TWO STOPS ON THE PATHWAY TOWARDS LIGNIN UTILIZATION:
A STORY OF SYNTHESIS AND ENZYMEOLOGY

By
Kevin P. Barry
Faculty Advisor: Dr. Erika A. Taylor

A Dissertation submitted to the Faculty of Wesleyan University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Middletown, Connecticut April 2015
Abstract

LigAB from *Sphingomonas paucimobilis* SYK-6 is the first structurally characterized dioxynogenase of the largely uncharacterized superfamily of Type II extradiol dioxygenases (EDOs). This enzyme catalyzes the oxidative ring opening of protocatechuate (3,4-dihydroxybenzoic acid or PCA) in a pathway allowing the degradation of lignin derived aromatic compounds (LDACs). In addition to PCA, LigAB has also been shown to catalyze the dioxygenation of gallate, and 3-O-methyl gallate (3OMG) from the same pathway; however, steady-state kinetic parameters (\(k_{\text{cat}}\), \(K_M\), and \(k_{\text{cat}}/K_M\)) had not been reported for these or any other compounds. The steady-state kinetics determined here found the dioxygenation of PCA by LigAB was highly efficient, with a \(k_{\text{cat}}\) of 216 s\(^{-1}\) and a \(k_{\text{cat}}/K_M\) of 4.26 \(\times\) 10\(^6\) M\(^{-1}\)s\(^{-1}\). LigAB also demonstrated the ability to use a variety of additional catecholic molecules as substrates including 3,4-dihydroxybenzamide, homoprotocatechuate, catechol, and 3,4-dihydroxybenzonitrile. Kinetic parameters for the dioxygenation of these substrates were also determined.

Additionally, we have observed the first evidence for allosteric feed-forward activation of a ring-cleaving extradiol dioxygenase in LigAB. Allosteric rate enhancement occurs in the presence of unreactive protocatechuate-like aldehydes such as vanillin. Despite wild-type LigAB’s (LigAB-WT) broad substrate utilization profile, the rate enhancement is only observed during the dioxygenation reaction with the natural substrate (protocatechuate, PCA) in the presence of the vanillin – increasing \(k_{\text{cat}}\) by 36 %. Computational docking has identified a potential site of allosteric binding near the entrance to the active site. Point mutants of Phe103\(\alpha\) and Ala18\(\beta\), two residues located in the putative allosteric pocket, display altered rate enhancement as compared to LigAB-WT, providing support for the computationally identified allosteric binding site. These results provide valuable catalytic insight into the reactions
catalyzed by LigAB and make it the first Type II EDO that is fully characterized both structurally and kinetically.

In addition to the studies of LDAC metabolism, syntheses of lignin model compounds capable of spectrophotometrically reporting on lignin depolymerization in living systems were pursued. The molecules were designed such that two fluorescent dyes were incorporated on opposite ends of a lignin dimeric core, and serve as a FRET-pair. Lignin depolymerization would cleave the dimer, resulting in a separation of the pair, and allow the fluorescent signal to be detected. While significant progress has been made towards the designed constructs, the project remains a work in progress.
Acknowledgements

Over the last eight years, I have spent countless hours in pursuit of my research and completing my dissertation. Neither of these endeavors would have been possible without the guidance from my committee members and other Wesleyan University faculty members, and endless support from my friends, family, and my fiancée (soon to be wife).

With a background in inorganic chemistry, I arrived at Wesleyan with very little biochemical research experience. However, Dr. Erika Taylor eagerly welcomed me into her laboratory, and expertly showed me the ropes. Joining her fledgling lab, and taking on a project involving metalloenzymes (of which neither of us had experience working with) was certainly not without challenges, but these challenges were made surmountable by her guidance, support, and friendship over the years.

I would like to thank my committee members Dr. Rex Pratt (chair) and Dr. T. David Westmoreland. Their advice and direction allowed me to overcome the many hurdles faced as a graduate student in the Wesleyan Chemistry Department. Additionally, Dr. Pratt’s expertise in biochemistry and kinetics has been an invaluable source of knowledge. I would also like to thank Dr. Ishita Mukerji and Dr. Manju Hingorani for their help in guiding experiments for my research.

I would like to thank all of the members of the Taylor lab, both past and present. I would like to thank Daniel Czyzyk for always being supportive, and being a sounding board for research ideas. I would especially like to thank Joy Cote, and Carlos Ramirez-Mondragon for their invaluable friendship, willingness to lend a helping hand when needed, and contributions to my research. They have always made working in the lab enjoyable. Additionally, I would like to thank all of the undergraduates that have contributed to this research, especially:
Ann-Marie Illsley, Abraham Ngu, and Erin Cohn for their outstanding contributions to the work on LigAB, and Matthew Long, Julie Hsia, Alex Givner, and Mackenzie Schlosser for their contributions to the synthesis of lignin model compounds. A strong undergraduate presence has always made the Taylor lab an enjoyable place to work. Thank you to all of them.

I would also like to thank the entire community of the Wesleyan Chemistry. The assistance and encouragement from the faculty and staff is unending, and a tremendous support system. I would especially like to thank Roslyn Carrier-Brault, Sarah Atwell, Caitlin Zinser Palmieri, Donald Albert, and Doug Allen for their help over the years.

Without the encouragement of my parents, I would likely not have transitioned to Wesleyan, and made it to this point. They have always been my biggest supporters, and I would not be who I am today without that support, and the sacrifices they have made for me. I will be eternally grateful for their presence in my life.

Lastly, and certainly not least, I would like to thank my fiancée for her support and love. When we met I had said I would be done with graduate school in another year, maybe two… over four years later she won’t let me forget those words, but without her this day would not have happened.
Table of Contents

Abstract .............................................................................................................. I
Acknowledgements .......................................................................................... III
Table of contents ............................................................................................... V
List of figures and tables ..................................................................................... VIII
Complete list of abbreviations and definitions ................................................ XI

Chapter 1 – A perspective on the historic use of fossil fuels, and an introduction to  
 lignin and the future of bio-derived alternatives ........................................... 1-38
  1.1. Global climate change ............................................................................. 4
  1.2. Returning to our roots ........................................................................... 7
  1.3. The case for renewables ......................................................................... 12
  1.4. A deeper understanding of lignin ............................................................ 17
      1.4.1. Lignin structure and biosynthesis ..................................................... 19
      1.4.2. Lignin depolymerization .................................................................. 24
      1.4.3. Microbial lignin degradation .............................................................. 26
      1.4.4. Metabolism of lignin derived aromatic compounds ......................... 28
      1.4.5. Ring-cleaving dioxygenases ............................................................... 33
      1.4.6. The PCAD superfamily .................................................................... 34
  1.5. Research aims ......................................................................................... 36

Chapter 2 – Characterizing the Promiscuity of LigAB, a lignin catabolite degrading  
 extradiol dioxygenase from Sphingobium sp. strain SYK-6 .............................. 39-77
  2.1. Introduction .............................................................................................. 40
  2.2. Materials and methods .......................................................................... 43
      2.2.1. General methods .............................................................................. 43
      2.2.2. Protein expression ........................................................................... 43
      2.2.3. Aerobic purification .......................................................................... 44
      2.2.4. Anaerobic purification ...................................................................... 45
      2.2.5. Steady-state kinetics assays ............................................................... 46
      2.2.6. Substrate induced enzyme inactivation ............................................ 48
      2.2.7. Metal dependence ........................................................................... 49
      2.2.8. Product determination ...................................................................... 50
      2.2.9. Substrate promiscuity assessment .................................................... 50
      2.2.10. Inhibition assessment .................................................................... 52
      2.2.11. pH effect on iron binding ............................................................... 52
  2.3. Results and discussion ........................................................................... 54
      2.3.1. pH effects on rate ............................................................................. 55
      2.3.2. Metal dependence ........................................................................... 57
      2.3.3. Iron dissociation ............................................................................. 59
      2.3.4. UV-Vis characterization of LigAB activity ....................................... 61
      2.3.5. Steady-state kinetics .................................................................... 65
      2.3.6. Substrate specificity and inhibition ................................................ 70
      2.3.7. Mechanism-based inactivation in steady-state kinetics ................. 75
Chapter 3 – Exploring allosteric activation of LigAB from Sphingobium sp. strain SYK-6 through kinetics, mutagenesis, and computational studies ...... 78-112
3.1. Introduction .................................................................................................................. 79
3.2. Materials and methods .................................................................................................. 82
  3.2.1. Site directed mutagenesis ....................................................................................... 82
  3.2.2. Protein purification ................................................................................................. 83
  3.2.3. Steady-state kinetics .............................................................................................. 83
  3.2.4. Activity enhancement assays ................................................................................ 85
  3.2.5. Product identification ............................................................................................ 87
  3.2.6. Substrate induced enzyme inactivation ................................................................. 87
  3.2.7. Computational docking ......................................................................................... 88
  3.2.8. Sequence alignment of PCAD superfamily members ........................................... 90
  3.2.9. Measurement of LigAB fluorescence spectra ....................................................... 91
  3.2.10. Calculation of the solvent accessible surface area .............................................. 92
3.3. Results and Discussion ................................................................................................. 93
  3.3.1. Allosteric activation of LigAB ................................................................................ 93
  3.3.2. Assessment of vanillin’s impact on LigAB stability ................................................ 100
  3.3.3. Identification of a putative allosteric binding site in LigAB .................................... 102
  3.3.4. Impact of mutants on allosteric activation of PCA dioxygenation .......................... 105
  3.3.5. Potential mechanistic implications ...................................................................... 108
3.4. Conclusions .................................................................................................................. 112

Chapter 4 – Improving alternative substrate utilization of LigAB from Sphingobium sp. SYK-6 through site directed mutagenesis .............................................. 113-132
4.1. Introduction .................................................................................................................. 114
4.2. Materials and methods ............................................................................................... 115
  4.2.1. General methods .................................................................................................. 115
  4.2.2. Site directed mutagenesis ................................................................................... 116
  4.2.3. Protein purification .............................................................................................. 116
  4.2.4. Steady-state kinetics ............................................................................................ 116
  4.2.5. Activity enhancement assays .............................................................................. 118
  4.2.6. Computational docking ....................................................................................... 119
4.3. Results and discussion ............................................................................................... 120
  4.3.1. Altered substrate utilization of LigAB mutants ..................................................... 120
  4.3.2. Impact of Phe103α mutants on allosteric activation of alternate substrate dioxygenation ................................................................. 129
4.4. Conclusions ................................................................................................................ 130
4.5. Future directions ........................................................................................................ 131
Chapter 5 – Design and Synthesis of FRET-pair Labeled Lignin Model Compounds
......................................................................................................................... 133-161
  5.1. Introduction........................................................................................................ 134
  5.2. Results and discussion ................................................................................... 141
    5.2.1. Synthesis .................................................................................................. 141
  5.3. Conclusions and future directions ............................................................... 147
    5.3.1. Conclusions ............................................................................................ 147
    5.3.2. Future directions ..................................................................................... 147
      5.3.2.1. Synthesis ......................................................................................... 147
      5.3.2.2. Biological Significance .................................................................. 148
  5.4. Experimental methods .................................................................................. 148
    5.4.1. Materials and general methods ............................................................. 148
    5.4.2. Chemical synthesis .............................................................................. 149

Chapter 6 – Conclusions and Future Directions .............................................. 162-165

References – ........................................................................................................ 166-185

Appendix – ........................................................................................................... 186-276
List of Figures and Tables

Chapter 1. A perspective on the historic use of fossil fuels, and an Introduction to lignin and the future of bio-derived alternatives
Figure 1.1 Total global energy consumption .................................4
Figure 1.2 Proved petroleum reserves ..................................................8
Figure 1.3 Energy consumption by major economic markets ......................9
Figure 1.4 Consumption of fossil fuels by source ....................................10
Table 1.1 Lignin derived products and applications ..................................18
Figure 1.5 Lignin structural units ..........................................................19
Table 1.2 Percent contribution of monolignols to lignin polymer structure .......20
Figure 1.6 Monolignol radical formation and major lignin subunit bonding motifs ....21
Figure 1.7 Lignin peroxidase oxidative mechanism and the structure of veratryl alcohol ..........................................................27
Table 1.3 Aromatic monomers derived from the depolymerization of lignin ........29
Scheme 1.1 Lignin derived aromatic compound degradation pathway of Sphingobium sp. SYK-6, and the β-ketoadipate pathway ........................................31
Table 1.4 Abbreviations for enzymes and compounds found in Scheme 1.1 ........32
Figure 1.8 Structural comparison of extradiol dioxygenase enzyme superfamilies .....33
Figure 1.9 Structural comparison of known protocatechuate dioxygenase superfamily members ..........................................................34

Chapter 2. Characterizing the promiscuity of LigAB, a lignin catabolite degrading extradiol dioxygenase from Sphingomonas paucimobilis SYK-6
Figure 2.1 Ring-cleavage reactions catalyzed by known Type II extradiol dioxygenase enzymes ..........................................................41
Scheme 2.1 Ring cleavage reactions catalyzed by known PCA dioxygenases ..........42
Scheme 2.2 General mechanism of LigAB inactivation ................................48
Figure 2.2 Steady-state kinetics of aerobically vs. anaerobically purified LigAB .......54
Figure 2.3 pH rate profiles of LigAB with PCA .......................................55
Figure 2.4 Active site of LigAB with PCA bound .....................................56
Table 2.1 The pH dependent binding of iron by LigAB ................................59
Figure 2.5 UV-Vis spectra of LigAB catalysis with PCA ...............................62
Figure 2.6 1H NMR of the LigAB-PCA reaction and product conversion to 2,4-lutidinic acid ........................................................................63
Figure 2.7 Substrate UV-Vis spectral changes in the presence of LigAB .............64
Table 2.2 LigAB steady-state kinetic parameters measured with PCA and analogous substrates ..................................................................................65
Figure 2.8 Steady-state kinetics of LigAB with PCA and substrate analogues ..........66
Figure 2.9 Steady-state kinetics of LigAB with varying O2 concentration .............66
Table 2.3 Steady-state kinetic parameters for Type I, II, and II extradiol dioxygenase enzymes ..........................................................................68
Table 2.4 Referenced enzyme and substrate abbreviations ................................69
Table 2.5  Library of alternative substrates: observed activity, inhibition, and calculated oxidation potentials for analogues of PCA ........................................71
Figure 2.10  Structures of molecules used for substrate specificity and inhibition studies .................................................................................................................................72

Chapter 3. Exploring allosteric activation of LigAB from Sphingobium sp. strain SYK-6 through kinetics, mutagenesis, and computational studies

Figure 3.1  Allosteric feedforward activation found in central metabolism .................80
Scheme 3.1  Abridged lignin derived aromatic compound metabolic pathway from Sphingobium sp. SYK-6 demonstrating the metabolic connection between vanillin, protocatechuic acid, and LigAB ...........................................................................81
Table 3.1  Primers used in the site-directed mutagenesis of LigAB .........................83
Figure 3.2  Chemical structure comparison of the physiological LigAB-WT substrate (PCA) and observed enhancers of LigAB-WT dioxygenation rate .................93
Table 3.2  Library of alternative substrates: observed activity, inhibition, and activation .................................................................................................................................94
Figure 3.3  Steady-state kinetic enhancement and inhibition of turnover by LigAB-WT and mutants in the presence of vanillin ........................................97
Table 3.3  Percentage enhancement or inhibition of turnover by LigAB-WT and mutants in the presence of vanillin ........................................98
Figure 3.4  Normalized steady-state kinetics of LigAB-WT in the presence and absence of vanillin .................................................................99
Table 3.4  Steady-state kinetic parameters for LigAB-WT in the presence and absence of vanillin .................................................................99
Figure 3.5  Oxygen consumption reaction progress curves for LigAB with PCA, gallate, and 3OMG in the presence and absence of vanillin .................100
Figure 3.6  LigAB-WT enzyme inactivation rates ($k_i$) in the absence and presence of vanillin .................................................................101
Figure 3.7  Shotgun computational docking of vanillin to the LigAB-WT crystal structure .................................................................102
Figure 3.8  View of the LigAB-WT active site and allosteric binding site ................103
Figure 3.9  Changes in the intrinsic fluorescence spectrum of LigAB by PCA and vanillin .................................................................103
Figure 3.10  Steady-state kinetics of LigAB mutants with PCA ..................................106
Table 3.5  Steady-state kinetic parameters of LigAB Phe103α mutants .................106
Figure 3.11  Comparison of intrinsic fluorescence spectra of LigAB-WT and the LigAB-A18W mutant in the presence of PCA and vanillin .................109
Scheme 3.2  Proposed catalytic mechanism of LigAB ..........................................111
Chapter 4. Improving alternate substrate utilization of LigAB from Sphingobium sp. SYK-6 through site directed mutagenesis

Figure 4.1 Computational docking of substrates onto LigAB and Phe103α mutants ................................................................. 121
Table 4.1 Steady-state kinetic parameters for LigAB-WT and Mutants with PCA, Gallate, and 3OMG ................................................................. 123
Figure 4.2 Steady-state kinetics of LigAB Phe103α mutants ................................................................. 125
Figure 4.3 Comparative fitting of substrate inhibition in F103H ................................................................. 126
Figure 4.4 Internal cavities of LigAB ................................................................. 128
Table 4.2 Percent enhancement or inhibition of turnover by vanillin ................................................................. 129
Figure 4.4 Plots of steady-state kinetic enhancement and inhibition in F103T and F103H ................................................................. 129
Figure 4.5 Two dimensional representation of the LigAB active site with PCA bound ................................................................. 132

Chapter 5. Design and synthesis of FRET-pair labeled lignin model compounds

Figure 5.1 Examples of lignin model compounds ................................................................. 136
Scheme 5.1 Retrosyntheses of FRET labeled lignin model compounds ................................................................. 138
Scheme 5.2 FRET-pair labeled lignin model compound concept ................................................................. 139
Figure 5.2 Chemical structures of TAMRA and FAM ................................................................. 140
Scheme 5.3 Synthesis of (4R*,5S*)-4-[4-(4-Benzylxyloxy-3-methoxyphenyl)-2,2-dimethyl-1,3-dioxan-5-yloxy]-3-methoxybenzaldehyde (18) ................................................................. 141
Scheme 5.4 Synthesis of (4R*,5S*)-4-(4-hydroxy-3-methoxyphenyl)-5-[4-(tert-butyldimethy1silyloxy)methyl]-2-methoxyphenoxy]-2,2-dimethyl-1,3-dioxane (102) from 18 ................................................................. 142
Scheme 5.5 Synthesis of ester-linked FRET-pair labeled lignin model compounds ................................................................. 143
Scheme 5.6 Synthesis of click-chemistry linked FRET-pair lignin model compounds ................................................................. 144
Scheme 5.7 Synthesis of FAM labeled control compounds ................................................................. 146
Figure 5.3 Proton and carbon labeling diagram for NMR assignments of compounds ................................................................. 149
# Complete List of Abbreviations and Definitions

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>3OMG</td>
<td>3-O-methyl gallic acid</td>
</tr>
<tr>
<td>AR5</td>
<td>Fifth Assessment Report (of the IPCC)</td>
</tr>
<tr>
<td>CHMS</td>
<td>4-Carboxy-2-hydroxymuconate-6-semialdehyde</td>
</tr>
<tr>
<td>DHBAm</td>
<td>3,4-Dihydroxybenzamide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EDO</td>
<td>Extradiol dioxygenase</td>
</tr>
<tr>
<td>GA</td>
<td>Gallic acid, gallate</td>
</tr>
<tr>
<td>GtC</td>
<td>Gigatonnes of carbon (10^{15}) g</td>
</tr>
<tr>
<td>IPCC</td>
<td>Intergovernmental Panel on Climate Change</td>
</tr>
<tr>
<td>LDAC</td>
<td>Lignin derived aromatic compound</td>
</tr>
<tr>
<td>mbbl/d</td>
<td>Thousands of barrels per day</td>
</tr>
<tr>
<td>OPEC</td>
<td>Organization of the Oil Exporting Countries</td>
</tr>
<tr>
<td>ppm</td>
<td>Parts per million</td>
</tr>
<tr>
<td>PCA</td>
<td>Protocatechuic acid, 3,4-dihydroxybenzoic acid</td>
</tr>
<tr>
<td>PCAD</td>
<td>Protocatechuate dioxygenase</td>
</tr>
<tr>
<td>PVC</td>
<td>Polyvinyl chloride</td>
</tr>
<tr>
<td>syngas</td>
<td>Synthesis gas</td>
</tr>
<tr>
<td>VOC</td>
<td>Vicinal Oxygen Chelate</td>
</tr>
<tr>
<td>yr.</td>
<td>Year</td>
</tr>
<tr>
<td>CDCl₃</td>
<td>Chloroform-d, deuterochloroform</td>
</tr>
<tr>
<td>DCC</td>
<td>N, N-dicyclohexylcarbodiimide</td>
</tr>
<tr>
<td>DCM</td>
<td>Dichloromethane, methylene chloride</td>
</tr>
<tr>
<td>DMAP</td>
<td>4-Dimethylaminopyridine</td>
</tr>
<tr>
<td>DIPEA</td>
<td>N, N-diisopropylethylamine</td>
</tr>
<tr>
<td>ESI-MS</td>
<td>Electron spray ionization – mass spectrometry</td>
</tr>
<tr>
<td>EtOAc</td>
<td>Ethyl acetate</td>
</tr>
<tr>
<td>EtOH</td>
<td>Ethanol</td>
</tr>
<tr>
<td>Et₂O</td>
<td>Diethyl ether</td>
</tr>
<tr>
<td>FAM-N₃</td>
<td>5-carboxyfluorescein azidotetramethylrhodamine carboxamide</td>
</tr>
<tr>
<td>FAM-NHS</td>
<td>5(6)-carboxyfluorescein succinamidyl ester</td>
</tr>
<tr>
<td>MeCN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>MEM-Cl</td>
<td>2-methoxyethoxymethyl chloride</td>
</tr>
<tr>
<td>MeOH</td>
<td>Methanol</td>
</tr>
<tr>
<td>MHz</td>
<td>Megahertz</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass to charge ratio for mass spectrometry</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance spectroscopy</td>
</tr>
<tr>
<td>PPTS</td>
<td>Pyridinium para-toluene sulfonic acid</td>
</tr>
<tr>
<td>p-TsOH</td>
<td>para-Toluene sulfonic acid</td>
</tr>
<tr>
<td>TAMRA-N₃</td>
<td>5-carboxy-5-carboxytetramethylrhodamine azidotetramethylrhodamine carboxamide</td>
</tr>
<tr>
<td>TAMRA-NHS</td>
<td>5(6)-carboxy-5-carboxytetramethylrhodamine succinamidyl ester</td>
</tr>
<tr>
<td>TBAF</td>
<td>Tetrabutylammonium fluoride</td>
</tr>
<tr>
<td>TBDMS-Cl</td>
<td>tert-Butyl dimethylsilyl chloride</td>
</tr>
<tr>
<td>THF</td>
<td>Tetrahydrofuran</td>
</tr>
</tbody>
</table>
Chapter 1:

A Perspective on the Historic Use of Fossil Fuels, and an Introduction to Lignin and the Future of Bio-derived Alternatives
“But, whatever our resources of primary energy may be in the future, we must, to be rational, obtain it without consumption of any material. Long ago I came to this conclusion, and to arrive at this result only two ways, as before indicated, appeared possible—either to turn to use the energy of the sun stored in the ambient medium, or to transmit, through the medium, the sun's energy to distant places from some locality where it was obtainable without consumption of material.”

- Nikola Tesla, The Century Magazine, 1900

“We are like tenant farmers chopping down the fence around our house for fuel when we should be using Nature’s inexhaustible sources of energy – sun, wind and tide. ... I’d put my money on the sun and solar energy. What a source of power! I hope we don’t have to wait until oil and coal run out before we tackle that.”

- Thomas Edison, 1931
Anthropologists believe that humans have had the knowledge and ability to regularly control fire for nearly three-hundred and fifty thousand years. Bio-based sources of carbon were the basic primary fuel source for the fires that have heated, cooked, and smelted throughout history. However, in roughly the past two hundred and fifty years, an eye blink on the timeline of hominid existence, the primary fuel source has changed. Fossil fuels extracted from the Earth have come to dominate the fires of industry and homes around the world. Non-stop worldwide use of coal, petroleum, and natural gas has led to over-reliance on these fuel sources not only for energy but also for raw materials. Because of this, humanity as a whole has in many ways forgotten how to produce the energy we need from what is provided renewably by Nature. As the effects of fossil fuel usage on the planet become more apparent, a re-investment and re-education in bio-based renewable energy and commodity chemicals are needed.

Petroleum, also known as crude oil, is often used as a representative of all fossil fuels due to its ubiquity and extent of use, providing the largest portion of human energy consumption (Figure 1.1). In addition to its high energy value, petroleum provides a wide variety of hydrocarbons and organic compounds. Chiefly known as the source of liquid fuel for home heating oil and automobiles, petroleum also provides many of the chemical building blocks (petrochemicals) used in chemical research and the manufacture of everyday commodities. The synthesis of many modern antibiotics and pharmaceuticals may not have been possible without these readily available petrochemical building blocks. Polymers such as polyester, polyethylene, polypropylene, polyvinyl chloride (PVC), and polystyrene found in many modern luxuries would likely not be as prevalent or may possibly not even exist without the refining of their respective monomers from petroleum. That is not to say that these commodities...
(plastics, pharmaceuticals, fuels, etc.) could never have been discovered in the absence of petrochemicals, but their equivalents derived from natural sources would have required even more human ingenuity.

1.1. Global Climate Change.

There is no doubt that harnessing energy and chemicals provided by petroleum and other fossil fuels has changed the course of human existence. However, extraction and combustion of fossil fuels by humans has also dramatically altered the natural cycle of carbon flux on Earth; several billion tons of carbon (~3.75 billion metric tons since yr. 1780), previously locked away deep underground in the form of hydrocarbons, have been returned to the atmosphere and oceans as carbon dioxide (CO$_2$). Effects from the anthropogenic inflation of the natural atmospheric carbon dioxide concentration have begun to manifest in the form of global climate change.

Initially observed as a steady rising of the average global temperature and an increased pace of glacial and polar ice cap receding, termed previously by many as Global Warming, climate change has already shown additional broader reaching implications beyond warmer temperatures. Shifting seasonal boundaries (beginnings and endings), more pronounced and

Figure 1.1. A) Total global energy consumption (quadrillions of BTUs, blue) of all energy sources with the percent contribution to that total from petroleum (red), coal (dark grey), natural gas (yellow), and the combination of all other sources (green). B) Total global energy consumption and the energy contribution (quadrillions of BTUs) to that total from petroleum, coal, natural gas, and all other energy sources. Data retrieved from Ref. 2.
longer periods of drought and flooding, increased rate of glacial melt, ocean level rise, and increased frequency of high intensity weather events are a few of the macroscopic long term observable changes occurring around the world.\textsuperscript{4, 5}

Global climate change has been a continuing topic of deeply divided political debates for well over a decade. Is climate change occurring? Is climate change anthropogenic (of human origin)? What, if anything, can or should be done? Yet among the scientific community these questions have largely been answered due to overwhelming data that support the occurrence of climate change and the role humans have played in its propagation. Two large scale studies, conducted by panels of hundreds of scientists, have produced the most detailed reports ever assembled on global climate change. In 2013, the Intergovernmental Panel on Climate Change (IPCC), a scientific body formed by the United Nations, released its Fifth Assessment Report (AR5), a comprehensive analysis and discussion of all data related to climate change, its impact on humanity, and steps that should be taken to mitigate impact and further exacerbation. The AR5 was followed a year later by the third U.S. National Climate Assessment report, a report funded directly by the United States government. Both reports came to the same conclusion: climate change is occurring and is directly linked to human activities.\textsuperscript{3-5}

The Physical Science Basis within the AR5 outlines the key contributing factors, or drivers, that have brought about what we now see manifested as global climate change. While not all of the drivers of climate change will be discussed here, the driver determined to have the largest impact on climate change, radiative forcing from greenhouse gases,\textsuperscript{3} is inextricably linked to the guiding principles and purpose of the research outlined in this document.

As defined by the IPCC, a radiative force is an external perturbation that results in a net change in the energy balance of the Earth.\textsuperscript{3} By this definition, this force can result in either warming (e.g., greenhouse gases) or cooling (e.g., reflecting incident light off Earth’s surface),
and can be naturally or anthropogenically derived. In the case of atmospheric greenhouse gases, the warming of the atmosphere is caused by the gases’ absorption of infrared radiation (heat) reflected from the Earth’s surface that would have otherwise been lost. The gases determined to have the largest greenhouse/radiative force effect are CO₂ and methane (CH₄), and atmospheric concentration increases of both gases can be directly attributed to fossil fuel extraction and use.\(^3\)

Carbon dioxide is a naturally occurring gas in the atmosphere, produced by volcanic eruptions, forest fires, and cellular respiration. It is the source of carbon which terrestrial plants chemically combine with water using the energy of sunlight to produce glucose and molecular oxygen – a process without which much of life could not be sustained. Despite its importance and natural occurrence, anthropogenic release of CO₂ by the burning of fossil fuels has increased to a point such that natural carbon sinks such as plant growth cannot offset the rate of production. The IPCC estimates that, from the year 1750 (the beginning of the Industrial Revolution) to 2012, approximately 555 gigatonnes of carbon (GtC, 1 GtC = \(10^{15}\) g) have been released as CO₂ into the atmosphere by human activities. Of that, approximately 375 GtC are of fossil fuel origin, and more than half of the total anthropogenic carbon release (240 GtC) has accumulated in the atmosphere while the remainder has largely been absorbed by the oceans.\(^3\) The result is a 112.5 ppm increase in the atmospheric CO₂ concentration over the past 260 years (278 ppm [yr. 1750] vs 390.5 ppm [yr. 2012]). Laying additional credence to these data are the observations that the atmospheric CO₂ concentration increases with increasing latitude north of the equator (correlating to increased development and fossil fuel consumption of nations in the Northern hemisphere), and a decrease in the atmospheric CO₂ isotopic ratio of \(^{13}\)C to \(^{12}\)C,\(^3\) a ratio that would not show large variation without external influence.
There is no question that the extraction and use of fossil fuels have been a boon to societal development and technological advancement. However, it is likely that the long term effects of constant, prolonged, and ever expanding global use of fossil fuels for energy and chemicals were never envisioned by pioneers of industry. Natural gas, coal and petroleum were simply energy rich natural resources which could be easily obtained and provided an advantage over less energy-dense bio-based sources (wood). Unfortunately, the unforeseen effects are now being seen as climates change around the world, and while fossil fuels will remain an integral part of society for the foreseeable future we can no longer ignore the global environmental price that is paid for their use. Instead, the technologies fossil fuels made possible should be used to find and advance the next generation of energy and chemical feedstocks to replace fossil fuels. A cause to which this research hopes to contribute.

1.2. Returning to our roots. Achieving energy independence is a goal that has been tossed around by politicians in the United States since the Organization of the Petroleum Exporting Countries (OPEC) oil embargo in 1973. However, the exact goals of energy independence have never had a cohesive focus from one Presidential administration to the next. Under President Carter, billions of dollars was funneled into renewable-energy research programs. These programs were subsequently dismantled by President Reagan. This trend of renewable-energy funding ebbs and flows has continued to present day, and presents possibly the largest barrier to significant progress in the development and implementation of practical (economically viable and scalable) alternatives to fossil fuels in an impactful timeframe. A United States that is no longer dependent on foreign oil imports is an admirable goal; however, this policy strikes more at the heart of National (U.S.) security and economics than environmental and climate issues as it tends to promote increased domestic fossil fuel
production over advancing alternatives. Arguably a more worthy goal, though admittedly loftier, is to also achieve fossil fuel independence. By utilizing a diversity of renewable energy (i.e., solar, hydroelectric, wind, bio-derived liquid fuels) and bio-based chemicals, we would effectively be returning to a pre-Industrial Revolution state of energy and chemical feedstock sources.

While the return to bio-based feedstocks is currently in its infancy, chemical producers have been some of the earliest large scale adopters of bio-based products due to a vastly improved understanding of sustainability, and also reduced availability and increased cost of petrochemical feedstocks caused principally by global political instability. While currently only 2 % of the global chemical industry, the bio- and renewable chemical industry is expected to grow to over 20 % of the total market in the next decade. Bio-engineering of microbial biochemical pathways has allowed the large scale production of commodity chemicals (solvents and starting materials) as well as natural products for use as drugs in the treatment of disease (e.g. Artemisinin).

Unlike renewable energy feedstocks, fossil fuels are – by definition – not renewable. Their formation by geological processes over millions of years is well beyond the renewable timeframe associated with biomass or bio-produced feedstocks and chemicals. While large quantities of coal, petroleum, and natural gas have been discovered and have fueled the growth of society for the better part of the last two centuries, the reserves are finite. Estimates of proved petroleum reserves (i.e. the total

Figure 1.2. Proved petroleum reserves continue to increase despite every increasing consumption. Data retrieved from Ref 2.
recovered resource will exceed the stated value with greater than 90% certainty\textsuperscript{15}, however, continue to increase (Figure 1.2).\textsuperscript{2} Exact values for worldwide proven oil reserves are inherently inaccurate due to the widespread belief that OPEC members overestimate their reserves in order to receive higher production quotas, and the practice of not adjusting reserve values to account for oil that has already been extracted.\textsuperscript{16}

The uncertainty in petroleum reserve totals therefore propagates into calculations of peak oil production – the determination of when conventional petroleum production will hit a maximum and subsequently begin to decline.\textsuperscript{17,20} Despite the inaccuracy of the data provided to them, most scientists studying peak oil production have been expecting a peak to occur in the near term, between 2010-2020.\textsuperscript{18} Declining production of conventional oil with current global demand will ultimately lead to higher prices, and a shift towards the development of petroleum sources that are more difficult (less accessible or require more energy intensive extraction processes) to develop (i.e. shale oil and tar sands).\textsuperscript{16}

Further complicating the situation is the continual growth in worldwide energy demand (observable in Figure 1.1) due to booming economies in emerging markets such as China and South Korea. China has recently over-taken the United States as the largest consumer of energy in the world (both nations individually consume more energy than all European nations combined) (Figure 1.3).\textsuperscript{2} The rate of energy consumption in China has increased nearly five and a half times since 1980. The rate of Chinese energy consumption increased dramatically post-2002. United States (blue), all European nations combined (green), China (dark red – 1980 to 2002, red – 2002 to 2010, India (orange), and South Korea (yellow). Dashed lines show the projected energy consumption of China pre- and post-2002. Data retrieved from Ref. 2.
fold (2002-2011) from the consumption rate pre-2002 (1980-2002), while U.S. consumption has begun to decrease in recent years.\textsuperscript{2} This new period of economic growth in developing nations has been fueled by the same consumption of cheap and easily obtainable sources of fossil fuel that supported economic growth in the United States and other currently developed nations. While the United States is beginning the process of weaning itself from fossil fuels and attempts to be more efficient in the consumption of fossil fuels that are used (petroleum and coal consumption are declining while natural gas consumption has begun increasing), worldwide consumption of petroleum, coal, and natural gas has continued to grow steadily over the past thirty years (Figure 1.4).

Despite the increasing energy demand, energy culture is changing and there has been growing focus on returning to renewable feedstocks for energy and chemicals.\textsuperscript{21} However, unlike the significant shifting towards renewables observed in the commodity chemical industry,\textsuperscript{7} when applied to the energy industry (especially liquid fuels dominated by transportation consumption), this statement is idyllic. In the U.S., there has not been a shift in fossil fuel consumption that would suggest replacement by or a “shift back” to renewable alternatives despite increasing public demand. Reduction in coal consumption is being largely offset by an increase in natural gas consumption, and there is little to suggest that direct replacement of petroleum by biomass derived alternative fuels (biofuels) is occurring. In 2011, approximately 899 thousand barrels of biofuel were consumed per day (mbbl/d).
in the United States (twenty-one times less than the amount of petroleum consumed); however, petroleum consumption decreased by only 231 mbbl/d from 2010 to 2011 according to data from the U.S. Energy Information Agency - a decrease of less than 2 %. As such, increased biofuel usage is likely not the largest contributing factor to the modest decrease in U.S. petroleum consumption. Instead, more stringent Federal standards imposed on vehicle manufacturers for emissions and efficiency, growing popularity of more fuel efficient vehicles amongst the public, economic market crashes (i.e. 2008), and increasing gasoline and fuel (heating) oil prices are all interconnected and all likely contribute more significantly to the reduction in United States petroleum consumption in recent years. Currently, it would seem, the demand for liquid fuel is not being replaced but instead supplemented by alternatives – few of which have the economic and infrastructure advantages of petroleum. I posit that without substantial increases in funding for research and infrastructure, and further cultural changes on a global level that petroleum and other fossil fuels will remain the dominant sources of energy for generations.

That is, at least, until we run out or the use of fossil fuel becomes too costly both economically and environmentally.

Fossil fuel based industries are highly entrenched in the world economy, and because of their status and “necessity”, the difficultly lies in the ability to envision a path to renewable energy and chemical sources that could lead to a significant reduction in the use of natural gas, coal and/or petroleum as energy sources. Switching primary energy sources to a renewable alternative sounds like a daunting task; however, there are several examples of nations who have attempted and succeeded in large scale transitions away from fossil fuels. France made a large scale transition to sustainable nuclear energy to reduce reliance on imported petroleum. Forty-three percent of France’s total energy consumption is now accounted for by nuclear
energy. While the United States produces more nuclear energy, it only accounts for 8% of total consumption.\textsuperscript{22} As well, in response to oil crises in the 1970’s and the current spiking of oil prices, Brazil has focused on large scale production of renewable bio-ethanol from sugarcane (Brazil’s largest crop) to replace petroleum.\textsuperscript{22} Support from government subsidies and incentives for sugarcane growers, ethanol producers, and consumers made the transition widely accepted and successful, though not without environmental impact from increased clearing of tropical/rain forests.\textsuperscript{22} These examples show that a large scale transition away from non-renewable primary fuel sources (petroleum) towards renewable energy is indeed possible, but require both public and government support.

1.3. The Case for Renewables. Renewable energy and chemicals are derived from resources that can be regenerated on the human time scale (tens of years or less) through either human cultivation or natural cycles. Fossil fuels may be considered renewable to some degree, forming from biomass over millions of years (the geological timescale). However, this cycle is unreliable for our needs. “Renewable” resources such as solar, wind, hydroelectric, tidal, or geothermal are able to produce sustainable energy by harnessing natural processes that are replenished by natural cycles. Yet, reliable access to these sources of energy is often dependent on global locale, the time of year and even the weather. Additionally, long term storage of energy derived from these sources is difficult. However, biomass (biological material derived from or generated by living organisms) is nearly ubiquitous and is a renewable resource in the truest sense – requiring growth, use, and regrowth. Additionally, biomass can be used as both an energy (biofuel) and chemical feedstock.

Energy in biomass is found in the same form as in fossil fuels – as chemical bonds. The breaking and rearranging of these bonds by combustion of the material releases the energy as
heat. Similarly, rearranging the bonds of molecules found in biomass with more finesse can allow for the production of a rich diversity of chemicals with consumer and industrial relevance. Biomass consists of a large and diverse collection of molecules of which not all are suitable for use as fuel sources or starting materials (feedstocks) from which to produce fuel. Energy rich carbon sources such as fatty acids, sugars, and polymers (starch, cellulose, and lignin) that are biosynthesized in large quantities are ideal as molecular feedstocks for the production of fuels. However, there are many other biosynthetic (enzymatic) pathways that can be utilized to mass produce bio-derived chemicals of interest for use as solvents (e.g., acetone and ethanol)\(^{23}\) and starting materials for plastics (e.g., 3-hydroxybutyrate used to produce a polyester).\(^{24}\) Both a solvent and fuel among other uses, ethanol has been produced for thousands of years by harnessing the ability of yeast to convert sugars to the simple alcohol.

Biofuels have historically been categorized into two groups – primary and secondary. Inarguably the oldest form of human controlled energy production, a primary biofuel is one where the biomass is unmodified and used in its natural form as fuel (e.g., wood, animal waste, forest litter and crop residues, some plant and animal oils, and landfill gas).\(^{25}\) Lower energy density and efficiency with which they burn (dirtier, producing particulates such as soot and ash) limits the use of primary biofuels in most modern applications. Despite this, many primary biofuels continue to play an important energy role around the world, especially in nations with significant rural or destitute populations. Secondary biofuels begin their cycle as primary biofuels (biomass), but are converted or modified through chemical and biochemical reactions to new forms depending on the desired fuel application. While the majority of primary biofuels are found as solids, secondary biofuels can be solid (charcoal), liquids (biodiesel, methanol, ethanol, etc.), or gas (syngas and hydrogen).\(^{25}\) Regardless of the chemical state, a secondary biofuel has undergone a transformation from the original biomass source. Secondary biofuels
are also further classified into generations (1\textsuperscript{st} to 4\textsuperscript{th}). Despite being classified into “Generations”, the term is somewhat misleading as there is no chronology associated to a higher-order “generation” vs. one of lower-order. Additionally, there is no inherent superiority of one generation over another.\textsuperscript{26} Instead, “generations” are classified by the biomass source and technology used in fuel production;\textsuperscript{25} the type, complexity, ease of extraction, or environmental impact of the fuel that is produced does not contribute to Generational classification.

First generation biofuels are defined as those that are derived from simple sugars, starches, or oils obtained from terrestrial food crops.\textsuperscript{25} Food crops (corn, sugarcane, soy, etc) are highly accessible, and the sugars that they provide are easily broken down by yeast or other microorganisms. The sugar, starch, and oils, however, are usually only a small portion of the total biomass (e.g., starch from corn kernels, and oil from soy beans).\textsuperscript{25, 27} Bioethanol (produced by the fermentation of starches and grains) and bio-diesel (produced by the transesterification of vegetable oil triglycerides with an alcohol) are the two main examples of first generation biofuels.\textsuperscript{25, 27, 28} The majority of the plant goes unused in the production of primary biofuels, and the remainder consists of the plant’s lignocellulosic structural material. Regardless, first generation biofuels are currently the most widely produced and distributed of all biofuel generations.\textsuperscript{27, 28} However, their production is controversial due the large percentage of biomass that is unused (reducing the energy yield per area of crop), the energy and water intensity required for the biomass growth and harvest (especially corn), and competition with food resources.\textsuperscript{25, 26, 28} All of which contribute to increased cost, and debatable efficacy towards the goal of reducing CO\textsubscript{2} emissions.

Production of second generation biofuels depends highly on the modification of either the feedstock biomass (e.g., genetic engineering of the biomass source to reduce lignin content).\textsuperscript{29}
the method of processing the material (e.g., genetic engineering of microbes to more efficiently degrade and convert lignocellulosic material), or a combination of both. Cellulose, long polymers of β(1→4) linked glucose, and lignin, a highly heterogeneous aromatic polymer, are the main components of the lignocellulosic material that goes unused in the production of first generation biofuels. These materials are more recalcitrant (resistant to degradation) causing their use as feedstocks for the production of biofuels to be more difficult. However, they are also the first and second most abundant (respectively) biologically produced polymers on Earth, making them highly valuable feedstocks for biofuel production. Second generation biofuel production seeks to use all of the biomass components: sugars, starches, oils, and lignocellulosic material. In order to remove the food vs. fuel competition, inedible sources of lignocellulosic material such as trees and grasses can be grown specifically for the purpose of biofuel production.

First and second generation biofuels are produced from terrestrial biomass sources and the biomass is often converted to its secondary form by microbes. However, in third generation biofuels, the microbial biomass (bacteria, yeast, algae) is harvested to obtain fatty acids which can be converted to biodiesel. While the third generation of biofuel technologies harvests the microbes as the fuel source, microbes still require a carbon source in order to grow and produce the desired fatty acids. When using microbes such as bacteria (Escherichia coli) and yeast the carbon source is lignocellulosic biomass or waste from food crop processing. Use of food crop sugars and starches as a carbon source would (or at least should) result in a first generation classification. Unfortunately, industrial scale production of biodiesel often uses food crop derived sugars as a carbon source in order to maintain consistency and reproducibility of growth conditions in order to maximize yield and cost effectiveness. Algae is an alternative to bacteria and yeast (which require a biomass carbon source) since it produces sugars and fatty
acids using CO$_2$ sequestered via photosynthesis.$^{25, 27, 34}$ Regardless of the carbon source used as a feedstock, all microbes used in biofuel or chemical production require external nutrients in order to grow effectively (especially nitrogen and phosphorus), a serious concern and expense for industrial scale production. Despite this consideration, these technologies do not compete with food crops for arable land due to the aquatic nature of the organisms.

Having multiple definitions, the fourth generation is the most loosely defined of the secondary biofuel classes. Researchers in both academia and industry seemingly define this generation to suit their needs - ranging from the use of carbon negative microbes$^{27}$ (consumes more CO$_2$ during production than is emitted during consumption), to the genetic engineering of algae, microorganisms and biomass sources,$^{31, 35}$ to production of bio-hydrogen and bioelectricity. $^{36}$ Due to the lack of a strict definition based on a biomass feedstock distinction (as is the case for first, second, and third generation biofuels), the term “fourth generation biofuel” can essentially be thought of as a buzz word.

The research conducted herein aims to contribute to the understanding of microbial lignocellulosic biomass degradation and metabolism, and so potentially enhance methods used in the production of second and (some) third generation biofuels, and commodity chemicals. Of lignocellulosic material’s main components, the microbial and enzymatic degradation of cellulose to glucose has been well studied; cellulose derived biofuel production is on the way to becoming an industrially relevant technology.$^{37, 38}$ However, lignin degradation chemistry and subsequent metabolism of lignin degradation products are less well characterized. A greater understanding of the depolymerization and metabolism of lignin could lead to more efficient use of this highly carbon rich material - transforming what was once a waste product into a value added commodity.
1.4. A Deeper Understanding of Lignin. As the second most abundant biopolymer and accounting for 30% of all biomass carbon,\textsuperscript{39, 40} one would rightly suspect that lignin plays an important biological role. Lignification of plants is believed to be a key distinction between terrestrial and aquatic species,\textsuperscript{41-44} however, lignin has recently been discovered in cells of an intertidal red algae.\textsuperscript{45} While this may suggest that lignin is a convergently evolved structure or remained conserved over some 1 billion years,\textsuperscript{45} lignin is found \textit{almost} exclusively in terrestrial plants, and depending on the species lignin can compose 13-35% of the plant’s mass.\textsuperscript{46, 47} In general, softwood plant species contain the highest lignin content (25-35%) followed by hardwoods (18-25%) and grasses (17-24%).\textsuperscript{46} Found intertwined with the cellulose fibers of terrestrial plants, lignin is the source of key crosslinks between secondary cell walls that gives plants the structural integrity and rigidity necessary to grow vertically.\textsuperscript{39, 45} In addition to structural support, lignin plays a key role in the transport of water throughout plants.\textsuperscript{39} The majority of a plant’s structure is composed of hydrophilic (water loving/attracting) polysaccharides, but lignin differs in that its structure is highly aromatic and therefore highly hydrophobic (water fearing/repelling). Without lignin, most of the plant’s molecular structure would attract water and therefore restrict efficient water movement to a plant’s periphery. Lastly, lignin’s heterogeneous structure creates a recalcitrant defensive barrier that helps prevent the degradation of cellulose and other polysaccharides by hungry or infectious organisms.\textsuperscript{48}

Proficient as a natural barrier, lignin’s heterogeneous structure is also a major industrial obstacle. The primary source of lignin production is the production of wood pulp and paper where lignin is separated from cellulose fibers in order to prevent the discoloring of the final paper product. Additionally, lignin is removed during the production of cellulosic biofuels due to the inability of the microbes to utilize the material, and the tendency of lignin and its
breakdown products to inhibit various metabolic steps in the fermentation of cellulosic sugars – resulting in reduced efficiency.\(^\text{49, 50}\) Despite high natural abundance and industrial scale production, lignin is largely underutilized as a feedstock for the commercial production of commodity chemicals and value added products. The vast majority of the 50 million tons of lignin produced worldwide is used as boiler fuel by pulp and paper mills or discarded.\(^\text{32, 46, 47, 51, 52}\)

\(^\text{51, 52}\) Only about 2 % (1,000,000 tons) of total isolated lignins is utilized commercially,\(^\text{32, 46, 51, 52}\) and mainly for the production of low value chemicals (Table 1.1).

**Table 1.1. Lignin Derived Products and Applications.**

<table>
<thead>
<tr>
<th>Lignin Type</th>
<th>Products and Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kraft Lignin</td>
<td>• Incineration for energy/steam</td>
</tr>
<tr>
<td></td>
<td>• Fertilizer and pesticide carrier</td>
</tr>
<tr>
<td></td>
<td>• Carbon fiber</td>
</tr>
<tr>
<td></td>
<td>• Blending with thermoplastic polymers</td>
</tr>
<tr>
<td></td>
<td>• Binders</td>
</tr>
<tr>
<td></td>
<td>• Resins</td>
</tr>
<tr>
<td></td>
<td>• Ion-exchange resins</td>
</tr>
<tr>
<td></td>
<td>• Activated carbon</td>
</tr>
<tr>
<td></td>
<td>• Production of low MW compounds</td>
</tr>
<tr>
<td></td>
<td>• vanillin</td>
</tr>
<tr>
<td></td>
<td>• hydroxylated aromatics</td>
</tr>
<tr>
<td></td>
<td>• quinones</td>
</tr>
<tr>
<td></td>
<td>• aliphatic acids</td>
</tr>
<tr>
<td>Soda Lignin (sulfur free)</td>
<td>• Phenolic resins'</td>
</tr>
<tr>
<td></td>
<td>• Dispersants'</td>
</tr>
<tr>
<td></td>
<td>• Animal Nutrition'</td>
</tr>
<tr>
<td>Lignosulfonates</td>
<td>• Animal feeds</td>
</tr>
<tr>
<td></td>
<td>• Particleboard</td>
</tr>
<tr>
<td></td>
<td>• Detergents/Surfactants</td>
</tr>
<tr>
<td></td>
<td>• Glues/Adhesives</td>
</tr>
<tr>
<td></td>
<td>• Cement additive</td>
</tr>
<tr>
<td></td>
<td>• Colloidal suspension stabilizers</td>
</tr>
<tr>
<td></td>
<td>• Dispersing agents</td>
</tr>
<tr>
<td></td>
<td>• Drilling agent binders</td>
</tr>
<tr>
<td>Organosolv Lignins</td>
<td>• Filler in inks, varnishes, and paints</td>
</tr>
<tr>
<td>Hydrolysis Lignin</td>
<td>• Absorbent material (sorbent)</td>
</tr>
<tr>
<td>Ionic Liquids Lignin</td>
<td>* Industrial scales not yet achieved.</td>
</tr>
</tbody>
</table>

\(^\dagger\) High purity and biocompatibility are necessary for these applications

Retrieved from Ref. 51.
1.4.1. Lignin Structure and Biosynthesis. The complex molecular structure of lignin is, in part, a result of the building blocks (monomers) used in the construction of the polymer. In order to find new and creative ways to utilize lignin by breaking it down, there must first be an understanding of the chemical and biochemical origins of lignin synthesis and the resulting polymer structure. As mentioned above, cellulose is a simple polymer composed of a single basic repeating subunit, glucose, that is chained together nearly exclusively by one bond type, a β(1→4) linkage. A lignin polymer, however, results from the polymerization of three primary 4-hydroxyphenylpropanoid monomers (monolignols): p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol (Figure 1.5.A).\textsuperscript{40, 46, 53-55} The combination of these three basic units results in three subunit types within the lignin. Due to a high degree of modification during polymerization, the subunits can be distinguished by the number of methoxy substituents on the aromatic ring: p-hydroxyphenyl (H) units have no methoxy groups, guaiacyl (G) units have one methoxy group adjacent (ortho) to the phenolic oxygen (i.e. C3 or C5), and syringyl (S) subunits have two methoxy groups flanking the phenolic oxygen (i.e. C3 and C5) (Figure 1.5.B).\textsuperscript{55} Simply increasing the number of unit types dramatically increases the complexity of lignin beyond that of cellulose.

The structure of lignin is also highly dependent on the species of plant,\textsuperscript{46} the plant tissue,\textsuperscript{56} which monolignols are used, and the ratio of their incorporation (Table 1.2). In softwood plant species, the vast majority of the lignin structure is composed of coniferyl alcohol (G subunits),

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure15.png}
\caption{Lignin structural units. A) Primary 4-hydroxyphenylpropanoid monomers (monolignols) incorporated into lignin polymers. B) Basic structural units found within lignin polymers. Dashed bonds represent potential sites of connection to adjacent subunits.}
\end{figure}
with sinapyl alcohol (S subunits) incorporated in about 10% of the structure. Hardwoods on the other hand produce lignin that is composed of roughly equivalent parts of these two monolignols. Grasses utilize all three primary monolignols, though coniferyl alcohol is again the primary building block. While coniferyl, sinapyl, and p-coumaryl alcohols are the primary monolignols, other phenolic compounds can also be incorporated in lignin formation.

Additionally, instead of resulting in an inability to biosynthesize lignin, genetic mutations affecting the biosynthesis of the primary monolignols can result in altered lignin composition – demonstrating an adaptive ability of plants to utilize alternative 4-hydroxyphenylpropanoids, mainly monolignol precursors.

The variable monolignol content of lignin would be enough to complicate the structure of lignin even if the building blocks were connected to one another by a small set of specific bonding motifs. However, lignin is formed (lignification) through radical (unpaired electron) coupling reactions between monolignol radicals and radicals located on the growing lignin polymer. After synthesis in the cytoplasm, the relatively toxic monolignols are glycosylated and transported to the secondary cell wall. Exactly which proteins are then responsible for the monolignol oxidation to form the radical species is still uncertain – although several protein classes have been implicated including peroxidases and laccases among others.

Regardless of the enzymatic source of the reaction, radical formation occurs on the C4-hydroxyl of the monolignol aromatic ring. Due to the aromatic and highly conjugated nature

Table 1.2. Contribution of the Three Primary Monolignols to Lignin in Major Plant Biomass Types

<table>
<thead>
<tr>
<th>Wood Type</th>
<th>Total Lignin Content (% mass)</th>
<th>Coniferyl Alcohol</th>
<th>Sinapyl Alcohol</th>
<th>p-Coumaryl Alcohol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Softwood</td>
<td>25 – 35</td>
<td>90 – 95</td>
<td>5 – 10</td>
<td>0</td>
</tr>
<tr>
<td>Hardwood</td>
<td>18 – 25</td>
<td>50</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Grasses</td>
<td>17 – 24</td>
<td>75</td>
<td>25</td>
<td>5</td>
</tr>
</tbody>
</table>

Adapted from Ref. 46.
of these compounds, monolignol radicals are stabilized by electron delocalization and resonance. In addition to stabilizing the monolignol radical and allowing a lifespan long enough to survive travel from the point of formation to growing lignin polymer, these properties also allow the unpaired electron to relocate within the molecule prior to coupling (Figure 1.6.B). When ultimately coupled to a radical on the lignin polymer (this radical can be similarly delocalized from its phenolic origin) the result is a rich diversity of racemic C-C and C-O-C bonding motifs (Figure 1.6.C).

The most prevalent bond motif observed in natural lignin is β-O-4 (also referred to as 8-O-4') where a bond is formed between the β carbon (C8) of the monolignol and a C4-oxygen (O4) of the lignin polymer (Figure 1.6.C); this bond motif comprises 46 % of all softwood and 60 % of all hardwood lignin linkages. However, the majority of radical

![Diagram](image)

**Figure 1.6.** Monolignol radical formation and major bonding motifs found in lignin. A) Monolignol carbon numbering scheme. R and R' = –H, or –OCH₃. B) Monolignol radical formation on O4, and possible resonance structures. C) Major bonding motifs found in lignin. Dashed bonds represent possible linkage sites to other lignin subunits, or the substituents –H or –OCH₃. The β-O-4 bond is the most prevalent bond motif found in natural lignins.
character is localized on the monolignol phenolic oxygen followed generally by C8, as
determined in a recent computational study of monolignol radicals.\textsuperscript{62} The $\beta$-O-4 aryl ether bond
has been shown to be the most thermodynamically favored bond motif,\textsuperscript{62} lending to its natural
prevalence. Radical character on C3 and C5, though generally less than O4 and C8 ($\beta$), allow
for the formation of additional bond motifs which are dependent on the monolignol. Radical
couplings at C1 or positions with methoxy substituents (C3 and/or C5), coniferyl and sinapyl
alcohols, occur extremely infrequently likely due to steric hindrance.\textsuperscript{62}

Despite its ubiquity, industrial importance and decades of study, the exact mechanisms of
lignin biosynthesis, the biochemical controls of the lignin assembly process, remain hotly
debated.\textsuperscript{55-57, 59, 60, 63} While other theories exist,\textsuperscript{64, 65} two models of lignin biosynthesis dominate
current discussion: the Random Assembly Model, and the Dirigent Protein Model. The most
widely accepted of the two models, the Random Assembly Model (also termed combinatorial
coupling)\textsuperscript{57} initially described by Freudenberg\textsuperscript{53} and supported by Adler,\textsuperscript{54} suggests that there
is no direct biochemical (enzymatic or protein) control over lignin composition (subunit ratios)
and polymerization. The assembly of all other major biopolymers (cellulose, DNA, proteins,
etc) is intricately controlled at all levels, including individual subunit coupling. Conversely,
proponents of Random Assembly believe, as the name implies, that lignin assembly is a random
process. Instead of biochemical control of polymer construction, Random Assembly suggests
that there is no primary sequence to lignin structure and assembly is instead controlled by
regulation of monolignol radical release into the secondary cell wall matrix, and again in the
rate of diffusion to the growing polymer – while the actual coupling (resulting bonding motif
and stereochemistry of the product) is unregulated. Since the resulting lignin is considered
racemic, with two possible conformations for each stereocenter, the total number of potential
isomers for even a small (20 unit) lignin polymer produced by Random Assembly is enormous,
ranging in the tens of billions.\textsuperscript{39, 55} The majority of \textit{in vitro} experiments attempting to mimic the controlled release of lignin monomer radicals do not produce lignins that represent natural polymerization.\textsuperscript{56} Despite this difficulty, there has been some success in reproducing bond motif distribution, and the composition of monolignol incorporation from natural lignin.\textsuperscript{55, 66-69}

The Dirigent Protein Model, however, places itself in strict contrast to Random Assembly. In the Dirigent Protein Model, lignin is placed in the same category as DNA, and proteins – biopolymers with a defined sequence governed by biochemical (protein and enzymatic) controls.\textsuperscript{59, 60, 63} However, in the biosynthesis of DNA and proteins coupling reactions are catalyzed by enzymes (DNA polymerase) or ribosomes respectively. These biochemical agents are intricately involved in the course of the reaction, for example providing or removing protons, as well as provide an environment which correctly orients starting materials. Dirigent proteins, however, do not play a catalytic role in the reactions they facilitate, but instead serve only to mediate the coupling of starting materials (i.e. two monolignol radicals) and direct both the regio- and stereochemistry of the resulting product.\textsuperscript{63, 70} Dirigent proteins were first implicated in the stereoselective \textit{in vitro} coupling of coniferyl alcohol radicals to form the lignan (phenylpropanoid dimer) (+)-pinoresinol via a $\beta-O-\beta$ (8,8') linkage (Figure 1.6.C).\textsuperscript{63, 71} Coupling of coniferyl alcohol radicals in the absence of the dirigent protein (i.e. random coupling) results in a racemic mixture of multiple coupling products. While all currently known dirigent proteins only produce lignan dimers (a major point of contention\textsuperscript{56}) and specific dirigent proteins involved in lignin polymer synthesis have not been identified,\textsuperscript{59, 60} immunoassays for the (+)-pinoresinol forming dirigent protein in live cells suggest the localization of similar proteins in lignin forming areas.\textsuperscript{71, 72}

Without significant advances, the debate over lignin biosynthesis is not likely to end in the near future. Neither model can conclusively reproduce all aspects of a natural lignin polymer
in vitro, nor explain lignin structural composition variations within different plant tissues. From a biochemical perspective, it is difficult to rationalize how the structure of such a major biopolymer, integral to the structural integrity of the terrestrial plants, is left to chance – an argument on which Dirigent Protein Model proponents rely heavily. Yet, the seemingly random nature of lignin is also the factor from which lignin derives its ability to act as a natural barrier to predation and disease – an aspect of plant-lignin physiology which the Dirigent Protein Model cannot explain. Additionally, the Dirigent Protein Model can currently only hypothesize protein involvement in lignin polymerization, as an extension of known dirigent proteins which bind monolignol radicals to direct lignan formation, since there is currently no direct evidence for dirigent protein involvement in lignin polymer formation.

1.4.2. Lignin Depolymerization. Despite lignin’s recalcitrance, lignin is not completely resistant to degradation. Just as organisms evolved to degrade and utilize cellulose as a carbon source (more resistant to degradation than starch), so too have they evolved to degrade and utilize the carbon found in lignin. Microbial lignin degradation is in fact an important part of the carbon cycle – without which a thick layer of lignin would likely cover the Earth’s surface. While lignin is mainly used industrially as a boiler fuel, its potential as a feedstock for aromatic compounds and other value added chemicals has also led to the development of a wide range of thermochemical methods to depolymerize this highly abundant renewable material. Nearly all of these methods require high temperatures (often over 400 °C), high pressures, catalysts or a combination of each. Unquestionably, pyrolysis is the most widely studied method for lignin depolymerization. Here, lignin is exposed to high temperatures under an oxygen limited environment which prevents combustion.46, 61, 73-75 Pyrolysis can be performed in the absence (thermolysis) or presence of H2 (hydrogenolysis),75, 76 and often in the presence of wide variety of catalysts.46, 61 Other common methods of lignin depolymerization include base catalyzed
depolymerization\textsuperscript{52} and oxidation,\textsuperscript{75} while methods such as depolymerization in supercritical solvents (most often water) or ionic liquids\textsuperscript{75, 77} are increasing in popularity.

These thermochemical methods are not without their challenges despite the use of what can be thought of as brute force in comparison to those employed by nature. Regardless of which method is used for the thermochemical depolymerization of lignin, the result is more often than not an expansive list of small aromatic chemicals\textsuperscript{78} with little control over selectivity in their production,\textsuperscript{52} as well as the formation of undesirable solid residue byproducts.\textsuperscript{61} Additionally, the catalysts used by some methods are often deactivated by reaction products,\textsuperscript{52} necessitating their replacement and increasing the cost of their use. These issues make the application of energy intensive thermochemical methods for the large scale production of chemicals from lignin difficult, and research into thermochemical depolymerization is largely focused on improving product uniformity through catalysts.\textsuperscript{61}

Despite the lack of focus on reducing the energy intensity of these processes, a recent study has demonstrated the ability to depolymerize lignin under much more mild conditions.\textsuperscript{79} The new depolymerization process first oxidizes natural lignin resulting in the conversion of the $\alpha$-hydroxyls of $\beta$-$O$-4 bonded lignin subunits into ketones. The oxidized lignin is subsequently heated in an aqueous formic acid solution with sodium formate at 110 °C, significantly lower temperature than nearly all thermochemical depolymerization methods. While approximately 30 % of the original lignin mass remains an insoluble polymeric material (though decreased in molecular weight relative to the lignin starting polymer), more than 52 % of the original lignin polymer mass is converted to well-defined monomeric aromatic compounds.\textsuperscript{79} These results are very significant and a large leap forward for chemical lignin depolymerization methods, but still rely heavily on chemical catalysts and corrosive acids.
1.4.3. Microbial Lignin Degradation. Since microorganisms (both fungi and bacteria) have had a much longer exposure time to lignin, means of enzymatic degradation have evolved with much greater energy efficiency than currently known industrial methods. The most prolific and well-studied of nature’s lignocellulose degrading microorganisms are fungi. Degradation occurs when the fungal mycelia (the fibrous and branching root-like structures of the fungus) penetrate the woody material and excrete enzymes (lignin and manganese peroxidases, and laccases, etc) capable of depolymerizing the generally insoluble polymeric materials into soluble low molecular weight sugar and aromatic units. Fungi with the ability to degrade lignocellulosic material are classified into three groups based on the appearance of the material subsequent to fungal degradation: soft rot, brown rot, and white rot. Each group is capable of degrading both lignin and cellulose, however, with distinctly different capacities.

Cellulose derived sugars are the primary target for both brown rot and soft rot fungi. These fungi degrade lignin to only a small extent in order to access and more efficiently depolymerize cellulose. The majority of the lignin is therefore left intact and remains as a brown brittle or crumbling material. Conversely, lignin depolymerization (also known as mineralization) by white rot fungi is extensive and cellulose remains largely, though not completely, unscathed – leaving behind characteristic white crystalline fibers. While lignin is efficiently degraded and to some extent metabolized by white rot fungi, the fungi cannot subsist on lignin alone and so it is believed that the hemicellulose (a polysaccharide composed of primarily of pentose, five carbon, sugars) entangled within lignin is the primary carbon source being harvested.

The most commonly implicated enzymes in fungal lignin depolymerization are lignin peroxidase and manganese peroxidase. As described above, both enzymes are extracellular in nature. However, neither enzyme directly binds to lignin polymers to catalyze depolymerization. Instead, lignin peroxidase and manganese peroxidase generate radicals on
mediator substrates which are more mobile, and capable of diffusing to the lignin polymer to transfer the radical. Lignin peroxidase contains an iron (III) heme which undergoes an initial two electron oxidation by hydrogen peroxide (H$_2$O$_2$) to generate an iron (IV) heme-radical intermediate (Figure 1.7.A).\textsuperscript{81} This oxidized lignin peroxidase intermediate is then capable of sequential one-electron oxidations of two organic substrate mediators.\textsuperscript{81} The organic radical mediator for lignin peroxidase is thought to be veratryl alcohol (Figure 1.7.B), a fungal metabolite.\textsuperscript{81-83} The heme iron of manganese peroxidase is similarly activated by H$_2$O$_2$ to form a oxo-ferryl intermediate.\textsuperscript{82} Oxidation of Mn$^{2+}$ generates the radical mediator Mn$^{3+}$,\textsuperscript{82} which is released from the enzyme in the presence of stabilizing chelating agents such as oxalate.\textsuperscript{81}

Fungi were originally believed to be the only organismal source of lignin degradation and depolymerization, while bacteria played a supporting role by further catabolizing the lignin derived aromatic compounds (LDACs) generated from fungal degradation. However, literature over the last several decades has been challenging this long held canon with increasing frequency, principally due to advances in DNA sequencing technologies and increased effort towards lignin utilization.\textsuperscript{84-86} Bacteria containing LDAC metabolic pathways are often associated with and isolated from sources with high lignin concentration such as the waste water effluent of pulp and paper mills, and the guts of wood eating insects (e.g., termites, Southern pine beetles, and Asian long-horned beetles among others).\textsuperscript{86} The enzymatic means by which bacteria depolymerize lignin are similar to those found in fungal systems.
(peroxidases, laccases, etc.), but have generally been found to be less efficient than their fungal counterparts.86, 87

Unfortunately, neither bacteria nor fungi are held to a production schedule other than that required for survival, and like their thermochemical depolymerization counterparts fungal and bacterial enzymatic depolymerization produce a large range of a lignin derived aromatic compounds (LDACs) (Table 1.3).78, 79, 88-90 Regardless of their source, currently understood enzymes involved in lignin depolymerization tend to function on much longer time scales than would be industrially relevant. Despite lower lignin depolymerization efficiency, bacterial systems bring a host of advantages to industrial chemical production routes, including: a more facile organism to manipulate, the ability to accept foreign (non-bacterial) DNA and produce foreign proteins, a tolerance of a wide range of environmental conditions, fast growth and reproduction, and the capacity to exploit highly diverse sets of compounds as carbon sources – including aromatics.

1.4.4. Metabolism of Lignin Derived Aromatic Compounds. Bacteria utilize several different enzymatic pathways to metabolize aromatic compounds.91-94 LDACs, however, largely fall into a group of compounds that can be classified as derivatives of catechol (structure shown in bold in Table 1.3) identifiable from their vicinal oxygen substituents. Examples of LDACS include vanillin, 1,2-dihydroxy-4-methoxybenzene, guiacol, eugenol, and syringic acid (Table 1.3). Compounds such as these are funneled through the well characterized β-ketoadipate pathway (present in both soil bacteria and fungi) via a host of enzymatic reactions which ultimately lead to the formation of two major molecular contributors to central metabolism: succinyl-CoA and acetyl-CoA (Scheme 1.1).95-97
Table 1.3. Aromatic Monomers Derived from the Depolymerization of Lignin

<table>
<thead>
<tr>
<th>Lignin Derived Aromatic Compounds (LDACs)</th>
<th>Reference</th>
<th>78</th>
<th>79</th>
<th>88</th>
<th>89</th>
<th>LDAC Structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenol</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Methoxytoluene</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Methylphenol</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Methylphenol</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Vinylphenol</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Hydroxybenzaldehyde</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Hydroxybenzaldehyde</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Methoxyvinylbenzene</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Hydroxy-3-nitrobenzaldehyde</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catechol</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,2-Dihydroxy-4-methoxybenzene</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,2-Dihydroxy-3-methylbenzene</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,2-Dihydroxy-4-methylbenzene</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guaiacol</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylguaiacol</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Ethylguaiacol</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Vinylguaiacol</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vanillin</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homovanillin</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vanillic Acid</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Carboxy-2-methoxyphenol methyl ester</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homovanillic acid</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eugenol</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isoeugenol</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propylguaiacol</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetovanillin</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guaiacylacitone</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coniferaldehyde</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Methoxy-4-(1-hydroxyprop-2-ethyl)phenol</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-Methoxy-4-(1-hydroxypropyl)phenol</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,5-Dinitroguaiacol</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferulic Acid</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Versatilaldehyde</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,4-Dimethoxytoluene</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,2-Dimethoxy-4-vinylbenzene</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,2-Dimethoxy-4-(1-propenyl)benzene</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syringol</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Methysyringol</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Ethylsyringol</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Vinylsyringol</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Allysyringol</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syringaldehyde</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syringic acid</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Propylsyringol</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Propanolsyringol</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sinapylalcohol</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sinapylaldehyde</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sinapylacetone</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetosyringone</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,2,3-Trimethoxy-5-vinylbenzene</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Structures are shown for compound names highlighted in bold.
Many LDACs, both mono- and biphenyl, have been found to enter a second complex catabolic pathway (Scheme 1.1), akin to the β-ketoadipate pathway, in the Gram-negative bacterium *Sphingobium* sp. SYK-6 (formerly known as *Sphingomonas paucimobilis* SYK-6) isolated from a waste liquor treatment pond of a kraft pulp mill. While there are many similarities, a small difference in the catabolism of protocatechuic acid (3,4-dihydroxybenzoic acid, PCA) results in the divergence of the *Sphingobium* sp. SYK-6 LDAC degradation pathway away from the β-ketoadipate pathway found in soil bacteria. A comprehensive review of these soil bacteria and the β-ketoadipate pathway has been compiled by Wells and Ragauskas.

Protocatechuic acid is a keystone metabolite in the bacterial degradation of many LDACs, and the aromatic ring opening of PCA, via insertion of dioxygen across a double bond, is critical to its fate in central metabolism. As can be seen in Scheme 1.1, the β-ketoadipate pathway results when a 3,4-cleavage (intradiol) of PCA occurs. The ring-opening of PCA in *Sphingobium* sp. SYK-6, however, occurs via a 4,5-cleavage (extradiol) to form 4-carboxy-2-hydroxymuconate-6-semialdehyde (CHMS), and therefore changing the downstream enzymatic reactions that occur as well as the products that enter central metabolism (pyruvate and oxaloacetate vs. succinyl-CoA and acetyl-CoA). Abbreviations found in Scheme 1.1 are defined in Table 1.4. The ring cleaving dioxygenation of LDACS is a key step in the generation of linear products that can enter central metabolism, and dioxygenase enzymes are found throughout pathways for the degradation of lignin derived aromatics.
Scheme 1.1. The lignin derived aromatic compound degradation pathway of *Sphingobium* sp. SYK-6, and the β-ketoadipate pathway. The β-ketoadipate pathway found in soil bacteria (blue arrows) and fungi (green dashed arrows) is shown with grey background to highlight the difference in PCA catabolism resulting from 4,5-cleavage by *Sphingobium* sp. SYK-6. Full names for enzyme and compound abbreviations are listed in Table 1.4.
Table 1.4. Referenced Enzymes involved in *Sphingobium* sp. SYK-6 LDAC degradation and the β-Ketoadipate pathway of soil bacteria and fungi

<table>
<thead>
<tr>
<th>Enzyme abbr. and Function</th>
<th>Compound abbr. and Name</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sphingobium</strong> sp. SYK-6</td>
<td></td>
</tr>
<tr>
<td>LigV Vanillin dehydrogenase</td>
<td>PCA Protocatechuate</td>
</tr>
<tr>
<td>LigM Vanillate/3OMG O-demethylase</td>
<td>CHMS 4-Carboxy-2-hydroxymuconate-6-semialdehyde</td>
</tr>
<tr>
<td>LigAB PCA 4,5-dioxygenase</td>
<td>PDC 2-Pyrene-4,6-dicarboxylate</td>
</tr>
<tr>
<td>LigC CHMS dehydrogenase</td>
<td>GA Gallate</td>
</tr>
<tr>
<td>LigI PDC hydrolase</td>
<td>3OMG 3-O-methylgallate</td>
</tr>
<tr>
<td>DesA Syringate O-demethylase</td>
<td>CHMOD 4-Carboxy-2-hydroxy-6-methoxy-6-oxohexa-2,4-dienoate</td>
</tr>
<tr>
<td>DesB Gallate 3,4-Dioxygenase</td>
<td>OMA 4-Oxolmesaconate</td>
</tr>
<tr>
<td>DesZ 3-O-methylgallate 3,4-dioxygenase</td>
<td>CHA 4-Carboxy-4-hydroxy-2-oxoadipate</td>
</tr>
<tr>
<td>LigJ OMA hydratase</td>
<td></td>
</tr>
<tr>
<td>LigK CHA aldolase</td>
<td></td>
</tr>
<tr>
<td><strong>β-Ketoadipate Pathway</strong></td>
<td></td>
</tr>
<tr>
<td>C1,2O Catechol 1,2-dioxygenase</td>
<td>ML Muconolactone</td>
</tr>
<tr>
<td>MLE cis,cis-Muconate lactonizing enzyme</td>
<td>β-CM β-Carboxymuconate</td>
</tr>
<tr>
<td>MI Muconolactone isomerase</td>
<td>γ-CML γ-Carboxymuconolactone</td>
</tr>
<tr>
<td>P3,4O PCA 3,4-dioxygenase</td>
<td>β-CML β-Carboxymuconolactone</td>
</tr>
<tr>
<td>CMLE β-Carboxy-cis,cis-muconate lactonizing enzyme</td>
<td>β-KAE-L β-Ketoadipate enol-lactone</td>
</tr>
<tr>
<td>CMD γ-Carboxy-muconolactone decarboxylase</td>
<td>β-KA β-Ketoadipate</td>
</tr>
<tr>
<td>ELH β-Ketoadipate enol-lactone hydrolase</td>
<td>β-KA-CoA β-Ketoadipyl-CoA</td>
</tr>
<tr>
<td>TR β-Ketoadipate succinyl-CoA transferase</td>
<td></td>
</tr>
<tr>
<td>TH β-Ketoadipyl-CoA thiolase</td>
<td></td>
</tr>
</tbody>
</table>
1.4.5. *Ring-cleaving Dioxygenases.* While difficult to conduct non-enzymatically under ambient conditions, these oxidative ring-opening reactions are catalyzed in a relatively facile nature by non-heme iron containing dioxygenase enzymes. Aromatic ring-cleaving dioxygenases are generally divided into three classes: intradiol (Class I), extradiol (Class II), and those capable of aromatic ring-cleavage without a diol (proposed Class III). The structural diversity of intradiol dioxygenases is limited, all of which utilize a non-heme iron (III) cofactor to cleave the aromatic C-C double bond between adjacent hydroxyl substituents. The necessity for vicinal hydroxyl substituents limits intradiol dioxygenases to three substrates – catechol, protocatechuic, and 2-hydroxyquinol. Extradiol dioxygenases (EDOs), however, are much more diverse than their intradiol counterparts, in both substrate and structural diversity. EDOs are further sub-divided into Types based on superfamily classification (homologous enzymes that maintain a common structural fold, and either catalyze the same chemical reaction or different reactions with common mechanistic features): Vicinal Oxygen Chelate (VOC, Type I), Cupin (Type III), and a currently undefined superfamily (Type II) hereafter referred to as the Protocatechuate Dioxygenase (PCAD) superfamily (Figure 1.8).

![Figure 1.8. Structural comparison of extradiol dioxygenase superfamilies.
A) PCAD (LigAB, PDB 1B4U).
B) Cupin (Quercetin 2,3-dioxygenase, PDB 1Y3T).
C) Vicinal Oxygen Chelate (Catechol 2,3-dioxygenase, PDB 1MPY).](image-url)
EDOs from the VOC and Cupin superfamilies have been well characterized by both enzymological and structural methods.\textsuperscript{105-109}

1.4.6. The PCAD Superfamily. This thesis describes the foundational work in the Taylor Lab to contribute to the characterization of the PCAD superfamily. The bacterium \textit{Sphingobium} sp. SYK-6 harbors genes for the production of three dioxygenase enzymes linked to the metabolism of LDAC phenolic monomers, each of which also belongs to the Type II PCAD Superfamily: Protocatechuate 4,5-dioxygenase (LigAB), gallate 3,4-dioxygenase (DesB), and 3-O-methylgallic acid 3,4-dioxygenase (DesZ). LigAB was the first, and for nearly a decade the only, structurally characterized member of the PCAD Superfamily. The structure of LigAB has recently been joined by DesB from \textit{Sphingobium} sp. SYK-6, 4,5-DOPA-extradiol dioxygenase (YgiD) from \textit{E. coli} K12 (Figure 1.9), and a phosphotyrosine binding protein (MEMO, not shown in Figure 1.9).\textsuperscript{110-113} In addition to these structurally characterized members, there are other enzymes in the PCAD superfamily that have been functionally characterized including protocatechuate 4,5-dioxygenases from the \textit{Comamonas testosteroni} strains Pt-L\textsuperscript{5114} and T-2,\textsuperscript{115} and the 2,3-dihydroxyphenylpropionate 1,2-dioxygenase (MhpB),\textsuperscript{116} and homoprotocatechuate 2,3-dioxygenase (HpcB)\textsuperscript{117} from \textit{E. coli}.
among others. Together these enzymes and proteins provide the foundation for understanding the PCAD superfamily.

While LigAB shares structural similarity with the other known PCAD superfamily members, it is in many ways unique amongst its peers. Structurally, LigAB is a dimer of heterodimers unlike the known structures of other PCAD members which are homodimers. Each heterodimer of LigAB consists of a small (α) and large (β) subunit resulting in an enzyme complex formula of $\alpha_2\beta_2$. Catalysis occurs in an active site located at the interface of each $\alpha$-$\beta$ pair. Comprised primarily of residues located in the β subunit, the active site contains a catalytic iron (II) cofactor bound within a two histidine and one glutamic acid residue (2His, 1 Glu) motif that is conserved not only within the Type II extradiol dioxygenase superfamily but also Types I and III (suggesting that evolution has converged upon this iron binding motif and catalytic function multiple times).

LigAB’s in-pathway relatives, DesB and DesZ, on the other hand are composed of two identical subunits (homodimeric) in which each subunit contains two distinct domains, N-terminal and a C-terminal, which maintain sequence, and, in the case of DesB, structural homology to the β and α subunits of LigAB, respectively. Amino acid sequence alignments reveal that DesB retains 39 and 30 % sequence identity to the LigAB β and α subunits while DesZ retains only 17 and 12 % sequence identity, suggesting that LigAB and DesB are more closely related to each other evolutionarily than they are to DesZ.

Another feature that makes LigAB an interesting system to understand in greater detail, is its ability to catalyze the oxidative ring cleavage of the in-pathway substrates gallate and 3-O-methyl gallate – reactions that are natively performed by DesB and DesZ (Scheme 1.1). From an enzymatic pathway engineering perspective, with the goal of chemical production using lignin depolymerization products as a feedstock, this “omnivore” feature of LigAB could
possibly be highly attractive due to the large diversity of LDACs that are produced in the lignin depolymerization process. An enzymatic pathway containing one enzyme capable of catabolizing many substrates would greatly reduce the complexity and time required to generate a functional pathway, in comparison to the one-substrate-one-enzyme alternative. Further understanding the underlying mechanisms, both structural and chemical, that govern the substrate selectivity of LigAB and other PCAD superfamily dioxygenases could greatly enhance the utility of these enzymes, possibly leading to enzyme mutants or chimeras with expanded or altered substrate range.

1.5. Research aims. Lignin depolymerization and oxidative aromatic ring-opening are only two of many enzymatic steps that would be necessary to produce a functioning metabolic pathway for the bio-production of chemicals from lignin. However, these steps also pose two of the largest hurdles to achieving this goal. The research described herein has been conducted in hopes of contributing to the enzymological knowledge of LigAB and the PCAD Superfamily, as well as our ability to detect new sources of lignin degradation in natural systems via the synthesis of lignin structural mimetic probes.

Although our initial understanding of LigAB, and dioxygenases in general, was in many ways limited, we were inspired to study this system due to the unique qualities that set it apart from other ring-cleaving dioxygenase enzymes. Despite having been isolated and studied for the better part of two decades prior to our involvement, little was known about the enzymology, and capabilities of LigAB beyond the context of its native pathway. In order to understand the role LigAB may be able to play in the larger scheme of bio-commodity production, the research herein was conducted to address several specific aims: (1) determine the kinetic parameters of LigAB with its native substrate; (2) determine the scope of substrate structures which inhibit LigAB ring-cleavage of the native PCA substrate, and which can or
cannot be cleaved by LigAB; (3) investigate the structural parameters that contribute to the substrate specificity and selectivity of LigAB; (4) study any previously unobserved phenomena unique to LigAB that present themselves while addressing Aims #1-3; and (5) design and synthesize lignin model compounds for the in vivo identification of lignin depolymerization by bacteria and fungi within difficult to isolate natural systems.

Aim #1 will be addressed by first obtaining anaerobically purified LigAB. Due to the sensitivity of LigAB’s iron (II) cofactor towards oxidation by dissolved oxygen, methods for anaerobic protein purification performed in a glove box containing an inert-N₂ atmosphere will be key to obtaining maximally active enzyme. The kinetic parameters (kₐₜ, Kₘ) for LigAB’s catalytic ring-cleavage will then be determined from assays measuring the rate of oxygen consumption in the presence of known substrates (PCA, gallate, and 3-O-methyl gallate) via an oxygen sensitive electrode. Due to the 1:1 stoichiometric ratio for the consumption of oxygen to organic substrate, oxygen consumption by LigAB can be used as a proxy to determine the rate of organic substrate ring-cleavage. Aim #2 will then be addressed by assessing LigAB’s oxygen consumption capabilities in the presence of both the native substrate PCA, and organic substrate-like compounds from a library of PCA structural analogues. Additionally, these PCA structural analogues will be further assessed for their ability to be substrates of LigAB themselves by observing LigAB’s oxygen consumption in the presence of each analogue.

Aim #3 will be addressed through the inspection of the LigAB crystal structure to identify non-catalytic amino acid residues that may play important substrate binding roles (through polar interactions or steric clashes) which can lead to the exclusion or hindrance of potential substrates from binding. Once identified, a series of mutations will be introduced at these residue positions in order to determine if modification allows for substrates that were
previously poor (low $k_{\text{cat}}/K_m$ relative to LigAB + PCA) to bind more efficiently to and/or be turned over more quickly by LigAB. Methods analogous to those used in the characterization of the wild-type enzymes will be used to characterize the LigAB mutants. By gaining a better understanding of how the structural components of LigAB impact substrate binding and utilization this information, the knowledge can direct future studies into the construction of enzyme mutants or chimeras for use in broad spectrum ring-cleavage of lignin derived aromatic compounds.

While not evident from the outset, Aim #4 will be addressed by studying the previously unobserved phenomena of allosteric feedforward activation of an extradiol dioxygenase which presented itself during investigations of Aims #1-3. Utilizing methods analogous to those used in Aims #1-3, the effect of co-incubating vanillin (a precursor to PCA, and non-substrate of LigAB) during the catalysis of known aromatic organic substrates by LigAB will be examined. The rate enhancement effects due to varied vanillin and constant substrate concentrations, and constant vanillin concentration with varied substrate concentration will be determined for both wild-type LigAB and mutants generated to address Aim #3.

In addition to Aims #1-4 which address the role and characterization of LigAB in the LDAC degradation pathway of *Sphingobium* sp. SYK-6, the design and synthesis of a lignin model compound for the *in vivo* identification of lignin depolymerization will be discussed. Utilizing the prevalent β-Ο-4 bond motif between aromatic lignin monomers as a scaffold, fluorescent dyes will be appended to opposite ends of a synthesized lignin dimer. Several methods to link the dyes to the dimer will be explored, including: ester formation through the coupling of activated carboxylic acids to a free hydroxyl group, and azide-alkyne copper catalyzed click chemistry.
Chapter 2:

Characterizing the Promiscuity of LigAB, a lignin catabolite degrading extradiol dioxygenase from *Sphingobium* sp. strain SYK-6

Adapted with permission from:


©2013 American Chemical Society
2.1. Introduction

The natural ability of microbes to process typically recalcitrant carbon sources has led to research focused largely on accessing the carbon locked in cellulose for the production of bio-derived fuels and chemicals, while lignin has been investigated less extensively. As has been previously discussed, lignin derived aromatic compounds (LDACs) have been shown to be accessible for use in central metabolism in some bacteria, which is suggestive of their potential as a possible carbon source in microbial fermentative processes. The ring-cleaving extradiol dioxygenases (EDOs) play a key role in the catabolism of these LDACs, and while EDOs from the VOC and Cupin superfamilies (Types I and III, respectively) have been well characterized by both enzymological and structural methods the EDOs of the Protocatechuate Dioxygenase (PCAD) superfamily are less well defined.

The existence of Type II EDOs, as postulated by protein sequence homology, has been known for the better part of two decades and yet only a handful of these enzymes have been studied beyond sequence alignments, or gene and pathway identification. Ten Type II EDOs have been isolated, and characterized in varying degrees either functionally or kinetically (2,3-dihydroxyphenylpropionate 1,2-dioxygenase from *E. coli* (MhpB) and *Alcaligenes eutrophus* (Mpl), protocatechuate 4,5-dioxygenases *Sphingobium* sp. SYK-6 (formerly *Sphingomonas paucimobilis*) (LigAB), *Comamonas testosteroni* Pt-L5, and *Comamonas testosteroni* T-2 (PmdAB), protocatechuate 2,3-dioxygenase (2,3-PCD) from *Paenibacillus* sp. (formerly *Bacillus macerans*), 2-aminophenol 1,6-dioxygenase (1,6-ApDO) from *Pseudomonas pseudoalcaligenes*, 2'-aminobiphenyl-2,3-diol-1,2-dioxygenase (CarBab) from *Pseudomonas stutzeri*, the polycyclic arene diol dioxygenase (PhnC) from *Burkholderia* sp., and gallate 2,3-dioxygenase (GDO) from *Pseudomonas putida*.
KT2440} (Figure 2.1). Further, only five other enzymes in this superfamily have been functionally identified by molecular biology techniques (homoprotocatechuate 2,3-dioxygenase (HpcB) from *E. coli*, an acid catechol dioxygenase (HppB) from *Rhodococcus globerulus* PWD1, an extradiol dioxygenase EdoD from *Rhodococcus* sp., and protocatechuate 4,5-dioxygenases from *Pseudomonas straminea* (originally *ochraceae* NGJ1) (ProOab) and *Anthrobacter keyseri* 12B (PcmA) (Figure 2.1). Additionally, the protocatechuate 4,5-dioxygenase LigAB from *Sphingobium* sp. SYK-6 was the first structurally characterized EDO from this largely undefined superfamily. Until now, full kinetic characterization and mechanistic investigation of Type II EDOs has only been performed for MhpB by Bugg, and 2,3-PCD by Lipscomb, though 2,3-PCD was only identified as a Type II dioxygenase a decade later by Masai. Kinetic analysis beyond specific activity is largely lacking for LigAB and other Type II EDOs.

Figure 2.1. Ring cleavage reactions catalyzed by known Type II extradiol dioxygenases.
LigAB, identified as a member of the bacterial LDAC catabolic pathway of *Sphingobium* sp. SYK-6,\textsuperscript{98, 119, 136-138} catalyzes the extradiol aromatic ring cleavage of protocatechuate by the insertion of O\textsubscript{2} across the C4-C5 bond of the aromatic ring to form 4-carboxy-2-hydroxymuconate-6-semialdehyde, CHMS (Scheme 2.1). LigAB has been shown to have some promiscuous activity with PCA substrate analogues gallate (GA) and 3-O-methyl gallate (3OMG), two compounds also found in the LDAC catabolic pathway of *Sphingobium* sp. SYK-6. This activity has been demonstrated through knock out studies and *in vitro* assays of specific activity,\textsuperscript{99, 139-141} however, the breadth of the substrate utilization profile was previously unknown.

In this chapter, we report the kinetic parameters for anaerobically purified LigAB with both of its physiological substrates, PCA and O\textsubscript{2}, as well as an investigation of other substrates containing the catechol scaffold including GA and 3OMG – with full kinetic characterization of the best substrates. From these analyses we aim to establish the structural and oxidation potential requirements of substrates for LigAB. Additionally, steady-state kinetic studies are used to establish the pH and metal dependence of LigAB. Furthermore, we demonstrate that anaerobic purification of this enzyme is necessary for optimal stability and kinetic efficiency.

\textbf{Scheme 2.1.} A) Ring cleavage reactions catalyzed by known PCA dioxygenases, and B) known substrates of LigAB.
2.2. Materials and Methods

2.2.1. General Methods. Commercially available reagents and solvents were purchased from Aldrich or Alfa Aesar and used without purification unless otherwise noted. 3-O-methyl gallate was purchased from ChromaDex. Several substrates were synthesized as previously described: 3,4-dihydroxybenzamide,\textsuperscript{142} 1,2-dihydroxy-4-aminobenzene hydrobromide,\textsuperscript{143} and methyl-3,4-dihydroxybenzoate.\textsuperscript{144} A Varian Mercury 400 MHz NMR was used to collect all \textsuperscript{1}H NMR spectra. Absorption spectra were collected with a Varian Cary 100 Bio UV-Visible Spectrophotometer (Palo Alto, CA). Bradford and Ferene-S assays were conducted in 96-well plate format and absorption values recorded on a Molecular Devices SpectraMax M5 plate reader (Sunnyvale, CA). MAX Efficiency DH5\textalpha and One Shot BL21 Star chemically competent \textit{E. coli} cells were purchased from Life Technologies (Carlsbad, CA). All enzymes used for DNA manipulations were purchased from New England Biolabs (Ipswich, MA). Centrifugation and ultracentrifugation were performed on a DuPont Instruments (Wilmington, DE) Sorvall RC-5B centrifuge and Beckman (Brea, CA) L7-80 Ultracentrifuge with type 60-Ti rotor, respectively. All cells were lysed using a SIMO-Aminco Industry Inc. (Rochester, NY) French press.

2.2.2. Protein Expression. Since we were unable to acquire genomic DNA for \textit{Sphingobium} sp. SYK-6 or the plasmids containing the genes for LigAB, the DNA sequence encoding the LigA-LigB gene cluster was retrieved from the NCBI data bank. A construct for gene synthesis was prepared with insertion of an upstream \textit{NdeI} endonuclease site, and additional stop codons followed by a \textit{BamHI} endonuclease site downstream of the LigA and LigB gene cluster. The modified sequence was synthesized by DNA2.0 (Menlo Park, CA). The synthetic gene sequence was subsequently removed from the provided pJ241 vector by endonuclease cutting with \textit{NdeI} and \textit{BamHI}, ligated into pET-15b (EMD Millipore, Billerica, MA) in the \textit{NdeI} and
BamHI cloning sites and transformed by heat shock separately into DH5α and BL21 competent 
*E. coli* cells.

His-tagged protein (α subunit N-terminally tagged, β subunit untagged) was expressed from 
a pET15b plasmid by induction of BL21 competent cells with a final concentration of 1 mM 
Isopropyl β-D-1-thiogalactopyranose (IPTG) at OD_{600} = 0.6 under aerobic conditions at 37 °C. 
After 20 h the cell cultures were centrifuged at 5,000 rpm. If not used immediately, cell pellets 
were stored at -80 °C.

2.2.3. Aerobic Purification. Pelleted cells were re-suspended in 50 mM Tris (pH 7.0) 
containing 1:9 glycerol/H_{2}O with one tablet of cOmplete-EDTA free protease inhibitor cocktail 
(Roche, Basel, Switzerland). Re-suspended cells were lysed by 3 passes through a French-press 
at 13,000 psi. The cell lysate was subsequently centrifuged at 25,000 rpm for 45 min in an 
ultracentrifuge.

The supernatant was applied to a gravity column loaded with HisPur Ni-NTA resin (10 mL, 
Thermo Scientific) previously charged with Ni^{2+}, and equilibrated with 5 column volumes (CV) 
of bind buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 7.5) followed by 5 CV 
of wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 7.5) to remove unbound 
protein. The tagged enzyme was eluted by a step gradient of a buffer containing 50 mM HEPES 
(pH 8.0) and 300 mM NaCl with, sequentially, 62.5 mM (2.5 CV), 125 mM (2.5 CV), and 250 
mM (5 CV) imidazole. Fractions (10 mL) were collected for all wash and elution steps. 
Aliquots (30 µL) of each fraction were analyzed for purity by SDS-PAGE.

Fractions containing the co-purified LigAB α and β subunits were pooled, concentrated 
using an Amicon concentrator stirred cell with a regenerated cellulose 10 kDa molecular weight 
cutoff ultrafiltration membrane (EMD Millipore), and buffer exchanged into 50 mM Tris (pH 
7.5), 0.5 mM DTT, and 0.5 mM Fe(NH_{4})_{2}(SO_{4})_{2}·6H_{2}O. Exchanged protein was filtered with a
0.45 µm syringe filter, and subsequently flash frozen in liquid nitrogen and stored at -80 °C. From 6 L of cells, 70 mg of protein was typically obtained and concentrated to ~7 mg/mL.

2.2.4. Anaerobic Purification. Anaerobic purification of LigAB was performed as described above with modification to maintain an oxygen free environment. After cell lysis, the lysate was transferred to centrifuge tubes and the head space was flushed with N₂ gas prior to centrifugation at 25,000 rpm for 45 min in an ultracentrifuge. The supernatant was decanted to a serum vial, flushed with N₂ gas, sealed, and transferred into a Vacuum Atmospheres Company (Hawthorne, CA) HE-493/MO-5 glovebox under N₂ atmosphere, operated at <5 ppm O₂.

All buffers, water, and column resins to be used in the glovebox were prepared as described and thoroughly degassed, to remove O₂, by three cycles of degassing under vacuum followed by N₂ gas bubbling. Degassed buffers were allowed to equilibrate in the glovebox atmosphere overnight prior to use.

The supernatant was applied to a gravity column loaded with HisPur Ni-NTA resin inside the glove box and protein was eluted from the column as described above. Fractions (10 mL) were collected for all wash and elution steps. Aliquots (30 µL) of each fraction were removed from the glove box and analyzed for purity by SDS-PAGE. Fractions containing the co-purified LigAB α and β subunits were pooled, concentrated using an Amicon concentrator stirred cell with a regenerated cellulose 10 kDa molecular weight cutoff ultrafiltration membrane (EMD Millipore), and buffer exchanged into 50 mM Tris, pH 8.

Phosphate buffered saline (PBS) was added to the protein solution to a concentration of 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄•2H₂O, and 2 mM KH₂PO₄ at pH 7.4 followed by bovine derived thrombin (MP Bio) to a concentration of 1 U/mg purified protein to remove the His-tag. His-tag cleaved protein was buffer exchanged into 50 mM Tris (pH 8) to remove
excess salt prior to application to a Source-Q ion exchange column (10 mL, GE Healthcare). LigAB was applied to the column with 50 mM Tris (pH 8), washed with 50 mM Tris (pH 8), 100 mM NaCl, and eluted with 50 mM Tris (pH 8), 350 mM NaCl. Fractions (5 mL) were collected for both wash and elution steps. Aliquots (30 µL) of each fraction were removed from the glove box and analyzed for purity by SDS-PAGE. Those fractions containing both the LigAB α and β subunits were pooled, concentrated, and buffer exchanged into 50 mM Tris (pH 7.5), 0.5 mM DTT, and 0.5 mM Fe(NH₄)₂(SO₄)₂•6H₂O. Exchanged protein was filtered with a 0.45 µm syringe filter and removed from the glovebox in an airtight syringe with the needle sealed by a rubber stopper, and subsequently flash frozen in liquid nitrogen and stored at -80 °C. From 6 L of cells, 70 mg of protein was typically obtained and concentrated to ~7 mg/mL.

2.2.5. Steady-State Kinetics Assays. The rate of the enzymatic reaction was determined by measuring O₂ consumption using an O₂-sensitive Clark-type electrode with computer integration via an Oxygraph electrode control unit (Hansatech, King’s Lynn, Norfolk, England). Prior to each assay, the electrode was standardized with air saturated water and water depleted of O₂ by addition of sodium hydrosulfite as described by the manufacturer. Stock solutions and buffers were prepared fresh daily. Prior to each experiment, 100 µL of LigAB (60 µM) was thawed and buffer exchanged into degassed 50 mM Tris (pH 7.5) containing 1:9 1-butanol/H₂O using a 3 mL Sephadex G-25 desalting gel (GE Healthcare) column in a glove-bag flushed with high purity N₂ for 1 h. Exchanged enzyme (3-7 µM) was kept under inert atmosphere (N₂) in a sealed vial on ice for the duration of each experiment.

The effect of pH on the rate of LigAB O₂ consumption was investigated in the pH range of 6-10 at 23 °C. The buffer solutions (50 mM) used to span this pH range were: phosphate (pH 6 to 8), Tris (pH 7.5 to 9), BICINE (pH 8 to 9), and CAPSO (pH 9 to 10). The initial rate assays
were performed in a final volume of 1 mL of air-saturated 50 mM buffer of desired pH, and 250 µM PCA. The reaction was initiated by addition of 1-2 µL LigAB (3-7 nM final concentration) after the equilibration of all other components for at least 1 min to obtain a constant background O₂ consumption rate of ± 0.5 µM/min. Reaction velocities were calculated from the slope of the first 30 s of data after LigAB addition and corrected for background O₂ consumption using 30 s of data immediately prior to LigAB addition. The precise enzyme concentration for each experiment was determined by Bradford assay (Bio-Rad) after completion. The pKa values of affected active site residues were determined by a fit of the reaction velocities as a function of pH to Equation 2.1 (1 acidic pKa, and 1 basic pKa) and Equation 2.2 (2 indistinguishable acidic pKa’s, and 1 basic pKa).

\[
\text{Eq. 2.1} \quad k = \frac{k_{\text{cat}}}{1 + [H^+] K_{a1} [H^+]} + K_{a2} [H^+] \\
\text{Eq. 2.2} \quad k = \frac{k_{\text{cat}}}{1 + \left( \frac{[H^+]}{K_{a1}} \right)^2 + K_{a2} [H^+]} 
\]

Steady-state kinetic parameters for LigAB with respect to the organic substrate were determined by measuring the rate of O₂ consumption in the presence of varying concentrations of organic substrate (1 µM to 5000 µM). An aqueous stock solution (25 mM) of the desired organic substrate (PCA, GA, or 3OMG) was prepared immediately prior to use in a 10 mL volumetric flask. Stock solutions of 3OMG (25 mM) were prepared in 1:9 DMSO/H₂O. The initial rate assays were performed, as described above, in air-saturated 50 mM Tris (pH 7.5) and initiated by the addition of 1-2 µL LigAB (3-7 nM). Steady-state kinetic parameters with respect to O₂ were measured in 50 mM Tris (pH 7.5), 1 mM PCA (20x Kₘ), and 40-450 µM O₂ and initiated by the addition of LigAB (3-7 nM). The buffer was equilibrated prior to each reaction with a fixed mixture of O₂ and N₂ gas using a Cole-Parmer gas proportioner (Vernon Hills, IL), and the reaction chamber was maintained under an atmosphere of the same O₂/N₂.
mixture. The precise enzyme concentration for each experiment was determined by Bradford assay (Bio-Rad) after completion. Kinetic parameters were determined by a least-squares fitting of the Michaelis-Menten equation to the data using KaleidaGraph (Synergy).

2.2.6. Substrate Induced Enzyme Inactivation. The partition ratios (the number of substrate turnovers per enzyme prior to enzyme inactivation) and apparent rate constants of enzyme inactivation \( j_{\text{inact}}^{\text{app}} \) in the presence of PCA, GA, 3OMG, and DHBAm were calculated using methods previously described for mechanism-based inactivation of dioxygenases.\(^{100, 145}\) Using the mechanism described in Scheme 2.2, the apparent rate constant of inactivation \( j_{\text{inact}}^{\text{app}} \) was determined. Although the exact mechanism of LigAB is unknown, product release has previously been determined to be the rate limiting step for the Type I EDO homoprotocatechuate 2,3-dioxygenase.\(^{146}\) If the same holds true for LigAB, then \( j_{\text{inact}}^{\text{app}} = \sum j_i \), and reports on all possible routes of inactivation. This assumption may change upon elucidation of LigAB’s mechanism of catalysis, and product release is not found to be rate limiting.

\[
\begin{align*}
& E_5^* \\
& j_5 \\
& AEA \\
& A \\
& E + A + O_2 \longrightarrow EA + O_2 \longrightarrow EAO_2 \longrightarrow EP \longrightarrow E + P \\
& O_2 \mid j_1 \\
& j_2 \\
& j_3 \\
& j_4 \\
& E_1^* \\
& E_2^* \\
& E_3^* \\
& E_4^*
\end{align*}
\]

Scheme 2.2. General mechanism of LigAB inactivation. The rate constants \( (j_i) \) indicate possible reactions that lead to inactive enzyme \( (E_i^*) \). Adapted from Ref. 145.

The rate constant of inactivation at a given substrate concentration \( (j_s, s^{-1}) \) was determined by the fitting of individual progress curves to Equation 2.3 using KaleidaGraph.\(^{100, 145}\) Prior to fitting, data from the progress curves were modified from consumption of \( O_2 \) to reflect product formation, using the 1:1 stoichiometric ratio of \( O_2 \) to product (Please see Appendix Figure A.1).
All data points from moment of enzyme addition (reaction initiation) to the termination of data collection were retrieved (the oxygen concentration of the reaction solution was recorded from the electrode every 0.1 s). The recorded time of reaction initiation was converted to \( t = 0 \), and the oxygen concentration of this point set to a product concentration of zero. The modified data were then fit using equation 2.3 where \( P_t \) is the concentration of product formed at time \( t \), \( P_\infty \) is the maximal concentration of product formed at reaction completion, and \( P_i \) is the concentration of the product at the start of the reaction.

\[
\text{Eq. 2.3} \quad P_t = P_\infty \left(1 - e^{-j_s t}\right) + P_i
\]

The apparent rate constant of inactivation \((j_{\text{inact}}^{\text{app}})\) was subsequently determined by a fit of Equation 2.4 to the values of \( j_s \) as a function of substrate concentration, for those substrates showing a substrate concentration dependent inactivation. In this equation, \( K_m^{\text{app}} \) is the \( K_m \) of the organic substrate in air-saturated buffer. For substrates not showing a concentration dependence of the apparent rate constant of inactivation, \( j_{\text{inact}}^{\text{app}} \) was calculated as the average of the \( j_s \) values over the range of substrate concentrations.

\[
\text{Eq. 2.4} \quad j_s = j_{\text{inact}}^{\text{app}} \frac{[S]}{K_m^{\text{app}}+[S]}
\]

The partition ratio was then calculated using the \( k_{\text{cat}}^{\text{app}} \) values determined from the steady-state kinetics assays and the \( j_{\text{inact}}^{\text{app}} \) value determined above, using Equation 2.5.

\[
\text{Eq. 2.5} \quad \text{partition ratio} = \frac{[\text{substrate consumed}]}{[\text{enzyme inactivated}]} = \frac{k_{\text{cat}}^{\text{app}}}{j_{\text{inact}}^{\text{app}}}
\]

### 2.2.7. Metal Dependence.

The influence of different divalent metals on the PCA dioxygenase activity was tested via a method adapted from Diaz, et al.\textsuperscript{129} Anaerobically purified enzyme was buffer exchanged into degassed 100 mM Tris (pH 7.5) containing 1:9 \( t \)-butanol/H\_2O, and subsequently incubated under N\_2 in a sealed vial with 2 mM 2,2\textsuperscript{'}-dipyridyl (2,2\textsuperscript{'}-bipyridine, bipy) at 4 °C for 1 h and 24 h. Apo-enzyme reactivation was attempted by
incubation of the apo-enzyme solution in a 2 mM solution of the desired divalent metal salt (FeSO$_4$, CuSO$_4$, CoSO$_4$, MnSO$_4$•H$_2$O, MgSO$_4$•7H$_2$O, NiSO$_4$, or ZnSO$_4$) under anaerobic conditions for 1 h and/or 24 h at 4 °C. The activities of untreated enzyme (no bipy added, after 1 h and 24 h), apo-enzyme samples, and samples incubated with metal salts were determined by O$_2$ consumption assays of 250 µM PCA in 50 mM Tris (pH 7.5). The concentration of enzyme in the buffer exchanged stock was determined by Bradford assay, and the concentration of enzyme in bipy and metal salt treated samples was calculated by calculating the dilution of the buffer exchanged stock.

2.2.8. Product Determination. LigAB reaction products were identified by two different methods. The reaction product of the PCA starting material was generated under O$_2$ saturated conditions in 1 mL of 50 mM phosphate buffer (pH 7.5), 2 mM PCA, and incubation with ~10 µg (50 µL of a 4.5 µM stock) of anaerobically purified LigAB added portion-wise over 1 h. The reaction was monitored by $^1$H NMR for disappearance of the PCA aromatic chemical shifts, with 3-(trimethylsilyl)-propionic-2,2,3,3-d$_4$ acid sodium salt (DSS, Aldrich) as a reference. Additionally the product, CHMS, was converted to 2,4-lutidinic acid as previously described by the addition of 200 µL of concentrated NH$_4$OH.$^{147-149}$ After stirring at room temperature for 2 h, the $^1$H NMR spectrum was recorded. All aqueous $^1$H NMR spectra were recorded in 1:9 D$_2$O/H$_2$O with water suppression. The UV-Vis spectra of the CHMS reaction product were recorded under reaction (pH 7.5), acidic (pH 1), and alkaline (pH 14) conditions by the addition of 4 M HCl or 4 M NaOH respectively. The UV-Vis spectrum of the product formed by the reaction of CHMS with ammonium hydroxide was also recorded.

2.2.9. Substrate Promiscuity Assessment. A wide range of potential substrates were screening using methods analogous to those previously described, using anaerobically purified LigAB and 5 mM compound. Reactions were performed in 1 mL of air-saturated 50 mM Tris
(pH 7.5) with 5 mM substrate at 25 °C and monitored by oxygen consumption. LigAB was added to initiate the reaction as previously described. In addition to stock solutions of PCA and gallate (25 mM in H$_2$O), and 3OMG (25 mM in 1:9 DMSO/H$_2$O), stock solutions of the organic substrates used for alternative substrate determination were prepared as 25 mM solutions in H$_2$O (3,4-dihydroxybenzaldehyde, 2,3,4-trihydroxybenzoic acid, 2,3-dihydroxybenzoic acid, homoprotocatechuate, 4-hydroxybenzoic acid, 3,4-dihydroxybenzamide, 1,2-dihydroxy-4-aminobenzene, 4-nitrocatechol, catechol, 4-methyl catechol, 3-methoxycatechol, 2-aminophenol, 3,4-dihydroxybenzonitrile), 1:9 DMSO/H$_2$O (benzoic acid, vanillic acid, syringic acid, 2-iodobenzoic acid, vanillin, 3,4-dimethoxy benzaldehyde, syringaldehyde, 3-ethoxy-4-hydroxy benzaldehyde, methyl-3,4-dihydroxybenzoate, 2-methoxy-4-methylphenol, 3-methoxyphenol, veratrol) or as 12.5 mM solutions in 1:4 DMSO/H$_2$O (caffeic acid, p-coumaric acid). In an effort to see activity from potentially poor substrates, varying concentrations of LigAB (3-350 nM) were added to assays of the organic substrates. Assays with an addition of anaerobic enzyme greater than 2 µL exhibited O$_2$ dilution effects upon addition. The dilution effect was corrected by subtraction of the progress curve for addition of an equivalent volume of degassed 50 mM Tris, 1:9 t-butanol/H$_2$O buffer containing no enzyme. Kinetic parameters were determined for substrates identified to have significant reactivity. Substrates found to have low levels of O$_2$ consumption activity are reported as a percentage of the activity of LigAB with 5 mM PCA.

The oxidation potentials of a series of PCA analogues, where the C1 substituent was varied, were calculated using Gaussian '09 (Wallingford, CT). Initial structures were minimized by a molecular mechanics procedure, starting from several different initial geometries of the dianion (deprotonation of the C1 carboxylic acid and C4 hydroxyl) or the monoanion (deprotonated C4 hydroxyl) to insure location of the global minimum, followed by complete
geometry optimization at the density functional B3LYP/6-31+g(d)-PCM level. Because solvation is especially important with charged species, solvation energies of all species were computed by the commonly used polarized continuum method of Tomasi. The computed free energies (in kcal/mol) of each parent anion and the corresponding neutral species formed by removal of an electron from the parent were computed. The difference ($\Delta G_{\text{comp}}$) between the two energies is a measure of the ease of electron removal from the anion. $\Delta G_{\text{comp}}$ was then converted to the absolute oxidation potential (independent of reference electrode) of the substance through the relationship $\Delta G_{\text{comp}} = -\mathcal{F}(\Delta G_{\text{comp}})$, in which the Faraday ($\mathcal{F}$, the proportionality constant between electrochemical and chemical free energies) = 23.06 kcal/volt. Absolute potentials have been shown to be highly linearly correlated with experimental oxidation potentials in other systems.

2.2.10. Inhibition Assessment. All molecules that exhibited no oxygen consumption upon incubation with LigAB were examined for their ability to bind to and inhibit the dioxygenation of PCA by LigAB. Stock solutions of all molecules were prepared as previously described. Reactions were performed in 1 mL of air-saturated 50 mM Tris (pH 7.5) with 500 µM PCA and 1 mM potential inhibitor at 25 °C, and monitored by oxygen consumption; LigAB was added to initiate the reaction (3-7 nM). Inhibition was determined by comparison of the rates of inhibitor containing assays to the rates obtained from assays performed directly before or after where only 500 µM PCA (and no potential inhibitor) was present.

2.2.11. pH Effect on Iron Binding. Thawed anaerobically purified enzyme was buffer exchanged using Sephadex G-25 desalting gel as above into degassed 50 mM Tris (pH 7.5), 1:9 t-butanol/H$_2$O to remove excess iron from storage. To account for differences in enzyme concentration from the separate desalting of multiple enzyme samples, a 30 µL aliquot of each enzyme sample was removed for use in a Bradford assay to determine the enzyme
concentration used in each of the following conditions. Desalted enzyme was subsequently buffer exchanged under ambient conditions into 50 mM buffer of pH 6 (phosphate), 7.5 (phosphate), and 10 (CAPSO) using Amicon Ultra 0.5 mL centrifugal filters with a 10 kDa molecular weight cutoff (Millipore). After buffer exchanging, the protein samples were diluted to their original volume with water, and a Ferene-S assay adapted for the 96-well plate format was performed, as described by Capyk, et al.,\textsuperscript{154} to determine the iron content of all samples. The iron concentration was divided by the concentration of enzyme determined from the Bradford assay to give the relative iron-enzyme ratios.
2.3. Results and Discussion

LigAB has been the subject of several studies since it was first reported as a protocatechuate 4,5-dioxygenase in 1990 by Noda, et al. The crystal structure, the first for this superfamily, and reaction kinetics (specific activity) of LigAB have been previously reported and are often cited by those investigating dioxygenase enzymes believed to be members of the PCAD superfamily; however, in these previous studies, LigAB was expressed and purified aerobically. Our attempts to mimic these purification conditions did indeed lead to the production of active enzyme. Despite this success, our initial experiments to confirm the kinetic parameters of the enzyme revealed inconsistent initial rates at a given organic substrate concentration as well as loss of activity (~20 % per hour, $T_{1/2} = 2.25$ h) after thawing from storage (Figure 2.2, assays for these initial studies were performed under previous literature conditions for LigAB using a buffer consisting of 50 mM 3,3-dimethylglutarate, 50 mM Tris, and 50 mM 2-amino-2-methyl-1,3-propanediol at a pH 8.5, knowing that assay conditions might be changed by the results of subsequent experiments). Incubation of aerobically purified enzyme with reductants, such as dithiothreitol or ascorbate, failed to recover or enhance the activity. The adoption of the anaerobic protein purification techniques described here and by...

**Figure 2.2.** (a) Steady-state kinetics of LigAB with PCA: anaerobically purified (○) and aerobically purified (■). (b) Steady-state kinetics of aerobically purified LigAB (●) and the time dependent inactivation of aerobically purified enzyme after thawing (■) (pH 7.5, 25 °C).
others\textsuperscript{155-157} has produced pure co-purified LigAB with greater than ten-fold higher activity over that of the aerobically purified enzyme and lower loss of activity over time after thawing from storage. This large difference in activity can likely be attributed to the anaerobically purified enzyme being properly occupied by the active Fe(II) cofactor, and the aerobic purification conditions resulting in premature exposure of bound Fe(II) to oxygen which ultimately leads to an oxidized and inactive Fe(III). Additionally, no loss of activity has been observed in the anaerobically purified enzyme stored at -80 °C over a period of 6 months.

2.3.1. pH Effects on Rate. Despite exhaustive searching, a previous determination of the pH rate maximum of LigAB was not found. The pH rate maximum of a related enzyme, gallate 3,4-dioxygenase (DesB), from the same LDAC degradation pathway was determined, by oxygen consumption, to be 8.5 in the presence of gallate.\textsuperscript{140} Kinetic parameters (specific activity) were previously determined for LigAB as well as DesB and DesZ (3-O-methylgallic acid 3,4-dioxygenase) using the pH optimum conditions for DesB. The data collected here were fit to both Equation 2.1 (which describes the presence of one acidic and one basic pKa) and Equation 2.2 (which describes the presence of two indistinguishable acidic pKa’s and one basic pKa). Here, LigAB was found to have a pH

\textbf{Figure 2.3.} pH rate profiles of LigAB with PCA at 25 °C, using buffers (▲) Tris, (●)Phosphate, (■) BICINE, (♦) CAPSO. (A) $k_{\text{cat}}$ vs. pH with fitting to eq. 1 (solid) and fitting to eq. 2 (dashed). (B) log($k_{\text{cat}}$) vs pH fit to the log(eq. 1) and the log(eq. 2), solid and dashed lines respectively.
rate maximum of 7.5 in phosphate buffer for the consumption of O_2 in the presence of PCA (Figure 2.3a) when fit to Equation 2.2 (R^2 = 0.97 compared to R^2 = 0.89 for Equation 2.1). This pH rate maximum is consistent with pH values found for other Type II EDOs, including PCA and homoprotocatechuate (HPCA) dioxygenases. The pK_a values determined from the fit of Equation 2.2 suggest that there are two residues with a pKa of 6.3 and a third residue with a pKa of 9.7. The slope of the acidic arm, 1.8, from the plot of log(k_cat) as a function of pH and fit to Equation 2 (Figure 2.3b) also supports the presence of two residues with acidic pK_a values, and the slope of the basic arm, 0.6, supports the presence of one residue with a basic pKa.

In addition to the pH rate effect, changes in buffering agents also influenced the reaction rate, as can be seen in the pH region of 7.5 to 9 (Figure 2.3a). The effect was most evident at pH 7.5 where the initial rate of reactions performed in phosphate buffer displayed a small rate enhancement (20 %) over the initial rate of reactions performed in Tris buffer. The small size and the diol like structure of Tris may contribute to the ability for Tris to possibly bind and occupy the LigAB active site, thus causing a mild inhibitory effect. Despite the rate enhancement in phosphate buffer, assays of LigAB activity were performed in Tris buffer to remain consistent with previous studies and the pH was adjusted to 7.5 corresponding to the pH rate maximum found here.

LigAB contains two active site histidine residues, His127β and His195β, which based upon homology to mechanistically characterized Type I and Type III EDOs are believed to aid in substrate alignment and play
a role as a catalytic base and catalytic acid, respectively (Figure 2.4). His127β is likely one of the two active site residues that becomes protonated and leads to decreased activity at lower pH. The second protonation at low pH may occur on one of the iron chelating residues, leading to a loss of iron from the active site and subsequently a decrease in activity (see section 2.3.3. for discussion of metal dissociation experiments). Additionally, this second protonation event could also correspond to the protonation of the substrate carboxylic acid which would also lead to decreased activity. The loss of activity at high pH corresponds to a single deprotonation event. Although His195β has been suggested to play a role in substrate alignment in the active site, it is possible that this residue plays an additional role as a catalytic acid, based upon homology to mechanisms of other dioxygenases.\textsuperscript{134, 158} Additional explanations for the decrease in activity at high pH corresponding to the relatively high pKa of 9.7 could include oxidation of the active site iron, or deprotonation of pKa depressed Ser269β. Deprotonation of Ser269β would create an unfavorable charge-charge interaction with the carboxylic acid of PCA, preventing proper alignment in the active site. Further mechanistic analysis, including mutagenesis of these residues, will be performed to test mechanistic assignments.

2.3.2. Metal Dependence. In metalloenzymes, the relative promiscuity of active site ligands to bind a variety of different metal ions of the same oxidation state makes it necessary to carefully determine which metal(s) are able to promote catalysis. In a variety of enzyme superfamilies, like the amidohydrolase superfamily, the number of metal ions necessary for catalysis and also the identity of the metal ion(s) can be an important diagnostic for membership in an enzyme subtype, which can be informative when trying to determine the reaction catalyzed.\textsuperscript{159, 160} Since non-heme iron extradiol dioxygenases have been observed to be catalytically active with a variety of metal ions, most commonly Fe\textsuperscript{2+}, but also with other divalent metals such as Mn\textsuperscript{2+}, Co\textsuperscript{2+}, Cu\textsuperscript{2+}, Zn\textsuperscript{2+}, or Ni\textsuperscript{2+}, we sought to determine the metal ion
dependency of LigAB. Since many non-heme iron dioxygenases are expressed and purified aerobically, and subsequently dosed with an iron(II) salt (or other divalent metal ions) to reconstitute the enzyme activity, LigAB was grown and purified in the absence of any exogenous metal. Despite the absence of added iron(II) to the growth media or lysis buffer, the enzyme co-purified with bound iron cofactor under both aerobic and anaerobic conditions, in a catalytically active state.

In order to obtain apo enzyme, LigAB was incubated with alternately EDTA and bipy (2,2′-bipyridine, a commonly used transition metal chelating agent) to remove the liganded metal. An incubation of 1 h with either bipy or EDTA, as described by Diaz, was insufficient to completely eliminate LigAB activity suggesting that LigAB has a more tightly bound and/or less accessible iron center than other dioxygenases (LigAB maintained 65 % activity upon incubation with EDTA and 50 % activity upon incubation with bipy, after 1 h incubation at pH 7.5). In the case of LigAB, complete loss of activity was only achieved by incubation of enzyme with 2 mM bipy for 24 h. Incubation of LigAB with EDTA (1 or 2 mM) for the same time frame yielded enzyme which maintained 35 % activity. Recovery of activity for the apo enzyme (generated with 24 h incubation with bipy) was only observed with a 1 h incubation of Fe²⁺ under anaerobic conditions; however, only 5 % of the native activity was recovered. Longer incubation times with Fe²⁺ had no positive effect on the recovery of activity. Additionally, no recovery of activity was observed with any of the other metal ions. We speculate that, based upon our observation that LigAB activity is lost over time for reasons other than Fe removal, the long incubation times needed to strip the enzyme of Fe²⁺ lead to a protein solution containing misfolded, collapsed or otherwise inactivated protein incapable of metal binding in an active complex, which would explain only a partial recovery (5 %) of enzyme activity. Interestingly, LigAB stripped of 50 % activity (generated by incubation with 1 mM EDTA for
2 hour) and reconstituted with 2 mM Cu$^{2+}$ for 1 h lost activity further to 8 % of that of the native enzyme. Incubation of partially active enzyme with either Co$^{2+}$ or Mn$^{2+}$ salts did not have this added effect of activity loss. This result suggests that, while not able to form an active complex, Cu$^{2+}$ may be able to out-compete Fe$^{2+}$ for the metal binding ligands in the active site. Inhibition by addition of non-native metal ions (Co$^{2+}$, Mn$^{2+}$, Ni$^{2+}$, or Cu$^{2+}$) has been observed in previous studies of other 4,5-PCDs.$^{149,164}$

2.3.3. Iron Dissociation. To investigate the potential contribution of Fe$^{2+}$ dissociation to the decreases in rate observed in the pH rate profile, the iron content of enzyme exchanged and incubated at pH 6.5, 7.5 and 10 was analyzed (Table 2.1) spectrophotometrically via a 96-well plate adapted Ferene-S assay. Binding of the Ferene ligand [3-(2-pyridyl)-5,6-di(2-furyl)-1,2,4-triazine-5′,5″-disulfonic acid disodium salt], initially yellow in aqueous solution, to Fe$^{2+}$ produces a complex with a sharp absorption maximum at 593 nm. Using a standard curve of known Fe$^{2+}$ concentrations, the Fe$^{2+}$ concentration of unknown samples can be quantitatively determined.

At these pH values, the iron to enzyme ratios were found to be 0:1, 1:5, and 1:1, respectively (note: enzyme assayed directly after the desalting column was found to have an iron to enzyme ratio of 4:1, suggesting some free or adventitiously bound iron is also present, leading to the routine use of the Bradford assay to determine the enzyme concentration). The results of the pH incubation and buffer exchange suggest that at low pH, a decrease in catalytic rate, in addition to the rate effect from a

<table>
<thead>
<tr>
<th>Condition</th>
<th>LigAB</th>
<th>Fe</th>
<th>Fe/LigAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desalted</td>
<td>3.4 ± 0.6</td>
<td>14.92 ± 0.5</td>
<td>4.4</td>
</tr>
<tr>
<td>pH 6</td>
<td>3.5 ± 0.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>pH 7.5</td>
<td>5.5 ± 0.5</td>
<td>1.0 ± 0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>pH 10</td>
<td>3.9 ± 0.2</td>
<td>4.1 ± 0.3</td>
<td>1.04</td>
</tr>
</tbody>
</table>
protonated catalytic base residue, may be caused by one or more iron binding residues being protonated leading to an increased rate of dissociation of Fe\(^{2+}\) from the active site. This is consistent with the slope of 1.8 observed in the acidic arm of the plot of pH vs log\(k_{\text{cat}}\). The slope greater than 1 suggests additional mechanisms of inactivation beyond the protonation of a catalytic base.

The 1:1 ratio of iron to enzyme found at pH 10 suggests that the iron binding residues remain deprotonated and able to bind Fe at higher pH. At this high pH, oxidation of Fe\(^{2+}\) to Fe\(^{3+}\) occurs much more rapidly than at a pH of 6 or even 7.5,\(^{165}\) and so the Fe remaining bound to LigAB at this pH is likely in the Fe\(^{3+}\) state. The reduced \(k_{\text{cat}}\) at high pH observed in the pH rate profile may then have a contribution from iron oxidation in addition to the rate reduction caused by a deprotonated catalytic acid residue. The slope corresponding to the basic arm of the pH rate profile is less than 1 and so the contributions of either mechanism of deactivation are unclear. Additionally, the 1:1 Fe/enzyme ratio at pH 10, in comparison to the ratio of 1:4 observed at pH 7.5, suggests that the inactive Fe\(^{3+}\) state is more tightly bound to LigAB than Fe\(^{2+}\). While the non-stoichiometric occupancy of Fe at pH 7.5 may seem surprising, the buffer exchanging process likely reduces the concentration of Fe\(^{2+}\) in the solution below LigAB’s \(K_d\) for Fe\(^{2+}\) allowing the ion to dissociate more freely. At pH 6 no residual iron was observed after buffer exchanging, suggesting that iron, when not bound to a complex larger than 10 kDa, is capable of being reduced to undetectable levels by this method. Thus, at pH 7.5 the remaining iron is likely enzyme bound and the reduced iron content is due to a \(K_d\) for Fe\(^{2+}\) that is higher than the residual iron concentration after complete buffer exchange.

To the best of our knowledge, the pH related dissociation of Fe from a dioxygenase of any type has not previously been measured; however, Takemori et al. previously analyzed the pH dependence of holoenzyme reconstitution for the catechol 2,3-dioxygenase from *Pseudomonas*
*putida* (XylE), and observed that enzyme reactivation occurred most quickly at pH 6.0 ± 0.2, as compared with the exchange rates observed at pH 6.6.\textsuperscript{166} We speculate that these differences in enzyme coordination behavior could be related to membership in different dioxygenase superfamilies, amongst other possible reasons. Further, $K_d$ values for Fe$^{2+}$ at constant pH are rare, and values for Fe$^{3+}$ have not been found in literature searches. The mononuclear non-heme dioxygenases cysteine dioxygenase (CDO) from rat and β-diketone-cleaving enzyme (Dke1) from *Acinetobacter johnsonii* were found to have $K_d$ values for Fe$^{2+}$ of 5.2 µM and 5.8 µM, respectively.\textsuperscript{167, 168} While these values may not be directly applicable to LigAB since CDO and Dke1 are Cupin dioxygenases with a 3-His iron binding motif in contrast with the 2-His-1-Glu motif found in LigAB, they provide a value for reference that is not commonly measured for dioxygenase enzymes. While an iron $K_d$ value for LigAB is not elucidated through this experiment, thawed enzyme buffer exchanged via the desalting gel column, has an $[\text{Fe}^{2+}]_{\text{Tot}} = 14.9 ± 0.5$ µM (higher than the $K_d$ values determined for CDO and Dke1) suggesting LigAB should be fully occupied assuming a similar $K_d$.

Much more in regards to the iron environment of LigAB remains to be elucidated. Further studies will be required to determine the spin state of the iron in addition to any oxidation state changes that may occur during the mechanism of dioxygenation. While not yet determined for LigAB, we can speculate that the iron(II) cofactor is in the high spin state similar to that determined for the Protocatechuate 4,5-dioxygenase from *Pseudomonas testosteroni*.\textsuperscript{169}

**2.3.4. UV-Vis Characterization of LigAB Activity and Product Determination.** The product of the dioxygenation of PCA by LigAB, 4-carboxy-2-hydroxymuconate-6-semialdehyde (CHMS), was identified by UV-Vis spectroscopy and $^1$H-NMR, aided by a secondary conversion of CHMS to 2,4-lutidinic acid. Since CHMS is involved in an extensive equilibrium of structural isomeric forms, its isolation is difficult. The UV-Vis spectra of CHMS under
multiple pH conditions have been previously reported, and the spectra recorded here (Figure 2.5a) are consistent with previous data showing a $\lambda_{\text{max}}$ of 298 nm and a second $\lambda_{\text{max}}$ of 410 nm at pH 7.5. Additionally, the spectrum of PCA at pH 7.5 is reported here. The spectral changes observed upon reducing the pH to 1 or increasing the pH to 14 are also consistent with previous observations of CHMS (Figure 2.5b). The conversion of CHMS to 2,4-lutidinic acid also provides a diagnostic UV-Vis spectral change that is consistent with the conversion to this product. In addition to the UV-Vis spectral changes, the reaction was monitored by $^1$H-NMR spectroscopy (Figure 2.6). The disappearance of PCA can be clearly observed on the NMR time scale; however, peaks corresponding to product formation were not observed in the $^1$H-NMR spectrum of the reaction at completion due to the equilibrium of multiple product isomers. Addition of NH$_4$OH to convert CHMS to 2,4-lutidinic acid results in

Figure 2.5. UV-Vis spectra of LigAB catalysis with PCA. (A) Spectra of PCA (solid) and CHMS at reaction completion (dashed), pH 7.5. (B) pH dependent UV-Vis spectral changes of CHMS (pH 7.5, dashed; pH 2, dots; pH 14, dash-dot) and the spectrum of 2,4-lutidinic acid (solid) formed upon the addition of NH$_4$OH. (C) Reaction scheme for the conversion of CHMS to 2,4-lutidinic acid.
the appearance of three aromatic proton signals consistent with a spectrum of 2,4-lutidinic acid from the Sigma-Aldrich spectral library.172

Prior to understanding of the need for anaerobically purified enzyme, monitoring of LigAB’s catalytic turnover of aromatic substrates (PCA, gallate, 3OMG, and catechol) was initially performed by recording the time dependent UV-Vis spectral changes of the substrates in the presence of enzyme. While changes in the absorption spectra were observed for some substrates (PCA, gallate, and catechol) but not observed for others (catechol) (Figure 2.7), the isolation and spectral characterization of the reaction products proved difficult. Low yields, the relatively long reaction time (~30 min) and subsequent extraction, and the isomerization of the product over this time frame prevented the isolation of a single reaction product. Without accurate extinction coefficient values for both substrates (obtainable) and products of LigAB catalysis, monitoring the catalytic rate of substrate consumption by UV-Vis proved ineffective due to a lack of differential extinction coefficients.

![Figure 2.6](image)

*Figure 2.6* $^1$H NMR of LigAB reaction and product conversion to 2,4-lutidinic acid. Top: 2 mM PCA in 50 mM phosphate buffer, pH 7.5 prior to addition of LigAB. Middle: Spectra after incubation with LigAB for 1 h (reaction completion); peaks corresponding to CHMS were not directly observed due to an extensive equilibrium. Bottom: 2,4-Lutidinic acid formed from direct addition of NH$_4$OH to the assay at completion.
Figure 2.7. Observed UV-Vis spectral changes during substrate turnover by LigAB. A) Reaction scheme showing the conversion of PCA (1) to CHMS (2), and a portion of the extensive equilibrium of tautomers which lead to the formation of a stable cyclic form (3). B.1) Absorption spectrum of PCA B.2) Absorption changes observed during the turnover of PCA to produce CHMS upon addition of LigAB. Absorption decreases at 253 and 292 nm corresponding to substrate consumption while the appearance of an absorption band at 410 nm corresponds to product formation (30 s between scans). B.3) Long term changes in the absorption spectrum of PCA•LigAB assays (beyond 10 min). The absorption band at 292 nm continues to decrease, while the absorption band at 410 nm ceases to increase and begins to decrease. This decrease corresponds to a decrease in the initial ring-opened form of CHMS due to the formation of a stable cyclic form. C) Absorption spectrum changes observed during the turnover of gallate by LigAB (initial spectrum, red). D) Absorption spectrum changes observed during the turnover of 3-O-methyl gallate by LigAB (initial spectrum, red).
2.3.5. Steady-State Kinetics. Although Masai and co-workers previously established the relative rates of reaction for LigAB with PCA, GA, and 3OMG, these experiments were conducted under less than optimal conditions with aerobically purified enzyme and a pH optimum determined for a different, yet related, dioxygenase.\textsuperscript{140} As well, the kinetic parameters of $k_{\text{cat}}$ and $k_{\text{cat}}/K_m$ were not determined. Results of the steady-state kinetics for LigAB with PCA, gallate, and 3OMG obtained in this study are shown in Figure 2.8. The steady-state kinetic parameters are listed in Table 2.2. The values of $K_{m}^{\text{app}}$ for these substrates are consistent with those previously determined; however, here we have found that 3OMG binds much less tightly than previously observed – 2319 ± 316 $\mu$M compared to 937 ± 18 $\mu$M, respectively. The values of $k_{\text{cat}}^{\text{app}}$ determined here for these substrates demonstrate a similar relative ratio of activities to the previously determined specific activity with PCA having the highest turnover,

### Table 2.2. LigAB Steady-State Kinetic Parameters Measured with PCA and Analogous Substrates in Air-Saturated Buffer

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_{m}^{\text{app}}$ (µM)</th>
<th>$k_{\text{cat}}^{\text{app}}$ (s$^{-1}$)</th>
<th>$k_{\text{cat}}^{\text{app}}/K_{m}^{\text{app}}$ (M$^{-1}$ s$^{-1}$)</th>
<th>partition ratio$^a$</th>
<th>$j_{\text{inact}}^{\text{app}}$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>51 ± 4</td>
<td>216 ± 3</td>
<td>4.26 x 10$^6$</td>
<td>~18000</td>
<td>0.012$^c$</td>
</tr>
<tr>
<td>Gallate</td>
<td>441 ± 55</td>
<td>53 ± 2</td>
<td>1.21 x 10$^5$</td>
<td>~2500</td>
<td>0.021$^b$</td>
</tr>
<tr>
<td>3OMG</td>
<td>2319 ± 316</td>
<td>7.3 ± 0.5</td>
<td>3.13 x 10$^3$</td>
<td>~5600</td>
<td>0.0013$^c$</td>
</tr>
<tr>
<td>DHBAm</td>
<td>3263 ± 327</td>
<td>104 ± 5</td>
<td>3.20 x 10$^4$</td>
<td>~3200</td>
<td>0.033$^d$</td>
</tr>
</tbody>
</table>

$^a$ Estimated from the $j_{\text{inact}}^{\text{app}}$ and $k_{\text{cat}}^{\text{app}}$ (eq. 2.5). $^b$ Calculated from a fit of Equation 2.4 to $j_i$ values determined from the fitting of progress curves (eq. 2.3) at different concentrations of gallate. $^c$ Concentration dependent inactivation was not observed for PCA or 3OMG. $j_{\text{inact}}^{\text{app}}$ was calculated as an average of the $j_i$ values determined from the fitting of Equation 2.3 to progress curves at different substrate concentrations. $^d$ Concentration dependent inactivation was observed for DHBAm; however, the behavior is inconsistent with Equation 2.4 due to highly inactivating product-inhibition. $j_{\text{inact}}^{\text{app}}$ was calculated from an average of the $j_i$ values determined from the fitting of Equation 2.3 to progress curves of reactions with 1000-5000 $\mu$M DHBAm.
followed by GA and then 3OMG. The C5-substituent effect on reactivity and substrate specificity is quite apparent from this series, likely due to steric clash with Phe103α (Figure 2.4); however, not all steric and electronic contributions to substrate specificity are evident from these data. Significantly, as described above, the LigAB purified anaerobically in this study is greater than ten-fold more active than enzyme expressed from the same source but purified aerobically. While the previously reported results for LigAB are significant and inspired our work with this enzyme, anaerobic purification techniques allow for a more accurate characterization of LigAB.

The steady-state kinetic parameters determined for LigAB with the native substrate, PCA, as compared to other dioxygenases with their native substrates are listed in Table 2.3 and Table 2.4. Dioxygenases of all types (I-VOC, II-PCAD, and III-Cupin) are quite proficient enzymes, with $k_{cat}/K_m$ values ranging from $10^5$ - $10^8$ M$^{-1}$s$^{-1}$, and LigAB (4.3 x $10^6$ M$^{-1}$s$^{-1}$) fits well...
within this trend. Additionally, the kinetic parameters of LigAB coincide well with those determined for 2,3-PCD, MhpB, and 4,5-PCD (C. testosteroni Pt-L5) - the only other Type II dioxygenases with a complete set of steady-state kinetic parameters determined.\textsuperscript{114, 122, 135} The $K_{mO_2}$ for LigAB (162 ± 16 µM), determined from Figure 2.9, was also found to be in the range of those determined for other extradiol dioxygenases (Table 2.3) and is nearly identical to that determined for the Type II dioxygenase 2,3-PCD (142 ± 14 µM).
Table 2.3. Steady-State Kinetic Parameters for Type I, II, and III Extradiol Dioxygenases

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Subst.</th>
<th>Type</th>
<th>Cleavage</th>
<th>(k_{cat}(s^{-1}))</th>
<th>(K_m(\mu M))</th>
<th>(K_{mO_2}(\mu M))</th>
<th>(k_{cat}/K_m(M^2s^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>LigAB</td>
<td>PCA</td>
<td>II</td>
<td>4,5</td>
<td>216 ± 3</td>
<td>51 ± 4</td>
<td>162 ± 16</td>
<td>4.3 x 10^6</td>
</tr>
<tr>
<td>2,3-PCD</td>
<td>PCA</td>
<td>II'</td>
<td>2,3</td>
<td>210 ± 2.1</td>
<td>24 ± 2.4</td>
<td>142 ± 14</td>
<td>8.8 x 10^6</td>
</tr>
<tr>
<td>MhpB</td>
<td>DHP</td>
<td>II</td>
<td>1,2</td>
<td>50</td>
<td>26</td>
<td>ND</td>
<td>1.9 x 10^6</td>
</tr>
<tr>
<td>C2,3O</td>
<td>Cat</td>
<td>I</td>
<td>2,3</td>
<td>180 ± 11</td>
<td>1 ± 0.09</td>
<td>ND</td>
<td>1.8 x 10^8</td>
</tr>
<tr>
<td>Mpc</td>
<td>Cat</td>
<td>I</td>
<td>2,3</td>
<td>278</td>
<td>1.87</td>
<td>7.5</td>
<td>1.5 x 10^8</td>
</tr>
<tr>
<td>DHBD</td>
<td>DHB</td>
<td>I</td>
<td>1,2</td>
<td>251 ± 6</td>
<td>12 ± 1</td>
<td>1350</td>
<td>2.1 x 10^7</td>
</tr>
<tr>
<td>BphC</td>
<td>DHB</td>
<td>I</td>
<td>1,2</td>
<td>99</td>
<td>0.46-</td>
<td>39.6</td>
<td>2.2 x 10^8</td>
</tr>
<tr>
<td>HPCD</td>
<td>HPCA</td>
<td>I</td>
<td>2,3</td>
<td>7.8 ± 0.3</td>
<td>31 ± 6</td>
<td>60</td>
<td>2.5 x 10^5</td>
</tr>
<tr>
<td>PcpA</td>
<td>diCIHQ</td>
<td>I</td>
<td>1,2</td>
<td>3.2 ± 0.3</td>
<td>3.05 ± 0.05</td>
<td>190 ± 40</td>
<td>1.0 x 10^6</td>
</tr>
<tr>
<td>QueD</td>
<td>QUE</td>
<td>III</td>
<td>2,3 (Int)</td>
<td>(Mn^{2+})</td>
<td>25 ± 1</td>
<td>4 ± 0.9</td>
<td>90 ± 10</td>
</tr>
<tr>
<td>GenDO_{Pi}</td>
<td>GEN</td>
<td>III</td>
<td>1,2</td>
<td>366</td>
<td>85</td>
<td>96</td>
<td>4.3 x 10^6</td>
</tr>
<tr>
<td>GenDO_{Pa}</td>
<td>GEN</td>
<td>III</td>
<td>1,2</td>
<td>321</td>
<td>74</td>
<td>55</td>
<td>4.3 x 10^6</td>
</tr>
<tr>
<td>P25X</td>
<td>GEN</td>
<td>III</td>
<td>1,2</td>
<td>40.25 ± 0.65</td>
<td>92 ± 5</td>
<td>ND</td>
<td>4.4 x 10^5</td>
</tr>
<tr>
<td>P35X</td>
<td>GEN</td>
<td>III</td>
<td>1,2</td>
<td>56.3 ± 0.7</td>
<td>143 ± 7</td>
<td>ND</td>
<td>3.9 x 10^5</td>
</tr>
<tr>
<td>HGO</td>
<td>HGA</td>
<td>III</td>
<td>1,2</td>
<td>56.0 ± 1.1</td>
<td>22.3 ± 0.9</td>
<td>1270 ± 130</td>
<td>2.5 x 10^6</td>
</tr>
<tr>
<td>ARD'</td>
<td>DHKH</td>
<td>III</td>
<td>n.a.</td>
<td>260</td>
<td>52</td>
<td>47</td>
<td>5.0 x 10^6</td>
</tr>
<tr>
<td>OsARD1</td>
<td>DHKP</td>
<td>III</td>
<td>n.a.</td>
<td>11.7</td>
<td>25.9</td>
<td>1100</td>
<td>4.5 x 10^5</td>
</tr>
</tbody>
</table>

For references to each of these systems, please see the supplementary material of the original Biochemistry (2013) 52, 6724-6736 publication.
### Table 2.4. Referenced Enzyme and Substrate Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Enzyme Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>LigAB</td>
<td>protocatechuate 4,5-dioxygenase from <em>Sphingobium</em> sp. SYK-6</td>
</tr>
<tr>
<td>2,3-PCD</td>
<td>protocatechuate 2,3-dioxygenase from <em>Paenibacillus</em> sp. (formerly <em>Bacillus macerans</em>)</td>
</tr>
<tr>
<td>MhpB</td>
<td>2,3-dihydroxyphenylpropionate 1,2-dioxygenase from <em>E. coli</em></td>
</tr>
<tr>
<td>4,5-PCD</td>
<td>protocatechuate 4,5-dioxygenase from <em>Comamonas testosteroni</em> Pt-L5 (formerly <em>Pseudomonas testosteroni</em> PtL-5)</td>
</tr>
<tr>
<td>C2,3O</td>
<td>catechol 2,3-dioxygenase from <em>Pseudomonas stutzeri</em> OX1</td>
</tr>
<tr>
<td>Mpc</td>
<td>catechol 2,3-dioxygenase from <em>Pseudomonas putida</em> mt-2</td>
</tr>
<tr>
<td>DHBD</td>
<td>2,3-dihydroxybiphenyl 1,2-dioxygenase from <em>Burkholderia cepacia</em> LB400</td>
</tr>
<tr>
<td>BpHC</td>
<td>2,3-dihydroxybiphenyl 1,2 dioxygenase from <em>Pseudomonas</em> sp. KKS102</td>
</tr>
<tr>
<td>HPCD</td>
<td>homoprotocatechuate 2,3-dioxygenase from <em>Brevibacterium fuscum</em></td>
</tr>
<tr>
<td>PcpA</td>
<td>2,6-dichlorohydroquinone 1,2-dioxygenase from <em>Sphingobium chlorophenolicum</em> ATCC 39723</td>
</tr>
<tr>
<td>QueD</td>
<td>quercetin 2,3-dioxygenase from <em>Bacillus subtilis</em></td>
</tr>
<tr>
<td>GenDOPt</td>
<td>gentisate 1,2-dioxygenase from <em>Pseudomonas testosteroni</em></td>
</tr>
<tr>
<td>GenDOPa</td>
<td>gentisate 1,2-dioxygenase from <em>Pseudomonas acidovorans</em></td>
</tr>
<tr>
<td>P25X</td>
<td>gentisate 1,2-dioxygenase from <em>Pseudomonas alcaligenes</em> NCIB 9867</td>
</tr>
<tr>
<td>P35X</td>
<td>gentisate 1,2-dioxygenase from <em>Pseudomonas putida</em> NCIB 9869</td>
</tr>
<tr>
<td>HGO</td>
<td>homogentisate 1,2-dioxygenase from humans</td>
</tr>
<tr>
<td>ARD²</td>
<td>aci-reductone dioxygenase from <em>Klebsiella pneumonia</em></td>
</tr>
<tr>
<td>OsARD1</td>
<td>aci-reductone dioxygenase from <em>Oriza sativa</em> (rice)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subst. Abbr.</th>
<th>Substrate Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>protocatechuate</td>
</tr>
<tr>
<td>DHP</td>
<td>2,3-dihydroxyphenylpropionate</td>
</tr>
<tr>
<td>Cat</td>
<td>catechol</td>
</tr>
<tr>
<td>DHB</td>
<td>2,3-dihydroxybiphenyl</td>
</tr>
<tr>
<td>HPCA</td>
<td>homoprotocatechuate</td>
</tr>
<tr>
<td>diCLHQ</td>
<td>2,6-dichlorohydroquinone</td>
</tr>
<tr>
<td>Que</td>
<td>quercetin</td>
</tr>
<tr>
<td>GEN</td>
<td>gentisate</td>
</tr>
<tr>
<td>HGA</td>
<td>homogentisate</td>
</tr>
<tr>
<td>DHKH</td>
<td>1,2-dihydroxy-3-ketohex-1-ene (aci-reductone analogue)</td>
</tr>
<tr>
<td>DHKP</td>
<td>1,2-dihydroxy-3-ketopent-1-ene (aci-reductone analogue)</td>
</tr>
</tbody>
</table>
2.3.6. Substrate Specificity and Inhibition. Intradiol and extradiol dioxygenases have long been known to exist for the preferential cleavage of a wide variety of substrates. Substrate classes include catechols, protocatechuates, gentisates, 2-aminophenols, quinones, aromatic amino acids, and even halogenated aromatics.\textsuperscript{103} Dioxygenase specificities range from the highly specific to the quite promiscuous. LigAB has been shown here and in previous studies to be a promiscuous enzyme capable of efficient catalytic turnover of several substrates in the LDAC catabolic pathway of \textit{Sphingobium} sp. (PCA, GA, and 3OMG). The promiscuity of LigAB with GA and 3OMG contrasts with another Type II dioxygenase, gallate dioxygenase (GDO) from \textit{Pseudomonas putida} KT2440 (39 \% sequence identity to LigB). GDO was found to have no oxygen consumption activity with PCA and is believed to be a gallate specific dioxygenase.\textsuperscript{129} Additionally, several PCA 4,5-dioxygenases (4,5-PCD) have been identified for which substrate specificity, substrate uptake by live cells, and the inhibition effect of other substrates have been studied; however, reaction kinetics remain largely unstudied for these 4,5-PCDs\textsuperscript{115, 117, 147, 161, 164, 179-181} with exception of the 4,5-PCD from \textit{Comamonas testosteroni} Pt-L5 (formerly \textit{Pseudomonas testosteroni} Pt-L5) (Table 2.3).\textsuperscript{114}

Despite the evidence of promiscuity in LigAB, the range of accepted substrates was never explored beyond the PCA analogues identified from within its pathway. Here we have tested multiple classes of substrates analogous to the basic catechol scaffold with substitutions primarily at C1, C5, and functional modification of the catecholic hydroxyl groups (Table 2.5 and Figure 2.10). In addition to the three known substrates, LigAB was demonstrated to utilize four new substrates: 3,4-dihydroxybenzamide (DHBAm), homoprotocatechuate (HPCA), catechol and 3,4-dihydroxybenzonitrile. DHBAm was the most active of the newly identified substrates, having steady-state kinetic parameters similar to those found for the previously identified substrates (Table 2.2). LigAB was found to have the following trend for specificity
LigAB was found to turn over \((k_{\text{cat}})^{\text{app}}\) DHBAm more quickly than GA, but with a larger \(K_{m}^{\text{app}}\) than that of 3OMG. Although DHBAm is a good substrate, \(O_2\) consumption kinetic traces displayed very fast deactivation of LigAB and in most cases only \(~10\) s of data was used to obtain a rate after the initiation of the reaction by addition of LigAB. The rate of inactivation \((j_i)\) during turnover of DHBAm shows substrate concentration dependence; however, DHBAm is more inactivating at lower substrate concentrations than at high substrate concentrations and so cannot be fit using Equation 4. Assays, performed in the presence of DHBAm product formed \textit{in situ}, show enhanced inactivation over assays performed with no product initially present. These observations are consistent with turnover of DHBAm leading to a product that competes with substrate to bind LigAB and resulting in an inactivated enzyme by a dead-end product inhibition mechanism.
Figure 2.10. Structures of molecules used for substrate specificity and inhibition studies. Molecules determined to be substrates of LigAB are shown in green with the corresponding catalytic efficiency (M$^{-1}$s$^{-1}$) or percent activity at a concentration of 5 mM. Molecules determined to be inhibitors of LigAB turnover of PCA are shown in red with the percent inhibition of the rate of LigAB with 500 µM PCA. Molecules shown in black were determined to be neither substrates nor inhibitors.
HPCA, 3,4-dihydroxybenzonitrile, and catechol were poorer substrates, displaying only a small fraction of the activity (less than 1%) as seen for LigAB with PCA (Table 2.5). Interestingly, a number of molecules were identified that were not substrates, but showed competitive inhibition of the LigAB reaction with PCA, including 4-nitrocatechol and 4-methylcatechol (Table 2.5). All of these molecules are analogues of PCA, where the C1 substituent (using the PCA numbering scheme) is varied. To explain these observations we initially performed calculations to determine if the electronic influence of the C1 substituent was correlated with the catalytic efficiency, using Gaussian ‘09 (Table 2.5). From the data, it is clear that a range of organic substrate oxidation potentials (5.40 – 6.30 eV) is tolerated for accepted substrates; however, the incorporation of O₂ is most facile when the oxidation potential of the organic substrate is 5.82 – 6.11 eV, as for PCA and 3,4-dihydroxybenzamide. 4-Methyl catechol and catechol, with oxidation potentials of 5.71 and 5.88 eV respectively, are close to or within the facile range, yet oxygen consumption is not seen or is very low. Substrate like molecules with oxidation potentials below this range were also examined for their ability to be turned over by LigAB; however, these molecules often quickly reacted with dissolved oxygen before LigAB could be introduced (e.g. 1,2-dihydroxy-4-aminobenzene). Some molecules on the fringe of the oxidation potential range such as HPCA (5.40 eV) and 3,4-dihydroxybenzonitrile (6.30 eV) are tolerated as substrates by LigAB but the activity is severely diminished. 4-Nitrocatechol with the highest oxidation potential of the C1 PCA analogues is not a substrate but is rather an inhibitor, suggesting that compounds that have a higher potential than 6.30 eV are not electronically accessible by LigAB.

Since a clear correlation between catalytic efficiency and oxidation potential for the library of C1 analogues of PCA could not be established for all of the compounds in this series, we then evaluated the likely binding interactions of the various molecules with the LigAB active
site residues. We noted that the two worst substrates pyrocatechol and 3,4-dihydroxybenzonitrile, as well as one of the inhibitors 4-methylcatechol would all be unable to have favorable hydrogen bonding interactions with the NH$_2$ group of Asn270β which is predicted to hydrogen bond to the carboxylate anion of PCA (Figure 2.4). Sterics could also explain the relatively low catalytic efficiency of LigAB with HPCA (a C1 substrate analogue which moves the carboxylate further into the pocket by addition of a methylene) and the absence of catalysis or inhibition with methyl-3,4-dihydroxybenzoate and caffeic acid which are likely just too big to fit into the binding site, perhaps due to steric conflict with Thr271β (Figure 2.4). Conversely, the inhibitory nature of 4-nitrocatechol, which is isosteric with PCA and could accept hydrogen bonds from the NH$_2$ group of Asn270β, is likely dominated by electronic effects. In other systems it has been observed that substrate/inhibitor scope is generally smaller than that observed for LigAB. In the 4,5-PCD from C. testosteroni, sulfonylcatechol (a C1 sulfonyl substitution of the PCA carboxylic acid) was the only known substrate analog with a C1 modification to be turned over.$^{114}$ Additionally, catechol (C1-H) and 4-methylcatechol (C1-Me) were observed to be inhibitors of the 4,5-PCD of Pseudomonas sp.$^{161}$ but non-substrates of the 4,5-PCD of C. testosteroni.$^{149,164}$

Interestingly, 3,4-dihydroxybenzaldehyde was neither a substrate nor an inhibitor of LigAB despite the structural similarity to PCA and its seemingly acceptable oxidation potential. The inability to turnover 3,4-dihydroxybenzaldehyde has been observed in 4,5-PCDs from Pseudomonas sp.$^{161}$ and C. testosteroni$^{164}$ however, in these instances 3,4-dihydroxybenzaldehyde does act as an inhibitor. Examination of other aldehyde containing compounds including vanillin, revealed that none of these compounds were substrates or inhibitors for LigAB, despite their structural similarity to viable LigAB substrates. Knowing that vanillin is the substrate of the enzyme two steps ahead of LigAB in the Sphingobium sp.
LDAC catabolic pathway, we hypothesize that perhaps LigAB has evolved to preclude the binding of aldehydes in the active site.

Taken together, the results of the library suggest that there are both structural and electronic requirements which are necessary for organic substrate activity with LigAB. The first requirement is an electron withdrawing group in the C1 position such that the oxidation potential of the substrate is 5.40-6.30 eV (~5.8 eV being the sweet spot). While likely not electron withdrawing in its charged form, the carboxylate of PCA is capable of accepting additional electron density through resonance after deprotonation of the C4-hydroxyl. Secondly, a 3,4- or a 4,5- diol is required. Functional modification of hydroxyls in these positions, such that hydroxyls ortho to one another are not present, eliminates activity. For example 3OMG remains a viable substrate because the 4,5-diol motif is conserved, yet vanillate and syringate which have a methylether rather than a hydroxyl ortho to the C4-hydroxyl displayed no activity. Mutagenesis of various active site residues will be pursued to test for possible gain of function for the various substrates that are believed to be electronically allowed, but sterically disfavored.

2.3.7. Mechanism-Based Inactivation in Steady-State Kinetics. As has been previously reported for other dioxygenase enzymes, LigAB is susceptible to inactivation during catalytic turnover under air-saturated conditions. While inactivation of a Type II dioxygenase has not previously been reported, the rates of inactivation for LigAB (Table 2.2) are similar to those reported for the well characterized Type I dioxygenases 2,3-dihydroxybiphenyl 1,2-dioxygenase (DHBD) and 2,6-dichlorohydroquinone 1,2-dioxygenase (PcpA). Inactivation was seen with each of the four most preferred substrates (PCA, GA, 3OMG, DHBAm); however, only gallate showed substrate concentration dependent inactivation rates that could be fit to Equation 2.4, resulting in a $j_{\text{inact}}^{\text{app}}$ of 0.021 s$^{-1}$. 
Inactivation during turnover of PCA (0.012 s\(^{-1}\)) and 3OMG (0.0013 s\(^{-1}\)) was independent of substrate concentration within the experimental concentration range. While there has been some debate over how mechanism-based inactivation occurs, the strongest evidence suggests that the active site Fe(II) becomes irreversibly oxidized to Fe(III) due to the release of superoxide.\(^{145}\) Superoxide release can occur before or after organic substrate binding – though tends to occur more slowly in the free enzyme due to a lower affinity (higher \(K_{mO2}\)) for oxygen in the absence of substrate as was observed for DHBD.\(^{145,155}\) Additionally, while inactivation tends to be more pronounced in the presence of poor substrates, inactivation has also been observed in the presence of the preferred substrate(s).\(^{145}\) While the data presented here does not strictly follow this trend (Table 2.2), the rate of inactivation for the three most preferred substrates (PCA, gallate, and DHBAm) is consistent with this observation - with PCA being the least inactivating and DHBAm being the most inactivating of these substrates. 3OMG, despite being the poorest of the preferred substrates, displayed the least inactivation. The increased sterics of 3OMG likely prevent this substrate from properly binding; however, once bound in a geometry suitable for catalysis, 3OMG may be more active towards C-O bond formation with superoxide thus preventing premature release and slowing the oxidation of the LigAB active site iron.

### 2.4. Conclusions

Steady-state kinetic studies have demonstrated that LigAB is highly specific for catecholic substrates with an electron withdrawing/hydrogen bond accepting substituent on C1. While LigAB was previously known to have promiscuous activity, this promiscuity has now been expanded beyond molecules of the LDAC degradation pathway of *Sphingobium* sp. SYK-6. Analysis of the structure and oxidation potentials of the accepted substrates suggests that
LigAB is highly specific for diol containing substrates with an oxidation potential between 5.40-6.30 eV. Additionally, steric and hydrogen bonding interactions by active site residues (such as Phe103α, Thr271β, or Ser269β) likely impart some of LigAB’s substrate specificity – modulating the catalysis of accepted substrates and preventing some substrate-like molecules that meet these parameters from being acceptable substrates; mutagenesis of LigAB will likely reveal the relative importance of these residues in controlling substrate specificity. LigAB, like other dioxygenases, is subject to mechanistic inactivation although not all substrates show a concentration dependence on the rate of inactivation. In the case of the newly discovered substrate DHBAm, dead-end product inhibition dominates the mechanisms of inactivation. While inhibition of LigAB is not directly applicable to our research goals, a number of competitive inhibitors were also identified in this study.

These results could provide insight toward identifying lead compounds in a system where dioxygenase inhibition is desired. Furthermore, the pH rate studies of this enzyme suggest the possibility of either (1) a catalytic acid and a catalytic base involved in catalysis for LigAB or (2) inactivation due to alterations of the Fe(II) bound in the active site. Further mechanistic analyses, including mutagenesis might shed light on these results and allow comparisons to mechanisms proposed for dioxygenases from other enzyme superfamilies. In total, these results make LigAB the first Type II dioxygenase to be fully characterized both structurally and kinetically, and solidify its position as the defining member of the Type II dioxygenase superfamily.
Chapter 3:

Exploring Allosteric Activation of LigAB from Sphingobium sp. strain SYK-6 through Kinetics, Mutagenesis, and Computational Studies

Adapted with permission from:
©2015 Elsevier
3.1. Introduction

LigAB is known for its natural propensity to turnover multiple catechol derivatives including protocatechuate (PCA), gallate (GA), and 3-O-methyl gallate (3OMG) (as shown in Ch. 2). While our previous LigAB competition studies identified several molecules capable of inhibiting PCA turnover, we also identified several catechol derivatives that enhance LigAB’s turnover ($k_{cat}$) when co-incubated with PCA. While inhibition of dioxygenases by catechol derivatives has been widely reported in ring cleaving extradiol dioxygenases, no reference to activity ($k_{cat}$) enhancement in these systems could be found.

We report in this chapter the first evidence for the heteroallosteric activation of LigAB by vanillin, a metabolically linked small molecule effector. Since vanillin is the pre-pre-substrate of LigAB (Scheme 3.1), this enzyme activation is hypothesized to be the result of feed-forward activation (FFA) – a rarely reported phenomena, never before observed for an extradiol dioxygenase. Feed-forward activation (FFA) is a generally understood principle in biochemistry (being the inverse of feedback inhibition); however, examples of enzymes that are feed-forward activated by small molecules are rare in the literature. Only a handful of well-known examples exist – most of which are found in human central metabolism: phosphofructokinase-1 is enhanced homoallosterically through binding fructose-6-phosphate (F6P, substrate) and also heteroallosterically through binding fructose-2,6-bisphosphate (a metabolite of F6P), and glycogen synthase activity is enhanced by glucose-6-phosphate (Figure 3.1). In addition to demonstrating activation kinetics for LigAB, the X-ray crystal structure of LigAB (PDB 1B4U) was used to computationally determine the likely physiological allosteric effector (vanillin) binding pocket, and mutations of the LigAB Phe103α and Ala18β were performed in vitro to support the proposed site. Both of these
mutants demonstrated altered activation profiles, providing support for the hypothetical allosteric binding pocket. These results provide unique insight into the inner workings of LigAB and drives speculation on what other unique features may be discovered by future examinations of PCADSF members.

\[
\begin{align*}
\text{Glucose} & \xrightarrow{\text{Glucokinase}} \text{Glucose-6-phosphate} \\
\text{Fructose-6-phosphate} & \xrightarrow{\text{PFK-1}} \text{Fructose-1,6-bisphosphate} \\
\text{Fructose-2,6-bisphosphate} & \xrightarrow{\text{PFK-2}} \text{Fructose-2,6-bisphosphate} \\
\text{Phosphoglucomutase} & \xrightarrow{\text{Glycogen Synthase}} \text{Glycogen} \\
\text{Glycogen Phosphorylase} & \xrightarrow{\text{Glycogen}} \text{Glycogen-1-phosphate} \\
\end{align*}
\]

**Figure 3.1.** Allosteric feedforward activation found in central metabolism. Shown are A) the activation of Phosphofructokinase-1 (PFK-1) by fructose-6-phosphate (homoallosteric) and by fructose-2,6-bisphosphate (heteroallosteric), and B) the feedforward activation of glycogen synthase by glucose-6-phosphate. Dashed green lines indicate activation (+) of enzyme activity while the dashed red line indicates a feedback inhibition loop.
Scheme 3.1. Abridged lignin derived aromatic compound metabolic pathway from *Sphingobium* sp. SYK-6 demonstrating the connection between vanillin (orange, boxed), protocatechuic acid (PCA, red), and LigAB (bold). Adapted from Masai et al. (Ref. 99). Enzyme abbreviations: DesA, syringate O-demethylase; DesZ, 3OMG 3,4-dioxygenase; DesB, gallate 3,4-dioxygenase; LigAB, protocatechuic acid 4,5-dioxygenase; LigC, 4-carboxy-2-hydroxymuconate-6-semialdehyde (CHMS) dehydrogenase; LigI, 2H-pyran-2-one-4,6-dicarboxylate (PDC) hydrolase; LigJ, 4-oxalomesaconate (OMA) hydratase; LigK, 4-carboxy-4-hydroxy-2-oxoadipate (CHA) aldolase/oxaloacetate decarboxylase; LigM, vanillin/3-O-methylgallate (3OMG) O-demethylase; LigV, vanillin dehydrogenase. Other abbreviations: CHMOD, 2-hydroxy-6-methoxy-6-oxohexa-2,4-dienoate.
3.2. Materials and Methods

Commercially available reagents and solvents were purchased from Aldrich or Alfa Aesar and used without purification unless otherwise noted. 3-O-methyl gallate was purchased from ChromaDex. Bradford assays were conducted in 96-well plate format and absorption values recorded on a Molecular Devices SpectraMax M5 plate reader (Sunnyvale, CA). MAX Efficiency DH5α and One Shot BL21 Star chemically competent E. coli cells were purchased from Life Technologies (Carlsbad, CA). QuikChange Lightening Site Directed Mutagenesis kit was purchased from Agilent (Santa Clara, CA). Centrifugation and ultracentrifugation were performed on a DuPont Instruments (Wilmington, DE) Sorvall RC-5B centrifuge and Beckman (Brea, CA) L7-80 Ultracentrifuge with type 60-Ti rotor, respectively. All cells were lysed using a SIMO-Aminco Industry Inc. (Rochester, NY) French press.

3.2.1. Site Directed Mutagenesis. Site directed mutagenesis was performed as described in the protocol for the QuikChange Lightning Mutagenesis Site-Directed Mutagenesis kit (Agilent Technologies, Santa Clara, CA) using pET-15b containing the LigA-LigB gene cluster subcloned into the NdeI and BamHI sites as a template. Primers for all mutagenesis reactions (Table 3.1) were designed using the QuikChange Primer Design Tool provided by Agilent Technologies, and were synthesized by Life Technologies (Carlsbad, CA). Colonies of E. coli XL10-Gold or DH5α ultracompetent cells resulting from transformation of PCR reaction products were harvested and grown in 15 mL of LB-Amp media. Mutated plasmids were extracted from the cells using the QIAprep Spin Miniprep kit (Qiagen, Limburg, Netherlands). Extracted plasmids were submitted to the DNA Analysis Facility on Science Hill at Yale University (New Haven, CT) for sequencing confirmation of mutation incorporation.
3.2.2. Protein Purification. All enzymes, LigAB-WT and mutants, were purified anaerobically as previously described via nickel affinity chromatography. The His₆-tag was subsequently removed anaerobically by thrombin cleavage of isolated enzyme samples of LigAB-WT, and the mutants F103V and F103L, also as previously described. A second sample of LigAB-WT purified by nickel affinity chromatography was stored without subsequent removal of the His₆-tag to use for comparison of the kinetic constants to untagged LigAB-WT. All remaining mutants were purified anaerobically, and stored without removal of the His₆-tag. All enzymes were flash frozen in liquid nitrogen, and stored at -80 °C.

3.2.3. Steady-State Kinetics. The rate of the enzymatic reaction was determined by measuring O₂ consumption using an O₂-sensitive Clark-type electrode with computer integration via an Oxygraph electrode control unit (Hansatech, King’s Lynn, Norfolk, England). Prior to each assay, the electrode was standardized with air saturated water and water depleted of O₂ by addition of sodium hydrosulfite as described by the manufacturer. Stock solutions and buffers were prepared fresh daily. Prior to each experiment, 100 µL of enzyme
(60 µM) was thawed under a flow of N₂ and buffer exchanged into degassed 50 mM Tris (pH 7.5) containing 1:9 t-butanol/H₂O using a 3 mL Sephadex G-25 desalting gel (GE Healthcare) column in a glove-bag flushed with high purity N₂ for 1 h. Exchanged enzyme (3-7 µM) was removed from the glove-bag in a rubber septum-sealed vial containing an inert atmosphere (N₂) and stored on ice for the duration of each experiment. Enzyme aliquots used to initiate assays were removed from the vial using a gas-tight syringe, previously purged with N₂, immediately prior to injection into the assay mixture.

Steady-state kinetic parameters for LigAB (with and without His₆-tag) and all mutants with respect to the organic substrate were determined by measuring the rate of O₂ consumption in the presence of varying concentrations of organic substrate (1 µM to 5000 µM). An aqueous stock solution (25 mM) of the desired organic substrate (PCA, GA, or 3OMG) was prepared immediately prior to use in a 10 mL volumetric flask. Stock solutions of 3OMG (25 mM) were prepared in 1:9 DMSO/H₂O. The initial rate assays were performed in air-saturated 50 mM Tris (pH 7.5) and initiated by the addition of 1-2 µL of enzyme (3-7 nM). Reaction velocities were calculated from the slope of the first 30 s of data after enzyme addition and corrected for background O₂ consumption using 30 s of data immediately prior to enzyme addition, as previously described. ¹⁰⁴

Steady-state kinetic parameters with respect to O₂ were measured in 50 mM Tris (pH 7.5), 1 mM PCA, and 40-450 µM O₂ and initiated by the addition of enzyme (3-7 nM). The buffer was equilibrated prior to each reaction with a fixed mixture of O₂ and N₂ gas using a Cole-Parmer gas proportioner (Vernon Hills, IL), and the reaction chamber was maintained under an atmosphere of the same O₂/N₂ mixture. The precise enzyme concentration for each experiment was determined by Bradford assay (Bio-Rad) after completion. Kinetic parameters for all steady-state data were determined by a least-squares fitting to either the Michaelis-
Menten equation (Eq. 3.1) or the Haldane equation for substrate inhibition (Eq. 3.2) to the data using KaleidaGraph (Synergy).

\[
\text{Eq. 3.1} \quad \nu = \frac{k_{\text{cat}}[S]}{[S] + K_m}
\]

\[
\text{Eq. 3.2} \quad \nu = \frac{k_{\text{cat}}[S]}{[S] + K_m + \frac{[S]^2}{K_i}}
\]

3.2.4. Activity Enhancement Assays. Initial detections of activity (rate) enhancement were performed as competition studies of the non-substrates (vanillin, 3,4-dihydroxybenzaldehyde, and methyl-3,4-dihydroxy benzoate, amongst others) in the presence of LigAB’s natural substrate, PCA. The concentrations of PCA (500 µM) and non-substrate (1000 µM) were fixed for all reactions. Kinetic values were determined by measuring the initial rate of LigAB-WT O₂ consumption with the non-substrate and PCA pre-mixed in 50 mM Tris buffer (pH 7.5, 25 °C). The reaction was initiated by the addition of enzyme (3-7 nM). A baseline, corresponding to 100 % of the native (un-stimulated) LigAB-WT catalysis rate (ν₀) of PCA, was obtained at 500 µM PCA in the absence of non-substrate. The percent increase in reaction rate over ν₀ was calculated for each non-substrate.

To determine the effect of vanillin on the steady-state kinetic parameters of LigAB-WT dioxygenation of PCA, the rate of O₂ consumption was measured at varying concentrations of PCA (10 µM to 5000 µM) in the absence of vanillin, and again in the presence of vanillin (5000 µM) immediately before or after. All reactions were initiated by addition of enzyme, and each set of kinetics for the dioxygenation of PCA (both with and without vanillin) were obtained using the same thawed and desalted enzyme preparation. An aqueous stock solution of PCA (25 mM) was prepared immediately prior to use in a 10 mL volumetric flask. The initial rate assays were performed in air-saturated 50 mM Tris (pH 7.5) and initiated by the addition of 1 µL of LigAB-WT (3-7 nM). Reaction initial velocities were calculated from the slope of
the first 30 s of data after enzyme addition and corrected for background O₂ consumption using 30 s of data immediately prior to enzyme addition, as previously described. These data (each set consisting of both LigAB-WT-PCA, and LigAB-WT-PCA-Vanillin) were collected in quadruplicate using a newly thawed and desalted preparation of anaerobic enzyme for each replicate. The precise protein concentration for each set of kinetics data was determined after completion of all kinetic experiments. The kinetic parameters were determined from a least-squares fitting of the Haldane equation (Eq. 3.2) to the data using KaleidaGraph (Synergy). Due to slight variability in activity for each enzyme preparation, each data set was separately normalized to the \( V_{\text{max}} \) value determined for LigAB-PCA for that given replicate.

The activity \((k_{\text{cat}})\) enhancement effects imparted by vanillin were further investigated for LigAB-WT and each of the LigAB Phe103α and Ala18β mutants. First, the steady-state kinetics of each mutant was determined for dioxygenation of PCA, and second, the influence of vanillin on catalysis was measured in methods analogous to those above. For the analysis of the impact of vanillin on catalysis, the concentration of PCA for each mutant enzyme was set at 10x \( K_m \) up to a concentration of 5 mM (5 mM was the maximal concentration used due to solubility limitations). The concentration of vanillin was varied from 10 µM to 5000 µM. The organic substrate and vanillin were pre-mixed in 50 mM Tris (pH 7.5, 25 °C), and the reaction was initiated by the addition of enzyme (3-7 nM). Additionally, a baseline corresponding to 100 % of the native (un-stimulated) enzyme catalysis rate \((v_o)\) was obtained at 10x \( K_m \) or 5 mM organic substrate in the absence of vanillin. The percent increase/decrease in reaction rate over \( v_o \) was calculated at each concentration of vanillin, and kinetic parameters were determined by a least-squares fitting of the data to the Michaelis-Menten equation modified to describe non-competitive inhibition and allosteric activation (Eq. 3.3). In this equation, \( k_{\text{cat1}} \) is the turnover number in the absence of activator and \( k_{\text{cat2}} \) is the turnover number in the presence of activator.
Eq. 3.3 was subsequently modified by a substitution to reflect the percent change in turnover rate from baseline assays containing no activator (%base = 100 %) to assays in the presence of activator (%act) (Eq. 3.4). In addition to the ultimate fractional percent enhancement of activity (%act), this allows determination of the equilibrium constant for binding of the activator (Kact) to LigAB.

Eq. 3.3  \( v = k_{cat1} - (k_{cat1} - k_{cat2}) \frac{[act]}{[act] + K_{act}} \)

Eq. 3.4  \( \% \text{ rate change} = \%_{base} - (\%_{base} - \%_{act}) \frac{[act]}{[act] + K_{act}} \)

3.2.5. Product Identification. The production of 4-carboxy-2-hydroxymuconate-6-semialdehyde (CHMS) by the dioxygenation of PCA was monitored by observing changes to the UV-Vis spectrum in methods analogous to those used previously. In brief, the absorbance between 200 and 600 nm was monitored in order to observe the disappearance of peaks at 254 and 297 nm (corresponding to PCA) and the appearance of a peak at 410 nm (corresponding to CHMS).

3.2.6. Substrate Induced Enzyme Inactivation. To evaluate the effect of vanillin on LigAB stability, the rate constants of enzyme inactivation (japp inact) during PCA turnover both with and without vanillin present were calculated using methods previously described for mechanism-based inactivation of dioxygenases. The rate constant of inactivation at a given substrate concentration (j_s) in was determined from a fitting of Eq. 3.5 in KaleidaGraph to the curvature of O_2 consumption progress curves (modified to reflect product formation as discussed in Chapter 2) performed in air-saturated buffer during catalytic turnover of varying PCA concentrations (10 – 5000 µM) in the absence or presence of vanillin (5000 µM) using anaerobically purified LigAB. In this equation, P_s is the concentration of product formed.
at time $t$, $P_\infty$ is the concentration of product formed at reaction completion, and $P_i$ is the concentration of the product at the start of the reaction.

$$\text{Eq. 3.5} \quad P_t = P_\infty(1 - e^{-j_st}) + P_i$$

The apparent rate enzyme of inactivation ($j_{\text{inact}}^{\text{app}}$) in the presence or absence of vanillin during PCA turnover was subsequently determined as the average of the $j_s$ values measured at or above 200 $\mu$M PCA.

3.2.7. Computational Docking. In order to identify a possible allosteric site locations on or within LigAB, the crystal structure of protocatechuate 4,5-dioxygenase (LigAB) with PCA bound in the active site (PDB 1B4U, 2.20 Å) was used as the template for all docking experiments. Due to the aerobic nature of the crystallization conditions, the structure of LigAB used for these docking experiments presents an active site with a catalytically inactive Fe$^{3+}$ cofactor, allowing for the bound and intact PCA substrate to be co-crystallized with the enzyme. While nearly identical to the crystal structure with PCA bound, the LigAB crystal structure with no bound substrate (PDB 1BOU, 2.20 Å) was not used as a template for docking as the assumption was made that the bound form (1B4U) more accurately represents the structure of an activated complex. The pKa values for all Asp, Glu, His, Arg, Lys, Ser, and Tyr were calculated using the PROPKA web interface. Hydrogen atoms were added and the protonation states of the active site Histidine residues were assigned using the addH feature of the UCSF Chimera package. Water molecules were removed from the structure and the coordinates of chains C and D were subsequently saved to a separate file. MGL Tools (v1.5.6) was used to merge non-polar hydrogens, add Gasteiger charges, and assign AutoDock4 atom types to prepare the protein for docking. Protocatechuate crystallographically bound within the active-site of LigAB was not removed in order to preclude this site from docking results. The active-site iron was manually assigned a 2+ charge in the AutoDock PDBQT file. The vanillin
ligand was built using Avogadro (v1.1.1): an open-source molecular builder and visualization tool. Atomic charges were assigned to vanillin using AutoDock Tools.

Initial docking experiments were conducted using AutoDock Vina v1.1.2. Docking was performed in a search grid of 40 (x) by 35 (y) by 40 (z) with a step spacing of 1 Å centered near the crystal structure coordinates of the active-site non-heme iron (x = 62.258, y = 43.574, z = 47.368). The search space was designed to exclude the interface at which the A-B and C-D chains meet (the dimer-dimer interface) as well as the end of the β subunit (chain D) furthest away from the α subunit (chain C). In addition to search grid size limitations, these spaces were excluded due to solvent inaccessibility at the interfaces (making these regions unlikely candidates for allosteric effector binding), and a general assumption that the allosteric site was likely located proximal to the active site. Each Vina docking calculation generated 20 docked conformations and was repeated ten-fold to generate a total of 200 docked conformations.

More intensive docking experiments were conducted using AutoDock v4.2 (AD4) and were performed on the Wesleyan University Swallowtail compute cluster. The docking search parameters generated 200 docked conformations using a genetic algorithm with a population size of 150 individuals, a maximum generation number of $2.7 \times 10^4$, and $2.5 \times 10^6$ energy evaluations with all other parameters left at their default values. Docking was performed in a search grid of 100 (x) by 90 (y) by 100 (z) with a step spacing of 0.375 Å centered as described above for the Vina calculations. The results from both calculations (Vina and AD4) were clustered separately by binding site conformation and binding energy.

In addition to this shotgun docking approach to determine possible binding sites for vanillin, docking of the substrates PCA, GA, and 3OMG to the active sites of LigAB-WT and LigAB with computationally mutated Phe103α (F103H, F103L, F103T, and F103L) was conducted, also using AD4. Crystallographically bound PCA was removed from the active site.
In all cases, the 103α residue was left to be flexible during the docking. These dockings were performed in a search grid of 40 (x) by 40 (y) by 40 (z) with a step spacing of 0.375 Å, and centered near the crystal structure coordinates of the active-site non-heme iron (x = 55.198, y = 44.339, z = 47.175). The docking search parameters generated 25 docked conformations for PCA, and GA and 200 docked conformations for 3OMG using a genetic algorithm as described earlier. Two hundred, as opposed to 25, conformations were generated for 3OMG due to an absence of what is believed to be the physiological conformation of binding (i.e. the C5 substituent facing the Phe103α residue) in the initial generation of 25 docked conformations.

Using a physiologically relevant docked conformation of each substrate onto each enzyme, vanillin was subsequently docked onto the identified allosteric binding site. The 103α residue was again left as flexible. These dockings were performed in a search grid containing the allosteric site and the 103α residue of 32 (x) by 40 (y) by 30 (z) with a step spacing of 0.375 Å, and centered at x = 57.612, y = 39.147, z = 45.690. The docking search parameters generated 50 docked conformations for vanillin using a genetic algorithm as described earlier.

3.2.8. Sequence Alignment. Homologs of the Sphingobium sp. strain SYK-6 LigB protein were retrieved through sequence profile-based and Hidden Markov Model (HMM) profile-based iterative searching performed using the PSI-BLAST\textsuperscript{191} and JACKHMMER web utilities (http://hmmer.janelia.org/search/jackhmmer), respectively. Queries were run against the non-redundant (nr) protein database of the National Center for Biotechnology Information (NCBI), beginning with the Sphingobium LigB sequence as the initial seed for the searches, with additional seeds selected from the results and queried in turn to retrieve sequence homologs from families outside of the PCAD LigB family. For the sequence-based homology searches, a cut-off e-value of 0.01 was used to assess significance. In each iteration, newly detected sequences included within the cut-off were evaluated via initiation of a new search with the
sequence in question as the query to guard against inclusion of false positives; searches were continued with the same e-value threshold only if the profile remained uncorrupted without false positives. Postulated relationships recovered using iterative searches were further confirmed with other aids such as concordance of predicted or known secondary structural elements.

Sequence-based homology clustering to identify individual PCAD families was performed on all sequences retrieved above using the BLASTCLUST program (http://ftp.ncbi.nih.gov/blast/documents/blastclust.html), using empirically-determined length and score threshold cut-off values. Multiple sequence alignments were constructed using the MUSCLE and Kalign alignment programs, followed by manual adjustment informed by sequence-based homology search results and experimentally-determined structures. In families lacking solved three-dimensional structures, secondary structure predictions were performed on multiple sequence alignments with the JPred program. Consensus for individual and combined family alignments was performed on each column in all alignments, testing for absolute residue conservation and conservation of shared biochemical properties.

3.2.9. Measurement of LigAB Fluorescence Spectra. The steady state intrinsic protein fluorescence spectra were measured using 1 μM LigAB-WT or A18W either alone or in combination with 100 μM PCA and/or 100 μM vanillin in 100 mM Tris buffer (pH 7.5, 25 °C) in quartz cuvettes of 200 μL final volume. Emission spectra were obtained between 310 – 450 nm using a Horiba SPEX Fluoromax-4 spectrofluorometer (Edison, NJ) with an excitation of 340 nm with slits of 2 and 4 nm band-pass for excitation and emission, respectively, with the polarizers off. Spectra for each condition were obtained from an average of six replicates; averaging for each data set was performed after spectral correction for background, buffer and
PCA and/or vanillin contributions. Spectral analyses were performed with Grams AI ver. 8 (Thermo Electron Corp.).

3.2.10. Calculation of the Solvent Accessible Surface Area (SASA). Calculation of the SASA for all LigAB aromatic residues in its various ligation forms was determined using the Sh Hale-Rupley algorithm (using the 1b4u pdb file and the vanillin docked structure).

Briefly, a 1.4 Å probe radius (approximate radius of water molecule) was employed to scan a spherical mesh composed of 500 equidistant points positioned beyond the van der Waals radius of each atom. Only those points that did not overlap with another atoms mesh, given the probe radius, were included in the SASA calculation. Only a single dimer (chains C and D) of the LigAB tetramer complex was included in the calculations.
3.3 Results and Discussion

Identification of molecules which can act as substrates or inhibitors of ring-cleaving dioxygenases (intradiol and extradiol) is well documented in the literature. From these studies it is clear that ring-cleaving dioxygenases, in general, have expansive substrate utilization profiles often being able to use 5 or more compounds as substrates in their respective dioxygenation reactions. Furthermore, two classes of inhibitors are described, those that function as chelators and oxidants that completely abolish dioxygenase activity by altering the nature of the iron cofactor, and organic substrate analogs which compete for binding to the active site. In our previous work characterizing LigAB, we identified numerous substrates and inhibitors (Table 3.2). Here we describe how those studies lead us toward identification of molecules that are activators of LigAB as well as investigation of the impact of mutagenesis on the activation profile for vanillin.

3.3.1. Allosteric Activation of LigAB. Based upon our work and the investigations of other dioxygenases, one would expect that methyl-3,4-dihydroxybenzoate (MDHB), 3,4-dihydroxybenzaldehyde (PCAld), and vanillin might act as inhibitors or substrates of LigAB due to their structural similarities to PCA (Figure 3.2); however, none of these molecules is turned over in the presence of LigAB, nor inhibits LigAB’s catalytic activity. Interestingly, however, at a concentration of 1 mM all stimulate an 11-28 % increase in the LigAB-WT rate ($k_{cat}$) of PCA (500 µM) dioxygenation (Table 3.2). The identification of small molecule...
allosteric activators for an EDO is interesting and raises mechanistic questions about the differences of LigAB from other previously described dioxygenases.

Table 3.2. Library of Alternative Substrates: Observed Activity, Inhibition, and Activation

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Activity</th>
<th>% Activity</th>
<th>% Inhibition</th>
<th>% Activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>Y</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,4-Dihydroxybenzamide</td>
<td>Y</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gallate</td>
<td>Y</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-O-Methylgallate</td>
<td>Y</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(HPCA) Homoprotocatechuate</td>
<td>Y</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catechol</td>
<td>Y</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,4-Dihydroxybenzonitrile</td>
<td>Y</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vanillin</td>
<td>N</td>
<td></td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>3,4-Dihydroxybenzaldehyde</td>
<td>N</td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Methyl-3,4-dihydroxybenzoate</td>
<td>N</td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>2,3,4-Trihydroxybenzoic Acid</td>
<td>N</td>
<td></td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>2,3-Dihydroxybenzoic Acid</td>
<td>N</td>
<td></td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>4-Nitrocatechol</td>
<td>N</td>
<td></td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>4-Methylcatechol</td>
<td>N</td>
<td></td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>3-Methoxycatechol</td>
<td>N</td>
<td></td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Caffeic Acid</td>
<td>N</td>
<td></td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Vanillic Acid</td>
<td>N</td>
<td></td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>3,4-Dimethoxybenzaldehyde</td>
<td>N</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2-Methoxy-4-methylphenol</td>
<td>N</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Veratrole</td>
<td>N</td>
<td></td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

*Values taken from Ref. 104.

*This work. Measured with 500 µM PCA and 1 mM activator in air saturated 50 mM Tris buffer at 25 °C, pH 7.5. Calculated as a percentage of the activity of LigAB with 500 mM PCA with no activator present.
Vanillin, the compound that causes the most pronounced rate enhancement (28 % at 1 mM), is part of the *Sphingobium* sp. SYK-6 LDAC degradation pathway and is the pre-pre-substrate of LigAB (Scheme 3.1). This suggests that a heteroallosteric feed-forward activation (FFA) mechanism may be at work in this pathway. Enzymes that are subject to allosteric activation described in the literature can have rate enhancements from 11 %\(^{206}\) ranging to systems that absolutely require the activator to have any activity.\(^ {207}\) The phenomena of allosteric activation by a metabolic precursor has not been previously reported for a non-heme iron intradiol or extradiol dioxygenase of any type (Type I, II or III), and rate enhancement by small molecule effectors of any kind has been reported in only two instances for dioxygenases in general. In mammalian cysteine dioxygenase, covalent modification of the enzyme by cysteine (the substrate) produces a non-allosteric activation which then effects an 11 – fold rate enhancement, but is not strictly required for activity.\(^ {208}\) In human indoleamine dioxygenase (IDO), activity (\(V_{\text{max}}\)) is enhanced through binding of synthetic indole derivatives to an auxiliary binding pocket near the active site, leading to a 35 to 60 % rate enhancement (depending upon the pH and effector molecule).\(^ {209, 213}\) Those studies suggested the existence of an auxiliary binding site for IDO allosteric effectors which was recently identified by computational docking studies using the IDO crystal structure.\(^ {214}\) While these two examples of activation of dioxygenases are both for mammalian enzymes, allosteric activation of non-dioxygenase bacterial enzymes has been previously reported in the literature, as exemplified by the *Escherichia coli* fructose-1,6-bisphosphatase which is subject to FFA by phosphoenolpyruvate.\(^ {215}\)

Despite limited examples of allosteric dioxygenase activation, transcriptional activation of metabolic pathways in response to the presence of small aromatic molecules is well documented. Examples of transcriptional activators of LDAC degradation that are sensitive to
small molecules include LigR from *Sphingobium* sp. SYK-6\textsuperscript{216} and PcaV from *Streptomyces coelicolor*.\textsuperscript{217} Both proteins regulate genes encoding enzymes directly responsible for the degradation of LDACs (the β-ketoadipate pathway), and are activated by the presence of PCA. Evolutionarily, tight regulation of mRNA transcription, and consequently protein expression, is important for cellular energy conservation, and is most often controlled by the presence or absence of small molecules from the associated enzyme pathway(s). Therefore, expression of the LigR and PcaV enzymes while only in the presence of LDAC compounds would reduce unnecessary energy expenditure by the bacterium (i.e., avoiding the synthesis of un-needed protein in the absence of substrates), and optimize carbon source utilization.

In contrast, we hypothesize that under high concentrations of vanillin in the cytoplasm, the activity ($k_{\text{cat}}$) of LigAB would be up-regulated, allowing for greater metabolic flux without necessitating the expense of additional protein production. While intracellular concentrations of vanillin (or other LDACs) are never reported in the literature, ample literature describes the amount of vanillin that can be isolated from pulping waste (reported yields of 2.6 g of vanillin from each liter of pulping sulfate liquor)\textsuperscript{218,219} which is the environmental origin for the bacteria *Sphingobium* sp. SYK-6 from which LigAB is encoded. This mass of isolated vanillin corresponds to a concentration of 17 mM in the pulping waste, suggesting that high micromolar and even millimolar cytoplasmic concentrations of vanillin might be physiologically possible. Additionally, transcriptionally activated LDAC degradation by LigR was demonstrated using 10 mM concentrations of either PCA or gallate.\textsuperscript{216} Furthermore, whole cell studies on *Comamonas testosteroni* strain BR6020 grown on various LDACs showed that when media was supplemented with vanillin at concentrations up to 12 mM no aromatic compounds could be detected in the medium when cultures reached mid- to late-log-phase, suggesting the active uptake of LDAC compounds. In these cultures, when *C. testosteroni* was grown in vanillin
containing media and subsequently exposed to PCA significant increases in oxygen consumption were observed suggesting that PCA and vanillin are acting synergistically to promote LDAC degradation.\textsuperscript{220} Curiously, despite the structural similarities between PCA and vanillin and these previous biological studies, the literature provided no papers describing the impact of vanillin on a LDAC dioxygenase. While other enzymes are needed to enable conversion of vanillin into a central metabolite, the \textit{in vivo} data above support our assertion that the sensitivity of LigAB to vanillin could enhance metabolic flux thereby maximizing carbon source utilization (if LigAB catalysis is rate limiting in the pathway), and we sought to be the first to describe the \textit{in vitro} relationship of vanillin to a dioxygenase.

Due to this physiological connection between LigAB and vanillin, our allosteric activation studies focused on vanillin to the exclusion of our other discovered activators. Varying vanillin concentration subsequently revealed a concentration dependence to the rate enhancement that can be fit to Eq.4\textsuperscript{190} (Figure 3.3), and results in a maximum rate ($k_{cat}$) enhancement of 36 \pm 3 \% and a $K_{act}$ for vanillin of 1.1 \pm 0.3 mM (Table 3.3). Additional experiments were then performed by varying the concentration of PCA (10 – 5000 µM) in the presence of a constant vanillin concentration (5000 µM) to allow for comparison to initial rates of LigAB-WT in the
absence of vanillin (Figure 3.4). While the observation was at first unexpected, as it had not been previously reported for ring-cleaving dioxygenases, these simple experiments demonstrated a consistently observed rate enhancement for LigAB in the presence of vanillin. Examination of Figure 3.4 clearly demonstrates that vanillin is a V-type allosteric activator, which alters $k_{\text{cat}}$, consistent with the values reported in Table 3.4, and by increasing the rates at concentrations of PCA greater than 100 µM it results in the simultaneous increase in the $K_m$ of LigAB for PCA. The rate enhancement is significant and reproducible, and extrapolates to an effective enhancement of $k_{\text{cat}}$ to nearly 230 s$^{-1}$ versus the unenhanced rate of 180 s$^{-1}$.104

| Table 3.3. Percentage Enhancement or Inhibition of Turnover of LigAB-WT and Mutants as a result of Co-incubating Enzyme with 5 mM PCA and Varying Concentrations of Vanillin |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | LigAB-WT        | F103V           | F103L           | F103T           | F103H           | A18W            |
| $\%_{\text{act}}$ | + 36 ± 3        | + 19 ± 4        | no change       | Inhib$^\dagger$ | − 23 ± 18       | + 30 ± 18       |
| $K_{\text{act}}$ (mM) | 1.1 ± 0.3      | 1.6 ± 0.8       | —               | —               | —               | 6 ± 6           |
| $K_{\text{inh}}$ (mM) | —               | —               | —               | —               | 7 ± 8           | —               |

$^\dagger$ Poor tightness of fit to Eq. 3.4 prevents an accurate estimation of the percent inhibition in these cases.
Figure 3.4. Normalized steady-state kinetics of LigAB-WT dioxygenation of PCA in the presence and absence of vanillin. No vanillin added (red filled, solid line), vanillin present (blue open, - -). The dependence of the initial rates on the PCA concentration in the presence of vanillin was determined at a vanillin concentration of 5 mM. All data were collected in air-saturated buffer (250-275 µM O₂, pH 7.5, 25 °C). The initial rates from each data set (denoted by shape) were subsequently normalized to the V_max determined for LigAB with PCA for that replicate, and subsequently fit to Eq. 3.2.

Table 3.4. Steady-State Kinetic Parameters for LigAB-WT in the Absence or Presence of the Allosteric Effector Vanillin

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Effector</th>
<th>k_cat (s⁻¹)</th>
<th>K_m (µM)</th>
<th>K_i (mM)</th>
<th>k_cat/K_m (x 10⁶ M⁻¹ s⁻¹)</th>
<th>f_inact (x 10⁻² s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td></td>
<td>178 ± 9</td>
<td>81 ± 12</td>
<td>18 ± 6</td>
<td>2.2 ± 0.3</td>
<td>1.0 ± 0.1</td>
</tr>
<tr>
<td>PCA</td>
<td>Vanillin</td>
<td>230 ± 5</td>
<td>130 ± 8</td>
<td>17 ± 2</td>
<td>1.8 ± 0.1</td>
<td>1.2 ± 0.2</td>
</tr>
</tbody>
</table>
With the previously observed gene regulatory activation of LigAB by PCA, and knowing that LigAB is naturally promiscuous, we could envision two scenarios for the newly observed LigAB rate enhancement in the presence of vanillin: 1) vanillin enhances the turnover rate of all acceptable substrates, or 2) rate enhancement by vanillin is specific to PCA. To investigate this, GA and 3OMG, LigAB’s other in-pathway natural but “lesser” substrates, were analyzed for rate enhancement in the presence of vanillin via the same assay which revealed the first observations of rate enhancement by vanillin. Under these conditions, no enhancement or inhibition (considered to be greater than 5 % change in rate) was observed, suggesting that the allosteric rate enhancement induced by vanillin binding is substrate specific (Figure 3.5). This observation is consistent with the activation being a regulatory feed-forward mechanism, since from the perspective of evolutionary selection for increased pathway flux, only the enhancement of PCA catabolism would lead to alteration of the concentration of vanillin in the environment.

3.3.2. Assessment of Vanillin’s Impact on LigAB Stability. Wondering if the observed activation by vanillin resulted from increased structural stability of the LigAB-WT protein complex through either (1) enhancing α/β subunit complex formation (preventing dissociation of the active enzyme complex) or (2) by restricting the generation of reactive oxygen species

![Figure 3.5](image-url)
in the absence of bound organic substrate, the concentration dependence of enzyme inactivation \((j_s)\) during PCA turnover was calculated in the presence and absence of vanillin. Both with and without vanillin present, LigAB-WT shows increasing rates of inactivation at low \((\leq 100 \ \mu M)\) PCA concentrations and a plateau in the rate of inactivation at higher \((\geq 200 \ \mu M)\) concentrations of PCA (Figure 3.6) – a trend that does not lend to the traditional determination of the apparent rate of inactivation \(j_{\text{inact}}^\text{app}\) by a fitting of Eq. 3.6 to the \(j_s\) vs. substrate concentration data.

\[
\text{Eq. 3.6} \quad j_s = \frac{j_{\text{inact}}^\text{app}[S]}{K_m^\text{app} + [S]}
\]

The \(j_{\text{inact}}^\text{app}\) values were therefore determined as the average of \(j_s\) values observed at PCA concentrations \(200 – 5000 \ \mu M\) (Table 3.4); yielding a LigAB-WT inactivation rate of \(1.0 \pm 0.1 \times 10^{-2} \text{ s}^{-1}\) in absence of vanillin and \(1.2 \pm 0.2 \times 10^{-2} \text{ s}^{-1}\) in the presence of vanillin (Table 3.4). The nearly identical \(j_{\text{inact}}^\text{app}\) values, where inactivation rate is slightly greater in the presence of vanillin, support our proposal that vanillin is acting as an allosteric effector of rate rather than increasing enzyme stability.
3.3.3. Identification of a Putative Allosteric Binding Site in LigAB. In order to determine a potential vanillin binding site, computational docking of vanillin onto the C (alpha) and D (beta) chains of the LigAB crystal structure (PDB 1B4U) was performed using a shotgun approach similar to that used to determine the indoleamine derivative allosteric binding site for IDO as mentioned earlier. The docked conformations, from a search centered near the active site, identified several potential binding sites; however, one binding pocket located near the entrance to the active site was occupied by 67% of all conformations generated by AD4 calculations (Figure 3.7) and by 86% of all conformations generated using Vina. Closer examination of the proposed allosteric binding pocket shows that residues Phe103α, His12β, and His127β are shared in common with the active site, and therefore may be involved in activating LigAB (Figure 3.8). Additionally, several docked conformations of vanillin show backbone contacts with secondary sphere active site residues. Of these residues, Glu86β is directly involved in a hydrogen bond network through the carboxylate with Nδ of His127β, believed to be the catalytic base that performs the

Figure 3.7. Shotgun style docking of vanillin onto the crystal structure of LigAB (PDB 1B4U): alpha subunit (Light grey) and beta subunit (dark grey). Each docking location is displayed as a different color. The identified potential allosteric binding pocket is represented by the bright green ensemble near the active site (active site Fe shown as an orange sphere) at the center of the image.
initial substrate deprotonation. The His127β methylene also contributes a small fraction to the identified allosteric binding pocket. Examination of the multiple sequence alignment of the LigB family (Appendix Figure A.2) revealed strong consensus conservation (typically exceeding 85%) of the biochemical properties of residues in and around the allosteric binding pocket of the LigB family. Inspection of this same region in other families of the PCADSF revealed striking family-specific sequence and structure conservation apparent even in the MhpB and HpaD families, two of the most closely-related families to LigB (Appendix Figure A.2). Through this comparison, two discrete stretches of residues displaying family-specific conservation were identified, both of which form small, insert-like extensions embedded within the core scaffold of the PCADSF domain: (1) residue positions 15-30 and (2) residue positions...
80-96 when mapped to the *Sphingomonas* 1B4U structure. Taken together, these observations suggest that the indented pocket identified in this study is conserved as a general feature across the PCADSF which, through family-specific sequence/structure diversification, could specifically mediate allosteric activation in other PCADSF families.

Examination of this pocket also revealed close contact between two of LigAB’s tryptophan residues (Trp30β and Trp88β) with docked vanillin. Examination of the binding location for all other docked vanillin residues (regions within 5 Å of all non-green vanillin residues displayed in Figure 3.8) revealed that none of the other binding sites had any nearby tryptophan residues. We therefore hypothesized that if vanillin binds to the above mentioned binding site that we might see a blue shift in the equilibrium intrinsic Trp fluorescence of the protein associated with the displacement of waters associated with Trp30β and Trp88β. Fluorescence spectra were then obtained for (1) apo LigAB (with Fe but no organic substrates), (2) LigAB∙PCA, (3) LigAB∙vanillin, and (4) LigAB∙PCA∙vanillin (Figure 3.9). Both PCA and vanillin give rise to a blue shift in the fluorescence when compared to the unliganded enzyme (the λ_max for LigAB∙Fe is 334 nm, while LigAB∙PCA and LigAB∙vanillin have λ_max of 329 nm). These observed blue shifts are likely due to each molecule interacting with Trp88β and the allosteric site regardless of the other’s presence. When both ligands are present an even larger blue shift is observed (λ_max for LigAB∙PCA∙vanillin is 327 nm). To further

![Figure 3.9. Changes in the intrinsic fluorescence spectra of LigAB induced by PCA and vanillin, individually and combined. Emission spectra of LigAB with PCA (blue), vanillin (green) and both (black) are blue shifted when compared with apo LigAB (red).](image-url)
support that we are observing a change in the solvation of Trp30β and/or Trp88β, we assessed the solvent accessibility of each of the Trp residues of LigAB based upon the crystal structures (1b4u.pdb for the LigAB-PCA and our docking results for the LigAB-PCA-vanillin complex). We observed that Trp88β was calculated to have the most significant decrease in solvent accessibility upon binding of vanillin to the LigAB-PCA complex (corresponding to a decrease in the SASA of 4.0 Å²). The magnitude of the blue shift observed is consistent with a slight decrease in SASA, providing additional support for the location of the predicted allosteric binding site.

3.3.4. Impact of Mutants on Allosteric Activation of PCA Dioxygenation. Based upon the studies described above we proceeded to look more closely at the residues at the interface of the allosteric pocket and the active site. As mentioned above, residues Phe103α, His12β, and His127β are shared in common with the allosteric pocket and the active site. Since we hypothesized that mutation of His12β and His127β would likely severely impact catalysis through their proposed roles in metal coordination and acid/base-catalysis, respectively, we decided to mutate Phe103α to examine its role in the FFA mechanism. Phe103α was mutated to a series of residues with varying size and polarity: alanine (F103A), serine (F103S), threonine (F103T), valine (F103V), leucine (F103L), and histidine (F103H). While all mutations were successfully generated and expressed, co-purification of the α and β subunits via nickel affinity chromatography, as is performed for LigAB-WT, was found to be effective for the purification all mutants except the alanine and serine mutants. In both of these cases, the α subunit containing both the mutation and the 6-His tag remained bound to the nickel affinity column matrix as expected, but the untagged β subunit flowed freely from the column with the rest of the supernatant. This suggests that the α/β dimer interface of LigAB is
sensitive to structural changes in this residue position – requiring a larger and more non-polar residue.

In addition to the mutation of Phe103α, we additionally mutated Ala18β, which is at the opening of the putative allosteric pocket. We hypothesized that allosteric binding of vanillin would be diminished or slowed in this mutant if we had correctly identified the location of the allosteric pocket. Additionally, since Ala18β is located over 10 Å from the closest portion of PCA mutation of Ala18β should not directly impact catalysis. Steady-state kinetic parameters for PCA dioxygenation were determined for each of the isolatable mutants described above (Table 3.5; Figure 3.10). Unsurprisingly, all mutants except A18W (within error) showed modestly diminished catalytic efficiency ($k_{cat}/K_m$) as compared to wild-type enzyme, consistent

**Table 3.5.** Steady-State Kinetic Parameters for LigAB Phe103α mutants

<table>
<thead>
<tr>
<th></th>
<th>F103V</th>
<th>F103L</th>
<th>F103T</th>
<th>F103H</th>
<th>A18W</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>89 ± 2</td>
<td>33 ± 2</td>
<td>112 ± 2</td>
<td>110 ± 3</td>
<td>122 ± 3</td>
</tr>
<tr>
<td>$K_m$ (µM)</td>
<td>260 ± 20</td>
<td>1600 ± 200</td>
<td>210 ± 20</td>
<td>108 ± 9</td>
<td>66 ± 5</td>
</tr>
<tr>
<td>$K_i$ (mM)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>10 ± 5</td>
<td>40 ± 20</td>
</tr>
<tr>
<td>$k_{cat}/K_m$ (x 10$^5$ M$^{-1}$s$^{-1}$)</td>
<td>3.4 ± 0.3</td>
<td>0.21 ± 0.01</td>
<td>5.3 ± 0.5</td>
<td>11 ± 1</td>
<td>19 ± 3</td>
</tr>
<tr>
<td>Fold Change in $k_{cat}/K_m$ from WT</td>
<td>-6.5</td>
<td>-105</td>
<td>-4.2</td>
<td>-2.0</td>
<td>-1.2</td>
</tr>
</tbody>
</table>
with other studies on the mutagenesis of non-catalytic residues in dioxygenases; however, all of the mutant enzymes were able to maintain $k_{\text{cat}}/K_m$ values in excess of $10^4$ M$^{-1}$s$^{-1}$. The oxygen $K_m$ for the mutants were largely unchanged (data not shown), however for F103V it was observed to be smaller than LigAB-WT ($94 \pm 8 \mu$M vs. $162 \pm 16 \mu$M).

Vanillin activation with PCA was also investigated for each of the mutants. Of the mutants examined, only the F103V and A18W mutants exhibited any vanillin induced catalytic rate enhancement of PCA turnover, as observed with LigAB-WT (Table 3.3; Figure 3.3). The allosteric activation for F103V is 19%, possibly indicating that the valine mutant maintains some of the ability to transmit the activation signal from the allosteric pocket to the active site. Interestingly, the A18W mutant, has a maximal activation of 30 %, which is nearly the same as that of LigAB-WT. From examination of the activation data (Figure 3.3 and Table 3.3) you can see that, while the fit predicts almost wild-type activation of LigAB, the $K_{\text{act}}$ is at a noticeably higher concentration, which is consistent with our hypothesis that this mutant would impede the binding of vanillin by partially closing off the opening to the allosteric pocket. The remaining three mutants showed either no change in rate compared with LigAB-WT (F103L) or show diminished turnover of PCA in the presence of vanillin (F103H and F103T). Like the observed rate enhancements for LigAB-WT, F103V and A18W, the inhibition observed with F103H could be fit using Eq. 3.4, suggesting a non-competitive inhibition mechanism with a maximum fractional inhibition by vanillin of $23 \pm 18 \%$ (Table 3.3). The inhibition observed with F103T, however, is instead linear with respect to vanillin concentration under the observed conditions suggesting that either saturating concentrations of vanillin were not reached such that the concave shape of the inhibition curve could be observed or that non-competitive inhibition may not adequately describe the mechanism of the rate decrease. We posit that the diminishment of activation (or appearance of non-competitive inhibition) observed for these
mutants provides support our assignment of the location of the allosteric binding pocket. Further mutagenesis of the proposed pocket could provide additional validation of the location of vanillin binding, while mechanistic studies of LigAB with and without vanillin could help reveal the mechanism by which this activation occurs.

Furthermore, upon seeing that the activation of the A18W mutant was diminished, but that catalysis was virtually unchanged, we proceeded to assess the intrinsic fluorescence to allow comparison with LigAB-WT. In most of the conditions tested the fluorescence of the iron bound enzymes, and the enzyme complexes with either PCA or Vanillin bound were effectively identical for A18W as compared to LigAB-WT (Appendix Figure A.3); however, we observed that the blue shift associated with forming the activated complex (LigAB-PCA-vanillin) was less for the mutant than for wild-type ($\lambda_{\text{max}}$ values of 327 and 330 nm for LigAB-PCA-vanillin and A18W-PCA-vanillin, respectively; Figure 3.11). This suggested to us that perhaps the presence of a large residue at the opening of our proposed allosteric site was preventing formation of the fully liganded LigAB-PCA-vanillin complex. While the difference in blue shift is small, we feel that this provides additional evidence for the case that we have identified the allosteric binding pocket.

### 3.3.5. Potential Mechanistic Implications

Allosteric regulation by small molecule effectors has the potential to modify enzyme function by altering either the enzyme’s affinity for the substrate (K-type, altered $K_m$), or the catalysis of the substrate (V-type, altered $k_{\text{cat}}$). The V-type allosterism observed with LigAB suggests that a rate limiting chemical step(s) of the reaction mechanism (Scheme 3.2), subsequent to substrate binding (PCA or $O_2$), or product release is altered upon allosteric binding of vanillin. While the catalytic mechanism of LigAB’s extradiol ring-opening of PCA is not known, the mechanism of ring-cleaving dioxygenation has been extensively investigated for both Type I and III EDOs. The mechanism described in
Scheme 3.2 is largely based on the mechanistic investigations, including single turn-over kinetics and crystallographically trapped intermediates, of the catalytic cycle of the Type I EDO homoprotocatechuate 2,3-dioxygenase (2,3-HPCD).\textsuperscript{106, 146, 222, 223} Single turn-over kinetic studies of 2,3-HPCD found that $O_2$ reacts in a minimum of three steps after binding, and optical spectra identified the final steps as ring opening and a rate limiting product release.\textsuperscript{146, 222}

Several of the oxygen reaction steps have been identified through crystallographically trapped intermediates, including 1) a substrate semiquinone radical-Fe(II) superoxo species (identified by a substrate ring pucker and side on binding of oxygen to the Fe(II) (Scheme 3.2, step B),\textsuperscript{106} 2) formation of an Fe(II)-alkylperoxo (Scheme 3.2, step C),\textsuperscript{106} and 3) the gem-diol adduct formed from homolytic cleavage of the peroxo species (Scheme 3.2, step D' red).\textsuperscript{223} It is unknown exactly how LigAB’s mechanism may vary from that of 2,3-HPCD.

Due to the location of the allosteric binding pocket proximal to the active site, binding of vanillin likely effects the structural conformation or dynamics of catalytically active residues resulting in the activation effect. As mentioned previously, the backbone of His127$\beta$, the

![Figure 3.11](image1.png)

**Figure 3.11.** Comparison of the intrinsic tryptophan fluorescence spectra of LigAB-WT to that of the LigAB-A18W mutant enzyme, apo and bound to both PCA and vanillin. (A) Emission spectra of LigAB apo (red line) and A18W apo (blue line) are compared with the respective complexes with both PCA and vanillin bound (red and blue dashed lines, respectively). (B) Zoom in of the fluorescence spectra to ease observation of the difference in blue shifts between LigAB-WT and A18W.
catalytic base responsible for the initial deprotonation of the substrate, contributes a small fraction to the surface area of the allosteric pocket. A second residue contributing more substantially to the allosteric pocket is Glu86β. This residue is potentially significant to due to its involvement in a hydrogen bonding cascade with the δN of the His127β sidechain. Vanillin binding may induce a slight conformational change placing His127β in a more favorable alignment for deprotonation to occur (although likely already a fast step) or aid in the acid assisted product release step, allowing hydrolysis to occur more quickly. In order to more precisely determine the both mechanistic steps effected by the allosteric binding of vanillin and how the allosteric signal is transmitted to the active site, additional mutants of the both the active site and allosteric site will need to be generated.
Scheme 3.2. Proposed catalytic cycle of LigAB, based on previously characterized extradiol dioxygenase mechanisms and structural information from LigAB.
3.4. Conclusions

We have reported here the first example of allosteric activation by a metabolic precursor identified in an intra- or extradiol dioxygenase of any type. While feed-forward activation is a well-known phenomenon in human metabolism, it is rarely reported in bacterial enzyme systems. Although not yet observed in any other members of the PCADSF, the probability exists that closely related PCAD superfamily members such as DesB from *Sphingobium* sp. SYK-6, PCA 4,5-dioxygenase from *Comamonas testosteroni* Pt-L5, gallate 3,4-dioxygenase from *Pseudomonas putida*, and PmdB from *Comamonas* sp. E6 may display similar feed-forward activation behavior due to their evolutionary relationship to LigAB. It is unknown, however, whether the substrates or pre-substrates of these enzymes attain the same environmental concentrations observed for vanillin. The Phe103α and Ala18β mutants provide support for the location we propose for the allosteric binding pocket. While this pocket is located adjacent to the active site, the mechanism by which this activation occurs is still largely unknown.

Additionally, we have demonstrated that Phe103α is a critical residue for LigAB; its location at the interface of LigA with LigB provides important hydrophobic interactions enabling the two proteins to be co-purified, and by being located between the allosteric pocket and the active site it influences enzyme activation. Detailed mechanistic studies of the activation of LigAB-WT and mutant enzymes are likely to reveal important insights into the mechanistic modes responsible for controlling the allosterism of this enzyme. The identification of allosteric activation of an extradiol dioxygenase involved in the catabolism of lignin might suggest that the enzymes of this pathway are responsive to changes in environmental concentration of LDAC molecules, and further investigations of this hypothesis are warranted.
Chapter 4:

Improving Alternate Substrate Utilization of LigAB from *Sphingobium* sp. strain SYK-6 through Site Directed Mutagenesis

Submitted as a manuscript for publication:
4.1. Introduction

We initiated research into the function and specificity of enzymes involved in the catabolism of lignin with the hope of improving the conversion of lignin into industrial products, including biofuels. When lignin is depolymerized a multitude of aromatic compounds are produced, and pathways for utilizing these lignin derived aromatic compounds (LDACs) have been identified in multiple organisms. These pathways often contain several branches which are capable of siphoning off a relatively small subset of LDACS produced during depolymerization and ultimately metabolize these compounds into molecules useful for central metabolism.

However, as can be seen in the LDAC degradation pathway of Sphingobium sp. SYK-6 (please refer to Scheme 3.1), each branch of the pathway contains a unique set of enzymes which metabolize one LDAC. Yet the enzymes in one branch are often evolutionarily related to those in other branches – resulting in multiple enzymes performing the same function on only slightly different substrates. For example, there are three dioxygenase enzymes (LigAB, DesB, and DesZ) in the abbreviated LDAC degradation pathway of Sphingobium sp. SYK-6 shown in Scheme 3.1, which metabolize aromatic substrates. The only difference in the substrate structure for these enzymes is a C5 substituent that varies from –H to –OH to –OCH$_3$ (PCA, gallate, and 3OMG, respectively). In addition to dioxygenase redundancy, there are also multiple demethylases (DesA and LigM), and amidohydrolases (LigI and LigJ).

The question then becomes, can one enzyme of each type be engineered to perform the function of all its relatives? One can imagine that if a microorganism were to be bioengineered to utilize a broad spectrum of substrates in the generation of a product that encoding one gene for each enzyme function would reduce the complexity of the resulting pathway in comparison to the one-substrate-one-enzyme alternative. By generating enzymes with broader substrate
scope, multiple branches in a pathway could effectively be condensed into one broad spectrum branch. However, in order to design enzymes to accomplish a task such as this, the relationships between an enzyme’s structure, substrate scope, and ultimate function need to be well understood.

Upon examining the crystal structure of LigAB, we hypothesized that mutagenesis of the active site residue Phe103α would enable enhanced turnover of the LDACs gallate (GA) and 3-O-methyl gallate (3OMG) due to its apparent steric interaction with bound substrate (being 3.8 Å away from the C5 carbon). Herein we report the generation of a series of Phe103α mutants, including Valine, Leucine, Threonine and Histidine, and the corresponding substrate utilization and activation kinetics for these mutants. Multiple mutants were identified that had enhanced catalysis for C5 substituted derivatives of PCA, with some also displaying catalytic activation in the presence of vanillin. These studies provide interesting insight into the mechanisms by which LigAB controls substrate specificity and how it can be modified to enhance the catalysis of alternate substrates.

4.2. Materials and Methods

4.2.1. General Methods. Commercially available reagents and solvents were purchased from Aldrich or Alfa Aesar and used without purification unless otherwise noted. 3-O-methyl gallate was purchased from ChromaDex. Bradford assays were conducted in 96-well plate format and absorption values were recorded on a Molecular Devices SpectraMax M5 plate reader (Sunnyvale, CA). MAX Efficiency DH5α and One Shot BL21 Star chemically competent E. coli cells were purchased from Life Technologies (Carlsbad, CA). QuikChange Lightening Site Directed Mutagenesis kit was purchased from Agilent (Santa Clara, CA). Centrifugation and ultracentrifugation were performed on a DuPont Instruments (Wilmington, DE) Sorvall RC-5B centrifuge and Beckman (Brea, CA) L7-80 Ultracentrifuge with type 60-
Ti rotor, respectively. All cells were lysed using a SIMO-Aminco Industry Inc. (Rochester, NY) French press.

4.2.2. Site Directed Mutagenesis. Site directed mutagenesis was performed as described previously. In brief, the QuikChange Lightning Mutagenesis Site-Directed Mutagenesis kit (Agilent Technologies, Santa Clara, CA) was used to alter our pET-15b-ligAB construct, with the primers being generated using the QuikChange Primer Design Tool. Mutated plasmids were extracted from the cells using the QIAprep Spin Miniprep kit (Qiagen, Limburg, Netherlands). All plasmids were submitted to the DNA Analysis Facility on Science Hill at Yale University (New Haven, CT) for sequencing confirmation of mutation incorporation.

4.2.3. Protein Purification. All enzymes, LigAB-WT and mutants, were purified anaerobically as previously described via nickel affinity chromatography. The His$_6$-tag was subsequently removed anaerobically by thrombin cleavage of isolated enzyme samples of LigAB-WT, for the mutants F103V and F103L, also as previously described. A second sample of LigAB-WT purified by nickel affinity chromatography was stored without subsequent removal of the His$_6$-tag to use for comparison of the kinetic constants to untagged LigAB-WT. All remaining mutants were purified anaerobically, and stored without removal of the His$_6$-tag. All enzymes were flash frozen in liquid nitrogen, and stored at -80 °C.

4.2.4. Steady-State Kinetics. The rate of the enzymatic reaction was determined by measuring O$_2$ consumption, as previously described, using an O$_2$-sensitive Clark-type electrode with computer integration via an Oxygraph electrode control unit (Hansatech, King’s Lynn, Norfolk, England). Prior to each assay, the electrode was standardized with air saturated water and water depleted of O$_2$ by addition of sodium hydrosulfite as described by the manufacturer. Stock solutions and buffers were prepared fresh daily. Prior to each experiment, 100 µL of enzyme (60 µM) was thawed under a flow of N$_2$ and buffer exchanged into degassed
50 mM Tris (pH 7.5) containing 1:9 t-butanol/H$_2$O using a 3 mL Sephadex G-25 desalting gel (GE Healthcare) column in a glove-bag flushed with high purity N$_2$ for 1 h. Exchanged enzyme (3-7 µM) was stored in a sealed vial containing an inert atmosphere (N$_2$) and on ice for the duration of each experiment. Enzyme aliquots used to initiate assays were removed from the vial using a gas-tight syringe, previously purged with N$_2$, immediately prior to injection into the assay mixture.

Steady-state kinetic parameters for LigAB and all mutants were determined by measuring the rate of O$_2$ consumption in the presence of varying concentrations of organic substrate (1 µM to 5000 µM) as previously described.$^{104}$ Aqueous stock solutions (25 mM) of the desired organic substrate (PCA, GA, or 3OMG) were prepared immediately prior to use in a 10 mL volumetric flask. 3OMG (25 mM) was prepared in 1:9 DMSO/H$_2$O. Assays were performed in air-saturated 50 mM Tris (pH 7.5) and initiated by the addition of 1-2 µL of enzyme (3-7 nM). Reaction velocities were calculated from the slope of the first 30 s of data after enzyme addition and corrected for background O$_2$ consumption using 30 s of data immediately prior to enzyme addition.

Steady-state kinetic parameters with respect to O$_2$ were measured in 50 mM Tris (pH 7.5), 1 mM PCA, and 40-450 µM O$_2$ and initiated by the addition of enzyme (3-7 nM). The buffer was equilibrated prior to each reaction with a fixed mixture of O$_2$ and N$_2$ gas using a Cole-Parmer gas proportioner (Vernon Hills, IL), and the reaction chamber was maintained under an atmosphere of the same O$_2$/N$_2$ mixture. The precise enzyme concentration for each experiment was determined by Bradford assay (Bio-Rad) after completion. Kinetic parameters for all steady-state data were determined by a least-squares fitting to either the Michaelis-Menten equation (Eq. 4.1) or the Haldane equation for substrate inhibition (Eq. 4.2)$^{189}$ to the data using KaleidaGraph (Synergy).
\[ \text{Eq. 4.1} \quad \nu = \frac{k_{\text{cat}} [S]}{[S]+K_m} \]

\[ \text{Eq. 4.2} \quad \nu = \frac{k_{\text{cat}} [S]}{[S]+K_m+\frac{100}{k_i}} \]

4.2.5. Activity Enhancement Assays. Initial detections of rate enhancement were performed as competition studies of non-substrates (vanillin, PCA aldehyde, and methyl-3,4-dihydroxy benzoate, amongst others) in the presence of LigAB’s natural substrate, PCA. In brief, the kinetic values were determined by measuring the initial rate of LigAB-WT O$_2$ consumption with non-substrate and PCA pre-mixed in 50 mM Tris buffer (pH 7.5, 25 °C). Each reaction was initiated by the addition of enzyme (3-7 nM) and a baseline corresponding to 100 % of the native (un-stimulated) LigAB-WT catalysis rate (\(\nu_0\)) of PCA was obtained at 500 µM PCA in the absence of non-substrate. The percent increase in reaction rate over \(\nu_0\) was calculated for each non-substrate.

The activity enhancement effects imparted by vanillin were further investigated here for each of the LigAB Phe103\(\alpha\) mutants. The steady-state kinetics of LigAB and mutants in the presence of both a known organic substrate (PCA, gallate, or 3OMG) and the non-substrate vanillin was determined by measuring O$_2$ consumption in the presence of varying concentrations of vanillin. The concentration of organic substrate (PCA, Gallate, or 3OMG) was set as 10x \(K_m\) of the enzyme for the respective substrate as determined from the least-squares fitting of steady-state kinetics data (measured in the absence of vanillin) to the Michaelis-Menten equation. For enzyme-substrate combinations where the value of 10x \(K_m\) exceeded 5 mM, 5 mM organic substrate was used. The concentration of vanillin was varied from 10 µM to 5000 µM. The organic substrate and vanillin were pre-mixed in 50 mM Tris (pH 7.5, 25 °C), and the reaction was initiated by the addition of enzyme (3-7 nM). Additionally, a baseline corresponding to 100 % of the native (un-stimulated) enzyme catalysis rate (\(\nu_0\)) was obtained at 10x \(K_m\) or 5 mM.
organic substrate in the absence of vanillin. The percent increase/decrease in reaction rate over 
v₀ was calculated at each concentration of vanillin, and kinetic parameters were determined by 
a least-squares fitting of the data to the Michaelis-Menten equation modified to describe non-competitive inhibition and allosteric activation, as previously described using Eq. 4.3 and 
Eq. 4.4.¹⁹⁰, ²²⁶

\[
\nu = k_{cat1} - (k_{cat1} - k_{cat2}) \frac{[act]}{[act] + K_{act}}
\]

\[
\% \text{ rate change} = \%_{base} - (\%_{base} - \%_{act}) \frac{[act]}{[act] + K_{act}}
\]

4.2.6. Computational Docking. Computational dockings were performed using AutoDock 
v4.2 (AD4) with methods analogous to those previously reported.²²⁶ For all dockings the 
structure of protocatechuate 4,5-dioxygenase (PDB 1B4U, 2.20 Å) was used as the template 
for docking experiments with pKa values for all charged residues being calculated using the 
PROPKA web interface.²²⁷ Docking experiments were conducted using AD4 and were 
performed on the Wesleyan University Swallowtail compute cluster. Each of the LigAB 
substrates described herein (PCA, GA, and 3OMG) was docked to the active sites of LigAB-
WT and LigAB with computationally mutated Phe103a (F103H, F103L, F103T, and F103L). 
In all cases, the 103a residue was left to be flexible during the docking. These dockings were 
performed in a search grid containing the active site of 40 (x) by 40 (y) by 40 (z) with a step 
spacing of 0.375 Å, and centered near the crystal structure coordinates of the active-site non-
heme iron (x = 55.198, y = 44.339, z = 47.175). The docking search parameters generated 25 
docked conformations for PCA, and GA and 200 docked conformations for 3OMG using a 
genetic algorithm with a population size of 150 individuals, a maximum generation number of 
2.7 x 10⁴, and 2.5 x 10⁶ energy evaluations with all other parameters left at their default values. 
Two hundred, as opposed to 25, conformations were generated for 3OMG due to an absence
of what is believed to be the physiological conformation of binding in the initial generation of 25 docked conformations.

Using a physiologically relevant docked conformation of each substrate onto each enzyme, vanillin was subsequently docked onto the identified allosteric binding site. The 103α residue was again left as flexible. These dockings were performed in a search grid containing the allosteric site and the 103α residue of 32 (x) by 40 (y) by 30 (z) with a step spacing of 0.375 Å, and centered at x = 57.612, y = 39.147, z = 45.690. The docking search parameters generated 50 docked conformations for vanillin using a genetic algorithm as described earlier.

4.3 Results and Discussion

4.3.1. Altered Substrate Utilization of LigAB Mutants. Our previous investigation of LigAB identified the ability to catalyze 4,5-dioxygenation reactions on seven different substrates with rates ranging over four orders of magnitude. This makes LigAB the most omnivorous of the protocatechuate 4,5-dioxygenases that have been investigated to date. Analysis of the LigAB crystal structure (PDB 1B4U) shows that Phe103α, in addition to being part of the previously identified allosteric pocket, is positioned within 4 Å of the C5-position of PCA. Steric interaction with the C5 functional groups of substrates was believed to impart the trend in catalytic efficiency \( \frac{k_{cat}}{K_m} \) observed for the natural substrates PCA, GA, and 3OMG – decreasing \( \frac{k_{cat}}{K_m} \) as the C5 functional group increases in size. To investigate this, each of the mutants described above were tested for their ability to utilize PCA, GA and 3OMG. Additionally, computational docking of PCA, GA and 3OMG was performed with LigAB-WT and with each of the mutants to allow examination of potential active site interactions (Figure 4.1).
Figure 4.1. Computational docking of substrates to LigAB-WT and F103Xα mutant active sites. A) Native substrate (PCA, brick red) docked to LigAB-WT (F103α, red). B) Docking of PCA to F103H (orange) showing non-productive binding of substrate possibly contributing to observed substrate inhibition. C) Docking of GA (light blue) to F103V (blue). D) F103T (purple) showing an additional hydrogen bond formed between the carboxylate of a docked GA and the Thr hydroxyl. E) LigAB-WT showing non-productive binding with docked 3OMG (green). F) F103T with 3OMG docked in a productive binding conformation.
It is interesting to note from the docking studies performed with LigAB-WT that PCA and GA consistently adopted a chemically productive geometry (with bidentate coordination to the bound iron by two vicinal hydroxyl groups; this can be seen for PCA in Figure 4.1a). However, 3OMG failed to adopt a geometry with bidentate hydroxyl coordination of iron within the wild-type enzyme, preferring instead to bind with the –OCH$_3$ facing away from Phe103α (Figure 4.1e) such that the iron was chelated by one hydroxyl and one ether oxygen. When 3OMG was docked onto any of the LigAB mutants (including F103L where little kinetic change was observed), it was able to find the physiologically relevant binding conformation (Figure 4.1f), suggesting that the steric bulk of Phe is indeed a contributor to the substrate specificity of LigAB. While Phe103α clearly plays a role in substrate recognition and specificity, there are very likely other contributing factors that are important for the discrimination of substrate functionality at positions other than C5.

Analysis of the mutant kinetics revealed that the activity ($k_{cat}$ and $k_{cat}/K_m$) with the preferred natural substrate (PCA) remains highest with wild-type enzyme (Table 4.1, Figure 4.2). The activity reductions with the Phe mutants are likely due to a loss of side-on π-stacking in the mutants as observed in the LigAB-WT crystal structure (PDB 1B4U), and additionally an increase in active-site volume which allows PCA to explore a larger conformational space. While mutations lead to decreases in $k_{cat}/K_m$ for PCA ranging from 4-fold to 200-fold, multiple mutants were identified in which $k_{cat}$ and/or $k_{cat}/K_m$ were enhanced for gallate and 3OMG. Substitution with the smaller hydrophobic valine, or its isosteric residue threonine, yielded mutant enzymes that are unchanged for GA dioxygenation, but that have greater turnover and lower $K_m$ values for the dioxygenation of 3OMG ($k_{cat}/K_m$ was increased in both cases by over an order of magnitude; Table 4.1). The similar catalytic profiles for F103V and F103T are likely due to the increased size of the active site pocket, while the ability of the Thr hydroxyl
**Table 4.1.** Steady-state Kinetic Constants for LigAB-WT and Mutants with PCA, Gallate and 3OMG

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Kinetic Constant</th>
<th>PCA</th>
<th>Gallate</th>
<th>3OMG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_m$ (µM)</td>
<td>51 ± 4</td>
<td>441 ± 55</td>
<td>2319 ± 316</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>216 ± 3</td>
<td>53 ± 2</td>
<td>7.3 ± 0.5</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}/K_m$ (M$^{-1}$s$^{-1}$)</td>
<td>4.26 x 10$^6$</td>
<td>1.21 x 10$^5$</td>
<td>3.13 x 10$^3$</td>
</tr>
<tr>
<td><strong>LigAB-WT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_m$ (µM)</td>
<td>261 ± 20</td>
<td>864 ± 158</td>
<td>1002 ± 54</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>89 ± 2</td>
<td>24 ± 1</td>
<td>49 ± 1</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}/K_m$ (M$^{-1}$s$^{-1}$)</td>
<td>3.41 x 10$^5$</td>
<td>2.78 x 10$^4$</td>
<td>4.89 x 10$^4$</td>
</tr>
<tr>
<td>Fold Change in</td>
<td>$k_{cat}/K_m$ from WT</td>
<td>- 12.5</td>
<td>- 4.4</td>
<td>+ 15.6</td>
</tr>
<tr>
<td><strong>F103V</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_m$ (µM)</td>
<td>1598 ± 222</td>
<td>6881 ± 2879</td>
<td>3943 ± 1304</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>33 ± 2</td>
<td>5 ± 1</td>
<td>7 ± 1</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}/K_m$ (M$^{-1}$s$^{-1}$)</td>
<td>2.07 x 10$^4$</td>
<td>7.26 x 10$^3$</td>
<td>1.78 x 10$^3$</td>
</tr>
<tr>
<td>Fold Change in</td>
<td>$k_{cat}/K_m$ from WT</td>
<td>- 205.8</td>
<td>- 166.7</td>
<td>- 1.8</td>
</tr>
<tr>
<td><strong>F103L</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_m$ (µM)</td>
<td>206 ± 17</td>
<td>212 ± 29</td>
<td>392 ± 55</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>112 ± 2</td>
<td>22.7 ± 0.6</td>
<td>15.1 ± 0.6</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}/K_m$ (M$^{-1}$s$^{-1}$)</td>
<td>5.44 x 10$^5$</td>
<td>1.07 x 10$^5$</td>
<td>3.85 x 10$^4$</td>
</tr>
<tr>
<td>Fold Change in</td>
<td>$k_{cat}/K_m$ from WT</td>
<td>- 7.8</td>
<td>- 1.1</td>
<td>+ 12.3</td>
</tr>
<tr>
<td><strong>F103T</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_m$ (µM)</td>
<td>108 ± 9</td>
<td>192 ± 29</td>
<td>926 ± 42</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>110 ± 3</td>
<td>41 ± 2</td>
<td>91 ± 1</td>
</tr>
<tr>
<td></td>
<td>$k_{cat}/K_m$ (M$^{-1}$s$^{-1}$)</td>
<td>1.02 x 10$^6$</td>
<td>2.14 x 10$^5$</td>
<td>9.83 x 10$^4$</td>
</tr>
<tr>
<td>Fold Change in</td>
<td>$k_{cat}/K_m$ from WT</td>
<td>- 4.2</td>
<td>+ 1.8</td>
<td>+ 31.4</td>
</tr>
</tbody>
</table>
to hydrogen bond with the C5-hydroxyl or carboxylate of 3OMG likely explains the catalytic differences between these mutants with this substrate. The increased stability of 3OMG in the F103T active site provided by a hydrogen bond such as this could result in the observed lower $K_m$ and reduced $k_{cat}$ if: (1) product release is or becomes rate limiting, or (2) the hydrogen bond reduces the electron density within the aromatic ring enough to result in slower oxidation. Computational docking of 3OMG into the F103T mutant supports this hypothesis as a hydrogen bond is formed between the enzyme and the C1 carboxylate of 3OMG (Figure 4.1).

In the Phe to His mutant, which preserves the side chain aromaticity but introduces potential for hydrogen bonding interactions in the active site, the catalytic efficiency is decreased by only a factor of four for the native substrate and increased for both alternate substrates. The F103H mutation results in a more polar and slightly more open active site environment in comparison to the wild-type Phe, which we hypothesize leads to the reduced $K_m$ for both GA and 3OMG and the overall improvements in $k_{cat}/K_m$ as compared to LigAB-WT. It should be noted, however, that significant substrate inhibition was observed for the F103H mutant at PCA or GA concentrations higher than 1 mM. Consistent with practice observed in the literature, only the data below 1.5 mM for PCA or 1 mM for GA was used for a least-squares fitting to the Michaelis-Menten equation in these cases (Figure 4.2) since fits to the data by a substrate inhibition equation (Eq. 2) did not adequately describe the data (Figure 4.3). Substrate inhibition of the magnitude seen with the F103H mutant was not observed for LigAB-WT, but has been observed with many wild-type dioxygenases for reactions with both native and alternate substrates. Since classical substrate inhibition (requiring the binding of a second substrate to an active enzyme-substrate complex) does not appear to adequately describe the observed reduction in turnover at high substrate concentrations, we hypothesize
that in this case the reduction in turnover at higher concentrations might arise through potential non-productive

**Figure 4.2.** Steady-state kinetics of LigAB Phe103α mutants with known LDAC metabolic pathway substrates. (A) Steady-state kinetics of F103H. (B) Steady-state kinetics of F103V. (C) Steady-state kinetics of F103L. (D) Steady-state kinetics of F103T. PCA (red circles, solid line), Gallate (blue squares, dashed line), 3OMG (green diamonds, dotted line). The dependence of the initial rates on the organic substrate concentration was determined in air-saturated buffer (250-275 µM O₂, pH 7.5, 25 °C) and the curves are fit to the Michaelis-Menten equation unless otherwise noted. The data for F103H with PCA and GA were each split at the substrate concentration where rate decreases with additional substrate (1500 µM, and 1000 µM respectively) and both sections were subsequently fit with the Michaelis-Menten equation.
Figure 4.3. Comparative fitting of substrate inhibition observed in the steady-state kinetics of the LigAB mutant F103H with A) PCA (red circles), and B) GA (red squares). The data are shown with least-squares fitting to both substrate inhibition (solid orange), and the Michaelis-Menten (blue dashed) kinetics. The Michaelis-Menten equation was used to fit the data twice – once for rates increasing and once for rates decreasing (inhibitory phase) with concentration. Since the equation for reversible substrate inhibition does not fit well to the data, it is possible that more complicated processes are occurring at concentrations above ~1 mM.
binding of PCA where the C3-hydroxyl group faces the 103α residue leading to a hydrogen bond forming with the introduced histidine residue. This scenario was indeed observed in 24% of bound conformations resulting from the computational docking of PCA to the F103H mutant (Figure 4.1b). A similar non-productive “flipped” binding mode of PCA occurred in 16% of docked conformations to LigAB-WT; however, hydrogen bonding to the native Phe103α side chain is not possible, and so this conformation is inherently less stable than in F103H.

While the rate profile observed for F103H was not consistent with the equation that describes classical substrate inhibition, it is interesting to note that our docking studies did reveal two solvent inaccessible cavities at the α-β heterodimer interface where PCA and vanillin were observed to bind during shotgun-style docking studies (Figure 4.4). It is possible that these cavities represent a channel through which either substrate enters or product leaves the active site. While there is currently no data other than the aforementioned computational docking to suggest a channel to funnel substrate/product in or out of the active site, these location would provide a sites to which a second substrate could bind at high concentrations allowing for substrate inhibition to occur. In this scenario, substrate inhibition would be more likely if the channel were an exit route for product, where binding of substrate would essentially cause a backup in the production line.
Figure 4.4. Internal cavities of LigAB shown to bind PCA and vanillin during shotgun-style computational docking studies. The active-site of LigAB (light grey = α subunit, dark grey = β subunit) is denoted by the crystallographically bound PCA (red sticks) and iron cofactor (orange). The hypothesized allosteric binding site is shown in green, while the two internal cavities representing a possible substrate/product channel are shown in aqua and magenta.
4.3.2. Impact of Phe103α Mutants on Allosteric Activation of Alternate Substrate Dioxygenation. Since each of the mutants had altered dioxygenation rates for the various substrates as compared to LigAB-WT, the mutants were also examined for alterations in allosteric activation with GA and 3OMG (Table 4.2). Like LigAB-WT, F103V and F103L also show no activation (or inhibition) of either GA or 3OMG turnover in the presence of vanillin. F103H and F103T, however, show a change in turnover rate under all substrate-vanillin combinations (Figure 4.5). For F103H, vanillin diminished the rate of turnover of all substrates, albeit to varying degrees. In addition to inhibiting the turnover of PCA by F103T, vanillin inhibits F103T turnover of 3OMG as well – though to a

Table 4.2. Percentage Enhancement or Inhibition of Turnover as a Result of Co-incubating Substrate and Vanillin

<table>
<thead>
<tr>
<th>Enz.</th>
<th>PCA</th>
<th>Gal</th>
<th>3OMG</th>
</tr>
</thead>
<tbody>
<tr>
<td>LigAB</td>
<td>+37 ± 3 %</td>
<td>no change</td>
<td>no change</td>
</tr>
<tr>
<td>F103V</td>
<td>+20 ± 4 %</td>
<td>no change</td>
<td>no change</td>
</tr>
<tr>
<td>F103L</td>
<td>no change</td>
<td>no change</td>
<td>no change</td>
</tr>
<tr>
<td>F103T</td>
<td>Inhib†</td>
<td>+56 ± 3 %</td>
<td>Inhib†</td>
</tr>
<tr>
<td>F103H</td>
<td>-23 ± 18 %</td>
<td>Inhib†</td>
<td>-20 ± 4 %</td>
</tr>
</tbody>
</table>

† Poor tightness of fit to Eq. 4.2 prevents an accurate estimation of the percent inhibition in these cases.

Figure 4.5. Steady-state kinetic enhancement and inhibition of LigAB Phe103α mutants in the presence of vanillin. (A) F103T. (B) F103H. Color coding for substrates is as follows: PCA (red circles, solid line), GA (blue squares, - -), and 3OMG (green diamonds, dotted line). The dependence of the initial rates on the vanillin concentration was determined in air-saturated buffer (250-275 µM O₂, pH 7.5, 25 °C), 10x Kₘ substrate concentration or 5 mM, and the curves are fit to Equation 4.2.
lesser extent yet still maintaining a linear relationship with vanillin concentration. Interestingly, despite the inhibition of PCA and 3OMG turnover, vanillin elicits a turnover rate enhancement, $56 \pm 3\% \ (K_{\text{cat}} = 0.4 \pm 0.1 \text{mM})$, of GA dioxygenation by F103T. This is the only enhancement observed with a substrate other than PCA for any of the examined mutants. Because the mechanism of the activation is thus far unknown, this result makes the F103T another interesting enzyme with which to pursue mechanistic studies in the future.

4.4. Conclusions

Understanding the naturally evolved mechanisms by which enzymes select or exclude substrates is important for the growing use of bioengineered microbial systems in commodity chemical production. The steady-state kinetic studies of LigAB-WT and Phe103α mutants have demonstrated one mechanism by which this enzyme imparts substrate specificity. Through these mutations, LigAB’s natural propensity to catalyze the dioxygenation of non-native substrates has been enhanced – reducing the relative specificity differences between protocatechuate, gallate, and 3-O-methyl gallate. Of note, three of the mutant enzymes (F103V, F103H and F103T) have $k_{\text{cat}}/K_m$ values that are greater than those reported for DesZ for the dioxygenation of 3OMG ($1.03 \times 10^4 \text{M}^{-1}\text{s}^{-1}$) which is the physiological catalyst for ring-opening 3OMG. The ability to utilize GA remained relatively constant regardless of mutation. These data demonstrate that, while PCA remains the preferred substrate in all cases, LigAB can be modified such that activity ($k_{\text{cat}}$ and/or $k_{\text{cat}}/K_m$) towards alternate substrates can be increased with only moderate reduction in catalytic efficiency of the natural substrate. Researchers attempting to engineer a microorganism for the fermentative conversion of LDACs into biofuels or fine chemicals would therefore be able to take advantage of the enzymatic proficiency of F103H for the dioxygenation of PCA, GA and 3OMG rather than needing to introduce multiple enzymes to perform these functions.
4.5. Future Directions

The mutations to Phe103α described in this chapter have shown that this non-catalytic residue contributes an integral steric role to LigAB’s substrate recognition and selection mechanisms. The C1, C2, C5, and C6 functional positions of the native PCA substrate have significant non-catalytic interactions with the active site (Figure 4.6). Mutation of the active site residues that interact with these position may allow the substrate scope of LigAB to be further expanded. We have shown that mutations of Phe103α modulate LigAB’s ability to catalyze the dioxygenation of substrates with varied C5 functionality – increasing the tolerance for larger functional groups. Future work to alter LigAB’s substrate reactivity profile will focus on active site residues that contribute steric or hydrogen bonding interactions with C1 functional groups of potential substrates (Figure 4.6), as LDACs rarely display C2 and C6 functional variability. Mutations to Asn270β and Thr271β could allow substrates with bulkier or extended C1 functionalities (e.g. homoprotocatechuate and methyl-3,4-dihydroxybenzoate) to be more acceptable substrates. Turnover of substrates with hydrophobic C1 functional groups (e.g. catechol and 4-methylcatechol) may benefit from replacement of Ser269β and/or Asn270β with more hydrophobic residues. Additionally, polar mutations of Ser269β may be able to increase turnover of substrates with uncharged polar C1 functional groups (e.g. 3,4-dihydroxybenzaldehyde, and 3,4-dihydroxybenzonitrile). Expanding our understanding of how LigAB active site structure components influence substrate recognition and selection will help in the creation of dioxygenases with broad substrate scope, as well as contribute to the growing collective knowledge of how enzyme structure begets function.
Figure 4.6. Two dimensional representation of the LigAB wild type active site derived from the LigAB crystal structure (1b4u.pdb). Hydrogen bonding (green dashed lines) and steric interactions (red arcs) between the native PCA substrate and active site residues are shown. Mutations that may lead to further alterations in substrate utilization are suggested. Active site image created using LigPlot+.
Chapter 5:

Design and Synthesis of FRET-Pair Labeled Lignin Model Compounds
5.1. Introduction

Up to this point, the catabolism of lignin derived aromatic compound monomers has been the main focus of our investigations. However, a more detailed understanding of the ring-opening dioxygenation of LDACs is one of many necessary steps towards successfully engineering enzymatic pathways for the streamlined conversion of feedstocks to bio-produced chemicals. The most daunting enzymatic barrier to the efficient use of lignin as an industrialized carbon source may be the initial depolymerization step. Lignin is the second most abundant naturally occurring biopolymer, and exhibits a recalcitrance to degradation that currently limits the utility of enzymes in industrial depolymerization. In order to overcome a barrier such as this, there is a need for new degradation strategies which more rapidly breakdown lignin polymers into monomeric and dimeric LDACs.

Fungi, especially white-rot fungi such as Phanerochaete chrysosporium, are the most prolific and well-studied of nature’s lignocellulose degrading microorganisms, and are the largest source of currently studied lignin depolymerizing enzymes. Often these fungi have multiple genes encoding both lignin peroxidases (LiP) and manganese peroxidases (MnP), while other fungi produce extracellular laccases. There are, however, several lesser explored organismal systems in which lignin degradation is known or believed to occur: soil bacteria, mixed fungal and bacterial systems, and bacteria in the guts of wood consuming insects such as termites, and the Asian longhorned beetle. Genome sequencing and bioinformatic analysis (genome mining) of new organismal systems is often the fastest way of determining the presence and identity of potentially useful enzymes.

Bacteria and fungi from these systems are often symbiotic, and rely on nutrients or environmental conditions provided by other bacteria, fungi, or the host – making growth under laboratory conditions and isolation of pure DNA samples challenging. Additionally, recent
meta-genomic analyses of the microflora in these challenging systems (mixed fungal communities, insect-guts, etc) have failed to show homologues of known lignolytic enzymes. Therefore, locating the exact source of (ex. location within an insect’s gut, or organism within a mixed system) and detecting lignin degradation in real time may be able to increase the success rate of identifying sources of lignin degradation from systems where lignin is degraded quickly. To accomplish this, a probe capable of producing a signal in the presence of lignin depolymerization, and detectable through living organic tissues without the necessity of isolating reaction products from the medium would be a highly valuable tool.

The characterization of lignocellulose degrading organisms and enzymes via the use of lignin model compounds is well established. These lignin model compounds range from the very simple mimics\textsuperscript{77, 232} to those that attempt to accurately resemble lignin inter-subunit bonding motifs,\textsuperscript{81, 233-245} or even those which are direct modifications of lignin polymer (Figure 5.1).\textsuperscript{80} Most often, low molecular weight lignin model compounds (small organic molecules), such as Alderol,\textsuperscript{234-240, 246-248} attempt to mimic the $\beta-O-4$ bonding motif found in natural lignin polymers.

However, very few lignin model compounds are able to provide a convenient method of real time data collection in assays of lignin depolymerization, and generally require the use of radioactivity and/or subsequent separation and analysis of degradation products by LC-MS. Recent efforts to solve this problem resulted in milled wood lignin polymer covalently modified by the addition of a fluorescent dye or the nitration of lignin subunit aromatic rings (Figure 5.1).\textsuperscript{80} These modified lignins provide a facile method for spectrophotometric determination of lignin degrading organisms or enzymes by changes in either fluorescence or absorption, respectively.
Figure 5.1. Examples of lignin model compounds used in previous studies of lignin depolymerization. The structure of Alderol, a commonly used lignin model compound, is shown on the top left. Numbers denote references in which the corresponding lignin model compounds were used.
The research conducted herein describes synthesis of low molecular weight lignin probes (Scheme 5.1) designed to utilize the phenomenon of fluorescence resonance energy transfer (FRET) to produce a detectable signal. FRET is a non-radiative process by which the energy of a chromophore (light sensitive molecule) in an excited state (referred to as the donor) is transferred to a second chromophore (acceptor) through dipole-dipole coupling. Importantly, the efficiency of energy transfer is sensitive to the distance between the donor and acceptor. Energy transfer is most efficient when the chromophores are close to each other and consequently decreases as the inter-chromophore distance increases. While this relationship can be used for highly accurate measurements of very small distances on the order of Ångstroms, our interests lie in its alternative function as an ON/OFF switch for reporting on the state the lignin model compound.

The aim of this study was to design and synthesize lignin model compounds that can accommodate fluorescent dyes (a donor and acceptor) located on opposite ends of a β-O-4 linked lignin dimer. Monitoring the fluorescence of the donor would allow for an “on or off” determination of the presence of lignin depolymerization as is depicted in Scheme 5.2. In the absence of lignin degradation, the β-O-4 bond motif and the dimer remain unaltered allowing FRET to occur between the donor and the acceptor. In this state, the observable fluorescence of the donor should be low since the excited donor will transfer its energy to the acceptor which effectively quenches donor fluorescence. This low donor fluorescence state can be considered the “off state”. When the probe encounters a lignin depolymerization agent such as lignin peroxidase, the β-O-4 subunit linkage is cleaved along the Cα-Cβ bond. Upon cleavage, the donor will diffuse away from the acceptor. The increase in distance would reduce or eliminate the occurrence of FRET, resulting in a high donor fluorescence state which can be considered the “on state”.
Scheme 5.1. Retrosyntheses of FRET labeled lignin model compounds. Left branch: ester linkage. Right branch: linkage via click chemistry.
Scheme 5.2. FRET-pair labeled lignin model compound concept. A) Chemical structure (top) and graphical representation (bottom) of a complete FRET-pair labeled lignin model compound. Important regions of the lignin model compound are color coded as follows: lignin dimer core (green), tetramethylrhodamine (TAMRA) dye (purple), Fluorescein (FAM) dye (yellow/orange). B) Graphic representation of the lignin model compound high-FRET/low-fluorescence state (left), and low-FRET/high-fluorescence state (right) resulting from lignin depolymerization activity. Incident light used to probe the molecule is represented by the blue wave. C) Chemical reaction of the lignin model compound demonstrating β-O-4 bond cleavage.
Ideally, excitation and emission spectra of the dyes incorporated into this probe would fall in the near-infrared (NIR) due to the ability of NIR wavelengths to penetrate living organic tissues. However, the probe is designed such that any FRET pair can be used depending on the desired application, and is demonstrated here using the FRET pair fluorescein (FAM, donor) and tetramethylrhodamine (TAMRA, acceptor/quencher) as shown in Figure 5.2. While the absorption spectrum of FAM does not allow for excitation by NIR wavelengths, the TAMRA-FAM pair is a commonly used in cellular imaging, and so will provide an initial proof of concept and working model. Additionally, preliminary fluorescence assays of TAMRA and FAM in the presence of lignin peroxidase have shown that these fluorophores are stable under the in vitro assay conditions that would be used for initial activity studies of labeled lignin model compounds. A lignin depolymerization probe such as the one described here provides the well-defined chemical structure of low molecular weight lignin model compounds while also including and expanding upon the capabilities of spectrophotometric abilities provided by fluorescent and nitrate modified lignin polymer probes (detection in or through living tissues, possible single molecule detection applications).

![Chemical structures of Tetramethylrhodamine (TAMRA, purple), and Fluorescein (FAM, orange).](image-url)
5.2. Results and Discussion

5.2.1. Synthesis

Several routes to synthesize the bi-functionalized low molecular weight lignin model compounds described here were explored. Compound 18 was synthesized as previously described by Ciofi-Baffoni, et al. (Scheme 5.3), and was used as the starting point from which our syntheses diverged. The aldehyde and benzyl protected phenolic oxygen of 18 provided two positions suitable to sequential modification, and additionally allowing for the attachment of a fluorescent dye to each aromatic ring of the β-O-4 linked lignin model dimer core.

Compound 18 of the original Ciofi-Boffani synthesis was reduced and yielded the benzyl alcohol which was subsequently protected by the addition of a TBDMS group. This then allowed the benzyl protecting group to be removed, revealing the phenolic alcohol in 102

![Scheme 5.3. Reagents and condition: (a) Methyl α-bromoacetate, anhydrous K$_2$CO$_3$, anhydrous acetone, N$_2$, reflux. (b) p-TsOH, ethylene glycol, benzene, N$_2$, reflux using a Dean-Stark trap. (c) LDA, anhydrous THF, N$_2$, -78 °C/room temperature. (d) NaBH$_4$, 3:1 THF/H$_2$O, room temperature. (e) PPTS, wet acetone, reflux. (f) Anhydrous p-TsOH, acetone, room temperature. Adapted from Ref. 250.](image-url)
(Scheme 5.4). The differential protection provided handles to sequentially and selectively attach the FRET agents. Initially, conjugation of the fluorescent dyes to the lignin model compound core was envisioned to proceed through Steiglich esterification utilizing a free core hydroxyl group and an activated carboxylic acid on the fluorescent dye (Scheme 5.5). Due to the presence of free phenolic hydroxyl groups on FAM as seen in Schemes 5.2 and 5.5, TAMRA was conjugated to the lignin model dimer first to provide 118 in order to avoid unwanted esterification of FAM hydroxyls. Addition of both FAM and TAMRA dyes to the lignin model dimer was successfully accomplished by the esterification route using the succinimidyl ester-activated form of each dye (compound 122). However, subsequent deprotection of the remaining acetonide protecting group to reveal the α- and γ-hydroxyls was unsuccessful. The acidic conditions explored to perform this deprotection, wet acetone/PPTS or Amberlyst catalyst, resulted in either the undesired removal of the ester linked FAM to reform 118 or no reaction, respectively. In either case, acetonide deprotection was not observed. These results led us to believe that the attached TAMRA may be hindering the acetonide from deprotection by possibly π-stacking with the core’s aromatic rings. Despite the unfortunate resilience of the acetonide protecting group, the inability of one of the dyes to

**Scheme 5.4.** Reagents and condition: (a) NaBH₄, 3:1 THF/H₂O, room temperature. (b) TBDMS-Cl, imidazole, anhydrous DMF, N₂, room temperature. (c) 10 % Pd/C, H₂, ethanol.
remain attached under mild acidic conditions, as demonstrated by the weakness of the FAM ester-linkage, is antithetical to the goal and purpose of a doubly labeled lignin model dimer.

In order to avoid further complications imparted by ester-linkages, an alternative synthetic route utilizing copper catalyzed click reactions (azide-alkyne Huisgen cycloaddition) was explored (Scheme 5.6). The click-reaction pathway was initially designed such that two propargyl groups were installed onto the lignin dimer prior to the addition of either fluorophore (Appendix Scheme A.1). This would have allowed for the possibility of a single-pot reaction for the sequential click-addition of the fluorophores as described by Aucagne and Leigh.\textsuperscript{25} A lignin dimer core with two terminal alkynes (one protected and one unprotected) would allow for one click reaction to be performed, followed by an \textit{in situ} deprotection of the second alkyne and second click reaction.

\textbf{Scheme 5.5.} Reagents and condition: (a) DCC, DMAP, TAMRA-NHS, anhydrous CH\textsubscript{2}Cl\textsubscript{2}, N\textsubscript{2}, reflux, dark. (b) TBAF, anhydrous THF, N\textsubscript{2}, 0 °C/room temperature, dark. (c) DCC, DMAP, FAM-NHS, anhydrous CH\textsubscript{2}Cl\textsubscript{2}, N\textsubscript{2}, reflux, dark. (d) PPTS, 3:2 acetone/water, reflux, dark.
After installation of the first propargyl to form 147 and subsequent removal of the TBDMS to form 148 (Scheme 5.6), attempts to install the TIPS-propargyl onto 148 were unsuccessful which lead to an attempt to first form the triazole linkage between 148 and TAMRA-N3, followed by the addition of a second unprotected propargyl group to the free alcohol as shown in Scheme 5.6. Both of these reactions were successful, resulting in compounds 149 and 150. The propargyl functionalization of 149 to give 150 installs the point at which the second fluorophore (FAM-N3) would be attached via a second click reaction. Click reactions using FAM-N3 have to this point been ineffective, having shown no evidence of FAM linkage. Under standard reaction conditions containing CuSO₄ and sodium ascorbate in THF with minimal water, addition of FAM-N3 to the reaction mixture

**Scheme 5.6.** Reagents and condition: (a) Propargyl bromide, anhydrous K₂CO₃, anhydrous acetone, N₂, reflux. (b) TBAF, anhydrous THF, N₂, 0 °C/room temperature. (c) CuSO₄, sodium ascorbate, TAMRA-N3, 3:1 THF/H₂O, N₂, room temperature, dark. (d) Propargyl bromide, anhydrous K₂CO₃, anhydrous acetone, N₂, reflux. (e) CuSO₄, sodium ascorbate, FAM-N3, dark. (f) PPTS, 3:2 acetone/water, reflux, dark.
containing 150 resulted in an immediate black discoloration of the copper salt or sodium ascorbate. The cause of the discoloration remains unclear, yet seems to result in catalyst inactivation. Additionally, 150 was unable to be recovered from or identified by ESI-MS in the resulting reaction mixtures. Control click reactions of FAM-N₃ with model compound 148, with no TAMRA attached, showed the same black discoloration, and resulted in no observable formation of product. Attempts to perform the click reaction under high temperature with no copper catalyst, or two fold excess of the copper (I) catalyst, [Cu(MeCN)₄]PF₆ (tetrakis(acetonitrile) copper (I) hexafluorophosphate), also proved unsuccessful. Several examples of successful azide-alkyne cycloadditions of FAM-N₃ have been reported. In these examples, the length of the linker between the core and the FAM-N₃ is longer than used here. Additionally, these examples perform the reactions in either DMSO or 9:1 DMSO/H₂O, and include the base N,N-diisopropylethylamine (DIPEA) and/or the chelating agent tris(benzyltriazolyl)methyl amine (TBTA). TBTA acts to stabilize the Cu(I) oxidation state of the active catalyst, and accelerate the catalysis in water containing systems. Utilizing a more amphipathic solvent system (DMSO), stabilizing the Cu(I) catalyst with TBTA, and reducing steric effects by using longer linkers may be beneficial variations in further attempts at the cycloaddition of FAM-N₃.

In addition to the design and synthesis of the functional lignin model compounds described above, control molecules were constructed. Compound 122, described previously, acts as a non-cleavable control molecule containing both FRET agents. Additionally, compounds 118 (Scheme 5.5), and 149 (Scheme 5.6) serve as appropriate TAMRA linked control molecules; however, a control containing only the FAM fluorophore does not directly result from the planned synthetic ester or azide-alkyne click chemistry linkage pathways. A control molecule containing FAM was therefore designed and synthesized as shown in Scheme 5.7.
Scheme 5.7. (a) DCC, DMAP, FAM-NHS, anhydrous CH$_2$Cl$_2$, N$_2$, reflux, dark. (b) PPTS, 3:2 acetone/water, reflux, dark.
5.3. Conclusions and Future Directions

5.3.1. Conclusions

Significant progress has been made towards the synthesis of a lignin model compound with novel functionality. Successful incorporation of a fluorescent dye FRET-pair into the lignin dimer model compound will ultimately allow for real time spectrofluorometric analysis of lignin depolymerization activity, both in vitro and in vivo. The discovery of new enzymes capable of more efficient and expedient lignin depolymerization is paramount to large scale industrial and commercial utilization of lignin, and more generally lignocellulose, as a carbon feedstock in bio-processing. This FRET-labeled lignin model compound will allow the exploration of complex living systems which are not easily investigated by more “traditional” genomic and bioinformatic approaches.

5.3.2 Future Directions

5.3.2.1. Synthesis. Although proving to be challenging, significant progress has been made toward the completion of the FRET-pair labeled lignin model compounds by both ester and click-chemistry linkage routes. Future work on and the ultimate completion of this project will be performed by undergraduate and/or graduate students under the guidance of Prof. Erika Taylor. This work is currently being pursued by Mackenzie Schlosser (Wesleyan University, Class of 2016), who is investigating alternate conditions for performing the copper catalyzed click-reaction of the dye FAM-N₃ to compound 150 and other propargyl labeled lignin model compounds. Further, she is pursuing alternative routes by which this complication can be avoided. To arrive at the acetonide-protected FRET-pair labeled lignin model compound as described above (Scheme 5.6), Mackenzie has explored the alternative of first performing the click-reaction of FAM-N₃ with propargyl bromide, which has shown to be successful, followed by the nucleophilic substitution utilizing the free alcohol of compound 149, which is still in
progress. If this method also proves unsuccessful, linkage of both dyes of the FRET-pair by the same methodology is likely not critical to a properly functioning probe, and so other methods for fluorescein labeling, including the use of fluorescein isothiocyanate as used in the single-dye labeling of milled wood lignin polymer, could to be explored.

5.3.2.2. Biological Significance. Upon final completion, the FRET-pair labeled lignin model compounds will be further analyzed in conjunction with the laboratory of Prof. Ming Tien (Pennsylvania State University, Department of Biochemistry and Molecular Biology). Initial studies will focus on in vitro assays to analyze the ability of the lignin model compound to qualitatively and quantitatively report on lignin depolymerization in the presence of the lignin peroxidase of Phanerochaete chrysosporium as proof of concept. Subsequent to confirmation of FRET-labeled lignin model compound function by in vitro studies, in vivo studies of these compounds will be pursued in mixed fungal systems and other unique lignin degrading niches such as the microbial communities of insect guts.

5.4. Experimental Methods

5.4.1. Materials and General Methods.

All chemicals were obtained from either Sigma-Aldrich or Alfa Aesar, and used without further purification. N-succinimidyl ester (NHS) activated carboxytetramethylrhodamine (TAMRA-NHS) and carboxyfluorescein (FAM-NHS) were obtained from Thermo Fisher Scientific (Waltham, MA). Tetramethylrhodamine-azide (TAMRA-N3) and fluorescein azide (FAM-N3) were obtained from Lumiprobe Corp. (Hallandale Beach, FL). Silica gel plates (SiliCycle Inc., Quebec City, Quebec, Canada) and silica gel columns (GRACE, 40 µm) were used for TLC analysis and column chromatography. All column chromatography was performed on a GRACE Reveleris flash chromatography system (Colombia, MD). The $^1$H and
\[ ^{13}C \text{NMR spectra were obtained with a Bruker 400 MHz NMR spectrometer. The chemical shifts are reported in } \delta \text{ (ppm) and are relative to the central peak of the residual solvent.} \]

All proton and carbon chemical shifts were assigned a letter as demonstrated in Figure 5.3. The coupling constants \( (J) \) are expressed in Hz. The following abbreviations are used for peak multiplicity or assignment: \( s = \text{singlet}, \ d = \text{doublet}, \ dd = \text{doublet of doublets}, \ t = \text{triplet}, \ m = \text{multiplet}, \ br. = \text{broad}, \ ar. = \text{aromatic}. \)

GC-MS spectra were obtained for compounds 8 and 4; however, GS-MS spectra were not obtainable for the majority of compounds. ESI-MS spectra were obtained for compounds containing attached fluorophores.

![Diagram](image)

**Figure 5.3.** Proton and carbon labeling diagram used for NMR assignments. A) Labeling for the lignin model compound core. B) Labeling of compound 90 as an example of phenolic and benzylic oxygen substituent assignments.

### 5.4.2. Chemical Synthesis

**Methyl (4-formyl-2-methoxyphenoxy) acetate (8)**

Methyl \( \alpha \)-bromoacetate (2.8 mL, 29.6 mmol) and anhydrous \( \text{K}_2\text{CO}_3 \) (4.08 g, 29.5 mmol) were added to a solution of vanillin (3 g, 19.7 mmol) in dry acetone (33 mL) at room temperature. The mixture was heated at reflux for 2 h, filtered, and the solid residue washed with EtOAc. The filtrate was concentrated in vacuo and the resulting yellow solid was recrystallized from hot EtOH to yield pure 8 as a white fibrous crystalline solid (3.65 g, 16.28 mmol, yield 86%).

\[^{1}H\text{ NMR (CDCl}_3, 300 \text{ MHz}) \delta \text{ ppm 3.81 (s, 3 H, k), 3.96 (s, 3 H, h), 4.80 (s, 2 H, i), 6.87 (d, J=7.92 Hz, 1 H, f), 7.38 - 7.48 (m, 2 H, c and g), 9.87 (s, 1 H, a).} \]

\[^{13}C\text{ NMR (CDCl}_3, 75 \text{ MHz}) \]

149
\[ \delta \text{ ppm } 52.47 \text{ (s, 1 C, k), 56.10 (s, 1 C, h), 65.82 (s, 1 C, i), 109.86 (s, 1 C, ar.), 112.32 (s, 1 C, ar.), 126.15 (s, 1 C, ar.), 131.20 (s, 1 C, ar.), 149.95 (s, 1 C, ar.), 152.43 (s, 1 C, ar.), 168.53 (s, 1 C, j), 190.84 (s, 1 C, a). } \]

GC-MS \( m/z \): 224 (M\(^+\)), 165 (M\(^+\) – CO\(_2\)CH\(_3\)), 151 (M\(^+\) – CH\(_2\)CO\(_2\)CH\(_3\)), 137 (M\(^+\) – CH\(_2\)CO\(_2\)CH\(_3\) – CH\(_3\)), 105 (M\(^+\) – OCH\(_2\)CO\(_2\)CH\(_3\) – OCH\(_3\)), 77 (M\(^+\) – CHO – OCH\(_2\)CO\(_2\)CH\(_3\) – OCH\(_3\)). Expected: 224.07 (M).

**Methyl [4-(1,3-dioxolan-2-yl)-2-methoxyphenoxy]acetate (4)**

To a solution of 8 (3.17 g, 14.1 mmol) in benzene (56 mL), ethylene glycol (1.2 mL, 21.5 mmol) and a catalytic amount of pTsOH•H\(_2\)O (0.09 mol%, 2.5 mg) were added to the solution. Using a Dean-Stark trap to distill off the benzene-H\(_2\)O azeotrope, the reaction mixture was refluxed for 2.5 h after the first droplet of water was collected in the trap. The reaction was allowed to cool, and the reaction mixture was washed thoroughly with water to remove ethylene glycol from the organic layer, and then dried over Na\(_2\)SO\(_4\), filtered and concentrated in vacuo. The resulting solid was recrystallized from hot EtOH to yield 4 as a white crystalline solid consisting of square iridescent plates (3.46 g, 12.89 mmol, yield 91 %). Difficulties in the recrystallization resulted if ethylene glycol was not completely removed during washing.

\(^1\)H NMR (CDCl\(_3\), 300 MHz) \( \delta \) ppm 3.78 (s, 3 H, k), 3.91 (s, 3 H, h), 3.98 - 4.19 (m, 4 H, l), 4.70 (s, 2 H, i), 5.75 (s, 1 H, a), 6.81 (d, \( J=8.21 \) Hz, 1 H, f), 7.00 (dd, \( J=8.36, 1.61 \) Hz, 1 H, g), 7.05 (d, \( J=1.47 \) Hz, 1 H, c). \(^{13}\)C NMR (CDCl\(_3\), 75 MHz) \( \delta \) C ppm 52.29 (s, 1 C, k), 55.96 (s, 1 C, h), 65.30 (s, 2 C, l), 66.45 (s, 1 C, i), 103.53 (s, 1 C, a), 109.92 (s, 1 C, ar.), 113.73 (s, 1 C, ar.), 119.15 (s, 1 C, ar.), 132.12 (s, 1 C, ar.), 147.92 (s, 1 C, ar.), 149.66 (s, 1 C, ar.), 169.37 (s, 1 C, j). GC-MS \( m/z \): 267 (M\(^+\) –H), 237 (M\(^+\) – OCH\(_3\)), 209 (M\(^+\) – CO\(_2\)CH\(_3\)), 196 (M\(^+\) – CHO – OCH\(_2\)CH\(_2\)CH\(_2\)), 123 (M\(^+\) – CH\(_2\)CO\(_2\)CH\(_3\) – CHO\(_2\)CH\(_2\)CH\(_2\)). Expected: 268.09 (M).
(2R*,3R*)- and (2R*,3S*)-Methyl 3-(4-benzyloxy-3-methoxyphenyl)-2-[4-(1,3-dioxolan-2-yl)-2-methoxyphenoxy]-3-hydroxypropionate (15a,b)

LDA was formed in situ by the addition of n-butyllithium (9.1 mL of a 1.6 M solution in hexanes, 14.56 mmol) to diisopropylamine (2.0 mL, 14.56 mmol) in dry THF (24 mL) at 0 °C under nitrogen. After 20 min, the LDA solution was cooled to -78 °C, and a solution of 4 (3.39 g, 12.6 mmol) in dry THF (20 mL) was added dropwise resulting in a bright yellow solution. After stirring for 10 min, a solution of benzylvanillin (3.06 g, 12.6 mmol) in dry THF (10 mL) was added dropwise. After stirring an additional 40 min at -78 °C the reaction solution was quenched by the addition of saturated aqueous NH₄Cl (40 mL), and allowed to warm to room temperature. The resulting mixture was extracted with EtOAc (3 x 50 mL), the combined organic layers were dried over Na₂SO₄, filtered, and the solvent was removed in vacuo to yield a viscous yellow-orange oil containing the 15a and 15b diastereomers in a 72:28 diastereomeric ratio as previously reported. The diasteriomers were separated by flash chromatography on silica gel using a gradient of 50 – 80 % EtOAc in hexanes. Compound 15a (2.18 g, 4.26 mmol, 34 % isolated yield) was isolated as a pale yellow viscous oil. ¹H NMR (CDCl₃, 300 MHz) δ ppm 3.51 - 3.62 (m, 1 H, l), 3.65 (s, 3 H, m), 3.88 (s, 3 H, h/n), 3.84 (s, 3 H, h/n), 3.96 - 4.19 (m, 4 H, v), 4.70 (d, J=5.28 Hz, 1 H, j), 5.05 - 5.12 (m, 1 H, i), 5.14 (s, 2 H, a₅), 5.72 (s, 1 H, u), 6.78 - 7.09 (m, 6 H, ar.), 7.26 - 7.46 (m, 5 H, a₁₃).

(1R*,2S*)-4-[1-(4-Benzyloxy-3-methoxyphenyl)-1,3-dihydroxy-propan-2-yloxy]-3-methoxybenzaldehyde (17)

To a solution of 15a (1.16 g, 2.27 mmol) in 3:1 THF:H₂O (20 mL), NaBH₄ (0.86 g, 22.66 mmol) was added portion-wise over 6 h at room temperature. The mixture was allowed to stir rapidly for 16 h. Excess NaBH₄ was then quenched by the addition of water (10 mL). The
reaction solution was extracted with Et₂O (3 x 35 mL), and the combined organic layers dried over Na₂SO₄. The solvent was then removed in vacuo to yield 16 as a milky-white oil (1.05 g, 2.18 mmol, 96 % yield). The product was used without further purification. ¹H NMR (CDCl₃, 300 MHz) δ ppm 2.76 (br. t, J=6.16 Hz, 1 H, m), 3.52 (d, J=2.64 Hz, 1 H, l), 3.58 - 3.72 (m, 1 H, j), 3.82 - 3.93 (s x 2, 6 H, h/n), 3.97 - 4.11 (m, 6 H, k/v), 4.93 (br. t, J=2.90, 2.90 Hz, 1 H, i), 5.13 (s, 2 H, a₅), 5.73 (s, 1 H, u), 6.76 - 7.09 (m, 6 H, ar.), 7.27 - 7.47 (m, 5 H, a₁-3).

To a solution of 16 (0.97 g, 2.00 mmol) in 4:1 acetone:water (22.5 mL), pyridinium p-toluenesulfonate (0.1503 g, 0.60 mmol) was added at room temperature. The mixture was then heated at reflux for 4 h, after which the acetone was distilled from the reaction solution, and a saturated solution of NaHCO₃ (50 mL) was added to the residue. The aqueous phase was extracted with Et₂O (3 x 50 mL). Additionally, the reaction flask contained a yellow residue insoluble in either water or Et₂O, but which dissolved in CH₂Cl₂ or EtOAc. After dissolving the residue in EtOAc it was combined with the Et₂O layers from the extraction. The resulting organics were washed with brine (3 x 50 mL), and dried over sodium sulfate. Flash chromatography on silica gel with a gradient of 60 – 100 % EtOAc in hexanes removed several small impurities, and yielded 17 (0.65 g, 1.50 mmol, 75 % yield) as a viscous white oil. ¹H NMR (CDCl₃, 400 MHz) δ ppm 3.16 (br. s., 1 H, m), 3.76 (s x 2, 6 H), 3.79 - 3.95 (m, 3 H, j and l), 4.35 (q, J=5.08 Hz, 1 H, k), 4.89 (d, J=5.47 Hz, 1 H, i), 5.04 (s, 2 H, a₅), 6.75 (d, J=8.20 Hz, 1 H, ar.), 6.79 (dd, J=8.20, 1.56 Hz, 1 H, ar.), 6.88 (d, J=8.20 Hz, 1 H, ar.), 6.95 (d, J=1.17 Hz, 1 H, ar.), 7.19 - 7.32 (m, 5 H, ar.), 7.32 - 7.38 (m, 2 H, ar.), 9.73 (s, 1 H, v). ¹³C NMR (CDCl₃, 75 MHz) δC ppm 56.10 (s, 1 C, h/n), 56.12 (s, 1 C, h/n), 61.43 (s, 1 C, k), 71.08 (s, 1 C, a₅), 73.44 (s, 1 C, j), 85.58 (s, 1 C, i), 110.10 (s, 1 C, ar.), 110.21 (s, 1 C, ar.), 113.88 (s, 1 C, ar.), 117.54 (s, 1 C, ar.), 118.70 (s, 1 C, ar.), 126.36 (s, 1 C, ar.), 127.33 (s, 2 C, ar.), 127.98 (s,
(4R*,5S*)-4-[4-(4-Benzylxy-3-methoxyphenyl)-2,2-dimethyl-1,3-dioxan-5-ylxy]-3-methoxybenzaldehyde (18)

To a solution of 17 (0.65 g, 1.50 mmol) in dry acetone (20 mL), anhydrous pTsOH (30 mg, 0.15 mmol) was added at room temperature. After stirring for 1 h, 4 Å molecular sieves (1.5 g) were added and the reaction solution was stirred vigorously for an additional hour, followed by the addition of anhydrous Na$_2$CO$_3$ (0.25 g) to quench acidity. The reaction mixture was then filtered, and the residues washed with EtOAc. The combined organics were dried over Na$_2$SO$_4$, filtered, and the solvent removed in vacuo. The resulting mixture of the product 18 and starting material 17 was separated by flash chromatography on silica gel with a gradient of 30 – 40 % EtOAc in hexanes to give 18 (0.40 g, 0.84 mmol, 57 % yield) as a white solid. $^1$H NMR (CDCl$_3$, 400 MHz) $\delta$H ppm 1.50 (s, 3 H, m$_{1/2}$), 1.61 (s, 3 H, m$_{1/2}$), 3.80 (s, 6 H, h/n), 3.98 (t, $J$=10.20 Hz, 1 H, j), 4.11 - 4.36 (m, 2 H, k), 4.91 (d, $J$=8.98 Hz, 1 H, i), 5.07 (s, 2 H, a$_5$), 6.58 (d, $J$=7.42 Hz, 1 H, ar.), 6.76 (d, $J$=7.03 Hz, 1 H, ar.), 6.94 (d, $J$=8.20 Hz, 1 H, ar.), 7.00 (s, 1 H, ar.), 7.18 (d, $J$=8.20 Hz, 1 H, ar.), 7.22 - 7.45 (m, 6 H, ar.), 9.74 (s, 1 H, u). $^{13}$C NMR (CDCl$_3$, 75 MHz) $\delta$C ppm 19.73 (s, 1 C, m$_{1/2}$), 28.45 (s, 1 C, m$_{1/2}$), 55.95 (s, 1 C), 55.96 (s, 2 C, h/n), 62.44 (s, 1 C, k), 70.90 (s, 1 C, a$_5$), 74.12 (s, 1 C, j), 76.45 (s, 1 C, i), 99.75 (s, 1 C, l), 109.85 (s, 1 C, ar.), 110.87 (s, 1 C, ar.), 113.75 (s, 1 C, ar.), 114.45 (s, 1 C, ar.), 119.39 (s, 1 C, ar.), 126.15 (s, 1 C, ar.), 127.19 (s, 2 C, ar.), 127.84 (s, 1 C, ar.), 128.56 (s, 2 C, ar.), 130.89 (s, 1 C, ar.), 131.86 (s, 1 C, ar.), 137.11 (s, 1 C, ar.), 148.00 (s, 1 C, ar.), 149.46 (s, 1 C, ar.), 150.40 (s, 1 C, ar.), 152.52 (s, 1 C, ar.), 190.77 (s, 1 C, u).
(4R*,5S*)-4-(4-benzyloxy-3-methoxyphenyl)-5-(4-hydroxymethyl-2-methoxyphenoxy)-2,2-dimethyl-1,3-dioxane (90)

18 (0.3284 g, 0.686 mmol) was dissolved in 3:1 THF:H2O (6.6 mL) at room temperature in a round bottom flask. To the solution of 18, NaBH₄ (0.2595 g, 6.86 mmol) was added portion-wise over 6 h with stirring. Water (5 mL) was added after 24 h of stirring to neutralize the remaining NaBH₄. The aqueous-organic mixture was extracted with Et₂O (3 x 20 mL). The combined organic layers were washed with water and then dried over Na₂SO₄. The organic solvent was removed in vacuo. The product 90 was isolated in near quantitative yield as a cloudy white oil and was used with no further purification (0.3279 g, 0.682 mmol, 99%). ¹H NMR (CDCl₃, 400 MHz) δH ppm 1.51 (s, 3 H, m₁₂), 1.63 (s, 3 H, m₁₂), 2.21 (br. s., 1 H, v), 3.70 (s, 3 H, h/n), 3.84 (s, 3 H, h/n), 3.94 - 4.05 (m, 1 H, j), 4.06 - 4.22 (m, 2 H, k), 4.49 (s, 2 H, u), 4.89 (d, J=8.59 Hz, 1 H, i), 5.11 (s, 2 H, a₅), 6.43 (d, J=8.20 Hz, 1 H), 6.62 (dd, J=8.20, 1.95 Hz, 1 H), 6.76 - 6.85 (m, 2 H), 7.00 (dd, J=8.20, 1.95 Hz, 1 H), 7.04 (d, J=1.95 Hz, 1 H), 7.28 (d, J=7.03 Hz, 1 H, a₁), 7.34 (t, J=7.23 Hz, 2 H, a₂), 7.40 (d, J=7.42 Hz, 2 H, a₃). ¹³C NMR δc ppm 19.68 (s, 1 C, m₁₂), 28.50 (s, 1 C, m₁₂), 55.69 (q, 1 C, h/n), 55.93 (q, 1 C, h/n), 62.85 (s, 1 C, k), 64.86 (s, 1 C, u), 70.93 (t, 1 C, a₅), 74.57 (s, 1 C, j), 77.21 (s, 1 C, i), 99.50 (s, 1 C, l), 110.99 (s, 1 C, r), 111.19 (s, 1 C, d), 113.74 (s, 1 C, c), 117.27 (s, 1 C, q), 119.16 (s, 1 C, t), 119.75 (s, 1 C, f), 127.22 (s, 1 C, a₂₃), 127.79 (s, 1 C, a₁), 128.50 (s, 1 C, a₂₃), 132.32 (s, 1 C), 135.67 (s, 1 C), 137.14 (s, 1 C), 146.42 (s, 1 C), 147.84 (s, 1 C), 149.32 (s, 1 C), 150.40 (s, 1 C).
(4R*,5S*)-4-(4-benzylxy-3-methoxyphenyl)-5-[4-(tert-butyldimethylsilyloxymethyl)-2-methoxyphenoxy]-2,2-dimethyl-1,3-dioxane (92)

To a solution of 90 (479.6 mg, 0.999 mmol), and imidazole (0.6008 g, 8.8 mmol) in dry DMF (4 mL), TBDMS-Cl (0.75 g, 5.0 mmol) was added portion-wise over 1 h. After an additional 1 h of stirring, water (10 mL) was added. The mixture was extracted with Et₂O (3 x 20 mL). The organic layers were combined and dried over Na₂SO₄. The organic solvent was removed in vacuo. The resulting oil was placed on under high vacuum and a heat gun was used to heat the vial in order to remove volatile silane impurities. The product (92) was isolated as a viscous cloudy white oil and was used without further purification (520 mg, 0.874 mmol, 87 %).

¹H NMR (CDCl₃, 400 MHz) δ ppm 0.07 (s, 6 H, ν₁), 0.92 (s, 9 H, ν₂), 1.50 (s, 3 H, m₁-2), 1.63 (s, 3 H, m₁-2), 3.72 (s, 3 H, h/n), 3.85 (s, 3 H, h/n), 3.96 - 4.05 (m, 1 H, j), 4.07 - 4.17 (m, 2 H, k), 4.60 (s, 2 H, u), 4.89 (d, J=8.59 Hz, 1 H, i), 5.13 (s, 2 H, a₅), 6.40 (d, J=8.20 Hz, 1 H), 6.61 (dd, J=8.20, 1.95 Hz, 1 H), 6.82 (d, J=8.59 Hz, 2 H), 6.99 (dd, J=8.20, 1.95 Hz, 1 H), 7.05 (d, J=1.56 Hz, 1 H), 7.28 (d, J=7.03 Hz, 1 H, a₁), 7.34 (t, J=7.42 Hz, 2 H, a₂), 7.40 (d, J=7.03 Hz, 2 H, a₃). ¹³C NMR δ ppm -5.13 (s, 1 C, ν₄), 18.48 (s, 2 C, ν₁), 19.76 (s, 1 C, m₁-2), 26.01 (s, 3 C, ν₂), 28.62 (s, 1 C, m₁-2), 55.71 (q, J=2.81 Hz, 1 C, h/n), 56.00 (q, J=2.81 Hz, 1 C, h/n), 63.00 (s, 1 C, k), 64.69 (t, J=3.80 Hz, 1 C, u), 71.03 (t, J=4.60 Hz, 1 C, a₅), 74.63 (s, 1 C, j), 77.54 (s, 1 C, i), 99.48 (s, 1 C, l), 110.21 (s, 1 C, r), 111.21 (s, 1 C, d), 113.86 (s, 1 C, c), 117.48 (s, 1 C, q), 118.14 (s, 1 C, t), 119.78 (s, 1 C, f), 127.25 (s, 3 C, a₂/3), 127.81 (s, 1 C, a₁), 128.56 (s, 2 C, a₂/3), 132.51 (s, 1 C), 136.24 (s, 1 C), 137.28 (s, 1 C), 146.01 (s, 1 C), 147.91 (s, 1 C), 149.46 (s, 1 C), 150.41 (s, 1 C).
(4R*,5S*)-4-(4-hydroxy-3-methoxyphenyl)-5-[4-(tert-butyldimethylsilyloxyethyl)-2-methoxyphenoxy]-2,2-dimethyl-1,3-dioxane (102)

A solution of 92 (520 mg, 0.87 mmol) in EtOH (10 mL) was added to 10 % Pd/C (15 mg) and hydrogenated at room temperature under an atmosphere of H₂ (30 psi) for 2.5 h. The solution was filtered through Celite, and concentrated in vacuo to give a viscous colorless oil. The oil was purified via flash chromatography on silica gel, the product 102 eluting with a gradient of 5-10 % EtOAc in DCM (407.4 mg, 0.80 mmol, 92 %). ¹H NMR (CDCl₃, 300 MHz) δH ppm 0.06 (s, 6 H, v₁), 0.92 (s, 9 H, v₂), 1.51 (s, 3 H, m₁-2), 1.63 (s, 3 H, m₁-2), 3.75 (s, 3 H, h/n), 3.83 (s, 3 H, h/n), 3.92 - 4.24 (m, 3 H, j), 4.60 (s, 2 H, k), 4.89 (d, J=8.50 Hz, 1 H, u), 5.60 (s, 1 H, i), 6.44 (d, J=8.21 Hz, 1 H, a), 6.62 (d, J=7.62 Hz, 1 H), 6.81 (s, 1 H), 6.86 (d, J=8.21 Hz, 1 H), 7.00 (s, 1 H), 7.03 (d, J=8.21 Hz, 1 H). ¹³C NMR δC ppm 5.10 (s, 1 C, v₃), 18.53 (s, 2 C, v₁), 19.80 (s, 1 C, m₁-2), 26.08 (s, 3 C, v₂), 28.70 (s, 1 C, m₁-2), 55.79 (s, 1 C, h/n), 55.99 (s, 1 C, h/n), 63.07 (s, 1 C, k), 64.75 (s, 1 C, u), 74.78 (s, 1 C, j), 77.58 (s, 1 C, i), 99.55 (s, 1 C, l), 110.19 (2 x s, 2 C, d-r), 114.16 (s, 1 C, c), 117.47 (s, 1 C, q), 118.22 (s, 1 C, t), 120.65 (s, 1 C, f), 131.32 (s, 1 C), 136.27 (s, 1 C), 145.55 (s, 1 C), 146.08 (s, 1 C), 146.35 (s, 1 C), 150.45 (s, 1 C)

(4R*,5S*)-4-[4-(tetramethylrhodamine-5(6)-carboxylate ester)-3-methoxyphenyl]-5-[4-(tert-butyldimethylsilyloxyethyl)-2-methoxyphenoxy]-2,2-dimethyl-1,3-dioxane (109)

A solution of DCC (0.4 mg, 2.0 µmol) in dry DCM (0.2 mL) was added to a solution of 102 (10.3 mg, 20 µmol), TAMRA-NHS (3 mg, 6 µmol), and DMAP (0.1 mg, 0.82 µmol) in dry DCM (1 mL). The solution was allowed to reflux for 16 h and then stirred at room temperature for an additional 48 h. The solution was concentrated, and the residue was purified via flash chromatography on silica gel, to yield 109 eluting with acetone (100 mL) (4.6 mg, 5.0 µmol,
83 % from TAMRA-NHS) as a solid pink residue. Product 109 formation was confirmed by ESI-MS. ESI-MS m/z 917.47 (M+H⁺), expected: 916.40 (M), 917.40 (M+H⁺).

(4R*,5S*)-4-[4-(tetramethylrhodamine-5(6)-carboxylate ester)-3-methoxyphenyl]-5-[4-(hydroxymethyl)-2-methoxyphenoxy]-2,2-dimethyl-1,3-dioxane (118)

TBAF (2 drops of 1 M in THF) was added to a solution of 109 (1.25 mg, 1.4 µmol) in dry THF (1 mL) and stirred at 0 °C under an atmosphere of N₂ in the dark for 10 min. The reaction was allowed to warm to room temperature and stirred for another 5 h. Reaction progress was monitored by TLC (10 % MeOH in acetone). The solvent was removed in vacuo, and preparative TLC performed (1:1 MeOH:acetone) on the resulting film. The slower running of two bands (Rf = 0.4) was removed, extracted with MeOH, and filtered through a medium fritted filter to produce 118 (0.49 mg, 0.61 µmol, 44 %) as a pink film. Product formation was confirmed by ESI-MS. ESI-MS m/z 803.2 (M+H⁺), expected: 802.31 (M), 803.31 (M+H⁺).

(4R*,5S*)-4-[4-(tetramethylrhodamine-5(6)-carboxylate ester)-3-methoxyphenyl]-5-[4-(fluorescein-5(6)-carboxylate methylester)-2-methoxyphenoxy]-2,2-dimethyl-1,3-dioxane (122)

DCC (1 mg, 5 µmol) was added to a solution of 118 (0.16 mg, 0.2 µmol), FAM-NHS (3 mg, 6.3 µmol), and DMAP (1 mg, 8.2 µmol) in dry THF (3 mL). The solution was allowed to reflux at 70 °C for 5 h. The solvent was removed in vacuo, and preparatory TLC on silica gel was performed on the resulting film (pure MeOH). The fastest moving pink band was removed, extracted with MeOH, and filtered through a medium fritted filter to produce 122 (< 1 mg) as a pink film. Product formation was confirmed by ESI-MS. ESI-MS m/z 1161 (M+H⁺), 1160 (M⁺), expected: 1160.36 (M), 1161.36 (M+H⁺).
(4R*,5S*)-4-(3-methoxy-4-propargyloxyphenyl)-5-[4-(tert-butyldimethylsilyloxy)methyl]-2-methoxyphenoxy]-2,2-dimethyl-1,3-dioxane (147)

Propargyl bromide (80 wt % in toluene, 0.106 mL, 0.96 mmol) and K$_2$CO$_3$ (0.1673 g, 1.2 mmol) were added to a solution of 102 (407.4 mg, 0.80 mmol) in dry acetone (10 mL). The solution was heated at reflux (60 °C) under nitrogen for 2.5 h. After cooling to room temperature, the reaction mixture was filtered and the solvent removed in vacuo. The dark, brown oil was purified by flash chromatography on silica gel, applied to the column with 10 % DCM in hexanes and eluted with pure DCM, to yield 147 (345.7 mg, 0.637 mmol, 80 %) as an amber colored oil that could not be separated from the starting material. $^1$H NMR (CDCl$_3$, 300 MHz) δH ppm 0.06 (s, 6 H, $v_1$), 0.77 - 1.01 (s, 9 H, $v_2$), 1.50 (s, 3 H, m$_{1/2}$), 1.63 (s, 3 H, m$_{1/2}$), 2.47 (t, $J$=2.30, 1H, a$_1$), 3.75 (s, 3 H, h/n), 3.84 (s, 3 H, h/n), 3.93 - 4.07 (m, 1 H, j), 4.13 (m, 2 H, k), 4.60 (s, 2 H, u), 4.73 (d, $J$=2.35 Hz, 2 H, a$_2$), 4.88 (d, $J$=8.80 Hz, 1 H, i), 6.42 (dd, $J$=8.21, 3.52 Hz, 1 H, ar.), 6.62 (d, $J$=8.21 Hz, 1 H, ar.), 6.72 - 6.92 (m, 2 H, ar.), 6.92 - 7.17 (m, 2 H, ar.)

(4R*,5S*)-4-(3-methoxy-4-propargyloxyphenyl)-5-[4-hydroxymethyl-2-methoxyphenoxy]-2,2-dimethyl-1,3-dioxane (148)

TBAF (1.28 mL of a 1 M solution in THF, 1.28 mmol) was added to a solution of 147 (345.7 mg, 0.637 mmol) in THF (1 mL) and stirred at 0 °C for 10 min. The solution was the allowed to warm to room temperature and allowed to stir overnight. The solvent was removed in vacuo and the resulting oil was purified by flash chromatography. The major silyl-deprotected product, 148, eluted with 20 % EtOAc in DCM. A minor product showed the expected removal of the silyl group as well as the removal of the propargyl moiety. $^1$H NMR (CDCl$_3$, 300 MHz) δH ppm 1.51 (s, 3 H, m$_{1/2}$), 1.63 (s, 3 H, m$_{1/2}$), 2.47 (t, $J$=2.30 Hz, 1 H, a$_1$),
3.76 (s, 3 H, h/n), 3.83 (s, 3 H, h/n), 3.94-4.08 (m, 1 H, j), 4.08 - 4.24 (m, 2 H, k), 4.54 (s, 2 H, u), 4.72 (d, J=2.35 Hz, 2 H, a2), 4.91 (d, J=8.50 Hz, 1 H, i), 6.45 (d, J=8.21 Hz, 1 H), 6.66 (d, J=6.74 Hz, 1 H), 6.82 (s, 1 H), 6.90 - 7.00 (m, 1 H), 7.00 - 7.13 (m, 2 H). 13C NMR δC ppm
19.82 (s, 1 C, m1-2), 28.67 (s, 1 C, m1-2), 55.89 (s, 1 C, h/n), 55.97 (s, 1 C, h/n), 56.94 (s, 1 C, a2), 63.01 (s, 1 C, k), 65.26 (s, 1 C, u), 74.63 (s, 1 C, j), 75.85 (s, 1 C, a1), 77.38 (s, 1 C, i), 78.73 (s, 1 C, a3), 99.64 (s, 1 C, l), 111.18 (2 x s, 2 C, d/r), 114.26 (s, 1 C, c), 117.41 (s, 1 C, q), 119.40 (s, 1 C, t), 119.72 (s, 1 C, f), 133.42 (s, 1 C), 135.65 (s, 1 C), 146.62 (s, 1 C), 146.69 (s, 1 C), 149.53 (s, 1 C), 150.60 (s, 1 C).

(4R*,5S*)-4-[4-(4-methoxy-1-[tetramethylrhodamine-5-carboxamide-(N)-propyl]-1,2,3-triazol)-3-methoxyphenyl]-5-[4-hydroxymethyl-2-methoxyphenoxy]-2,2-dimethyl-1,3-dioxane (149)

To a solution of 148 (5 mg, 12 µmol) in 3:1 THF:H2O (1 mL), TAMRA-N3 (1 mg, 20 µmol), CuSO4•5H2O (0.6 mg, 2.4 µmol), and sodium ascorbate (0.9 mg, 5 µmol) were added and stirred at room temperature under an atmosphere of N2 overnight in the dark. The solvent was removed in vacuo, the residue re-dissolved in acetone, and filtered through a paper disk. The product was further purified by preparative TLC on silica gel plates using pure methanol as a mobile phase. All resulting bands were removed from the plate, the adsorbed product was removed from the silica gel using methanol, filtered through a fine glass frit and further rinsed with methanol. The resulting residues, formed after evaporation of methanol from the samples, were analyzed by ESI-MS for the presence of product. Compound 149 was identified as a solid pink film (< 1 mg), and formation was confirmed by ESI-MS. ESI-MS m/z 941 (M+), expected: 940.40 (M), 941.40 (M+H+).
(4R*,5S*)-4-[4-(4-methyloxy-1-[tetramethylrhodamine-5-carboxamide-(N)-propyl]-[1,2,3]triazol)-3-methoxyphenyl]-5-[4-propargyloxymethyl-2-methoxyphenoxy]-2,2-dimethyl-1,3-dioxane (150)

To a solution of 149 (< 1 mg) in acetone (1 mL), K$_2$CO$_3$ and propargyl bromide (80 wt % in toluene, 20 µL, 0.2 mmol) were added under a flow of N$_2$ gas. The mixture was refluxed for 2.5 h at 64 °C in the dark. The solids were removed via filtration, the flow through collected and the solvent evaporated. The product was further purified by preparative TLC on silica gel plates using pure methanol as a mobile phase. All resulting bands were removed from the plate, the adsorbed product was removed from the silica gel using methanol, filtered through a fine glass frit and further rinsed with methanol. The resulting residues, formed after evaporation of methanol from the samples, were analyzed by ESI-MS for the presence of product. Product formation, 150, was confirmed by ESI-MS; starting material (149, $m/z$ 941) was not observed after purification. ESI-MS $m/z$ 979.10 (M+H$^+$), expected: 978.42 (M), 979.40 (M+H$^+$).

(4R*,5S*)-4-(4-benzyloxy-3-methoxyphenyl)-5-[4-(fluorescein-5(6)-carboxylate methylester)-2-methoxyphenoxy]-2,2-dimethyl-1,3-dioxane (111)

FAM-NHS (3 mg, 6.3 µmol) was added to a solution of 90 (20 mg, 41 µmol), DCC (1 mg, 5 µmol), and DMAP (1 mg, 8 µmol) in THF (2.5 mL). The solution was refluxed for 5 h at 70 °C in the dark. The solvent was removed in vacuo and the organic soluble products were redissolved in 1:1 CHCl$_3$:acetone; unreacted FAM-NHS remained as an insoluble film. The organic soluble fraction was purified by flash chromatography on silica gel and eluted with pure acetone (100 mL) to yield 111 as a solid orange film (2.1 mg, 2.5 µmol, 40 % from FAM-NHS). Product formation was confirmed by ESI-MS. ESI-MS $m/z$ 837.3 (M–H$^+$), expected: 838.26 (M), 837.26 (M–H$^+$).
(1R*,2S*)-1-(4-benzyloxy-3-methoxyphenyl)-2-[4-(fluorescein-5(6)-carboxylate methylester)-2-methoxyphenoxy]propane-1,3-diol (115)

PPTS (1.4 mg, 5.6 µmol) was added to a solution of 111 (1.1 mg, 1.3 µmol) in 1:4 H2O:acetone (2.5 mL). The solution was refluxed for 17 h at 50 °C in the dark. The solvent was removed in vacuo, and the residue washed with CHCl₃ (3 x 3 mL). The organic layers were combined and the solvent removed in vacuo. Preparatory TLC on silica gel was performed on the resulting organic soluble film. The product 115 (Rf = 0.47) was removed from the plate, extracted from the silica gel with MeOH, and filtered through a medium fritted filter. Product formation was confirmed by ESI-MS. ESI-MS m/z 797.2 (M–H⁺), expected: 798.23 (M), 797.23 (M–H⁺).
Chapter 6:

Conclusions and Future Directions
6.1. Conclusions and Future Directions

The ring-cleaving extradiol dioxygenase LigAB, has been the subject of molecular biology, biochemical, and metabolic pathway research since 1987 – twenty years prior to our discovery of this enzyme in the literature. Despite this, the extent to which LigAB had been kinetically and functionally characterized relative to other ring-cleaving dioxygenase enzymes was very low. Now, however, the wild type activity of LigAB has been fully characterized since we have determined kinetic parameters for the dioxygenation of the native in-pathway substrates PCA, gallate, and 3OMG, and established the importance of anaerobic purification of this enzyme. Additionally, the scope of LigAB’s substrate promiscuity has been better defined by the identification of additional substrates and inhibitors. Analysis of the substrate structures suggests that LigAB is highly specific for diol containing substrates with oxidation potentials between 5.40-6.30 eV.

Despite this, substrate-like molecules meeting these criteria were also identified as either non-substrates or inhibitors, suggesting that active site residues further modulate the catalysis of substrates through hydrogen bonding or steric. To support this, we have shown that mutations of Phe103α modify LigAB’s catalytic profile such that previously poor substrates, for example 3OMG, are utilized more efficiently. To expand upon the findings in this dissertation, additional mutation studies will be needed to determine the catalytic mechanism of LigAB as well as further define the enzyme structural parameters that lead to substrate selection. Many of the mutants to complete these studies have already been designed, and constructed. Mutations to investigate the mechanism of LigAB include the catalytic base and acid (His127β to Ala, Asp, and Asn, and His195β to Ala, and Gln), and the iron binding residue Glu242β (Ala, and Gln). Structural control over substrate binding will be further investigated utilizing mutations S169β to Gln (additional mutants to Val, Lue, Ile, Asp, and Asn should also be considered), and T271β to Ala and Ser.
In addition to the identification of substrates and inhibitors, allosteric activators of LigAB catalysis were also identified during competition studies between known substrates and substrate-like molecules. Significantly, vanillin was found to be among the activators, and we hypothesize that the enhancement of substrate catalysis is the result of feed-forward allosterism – a rarely observed phenomenon, and not previously reported in dioxygenases. The rate enhancement effected by vanillin in the wild-type LigAB was found to be highly substrate dependent – present only during the catalysis of the native substrate PCA. However, mutations of Phe103α showed that this activation specificity could be altered to allow for activation of gallate dioxygenation, removal of the activation effect entirely, or inhibition of catalysis dependent on which substrate was presented to the enzyme. The presence of this phenomena in LigAB is exciting and could have far reaching implications for the selection of enzymes used in engineered microorganisms for the bio-production of chemicals. To better define the allosteric activation of LigAB, further mutation studies of residues contributing to the allosteric pocket (ex. Glu86β to Asp, Gln, Asn, and/or Ala) will be needed to understand how the allosteric signal is transmitted to the active site as well as to determine which step(s) of the catalytic mechanism is(are) impacted by the allosteric binding of vanillin.

While effective degradation of lignin derived aromatic compounds into central metabolites is critical for the bio-utilization of lignin, efficient lignin depolymerization must first occur. Lignin’s chemical and biochemical recalcitrance must be overcome by either highly efficient chemical processes or through the discovery of more efficient enzymes. We have described the design of lignin mimetics modified with a FRET-pair of fluorophores utilizing the lignin β-O-4 bond motif for use in the discovery of such enzymes. Although the synthesis of these probes is not yet complete, hampered by unsuccessful reaction steps near pathway’s end, several precursors have been successfully synthesized with one or both of the fluorophores
linked to the lignin dimer core. Multiple avenues by which to reroute or rework the synthesis of the FRET-labeled lignin mimetic will continue to be pursued by members of the Taylor lab. Mackenzie Schlosser is currently attempting to modify reaction and solvent conditions for the click-linkage of the FAM-N$_3$ fluorophore to the lignin dimer core structure, working under a hypothesis that steric interference due hydrophobic collapse of the TAMRA-lignin dimer complex under the solvent conditions that have been attempted may be preventing forward progress of the reaction. Changing the solvent and other reaction conditions (i.e. addition of TBTA) will hopefully lead to the successful completion of the labeled lignin model compound.

Production of fuels and chemicals from under-utilized carbon sources such as lignin will become increasingly important as petroleum sources dwindle, becomes more difficult and costly to extract in the near future, and public opinion shifts due to global climate change. Efficient bio-utilization of alternative carbon feedstocks will require microorganisms engineered to utilize a diverse range of carbon feedstock molecules, including aromatics. Dioxygenases will therefore play a critical role in the facile linearization of aromatic molecules so that they can be used in central metabolism. A complete understanding of how LigAB and other closely related dioxygenase enzymes (DesB, DesZ, and LigZ also from the Sphingobium sp. SYK-6 LDAC degradation pathway) employ enzyme structural composition to select substrates and enhance catalysis will help in the future design of dioxygenases. Dioxygenases such as these that are capable of accepting a wide range of substrates will be necessary for the efficient utilization of myriad aromatic compounds derived from lignin depolymerization. As well, successful synthesis of a FRET-pair labeled lignin model compound may help identify new and more efficient sources of enzymatic lignin degradation. These lignin depolymerizing enzymes could also be incorporated, resulting in engineered microorganisms capable of feedstock-to-product production of chemicals and fuels from lignin.
References


to related biochemical processes in allyl/propenyl phenol and lignan biosynthesis, *Natural product reports* 25, 1015-1090.


178

Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.;
Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman,
J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.


methods based on continuous distributions of the solvent, Chem. Rev. (Washington,
DC, U.S.) 94, 2027-2094.

potentials of polycyclic aromatic hydrocarbons are highly linearly correlated over a

ketosteroid 9a-hydroxylase, a Rieske-oxygenase in the cholesterol degradation
pathway of Mycobacterium tuberculosis, J. Biol. Chem. 284, 9937-9946.

Molecular basis for the stabilization and inhibition of 2,3-dihydroxybiphenyl 1,2-
dioxygenase by t-butanol, J. Biol. Chem. 273, 34887-34895.

L., Mekmouche, Y., Spiro, T. G., Lipscomb, J. D., Blades, M. W., Turner, R. F. B.,
and Eltis, L. D. (2005) Spectroscopic studies of the anaerobic enzyme - Substrate
complex of catechol 1,2-dioxygenase, J. Am. Chem. Soc. 127, 16882-16891.

Swapping metals in Fe- and Mn-dependent dioxygenases: Evidence for oxygen
activation without a change in metal redox state, Proc. Natl. Acad. Sci. U.S.A. 105,
7347-7352.

cleaving homoprotocatechuate 2,3-dioxygenase into an intradiol cleaving enzyme, J.

[159] Seibert, C. M., and Raushel, F. M. (2005) Structural and Catalytic Diversity within the
Amidohydrolase Superfamily, Biochemistry 44, 6383-6391.

Deacylation of d-Amino Acids by Members of the Amidohydrolase Superfamily,
Biochemistry 48, 6469-6481.

protocatechuate 4,5-dioxygenase, Biochim. Biophys. Acta 220, 224-238.

Manganese(II)-dependent extradiol-cleaving catechol dioxygenase from Arthrobacter


from the thermoacidophilic archaeon Picrophilus torridus, *Extremophiles* 18, 641-651.


Appendix
Figure A.1. Representation of the process to calculate $j_s$ values used in the analysis of mechanistic inactivation of LigAB and mutants. A) Raw data collected in the Oxygraph user interface showing the progress curve for oxygen consumption upon enzyme addition. B) The raw data is extracted from the point of enzyme addition to the end of data collection, and C) modified to reflect product formation. D) The modified data is then plotted in KaleidaGraph and fit to the exponential equation below where $j_s$ is taken as the value determined as the variable m3.

$$P_t = P_\infty (1 - e^{-j_s t}) + P_i$$
Figure A.2. Multiple sequence alignment of representative sequences of the LigB, MhpB-like, and HpaD-like PCADSF families. Figure caption continued on the following page.
**Figure A.2. (continued)** Multiple sequence alignment of representative sequences of the LigB, MhpB-like, and HpaD-like PCADSF families. Secondary structure provided at the top of the alignment, assigned by comparison with the 1B4U structure; the primary sequence of the 1B4U structure is provided as the first row in the alignment. Sequences are labeled to the left with gene name, organism abbreviation, and ncbi gene identifier (gi) number; families are specified at the right of the alignment. Numbers bookending sequences represent positions of the domain within the sequence. Numbers within alignment represent the number of amino acids excised in regions of poor conservation. Numbers at the top of the alignment denote the two variable regions (VR1/VR2) predicted to generally function in allosteric activation across the PCADSF. The alignment consensus across all three superfamilies is calculated at the 90% level, consensus values are calculated for all sequences in each family and not just the sequences selected for this alignment; therefore, values assigned to each column may reflect the diversity observed across all families. Consensus coloring and abbreviation coding as follows: h, hydrophobic shaded in yellow; p, polar in cyan; s/u, small/tiny in green; l, aliphatic in yellow; o, hydroxylic in orange; a, aromatic in yellow; b, big in gray; c/- charged/negatively-charged in pink. Absolutely conserved residues with predicted roles in active site catalysis are shaded in red and colored in white. Residue positions displaying lineage-specific consensus patterns are shaded in dark blue (greater than 85% consensus) or light blue (80-85% consensus) with the consensus specific to each subfamily provided below each variable region. Residues of interest in variable regions specifically mentioned in the main text are denoted with red asterisks above the relevant alignment column on the secondary structure line. Organism abbreviations as follows: Abau: Acinetobacter baumannii; Aboh: Acinetobacter bohemicus; Acor: Amorphus coralli; Afae: Alcaligenes faealci; Ajap: Amphitricha japonica; Avsp.: Acidovorax sp.; Absp.: Acinetobacter sp.; Bbac: Burkholderiales bacterium; Bbat: Bacillus bataviensis; Bbry: Burkholderia bryophila; Been: Burkholderia cenocepacia; Brad: Bradyrhizobium sp.; Bsub: Brevundimonas subvibrioides; Burk: Burkholderia sp.; Carg: Corynebacteriu argenteratense; Cdef: Castellaniella defragrans; Cdpi: Corynebacterium diptheriae; Cmsp.: Citromicrobium sp.; Cvsp.: Cupriavidus sp.; Ctes: Comamonas testosteroni; Daro: Dechloromonas aromatic; Dsp.: Dietzia sp.; Enor: Enterovibrio norvegicus; Fsp.: Frankia sp.; Gpro: gamma proteobacterium; Grub: Gordonia rubripertincta; Gsp.: Gordonia sp.; Hbal: Hirschia baltica; Hsp.: Herbaspirillum sp.; Lagg: Labrenzia aggregata; Lsp.: Limnohabitans sp.; Malk: Maritimibacter alkaliphilus; Mhas: Mycobacterium hassiaca; Mmag: Magnetospirillum magneticum; Mrhi: Marinobacterium rhizophilum; Msan: Marinobacter santoriniensis; Mspr.: Magnetospira sp.; Msta: Marinobacterium staniertii; Nosp.: Nocardia sp.; Osp.: Oceanicola sp.; Paer: Pseudomonas aeruginosa; Pflu: Pseudomonas fluorescens; Pgil: Polymorphism gilvum; Pmul: Pasteurella multocida; Pnoe: Pusillimonas noertemannii; Ptem: Photorhabdus temperata; R: Roseomonas; Rbac: Rhodobacteraceae bacterium; Ropa: Rhodococcus opacus; Rorn: Raoultella ornithinolytica; Rqin: Rhodococcus qingshengii; Rspp.: Rhodobacter sphaeroide; Rwra: Rhodococcus wratislaviensis; Samb: Streptomyces ambofaciens; Sesp.: Saccharothrix espanaensis; Sful: Sphingobium fuliginis; Sgrl: Streptomycyes griseoauranticus; Spau: Sphingomonas paucimobilis; Sros: Streplosporangium roseum; Ssp.: Serratia sp.; Ssp.: Sphingobium sp.; Ssul: Streptomyces sulphureus; Tdis: Thiothrix disciformis; Tsp.: Terrabacter sp.; Umar: uncultured marine; Vpar: Variovorax paradoxus; Vvul: Vibrio vulnificus; Xbov: Xenorhabdus bovienii.
Figure A.3. Comparison of the intrinsic tryptophan fluorescence emission spectra of LigAB-WT to that of the LigAB-A18W mutant enzyme bound to either PCA or vanillin. (A) Emission spectra of LigAB (red line) and A18W (blue line) in the absence of organic substrates are compared with the respective complexes with PCA is bound (red and blue dotted lines). (B) Zoom in of the fluorescence emission maxima to ease observation of the difference in blue shifts of the apo and PCA-complexed enzymes. (C) Emission spectra of LigAB (red line) and A18W (blue line) in the absence of organic substrates are compared with the respective complexes with vanillin is bound (red and blue dotted lines). (D) Zoom in of the fluorescence emission maxima to ease observation of the difference in blue shifts of the apo and vanillin-complexed enzymes.
Figure A.4. FPLC chromatograph showing the elution of LigAB from a Ni-NTA agarose resin using a linear gradient of imidazole. Purification of LigAB by FPLC was eventually abandoned due to the need to purify LigAB anaerobically.
Figure A.5. SDS-PAGE analysis of fractions obtained from the purification of LigAB via an Ni-NTA resin column
Figure A.6. High resolution mass spectrum of LigAB. Theoretical molecular weights: LigA - 17,712 Da; LigB – 33,292 Da; Dimer – 50,986 Da. Theoretical molecular weights computed using the ExPASy ProtParam webtool (http://web.expasy.org/cgi-bin/protparam/protparam)
Figure A.7. High resolution mass spectrum of LigAB. Theoretical molecular weights: LigA - 17,712 Da; LigB – 33,292 Da; Dimer – 50,986 Da. Theoretical molecular weights computed using the ExPASy ProtParam webtool (http://web.expasy.org/cgi-bin/protparam/protparam)
Figure A.8. High resolution mass spectrum of A) LigA and B) LigB. Theoretical molecular weights: LigA - 17,711.8 Da; LigB – 33,292 Da; Dimer – 50,986 Da. Theoretical molecular weights computed using the ExPASy ProtParam webtool (http://web.expasy.org/cgi-bin/protparam/protparam)
Figure A.9. SDS-PAGE analysis of elution fractions obtained from a Ni-NTA column purification of LigAB using a gravity column with N\textsubscript{2} back pressure. A step gradient of imidazole (yellow) was used to elute the protein from the resin.

Figure A.10. SDS-PAGE analysis of thrombin cleavage of the His\textsubscript{6}-tag from a sample of LigAB using 1, 2, or 3 units of thrombin.
**Figure A.11.** SDS-PAGE analysis of the time dependent thrombin cleavage of the His6 tag from LigA with 2 units of thrombin.

**Figure A.12.** SDS-PAGE analysis of eluted fractions obtained from the anaerobic Ni-affinity purification of LigAB using a gravity column with N₂ back pressure. Crude (C), pellet (P), wash (W), and elution (E) fractions are shown.
Figure A.13. SDS-PAGE analysis of eluted fractions obtained from the anaerobic Source-Q purification of LigAB after thrombin cleavage. Fractions containing LigA and LigB were combined.

Figure A.14. SDS-PAGE analysis of LigAB after each step of the full anaerobic purification. A) Combined nickel affinity column fractions. B) LigAB after thrombin cleavage. C) Combined fractions after Source-Q purification.
Figure A.15. SDS-PAGE analysis of eluted fractions obtained from the anaerobic Ni-affinity purification of F103L using a gravity column with N$_2$ back pressure. Crude (C), pellet (P), bind (B), wash (W), and elution (E) fractions are shown.
Figure A.16. SDS-PAGE analysis of eluted fractions obtained from the anaerobic Source-Q purification of LigAB after thrombin cleavage. Fractions containing the α and β subunits of F103L were combined.
Figure A.17. SDS-PAGE analysis of eluted fractions obtained from the anaerobic Ni-affinity purification of F103V using a gravity column with N$_2$ back pressure. Crude (C), pellet (P), bind (B), wash (W), and elution (E) fractions are shown.
Figure A.18. SDS-PAGE analysis of eluted fractions obtained from the anaerobic Source-Q purification of LigAB after thrombin cleavage. Fractions containing the α and β subunits of F103L were combined.
Figure A.19. SDS-PAGE analysis of eluted fractions obtained from the anaerobic Ni-affinity purification of F103S using a gravity column with N₂ back pressure. Crude (C), pellet (P), bind (B), wash (W), and elution (E) fractions are shown. Co-purification of the a and b subunits was not observed.
**Protein Sequences**

**LigA (native sequence) (15548 Da)**

MTEKKERIDVHAYLAEFDDIPGTRVFTAQRARKGYNLNQFAMSMLKAENRERFKADESAYLD
EWNLTPAAKAAVLARDYNAMIDEGGNVYFLSKLFSTDGKSFQFAAGSMTGMTQEEYAQMMID
GGRSPAGVRSIKGY

**LigA with His-tag (red) (17711 Da)**

MGSSHHHHHHSSGLVPRGSHMTEKKERIDVHAYLAEFDDIPGTRVFTAQRARKGYNLNQFAM
SLMKAENRERFKADESAYLDEWNLTPAAKAAVLARDYNAMIDEGGNVYFLSKLFSTDGKSFQ
FAAGSMTGMTQEEYAQMMIDGGRSPAGVRSIKGY

**LigB (native sequence) (33292 Da)**

MARVTTGITSHIPALGAAIQTGTSNDNYWGPVFKGYQPIRDVIQPQGNNMPDVVILVYNHDA
SAFDMMIPTFAIGCAETFPADEGWGPRVPVDPVKHPDLAWHIAQSLILDEFDMTNQMD
VDHGCTVPLSIMIGEPEEWPCKVIPFVTVPSSGKRCFLGDSIRAAVESFEDLNVH
VWGTGGMSHQLGFRAGLKEFDLNFIDKLISDPEELSKMPHIQLRESGSEGVELVMWLI
MRGALPEKXDLYTFYHIPASNTALGMLQPEETAGTPLEPRKVMGHSLQA
DNA Sequences and Sequencing Results

LigAB Gene Pair

**LigA:**

ATGACCGAGAAGAAAGAGAATCGACGTTCACGCCTATCTCGCCGAGTTTGACGACATTCCCGGCACCC
GCGTGTTCACCGCCAGGCAGGGCTATAATCTCAACCAGTTCGCGATGAGCCTGATGAAGGC
CGAGAACCGCGACTGCTGCTCCGCCAATGCGATGATCGACGAGGGCGGGAATGTCTATTTCCTGTCCA
AGCTGTTCTCGACCGACGGCAAGAGCTTCCAGTTCGCCGCCGGCTCGATGACCGGCATGACTCAGGAAGA
ATATGCACAGATGATGATCGATGGCGGCCGTTCGCCCGCGGGTGTGCGCTCGATCAAGGGAGGCTACTGAG

**LigB:**

ATGGACCGAGAAGAAAGAGAATCGACGTTCACGCCTATCTCGCCGAGTTTGACGACATTCCCGGCACCC
GCGTGTTCACCGCCAGGCAGGGCTATAATCTCAACCAGTTCGCGATGAGCCTGATGAAGGC
CGAGAACCGCGACTGCTGCTCCGCCAATGCGATGATCGACGAGGGCGGGAATGTCTATTTCCTGTCCA
AGCTGTTCTCGACCGACGGCAAGAGCTTCCAGTTCGCCGCCGGCTCGATGACCGGCATGACTCAGGAAGA
ATATGCACAGATGATGATCGATGGCGGCCGTTCGCCCGCGGGTGTGCGCTCGATCAAGGGAGGCTACTGAG
**Synthetic Gene Constructs**

For all constructs the following coloring scheme applies:

- **Upstream XbaI site**;
- **Ribosome binding site**; **NcoI site**; **thrombin cleavable His$_6$ tag**; **NdeI site**; **stop codons**; **BamHI site**

**LigAB:**

```plaintext
LctagdAAGAACAGGAGAATAACCCATGGGACGCTTCTCATACCCACCACCATCCACACTCTAGGGGGCTG
GGCCCGAGGGAGGATCATGACCGGAAGAACGGATCTCCACCGCGAGGGCCATCCGGACCTTGCACTGACCATCATGAGG
TCATGGCTGCACCGTGCCGCTCTCGATGATCTTCGGCGAGCCCGAGGAATGGCCGTGCAAGGTCATCCCTTCCCGGTCAATGTCGTCACTTATCCGCCGCCGTCGGGCAAGCGCTGCTTCGCGCTCGGTGACAGCATTCCGCGCCGCGGTCGAGAGCTTCCCGGAAGACCTCAACGTCCATGTCTGGGGCACCGGCGGCATGAGCCACCAGCTTCAGGGCCCGCGC
GCCGGCCTCATCAACAAGGAGTTCGACCTGAACTTCATCGACAAGC
TGATCAGCGACCCCGAGGAGCTGAGCAAGATGCCGCACATCCAGTATCTGCGCGAAAGCGGATCGGAAGGGCGTCGAGCTGGTCATGTGGCTCATCATGCGCGGCGCGCTGCCGGAGAAGGTGCGGGATCTCTACACCTTCTATCACATCCCGGCCTCCAACACCGCGCTCGGCGCGATGATCCTGCAGCCGGAGGAGACCGCAGGTACACCGCTCGAACCGCGCAAGGTGATGAGCGGACACAGCCTGGCCCAGGCC
TAATA
GATCC
```
DesB:

DesZ:
**LigC:**

```
<table>
<thead>
<tr>
<th>LigC:</th>
<th>tctagaAAGAACAGGAGGAATTAACCATGGCCAGCTTCATCACCACCATCCACCACCTCTAGCCGGCTG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GTCCCCGGAGGGAGTCATATGAGAATTAGACCTGACGCGGCGACCACCCATCCACCTCTAGCCGGCTG</td>
</tr>
<tr>
<td></td>
<td>CCGTCTCAAAGATATATCAGCGGGCATCAGTGTCCGTCTCAGCCGCTCCAGCGGAGGCCGAGACCAC</td>
</tr>
<tr>
<td></td>
<td>AAGTCCGAGGAAATATGGCCGAGACCAAGGGCACCCGACCGGATTTTCAAGCGCTTGCCTGCCGAGAT</td>
</tr>
<tr>
<td></td>
<td>GTGCCAGGCTGATCTCTGTGACCGGCGAGAAGGCCGAGCGGAGATGGATCTGGAAGCTGCTAGGATA</td>
</tr>
<tr>
<td></td>
<td>GCCGCAACAGCATGCAAGGGCAAGGGCCGCTGGACGGGACCATCTGCT</td>
</tr>
</tbody>
</table>
```

**LigK:**

```
<table>
<thead>
<tr>
<th>LigK:</th>
<th>tctagaAAGAACAGGAGGAATTAACCATGGCCAGCTTCATCACCACCATCCACCACCTCTAGCCGGCTG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GTCCCCGGAGGGAGTCATATGAGAATTAGACCTGACGCGGCGACCACCCATCCACCTCTAGCCGGCTG</td>
</tr>
<tr>
<td></td>
<td>CCGTCTCAAAGATATATCAGCGGGCATCAGTGTCCGTCTCAGCCGCTCCAGCGGAGGCCGAGACCAC</td>
</tr>
<tr>
<td></td>
<td>AAGTCCGAGGAAATATGGCCGAGACCAAGGGCACCCGACCGGATTTTCAAGCGCTTGCCTGCCGAGAT</td>
</tr>
<tr>
<td></td>
<td>GTGCCAGGCTGATCTCTGTGACCGGCGAGAAGGCCGAGCGGAGATGGATCTGGAAGCTGCTAGGATA</td>
</tr>
<tr>
<td></td>
<td>GCCGCAACAGCATGCAAGGGCAAGGGCCGCTGGACGGGACCATCTGCT</td>
</tr>
</tbody>
</table>
```
LigM:

```
ctagAAGAACGAGGGAATTAACCTAGGCAGCTTCACATCACCCAGCTACCCATCTTAGCCGGGCTG
GTCCCGAGAGGAGCAAGTATGTGGCTGACCTCTACATCAATCCGGTGACGGGAGCTG
CTGCCCCGAGGGAGTGATGTGGCCGAAATCGTGCTGGGTATGTGGACGTCGCACGGGCCGACCCTGTC
CACCACGCCGGAGGCAATCCGGTGACGGGAGCTG
```

LigZ:

```
ctagAAGAACGAGGGAATTAACCTAGGCAGCTTCACATCACCCAGCTACCCATCTTAGCCGGGCTG
GTCCCGAGAGGAGCAAGTATGTGGCTGACCTCTACATCAATCCGGTGACGGGAGCTG
CTGCCCCGAGGGAGTGATGTGGCCGAAATCGTGCTGGGTATGTGGACGTCGCACGGGCCGACCCTGTC
CACCACGCCGGAGGCAATCCGGTGACGGGAGCTG
```

209
Sequencing results for F103V

LigAB Gene Pair
F103V Sequencing Data
TAATTCCTCCCTAGAAATAATTGGTTTTAACATTGAAGAGGATAGTACGACATTCACAGCGAGCGCAGCCAGC

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG

LigAB Gene Pair
F103V Sequencing Data
ATGACCGAGAAG
Sequencing Results for F103L

F103L Seq.  CCTTTCCCCTAGAAATTTGGTTTAACTTTAAGAGGAGATATACCATGGGCAGCAG
LigAB Gene Pair  

F103L Seq.  CCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCAT
LigAB Gene Pair  

F103L Seq.  GAAAGAGAGAACTCGACGTTCACGCCTATCTCGCCGAGTTTGACGACATTCCCGGCACCCG
LigAB Gene Pair  

F103L Seq.  CGTGTTCACCCGCCACGCAGCGGCCTGGTGCCGCGCGGCAGCCAT
LigAB Gene Pair  

F103L Seq.  GATGAAGGCCGAGAACCGCGAGCGGTTCAAGGCCGACGAGAGCGCCTATCTGGACGAGTG
LigAB Gene Pair  

F103L Seq.  GAACCTCACGCCGCCACAGGCCGCCGTGCTGGCCCGCGACTACAATGCGATGATCGA
LigAB Gene Pair  

F103L Seq.  CGAGGGCGGGAATGTCTATTTCCTGTCCAAGCTGTTCTCGACCGACGGCAAGAGC
LigAB Gene Pair  

F103L Seq.  GTTCGCCGCCGGCTCGATGACCGGCATGACTCAGGAAGAATATGCACAGATGATGATCGA
LigAB Gene Pair  

F103L Seq.  TGGCGGCCGTTCGCCCGCGGGTGTGCGCTCGATCAAGGGAGGCTACTGATATGGCACGCG
LigAB Gene Pair  

F103L Seq.  TTACCACGGGCATACGTACGCAGCACATTCCCCGCCGCTGGCCCGCGCCCATACAGCAGCGCA
LigAB Gene Pair  

F103L Seq.  CCGACGACAATATGATCTCCTGGCCTTGGCAAGGACATGGCAGTC-TGAGCTTCTGGCTTATAACGACCATGTGGCTCG
LigAB Gene Pair  

F103L Seq.  CCGACGACAATATGATCTCCTGGCCTTGGCAAGGACATGGCAGTC-TGAGCTTCTGGCTTATAACGACCATGTGGCTCG
LigAB Gene Pair  

F103L Seq.  CCGACCAGGGAGCGCCCGCCGCCCCC--
LigAB Gene Pair  

***** ***** ***** *
Sequencing Results for F103A

F103A Seq. Data: TCTAGAAATTATTTATTTATTTATTTAAGAGGAGATATACCATGTCGGCACTCCACCTCATC
LigAB Gene Pair: --------------------------------------------------

F103A Seq. Data: ATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATGCAGCCGCTGTTCA
LigAB Gene Pair: ATGACCGAGAAGAAAGAGA

F103A Seq. Data: ATGACCGAGAAGAAAGAGA
LigAB Gene Pair: ----------------------------

F103A Seq. Data: GAATCGACGTTCACGCCTATCTCGCCGAGTTTGACGACATTCCCGGCACCCGCGTGTTCA
LigAB Gene Pair: GAATCGACGTTCACGCCTATCTCGCCGAGTTTGACGACATTCCCGGCACCCGCGTGTTCA

F103A Seq. Data: CCAGAGAACCGCGAGGTTCAAGGCCGACGAGAGCGCCTATCTGGACGAGTGGAACCTCA
LigAB Gene Pair: CCAGAGAACCGCGAGGTTCAAGGCCGACGAGAGCGCCTATCTGGACGAGTGGAACCTCA

F103A Seq. Data: CGCCCGCCGCCAAGGCCGCCGTGCTGGCCCGCGACTACAATGCGATGATCGACGAGGGCG
LigAB Gene Pair: CGCCCGCCGCCAAGGCCGCCGTGCTGGCCCGCGACTACAATGCGATGATCGACGAGGGCG

F103A Seq. Data: GGAATGTCTATTTCCTGTCCAAGCTGTTCTCGACCGACGGCAAGAGC
LigAB Gene Pair: GGAATGTCTATTTCCTGTCCAAGCTGTTCTCGACCGACGGCAAGAGC

F103A Seq. Data: GTTCGCCCGCGGGTGCTCAGCTCAGGAAGAATATGCACAGATGATGATCGATGGCGGCC
LigAB Gene Pair: GTTCGCCCGCGGGTGCTCAGCTCAGGAAGAATATGCACAGATGATGATCGATGGCGGCC

F103A Seq. Data: GGGATGGGGCCCGCGCCCG----------------------------------------
LigAB Gene Pair: GGGATGGGGCCCGCGCCCGGTGCCCGACGTGAAGGGCCATCCGGACCTTGCCTGGCAC

F103A Seq. Data: TCTAGAAATTATTTATTTATTTATTTAAGAGGAGATATACCATGTCGGCACTCCACCTCATC
LigAB Gene Pair: --------------------------------------------------
A18W Sequencing Data

LigAB Gene Pair  
A18W-C3  
TTCCCTTAGAATTTTTGTTTAACTTTAAGAAGGAGAGATAACATGAGGCGAAGCCACATC

LigAB Gene Pair  
A18W-C3  
ATCATCATCATCCAGCAGCGCGCTGGTGCCGCAGCAGCCTATGACCGGAGAAGAA

LigAB Gene Pair  
A18W-C3  
TCACCGCCAGCGCGCGAGCGCTGCTGGCCGCGCGAGCGCTGCTGGCCGCGCACTAACTCAGGATGATCAGAGG

LigAB Gene Pair  
A18W-C3  
GCCGTTCGCCCGCGGGTGCGCTCGATCAAGGGAGGCTACTGATATGGCACGCGTTACC

LigAB Gene Pair  
A18W-C3  
AGCGGCATCACGTCCAGCCACATTCCCGCGCTCGGC

LigAB Gene Pair  
A18W-C3  
GACAATGATTACTGGGGCCCGGTGTTCAAGGGCTACCAGCCGATCCGCGACTGGATCAAG

LigAB Gene Pair  
A18W-C3  
GATCATGGCTGCACCGTGCCGCTCTCGATGATCTTCGGCGAGCCCGAGGAATGGCCGTGC
Scheme A.1. Proposed alkyne-azide click-chemistry synthesis of a fluorophore labeled lignin model compound.
Synthesis NMR and Mass Spectral Data

Figure A.20. Full $^1$H NMR spectrum of compound 90 (0 – 8 ppm).
Figure A.21. $^1$H NMR spectrum of compound 90 (6 – 8 ppm).
Figure A.22. $^1$H NMR spectrum of compound 90 (3.5 – 5.25 ppm).
Figure A.23. $^{13}$C NMR spectrum of compound 90 (-10 – 200 ppm).
Figure A.24. Full $^{13}$C NMR spectrum of compound 90 (95 - 160 ppm).
Figure A.25. $^{13}$C NMR spectrum of compound 90 (50 – 80 ppm).
Figure A.26. Full $^1$H NMR spectrum of compound 92 (−0.5 – 8 ppm).
Figure A.27. $^1$H NMR spectrum of compound 92 (6–8 ppm).
Figure A.28. $^1$H NMR spectrum of compound 92 (3.5–5.25 ppm).
Figure A.29. Full $^{13}$C NMR spectrum of compound 92 (-10 – 200 ppm).
Figure A.30. $^{13}$C NMR spectrum of compound 92 (95 – 155 ppm).
Figure A.31. $^{13}$C NMR spectrum of compound 92 (-10 – 80 ppm).
Figure A.3. Full $^1$H NMR spectrum of compound 102 (−0.5 – 8 ppm).

dichloromethane

chloroform

acetone

m, n
Figure A.33. $^1$H NMR spectrum of compound 102 (5.5 – 7.5 ppm).
Figure A.34. $^1$H NMR spectrum of compound 102 (3.5 – 5 ppm).
Figure A.35. Full $^{13}$C NMR spectrum of compound 102 (-10 – 200 ppm).
Figure A.36. $^{13}$C NMR spectrum of compound 102 (95 – 155 ppm).
Figure A.37. \textsuperscript{13}C NMR spectrum of compound \textbf{102} (-10 – 80 ppm).
Figure A.38. Full $^1$H NMR spectrum of compound 147 crude (-0.5 – 7.5 ppm).
Figure A.39. $^1$H NMR spectrum of compound 147 crude (6 – 7.5 ppm).
Figure A.40. $^1$H NMR spectrum of compound 147 crude (3.5 – 5.75 ppm).
Figure A.41. $^1$H NMR spectrum of compound 147 (-0.5 - 2.75 ppm).
Figure A.42. Full $^1$H NMR spectrum of compound 148 (0 – 8 ppm).
Figure A.43. Full $^{13}$C NMR spectrum of compound 148 (0 – 200 ppm).
Figure A.44. Full $^{13}$C NMR spectrum of compound 148 (95 – 155 ppm).
Figure A.45. $^{13}$C NMR spectrum of compound 148 (95 – 155 ppm).
Figure A.46. $^{13}$C NMR spectrum of compound 148 (15 – 65 ppm).
Figure A.47. $^{13}$C NMR spectrum of compound 148 (70 – 80 ppm).
Figure A.48. ESI-Mass spectrum of the purified compound 109, taken in positive ion mode. MW: 917.14, m/z: 917.47 (M+).
Figure A.49. ESI-Mass spectrum of the purified compound 118, taken in positive ion mode. MW: 802.88, m/z: 803.47 (M+H\(^+\)).
**Figure A.50.** ESI-Mass spectrum of the purified compound **122**, taken in negative ion mode. MW: 1161.18, m/z: 1160.13 (M–H⁺).
Figure A.51. ESI-Mass spectrum of the purified compound 149, taken in positive ion mode. MW: 941.05, m/z: 941.63 (M+H\(^+\)).
Figure A.52. ESI-Mass spectrum of the reaction mixture containing compound 149 (starting material) and propargyl bromide, taken in positive ion mode. Both the starting material (149, blue) and the product (150, green) can be seen in the spectra. Starting material (149) MW: 941.05, m/z: 941.73 (M+H⁺). Product (150) MW: 979.10, m/z: 979.67 (M+H⁺).
Figure A.53. ESI-Mass spectrum of the purified compound 150, taken in positive ion mode. MW: 979.10, m/z: 979.53 (M+H⁺).
Figure A.54. ESI-Mass spectrum of the purified compound 111, taken in negative ion mode. MW: 838.27, m/z 837.27 (M–H⁺, green), 879.33 (M–H⁺ + LiCl, blue), 418.12 (M–2H⁺, red).
Figure A.55. ESI-Mass spectrum of the purified compound 115, taken in negative ion mode. MW: 798.80, m/z: 797.20 (M–H⁺).
Grace Chromatographs

Column: Reveleris® Silica 40 g
Flow Rate: 40 mL/min
Equilibration: 2.4 min
Run Length: 28.3 min
Solvent A: Hexane
Solvent B: Ethyl acetate
Solvent C: Methanol
Solvent D: Methylene chloride

Gradient Table

<table>
<thead>
<tr>
<th>Min</th>
<th>Solvents</th>
<th>% 2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>AB 55</td>
</tr>
<tr>
<td>2</td>
<td>1.0</td>
<td>AB 55</td>
</tr>
<tr>
<td>3</td>
<td>14.4</td>
<td>AB 78</td>
</tr>
<tr>
<td>4</td>
<td>4.8</td>
<td>AB 78</td>
</tr>
<tr>
<td>5</td>
<td>0.0</td>
<td>BC 0</td>
</tr>
<tr>
<td>6</td>
<td>2.4</td>
<td>BC 35</td>
</tr>
<tr>
<td>7</td>
<td>4.8</td>
<td>BC 35</td>
</tr>
<tr>
<td>8</td>
<td>1.4</td>
<td>BC 35</td>
</tr>
</tbody>
</table>

Figure A.56. Flash chromatography separation of 15a and 15b. Performed on a GRACE Reveleris
Column: Reveleris® Silica 4g
Flow Rate: 18 mL/min
Equilibration: 4.0 CV
Run Length: 51.8 CV
Solvent A: Methylene chloride
Solvent B: Hexane
Solvent C: Ethyl acetate
Solvent D: i-Propanol

Gradient Table

<table>
<thead>
<tr>
<th>CV</th>
<th>Solvents</th>
<th>% 2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>AB</td>
</tr>
<tr>
<td>2</td>
<td>48</td>
<td>AB</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>AB</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
<td>BC</td>
</tr>
<tr>
<td>5</td>
<td>3.0</td>
<td>BC</td>
</tr>
<tr>
<td>6</td>
<td>3.0</td>
<td>BC</td>
</tr>
<tr>
<td>7</td>
<td>6.0</td>
<td>BC</td>
</tr>
<tr>
<td>8</td>
<td>1.4</td>
<td>BC</td>
</tr>
<tr>
<td>9</td>
<td>8.1</td>
<td>BC</td>
</tr>
<tr>
<td>10</td>
<td>18.5</td>
<td>BC</td>
</tr>
<tr>
<td>11</td>
<td>20.0</td>
<td>BC</td>
</tr>
<tr>
<td>12</td>
<td>0.0</td>
<td>BC</td>
</tr>
<tr>
<td>13</td>
<td>3.0</td>
<td>BC</td>
</tr>
<tr>
<td>14</td>
<td>3.0</td>
<td>CD</td>
</tr>
</tbody>
</table>
Flash chromatography purification of 18. Performed on a GRACE Revelers.

Column: Revelers® Silica 12g
Flow Rate: 40 mL/min
Equilibration: 5.0 CV
Run Length: 49.8 CV
Solvent A: Methylene chloride
Solvent B: Hexane
Solvent C: Ethyl acetate
Solvent D: i-Propanol

<table>
<thead>
<tr>
<th></th>
<th>CV</th>
<th>Solvents</th>
<th>% 2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>AB</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3.0</td>
<td>AB</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>BC</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3.0</td>
<td>BC</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>3.0</td>
<td>BC</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>10.0</td>
<td>BC</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>10.0</td>
<td>BC</td>
<td>39</td>
</tr>
<tr>
<td>8</td>
<td>10.0</td>
<td>BC</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>3.0</td>
<td>BC</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>BC</td>
<td>100</td>
</tr>
<tr>
<td>11</td>
<td>3.0</td>
<td>BC</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>1.6</td>
<td>BC</td>
<td>100</td>
</tr>
<tr>
<td>13</td>
<td>1.9</td>
<td>BC</td>
<td>100</td>
</tr>
<tr>
<td>14</td>
<td>0.0</td>
<td>CD</td>
<td>100</td>
</tr>
<tr>
<td>15</td>
<td>1.4</td>
<td>CD</td>
<td>100</td>
</tr>
</tbody>
</table>
Column: Reveleris® Silica 12g
Flow Rate: 36 mL/min
Equilibration: 6.0 CV
Run Length: 15.0 CV
Solvent A: Hexane
Solvent B: Methylene chloride
Solvent C: Ethyl acetate
Solvent D: i-Propanol

Gradient Table

<table>
<thead>
<tr>
<th></th>
<th>CV</th>
<th>Solvents</th>
<th>% 2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>AB</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>2.0</td>
<td>AB</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>8.0</td>
<td>AB</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
<td>BC</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>5.0</td>
<td>BC</td>
<td>50</td>
</tr>
</tbody>
</table>

Figure A.59. Flash chromatography purification of 92. Performed on a GRACE Reveleris
Column: Reveleris® Silica 4g
Flow Rate: 18 mL/min
Equilibration: 4.0 CV
Run Length: 35.0 CV
Solvent A: Methylene chloride
Solvent B: Ethyl acetate
Solvent C: i-Propanol
Solvent D: Hexane

Gradient Table

<table>
<thead>
<tr>
<th>CV</th>
<th>Solvents</th>
<th>% 2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>AB</td>
</tr>
<tr>
<td>2</td>
<td>5.0</td>
<td>AB</td>
</tr>
<tr>
<td>3</td>
<td>2.0</td>
<td>AB</td>
</tr>
<tr>
<td>4</td>
<td>8.0</td>
<td>AB</td>
</tr>
<tr>
<td>5</td>
<td>2.0</td>
<td>AB</td>
</tr>
<tr>
<td>6</td>
<td>8.0</td>
<td>AB</td>
</tr>
<tr>
<td>7</td>
<td>2.0</td>
<td>AB</td>
</tr>
<tr>
<td>8</td>
<td>2.0</td>
<td>AB</td>
</tr>
<tr>
<td>9</td>
<td>0.0</td>
<td>AB</td>
</tr>
<tr>
<td>10</td>
<td>3.0</td>
<td>AB</td>
</tr>
<tr>
<td>11</td>
<td>0.0</td>
<td>BC</td>
</tr>
<tr>
<td>12</td>
<td>3.0</td>
<td>BC</td>
</tr>
</tbody>
</table>
Figure A.61. Flash chromatography purification of 109. Performed on a GRACE Reveleris

Column: Reveleris® Silica 4g
Flow Rate: 18 mL/min
Equilibration: 4.0 CV
Run Length: 48.8 CV
Solvent A: Methylene chloride
Solvent B: Acetone
Solvent C: Ethyl acetate
Solvent D: i-Propanol

Gradient Table

<table>
<thead>
<tr>
<th></th>
<th>CV</th>
<th>Solvents</th>
<th>% 2r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>AB</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>5.0</td>
<td>AB</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3.0</td>
<td>AB</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>3.0</td>
<td>AB</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>6.0</td>
<td>AB</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>5.0</td>
<td>AB</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>2.0</td>
<td>AB</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>4.4</td>
<td>AB</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>6.8</td>
<td>AB</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>3.6</td>
<td>AB</td>
<td>100</td>
</tr>
<tr>
<td>11</td>
<td>0.0</td>
<td>AD</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>10.0</td>
<td>AD</td>
<td>100</td>
</tr>
</tbody>
</table>
Figure A.62. Flash chromatography purification of 147. Performed on a GRACE Revelers.

Column: Revelers® Silica 4g
Flow Rate: 18 mL/min
Equilibration: 5.0 CV
Run Length: 63.1 CV
Solvent A: Methylene chloride
Solvent B: Ethyl acetate
Solvent C: i-Propanol
Solvent D: Hexane

Gradient Table

<table>
<thead>
<tr>
<th>CV</th>
<th>Solvents</th>
<th>% 2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>AB</td>
</tr>
<tr>
<td>2</td>
<td>5.0</td>
<td>AE</td>
</tr>
<tr>
<td>3</td>
<td>6.0</td>
<td>AB</td>
</tr>
<tr>
<td>4</td>
<td>20.0</td>
<td>AB</td>
</tr>
<tr>
<td>5</td>
<td>10.0</td>
<td>AB</td>
</tr>
<tr>
<td>6</td>
<td>6.0</td>
<td>AB</td>
</tr>
<tr>
<td>7</td>
<td>5.0</td>
<td>AB</td>
</tr>
<tr>
<td>8</td>
<td>5.0</td>
<td>AB</td>
</tr>
<tr>
<td>9</td>
<td>0.0</td>
<td>BC</td>
</tr>
<tr>
<td>10</td>
<td>6.0</td>
<td>BC</td>
</tr>
</tbody>
</table>
Figure A.63. Flash chromatography purification of 148. Performed on a GRACE Reveleris

Column: Reveleris® Silica 12g
Flow Rate: 36 mL/min
Equilibration: 3.0 CV
Run Length: 27.0 CV
Solvent A: Hexane
Solvent B: Methylene chloride
Solvent C: Ethyl acetate
Solvent D: i-Propanol

<table>
<thead>
<tr>
<th>CV</th>
<th>Solvents</th>
<th>% 2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>AE</td>
</tr>
<tr>
<td>2</td>
<td>1.0</td>
<td>AE</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>AE</td>
</tr>
<tr>
<td>4</td>
<td>2.0</td>
<td>AE</td>
</tr>
<tr>
<td>5</td>
<td>0.0</td>
<td>BC</td>
</tr>
<tr>
<td>6</td>
<td>8.0</td>
<td>BC</td>
</tr>
<tr>
<td>7</td>
<td>8.0</td>
<td>BC</td>
</tr>
<tr>
<td>8</td>
<td>2.0</td>
<td>BC</td>
</tr>
<tr>
<td>9</td>
<td>2.0</td>
<td>BC</td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>CD</td>
</tr>
<tr>
<td>11</td>
<td>4.0</td>
<td>CD</td>
</tr>
</tbody>
</table>
NMR and GC-Mass Spectra of Previously Synthesized Compounds

Figure A.64. Full $^1$H NMR spectrum of compound 8 (0 – 11 ppm).
Figure A.65. Full $^{13}$C-$^1$H NMR spectrum of compound 8 (0 – 200 ppm).
**Figure A.66.** GC-MS Analysis of the recrystallized compound 8. A) Gas chromatograph. B) Mass spectrum from a retention time of 19.651 min. MW 224.07, m/z: 224 (M⁺), 165 (M⁺ – CO₂CH₃), 151 (M⁺ – CH₂CO₂CH₂), 137 (M⁺ – CH₂CO₂CH₃ – CH₃), 105 (M⁺ – OCH₂CO₂CH₃ – OCH₃), 77 (M⁺ – CHO – OCH₂CO₂CH₃ – OCH₃).
Figure A.67. Full $^1$H NMR spectrum of compound 4 (0 – 11 ppm).
Figure A.68. Full $^{13}$C/$^1$H NMR spectrum of compound 4 (0–180 ppm).
Figure A.69. GC-MS Analysis of the recrystallized compound 4. A) Gas chromatograph. B) Mass spectrum from a retention time of 23.177 min. MW 268.09, m/z: 267 (M+–H), 237 (M+–OCH3), 209 (M+–CO2CH3), 196 (M+–CHO2CH2CH2), 123 (M+–CH2CO2CH3–CHO2CH2CH3).
Figure A.70. Full $^1$H NMR spectrum of compound $15$ (0 – 8 ppm).
Figure A.71. Full $^1$H NMR spectrum of compound 16 (0 – 8 ppm).
Figure A.72. Full \textsuperscript{1}H NMR spectrum of compound 17 (0 – 11 ppm).
Figure A.73. $^1$H NMR spectrum of compound 17 (6 – 8 ppm).
Figure A.74. $^1$H NMR spectrum of compound 17 (3 – 5.25 ppm).
Figure A.75. Full \(^1\)C{\(^1\)H} NMR spectrum of compound 17 (0–200 ppm).
Figure A.76. Full 1H NMR spectrum of compound 18 (0–11 ppm).
Figure A.77. $^1$H NMR spectrum of compound 18 (6 – 8 ppm).
Figure A.78. $^1$H NMR spectrum of compound 18 (3.5 – 5.5 ppm).
Figure A.79. Full $^{13}\text{C}\{^1\text{H}\}$ NMR spectrum of compound 18 (0 – 200 ppm).