amounts of protein may be obtained. Lig W also has a C-terminal (his)6-tag which allows it to be purified from cell lysate using an affinity chromatography column charged with nickel ions as described for Lig I in section 4.1.A.ii.

Figure 21 shows the gel visualized fractions of affinity column purified Lig W. The presence of bands at the same height as the 37 kDa band of the ladder in both the crude and to a lesser extent the pellet fractions suggests that Lig W was expressed and is present in both the soluble and insoluble
fractions, though not in large amounts. While only a small amount of pure protein appears in the third and fourth wash fractions, the thinness and faintness of bands at all molecular weights in the pellet and crude suggests that the entire growth was not particularly dense. Fractions W3-W4 were collected and concentrated using centrifugation for further characterization of Lig W.

**iii. Characterization**

Lig W was further characterized by investigating the enzyme’s multimeric state using size exclusion chromatography (SEC). The expected molecular weight of a single Lig W protein is 39072.5 Da. This means that if Lig W’s native state is monomeric, a peak would be expected between the peaks of the myoglobin marker (17 kDa) and the ovalbumin marker (44 kDa) which appear at 17 mL and 15 mL, respectively. The single column run of Lig W produced a single, fairly strong peak presenting between the myoglobin (17 kDa) and ovalbumin (44 kDa) standards suggesting that Lig W exists natively as a monomer.
Subtype (within the amidohydrolase superfamily) determination was also attempted for Lig W. Clustal W alignments of Lig W with protein from the seven subtypes described by Raushel were performed. However, no successful alignments were made. Raushel has suggested (in personal communication with Professor Taylor) that recent research by his lab indicates the existence of additional subtypes. It is possible that Lig W belongs to one of those subtypes.

iv. Conclusion

Lig W, which catalyzes the conversion of 5CVA to vanillate in the lignin degradation pathway, expresses solubly when grown in BL-21AI cells at
37 °C in LB-media with no metal and induced with 0.1 mM IPTG and 0.2 % L-arabinose followed by 24 hours of growth. It was successfully purified with affinity column chromatography on a column charged with nickel ions. When characterized by size exclusion chromatography, Lig W displayed a single peak in the location consistent with the size of a single Lig W protein, suggesting that Lig W exists as a stable monomer. The amidohydrolase subtype of Lig W was also examined by aligning its sequence with known members of the other subtypes. However, no matching subtype was found, suggesting that Lig W may belong to a subtype that has yet to be defined.
D. Lig Y

i. Expression Screenings

Just as with the other three amidohydrolase enzymes, Lig Y was also screened for an optimal expression condition. Figure 23 displays the SDS-PAGE visualized results of that screening. The left side of the Figure, below the green rectangle shows growths induced with 1.0 mM IPTG while growths induced with 0.1 mM IPTG are shown on the right below the lighter green rectangle. L-arabinose is decreased (0.2 %, 0.002 %, and 0.0002 % moving from left to right) in both IPTG concentrations. The soluble fractions for each condition are shown in the top row of gels while the corresponding insoluble fractions are shown in the lower row.

Figure 23 – Gel results of the small expression screening of Lig Y from BL-21AI cells at 37 °C with varied IPTG and L-arabinose. Optimal condition circled in black.
Lig Y has an expected molecular weight of 38,491.7 Da, and therefore would be expected to display a band near the 37 kDa band of the ladder. A large amount of time-dependent protein growth can be seen across multiple conditions in the insoluble fractions while the soluble fractions showed a more modest amount of Lig Y. Of those, 0.1 mM IPTG and 0.2 % L-arabinose followed by 24 hours of growth were chosen as the preferable inductions conditions as it showed the largest amount of growth, both soluble and otherwise. (Figure 23 circled)

**ii. Isolation and Purification**

As with the other three amidohydrolases, determining expression conditions on a small scale allows expression to be scaled up so that greater amounts of protein can be obtained. Lig Y is purified using the same process as described for Lig I, J and W. Cells from a larger growth (induced to promote Lig Y expression) are harvested, broken open with a French Press Pressure Cell, and centrifuged to separate the soluble and insoluble fractions. Lig Y also possess a C-terminal (his)_6-tag and therefore is purified using an affinity chromatography column charged with nickel ions.

Figure 24 shows the SDS-PAGE visualized fractions of expressed Lig Y collected from the column. The crude cell lysate and pellet expressed can be observed, whether it is soluble or insoluble. Similarly to Lig W, the column fractions show very little presence of Lig Y, demonstrated by thin faint bands.
Figure 24 – Nickel affinity column fractions showing purification of Lig Y enzyme from cell lysate. Pure fractions are enclosed in black.

near the 37 kDa band of the ladder. However, neither the pellet nor the crude cell lysate show large amounts of Lig Y or any other E. coli proteins, suggesting that while Lig Y was expressed, there was less cellular growth overall. Though faint, Lig Y shows a band at the expected location in several fractions (B3-4, W1-4, and E1-2). Because isolation of Lig Y is desired, fractions with bands at other molecular weights (i.e. containing other proteins) are not selected. Wash fractions 1-4 and elution fractions 1 and 2 (Figure 24, circled in black) were selected and concentrated via centrifugation.

**iii. Characterization**

As with the other three amidohydrolase enzymes, the multimeric state of Lig Y was examined by size exclusion chromatography (SEC). SEC allows for the separation of macromolecules by size and does not denature enzymes, allowing the quaternary structure to be determined. Lig Y has an expected molecular weight of 38,491.7 Da, meaning that a Lig Y monomer would appear as a peak between the myoglobin (17 kDa) and ovalbumin (44 kDa) standards, which show peaks at 17 mL and 15 mL respectively. If Lig Y existed as a larger multimeric protein, such as a dimer, with an
expected molecular weight of 76,983.4 Da, a peak would likely display to the left of the BSA standard (66 kDa) at 14 mL.

Three separate runs of the column were performed for purified for Lig Y (Figure 25). The first two injections were from the same purification. Both showed peaks at 8 mL, 13 mL (the largest), and 15 mL, though the second injection was at a lesser concentration. Given that 8 mL falls within the range of the void volume, that peak is likely the aggregated protein. On the second injection, the same peaks reappeared (at a lesser concentration). The peak at 13 mL falls just to the left of the γ-globulin (158 kDa) peak indicating a possible pentamer (expected molecular weight for a Lig Y pentamer would be 192458.5 Da). This is the larger peak, suggesting that it may be the primary state that Lig Y exists in. The peak at 15 mL falls to the left of ovalbumin (44 kDa), ruling out the presence of a monomer. However, dimeric Lig Y (76983.4 Da) would like appear in line with or slightly to the left of the BSA peak (66 kDa) at 14.5 mL. This suggests that the peak at 15 mL may represent a Lig Y population that is transitioning between two states, possibly monomer and dimer. To further examine this, a fractions corresponding to the 13 mL peak of the 2nd run were concentrated and reinjected. The reinjected sample was very dilute, despite efforts to increase its concentration, but close magnification of the spectrum shows a similar peak at 13 mL. However, given that the concentrations of Lig Y used were low, especially in the second run and the reinjection, further study would
Figure 25 – A. Size exclusion column results for Lig Y, showing UV absorption (mAU, y-axis) of Lig Y solution as it exits the column (mL, x-axis). The blue line is a standard comprising several protein markers (thyroglobulin 670 kDa – 9-11 mL, γ-globulin 158 kDa – 13 mL, ovalbumin 44 kDa – 15 mL, myoglobin 17 kDa – approx. 17 mL, and vitamin B₁₂ 1.35 kDa -20 mL). The red line is a bovine serum albumin (BSA) standard 66 kDa – 14 mL. The green line shows the first injection of Lig Y (8 mL, 12 mL and 15 mL) and the purple line is the second injection of Lig Y (8 mL, 12 mL, 15 mL). The lighter blue line shows the reinjection of the 2nd run (12 mL). B. A magnification of plot shown in A. focused on y-values between 0 and 60 mAU. C. A magnification of A. focused on y-values between 0 and 6 mAU.
be needed in order to be conclusive.

Further characterization of Lig Y was attempted by examining its potential amidohydrolase subtype. When the sequence of Lig Y was compared to enzymes from other subtypes, no alignments showed particular similarity of conserved loop residues. However, an alignment of Lig Y with Lig J (which share 30 % sequence identity) did display similarity between the

LIG_J
LIG_Y

MSLLMIIDCHGHYTV---LPKAHDEWREQQKAAFKAGQPAPPYEPISDDEIRETIEANQL
---MSLLMIIDCHGHVSAVPELWAYKASLLAHRSGHGRGGVKTDEQIIAAAHKETWPDGHI

RLIKERGADMTIFSRAIASMAHVGDSVAVPWAQACNNLIARVVVLFPETFTAGVCMLPQ
ELLHNHGTDMQLISRTPFQMNSAKPARVHVWCEEVNTLIHRQCTLIPEMFIPVAGLPQ

SPEADMTSSIAELECRVPMGFLCNPLNPDP-GGGHFKHPPTDFRWFYFPFYEKMVELDVP
VAGEPIENVFE-MDRCV-SMGFKGFLNPDPYENGAEEAPPLGDRYWYPYELKCELDDLKP

AMIHVSGCNPAMHATGAYYLAADTIAFMQLCQLGFADFPTLRFIIPHGGA VenWGR
AHIATGQS-QERSYSHFINETIATYNLCTSSVFDDFQPLKVVSHGGAGAIYQLGR

FRGLADMLKQPSLDTLMMNVVFDTVCYYHQPGLNLLADVIDNKNILFGSEMVGAAGRCIP
FE---SQRSSKHFSERMALYFDTVLYTEGALRLLIETVPERCLFGSECPGVGSTLIP

TTGHYFDDTKRIADLD-ISDQERHAFEGRTRVFPRLDALKAREGHHHHHHHH
ATGKQMDHIAPFIQKDFDSLADKLLIFEDNARKVFN-----LEVEGHHHHHH

Figure 26 – Clustal W alignment of Lig Y and Lig J (a possible member of subtype II). Potentially conserved residues are highlighted.
conserved residues. (Figure 26) This indicates that Lig Y is possibly part of subtype II (the subtype Lig J is believed to be part of).

iv. Conclusions

Lig Y, which catalyzes the conversion of OH-DDVA meta-cleavage product into 5CVA in the lignin degradation pathway, was found to solubly express when grown in BL-21AI cells at 37 °C in LB-media with no metal and induced with 0.1 mM IPTG and 0.2 % L-arabinose followed by 24 hours of growth. Using a nickel charged affinity chromatography column, Lig Y was purified from other cellular proteins. Size exclusion chromatography (SEC) was used to examine the native state of the enzyme. When characterized by SEC, Lig Y displayed two major peaks in addition an aggregate peak, one indicating a possible pentamer and another that fell between the expected location for a dimer and monomer. However, given the very low concentrations of the enzyme further study of this is needed. The amidohydrolase subtype of Lig Y was also examined by aligning its sequence with known members of the other subtypes. While Lig Y did not align well with any of the subtype sample enzymes described by Raushel21, it did align well with the conserved loop residues of Lig J, which appears to be a member of Subtype II.
4.2 Dioxygenase Enzymes

A. Des B

i. N-terminal His-tag Expression Screening

Further characterization and kinetic studies of the four amidohydrolase enzymes were limited by the lack of commercially available or easily synthesized substrates, and so focus turned to the preceding enzymes in the pathway. Des B converts gallate to 4-OMA (the product of the Lig I’s reversible reaction and Lig J’s substrate). Gallate is easily obtained and thus functionally expressed Des B could be used to make Lig I substrate for kinetic studies.

Initially a small scale expression screen was performed for Des B. A desB containing p-TOM15b vector is transformed first into in BL-21 cells (E. coli cells engineered for high protein output when induced with IPTG). These cells are grown in media and aliquoted into 4 mL portions after reaching confluency. Aliquots are exposed to varying growth and induction conditions and time points are collected. Each time-point is processed to separate the soluble and insoluble fractions which are visualized by SDS-PAGE to determine the presence of the target protein.

Figure 27 displays the SDS-PAGE visualized results of Des B induced at 37 °C in a metal-less LB-media solution with varied IPTG. Lanes below the dark green rectangle on the left-side have been induced with 1 mM IPTG, their counterparts on the right below the light green rectangle have been induced with 0.1 mM IPTG. Soluble and insoluble fractions of both conditions
are placed side by side for comparison purposes. Des B has an expected molecular weight of 49,259.8 Da and would be expected to display a band at the same height as the 50 kDa band on the ladder (L). Both conditions in Figure 26 appear to have bands at the expected height in the soluble fractions much more so than the insoluble.

Figure 27 – Gel results of the small expression screening of Des B from BL-21 cells at 37 °C in media without metal with varied IPTG. Optimal condition circled in black.

Of the two, 0.1 mM IPTG followed by 24 hours of growth is chosen as the optimal condition because it showed slightly larger growth relative to E. coli proteins at other molecular weights.

Expression of Des B was scaled up to a larger growth (2 liters) in order to obtain a greater yield of protein. SDS-PAGE and the process used to separate the soluble and insoluble fractions for the small scale expression screens denature the proteins so an alternative purification approach was utilized. As described for the amidohydrolase enzymes, E. coli cells in the larger growths are harvested by centrifugation and broken open using a French Press Pressure Cell. The insoluble fraction is then removed from the cell lysate by further centrifugation and the soluble fraction is run over an affinity chromatography column charged with a metal ion. Des B has
(his)\textsubscript{10}-tag close to the N-terminus, which should allow it to be separated from other \textit{E. coli} proteins. The cell lysate was divided into two portions and run over a column charged with nickel ions as well as a column charged with zinc ions. The fractions from each were collected and visualized using SDS-PAGE. (Figure 28) Des B has an expected molecular weight of 49259.8 Da and should display a band near the 50 kDa band of the ladder. However, no such bands are visible in the fractions of either column.

![Figure 28](image)

**Figure 28** – A. Nickel affinity column fractions showing purification of Des B enzyme from cell lysate. B. Zinc affinity column fractions showing purification of Des B enzyme from cell lysate.

Bands at 50 kDa in the pellet, crude and flow through fractions of the nickel column and in the pellet crude and second binding buffer of the zinc column suggest that the bands believed to be Des B in the expression screen were not the target protein, but rather an \textit{E. coli} protein showing noticeable growth. While there appears to be a band near 50 kDa in the first elution
fraction, it is not pure. Because Des B was not expressed well in the original conditions tried, further small scale screenings were attempted with expanded growth and induction conditions. Figure 29 summarizes the expanded conditions as well as the original.

Figure 29 – Growth and induction conditions tested for Des B expression in pTOM-15b.

This included transforming the desB containing vector into BL-21AI cells which are engineered for high-protein output when induced with IPTG and L-arabinose as well as adding iron to the growth media (to promote the formation of soluble proteins by correct folding) and growing cultures at three different temperatures after induction (37 °C, 30 °C, and 16 °C).
A small scale expression screening using these expanded conditions was performed for Des B. Figure 30 displays the SDS-PAGE visualization of the timepoints for each condition. The gels are divided into three sections by post-induction growth temperatures, A. 37 °C, B. 30 °C, and C. 16 °C. Within each lettered section, the gels are further organized by conditions. The left side of each gel, below the green rectangle shows growths induced with 1.0 mM IPTG while growths induced with 0.1 mM IPTG are shown on the right below the lighter green rectangle. L-arabinose is decreased (0.2 %, 0.002 %, and 0.0002 % moving from left to right) in both IPTG concentrations. The soluble fractions for each condition are shown in the top row of gels while the corresponding insoluble fractions are shown in the lower row.

The presence of Des B is indicated by a band near the 50 kDa band of the ladder (expected molecular weight of 49,259.8 Da). The soluble fractions of Figure 30 do show a band at 50 kDa. However, it does not show increasing time-dependent growth, meaning that the band size and intensity remain relatively constant across the time points and any increase occurs on the same level as the *E. coli* (a result of increasing *E. coli* cells as the growth progresses). There is, however, a large amount of Des B present in the insoluble fraction of several conditions (virtually all in A., the 24 hour time point of the 0.1 mM IPTG – 0.2 % L-arabinose induced condition in B. and C.). This suggests that Des B is being expressed at levels greater than the
Figure 30 – Gel results of the small expression screening of Des B with varied IPTG and L-arabinose from BL-21 AI cells in media containing iron and cysteine, grown at A. 37 °C, B. 30 °C, and C. 16 °C.
native proteins of *E. coli*, but that the expressed protein is not functional. It was hypothesized that some combination of protein misfolding and inclusion body formation was likely occurring. When possible causes for the insolubility of the protein, the (his)$_{10}$-tag used for purifying Des B was identified as a possible culprit. The pTOM-15b vector containing *desB* encodes a ten histidine tag near the N-terminus of the resulting enzyme. This functions as an easy means of isolating the induced target protein from other *e. coli* based proteins by affinity column chromatography. While this provides a useful means of purification, there was concern that the tag might be interfering protein folding and possibly negatively interacting with the Des B’s active site (also found near the N-terminus), which in turn was leading to the aggregation of misfolded proteins (inclusion body formation).

**ii. Creating a C-terminal His-tag through Genetic Manipulations**

To remove the (his)$_{10}$-tag from the N-terminal end of the protein, it was necessary to clone *desB* into a different vector. A vector with a C-terminal (his)$_{6}$-tag, pET-21b, was chosen. Moving the location of the tag would allow Des B to still be purified with affinity chromatography while hopefully not adversely affecting folding or solubility.

Cloning *desB* into another vector requires digestion of *desB* in pTOM-15b as well as the pET-21b vector by restriction enzymes. After arresting both digestions, the reaction mixtures are run through an agarose DNA gel stained with ethidium bromide (EtBr). This both allows the results of the reactions to be visualized, but also separates the fragments which may
then be harvested from the gel. Once purified, the desired desB and pET-21b fragments are ligated together. After desB in pET-21b is obtained, mutating the adjacent stop codons at the end of desB is required for successful translation of the (his)_6-tag. Thus, cloning into pET-21b would in fact provide two different Des B enzymes; one with no tag at all (unmutated stop codon) and one with a C-terminal His-tag (mutated stop codon). While the primary focus was the latter, this provided an alternative path. If inclusion bodies were also present in Des B with a C-terminal his-tag, the former could be used and potentially purified using non-affinity based methods such as anion exchange chromatography and size exclusion columns.

A sample of pET-21b vector containing a hepI gene was obtained from Daniel Czyzyk. Both desB and hepI had been cloned into their respective plasmids using NdeI and BamHI cut sites. This allows for digestion of desB out of pTOM-15b and religation into pET-21b without any modification of the fragment ends.

The successful digestion of desB in pTOM-15b by NdeI and BamHI is expected to create two fragments. The first, containing desB is 1260 base pairs (bp) long and would be expected to appear as a band in between the 1 kb and 1.5 kb bands of the ladder. The second, containing pTOM-15b, is 5699 bp long and should display a band slightly below the 6 kb band of the ladder. The successful digestion of hepI in pET-21b by restriction enzymes NdeI and BamHI also creates two fragments, one containing the hepI gene with an expected length of 960 bp and the remaining pET-21b plasmid with
and expected length of 5405 bp. Visualized on an EtBr stained agarose gel, *hepl* and pET-21b should display bands near the 1 kb band and slightly below the 6 kb band of the ladder, respectively. Digestion via the restriction enzymes NdeI and BamHI was confirmed by the clear presence of individual bands at the expected fragments sizes following separation by gel electrophoresis. (Figure 31)

Figure 31 – EtBr stained agarose gel results of the restriction enzyme digestion of A. *desB* in pTOM-15b and B. *hepl* in pET-21b.
Initially, the desired fragments were to be isolated using a Qiagen gel extraction kit. However, after several attempts resulting in very low extraction yields and unsuccessful ligations, a different plan was devised. Figure 32 shows the modified cloning method for obtaining desB in pET-21b. Because extraction of the desired fragments was not fruitful, ligation reactions of all fragments were performed using conditions that favored the creation of desB in pET-21b.

Figure 32 – Modified plan for cloning desB into pET-21b vector. Blue circles indicate pTOM-15b vectors. Red circles indicate pET-21b vectors. Green insert indicates the desB gene and purple insert indicates the desZ gene.
In this new attempt, the only alteration to the original digests was that SapI was added to the desB in pTOM-15b digest instead of the hepl in pET-21b digest. Rather than separating fragments by agarose gel, the enzymes were removed by phenol-chloroform extraction and the DNA fragments were ethanol precipitated and used directly in ligation reactions. Because the gene and plasmid fragments were not separated prior to ligation, this method did allow for the formation additional undesired ligation products, as shown in Figure 31. It is possible for hepl to ligate into pTOM-15b or relegate into pET-21b, creating three potential products. Because however, the desB in pET-21b digestion was exposed to SapI (which removes phosphates from the end of DNA strands), religation of those fractions was highly unlikely. Moreover, 10 times the amount of desB (as compared to hepl and pET-21b) was digested and added to each ligation reaction to create a higher possibility of ligation of the desired product.

The products of the ligation reaction are then transformed into DH5α E. coli cells engineered for high plasmid production and seeded onto agar plates with ampicillin. Several colonies from each plate are tested for the presence of the desired gene in the vector. Colony PCR is used to identify the desired ligation product (desB in pET-21b). The pET-21b vector has a T7 promoter and terminator sequence on either side of the gene ligation sites. Using the associated T7 pro and term primers, PCR of the isolated plasmids produces linear DNA sequences with a length approximately 200 base pairs longer than the size of the gene ligated between the NdeI and BamHI cut
sites. Because *desB* is 1260 base pairs long and *hepl* is 960 base pairs long, the resulting PCR products (approximately 1460 and 1190, respectively) can be easily identified. Figure 33 shows the colony PCR product (run in triplicate) of *desB* in pET-21b. All three colony PCR reactions show bands slightly below the 1517 bp band and well above the 1200 bp band of the ladder, indicating that *desB* has been successfully ligated into pET-21b.

![Figure 33 – Colony PCR results of desB in pET-21b in EtBr stained agarose gel.](image)

As previously mentioned, pET-21b encodes a C-terminal histidine tag. However, *desB* ends with a STOP codon sequence of –TAATAG that, unless altered, would prevent transcription from continuing onto the histidine codons. To introduce the desired mutations, Stratagene’s Quikchange Lightning Mutagenesis Kit was used. This process involves designing primers that complement the sequence surrounding the targeted area with point mutations of the desired bases. Figure 34 shows both primers designed to introduce
Figure 34 – Two primers designed to use with Stratagene’s Quikchange Lightning Mutagenesis Kit to introduce point mutations. Both are forward primers (reverse complement primers were also synthesized for each). Point mutations are highlighted.

point mutations in the desB gene. Thermal cycling with desB in pET-21b and the designed primers is utilized to create many copies of the plasmid that contain the point mutations. The reaction product is then digested with the restriction enzyme DpnI, which specifically reacts with methylated DNA. Newly synthesized DNA plasmids would not be methylated, whereas any parental DNA would be at least partially methylated and would be subsequently digested. Finally, the newly made, point-mutated plasmids are translated into competent cells, seeded into growths, and harvested for future use.

After initial attempts (using the first primer and the recommended QuikChange protocol, several method variations were also tested. After eight unsuccessful attempts comprising multiple conditions, the decision was made to stop pursuing mutation of this gene.

iii. Conclusion

Des B catalyzes the conversion of gallate to 4-OMA, the product of Lig I’s reversible reaction and substrate of Lig J’s reaction. Expression conditions were initially screened on a small scale for the desB gene in a pTOM-15b
vector, grown in BL-21 cells with varied IPTG. Within that small scale expression, inducing with 0.1 mM IPTG and growing for 24 hours appeared to produce soluble time-dependent growth. However, when grown, induced and purified with affinity column chromatography on a larger scale, no pure soluble protein was present. Expression conditions were then expanded to growth of desB containing BL-21AI cells in LB-media containing iron and cysteine at three temperatures ((37 °C, 30 °C, and 16 °C) with varied IPTG and L-arabinose. The results of this screening showed large amounts of insoluble Des B in a number of conditions, but little to no soluble Des B.

Concern that the N-terminal (his)_{10}-tag from pTOM-15b was leading to misfolding, cloning desB into pET-21b (which possesses a (his)_{6}-tag on the C-terminal). Separate restriction enzyme digestions of desB in pTOM-15b and hepI in pET-21b with Ndel and BamHI were successfully performed and confirmed with gel electrophoresis. When obtaining the desired fragments (pET-21b and desB) directly from the gel proved unfruitful, the total DNA fragments were separated from digestion reactions by phenol-chloroform extractions and combined for ligation and transformation, with an excess of desB and pTOM-15b fragments. The desired product, desB in pET-21b was then identified from colonies of transformed cells using colony PCR.

Once pET-21b containing desB had been obtained, it was necessary to mutate the stop codons separating desB and the C-terminal (his)_{6}-tag. Stratagene’s Quikchange Lightning Mutagenesis Kit was used to mutate the stop codons into two serine residues. However, several attempts including
multiple thermal cycling, DpnI digestion, and transformation conditions as well as two different mutation primers were not successful and it was decided to discontinue mutagenesis attempts.

B. Des Z

i. N-terminal His-tag Expression Screening

Just as with Des B, Des Z was focused on due to its proximity to the amidohydrolases in the lignin degradation pathway. Des Z converts 3-O-methylgallate (3MGA) to 2-pyrone-4,6-dicarboxylate (PDC), the substrate of Lig I’s reversible reaction, and 4-carboxy-2-hydroxy-6-methoxy-6-oxohexa-2,4-dienoate (CHMOD) which is converted then to 4-OMA (Lig J’s substrate). 3-O-methylgallate is commercially available and so functionally expressed Des Z could be used to obtain substrate for Lig I and Lig J.

A small scale expression screen is initially performed for Des Z beginning with a desZ containing pTOM-15b vector being transformed first into BL-21 (E. coli cells that produce high protein outputs when induced with IPTG). These cells are grown in media until confluency is reached and then aliquoted into 4 mL portions. Each aliquot is exposed to a different growth and induction condition and timepoints are collected. Each timepoint is processed to separate the soluble and insoluble fractions which are visualized with SDS-PAGE.

Figure 35 displays the SDS-PAGE visualized results of Des Z induced with two different concentrations of IPTG at 37 °C in a metal-less LB-media solution. Results of the left side of the Figure, below the dark green rectangle
have been induced with 1 mM IPTG whereas results on the right below the lighter green rectangle have been induced with 0.1 mM IPTG. Soluble and insoluble fractions of both conditions are placed side by side for comparison purposes. Des Z has an expected molecular weight of 38,756.8 Da and should display a band at the same level as the 37 kDa band of the ladder.

<table>
<thead>
<tr>
<th>Timepost</th>
<th>L</th>
<th>0</th>
<th>5</th>
<th>24</th>
<th>0</th>
<th>5</th>
<th>24</th>
<th>0</th>
<th>5</th>
<th>24</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 mM IPTG</td>
<td>Soluble</td>
<td>Insoluble</td>
<td>Soluble</td>
<td>Insoluble</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1 mM IPTG</td>
<td>Des Z - BL-21 cell</td>
<td>Molecular Weight = 38756.8 Da</td>
<td>No Metal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 35 – Gel results of the small expression screening of Des Z from BL-21 cells at 37 °C in media without metal with varied IPTG. Optimal conditions circled in black.

Of the conditions tested, 1 mM IPTG and 5 hours growth (post-induction) was selected as the optimal condition because there appeared to be a band at the desired location showing time-dependent growth greater than that of the surrounding *E. coli* proteins.

Expression of Des Z, using the identified conditions from Figure 35, is scaled up in order to obtain larger amounts of protein. As was described previously for Des B, Lig I, Lig J, Lig W, and Lig Y, affinity column chromatography is used to isolate expressed Des Z. Based on those criteria, a large growth at 37 °C with no metal in solution the media was induced with 1 mM IPTG and allowed to grow for 5 hours prior to harvesting as suggested by the highlighted condition in Figure 35. Cell lysate from that growth was run separately over both nickel and cobalt charged columns (Figure 36). Small
amounts of soluble protein were lightly visible in the wash and elution fractions of both columns (black box, Figure 36). However, these were not pure and fell out of solution during concentration. The large pellet (insoluble fraction) present suggests that Des Z production is being induced in

![Figure 36 - A. Nickel affinity column fractions showing purification of Des Z enzyme from cell lysate. B. Cobalt affinity column fractions showing purification of Des Z enzyme from cell lysate.](image)

increased amounts. Additionally, the relative lack of insoluble Des Z protein in the small scale screening suggests that the increase in growth size (from 10-20 mL to 0.5-1 L) may play a role in the aggregations into inclusion bodies.

Because Des Z was not expressed well in the original conditions tested, further small scale screenings were attempted with expanded growth and induction conditions which are shown in Figure 29. This included transforming Des Z into BL-21AI cells which are induced by IPTG and L-arabinose, adding iron and cysteine to the growth media and growing
cultures at three different temperatures after induction (37 °C, 30 °C, and 16 °C). Figure 37 displays the SDS-PAGE visualization of the results of each condition. Gels are divided into three sections by post-induction growth temperatures, A. 37 °C, B. 30 °C, and C. 16 °C. Within each lettered section, the gels are further organized by conditions. The left side of each gel, below the green rectangle shows growths induced with 1.0 mM IPTG while growths induced with 0.1 mM IPTG are shown on the right below the lighter green rectangle. L-arabinose is decreased (0.2 %, 0.002 %, and 0.0002 % moving from left to right) in both IPTG concentrations. The soluble fractions for each condition are shown in the top row of gels while the corresponding insoluble fractions are shown in the lower row.

The presence of Des Z is indicated by a band roughly in the location of the 37 kDa band of the ladder. Des Z is present in large quantity in a number of the insoluble fractions. However, while there are bands present in the soluble fraction near the 37 kDa band of the ladder, it does not appear to be Des Z, but rather an *E. coli* protein, as it does not show time-dependent growth greater than that of the surrounding protein bands.
Figure 37 – Gel results of the small expression screening of Des Z with varied IPTG and L-arabinose from BL-21 AI cells in media containing iron and cysteine, grown at A. 37 °C, B. 30 °C, and C. 16 °C.
The significant amount of Des Z in the insoluble fractions suggests that while excess production of Des Z is occurring, misfolding or the formation of inclusion bodies is also occurring. Because the pTOM-15b vector (a modified version of pET-15b) containing Des Z also encodes an N-terminal ten histidine-tag and Des Z is believed to have a N-terminal located active site, cloning into pET-21b (which has a C-terminal tag) was pursued.

**ii. Creating a C-terminal His-tag through Genetic Manipulations**

Removing the (his)$_{10}$-tag from the N-terminal of the protein requires cloning the \textit{desZ} gene out of pTOM-15b into a different vector. The pET-21b vector was chosen because it possessed a C-terminal (his)$_6$-tag, which would allow purification by affinity column chromatography but would hopefully not interfere with folding as the N-terminal tag may have.

The cloning process began with the digestion of \textit{desZ} in pTOM-15b by restriction enzymes NdeI and BamHI. That digestion is expected to create two fragments, one 990 bp fragment containing \textit{desZ} and another 5699 bp fragment containing the rest of the pTOM-15b. Figure 38 shows the successful digestion of \textit{desZ} in pTOM-15b visualized on an EtBr stained agarose gel, evidenced by the presence of a lower band at the level of the 1kb band of the ladder and another band just below the 6 kb band of the ladder, in concordance with the expected fragment sizes.
Initially, the same digestion of *hepl* in pET-21b was to be used for obtaining the needed pET-21b fragment. However, because obtaining separate digestion fragments from the agarose gel was equally unsuccessful for this enzyme, the cloning plan was altered as with *desB*. This presented another challenge, however. The presence of *desB* in pET-21b in colonies of transformed ligation products was tested for using colony PCR. The *desB* gene is 300 base pairs longer than *hepl* (the gene in the alternate ligation product), so the two are easily discernible on an agarose gel. However, the *desZ* gene is only 30 base pairs longer than *hepl*, which could not be differentiated on an agarose gel. Thus colony PCR is not a feasible option for confirming the presence of *desZ* in pET-21b. Initially, digestion with restriction enzymes to create unique fragment sizes was attempted. Multiple digests were performed with different restriction enzymes. However, none produced clear visible gel results. After several unsuccessful attempts, pure,
un-cut pET-21b (with no genes cloned in) became available. At that point, the cloning plan was modified (Figure 39). This plan closely resembles the method used for cloning desB into pET-21b, the major difference being that unmodified pET-21b vector was used (instead of hepl containing pET-21b). This method also produced multiple ligation products in addition to the desired desZ in pET-21b. However, none of the additional potential products contained gene inserts, making colony PCR an effective method for identifying the target ligation product in this plan. Because the only undesired products possible have no gene inserts, PCR products will be easily
distinguishable (either approximately 200 base pairs for undesired products or approximately 1200 base pairs for the desired product). Figure 40 shows the colony PCR results of two ligation products where the presence of desZ in pET-21b is confirmed by the presence of a band at the same height as the 1200 bp band of the ladder. Sample 4.II.b was carried forward.

As previously mentioned, desB and desZ in pET-21b both have two adjacent STOP codons (– TAATAG) at the end of their sequences that must be modified to allow the transcription and translation of a C-terminal (his)$_6$-tag. Stratagene’s Quikchange Lightning Mutagenesis Kit was used to introduce point mutations into the STOP codons. Forward and reverse primers were designed to convert the stop codons into two serine residues. (Figure 41)

<table>
<thead>
<tr>
<th>100 bp Ladder</th>
<th>DesZ Colony PCR Sample 4.I.A</th>
<th>Sample 4.II.B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1517</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 40 – Colony PCR results in triplicate of desZ in pET-21b visualized by an EtBr stained agarose gel.

Figure 41 – Primer designed to use with Quikchange Mutagenesis kit, with point mutations highlighted. Its counterpart is the reverse complement.
After a few unsuccessful mutagenesis attempts, variations of the procedure were attempted. During the thermal cycling phase, the annealing temperature was tested at 50 °C, 55 °C, and 60 °C and the extension time lengthened from 3.5 minutes to 5 minutes. Additionally, the DpnI digest (which removes the un-mutated parent DNA) was lengthened and the amount of enzyme increased. These variations produced five (out of six reactions) successful mutations (shown aligned with un-mutated des Z in pET-21b in Figure 38.) Mutations were deemed successful, because the two stop codons –TAATAG were successfully changed to two adjacent serine residues TCTTCT.

iii. Expression Screening with C-terminal His-tag

With the successful cloning of the desZ gene into pET-21b, another small screening for optimal expression conditions was performed. Of the successfully mutated samples of desZ in pET-21b, sample DesZ_60A was
chosen for use in the screening. Figure 43 summarizes all of the conditions tested in this screening. The pET-21b vector containing desZ was transformed into both BL-21 and BL-21AI. Growths of BL-21 cells were cultured in media containing cysteine and no metal, iron, or manganese and after induction with one of two concentrations of IPTG (1 mM or 0.1 mM) were grown at one of three temperatures (37 °C, 30 °C, and 16 °C). Growths of BL-21AI cells were exposed to the same conditions with the addition of four different L-arabinose concentrations (0.2 %, 0.02 %, 0.002 %, or 0.0002 %).

Figure 43 – Expression conditions for Des Z with a C-terminal his-tag.
The small scale screening is performed as previously described. LB-media containing iron, manganese or no metal are inoculated and grown at 37 °C until confluency is reached. Then each growth is aliquoted into several 4mL aliquots which are induced with IPTG or IPTG and L-arabinose. Each aliquot is then grown for 24 hours at one of three denoted temperatures with a 1 mL timepoint taken immediately after induction and then at 5 and 24 hours. Centrifugation is used to harvest cells from the growth media for each timepoint. The soluble and insoluble fractions are then separated and visualized with SDS-PAGE.

Figure 44 shows the results of all expression conditions tested with BL-21 cells. Soluble fractions are shown in the top three rows while the insoluble fractions are shown on the bottom three rows of gels. The top row of both the soluble and insoluble fractions shows conditions grown at 37 °C, with those grown at 30 °C and 16 °C in the next two rows below. Metal content is organized into three columns for both the soluble and insoluble fractions with growths containing no metal, iron, and manganese arranged from left to right (as shown by the red rectangles). Both IPTG conditions for each metal and temperatures combination are shown directly below the metal in the green rectangles with 1 mM IPTG on the left and 0.1 mM on the right. Des Z has an expected molecular weight of 38,756.8 Da and its presence can be determined by a band at the same height as the 37 kDa band of the ladder. While the insoluble fraction appears to show several bands at the
Figure 44 – Gel results of the small scale expression screening of Des Z for BL-21 cells at 37 °C, 30 °C, and 16 °C in LB-media containing no metal, iron or manganese with varied IPTG.
expected location, no such bands with time-dependent growth appear in the soluble fractions.

Figures 45-47 show the results of all expression conditions tested with BL-21Al cells. Each Figure shows the results of a different temperature, Figure 45 – 37 °C, Figure 46 – 30 °C, and Figure 47 – 16 °C. For all three Figures, soluble fractions are shown in the top two rows while the insoluble fractions are shown on the bottom two rows of gels. Metal content is organized into three columns for both the soluble and insoluble fractions with growths containing no metal, iron, and manganese arranged from left to right (as shown by the red rectangles). The top row for both the soluble and insoluble fractions show results of conditions induced with 1 mM IPTG while the row below shows 0.1 mM IPTG. The L-arabinose concentrations are decreased from left to right across the top for each metal variation. While timepoints for conditions induced with 0.1 mM IPTG were collected, they were not processed or subjected to SDS-PAGE because the results for conditions induced at 1mM IPTG for the same temperature so no soluble growth at all. Therefore, Figure 47 does not show conditions induced with 0.1 mM IPTG.

Des Z has an expected molecular weight of 38,756.8 Da and its presence can be determined by a band at the same height as the 37 kDa band of the ladder and distinguished from the E. coli gene in the same location by time-dependent growth greater than that of the surrounding bands.
Figure 45 – Gel results of the small scale expression screening of Des Z for BL-21AI cells at 37 °C in LB-media containing no metal, iron or manganese with varied IPTG and L-arabinose.
Figure 46 – Gel results of the small scale expression screening of Des Z for BL-21 cells at 30 °C in LB-media containing no metal, iron or manganese with varied IPTG and L-arabinose. Optimal conditions circled in black.
at other molecular weights. While the insoluble fraction appears to show several bands at the expected location, most soluble fractions do not show time-dependent growth. There were, however, a few potential exceptions. These variations came from Des Z in BL-21AI cells at 30 °C. A number of conditions seemed favorable, especially among those induced with 1 mM IPTG. Three in particular were chosen as particularly favorable: 1. No metal, 1 mM IPTG, and 0.2 % arabinose, 2.) Fe, 1 mM IPTG 0.2 % arabinose, 3.) Fe, 0.1 mM IPTG, and 0.02 % arabinose (Figure 45, circled). The three

---

<table>
<thead>
<tr>
<th>Temperature</th>
<th>BL-21AI cell</th>
<th>Molecular Weight = 38756.8 Da</th>
<th>Insoluble Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal</td>
<td>No Metal</td>
<td>Iron (Fe)</td>
<td>Manganese (Mn)</td>
</tr>
<tr>
<td>L-arabinose</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>1 mM IPTG</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

Figure 47 – Gel results of the small scale expression screening of Des Z for BL-21 cells at 16 °C in LB-media containing no metal, iron or manganese with 1mM IPTG and varied L-arabinose.
Figure 48 – Nickel affinity column fractions showing the purification of Des Z enzyme from cell lysate expressed at 30 °C from three different conditions. A. No metal, 1 mM IPTG, 0.2 % L-arabinose. B. Iron, 1 mM IPTG, 0.2 % L-arabinose. C. Iron, 1 mM IPTG, 0.02 % L-arabinose.

chosen conditions were then tested on large media growths, purified with a nickel charged affinity chromatography column and the results fractions were visualized with SDS-PAGE (Figure 47).

Similarly to other large growths previously attempted, the solubility shown at a smaller scale completely disappears and the pellet (insoluble fraction) contains a larger amount of the protein (Figure 47). This suggests that the formation of inclusion bodies increasing with growth size occurs whether the His-tag is placed on the N-terminal or C-terminal end.
3.3. Conclusion

Des Z catalyzes the conversion of 3MGA to CHMOD and PDC. Expression conditions were initially screened on a small scale for the desZ gene in a pTOM-15b vector, grown in BL-21 cells with varied IPTG. Within that small scale expression, inducing with 1 mM IPTG and growing for 5 hours appeared to produce soluble time-dependent growth. However, when grown, induced and purified with a nickel and cobalt affinity column chromatography on a larger scale, no pure soluble protein was present. Expression conditions were then expanded to growth of desZ containing BL-21Al cells in LB-media containing iron and cysteine at three temperatures (37 °C, 30 °C, and 16 °C) with varied IPTG and L-arabinose. The results of this screening showed large amounts of insoluble Des Z in a number of conditions, but no soluble Des Z.

Concern that the N-terminal (his)10-tag from pTOM-15b was leading to misfolding, cloning desZ into pET-21b (which possesses a (his)6-tag on the C-terminal was pursued. A restriction enzyme digestion of desZ in pTOM-15b with NdeI and BamHI was successfully performed and confirmed with gel electrophoresis. Digestion of uncut pET-21b was performed because hepl and desZ have only 30 bp difference in length and cannot be discerned on a gel. All DNA fragments were separated from digestion reactions by phenol-chloroform extractions and combined for ligation and transformation, with an
excess of desZ and pTOM-15b fragments. The desired product, desZ in pET-21b was then identified from colonies of transformed cells using colony PCR.

Once pET-21b containing desZ had been obtained, it was necessary to mutate the stop codons separating desZ and the C-terminal (his)\(_6\)-tag. Stratagene’s Quickchange Lightning Mutagenesis Kit was used to mutate the stop codons into two serine residues. After varying the thermal cycling step by lengthening the extension period and varying the annealing temperatures (50 °C, 55 °C, and 60 °C), as well as lengthening the DpnI digestion time and amount of enzyme used, the stop codons at the end of desZ were mutated, allowing the C-terminal (his)\(_6\)-tag to be translated. DesZ was then transformed into both BL-21 and BL-21AI cells and a small scale expression screening was performed with several conditions. While some showed minor soluble growth in the small scale, they showed only insoluble growth when scaled up to a larger amount.
3.4 References


IV. CONCLUSIONS AND FUTURE DIRECTIONS

The four enzymes of the amidohydrolase superfamily, Lig I, Lig J, Lig W, and Lig Y have been successfully tested for soluble expression conditions, purified in larger amount and minimally characterized. The dioxygenase enzymes, Des B and Des Z, were tested over a number of conditions for soluble expression, but no successful conditions were found that produced soluble enzyme on a large scale. Future study of these enzymes in the Taylor lab will involve expression and purification without (his)-tags to investigate whether the tags are the cause of insolubility. However, the primarily focus will be in expressing and purifying Des B and Des Z anaerobically.

Lignin has the potential to be a valuable biofuel pre-cursor, when broken down into its central metabolites. Lig I, Lig J, Lig W, Lig Y, Des B, and Des Z play a vital role in this catabolic degradation of small lignin derived molecules. The enzymes in this catabolic pathway catalyze reactions that in many cases would require harsh chemicals or intense conditions to perform synthetically. Harnessing the power of these enzymes will be critical for sustainable future biofuel applications, whether it is as a means to easily separate lignin from cellulose or as a source for small hydrocarbons.

In order to be able to achieve that, it is necessary that the enzymes be better understood. While the majority of substrate and product reactions in the pathway have been defined, there is much that has yet been discovered about the mechanisms of catalysis, as well as the structures of the active site.
This would be especially important for the enzymes discussed in this paper because they are primarily located towards the end of the pathway. The Raushel group has already begun this process for the amidohydrolases\textsuperscript{28} Solving and examining the crystal structures of the wild types as well as mutants which have had key active site residues altered would be a possible first step.
V. Appendix

Amino Acid Sequences:

Lig Y

MIIDCHGHVSAIPVELWAYKASLLAHRGSHGRGGVKTDEQIIAAAHKETWDGDHIELLLNHGTDMQLSPRTFQMMNSAKPARVWFWCDEVNLTILHRQCTLIPEMFIPVAGLPQVAGPIENVFAEMDCVSVMFGKGFLLNPDPYENGAEAAPLLGDARYWPILYEKLELCILDPAHIAHTGSGQERSPSYSLHGFNENETIATYNLCTSSVFDDFPQLKVVSHGGAIPYQLGRFESQSRRSKHLSERMAKLYFDTVLYTEGALRLIEFTVGPERCLFGSECPSVGSTIDPATEGKQMDHIAPFIQKFDFLSDADKLIPEFNAKVFNL

Lig W

MRLIATEEAVTFQPVVDLARAHRTDASSDMILVDRDYGVDEPARPAMIGRLSDVTVLRALEMSNGVDMHLSLTAPGQVMFDATEGTRLARIANDLMAVTVAANPRTFAGLGTFAPQPDASAAREIERVATQLRLNLVINSHTNDLYYDDPFHFVPEAIEASGLAYIPRAPSQIDRAFRDIGMNSAIWGYGIESNAVRMLSGLDFRPRKLILVGHMGEAIPFWLWRLDYMHNATTFFGAPKLLKPEYEYRNFAITTSGVESHAALRYSIEVLGPEVMWAIDYPYQMPAVQFIRTAPIPEDVKAMVAGGNAARIKRIT

Lig I

MTNDERILSWNTPSKPRYTPPAIGADAHCHVFGPMAQFPFPKAKYLPRAAGPDMFLRLRDHLGFARNVIVQASCHGTDNAATLDAIARAQGKARGIAVVDPAIDEAELIALHEGMRGRFNLKRLVDDAPKDFKLEVAGRLPAGWHVIYFEADILEEFLREFPMDAIPIDHMRGPRDVQGPDGADMKAFRRRLDREDIFKATCPDRLDPAGPPWDFFARSVAPLVAADYADVRIWGTWPHNPMPQDAIPDDGLVVDMPRIAPTPELQHKMLVTPNMRLYWSEEM

Lig J

MMMIIDCHGHYTVLPAHDWEQKAAFKAGQQPAPPYEPISDEIREDIEAQLRLIKERGADMTIFSRSASPAMAPHVDQSVAPWACNNLIARVVDLFPETFAGVCMLPQSEADMTSSIAELERCVNELKGICGLNLNPDPGHHFHKHPPPLTDWRYPFYEKMEVLDVPAMIHSGCNPMAMHTGAYLYAATDIAFMQLQCNLFADDFPLRFIPPHGGAIVYHWGFRGGLADMLKQPSLDLMLMNVFDDTCVHYQPGINLLADVIDNOILCFGSEMVGAVGIPDGTTGHYFDDTKEYDALDISQERHAIFEGNTRVFPRLDAKLKARGL
Des B

MAKIIGFAVSHTPTIAFAHDANKYDDPDVWAPIFQGFEPVKQWLAEQKPDVT
FYVYNDHMTSFFEHYSHFALGVGEEYSPADEGGGQRDLPPIKGDPELAKHI
AECLVADEFDLAYWQGMLDHGAFSPLSVLLPHEHGWPVCRIVPLQCGVLQ
HIPKARRFWNFGSLRRAIQSYPRDKVAIAGTTGLSHQVHGERAGFNNT
EWDMEFMERLANDPESLLGAVTDLAKKGWEGAEVVMWLLMRGALSPE
VKTILHQSYFLPSMTAIAHMLFEDQGDAAPPAESDEALRARKRELAVEEIE
GTPFTIDRAVKGFRINHFLHRLEPDFRFKRFVEDPEGLFAESDLTEEKLIR
NRDWIGMIHYGVIFMKEKMAAVLGIGNIDVYAFRGLSVPEFKTRNAAITY
SVAGKQ

Des Z

MAEIVLGIGTSHGPMLVTQETQWRSRLAFDCSVNHAWRGGSWSYDQLVA
ERADQNFAAQITPEAMTAHNARCQASLDQLAEIFSEAKIDVAVILGNDQMEIF
DERLVPASFVYFDITNYEFPPPERMAALPPGINLSVAGYIPSGGAETYAGQP
ELARSIIAQAMADEFDVAAMKAPPKPETPHAFGFVYRRIMRDNPVPVPLV
NTFYPPNQPTVRRCYEGKSVLREIGAWESDARVAVLASGGLTHFVIDEIID
RLEFQAMEDRDIARLADLGEAIFFQDGTSKLNWIPLAGMMAELGLDHEILDY
VPCYRSEAGTGNAMGFVCWR
MALDI-TOF Mass Spectometry results for Lig I (33,320 Da)