Cloning and Characterization of New Protocatechuate Dioxygenase (PCAD) Superfamily Members

By
Ling Xie
Faculty Advisor: Prof. Erika Anne Taylor

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

Middletown, Connecticut     May 2013
Abstract

Genome sequence data has been rapidly accumulated with the development of new generations of sequencing techniques. A vast gap exists between the fast growth of sequence data and the comparatively slow progress on functional characterization of sequenced genes. Due to this vast gap, gene functional annotation mostly relies on extrapolations based upon functionally characterized genes with high sequence homology to newly sequenced open reading frames (ORFs). The understanding of the relationship between protein sequence, structure and function, which allows robust classification of protein superfamilies, will ultimately yield more accurate gene annotation. A new superfamily of extradiol dioxygenases, which includes aromatic-ring-cleaving dioxygenases involved in degradation of lignin derived aromatic compounds, has been identified. This superfamily, the procatechuate dioxygenase (PCAD) superfamily is largely uncharacterized with few structures, substrate specificities, or catalytic mechanism having been determined. Genes with less than 20 % pair-wise identity to the first structurally characterized enzyme in the superfamily, protocatechuate 4,5-dioxygenase from Sphingomonas paucimobilis SYK-6, also named as LigAB, have been selected through a BLAST search in order to identify potential dioxygenases in other bacterial strains and explore the boundaries of the PCAD superfamily. Among the resulting genes, YgiD in Escherichia coli K12 has been cloned, expressed, and
purified. YgiD has been shown to bind tightly to Fe(II), and to lack the ability to react with PCA. In the search for the native substrate of YgiD, a differential analysis of 2D NMR spectra was being exploited to investigate what compounds accumulate in a YgiD knockout strain of *E. coli*. The study of YgiD will help in understanding the functional diversity of this superfamily. Six other putative PCAD superfamily members from a diversity of organisms have also been cloned, of them three have been expressed and purified. Further examination of these to the PCAD superfamily members will likely reveal insights into the linkages between structure, sequence and function in this superfamily.
Acknowledgement

I would like to take this opportunity to acknowledge all the people who helped me complete my Master’s Thesis work. First, I would like to thank my advisor Professor Taylor for your mentorship and guidance in the past years. I have become a better scientist than before through your brilliant suggestions and guidance for my professional knowledge, experimental skills, and presentation ability. Thank you.

I would next like to thank my committee members Professor Pratt and Professor Novick for your professional feedback and guidance received from each committee meeting. Especially Professor Pratt, my knowledge of kinetics increased greatly from the two courses with you. It has been very enjoyable to TA with you, too. I also appreciate all of the other faculty and staff who have helped me in my time here.

I am most grateful to my labmates especially Kevin and Dan. You both helped me a lot in almost every aspect of school life, from course assignments, learning new experimental techniques, troubleshooting failed experiments and encouragement through tough times. I’d also like to thank Jagadash, Julie, Abraham, Noreen, Shu, Ann-Marie and other labmates for their support and entertainment from chatting.

To my family, much love and thank you very much. Your constant encouragement and immense mental support enabled me to finish my studies abroad. Without your love and blessings, this would have not been possible. Thank you to my uncle’s family. You helped me
to settle down when I had just arrived in the USA, and provided me family warmness at Christmas.

I also want to thank my friends especially Yayan who kept me company, helped me indulge in a break, entertained me with funny conversations, and gave me useful suggestions on both study and life. Thanks to everyone for supporting me in this journey.
Contents

Abstract .................................................................................................................................................. i
Acknowledgement ................................................................................................................................... iii
Contents .................................................................................................................................................... v
Table of Abbreviation .......................................................................................................................... viii
Table of Figures ......................................................................................................................................... xi
Chapter 1 Introduction ............................................................................................................................... 2
  “Complete” Gene Sequence Understanding ......................................................................................... 2
    From Sequence to Function ................................................................................................................... 2
    Protein Evolutionary Classification ...................................................................................................... 3
Superfamilies of Dioxygenases ............................................................................................................... 5
  Dioxygenase Evolutionary Classification ............................................................................................... 5
  VOC Superfamily .................................................................................................................................... 10
  Cupin Superfamily ................................................................................................................................. 14
  PCA superfamily ................................................................................................................................... 19
Chapter 2 Characterization of a putative extradiol dioxygenase YgiD in Escherichia coli ...................... 22
  Materials and Methods .......................................................................................................................... 25
    Cloning of YgiD from E. coli. ................................................................................................................ 28
    Expression and purification of putative YgiD dioxygenase. ............................................................... 28
    Isothermal Titration Calorimetry experiment ..................................................................................... 30
    UV-Vis Spectrophotometric Assays of YgiD ....................................................................................... 31
    Knock out ygiD from E. coli ................................................................................................................ 31
    ADP-heptose extraction from WBBO6 ............................................................................................... 33
    Differential analysis of knockout and wild type extractions ............................................................. 34
Results and Discussion ............................................................................................................................ 34
  Expression and purification of putative YgiD dioxygenase ................................................................. 35
  Calorimetric Titration Curves ................................................................................................................. 36
UV-Vis Spectroscopy Characterization ................................................. 39
Knockout Experiment of YgiD ............................................................ 41
Differential analysis of knockout and wild type extractions ................. 42
Conclusions and Future Directions ..................................................... 43

Chapter 3 Expression and Purification of 3-methoxy gallic acid 3,4- dioxygenase (DesZ) from *Sphingomonas paucimobilis* SYK-6 ........................................ 45

Materials and Methods........................................................................... 50
   Cloning of DesZ ................................................................................. 51
   Soluble protein expression screening ............................................... 52
   Expression and Purification of DesZ .................................................. 52
   UV-Vis spectroscopy ......................................................................... 55

Results and Discussion .......................................................................... 55
   Soluble protein expression screening ............................................... 50
   Purification of non-tagged version of DesZ ........................................ 51
   UV-Vis spectroscopy characterization ............................................... 52
   Purification of His-tagged version of DesZ ........................................ 53

Conclusions and Future Directions ....................................................... 61

Chapter 4 Other putative homologs in the PCAD superfamily .............. 62

Materials and Methods........................................................................... 66
   *Pseudomonas syringae* DC3000 (PSPT_1776) .................................. 69
   Cloning of putative PSPT_1776 dioxygenase .................................... 69
   Soluble protein expression screening ............................................... 70
   Expression and purification of putative PSPT_1776 dioxygenase ........ 64
   *Bacillus cereus* ATCC 10987(BCE_1944) ........................................ 73
   Cloning of putative BCE_1944 dioxygenase ...................................... 73
   *Chromobacterium violaceum* ATCC 12472 (CV3550) ..................... 74
   Cloning of putative CV3550 dioxygenase ........................................ 74
   Soluble protein expression screening ............................................... 76
   Expression and purification of putative CV3550 dioxygenase ............ 69
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCAD</td>
<td>procatechuate dioxygenase</td>
</tr>
<tr>
<td>ORFs</td>
<td>open reading frames</td>
</tr>
<tr>
<td>PCA</td>
<td>procatechuate or 3,4-dihydroxybenzoic acid</td>
</tr>
<tr>
<td>3MGA</td>
<td>3-0-methylgallate</td>
</tr>
<tr>
<td>EC</td>
<td>Enzyme Commission</td>
</tr>
<tr>
<td>SCOP</td>
<td>Structure Classification of Protein</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Search Alignment Tool</td>
</tr>
<tr>
<td>DHBD</td>
<td>2,3-dihydroxybiphenyl 1,2-dioxygenase</td>
</tr>
<tr>
<td>MhpB</td>
<td>2,3-dihydroxyphenylpropionate 1,2-dioxygenase</td>
</tr>
<tr>
<td>VOC</td>
<td>Vicinal oxygen chelate</td>
</tr>
<tr>
<td>3,4-PCD</td>
<td>protocatechuate 3,4-dioxygenase</td>
</tr>
<tr>
<td>C12O</td>
<td>catechol 1,2-dioxygenase</td>
</tr>
<tr>
<td>HQ12O</td>
<td>hydroxyquinol 1, 2-dioxygenase</td>
</tr>
<tr>
<td>C23O</td>
<td>catechol 2,3-dioxygenase</td>
</tr>
<tr>
<td>HPCD</td>
<td>homoprotocatechuate 2,3-dioxygenase</td>
</tr>
<tr>
<td>CHQO</td>
<td>chlorohydroquinone dioxygenase</td>
</tr>
<tr>
<td>DHPPD</td>
<td>2, 3-dihydroxyphenylpropionate 1,2-dioxygenase</td>
</tr>
<tr>
<td>AkbC</td>
<td>methylcatechol 2,3-dioxygenase</td>
</tr>
<tr>
<td>4,5-PCD (LigAB)</td>
<td>protocatechuate 4,5-dioxygenase</td>
</tr>
<tr>
<td>APD</td>
<td>2-aminophenol 1,6-dioxygenase</td>
</tr>
<tr>
<td>GDO</td>
<td>gentisate dioxygenase</td>
</tr>
<tr>
<td>PCB</td>
<td>polychlorinated biphenyls</td>
</tr>
<tr>
<td>PcpA</td>
<td>2,6-dichlorohydroquinone 1,2-dioxygenase</td>
</tr>
<tr>
<td>HAD</td>
<td>3-hydroxyanthranilate 3,4-dioxygenase</td>
</tr>
<tr>
<td>AHD</td>
<td>4-amino-3-hydroxybenzoate 2,3-dioxygenase</td>
</tr>
<tr>
<td>HGDO</td>
<td>homogentisate dioxygenase</td>
</tr>
<tr>
<td>HNDO</td>
<td>1-hydroxy2-naphthoate 1,2-dioxygenase</td>
</tr>
<tr>
<td>CDO</td>
<td>cysteine dioxygenase</td>
</tr>
<tr>
<td>QDO</td>
<td>quercetin dioxygenase</td>
</tr>
<tr>
<td>ADO</td>
<td>acireductione dioxygenase</td>
</tr>
<tr>
<td>SDO</td>
<td>salicylate 1,2-dioxygenase</td>
</tr>
<tr>
<td>DesZ</td>
<td>3-methoxy gallic acid 3,4- dioxygenase</td>
</tr>
<tr>
<td>LigD</td>
<td>NAD-dependent β-aryl ether dehydrogenase</td>
</tr>
<tr>
<td>LigE and LigF</td>
<td>β-etherase</td>
</tr>
<tr>
<td>LigG</td>
<td>glutathionelyase</td>
</tr>
</tbody>
</table>
LigX, 5,5’-dehydrodivanillate O-demethylase
LigZ, 2,2’-trihydroxy-3’-methoxy-5,5’-dicarboxyphenyl dioxygenase
LigY, C-C hydrolase
LigW/LigW2, carboxyvanillate decarboxylase
FerA, feruloyl-CoA synthetase
FerB/FerB2, feruloyl-CoA hydratase/lyase
LigM, tetrahydrofolate (H4folate)-dependent O-demethylase
LigC, 4-carboxy-2-hydroxymuconate-6-semialdehyde dehydrogenase
LigJ, 4-oxalomesaconate hydratase
DesB, gallate dioxygenase
LigK, 4-carboxy-4-hydroxy-2-oxoadipate aldolase/oxaloacetate decarboxylase
DesA, syringate O-demethylase
DDVA, 5,5’-dehydrodivanillate
OH-DDVA, 2,2’-trihydroxy-3’-methoxy-5,5’-dicarboxyphenyl
SCVA, 5-carboxyvanillate
CHMS,4-carboxy-2-hydroxymuconate-6-semialdehyde
CHA, 4-carboxy-4-hydroxy-2-oxoadipate
OMA, 4-oxalomesaconate
PDC, 2-pyrene-4,6-dicarboxylate
CHMOD, 4-carboxy-2-hydroxy-6-methoxy-6-oxohexa-2,4-dienoate
Rmsd, root mean square deviation
IPTG, Isopropyl β-D-1-thiogalactopyranoside
Tris-HCl, tris-(hydroxymethyl)aminomethane hydrochloride
EDTA, Ethylenediaminetetraacetic acid
SDS, sodium dodecyl sulfate
APS, ammonium persulfate
TEMED, N, N, N, N’-tetra-methylethylenediamine
SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis
LB, Luria-Bertani
Amp, ampicillin
SOC, Super Optimal broth with catabolite repression
TAE, Tris-acetate-EDTA
ADP-heptose, ADP-L-glycerol-D-manno-heptose
FPLC, fast protein liquid chromatography
PCR, polymerase chain reaction
ITC, isothermal titration calorimetry
NMR, nuclear magnetic resonance
MALDI, matrix-assisted laser desorption ionization
DTT, dithiothreitol
DQCOSY, double-quantum filtered correlation spectroscopy
ESI-MS, electrospray ionization mass spectrometry
SAP, Shrimp Alkaline Phosphatase
TSB, Tryptic Soy Broth
Table of Figures

Chapter 1 Introduction

Figure 1. Hypothetical model for domain structures of three classes of extradiol catechol dioxygenases

Figure. 2. Families of ring-cleaving enzymes based on structural folds

Figure 3. Ribbon structures of A. HPCD, B. GDO and C. LigAB

Figure 4. Ribbon structures of HPCD and DHBD

Figure 7. Structures of A. GDO, B. QDO, and C. HAD

Figure 8. Reactions catalyzed by ring-cleavage dioxygenases in the Cupin superfamily

Figure 9. A. Crystal structure of LigAB; B active site view of LigAB

Chapter 2 Characterization of a putative extradiol dioxygenase YgiD in Escherichia coli

Figure 1. A. Structure of the YgiD dimer; B. Superposition of YgiD and LigB

Figure 2. A. View of the Zn coordination residues in YgiD; B. superposition of active sites of YgiD monomer and LigB; C. active site view of YgiD monomer; D. active site view of YgiD dimer

Figure 3. A. Batch purification result of YgiD through a Sepharose Fast-Flow column charged with Ni$^{2+}$ by step wash

Figure 4. Calorimetric titration profiles of Fe and Zn titrated into YgiD

Figure 5. Interactions between bound Zn ions and YgiD residues

Figure 6. A. Time dependent UV-Vis spectrum of PCA with YgiD

Figure 7. Gradient PCR result of Kan cassette from pKD13 on a 1% Agarose gel

Chapter 3 Expression and Purification of 3-methoxy gallic acid 3,4- dioxygenase

DesZ from Sphingomonas paucimobilis SYK-6

Figure 1. Proposed catabolic pathway for the degradation of lignin-derived aromatic compounds in S. paucimobilis SYK-6

Figure 2. DesZ in pET21b vector expression screening

Figure 3. The chromatograph of non-tagged version of DesZ from DEAE column

Figure 4. The SDS-PAGE gels of the non-tagged version of DesZ from DEAE column
Figure 5. A. Time dependent UV-Vis spectrum of 3MGA with DesZ..................58
Figure 6. The chromatograph of C terminal tagged version of DesZ from the DEAE column..........................................................60
Figure 7. The SDS-PAGE gels of the C-terminal tagged version of DesZ from the DEAE column..........................................................60
Figure 8. The Western-blot of the C-terminal tagged version of DesZ off the DEAD column........................................................................61

Chapter 4  Other putative homologs in the PCAD superfamily
Figure 1. Pair-wise sequence identity determined for comparison of PCA dioxygenase superfamily members.............................................62
Figure 2. Sequence alignment with selected LigB homologs and two structurally characterized enzymes LigB and YgiD .........................57
Figure 3. Gradient PCR result of PSPT_1776 on a 1% Agarose gel ..............82
Figure 4. Colony PCR product of PSPT_1776..............................................82
Figure 5. ps-pET15b expression screening.................................................83
Figure 6. The SDS-PAGE gels of ps-pET15b purified using a Sepharose Fast-Flow column..............................................................83
Figure 7. A. Size exclusion column chromatography of purified ps-pET15b from Sepharose Fast-Flow column........................................84
Figure 8. Gradient PCR result of BCE_1944 on agarose gel ......................84
Figure 9. Colony PCR product of BCE_1944..............................................85
Figure 10. Gradient PCR result of CV3550 on agarose gel..........................86
Figure 11. Colony PCR result of CV3550 .....................................................86
Figure 12. Protein expression of cv-pTOM construct in BL-21 AI cells.........87
Figure 13. The SDS-PAGE gels of cv-pTOM in BL-21 AI cells from a Sepharose Fast-Flow column .........................................................88
Figure 14. Gradient PCR result of SSO0066 on agarose gel........................88
Figure 15. DesB in pET21b vector expression screening..............................89
Chapter 1 Introduction

“Complete” Gene Sequence Understanding

From Sequence to Function. With the development of new generations of sequencing techniques dramatically lowering the cost of genome sequencing, genome sequence data has been rapidly accumulating, thus allowing us to address fundamental biological questions that could not have been asked in the not so distant past. According to a review by Galperin, “by the end of 2009, 1052 genomes representing 720 individual species (636 bacteria, 61 archaea, and 23 eukaryotes) were completely sequenced, deposited in the public nucleotide sequence databases (GenBank\EMBL\DDBJ) and made freely available over the internet.”(1)

However, a vast gap exists between the fast growth of sequence data and the comparatively slow progress on functional characterization of sequenced genes. For example, *Escherichia coli* K12 MG1655 strain is the most well studied organism, since its genome was sequenced in 1997 by Blattner and his coworkers.(2) However, only about 54 % of the protein coding gene products had been experimentally assigned a biological function by 2005.(3) Predicted functions have been assigned to an additional 26 % of coding regions, based on (a) the sequence similarity to proteins with known functions, (b) the operon location, and (c) the phenotypes of mutants.(4) Over 10 % of the gene coding regions have no putative assignments of function at all. Considering all the other sequenced but far less well studied organisms, filling the gap described above with experimental indication of protein functions
is both urgent and crucial.

Due to the vast gap between the numbers of sequenced genes and experimentally verified protein functions, gene functional annotation mostly relies on extrapolations of functionally characterized genes with high sequence homology, based on the assumption that if the two proteins are similar in sequence, their functions are also similar. However, when the sequence identity is below 50%, which is a common situation, the two genes do not necessarily perform similar functions, even though they are in the same protein superfamily. With the rapid growth of sequence and structure data, it is found that notable functional promiscuity exists in many protein superfamilies: one structural fold may exhibit multiple functions, and one function may correspond with multiple structures. These complexities necessitate the understanding of the relationship between protein sequence, structure and function, which is reliant on robust classification of protein superfamilies, which will ultimately yield more accurate gene annotation. (5)

**Protein Evolutionary Classification.** An enzyme superfamily is a basic and popular term in biology and enzymology for protein classification, attributed to a group of enzymes hypothesized to have a common ancestor. The term “superfamily” has experienced a series of changes, and has evolved to have a clearer and more explicit definition as a deeper understanding of protein structure and function has developed. (6-10)

The first person to use superfamily and family to systematically classify proteins is
Margaret O. Dayhoff for “Atlas of Protein Sequence and Structure” in 1965. Without the invention of techniques to obtain protein structures, proteins are categorized only based upon sequence identities. Proteins with more than 50 % sequence identity were organized into one family. Proteins in a sub-family usually have more than 80 % sequence identity. Superfamilies were groups of families set up by statistical methods and search procedures.(11)

Proteins were subsequently reorganized on the basis of either function or structure, along with the help of sequence similarities. This restructuring leads to the formation of Enzyme Commission (EC), the best developed and mostly widely used protein functional classification system. The EC organizes enzymes into six classes of oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases based on their catalyzed chemical reactions.(12) Meanwhile, the Structure Classification of Protein (SCOP, http://scop.mrc-lmb.cam.ac.uk/scop/) database exclusively relies on visual comparison of structures evolved from a common progenitor to construct the protein classification system.(13) After more extensive exploration of evolutionary relationships of superfamily members hidden beyond the sequence level was performed, people found that the above protein classifications were not enough to describe the global relationships between protein sequence, structure and function. By considering structure, function, and sequence together, it was discovered that: (a) enzymes with chemistry-based homology may be partitioned into
different superfamilies in SCOP due to differences in structures, (b) enzymes which catalyze different categories of overall reactions in EC system can be closely related based on their similar structures.

This necessitates the development of a novel definition of superfamily which better describes the divergent evolution of enzyme functions.(14) Gerlt and Babbit clarified the definition for a superfamily in 2001, suggesting that it describes a group of homologous enzymes that catalyze either (a) the same chemical reaction with differing substrate specificities or (b) different overall reactions that share a common mechanistic attribute (partial reaction, intermediate, or transition state) enabled by conserved active site residues that perform the same functions in all members of the superfamily.(15) As a result of this definition, members of a superfamily typically share less than 50 % sequence similarity, and often share less than 20 % sequence identity. To the best of my knowledge, no better definition of superfamily has been proposed to systematically consider the protein structure, function, and sequence together, other than this one. With more complete understanding of superfamily members, the definition of superfamily may further evolve in the future.

Superfamilies of Dioxygenases

Dioxygenase Evolutionary Classification. Dioxygenase enzymes are significant participants in microbial pathways for degrading aromatic molecules.(16) The degradation of
aromatic compounds by dioxygenases includes two significant reactions: ring dihydroxylation and ring cleavage (Scheme 1). Catecholic ring cleavage is generally catalyzed with one of two regiochemical ring cleavage patterns: (a) intradiol dioxygenases which cleave between two adjacent hydroxyl substituents, typically involving a non-heme Fe(III); and (b) extradiol dioxygenases that cleave the bond adjacent to the hydroxyl substituents typically involving a non-heme Fe(II).(17) Intradiol dioxygenases investigated to date utilize substrates possessing vicinal hydroxyl groups with mildly electron-withdrawing substituents, while extradiol enzymes have demonstrated greater versatility, cleaving a wider variety of substrates with either electron withdrawing or electron-donating substituents. (18)

Scheme 1. Typical reactions catalyzed by ring hydroxylating and cleaving dioxygenases. Intradiol and extradiol cleavage patterns are indicated by blue and magenta respectively. Sequence and structural homology indicates that all the intradiol dioxygenases belong to a single evolutionarily related superfamily, (19) while the evolutionary relationships within extradiol dioxygenase superfamilies are less well defined. Limited by the numbers of extradiol dioxygenases found and the finite understanding of superfamily at the time, several
wrong classifications for extradiol dioxygenases have been proposed before. In 1989 Harayama and Rekik divided the extradiol dioxygenases into two families based on sequence alignment and substrate specificity: those showing a preference for bicyclic substrates and those showing a preference for monocyclic substrates. (19) Later in 1993 Bugg suggested an expanded classification system for extradiol dioxygenases using sequence alignment (Figure 1). (20) Class I are single-domain enzymes typified by the small 2,3-dihydroxybiphenyl 1,2-dioxygenase (DHBD) III in *Rhodococcus globerrulus* P6; (21) Class II, which most extradiol enzymes are belonged to, are two-domain enzymes in which the catalytically active C-terminal domain binds Fe(II), as the DHBD from *Pseudomonas cepacia* LB400 exemplifies. (21) Class III are two-domain enzymes where the N-terminal domain contains conserved histidine residues, typified by the 2,3-dihydroxyphenylpropionate 1,2-dioxygenase (MhpB) in *E. coli*. (22)

| Class I (DHBD III in *R. globerrulus* P6) | ... AHV ... HHR ... HHT ... QHG ... SLY ... VEL |
| Class II (DHBD in *P. cepacia* LB400) | Domain 1 ... GYM ... AWR ... AFA ... ASL ... LIT ... LE I |
| | Domain 2 ... GHF ... HHT ... HHT ... RHT ... SFY ... VEY |
| Class III (MhpB in *E. coli*) | Domain 1 ... SHS ... DHY ... GDF ... DHG ... PLE ... LDK |
| | Domain 2 ... SHQ ... ERE ... EAQ ... LSA ... KST ... IKT |

Figure 1. Hypothetical model for domain structures of three classes of extradiol catechol dioxygenases (adapted from Ref. (20)). The Fe(II) ligands identified in the dioxygenase DHBD by X-ray crystallography (His-146, His-210, and Glu-260) are underlined. (21) The positions marked by asterisks correspond to seven amino acid residues as active residues in DHBD from *Pseudomonas cepacia* LB400, namely, the three iron (II) ligands and His-195, His-241, Ser-248, and Tyr-250. (21) Residues which are identically conserved within each dioxygenase enzyme class are in bold face.
Recent studies have shown that the relationships of enzymes within a protein superfamily can be understood not only in terms of their sequence similarities and three-dimensional structures, but also by chemical threads that relate their functional attributes. Based on this, Eltis gave a revised and renamed classification system for dioxygenases. (18) Extradiol dioxygenases are divided into three superfamilies on the basis of this classification system (Figure 2): type I, Vicinal oxygen chelate (VOC) superfamily which includes two-domain and one-domain enzymes; (23) type II, an unknown superfamily which we have named the PCAD superfamily includes enzymes consisting of one or two different subunits; and type III, the Cupin superfamily which also are one or two domain dioxygenases. (24) Each type of dioxygenases has their own structural, which will be elucidated below characteristics (Figure 3).
Figure 2. Families of ring-cleaving enzymes based on structural folds (Adapted from ref.(18)). 3,4-PCD\(\text{B-10}\), protocatechuate 3,4-dioxygenase from \(P.\ putida\ \text{B-10} ;\) (25) C12O\(\text{ADP}_1\), catechol 1,2-dioxygenase from \(Acinetobacter\ sp.\ \text{ADP}_1;\) (26) DHBD\(\text{LB400}_2\), 2,3-dihydroxybiphenyl 1,2-dioxygenase from \(Pseudomonas cepacia\ \text{LB400} ;\) (21) HQ12O\(\text{3E}_2\), hydroxyquinol 1,2-dioxygenase from \(Nocardioides\ simplex\ \text{3E}_2;\) (27) C23O\(\text{mt2}_2\), catechol 2,3-dioxygenase from \(P.\ putida\ \text{mt-2} ;\) (28) HPCD\(\text{Bfu}_4\), homoprotocatechuate 2,3-dioxygenase from \(B.\ fuscum ;\) (29) CHQO, chlorohydroquinone dioxygenase; DHBD\(\text{P6-III}_2\), 2,3-dihydroxybiphenyl 1,2-dioxygenase III from \(R.\ globerulus\ \text{P6} ;\) (30) L-DOPA dioxygenase (LmbB1) from \(Streptomyces\ lincolnensis ;\) (31) gallate dioxygenase from \(P.\ putida\ \text{KT2440} ;\) (32) DHPPD\(\text{Eco}_4\), 3-dihydroxyphenylpropionate 1,2-dioxygenase from \(E.\ coli ;\) (33) HPCD\(\text{EcoC}_x\), homoprotocatechuate 2,3-dioxygenase from \(E.\ coli\ \text{C} ;\) (34) 4,5-PCD\(\text{SYK6,LigAB}_2\), protocatechuate 4,5-dioxygenase from \(Sphingomonas paucimobilis\ \text{SYK-6} ;\) (35) 4,5-PCD\(\text{T2}_4\), protocatechuate 4,5-dioxygenase from \(Comomona testosteroni\ \text{T-2} ;\) (36) C23O\(\text{JMP222-I}_4\), catechol 2,3-dioxygenase I from \(Alcaligenes eutrophus\ \text{JMP222} ;\) (37) APD\(\text{JS45}_2\), 2-aminophenol 1,6-dioxygenase from \(P.\ pseudoalcaligenes\ \text{JS45} ;\) (38) 1-hydroxy-2-naphthoate dioxygenase from \(Nocardioides\ sp.\ \text{KP7} ;\) (39) GDO, gentisate dioxygenase from \(P.\ testosterone ;\) (40) An “x” indicates an unknown oligomeric state.
Figure 3. Ribbon structures of A. HPCD$_{Bfu}$ (2qjt.pdb, VOC superfamily), B. GDO (3bu7.pdb, cupin superfamily), and C. LigAB (1b4u.pdb, unknown superfamily). Metal ions are shown as orange spheres.

**VOC Superfamily.** The VOC superfamily was named from the fact that the first identified three types of enzymes in this superfamily (glyoxalase I, extradiol dioxygenase, and fosfomycin resistance proteins) all have substrates or intermediates that coordinate to the metal center via vicinal oxygens.(41) Sequence alignments, three-dimensional structures, and functional analysis reveal that there are at least seven functionally distinct subclasses of this superfamily: type I extradiol dioxygenases, 4-hydroxyphenylpuruvate dioxygenase,(42) glyoxalase I,(43) fosfomycin resistance proteins,(44) methylmalonyl-CoA epimerases,(45) bleomycin resistance proteins,(46) and aminocarboxycyclopropane forming enzyme.(47)

Among the VOC superfamily members, type I extradiol dioxygenases are most often found as part of metabolic pathways. Catechol 2,3-dioxygenases (C23O) are present in pathways which degrade a variety of substituted aromatic compounds like toluene, nitrobenzene and aniline.(48-50) 2,3-dihydroxybiphenyl 1,2-dioxygenases (DHBD) are in
pathways degrading polychlorinated biphenyls (PCB). Homoprotocatechuate
2,3-dioxygenase (HPCD) is found in tyrosine catabolism.

All type I extradiol dioxygenases identified to date comprise one single subunit with
divergence into one- and two-domain enzymes. The majority of type I enzymes are
two-domain enzymes (Figure 2, DHBD_{LB400}, C23O_{mt2}, and HPCD_{Bfu}) within a range of
oligomeric states, while the one-domain enzymes (DHBD_{P8-III} and L-DOPA
dioxygenase_{lincolnensis}) are usually homodimers. Phylogenetic analysis based on sequence
alignment indicates that the ancestral type I dioxygenase was a one-domain enzyme, and two
duplication events were involved in the divergence of type I dioxygenases into one- and
two-domain enzymes. The majority of type I enzymes are two-domain, suggesting that
the catalytically inactive N-domain may bring some sort of advantage. No one-domain type I
extradiol dioxygenases have been structurally studied. Biochemical and crystallographic
characterization of VOC superfamily dioxygenases has been concentrated on the larger
two-domain enzymes, such as DHBD and HPCD.

Although sequence alignment indicates that the identity between HPCD_{Bfu} and DHBD
from *Pseudomonas sp.* strain KKS102 is no more than 15 %, their tertiary structures are
very similar (Figure 4). Both of monomers have their active site at the C-terminus. Each
monomer contains four conserved βαβββ motifs, which is conserved in all VOC superfamily
members. In the active site, the observed ligands for the metal ion are two histidines and
one glutamic acid residues. This 2-His-1-carboxylate facial triad is strictly conserved in all the type I dioxygenases.(54) The glutamate coordinates with the iron monodentately.

Sequence alignments of 23 mostly type I extradiol dioxygenases by Eltis show that there are another six conserved residues besides the 2-His-1-carboxylate facial triad: two histidines and one tyrosine in the active site pocket whose function are yet unknown; one glycine, one leucine, and one proline at the N- and C-terminal domain interface which are proposed to play a role in protein structure or folding.(17)

Figure 4. Ribbon structures of HPCD from *Brevibacterium fuscum* (left, 2qjt.pdb) and DHBD from *Pseudomonas sp.* strain KKS102 (right, 1eir.pdb). The $\alpha$ helices involving in the motif $\beta\alpha\beta\beta$ are colored in red. Fe (II) is highlighted in orange. Ligands coordinated with Fe are colored in cyan. The N- and C-terminal residues of each chain are labeled with N and C.

Some typical reactions catalyzed by type I dioxygenases and their metal contents are displayed in Figure 5. The substrates of most type I dioxygenases are catecholic compounds, except 2,6-dichlorohydroquinone 1,2-dioxygenase (PcpA), which catalyzes the substrate consisting of a hydroxyl-substituted aromatic ring with a hydroxyl at the *para* instead of
Figure 5. Reactions catalyzed by some type I dioxygenases. A. 2,3-dihydroxybiphenyl 1,2-dioxygenase (DHBD);(55, 56) B. methylcatechol 2,3-dioxygenase (AkbC);(57) C. catechol 2,3-dioxygenase (C23O);(58) D. 2,6-dichlorohydroquinone 1,2-dioxygenase (PcpA);(59) E. homoprotocatechuate 2,3-dioxygenase (HPCD);(60-62) F. L-DOPA dioxygenase.(63)

ortho position. The majority of the dioxygenases in VOC superfamily use Fe(II) as an active metal, only HPCD exhibited interesting “metal promiscuity”. HPCD from *Brevibacterium fuscom* showed highest catalytic activity under the presence of Co(II), and comparable activity with Mn(II) and Fe(II). HPCD from *Klebsiella pneumonia* is
the only one in type I dioxygenases which uses Mg(II) as an active metal, to the best of my knowledge. Further investigation is required to probe the cause of substrate diversity and metal promiscuity, and the effect to the catalytic mechanism.

**Cupin Superfamily.** The members of the cupin superfamily all share a six \( \beta \)-strands barrel core structure, from which the term cupin (derived from the Latin word *cupa* for barrel) comes. The cupin superfamily was discovered from an amino acid sequence analysis in which the wheat protein germin shared high level of similarity with fungal spherulins.(64) The cupin superfamily has members from all three domains of life including archaea, bacteria and eukaryotes, and is among the most functionally diverse superfamilies of proteins known.(65) A review by Dunwell in 2001 estimated that there were a minimum 18 different functional subclasses, including both enzymatic and non-enzymatic members, such as dioxygenases, decarboxylases, epimerases, isomerese, epoxidases, auxin-binding proteins, transcription factors, and seed storage globulins.(24) As more cupin members have been identified, the number of functional subclasses has continued to grow.

As with the other cupins, the dioxygenases in the cupin superfamily have two conserved motifs based on sequence alignment (Figure 6). Each motif, \( G(X)_3HXH(X)_3E(X)_bG \) (Motif 1) and \( G(X)_3PXG(X)_2H(X)_3N \) (Motif 2), contains two \( \beta \)-strands. The two motifs together form one metal binding site, and are separated by another two \( \beta \)-strands with an intervening variable length loop (Figure 7). The cupin superfamily
Figure 6. Multiple sequence alignment of selected dioxygenases in cupin superfamily (adapted from Ref.(65)) (some of structures are shown in Figure 6), showing the two diagnostic conserved domains (boxed), the four \( \beta \)-strands (shaded), the numbers of amino acids in the inter-motif region (IMR), the consensus sequence, the four conserved active-site residues (in red), and the conserved structurally important residues (in blue). The four \( \beta \)-strands, C, D, G, and H, are labeled using the convention based on the structure of phaseolin.\(^{(66)}\) The A or B suffix following the protein abbreviations designates either the N- or C-terminal domain of two-domain bicups respectively. The protein abbreviations are as follows: HADsce, 3-hydroxyanthranilate dioxygenase from *Saccharomyces cerevisiae*;(67) HADhsa, 3-hydroxyanthranilate 3,4-dioxygenase from *Homo sapiens*;(68) HGDOhsa, homogentisate dioxygenase form *Homo sapiens*;(69) GDOspo, gentisate 1,2-dioxygenase from *Silicibacter pomeroyi*;(70) GDOhs, gentisate 1,2-dioxygenase from *Halofex sp.* D1227;(71) CDOMmu, cystein dioxygenase in *Mus musculus*;(72) QDOaja, quercetin dioxygenase in *Aspergillus japonicas*;(73) ADOKpn, aciredoxime dioxygenase from *Klebsiella pneumonia*.(74)

Ring-cleaving dioxygenases can be subdivided into monocupins and bicups (Figure 8), based upon whether the monomer contains a single cupin domain such as 3-hydroxyanthranilate dioxygenase (HAD) and 4-amino-3-hydroxybenzoate 2,3-dioxygenase (AHD), or two cupin domains like homogesitate 1,2-dioxygenase (HGDO), gentisate 1,2-dioxygenase (GDO), and qercetin dioxygenase (QDO). The evolution of bicups could be from one or more gene duplication events originating from a monocupin ancestor.(75)

Structures representing the known scaffolds of cupin ring-cleaving dioxygenases are shown in Figure 7. Even though the various structures displayed below demonstrate a low
Figure 7. Structures of A. GDO in *Silicibacter pomeroyi* (bicupin, 3BU7.pdb), B. QDO in *Aspergillus japonicas* (bicupin, 1GQG.pdb), and C. HAD from *Saccharomyces cerevisiae* (monocupin, 1ZVF.pdb). Motif 1 and 2 are colored in magenta and orange respectively. Ni(II), Fe(II), and Cu(II) are colored in green, red, and deep red respectively. Residues coordinating with metal ions are in cyan.

level of overall structural identity, each has a β-barrel structure with the metal binding site at the center. Instead of using the conserved 2-His-1-carboxylate facial triad at the active site, some type III dioxygenases substitute the glutamate with a third histidine. This substitution has been discussed to possibly represent an adaption for binding of both the aromatic
carboxylate and O$_2$ to the metal.(76) Even still keeping the 2-His-1-carboxylate facial triad like HADsce, the glutamate residue binds in a bidentate manner, contrasting with the monodentate manner seen in the VOC superfamily.

All the reactions catalyzed by ring-cleaving dioxygenases found in the Cupin superfamily to date, as well as their metal contents, are shown in Figure 8. Instead of mainly using catecholic substrates as their VOC superfamily counterparts, type III dioxygenases utilize non-catecholic hydroxyl-substituted aromatic carboxylic acids, mostly monohydroxylated substrates, such as getisate, salicylate, 1-hydroxy 2-naphthoate, or aminohydroxybenzoates. Most type III dioxygenases use Fe(II) for catalysis, in rare cases other metals are used like Ni(II) in HAD from Saccharomyces cerevisiae, and Cu(II) in QDO from Aspergillus japonicas. Interestingly, QDO from some organisms display important “metal promiscuity”. They exhibit different metal selectivity, or are cambialistic, which means that they are able to display comparatively significant activities with either of the metals bound at the same active site. For example, the activities of QDO from Bacillus subtilis under the presence of different divalent metals are ordered in following sequence: Co(II), Cu(II)>Mn(II)>Ni(II). Further investigation is required to probe the effect of the the 3-His metal coordination. Investigating the substrate differences between the type I and III dioxygenases besides the primary structural differences and the reasons to cause these differences could increase our understanding about the whole extradiol dioxygenase group.
<table>
<thead>
<tr>
<th>Reaction</th>
<th>Metal Content</th>
<th>Domain, Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Fe(II)</td>
<td>Bicupin, homotetramer</td>
</tr>
<tr>
<td>B</td>
<td>Fe(II)</td>
<td>Bicupin, homohexamer</td>
</tr>
<tr>
<td>C</td>
<td>Fe(II)</td>
<td>Bicupin, homotetramer</td>
</tr>
<tr>
<td>D</td>
<td>Fe(II)</td>
<td>Bicupin, homotetramer</td>
</tr>
<tr>
<td>E</td>
<td>Fe(II), Ni(II)</td>
<td>Monocupin, homodimer</td>
</tr>
<tr>
<td>F</td>
<td>Fe(II)</td>
<td>Monocupin, homodimer</td>
</tr>
<tr>
<td>G</td>
<td>Cu(II), Mn(II), Co(II), Fe(II), Ni(II)</td>
<td>Bicupin, homodimer</td>
</tr>
</tbody>
</table>

Figure 8. Reactions catalyzed by ring-cleavage dioxygenases in cupin superfamily (adapted from Ref. (76)). A. gensitate 1,2-dioxygenase (GDO);(70, 77) B. 1-hydroxy2-naphthoate 1,2-dioxygenase (HNDO);(39) C. salicylate 1,2-dioxygenase (SDO);(78, 79) D. homogesitate 1,2-dioxygenase (HGDO);(69) E. 3-hydroxyanthranilate 3,4-dioxygenase (HAD);(67, 80) F. 4-amino-3-hydroxybenzoate 2,3-dioxygenase (AHD);(81) G. qercetin dioxygenase (QDO).(73, 82, 83)
**PCA superfamily.** The evolution of extradiol dioxynegase classification has resulted in the identification of an as yet unnamed superfamily in the literature which our lab temporarily calls the PCAD superfamily, because PCA is the substrate of LigAB, the first structurally characterized member. This new superfamily was first proposed to be part of the class III two-domain dioxynegases with the conserved histidine residues at the N-terminal domain by Bugg in 1996 (Figure 1).(20) However, a 35 dioxynegases protein sequence alignment conducted by Eltis showed no sequence similarities between the unknown superfamily members and their counterparts in the other classes.(17) No details of the catalytic mechanism were available because of a lack of any structural information for these dioxynegases, until Senda solved the first structure in this superfamily for protocatechuate 4,5-dioxynegase (LigAB) from *Sphingomonas paucimobilis* SYK-6 (Figure 9).(35)

![Figure 9. A. Crystal structure of LigAB (1B4U.pdb). The α subunit is colored in aqua, and the β subunit is colored in blue, with the Fe at the bottom of the cleft between two subunits (in green). B. active site view with the substrate protocatechuate (PCA, in magenta) bound. α and β after the numbers designate the subunit from where the corresponding residue comes.](image-url)
The determination of the LigAB structure validated the prediction that there are three distinct superfamilies containing extradiol dioxygenases from a structural perspective. LigAB is a heterodimer, with the active site at the interface of the two protein subunits (Figure 9A). Neither of the two subunits contains a βαββ domain as the VOC superfamily, nor has a six β-strands core barrel structure, as in the cupin superfamily. The LigB subunit is a single entity with nine β-strands forming a central core which is sandwiched by six α helices. This seems to be a novel type of folding in dioxygenases, because this folding pattern has not been observed through searches using SCOP. At the active site, LigAB keeps the 2-His-1-carboxylate facial triad and uses Fe(II) as the active metal (Figure 9B). Similar to counterparts in the VOC superfamily, the glutamic acid binds to the metal in a monodentate manner and the substrate PCA binds bidentately. Interestingly, some common features have been found around the active sites of LigAB and DHBD_{KKS102} in the VOC superfamily, such as the iron coordination spheres, their substrate stabilization environments, and a putative catalytic acid/base residue (His195 in LigAB, His 194 in DHBD). This suggests that they may possess the same catalytic mechanism despite their completely different tertiary and quaternary structures.

Based on the above information, Eltis and his coworkers proposed that the unknown superfamily includes LigAB and other dioxygenases whose sequences do not align well with their VOC and cupin counterparts (Figure 2). Since this superfamily is newly
assembled, it is far less well studied than the other two dioxygenase superfamilies. No other common features have been confirmed except that the sequences of the unknown superfamily members do not align well with those of VOC or cupin superfamily members. More enzymes must be characterized by crystallography, substrate specificity, gene context, and catalytic mechanism in order to better understand this superfamily.

This thesis describes the exploration of remote members in this unknown PCAD superfamily. The evolutionary diversity of this enzyme superfamily was assessed starting with the characterization of YgiD in *E. coli* and DesZ in *Sphingomonas paucimobilis* SYK-6, followed by the investigation of other homologues. Identification of new superfamily members, especially the ones at the edge of superfamily, will likely reveal homologues with diverse function and will allow further understanding of the overall characteristics of this superfamily.
Chapter 2 Characterization of a putative extradiol dioxygenase YgiD in *Escherichia coli*

In the search for new members of the PCA dioxygenase superfamily, our research led us to YgiD, a protein of unknown function in *E. coli*, which has been annotated to be a dioxygenase in all the sequence databases. Using the YgiD protein sequence as a BLAST entry, the most related sequence is annotated as the LigB subunit of LigAB, even though the pairwise identity between them is only 22.4%. Based on the crystal structure characterized by the Southeast Collaboratory for Structure Genomics in 2007, YgiD crystallized as a dimer with Zn as the bound metal instead of Fe (Figure 1A). While there are two monomers in the asymmetric unit, the interface of the dimer might be a crystallographic artifact. Based on the analysis from PDBePISA ([http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html](http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html), which is a sophisticated service that analyses the interfaces between macromolecules in their crystal

![Figure 1.A. Structure of the YgiD dimer (2PW6.pdb). The two subunits are colored in magenta and red separately, with Zn metal ions in green.](image)

![Figure 1.B. Superposition of YgiD in orange and LigB (B subunit of LigAB, 1B4U.pdb) in gray, with the Fe metal ion in gray and Zn metal ions in green.](image)
structures. The score for YgiD is 10 %, which means it’s only 10 % chance to be a real dimer for YgiD, while the score for LigAB is 65.2 %. The subunit Rossmann-like fold with has eight parallel β-strands located at the central core, surrounded by seven α helices. The root mean square deviation (rmsd) of YgiD superposed on LigB is only 2.223 Å (Figure 1B), showing that the structural folds are very similar.

Examining the active site of YgiD, the Zn ion is directly coordinated by both oxygens of the carboxylate group of Asp86α, a nitrogen from His22β, a nitrogen from His57β, and a nitrogen from His234β (Figure 2A). In addition, a water molecule can be assigned at the distance of 2.692 Å from the zinc center. The coordination sphere could be described as a distorted octahedral geometry (Figure 2A), with the His234β and one of the oxygens from Asp86α’s carboxylate in the axial positions and the other four proteinaceous ligands in equatorial positions.

Generally, the active site of the YgiD monomer superimposes well with that of LigB, especially the metal coordinating residues (Figure 2B). This suggests that YgiD may catalyze a similar type of reaction as LigAB, or different type of reactions but with the same mechanism. Nevertheless, there are some differences between the two active sites (Figure 2C and 2D). The Asp86α from the the scond YgiD subunit occupies the space where PCA is located in LigAB, which blocks the YgiD substrate from accessing the active site from the same direction as PCA in LigAB. This may have several important implications for the
Figure 2. A. View of the Zn coordination residues in YgiD. Distance between the Zn and the coordinating atoms from residues are colored in purple. B. Superposition of active sites of YgiD monomer (orange) and LigB (gray). L and Y after residue numbers stand for LigB and YgiD respectively. C. Active site view of LigAB. D. Active site view of YgiD. Residues are labeled as shown in the pictures. α and β after the residue numbers denote the residues subunit of origin.

function of the protein. His121β in YgiD adopts a vastly different conformation from that of the His127β in LigAB, possibly creating an empty cavity neighboring the Zn metal of YgiD for substrate binding. It’s also possible that the Asp86α may act as a catalytic cavity lid. Upon the binding of the substrate with YgiD, conformation changes may occur, leading Asp86α to be replaced by the substrate. Whether this assumption is true or not will be determined by a
crystal structure characterization of the YgiD-substrate complex. If the YgiD dimer interface is artificial, then that allows for a binding site.

Based upon gene context analysis, investigation carried out by Klemm and Hancock revealed that ygiD gene was highly up-regulated in biofilm formation of *E. coli* 83972 and VR50. Compared with the wild type, knockouts of ygiD in these two strains caused reduced biofilm formation in urine by 26-43%. However, the function of YgiD either as an enzyme or a regulatory protein related to biofilm formation has yet to be fully explored. The full characterization of YgiD may unveil new functions within the PCA dioxygenase superfamily, and thereby increase our knowledge of this new dioxygenase containing superfamily.

**Materials and Methods**

All primers (Table 1), AccuPrime Pfx SuperMix, DH5α, and BL-21(DE3) cells were purchased from Invitrogen (Carlsbad, CA). Plasmid pKD13 and strain BW25113 for generating a ygiD deletion were kindly donated by Dr. John Gerlt’s Lab in Department of Biochemistry in University of Illinois at Urbana-Champaign. *E. coli* K12 ygiD knockout strain JW3007 was purchased from the Coli Genetic Stock Center in Yale University (New Haven, CT). The *E. coli* heptosyltransferase-I knockout strain WBBO6 was kindly donated by Dr. Christian Raetz’s Lab at the Duke University School of Medicine. The *E. coli* K12 mb1760 strain, used to clone ygiD, was obtained from ATCC (ATCC19215, Manassas, VA).
*Dpn*I was obtained from New England Biolabs (Ipswich, MA). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was bought from Gold Bio Technologies (St. Louis, MO). B-PER II Bacterial Protein Extraction Reagent was purchased from Thermo Scientific (Rockford, IL). Ampicillin (Amp), ferrous ammonium sulfate hexahydrate, sodium hydroxide, catechol, triethylamine, tetracycline, 3,4-dihydroxybenzoic acid (PCA), L-cysteine, tris-(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), EDTA, sodium dodecyl sulfate (SDS), glycine, nickel sulfate hexahydrate, zinc sulfate, and 3,3-dimethylglutarate, 2-amino-2-methyl-1,3-propanediol were obtained from Sigma (St. Louis, MO). Imidazole, kanamycin, sodium chloride, sodium biphosphate, Bacto tryptone, Bacto yeast extract, and Amicon 3 k concentrators were purchased from Fisher Scientific (Pittsburgh, PA). Ammonium persulfate (APS), 30 % acrylamide, Coomassie Brilliant Blue R-250, and N, N, N, N'-tetra-methylethylenediamine (TEMED) were obtained from Bio-Rad (Hercules, CA).

The *Qiaquick* Gel Extraction Kit and *Qiaquick* PCR Purification Kit were purchased from Qiagen (Valencia, CA). The *pTOM* vector is a modified version of Novagen pET-15b vector in which the two Ser codons both before and after the His-tag were mutated to His codons by Toomas Haller in John Gerlt’s lab. The modified *pTOM* vector has an NdeI/XhoI/BamHI cloning site and encodes Amp resistance, an N-terminal 10-His-tag and a thrombin protease cleavage site for the cleavage of the His-tag.

Luria-Bertani (LB) media was prepared from 5 g/L sodium chloride, 10 g/L tryptone,
5 g/L yeast extract, and 1 mM sodium hydroxide. SOC media was made using 20 g/L Bacto Tryptone, 5 g/L Bacto Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, and 20 mM glucose. The concentration of ampicillin (Amp) in all growth media was 100 μg/mL, while the concentration of kanamycin in all growth media was 50 μg/mL. All liquid growths were performed at 37 °C, with 200 rpm shaking speed unless otherwise indicated. 1 % agarose gels for DNA electrophoresis were prepared using 10 mg/mL agarose, and 50 mL TAE electrophoresis buffer (40 mM tris, 20 mM glacial acetic acid, and 1 mM EDTA (pH 8.0), and 0.1 mg/mL ethidium bromide). 7.5 % polyacrylamide protein gels were prepared using 30% acrylamide, 0.01 mM EDTA, separating buffer (1.5 M Tris-HCl, pH 8.8), stacking buffer (0.5 M Tris-HCl, pH 6.8), APS and TEMED. Protein running buffer for sodium dodecylsulfide-polyacrylamide gel electrophoresis (SDS-PAGE) was prepared using 0.1 % w/w SDS, 192 mM glycine, and 25 mM Tris (pH 8.2). All DNA agarose and SDS-PAGE protein gels were run at 100 V using a PowerPac Basic power supply from Bio-Rad (Hercules, CA). Protein gels were stained using Coomassie Brilliant Blue R-250 for 22 min and destained for 30 min.

An AKTA FPLC system (GE Healthcare, Piscataway, NJ) was used for ADP-L-glycerol-D-manno-heptose (ADP-heptose) purification. A batch benchtop method was adopted for purification of proteins. Extinction coefficients and predicted masses of all proteins were calculated based on the amino acid sequence using the web application
ProtParam (http://www.expasy.ch/tools/protparam.html). PCR reactions were accomplished utilizing a S1000 Thermal Cycler (Bio-rad, Hercules, CA). Cells were lysed using a French press cell (Aminco, Urbana, Ill.). The presence of ADP-heptose was confirmed by a Finnigal LCQ Advantage MAX mass spectrophotometer (Thermo, Newington, NH). The isothermal titration calorimetry (ITC) experiments were accomplished using a VP-ITC apparatus (MicroCal, Northhampton, MA). Electroporation was carried out using a BTX electro cell manipulator 600 (BTX inc, San Diego, CA). Spectrophotometric assays were taken using a Varian Cary 100 Bio UV-Visible Spectrophotometer (Agilent Technologies, Santa Clara, CA). NMR was performed using Varian Unity Inova 400 MHz NMR spectrometer (Agilent Technologies, Santa Clara, CA). NMR data was processed using either VARIAN’s VNMR software or ACD/NMR processor (ACD/Labs, Toronto, Ontario, Canada).

Cloning of YgiD from E. coli. The YgiD gene was PCR amplified from E. coli K12 mb1760 and subcloned into pTOM vector in the NdeI and BamHI cloning sites by Sam Berman.

Expression and purification of putative YgiD dioxygenase (32.2 kDa). The YgiD-pTOM construct was transformed into BL-21 cells using the heat shock method, and stored in a 40% glycerol solution as a frozen stock in -80 °C freezer. 1 mL overnight LB-amp growth of transformed BL-21 cells was inoculated into 1 L LB-amp media. After the absorbance of the cell growth at a wavelength of 600 nm (OD600) reached 0.6, 1 mL of 1 M
IPTG was added into the growth medium to make a final concentration of 1 mM to induce the cells. Cells were harvested by centrifugation after 24 h growth from the time of induction at 5,000 rpm for 15 min. Resulting cell pellets were lysed in FE buffer (consisting of 50 mM Tris-HCl and 2 mM L-cysteine (pH 7.0)). The cells were passed three times through a French Press cell and followed by centrifugation at 11,000 rpm for 30 min at 4 ºC to remove the cell debris. The resulting supernatant was loaded onto a IMAC Sepharose Fast-Flow column (16 × 50 mm) charged with Ni²⁺, and purified by step wash: first with 80 mL binding buffer (consisting of 50 mM NaH₂PO₄ (pH 7.42), 300 mM NaCl, and 10 mM imidazole), then with 80 mL washing buffer (containing 50 mM NaH₂PO₄ (pH 7.45), 300 mM NaCl, and 50 mM imidazole), eluted by 40 mL eluting buffer (consisting of 250 mM imidazol, 50 mM NaH₂PO₄ (pH 8.0), and 300 mM NaCl). The Ni²⁺ and all residual proteins were stripped from the column using 80 mL strip buffer (containing 200 nM EDTA (pH 8.0), 500 mM NaCl). The fractions containing YgiD were identified by SDS-PAGE. The identified fractions were combined, and subjected to MALDI mass analysis (32073.1 g/mol compared with expected mass 32266.8 g/mol).

To determine the oligomeric state of the purified YgiD, a native gel and size exclusion chromatography methods were exploited. All preparations for running native gels were the same with SDS-PAGE except that (1) SDS was excluded from running buffer, (2) gels were run at 100 V running voltage at 4 ºC, (3) protein samples were not boiled, and (4) a 5X native
gel loading dye without dithiothreitol (DTT) was used. The 5X native gel loading dye was prepared with 0.313 M Tris-HCl (pH 6.8), 0.05 % m/w bromophenol blue, and 50 % m/w glycerol. The size exclusion chromatography was accomplished with a HiLoad 16/60 Superdex 200 prepgrad column (16 × 600 mm, 120 mL) in Dr. Rich Olson’s lab at Wesleyan University. Before loading the protein samples, the column was washed with 40 mL degassed deionized water at 1 mL/min, 45 mL degassed 20 % ethanol at 1.5 mL/min, and then equilibrated with 200 mL degassed equilibration buffer (containing 20 mM Tris (pH 7.4), and 150 mM NaCl). 1 mL of 8 mg/mL purified YgiD was spun at 15,000 rpm for 10 min to remove aggregates, filtered, and loaded onto the column with a superloop. The column was isocratically eluted with 30 mL equilibration buffer.

*Isothermal Titration Calorimetry experiment.* 3 mL purified 4.8 mg/mL YgiD was dialyzed in 1 L of 20 mM Tris-HCl (pH 7.0) with 1 mM EDTA three times to remove trace metals which may bind to the isolated protein, and then dialyzed into 1L of 20 mM Tris-HCl (pH 7.0) three times to remove EDTA. Each dialysis was performed for approximately 4 h. The ITC experiments were accomplished using a VP-ITC apparatus. In each experiment, 48 injections (the volume of each injection was 5 μL, but the first injection is 3 μL) of a solution containing 250 mM ZnSO₄ or Fe(NH₄)₂(SO₄)₂ were titrated into 2.5 mM YgiD (cell volume = 1.43 mL, stirring speed 374 rpm) using a computer controlled 310 μL microsyringe. The reference cell was filled with distilled water. Before each experiment, the protein and metal
solution were degassed for 2-5 min to eliminate air bubbles, using the ThermoVac accessory of the microcalorimeter. The first addition started after baseline stability was achieved. To allow the system to reach equilibrium, a spacing of 350 s between each injection was applied. Data analysis was performed using the Origin software supplied by MicroCal.

**UV-Vis Spectrophotometric Assays of YgiD.** Spectrophotometric Assays of YgiD were carried using a Varian Cary 300 Bio UV-visible Spectrophotometer at room temperature. The reaction was recorded with 21 scans with the range of 200-800 nm at 600 nm/min, with a 30 min interval between scans. Each 2 mL assay mixture contained 50 mM GTA buffer (consisting of 50 mM 3,3-dimethylglutarate, 50 mM 2-amino-2-methyl-1,3-propanediol, and 50 mM Tris, pH 8.5), purified YgiD (about 20 μg), 100 μM Fe(NH₄)₂(SO₄)₂, and 100 μM substrate (PCA or catechol). YgiD was pre-incubated with Fe(NH₄)₂(SO₄)₂ for 10 min before the assay began, and wasn’t added into the mixture consisting of buffer and substrate after the first scan.

**Knock out ygiD from E. coli.** Generating the knockout construction of the ygiD gene in E. coli was attempted based on experiments performed by Wanner (85). The Kanamycin-resistance cassette was generated by PCR from pKD13 using specifically designed primers (Table 1). PCR reactions (25 μL) were prepared using 22.5 μL AccuPrime SuperMix, 1.5 μL pKD13 (116 ng), 0.5 μL primer (50 ng, each). The reaction mixture was incubated at 95 °C for 5 min to denature the DNA, followed by 45 cycles consisting of 95 °C
Table 1  Primers used in this chapter

<table>
<thead>
<tr>
<th>Oligo name</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pKD13</td>
<td></td>
</tr>
<tr>
<td>Forward</td>
<td>GTTCTTACTCCGCGGCAACAGCTATTGTG</td>
</tr>
<tr>
<td></td>
<td>AGGATATCAAAAATGACACCTTTAGTAAAGG</td>
</tr>
<tr>
<td>Reverse</td>
<td>ATTCGGGGGATCCGTGACC</td>
</tr>
<tr>
<td></td>
<td>GCCACGATACCGGATGGCCTCAGCCATCCG</td>
</tr>
<tr>
<td></td>
<td>GTAATTGTTAGCCTATCTGCACCGACAGCT</td>
</tr>
<tr>
<td></td>
<td>GTAGGCTGGAGCTGCTTCG</td>
</tr>
<tr>
<td>ygiD</td>
<td></td>
</tr>
<tr>
<td>knockout</td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>GGGTTTTTGCGAGGCAAAGCG</td>
</tr>
<tr>
<td></td>
<td>GCAGCGGAAGGTCCGTATGGG</td>
</tr>
</tbody>
</table>

for 15 s, 45-65 °C for 30 s for annealing, and 68 °C for 1.5 min for elongation. The PCR products were stored at 4 °C and subsequently purified from agarose gel-electrophoresis by following Qiaquick Gel Extraction Kit procedures. The purified PCR products were digested with restriction enzyme *DpnI* at 37 °C overnight using the following conditions: 1 μg PCR product, 2 μL *DpnI* (20 U), 3 μL 10X NEB buffer 4 for a total volume of 30 μL. The digested products were purified utilizing Qiaquick PCR Purification Kit.

To make electroporation competent BW25113 cells, transformants BW25113 were grown in 100 mL LB-Amp media with 2 mM L-arabinose at 30 °C to an OD$_{600}$ of 0.6,
incubated in ice-water bath for 15 min, and then centrifuged at 6 k rpm for 15 min at 4 °C. The resulting cell pellets were resuspended in 18.75 mL ice-cold ddH₂O, and spun down at 1500 rpm for 10 min at 4 °C. The cell pellets were washed three times with 11.25 mL, 5 mL, and 1 mL ice-cold 10 % glycerol separately, resuspended in 600 μL ice-cold 10 % glycerol, and 50 μL aliquots were dispensed into sterile microcentrifuge effendorf tubes. Electroporation was performed by using 50 μL competent BW25113 cells and 400 ng digested PCR products at 2.5 kV with 25 μF and 200 ohm. 1 mL SOC and 1 mM L-arabinose was added into shocked cells, and shake at 200 rpm for 2 h at 37 °C. 400 μL incubated solution was spread onto LB-Kan Agar plates to select successful transformation. Viable colonies found on the LB-Kan plates were checked for successful knock out of ygiD via PCR.

**ADP-heptose extraction from WBBO6.** 0.5 mL overnight growth of WBBO6 E. coli Hep I and II knockout cells was inoculated into 1 L LB media containing 12 μg/mL of tetracycline, and grown at 37 °C until OD₆₀₀ reached 1.0. Cells were harvested at 6 k rpm for 15 min, and stirred in 20 mL 50 % ethanol for 1 h on ice. The mixture was then centrifuged at 5 k rpm for 10 min, and the resulting supernatant was placed in a Speedvac to remove the ethanol and reduce the sample volume to half. The remaining undisolved cell debris and high molecular weight compounds were removed by using Amicon 3 kDa cut off concentrators. To obtain purer ADP-heptose as a standard, the filtered aqueous solution was further purified through a HiPrep DEAE Sepharose Fast Flow column (16 × 320 mm, 64 mL).
The column was equilibrated with 2 CV of H$_2$O. The filtered solution was loaded onto the column at 3 mL/min after the equilibration, followed by a 1.5 CV of H$_2$O wash. The ADP-heptose was eluted by 5 CV of a linear gradient from 0 % to 52 % of 0.97 M triethylammonium bicarbonate (pH 8.0). The remaining compounds were washed off by 2 CV of 50 % 0.97 M triethylamine. 2 CV of H$_2$O was used to wash off the triethylamine. 150 μL of each fraction was lyophilized twice to remove the triethylamine, and subsequently dissolved in 50 % acetonitrile in water to confirm the presence of ADP-heptose by ESI-MS. Fractions containing ADP-heptose were combined, lyophilized and stored at -80 °C.

**Differential analysis of knockout and wild type extractions.** The *E. coli* wild type extraction of *E. coli* K12 mb1760 was performed using the same procedures as described for the extraction of ADP-heptose from WBB06. Both the ADP-heptose and *E. coli* wild type extraction were lyophilized and dissolved in D$_2$O. DQCOSY spectra were acquired using a VARIAN INOVA 400 MHz NMR spectrometer with following parameters: 512 complex increments (ni) in F1, and 16 scans per increment (nt). Spectra were processed using either VARIAN’s VNMR software or ACD/NMR Processor. Subtraction of *E. coli* wild type DQCOSY spectrum from that of ADP-heptose extraction was prepared via ACD/NMR processor.

**Results and Discussion**

Annotated to be a dioxygenase, the actual function of YgiD is still veiled. The
similar structural folds of YgiD and LigAB suggest YgiD may catalyze a similar type of reaction as LigAB, or different type of reactions but with the same mechanism. However, the differences in their active sites and preliminary gene context analysis both advise YgiD may have a very different function with LigAB. Thus different assays have been applied to explore the substrate of YgiD and its metal content.

**Expression and purification of putative YgiD dioxygenase.** The gene ygiD was subcloned directly from *E. coli* K12 MB1760 into pTOM vector using *NdeI* and *BamHI* digestion sites for the expression of His-tagged enzymes. Successful purification by immobilized nickel affinity chromatography afforded about 200 mg purified proteins per liter of culture. The protein was eluted using the elution buffer (consisting of 50 mM NaH$_2$PO$_4$ (pH 8.0), 300 mM NaCl, and 250 mM imidazole). The purity of the enzyme was confirmed by SDS-PAGE gels (Figure 3A). A single protein band was detected on the denaturing gel corresponding to about 32 kDa. The size exclusion chromatography, in which the molecules in solution were separated by their sizes, was used to investigate the oligomeric state of YgiD. The smaller the protein, the faster it passes through the column. Different oligomeric states of protein will be displayed as separated peaks on the chromatogram because of their variable sizes. Both size exclusion column chromatography and the native gel experiments clearly indicated the YgiD was a monomer (Figure 3B). If YgiD was a dimer then it would elute around the position of BSA on the size exclusion chromatography.
Figure 3. A. Batch purification result of YgiD through a Sepharose Fast-Flow column charged with Ni\(^{2+}\) by step wash: protein loading flow through (F), binding step (B), washing step (W), eluting step (E), and stripping step (S). Numbers besides lane under standard (std) are indicating corresponding molecular weight (kDa). YgiD was eluted and shown in red box. Fractions are indicated above each well. B. Size exclusion column chromatography of 1 mL 8 mg/mL YgiD from Sepharose Fast-Flow column. Traces of standards BSA and Lysozyme were indicated with labels besides them. C. Native gel of YgiD eluted from size exclusion column.

**Calorimetric Titration Curves.** ITC experiments were exploited to determine the metal binding information of YgiD. Upon the injection of ligand aliquots into the ITC sample cell, heat was either taken up or evolved depending on the nature of the reaction. Measurements consist of the time-dependent input of power required to maintain equal
temperatures between the sample and reference cells, causing signals return to baseline before the next injection. Then observations are plotted as the power needed to maintain the reference and sample cell at an identical temperature against the time. (86, 87) Two divalent metals Fe and Zn were chosen, because most extradiol dioxygenases contain iron and the crystal structure of YgD shows zinc. The binding of the divalent metal cations to YgiD is exothermic, the downward peaks indicating heat released on injection of cations. The binding profile of Fe$^{2+}$ binding to YgiD was shown in Figure 4A. It displayed a primary exothermic binding event, following by a second weaker exothermic binding event. A two sites binding mode was applied to fit the curve. The exothermic association parameters were

![Figure 4](image)

Figure 4. Calorimetric titration profiles of aliquots of 250 mM metal ions into 2.5 mM treated YgiD in 20 mM Tris, pH 7.0. A two sites binding mode carried out by Origin software supplied by MicroCal. A. FeNH$_4$SO$_4$; B. ZnSO$_4$
summarized in Table 2. The first association constant $K_1$ ($2.03 \times 10^8 \text{ M}^{-1}$) suggests a very tight binding. While the second association constant $K_2$ ($4.79 \times 10^6 \text{ M}^{-1}$) revealed a weaker but still relatively strong binding. The two enthalpies -83.8 kcal/mol and -31.3 kcal/mol were higher than the enthalpies for other protein metal binding events, which are generally range from 2-25 kcal/mol.\(^{(88-90)}\) The entropy changes of -243 and -76 cal/mol/deg are comparable with entropies from other protein metal binding.\(^{(91, 92)}\) According to the crystal structure, there are three metal binding sites at the YgiD monomer (Figure 5). It’s assumed that the stronger binding event correlates with the interaction between the metal cation and the active site (the middle metal in Figure 5), and the weaker exothermic event perhaps relates to the interactions

Figure 5. Interactions between bound Zn ions and YgiD residues, including integrated and enlarged graph.
between metal cations and one of the other two binding sites (the first, and third metal in Figure 5).

The binding of Zn$^{2+}$ to YgiD was found to be weaker than Fe$^{2+}$. Generally the trends were similar, but more noise was observed from two exothermic binding events (Fig 4B). The enthalpy and entropy are located in normal range, (89, 91, 92) but the rmsd are in ridiculers large (Table 2). The main association constant $K_1$ of Zn$^{2+}$ is much smaller than that of Fe$^{2+}$, suggesting Zn$^{2+}$ is not preferred by YgiD. This is consistent with the observation that dioxygenases tend to contain Fe$^{2+}$ rather than Zn$^{2+}$. (61, 93)

<table>
<thead>
<tr>
<th></th>
<th>N1</th>
<th>K1 (M$^{-1}$)</th>
<th>ΔH (cal/mol)</th>
<th>ΔS (cal/mol/deg)</th>
<th>N2</th>
<th>K2 (M$^{-1}$)</th>
<th>ΔH (cal/mol)</th>
<th>ΔS (cal/mol/deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
<td>0.676 ± 0.01</td>
<td>2.03 ± 0.65E8</td>
<td>-8.38 ± 97E4</td>
<td>-243</td>
<td>0.631 ± 0.01</td>
<td>4.79 ± 1.53E6</td>
<td>-3.14 ± 0.21E4</td>
<td>-74.6</td>
</tr>
<tr>
<td>Zn</td>
<td>3.46 ± 2.42</td>
<td>1.31 ± 47E4</td>
<td>-9188 ± 9886</td>
<td>-12.0</td>
<td>0.56 ± 24.6</td>
<td>4.39 ± 87.5E11</td>
<td>-1.13 ± 983E4</td>
<td>15.3</td>
</tr>
</tbody>
</table>

Table 2. Exothermic association parameters of Fe$^{2+}$ and Zn$^{2+}$ with YgiD obtained from two sites binding mode.

**UV-Vis Spectroscopy Characterization.** UV-Vis spectroscopy is exploited to determine whether YgiD can react with PCA derivatives, since LigAB, an extradiol dioxygenase that is related to YgiD, uses PCA as the native substrate. PCA has a UV spectrum with two peaks at 256 and 292 nm (Figure 6). There was no peak shift after PCA was incubated with YgiD for about 10 h. The intensity change could not be considered as persuasive proof of reaction because PCA itself also increases intensity due to autooxidation.
These scans look different from PCA-LigAB incubation (data not shown), thus implying that PCA is not the substrate of YgiD. Catechol has also been investigated in the same way, and no apparent spectral changes were observed (data not shown). Since YgiD failed to demonstrate any ability to utilize PCA or catechol, we hypothesize that either (1) the enzyme needs to be purified differently in order to maintain activity, or (2) that YgiD is not an extradiol dioxygenase, like LigAB.

![Figure 6](image)

Figure 6. A. Time dependent UV-Vis spectrum of PCA with YgiD. The curves were determined using 50 mM GTA buffer (pH 8.5) consisting of 50 mM 3,3-dimethylglutarate, 50 mM Tris, 50 mM 2-amino-2-methyl-1,3-propanediol, purified YgiD (about 20 μg), 100 μM Fe(NH₄)₂(SO₄)₃, and 100 μM PCA. YgiD was pre-incubated with Fe(NH₄)₂(SO₄)₃ for 10 min before the assay began, and wasn’t added into the mixture consisting of buffer and substrate until the first scan accomplished. The reaction was recorded with 21 scans with the range of 200-600 nm at 600 nm/min, with a 30 min interval between scans. B. UV-Vis spectrum of PCA autooxidation. The assay concentrations and parameters of B are exactly the same with A, except ddH₂O was used to substitute the enzyme.
**Knockout Experiment of YgiD.** Since PCA and catechol are not the substrate of YgiD, in an attempt to identify its natural substrate, a NMR difference experiment was performed using 2D DQCOSY NMR spectroscopy. If YgiD is an enzyme, a strain of *E. coli* that is deficient in YgiD should show differences with their counterparts from the wild type strain by 2D DQCOSY NMR spectroscopy because of the accumulation of substrate. (94) Attempts to generate a YgiD knockout strain were made according to Wanner’s paper,(23) in which the target gene has been replaced by a Kanamycin-resistance cassette. Gradient PCR was applied to obtain Kanamycin-resistance cassette from the vector pKD13 (Figure 7) However, generation of a YgiD knockout failed over multiple attempts. A YgiD knockout strain was ultimately purchased from the Coli Genetic Stock Center at Yale University.

![Figure 7. Gradient PCR result of Kan cassette from pKD13 on a 1% Agarose gel. The object bands are restricted by red box.](image)
**Differential analysis of knockout and wild type extractions.** As a control to ensure the ability to perform the experiment, a heptosyltransferase I (Hep I) deficient strain of *E. coli* (WBB06) was used. A cell extract solution containing the natural substrate of Hep I, ADP-D-glycero-D-manno-heptose, was prepared and analyzed by 2D DQCOSY NMR spectroscopy. The wild type strain *E. coli* mb 1760 strain was also been processed through the same procedures. Figure 8 was obtained by subtracting the DQCOSY spectrum of wild type extraction from that of WBB06 extraction. The yellow dots from the WBB06 extraction correspond to the difference.

Figure 8. Subtraction of DQCOSY spectrum of *E. coli* K12 mb1760 extraction from WBB06 extraction. The red dots are overlapping part. The yellow points correspond to the difference from WBB06 extraction.
spectrum are the ones which are not shown on the wild type extraction spectrum. However, there were more yellow points shown than expected from ADP-heptose alone. They may come from the background noise from the WBB06 extraction spectrum, or extra substance from the extraction process.

**Conclusions and Future Directions.**

While the protein sequence similarity suggests that it is most closely related to enzymes that are extradiol dioxygenases, like LigAB, this project failed to demonstrate any catalytic activity for YgiD. Interestingly, Fe was demonstrated to bind more tightly with YgiD than Zn which was observed to co-crystallize with the enzyme in its crystal structure. Additional experiments need to be performed to demonstrate success with the control reaction for the NMR difference experiment, and once successful attempts could be made to perform the same experiment with wild-type *E. coli* and the YgiD knockout strain. Consider the possibility of YgiD as a regulatory protein, but regulatory proteins can be divided to two groups, gene regulatory proteins which bind with DNA and non-genetic regulatory proteins which generally bind with proteins. An experiment using microscopy could be exploited to determine whether YgiD is a genetic regulatory protein or not. Label the YgiD with the GFP tag, transfer the labeled protein back to the YgiD knockout cell whose chromosome was labeled with DAPI, and visualize the whole complex under the microscope. If YgiD binds with DNA, an overlapping of the two colors resulted from the two fluorophors could be
observed. If YgiD binds is a non-genetic regulatory protein, a protein complex immunoprecipitation experiment using YgiD antibody could be adopted.(95, 96)
Chapter 3 Expression and Purification of 3-methoxy gallic acid 3,4- dioxygenase (DesZ) from *Sphingomonas paucimobilis* SYK-6

Lignin, which comprises 10-35 % of the dry weight in plant cell walls, is a low cost waste from the pulp and paper industry.(97, 98) Lignin degradation is an important area of research due to its role in nature as well as for commercial application.(99) Methods to optimize lignin degradation could provide access to lignin based products or allow enhancement of cellulose and hemicelluloses accessibility.(100) Currently, lignin brings enormous obstacles to the utilization of lignocelluloses for biofuels, and other commercial chemical feedstock production, due to its high molecular weight and the presence of various relatively stable carbon-to-carbon and ether links.(101) As a potentially valuable source of renewable aromatic chemicals, biodegradation of lignin to biofuel molecules could aid the biofuel production, provide a solution for the “food vs. fuel debate” risen from the first generation biofuel production technologies, and ensure overall sustainability and cost effectiveness of the biomass to biofuel production process.(102)

To microbiologically convert lignin to biofuels, three key steps are required: a scalable process for lignin depolymerization, microbiological pathways that can efficiently degrade the aromatic components obtained from the depolymerization, and the development of the microorganisms that are capable of producing fuel-like molecules from available carbon sources. The first and third steps of this process have been the focus of research in both industrial and academic labs; the second step is one focus of the Taylor Lab.
The first step, the depolymerization of lignin, is being extensively pursued through either chemical or microbial processes. (103-105) Much progress has been achieved in the chemical processes including the kraft process using NaOH and Na₂S, the sulfite process utilizing acidic sulfite, the soda process using NaOH, the quatam process employing tetramethylammonium hydroxide as digestion solutions, and several others. (106) The microbial degradation of lignin has been most extensively studied in white-rot (notably *Phanerochaete chrysosporium*) and brown-rot fungi. (107) Several types of extracellular oxidative enzymes found in both types of fungi, such as laccases, lignin peroxidases, manganese peroxidases, and versatile peroxidases, are involved in lignin degradation in plant cell walls. (108, 109) Besides the basidiomycetes, a number of bacteria have been reported to break down lignin. (110, 111) The extensive studies of lignin depolymerization provide abundant choices and possibilities for the further utilization of lignin to more valuable molecules.

For step three, the biosynthesis of fuel-like molecules from available carbon sources, research is ongoing toward the development of bacteria which efficiently convert sugar molecules into both ethanol and other higher molecular weight potential biofuels like butanol. (112-116) This includes the first generation of biofuel production technologies, such as biodiesel production from vegetable oils, bio-ethanol production from energy crops, and biogas produced by anaerobic digestion of biomass. (117-119)
Studies on microbial pathways which are capable of degrading lignin derived aromatic compounds (the second step mentioned above), and thus bridging the gap between the research areas of step one and step three are still rare.(120-123) Among them, a catabolic pathway in *S. paucimobilis* SYK-6 has been identified for funneling several lignin derived monomeric and dimeric components into the citric acid cycle compounds.(124) This pathway was determined on the basis of knock-out complementation studies and the mass spectrometric analysis of reaction products. Most of enzymes involved in the pathway are not fully characterized enzymologically or are just putatively assigned and completely uncharacterized. In the pathway (Figure 1),(124) compounds with a syringyl (4-dydroxy-3,5-dimethoxyphenyl) moiety and guaiacyl (4-hydroxy-3-methoxyphenyl) moiety are degraded to syringate and vanillate respectively. Subsequently, syringate and vanillate are degraded into 3-O-methylgallate (3MGA) and protocatechuate (PCA), which are further converted to pyruvate and oxaloacetate which are part of the citric acid cycle.(125)

Masai and his colleagues investigated the role of DesZ in the pathway through analyzing the ability of *desZ* mutant strain of *S. paucimobilis* SYK-6 to grown on syringate.(125) They found that mutation in *desZ* alone didn’t affect the syringate degradation much, however, the *desB desZ ligB* triple disruption mutant and the *desB ligB* double disruption mutant completely lost the ability to grow on syringate. Meanwhile *desB desZ* disruption mutant has a similar growth rate on syringate as *desB* mutant. It is hypothesized
that the desB ligB double disruption mutant couldn’t grow on syringate, because of the accumulation of gallate which completely inhibits the growth of the SYK-6 strain on syringate completely. To confirm this hypothesis and with the functions of LigM and LigI already determined,(126, 127) the authors built a desB ligB ligM triple inactivation mutant which was able to grow on syringate. To further validate the role of DesZ, a ligM disruption but ligI insertion strain was built and able to grow on syringate, because this mutant closed the route from 3MGA to gallate and activated the route from 3MGA to 4-oxalomesaconate (OMA) through 2-pyrone-4,6-dicarboxylate (PDC), thus allowing the mutant to grow. All of these factors indicated that there was a pathway converting 3MGA through 4-carboxy-2-hydroxy-6-methoxy-6-oxohexa-2,4-dienoate (CHMOD) to OMA catalyzed by DesZ and a hydrolase. The reaction products are validated through GC-MS. The $K_m$ and $V_{max}$ values of DesZ for 3MGA were determined to be $210 \pm 24 \, \mu M$ and $3.6 \pm 0.2 \, \text{U/mg}$, respectively.(128) Complete characterization of DesZ would allow for a full understanding of this pathway and deeper knowledge of the unknown dioxygenase superfamily.

This work is a continuation of the preliminary research by Ann-Marie Illsley toward obtaining DesZ for kinetic and crystallographic investigations.
Figure 1. Proposed catabolic pathway for the degradation of lignin-derived aromatic compounds in *S. paucimobilis* SYK-6 (adapted from Ref. 22). Enzymes abbreviations: LigD, NAD-dependent β-aryl ether dehydrogenase; LigE and LigF, β-etherase; LigG, glutathionelyase; LigX, DDVA O-demethylase; LigZ, OH-DDVA dioxygenase. LigY, C-C hydrolase; LigW/LigW2, carboxyvanillate decarboxylase; LigC, 4-carboxy-2-hydroxymuconate-6-semialdehyde dehydrogenase; FerA, feruloyl CoA synthetase. FerB/FerB2, feruloyl CoA hydratase/lyase; LigAB, PCA 4,5-dioxygenase; LigM, tetrahydrofolate (H₄folate)-dependent O-demethylase; LigI, PDC hydrolase; LigJ, OMA hydrolase; DesB, gallate dioxygenase; LigK, 4-carboxy-4-hydroxy-2-oxoadipate aldolase/oxaloacetate decarboxylase; DesA, syringate O-demethylase; DesZ, 3MGA 3,4-dioxygenase. Abbreviations: 5CVA, 5-carboxyvanillate DDVA, 5,5′-dehydrodivanillate; CHMS, 4-carboxy-2-hydroxymuconate-6-semialdehyde; CHA, 4-carboxy-4-hydroxy-2-oxoadipate; OH-DDVA, 2,2′-trihydroxy-3′-methoxy-5,5′-dicarboxybi phenyl;
Materials and Methods

BL-21 AI cells were purchased from Invitrogen (Carlsbad, CA). pET-21b vector was obtained from Novagen (EMD4 Biosciences, San Diego, CA). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was bought from Gold Bio Technologies (St. Louis, MO). B-PER II Bacterial Protein Extraction Reagent was purchased from Thermo scientific (Rockford, IL). Ampicilliam (Amp), lysozyme, ferrous ammonium sulfate, sodium hydroxide, EDTA, L-cysteine, tris-(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), sodium dodecyl sulfate (SDS), 3-methoxygallic acid, glycine, Triton X, and manganese sulfate monohydrate were obtained from Sigma (St. Louis, MO). Tris-base and sodium chloride were purchased from Fisher Scientific (Pittsburgh, PA). Glacial acetic acid and isopropanol were purchased from Phamaco-AAPER (Brookfield, CT). 30% acrylamide, ammonium persulfate (APS), Coomassie Brilliant Blue R-250, and N, N, N’ N’-tetra-methylene diamine (TEMED) were obtained from Bio-Rad (Hercules, CA). Bacto tryptone and Bacto yeast extract were obtained from Difco Laboratories (Lawrence, KS). pTOM vector is a modified version of Novagen pET-15b vector in which the two Ser codons both before and after the His-tag were mutated to His codons by Toomas Haller in John Gerlt’s lab, which encodes Amp resistance, an N-terminal 10-His-tag and a thrombin protease cleavage site for the cleavage of the His-tag.

The Luria-Bertani (LB) media was prepared from 5 g/L sodium chloride, 10 g/L
tryptone, 5 g/L yeast extract, and 0.001 M sodium hydroxide. The concentration of Amp in all 
growth media was 100 μg/mL. All liquid growths were performed at 37 °C, with 200 rpm 
shaking speed unless otherwise indicated. The polyacrylamide protein gels were prepared 
using 30 % acrylamide, separating buffer (1.5 M Tris-HCl, pH 8.8), stacking buffer (0.5 M 
Tris-HCl, pH 6.8), 0.01 mM EDTA, APS and TEMED. The protein running buffer was 
prepared using 0.1 % SDS, 192 mM glycine, and 25 mM Tris (pH 8.2). All SDS-PAGE 
protein gels were run at 100 V using a PowerPac Basic power supply from Bio-Rad. Then the 
protein gels were stained using Coomassie Brilliant Blue R-250 for 22 min and destained for 
30 min.

An AKTA FPLC system (GE Healthcare, Piscataway, NJ) was used for protein 
purification. The extinction coefficients and predicted masses of all proteins were calculated 
based on the amino acid sequence using the web application ProtParam 
(http://www.expasy.ch/tools/protparam.html). The cells were burst open using a French press 
cell (Aminco, Urbana, Ill.). Spectrophotometric assays were taken at a Varian Cary 100 Bio 
UV-visible Spectrophotometer (Agilent Technologies, Santa Clara, CA).

**Cloning of DesZ.** The desZ DNA sequence was synthesized by DNA2.0 (Menlo Park, 
CA), and provided in vector pJ201. It was cut out of pJ201 and ligated into pTOM vector 
with the NdeI and BamHI cloning sites by Julie Huang. Later the desZ was removed out of 
pTOM, subcloned into pET-21b in the NdeI and BamHI cloning sites, mutated out the stop
codons at the C-terminal, and transferred into BL-21 AI cells by Ann-Marie Illsley.

**Soluble protein expression screening.** The DesZ-pET21b with stop codons version (shorten as DesZ-pET21b) was transferred into BL-21 AI cells and stored as frozen stock. 100 µL overnight growth media was inoculated into 10 mL LB-Amp media. After the OD$_{600}$ reached 0.6, different concentration combinations of IPTG (1 mM and 0.1 mM) and L-arabinose (0.2 %, 0.02 %, 0.002 %, and 0.0002 %) were added into the growth media to induce the protein production. The incubation temperature was changed to 30 °C. 1 mL aliquot of growth media was removed out at 0 h, 5 h, and 24 h intervals after induction. The aliquots were centrifuged at 11,000 rpm for 10 min at room temperature to save the cell pellets and stored at -20 °C. To extract the soluble proteins, the thawed cell pellets were lysed utilizing 150 µL B-PER II Bacterial Protein Extraction Reagent. After centrifugation at 11,000 rpm for 5 min, supernatants were saved as the soluble fractions. The remaining cell pellets were re-suspended in 150 µL B-PER II Bacterial Protein Extraction Reagent as insoluble fractions. Both the soluble and insoluble proteins were visualized on SDS-PAGE gels and compared. To determine the best growth condition, the soluble and insoluble proteins obtained from the exact same growth conditions except changing the incubation temperature after induction from 30 °C to 37 °C and 16 °C, or adding 0.1 mM Fe(NH$_4$)$_2$SO$_4$ or MnSO$_4$ to induce were compared with each other.

*Expression and Purification of DesZ (DesZ-pET21b 36.5 kDa, DesZ-C-pET21b*
39.0 kDa) 16 mL overnight LB-amp growth of transformed DesZ-pET21b BL-21 AI cells was inoculated into 4 L LB-amp media. After the absorbance of the cell growth at 600 nm reached 0.6, 0.4 mL 1 M IPTG and 20 % L-arabinose were added into the growth media to make the final concentration of IPTG and L-arabinose to 0.1 mM and 0.002 % respectively to induce the cells. After 24 h incubation at 30 °C from induction, the cells were harvested at 6,000 rpm for 15 min. Resulting cell pellets were lysed in buffer consisting of 20 mM Tris, 0.1 mM Fe(NH₄)₂(SO₄)₂ and 10 mM L-cysteine (pH 8.0). The cells were passed three times through a French Press cell, followed by centrifugation at 11,000 rpm for 30 min at 4 ºC to remove the cell debris. The resulting supernatant was purified on an IEX DEAE Sepharose Fast-Flow column (26 × 530 mm, 244 mL). The column was first equilibrated with 3 CV of lysis buffer at 1 mL/min. After the proteins were loaded onto the column, 1.5 CV lysis buffer was used to remove the unbound proteins. The column was eluted with a linear gradient of 0-50 % of eluting buffer (containing 20 mM Tris base, 0.1 mM Fe(NH₄)₂(SO₄)₂, 1 M NaCl, and 10 mM L-cysteine (pH 8.0)) over 6.5 CV, followed by 1.5 CV of 50 % eluting buffer to wash off all the remaining proteins. After the elution, the column was washed by 2 CV of H₂O. 3 CV of 0.5 % Triton-X and 0.1 M acetic acid were utilized to wash off the Fe²⁺ ion at 0.5 mL/min. 3 CV of 30 % isopropanol was adopted to wash off the acid. The column was regenerated with 1 CV of 1 M NaCl, and the extra salt was washed off with 2 CV of H₂O. All the buffers used during this purification were degassed. The fractions containing DesZ were
identified using SDS-PAGE, and combined.

The purification for DesZ C-terminal His-tag version (shorten as DesZ-C-pET21b) was described as below. 8 mL overnight LB-amp growth of DesZ-C-pET21b BL-21 AI cells was inoculated into 3 L LB-amp media. After the absorbance of the cell growth reached 0.6, 0.3 mL 1 M IPTG and 30 mL 20 % L-arabinose were added into the growth media to make the final concentration of IPTG and L-arabinose to 0.1 mM and 0.2 %. 0.1 mM Fe(NH₄)₂(SO₄)₂ was also added to the induced cells. After 24 h incubation at 30 °C with shaking at 145 rpm, the cells were harvested at 6,000 rpm for 15 min. 3 L of non-His-tagged version DesZ-pET21b BL-21 AI cells were combined with the cell pellets of DesZ-C-pET21b, and the combined cell pellets were lysed in the buffer consisting of 20 mM Tris, 0.1 mM Fe(NH₄)₂(SO₄)₂ and 10 mM L-cysteine (pH 7.0) with a small spatula tip worth of lysozyme. The cells were passed three times through a French Press cell and followed by centrifugation at 11,000 rpm for 30 min at 4 °C to remove the cell debris. The resulting supernatant was purified through an IEX DEAE Sepharose Fast-Flow column (26 × 530 mm, 244 mL). The column was first equilibrated with 3 CV of lysis buffer at 1 mL/min. After the proteins were loaded onto the column, 1.5 CV lysis buffer was used to remove the unbound proteins. The column was eluted with a linear gradient of 0-50 % of 6.5 CV eluting buffer (containing 20 mM Tris base, 0.1 mM Fe(NH₄)₂(SO₄)₂, 1 M NaCl, and 10 mM L-cysteine (pH 7.0)), followed by 1 CV of 50 % eluting buffer to wash off all the remaining proteins. After elution,
the column was washed by 2 CV of H₂O. The fractions containing DesZ were identified using SDS-PAGE based on size, confirmed the presence of DesZ by anti-His tag antibody through Western-blot, combined and sent for Mass spectroscopy (36246 g/mol compared with expected mass 36530 g/mol).

**UV-Vis spectroscopy.** Spectrophotometric assays were taken using programs “kinetics” and “scanning kinetics” by a Varian Cary 100 Bio UV-visible Spectrophotometer at room temperature. Right before use, DesZ was buffer exchanged into freshly made buffer (consisting 20 mM Tris base, 0.1 mM Fe(NH₄)₂(SO₄)₂, and 10 mM L-cysteine (pH 8.0)). Each 1 mL kinetics assays was prepared with 2.24 mM 3-methoxygallic acid, 80 μL DesZ (320 μg), DA buffer (50 mM Tris, 50 mM 2-amino-2-methyl-1,3-propanediol, and 50 mM 3,3-dimethylglutarate (pH 8.0)). DesZ was not added into the cuvette until a flat base line was obtained. For scanning kinetics assays, each 2 mL assay mixture contained purified DesZ (about 250 μg), and 50 μM 3-methoxy gallic acid and DA buffer. DesZ wasn’t added into the mixture consisting of buffer and substrate after the first. The spectra were recorded with 21 scans with the range of 200-800 nm at 600 nm/min, and 30 min intervals between each scan.

**Results and Discussion**

As an important member in a catabolic pathway in *S. paucimobilis* SYK-6 for funneling several lignin derived monomeric and dimeric components into the citric acid cycle compounds, this study of DesZ is a continuation of the preliminary research by Ann-Marie
Illsley toward obtaining DesZ for kinetic and crystallographic investigations.

**Soluble protein expression screening.** After the DesZ gene was cloned into the pTOM vector by Julia Huang, it was found that little soluble protein was produced during protein expression screens (data not shown), and what little protein was expressed solubly failed to stick to a nickel affinity column to allow protein purification. Suspecting that the N-terminal His₆ tag might be causing the protein to misfold leading to the solubility problems, the DesZ gene was then subcloned into the pET21b vector, in which the stop codon prevents the His tag production, for new expression screening (Figure 2). Different combinations of temperatures and concentrations of inducing agents (IPTG and arabinose) and cell lines were tried to get the optimal expression conditions. 0.1 mM Fe(NH₄)₂(SO₄)₂ or MnSO₄ were also

![Figure 2. DesZ in pET21b vector expression screening. BL-21 AI cells with DesZ-pET21b were grown at 30 °C and induced by 0.1 mM IPTG and four different concentrations of arabinose (numbers under the red lines) separately. Aliquots were taken out at different time point as shown on top of each lane to check the protein solubility. The band corresponding to the DesZ molecular weight and showing the most soluble portion was circled in red.](image-url)
added at induction to check whether metal ions would help soluble protein expression. It was shown that DesZ was produced most solubly when the cells were grown at 30 °C for 24 h after being induced by 0.1 mM IPTG and 0.002 % arabinose.

**Purification of non-tagged Version of DesZ.** Since there is no His tag attached to DesZ from DesZ-pET21b, the protein was purified through a DEAE Sepharose Fast-Flow column by salt gradient wash (Figure 3). The complicated chromatograph hardly gives hints to suggest the location of DesZ. SDS-PAGE gel analysis of column fractions did not immediately reveal the position of DesZ (Figure 4). Based on the predicted molecular weight of 36.5 kDa for DesZ, there were two possible locations of DesZ, and addition tests were needed to identify where DesZ elutes.

---

**Figure 3.** The chromatograph of non-tagged version of DesZ from DEAE column. Numbers besides the vertical lines represent the fractions. The UV 280 nm reading is the Y axis, and being denoted as the blue curve. The red curve shows the UV 254 nm reading, and the green line displays the supposed salt concentration. The actual salt conductivity is demonstrated as brown line. The column was eluted with a linear gradient of 0-50 % of 6.5 CV eluting buffer containing 20 mM Tris base, 0.1 mM Fe(NO₃)₂SO₄, 1 M NaCl, and 10 mM L-cysteine (pH 7.0), followed by 1 CV of 50 % eluting buffer to wash off all the left proteins.
Figure 4. The SDS-PAGE gels of the non-tagged version of DesZ from DEAE column. The fractions are labeled on top of each lane. L represents standards, and numbers besides the lane “L” in the third gel on the top indicate the corresponding molecular weights. The possible DesZ are circled in red.

**UV-Vis Spectroscopy Characterization.** The purer fractions from the above gels were used to check whether it can react with DesZ’s natural substrate 3MGA, as monitored by UV-Vis spectroscopy (Figure 5). Compared with 3MGA autoxidation, there is an obvious

![UV-Vis Spectroscopy Characterization](image)

Figure 5. A. Time dependent UV-Vis spectrum of 3MGA with DesZ. The curves were determined using 50 mM GTA buffer (pH 8.5) consisting of 50 mM 3,3-dimethylglutarate, 50 mM Tris, 50 mM 2-amino-2-methyl-1,3-propanediol, DesZ (about 20 μg), 100 μM Fe(NH₄)₂(SO₄)₂, and 100 μM 3MGA. DesZ was pre-incubated with Fe(NH₄)₂(SO₄)₂ for 10 min before the assay began, and wasn’t added into the mixture consisting of buffer and substrate until the first scan completed. The reaction was recorded with 21 scans with the range of 200-600 nm at 600 nm/min, with a 30 min interval between scans. B. UV-Vis spectrum of 3MGA autoxidation. The assay concentrations and parameters of B are exactly the same with A, except ddH₂O was used to substitute the enzyme.
intensity increment at 278 nm for the DesZ and 3MGA mixture, which may be the result of protein addition. The increasing differences at 340 nm for both pictures are almost the same. No more other differences suggesting reaction happening could be observed. Therefore, UV-Vis spectra are not evident enough to support the fact that the purified factions could react with 3MGA.

_Purification of His-tagged version of DesZ._ Due to the fact that activity screens of fractions didn’t allow confirmation of the production and location of non-tagged version of DesZ off the column, a new purification method using C-terminal His-tagged DesZ was proposed (N-terminal His-tagged DesZ was not soluble). The stop codons between the DesZ gene and the His tag in DesZ-pET21b were mutated to two serines for the continuing production of the His tag at the C-terminal. The C terminal tagged version of DesZ was over-expressed together with the same of the non-tagged version. The chromatograph was very similar with the one of non-tagged version (Figure 6), giving complicated gel bands (Figure 7). The western-blot using His tag antibody confirmed the appearance and accurate position of the C terminal His tagged DesZ (Figure 8).
Figure 6. The chromatograph of C terminal tagged version of DesZ from the DEAE column. Numbers besides the vertical lines represent the fractions. The UV 280 nm reading is the Y axis, and being denoted as the blue curve. The red curve shows the UV 254 nm reading, and the green line displays the supposed salt concentration. The actual salt conductivity is demonstrated as brown line. The column was eluted with a linear gradient of 0-50 % of 6.5 CV eluting buffer containing 20 mM Tris base, 0.1 mM Fe(NH4)2SO4, 1 M NaCl, and 10 mM L-cysteine (pH 7.0), followed by 1 CV of 50 % eluting buffer to wash off all the left proteins.

Figure 7. The SDS-PAGE gels of the C-terminal tagged version of DesZ from the DEAE column. The fractions are labeled on top of each lane. L represents standards. The possible DesZ bands are circled in red box.
Figure 8. The western-blot of the C-terminal tagged version of DesZ off the DEAD column. The fractions are labeled on top of each lane. L represents standards.

Conclusions and Future Directions.

As an important member in the lignin-derived aromatic compounds degradation pathway in *S. paucimobilis* SYK-6, further efforts to obtain DesZ will have to be pursued. The N-terminal His tag causes solubility issues, while the non-tagged version of DesZ brought purification verification problems. Finally, western-blot noticeably helps confirm the existing of C-terminal His tagged DesZ. We believe that the anaerobic purification of DesZ will significantly increase the activity of DesZ, thus shed light on the fully characterization of DesZ.
Chapter 4  Other putative homologs in the PCAD superfamily

To identify other possible members in the PCAD dioxygenase superfamily, a psi-BLAST search using the LigB protein sequence from *Sphingomonas paucimobilis* SYK-6 as the query sequence has been conducted. LigB homologs have been identified in the genomes of about 1274 different organisms, and some of them have multiple hits in a single organism. From among these thousands of homologs, eight proteins (those not highlighted in yellow in Figure 1) have been selected, and cloned. Three of them have been expressed, and purified (colored in magenta). The selected proteins only have less than 15% pair-wise identity to any previously identified members of the PCAD superfamily. Identification of new superfamily members, especially the ones at the edge of superfamily, will likely reveal members with diverse substrate specificity and will allow further understanding of the overall characteristics of this PCAD superfamily.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. bronchiseptica</em></td>
<td>1</td>
<td>100</td>
<td>48.8</td>
<td>43.2</td>
<td>34.8</td>
<td>35.1</td>
<td>13.3</td>
<td>16.1</td>
<td>13.7</td>
<td>10.5</td>
<td>9.8</td>
</tr>
<tr>
<td><em>P. syringae</em></td>
<td>2</td>
<td>48.8</td>
<td>100</td>
<td>47</td>
<td>39</td>
<td>38.2</td>
<td>12.8</td>
<td>14.9</td>
<td>16</td>
<td>12.6</td>
<td>14.3</td>
</tr>
<tr>
<td><em>C. violaceum</em></td>
<td>3</td>
<td>43.2</td>
<td>47</td>
<td>100</td>
<td>38.6</td>
<td>37.1</td>
<td>14.1</td>
<td>14.8</td>
<td>12.7</td>
<td>9.9</td>
<td>10</td>
</tr>
<tr>
<td><em>B. cereus</em></td>
<td>4</td>
<td>34.8</td>
<td>39</td>
<td>38.6</td>
<td>100</td>
<td>37.8</td>
<td>11.1</td>
<td>13.7</td>
<td>12.4</td>
<td>11.8</td>
<td>11.5</td>
</tr>
<tr>
<td><em>E. coli K12 YgiD</em></td>
<td>5</td>
<td>35.1</td>
<td>38.2</td>
<td>37.1</td>
<td>37.8</td>
<td>100</td>
<td>9.6</td>
<td>13.9</td>
<td>12.6</td>
<td>10.7</td>
<td>12</td>
</tr>
<tr>
<td>DesZ</td>
<td>6</td>
<td>13.3</td>
<td>12.8</td>
<td>14.1</td>
<td>11.1</td>
<td>9.6</td>
<td>100</td>
<td>22.7</td>
<td>22.4</td>
<td>11.3</td>
<td>12.1</td>
</tr>
<tr>
<td>DesB</td>
<td>7</td>
<td>16.1</td>
<td>14.9</td>
<td>14.8</td>
<td>13.7</td>
<td>13.9</td>
<td>22.7</td>
<td>100</td>
<td>38.3</td>
<td>11.8</td>
<td>9.6</td>
</tr>
<tr>
<td>LigB</td>
<td>8</td>
<td>13.7</td>
<td>16</td>
<td>12.7</td>
<td>12.4</td>
<td>12.6</td>
<td>22.4</td>
<td>38.3</td>
<td>100</td>
<td>11.6</td>
<td>9.4</td>
</tr>
<tr>
<td><em>S. solfataricus</em></td>
<td>9</td>
<td>10.5</td>
<td>12.6</td>
<td>9.9</td>
<td>11.8</td>
<td>10.7</td>
<td>11.3</td>
<td>11.8</td>
<td>11.6</td>
<td>100</td>
<td>70.2</td>
</tr>
<tr>
<td><em>M. sedula</em></td>
<td>10</td>
<td>9.8</td>
<td>14.3</td>
<td>10</td>
<td>11.5</td>
<td>12</td>
<td>12.1</td>
<td>9.6</td>
<td>9.4</td>
<td>70.2</td>
<td>100</td>
</tr>
<tr>
<td><em>S. acidocaldarius</em></td>
<td>11</td>
<td>10.9</td>
<td>14.6</td>
<td>9.5</td>
<td>12.7</td>
<td>11.9</td>
<td>10.5</td>
<td>8.7</td>
<td>11.6</td>
<td>68.2</td>
<td>66.2</td>
</tr>
</tbody>
</table>

Figure 1. Pair-wise sequence identity determined for comparison of PCA dioxygenase superfamily members. Previously identified enzymes (indicated by yellow) are included for comparison to new putative superfamily members with no known functions. Enzymes which are cloned are highlighted in green; enzymes which have been purified are highlighted in magenta. The pair-wise identity of proteins in this project to the three previously identified PCA dioxygenases superfamily members is indicated by the black box.
An amino acid sequence alignment was performed using CLUSTALW, to explore sequence similarity of these putative dioxygenases (Figure 2). Instead of the consensus 2-His-1-carboxylate facial triad for metal binding, a three histidine metal coordination sphere is found in the active site of the selected enzymes, similar to the motif found in some Cupin

![Sequence Alignment](image)

Figure 2. Sequence alignment with selected LigB homologs and two structurally characterized enzymes---LigB and YgiD. Residues highlighted in red are ligands coordinated with metal ions (for both LigB and YgiD), in green are other conserved residues in the active site of LigB and YgiD, in blue are 60-80 % conserved in LigB superfamily provided by Pfam,(129) in yellow are 90 % conserved in LigB superfamily,(130) and in magenta are highly conserved in this alignment.
superfamily members. Also two histidines and one leucine (green highlighted in Figure 2) which are conserved residues in the active sites of LigB and YgiD may help in substrate binding. A consensus sequence GXPXLA (yellow highlighted) has been found, whose function is yet unknown, but be believed to help in new members identification in sequence alignment. Overall, sequences of orthologous from \textit{M. sedula}, \textit{S. acidocaldarius}, and \textit{S. solfataricus} are more similar to each other than counterparts from other strains, likely due to the fact that they are all Archaea.

The rest of the selected organisms are bacteria, none of these gene are in a context surrounded by genes related to lignin degradation. It’s reasonable to deduce that these PCAD superfamily members may have more diverse functions. \textit{Bordetella bronchiseptica} RB50 (\textit{B. bronchiseptica}) is a small Gram-negative bacterium that causes chronic respiratory infections in a wide range of animals like pigs, and rabbits, but rarely infects humans.(131) Current research involving \textit{B. bronchiseptica} are primarily focused on vaccine development.(132)

\textit{Pseudomonas syringae} pv tomato DC3000 (\textit{P. syringae}) is a Gram-negative plant pathogen that causes bacterial speck on tomato, and is an important model in molecular plant pathology.(133) Bioavailable iron for plants was shown to be strongly associated with a variety of pathways in this bacterium. Many previously uncharacterized genes were believed to be related with iron regulation.(134)
Bacillus cereus ATCC 10987 (B. cereus) is a non-lethal dairy Gram-positive bacterium which was isolated from a study on cheese spoilage. Some unique metabolic mechanism capabilities like urease and xylose have been identified, but the bacterium can’t utilize nitrite and nitrate as other B. cereus substrains.

Chromobacterium violaceum ATCC 12472 (C. violaceum) is a gram-negative bacterium that grows in soil and water. It’s found that human and animal infection caused by C. violaceum is rare, but it is associated with a high mortality rate when it happens. It characteristically produces violacein, a water-insoluble purple pigment with antibacterial activity. It’s also has potential for use in the production of other compounds of biotechnological and medical interest such as cyanide and chitinase.

Sulfolobus solfataricus (S. solfataricus), Sulfolobus acidocaldarius (S. acidocaldarius), and Metallosphaera sedula (M. sedula) are hyperthermophilic microorganisms belonging to the Archaea. Hyperthermophilic Archaea are worthy of attention because of their capacity to survive and reproduce at temperatures near the boiling point of water, and are of extreme biotechnological interest not only because of the exceptional stability of their proteins, but also because of their unique enzymatic activities. Many genes found within this domain are unique to Archaea.

Based on above information, all of the selected bacteria are sources of disease caused in human, animals or plants. None of the organisms, including the three Archaea, are
obviously related to lignin degradation. The putative dioxygenases from these bacteria are likely to utilize various substrates, or catalyze different types of reactions. Therefore, the characterization of these remote homologs may reveal various functional and structural diversities, and gain more insight into the structural and mechanism consensus of the whole PCAD superfamily, thus clarifying the boundaries.

Materials and Methods

All the genomic DNA and bacterial cells lines were purchased from ATCC (Manassas, VA). All the primers (Table 1), *E. coli* strains DH5α, BL-21, and BL-21 AI and AccuPrime pfx SuperMix polymerase were purchased from Invitrogen (Carlsbad, CA). *NdeI, BamHI, XhoI, DpnI, Taq* DNA Polymerase with Standard *Taq* Buffer and Quick Ligation Kit were obtained from New England Biolabs (Ipswich, MA). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was bought from Gold Bio Technologies (St. Louis, MO). DNAzol was obtained from MRC Inc. (Cincinnati, OH). B-PER II Bacterial Protein Extraction Reagent was purchased from Thermo scientific (Rockford, IL). Ampicillin (Amp), L-arabinose, tris-(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), ferrous ammonium sulfate, sodium hydroxide, EDTA, ethidium bromide, and sodium dodecyl sulfate (SDS) were obtained from Sigma (St. Louis, MO). Sodium biphosphate, imidizol, and sodium chloride are purchased from Fisher Scientific (Pittsburgh, PA). Agarose and Shrimp Alkaline Phosphatase (SAP) were obtained from Promega (Madison, WI). 30% acrylamide,
ammonium persulfate (APS), Coomassie Brilliant Blue R-250, and N, N, N, N’-tetra-methylethylenediamine (TEMED) were obtained from Bio-Rad (Hercules, CA). Bacto tryptone, tryptic soy broth, nutrient broth and Bacto yeast extract were obtained from Difco Laboratories (Lawrence, Ks). Qiaquick Gel Extraction kit and QIAprepSpin Mini-prep Kit were obtained from Qiagen (Valencia, CA). QuikChange Lightning Site-Directed Mutagenesis Kit was purchased from Agilent technology (Santa Clara, CA). All DNA sequencing results were obtained from ATCG (Wheeling, IL).

pET-15b and pET21b vectors were obtained from Novagen (Madison, WI). pTOM vector is a modified version of Novagen pET-15b vector in which the two Ser codons both before and after the His-tag were mutated to His codons by Toomas Haller in John Gerlt’s lab in Biochemistry Department at University of Illinois at Urbana-Champaign. The modified pTOM vector has an Ndel/XhoI/BamHI cloning site and encodes Amp resistance, an N-terminal 10-His-tag and a thrombin protease cleavage site for the cleavage of the His-tag. pET-15b and pTOM vector are both digested with the same endonucleases with corresponding inserts’ restriction sites, following a Shrimp Alkaline Phosphatase (SAP) at 37 °C for 1 h.

The Luria-Bertani (LB) media was prepared from 5 g/L sodium chloride, 10 g/L tryptone, 5 g/L yeast extract, and 0.001 M solidum hydroxide. The super broth media was prepared using 5 g/L sodium chloride, 32 g/L tryptone, 20 g/L yeast extract, and 0.001 M
sodium hydroxide. The SOC media is made of 20 g/L Bacto Tryptone, 5 g/L Bacto Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose. The concentration of Amp in all growth media was 100 μg/mL. All liquid growths were performed at 37 °C, with 200 rpm shaking speed unless otherwise indicated. The 1 % agarose gel for electrophoresis was prepared using 10 mg/mL agarose, 50 mL TAE electrophoresis buffer (including 40 mM Tris, 20 mM glacial acetic acid, and 1 mM EDTA (pH 8.0)), and 0.1 mg/mL ethidium bromide. The polyacrylamide protein gels were prepared using 30 % acrylamide, separating buffer (1.5 M Tris-HCl, pH 8.8), stacking buffer (0.5 M Tris-HCl, pH 6.8), 0.01 mM EDTA, APS and TEMED. The protein running buffer was prepared using 0.1 % SDS solution. All DNA agarose gels and SDS-PAGE protein gels were run at 100 V using a PowerPac Basic power supply from Bio-Rad. Then the protein gels were stained using Coomassie Brilliant Blue R-250 for 22 min and destained for 1 h.

The gradient PCR reactions were performed using a MJ Reserch PTC-200 Peltier Thermal Cycler and (Bio-rad, Hercules, CA). The colony PCR reactions were utilized AB Applied Biosystems Gene Amp PCR system 9700 (Life Technologies, grand Island, NY) or a S1000 Thermal Cycler. The extinction coefficients and predicted masses of all proteins were calculated based on the amino acid sequence using the web application ProtParam (http://web.expasy.org/protparam/). The cells were lysed using a French press cell (Aminco, Urbana, Ill.). The proteins from Chromobacterium violaceum ATCC 12472 and Pseudomonas
*Pseudomonas syringae* DC3000 were purified using an AKTA FPLC system.

**Pseudomonas syringae** DC3000 (PSPT_1776)

*Cloning of putative PSPT_1776 dioxygenase (705 b.p.)*. The genomic DNA of *P. syringae* was purchased from ATCC, dissolved in 10 µL H₂O, and stored in a 40 % glycerol solution at -80 °C. A gradient PCR using 22.5 µL AccuPrime SuperMix, 1 µL genomic DNA (38.4 ng), 0.75 µL primer (750 ng each, Table 1) was adopted to isolate the putative dioxygenase. The reaction mixture was incubated at 95 °C for 5 min to denature the DNA, followed by 35 cycles consisting of 95 °C for 15 s, 45-65 °C for 30 s for annealing, and 68 °C for 1 min for elongation, finally 68 °C for 5 min after the final cycle. The PCR products at the right size in the 1 % Agarose gel were isolated to extract the DNA by following the Qiaquick Gel Extraction kit protocol. The gel extraction product was incubated overnight with *NdeI* endonuclease at 37 °C with the following recipe: 1 µg PCR product, 2 µL *NdeI* (40 U), and 3 µL 10X NEB buffer 3 for a total volume of 30 µL. The *NdeI* digested products were followed by 2 h incubation with 2 µL *BamHI* (40 U), 0.6 µL 10X NEB buffer 3, and 0.36 µL BSA to make a total volume of 36 µL. The digested products were checked by 1 % Agarose gel and purified utilizing Qiaquick PCR Purification Kit. The endonuclease digestion product was ligated successfully with pET-15b and pTOM vectors via Quick Ligation Kit with a 3 to 1 insert to vector concentration ratio, and then transformed into DH5α cells.

The LB-Amp plates were inoculated with the transformation products and incubated
overnight at 37 °C. A single colony from the plates was used to inoculate 20 mL LB-Amp media. After overnight growth, 4 μL culture was used for colony PCR using 5 μL 10X standard Taq reaction buffer, 1 μL 10 mM dNTPs, 1 μL each T7 promoter and terminator (40 ng), 1 μL growth media, 0.25 μL Taq DNA polymerase to make a total volume of 50 μL to check the insert’s size. The remaining cells were harvested by centrifugation at 4000 rpm for 20 min. The plasmid DNA was extracted via QIAprepSpin Mini-prep Kit. Mini-prep protocol. 10 μL products were sent to sequencing. Another 5 μL pET-15b and pTOM-gene constructs (ps-pET15b and ps-pTOM respectively) were transformed into BL-21 AI cells. The transformation product was plated onto LB-Amp plates and incubated overnight at 37 °C. A single colony from the plates was used to inoculate in 20 mL LB-Amp media to make frozen stock.

**Soluble protein expression screening.** 200 μL LB-Amp overnight growth media of ps-pET15b was used to inoculate 20 mL Super Broth-Amp media. After incubation for about 4 h at 37 °C, the absorbance of the new growth at OD_{600} reached 0.4375. Different concentration combinations of IPTG (1 mM and 10 mM) and L-arabinose (0.2 %, 0.02 %, 0.002 %, and 0.0002 %) were added into the growth media to induce the protein production. Then 1 mL aliquots were removed at 0, 5, and 24 h after induction. The aliquots were centrifuged down at 11,000 rpm for 10 min at room temperature and the resulting cell pellets were stored at -20 °C. To extract the soluble proteins, the thawed cell pellets were lysed
utilizing 150 µL B-PER II Bacterial Protein Extraction Reagents. After centrifuged at 11 k rpm for 5 min, supernatants were removed from the cell pellets. The cell pellets were suspended in 150 µL B-PER II Bacterial Protein Extraction Reagents to allow analysis of the insoluble fraction. Both the soluble and insoluble protein fractions were visualized on SDS-PAGE gels and compared. To determine the best growth condition, the soluble and insoluble protein fractions were analyzed from growth with 37 °C, 30 °C or 16 °C incubation after induction.

Expression and purification of putative PSPT_1776 dioxygenase (29.9 kDa). 4 mL overnight superbroth-amp growth of ps-pET15b BL-21 AI cells frozen stock was inoculated into 4 L superbroth-amp medium with 0.1 mM Fe(NH₄)₂(SO₄)₂. After the absorbance of the cell growth reached 0.6, 40 mL of 1 M IPTG and 0.4 mL of 20 % L-arabinose were added into the growth media to make the final concentrations of IPTG and L-arabinose to 10 mM and 0.002 % to induce the cells. The cells were harvested by centrifugation at 5,000 rpm for 15 min after 24 h growth at 16 °C from induction. Resulting cell pellets were lysed in FE buffer (consisting of 50 mM Tris-HCl, 0.1 mM Fe(NH₄)₂SO₄, 10 % glycerol and 2 mM L-cysteine (pH 8.0)). The cells were passed three times through a French Press cell followed by centrifugation at 11,000 rpm for 30 min at 4 °C to remove the cell debris. The resulting supernatant was loaded onto an IMAC Sepharose Fast-Flow column (26 × 60 mm, 44 mL) charged with Ni²⁺. The column was first equilibrated with 2 CV of Fe buffer at 2 mL/min.
After the proteins were loaded onto the column, 2 CV of binding buffer (containing 20 mM Hepes, 0.5 M NaCl, and 20 mM imidazole) was used to remove the unbound stuff. The column was eluted with a linear gradient of 0-50 % of 25 CV of eluting buffer (containing 20 mM Hepes, 0.5 M NaCl, and 0.5 M imidazole (pH 8.0)), followed by 6 CV of 100 % eluting buffer to wash off all the remaining proteins. After elution, the column was washed by 6 CV of stripping buffer (consisting of 0.2 mM EDTA and 0.5 M NaCl), followed by 6 CV of H₂O. The column was consequently washed with 2 CV of 2 M Urea, 1 CV of H₂O at 1 mL/min, and 6 CV of H₂O at 2 mL/min. 2 CV of 0.5 % Triton-X and 0.1 M acetic acid were utilized to wash off the Fe²⁺ ion. 4 CV of 80 % ethanol was used to wash off the acid at 1 mL/min. Then the column was washed with 2 M NaCl at 0.5 mL/min, followed by 1 CV H₂O at 1 mL/min, and 6 CV of H₂O at 2 mL/min. 2 CV of 30 % isopropanol was adopted to clean the column, followed by 1 CV of H₂O at 1 mL/min, and 6 CV of H₂O at 2 mL/min. The column was regenerated with 3 CV of 0.2 M NiSO₄, and washed off the extra metal ion with 6 CV of binding buffer. The fractions containing the target protein were identified using SDS-PAGE, and combined.

The size exclusion chromatography was performed using a HiLoad 16/60 Superdex 200 prepgrad column (16 × 600 mm, 120 mL) in Prof. Rich Olson’s lab at Wesleyan University. Before loading the protein samples, the column was washed with 40 mL degassed deionized water at 1 mL/min, 45 mL degassed 20 % ethanol at 1.5 mL/min, and equilibrated
with 200 mL degassed equilibration buffer (containing 20 mM Tris (pH 7.4), and 150 mM NaCl). 1 mL of about 2 mg/mL Ni affinity column purified PSPT_1776 dioxygenase was spun at 15,000 rpm for 10 min to remove aggregates, filtered, and loaded onto the column with a superloop. The column was eluted with 30 mL equilibration buffer. The identified fractions were combined, and subjected to MALDI mass analysis (29260 g/mol compare with the expected mass 29911.8 g/mol).

**Bacillus cereus ATCC 10987(BCE_1944)**

**Cloning of putative BCE_1944 dioxygenase (757 b.p.)** After rehydration with 1 μL Tryptic Soy Broth (TSB) medium made from dissolving 30 g TSB powder into 1 L H₂O, the B. cereus cells were grown on a TSB plate at 30 °C for 24 h, followed by liquid overnight growth in 20 mL TSB, stored in a 40 % glycerol solution in -80 °C, and then inoculated into 500 mL TSB for 24 h. Cells were harvested by centrifugation at 4,000 rpm for 20 min. The genomic DNA was extracted by following DNAZol protocol. Then a phenol-chloroform extraction (phenol: chloroform: isoamylalcohol= 25:24:1) was adopted to purify 200 μL of genomic DNA. Gradient PCR was conducted for gene amplification using using 22.5 μL AccuPrime SuperMix, 1 μL genomic DNA (212 ng), 0.75 μL primer (750 ng each, Table 1). The reaction mixture was incubated at 95 °C for 5 min to denature the DNA, followed by 35 cycles consisting of 95 °C for 15 s, 45-65 °C for 30 s for annealing, and 68 °C for 1 min for elongation, finally 68 °C for 5 min after the final cycle. The PCR products at the right size in
the 1 % Agarose gel were isolated to extract the DNA by following the Qiaquick Gel
Extraction kit protocol.

Table 1 Sequences of the oligonucleotides used in the study. Restriction sites are underlined. The
mutation were indicated in red.

<table>
<thead>
<tr>
<th>Oligo name</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P. syringae</strong></td>
<td>Forward: GGAGCCC<strong>CATATG</strong>TTTCCCAGC</td>
</tr>
<tr>
<td></td>
<td>Reverse: CTGGATCTCAGTCGAACCGATAG</td>
</tr>
<tr>
<td><strong>B. cereus</strong></td>
<td>Forward: GGAGGTT<strong>CTCGAG</strong>ATGATGCCATCAC</td>
</tr>
<tr>
<td></td>
<td>Reverse: CATCCTTG<strong>GATC</strong>CTTAAAATGGGAGACAAAGATAAC</td>
</tr>
<tr>
<td><strong>C. violaceum</strong></td>
<td>Forward: GCAGGCAAACAGCATATGAC CACCC</td>
</tr>
<tr>
<td></td>
<td>Reverse: GCGGATCCTCAGTCGAAGCGCAG</td>
</tr>
<tr>
<td><strong>S. solfataricus</strong></td>
<td>Forward: GAGATGAGGCTTTG<strong>CATATG</strong>AGGAGATTACCAG</td>
</tr>
<tr>
<td></td>
<td>Reverse: GAAACG<strong>GATCC</strong>TTAACCTCACAACCGCAGCTGC</td>
</tr>
<tr>
<td>DesB</td>
<td>Forward: CGGTAGCCGGAACAGCT<strong>CCT</strong>CTGATCCGAATT</td>
</tr>
<tr>
<td></td>
<td>Reverse: ACTGTTTGGCCGGCTACCG</td>
</tr>
<tr>
<td>QuikChange</td>
<td>Forward: CGAGCTCCGT CGACAAGC</td>
</tr>
<tr>
<td></td>
<td>GCTTGTCGACGGAGCTCGAATTCGGATC<strong>AG</strong>AGG</td>
</tr>
<tr>
<td>T7 promoter</td>
<td>TAATACGACTCACTATAGG</td>
</tr>
<tr>
<td>T7 terminator</td>
<td>GCTAGTTATTTGCTCAGCGG</td>
</tr>
</tbody>
</table>

**Chromobacterium violaceum ATCC 12472 (CV3550)**

**Cloning of putative CV3550 dioxygenase (767 b.p.).** After the rehydration with 1 μL
Nutrient broth media made from dissolving 8 g nutrient broth powder in 1 L H₂O, the
CV3550 cells were grown on a Nutrient broth plate at 26 °C for 24 h, followed by liquid
overnight growth in 20 mL nutrient broth, stored in a 40 % glycerol solution in -80 °C, and then inoculated into 500 mL nutrient broth media for 24 h. Cells were harvested by centrifugation at 4000 rpm for 20 min. The genomic DNA was extracted by following DNAZol protocol. Then a phenol-chloroform extraction (phenol: chloroform: isoamylalcohol= 25:24:1) was adopted to purify 200 μL of genomic DNA. Gradient PCR was conducted for gene amplification using 22.5 μL AccuPrime SuperMix, 1 μL genomic DNA (106 ng), 0.75 μL primer (750 ng each, Table 1). The reaction mixture was incubated at 95 °C for 5 min to denature the DNA, followed by 35 cycles consisting of 95 °C for 15 s, 45-65 °C for 30 s for annealing, and 68 °C for 1 min for elongation, finally 68 °C for 5 min after cycles end. The PCR products at the right size in the 1 % Agarose gel were isolated to extract the DNA by following the Qiaquick Gel Extraction kit protocol. The gel extraction product was incubated overnight with XhoI endonuclease at 37 °C with the following recipe: 1 μg PCR product, 2 μL XhoI (40 U), 3 μL 10X NEB buffer 4, and 0.3 μL BSA for a total volume of 30 μL. The XhoI digested products were checked by 1 % Agarose gel and purified utilizing Qiaquick Gel Extraction Kit. The extracted DNA was digested with BamHI by 2 h incubation at 37 °C with exact the same recipe with XhoI except substituting 10X NEB buffer 4 with buffer 3. The endonuclease digestion product was ligated successfully into pTOM vector via Quick Ligation Kit with a 3 to 1 insert to vector concentration ratio, and then transformed into DH5α cells.
The LB-Amp plates were inoculated with the transformation products and incubated overnight at 37 °C. A single colony from the plates was used to inoculate 20 mL LB-Amp media. After overnight incubation, 4 μL culture total was used for colony PCR using 5 μL 10X standard Taq reaction buffer, 1 μL 10 mM dNTPs, 1 μL each T7 promoter and terminator (40 ng), 1 μL growth media, and 0.25 μL Taq DNA polymerase to make a total volume of 50 μL to check the insert’s size. The remaining cells were harvested by centrifugation at 4000 rpm for 20 min. The plasmid DNA was extracted via QIAprepSpin Mini-prep Kit. 10 μL Mini-prep products were sent to sequencing. Another 5 μL pTOM-gene constructs (shorten as cv-pTOM) were transformed into BL-21 AI cells. The transformation product was plated onto LB-Amp plates and incubated overnight at 37 °C. A single colony from the plates was used to inoculate in 20 mL LB-Amp media to make frozen stock.

**Soluble protein expression screening.** 200 μL LB-Amp overnight growth medium of cv-pTOM in BL-21 cells was used to inoculate 20 mL LB-Amp media. After incubation for about 4 h at 37 °C, the absorbance of the new growth at OD$_{600}$ reached 0.4153, at which point IPTG was added to a final concentration of 1 mM or 10 mM to induce the protein production. Then 1 mL aliquots were removed at 0, 5, and 24 h after induction. The aliquots were centrifuged down at 11,000 rpm for 10 min at room temperature and the resulting cell pellets were stored at -20 °C. To extract the soluble proteins, the thawed cell pellets were lysed utilizing 150 μL B-PER II Bacterial Protein Extraction Reagents. After centrifugation at
11,000 rpm for 5 min, supernatants were removed from the cell pellets. The cell pellets were suspended in 150 µL B-PER II Bacterial Protein Extraction Reagents to allow analysis of the insoluble fraction. Both the soluble and insoluble protein fractions were visualized on SDS-PAGE gels and compared. To determine the best growth condition, the soluble and insoluble protein fractions were analyzed from growth at 30 °C or 16 °C in LB-Amp meida, 16 °C in Super Broth-Amp medium, and 16 °C in Super Broth-Amp media with 0.5 mM Fe(NH$_4$)$_2$(SO$_4$)$_2$ incubation after induction. To improve the yield of soluble protein, the cv-pTOM construct was also transformed into BL-21 AI cells and the same protein screening process was repeated with different concentrations combinations of IPTG (1 mM and 10 mM) and L-arabinose (0.2, 0.02, 0.002, and 0.0002 % by molecular weight).

*Expression and purification of putative CV3550 dioxygenase (31.4 kDa).* 1 mL overnight superbroth-amp growth of cv-pTOM BL-21 AI cells frozen stock was inoculated into 1 L superbroth-amp media with 0.5 mM Fe(NH$_4$)$_2$(SO$_4$)$_2$. After the absorbance of the cell growth reached 0.6, 1 mL of 1 M IPTG and 20 % L-arabinose were added into the growth media to make the final concentrations of IPTG and L-arabinose to 1 mM and 0.02 % to induce the cells. The cells were harvested by centrifugation at 5 k rpm for 15 min after 24 h growth at 16 °C from induction. The resulting supernatant was loaded onto an IMAC Sepharose Fast-Flow column (26 × 60 mm, 44 mL) charged with Ni$^{2+}$. The column was first equilibrated with 2 CV of Fe buffer at 2 mL/min. After the proteins were loaded onto the
column, 2 CV of binding buffer (containing 20 mM Hepes, 0.5 M NaCl, and 20 mM imidazole) was used to remove the unbound stuff. The column was eluted with a linear gradient of 0-50 % of 25 CV eluting buffer (containing 0.5 M imidazole (pH 8.0), 20 mM Hepes, and 0.5 M NaCl), followed by another 6 CV of 100 % eluting buffer to wash off all the remaining proteins. After elution, the column was washed by 6 CV of stripping buffer (consisting of 0.2 mM EDTA and 0.5 M NaCl), followed by 6 CV of H2O. The column was consequentially washed with 2 CV of 2 M Urea, 1 CV of H2O at 1 mL/min, and 6 CV of H2O at 2 mL/min. 2 CV of 0.5 % Triton-X and 0.1 M acetic acid were utilized to wash off the Fe2+ ion. 4 CV of 80 % ethanol was adopted to wash off the acid at 1 mL/min. Then the column was washed with 2 M NaCl at 0.5 mL/min, followed by 1 CV H2O at 1 mL/min, and 6 CV of H2O at 2 mL/min. 2 CV of 30 % isopropanol was adopted to clean the column, followed by 1 CV of H2O at 1 mL/min and 6 CV of H2O at 2 mL/min. The column was regenerated with 3 CV of 0.2 M NiSO4, and washed off the extra metal ion with 6 CV of binding buffer. The fractions containing the target protein were identified using SDS-PAGE, combined, and subjected to MALDI mass analysis (31175.9 g/mol compared with the expected mass 31404.5 g/mol).

*Sulfolobus solfataricus P2* (SSO0066)

Cloning of putative SSO0066 dioxygenase (850 b.p.). The genomic DNA of *S. solfataricus* was purchased from ATCC, dissolved in 10 µL H2O, and stored in a 40 %
glycerol solution in -80 °C. A gradient polymerase chain reaction (gradient PCR) using 22.5 μL AccuPrime SuperMix, 1 μL genomic DNA (38.4 ng), 0.75 μL primer (750 ng each, Table 1) was adopted to isolate the putative dioxygenase. The reaction mixture was incubated at 95 °C for 5 min to denature the DNA, followed by 35 cycles consisting of 95 °C for 15 s, 45-65 °C for 30 s for annealing, and 68 °C for 1 min for elongation, finally 68 °C for 5 min after cycles end. The PCR products at the right size in the 1 % Agarose gel were isolated to extract the DNA by following the k Gel Extraction kit protocol.

**DesB**

**Cloning of DesB.** The desB DNA sequence was synthesized by DNA2.0 (Menlo Park, CA), and provided in vector pJ201. It was cut out of pJ201 and ligated into pTOM vector with the NdeI and BamHI cloning sites by Julie Huang. Later the DesB was removed out of pTOM, subcloned into pET-21b in the NdeI and BamHI cloning sites, and transferred into BL-21 AI cells by Ann-Marie Illsley.

**Soluble protein expression screening.** The DesB-pET21b with stop codons version (shorten as DesB-pET21b) was transferred into BL-21 AI cells and stored in a 40 % glycerol solution in -80 °C. 100 μL overnight growth media was inoculated into 10 mL LB-Amp media. After the OD$_{600}$ reaches 0.6, different concentration combinations of IPTG (1 mM and 0.1 mM) and L-arabinose (0.2 %, 0.02 %, 0.002 %, and 0.0002 %) were added into the growth media to induce the protein production. The incubation temperature was changed to
30 °C. Then 1 mL aliquots were removed at 0, 5, and 24 h after induction. The aliquots were centrifuged down at 11,000 rpm for 10 min at room temperature and the resulting cell pellets were stored at -20 °C. To extract the soluble proteins, the thawed cell pellets were lysed utilizing 150 µL B-PER II Bacterial Protein Extraction Reagents. After centrifugation at 11,000 rpm for 5 min, supernatants were removed from the cell pellets. The cell pellets were suspended in 150 µL B-PER II Bacterial Protein Extraction Reagents to allow analysis of the insoluble fraction. Both the soluble and insoluble protein fractions were visualized on SDS-PAGE gels and compared. To determine the best growth condition, the soluble and insoluble protein fractions were analyzed from growth at 30 °C or 16 °C, additionally adding 0.1 mM Fe(NH₄)₂(SO₄)₂ or MnSO₄ after induction were compared with each other.

*Site-directed mutagenesis of DesB-pET21b to generate C-terminal His₆ tagged proteins.* A 50 µL of PCR reaction was prepared utilizing 5 µL 10X reaction buffer, 125 ng of each primer, 1 µL of dNTP mix, 1.5 µL of QuikSolution reagent, 100 ng DesB-pET21b, and 1µL of QuikChange Lightning Enzyme. The PCR reaction mixture was incubated at 95 °C for 2 min, followed by 18 cycles containing 20 s at 95 °C, 10 s at 60 °C for annealing and 3.5 min at 68 °C for elongation, finally at 68 °C for 5 min after the final cycle. Annealing temperature at 55 and 65 °C had been tried with other same conditions. A control reaction using pWhitescript 4.5-kb control plasmid also performed according to the manual. 2 µL of DpnI was directly added to the PCR reaction mixtures, and incubated at 37 °C for 1 h. 10 µL
of digestion products was transferred into DH5α cells or XL 10-Gold cells, and plated on LB-Amp plates. Four colonies were selected for overnight growth at 37 °C in 15 mL LB-Amp medium. Cell pellets were harvested at 4,000 rpm for 15 min. Plasmids were extracted from cell pellets using QIAprepSpin Mini-prep Kit, and sent for sequencing.

Results and Discussion

To identify possible members in the PCA superfamily, a PSI-BLAST search using the LigB protein sequence has been conducted. Five proteins have been selected for cloning to obtaining proteins for further kinetic and crystallographic investigation.

**Pseudomonas syringae DC3000 (PSPT_1776)**

The gradient PCR was used to clone the designed target from the genomic DNA using the forward and reverse primers shown in Table 1. DNA gel analysis revealed that the molecular mass of this product was similar to the predicted molecular mass of the putative dioxygenase PSPT_1776 (705 b.p.) (Figure 3). The gene was subcloned into pET-15b and pTOM vectors. The successful ligations were confirmed by the bands around 1000 b.p. in the 1 % agarose gel after colony PCR (Figure 4), due to the 200 b.p.between T7 promoter and terminator regions. All cloned gene constructs were confirmed through sequencing.
Figure 3. Gradient PCR result of PSPT_1776 on a 1% Agarose gel. The left column is a DNA ladder with the corresponding size label. The estimated size of PSPT_1776 is 705 b. p.

Figure 4. Colony PCR product of PSPT_1776. The left column is DNA ladder besides with corresponding size label.

After the PSPT_1776 gene was cloned into the pET15b vector, and transformed into BL-21 cells, different combinations of temperatures and IPTG concentrations and were screened to reveal the optimal expression conditions. Unfortunately, it was found that little soluble protein was produced during protein expression screens (data not shown). So the ps-pET15b was re-transformed into a BL-21 AI cells. Different combinations of IPTG and arabinose concentrations were tried for induction, in addition to different temperatures (Figure 5). It was shown that the most soluble protein was produced when the cells were grown at 30 °C for 24 h after induced by 0.1 mM IPTG and 0.002 % arabinose.
Figure 5. *ps*-pET15b expression screening. BL-21 Al cells with *ps*-pET15b were grown at 16 °C in Super Broth-Amp media, and induced by 10 mM IPTG and four different concentrations (numbers above the lines) of arabinose separately. Aliquots were taken out at different time point as shown on top of each. The band corresponding to the *ps*-PET15b molecular weight was circled in red. The most soluble protein growth condition was in red, too.

Successful purification by immobilized nickel affinity chromatography afforded about 18 mg purified proteins per liter of culture. The enzyme was confirmed by SDS-PAGE gels corresponding to about 30 kDa (Figure 6). The size exclusion column chromatography displayed several peaks, suggesting that either the desired protein was not pure, or multiple oligomeric states existing (Figure 7).

Figure 6. The SDS-PAGE gels of *ps*-pET15b purified using a Sepharose Fast-Flow column charged with Ni²⁺. The fractions are labeled on top of each lane. Std represents standards, and numbers besides the standard lane indicate the corresponding molecular weights. The expected proteins are in red box.
**Bacillus cereus ATCC 10987 (BCE_1944)**

The gradient PCR was exploited to clone the expected gene from genomic DNA using the forward and reverse primers shown in Table 1. DNA gel analysis revealed single bands at the predicted DNA size of the putative dioxygenase *BCE_1944* (757 b.p.) (Figure 8). The gene was subcloned into the pTOM vector. The ligations were confirmed by the bands around 1000 b.p. in the 1% agarose gel after colony PCR (Figure 9). However, the sequencing result showed that the cloned insert has less than 30% similarity to the predicted sequence (data not shown).

---

Figure 7. A. Size exclusion column chromatography of purified ps-pET15b from Sepharose Fast-Flow column. B. Overlap of ps-pET15b with standards BSA and Lysozyme, which were indicated with labels besides them.

Figure 8. Gradient PCR result of BCE_1944 on agarose gel. The left column is DNA ladder besides with corresponding size label. The estimated size of BCE_1944 is 757 b.p.
shown). This mistake occurred probably due to mis-cut during endonuclease digestion or wrong primer binding during gradient PCR. So primers will be redesigned and the whole process will be repeated until obtaining correctly sequenced insert.

**Chromobacterium violaceum ATCC 12472 (CV3550)**

The gradient PCR was used to clone the expected enzyme DNA from genomic DNA using the forward and reverse primers shown in Table 1. DNA gel analysis revealed that the molecular mass of this product was similar to the predicted molecular mass of the putative dioxygenase CV3550 (767 b.p.) (Figure 10). The gene was subcloned into pET-15b and pTOM vectors.

The successful ligations were confirmed by the bands around 1000 b.p. in the agarose gel after colony PCR (Figure 11). All cloned gene constructs were sequenced. The successful cloning was confirmed by the sequencing results.
Figure 10. Gradient PCR result of CV3550 on agarose gel. The left column is DNA ladder besides with corresponding size label. The estimated size of BCE_1944 is 767 b. p., The isolated gene is circled orange rectangle.

Figure 11. Colony PCR result of CV3550. The left column is DNA ladder besides with corresponding size label.

To determine the best conditions to obtain soluble protein, all the parameters related to protein expression, such as vectors, temperature, media, and time were screened. Soluble and insoluble protein expression was compared for cv-pET-15b and cv-pTOM after transformation into BL-21 cells in LB-Amp media at 37 °C, 30 °C or 16 °C with induction with 1 mM or 10 mM IPTG, in addition with Super broth and Fe^{2+}. Unfortunately the SDS-PAGE gels showed that cv-pTOM dioxygenase produced very little soluble protein of the correct size under each growth condition (data not shown). Therefore, the cv-pTOM was transformed into BL-21 AI cells, and repeated the screening process (Figure 12). Under combinations of different concentrations of IPTG (1mM and 10 mM) and L-arabinose (0.2,
Figure 12. Protein expression of cv-pTOM construct in BL-21 AI cells with superbroth-Amp media with 0.05 mM Fe\(^{2+}\). BL-21 AI cells were grown at 30 °C and induced by 0.1 mM IPTG and four different concentrations (numbers on the middle of the gel) of arabinose separately. Aliquots were taken out at different time point as shown on top of each lane to check the protein solubility. The bands corresponding to the cv-pTOM molecular weight were in red rectangle, while the arabinose concentration at which cells produced the most soluble protein are circled in red oval.

0.02, 0.002, and 0.0002 % by weight), SDS-PAGE protein gel conformed super broth-Amp media with 0.05 mM ferrous ammonium sulfate at 16 °C for 24 h, induced with 1 mM IPTG and 0.02 % L-arabinose produced the highest level of soluble protein.

Successful purification of cv-pTOM was accomplished using immobilized nickel affinity chromatography. The protein was eluted using a lineal gradient of 0-50 % eluting buffer (containing 20 mM hepes, 0.5 mM NaCl, and 0.5 M imidazole (pH8.0)). The purity of the enzyme was confirmed by SDS-PAGE gels (Fig. 13). A single protein band was detected on the denaturing gel corresponding to about 31.3 kDa.
**Sulfolobus solfataricus P2 (SSO0066)**

Gradient PCR was used to clone the expected enzyme gene from genomic DNA using the forward and reverse primers shown in Table 1. DNA gel analysis revealed single bands at the predicted DNA size of the putative dioxygenase *SSO0066* (850 b.p.) (Figure 14). However, the attempt to ligate the *SSO0066* into pET-15b and pTOM vectors failed, and new attempts will be performed in the future.
**DesB protein expression screening and plasmid mutagenesis**

After the DesB gene was cloned into the pTOM vector by Julie Huang, it was found that little soluble protein was produced during protein expression screens (data not shown). Suspecting that the N-terminal His\_6 tag might be causing the protein to misfold leading to the solubility problems, the DesZ gene was re-subcloned into the pET21b vector in which the stop codon prevents the His tag production. Different combinations of temperatures and concentrations of inducing agents (IPTG and arabinose) including adding Fe\(^{2+}\) and Mn\(^{2+}\) were tried to get the optimal expression. However, none of these conditions provided enough soluble proteins, compared with the bulky bands in insoluble fractions as shown in Figure 15.

![Figure 15. DesB in pET21b vector expression screening. BL-21 AI cells with DesZ-pET21b were grown at 30 °C and induced by 0.1 mM IPTG and four different concentrations of arabinose (numbers under the red lines) separately. Aliquots were taken out at different time point as shown on top of each lane to check the protein solubility. The bands corresponding to the DesB molecular weight in the insoluble gels were very obvious and circled in red.](image-url)
Since the expression screening for DesB in pET21b without His tag was not optimal, using C-terminal His tag was considered to improve the solubility, thus site-directed mutagenesis was adopted to remove the stop codon in the DesB-pET21b construct. Many means have been tried including designing new primers, changing annealing temperature from 60 °C to 55 or 65 °C, increasing elongation time for 30 s, using better competent cells XL-10 gold. All of these attempts failed, giving a random sequence shown by sequencing. It was suspect that the batch of DesB-pET21b construct may have problems.

**Conclusions and Future Directions**

Among the selected five LigB divergent homologs, *C. violaceum* (CV3550) and *Pseudomonas syringae* DC3000 (PSPT_1776) were cloned and purified, DesB has solubility difficulty, and the other two homologs were still waiting to be inserted into a vector. Upon obtaining the proteins, the substrates need to be identified for further kinetics and other characteristic assays. The mutation of stop codon for the DesB-pET21b construct will be continued with other parameter changes, or an inclusion body purification may be applied. The cloning of the remaining two proteins will be focused on the ligation step, since the PCR amplification conditions have already been determined.
Chapter 5 Conclusion

Considering all the sequenced but far less well studied organisms, it is urgent and crucial to fill the vast gap between the fast growth of sequence data resulting from the development of sequencing techniques and the comparatively slow progress of functional characterization of sequenced genes. However, accurate gene functional annotation which mostly depends on extrapolations of functionally characterized genes with high sequence homology has been hindered, because notable functional promiscuity exists in many protein superfamilies. Thus, understanding of the relationship between protein sequence, structure and function, which is reliant on robust classification of protein superfamilies, will ultimately yield more accurate gene annotation.

In the family of aromatic-ring-cleaving dioxygenases, extradiol dioxygenases are involved in enzyme pathways that funnel lignin derived aromatic compounds into the TCA cycle. Enzymes belonging to the PCA superfamily of extradiol dioxygenases need additional characterization to reveal important aspects of structures, substrates specificities and catalytic mechanism. Genes with less than 20 % pair-wise identity to the first structurally characterized enzyme in the superfamily, protocatechuate 4,5-dioxygenase in Sphingomonas paucimobilis SYK-6, LigAB, are under investigation to reveal the boundaries of the PCAD superfamily.

Identification of potential diverged PCAD superfamily members will likely reveal enzymes with diverse substrate specificity. Further characterization of these and other related enzymes, including analysis of substrate specificity, mechanism and physiological structure,
should not only provide the foundation for understanding the specificity determinants for lignin-derived aromatic compound degradation, but also reveal more clearly the basis for differences between this new superfamily. This will also allow comparison to the VOC and cupin superfamilies.

The exploration and solidification of the new extradiol dioxygenase containing superfamily will facilitate discovery of new functions and help define the scope of possible reactions that can be catalyzed by this new structural fold, thus increasing our understanding between protein sequence, structure and function. This accumulative understanding will help filling the gap between sequenced genome data and experimentally verified functions, and benefit more accurate genome annotation.
Appendix 1 Sequences of proteins studied in this thesis.

Species of origin and putative function attributed by genome annotation are listed

*Escherichia coli* K12

Predicted dioxygenase YgiD
>F - unk
MTPLVVDIIMSSTRMPALFLGHGSPMNLEDNLYTRSWQKLGMMLPRPQAIVVVSAMH
WFRTRGTGVAMETPTIHDFGGFPQALYDTHYAPAGSPALQRLVELLAPIPVTLDKE
AWGFDHGSGVHLKMKYPMADIPMVQLSSKPAAWHFMGRKLALRDEGMILVA
SGNVVHLRTVWKHDSSPYPWATSFNEYVKANLTVQGPGVEQHPVLVNYLDHEGGET
LSNPTPEHLPPLYVLAGWDGQEPFTIPVEGIESMGSLSMVQIG

*Sphingomonas paucimobilis* SYK-6

3-methoxy gallic acid 3,4-dioxygenase

MAEIVLGIIGTSHPMLVTQTEQWRSLAFDQSVNHAWRGGWSYDQLVAERADQNFF
AAQITPEAMTAHNARQASLDQLAEIFSEAKIDVAVILQNDQMEEIFDERLVPAPFSFYGG
DTITNYEFPFAPECMLPGINLSVAGYIPSGGEAYAGQPELARSIIAQQAMADEFDVAA
MKALPDPETPHAFGQVYRRMRDNPVPSPVLPNYPPTVRRCYEFGKSLRG IQAWSERAVALASGGLTHFVIDEIDLFFQAMERDRIARLADLGEAIFQDGTSELK
NWIPLAGMMAELGLDEILDYVPCYRSEAGTGNAMGFVCWRMGHWHHHHHHHHHHG
LVPRGSH

*Bordetella bronchiseptica* RB50 (ATCC# BAA-588 BSL 2)

>Hypothetical_protein_BB1843_[Bordetella_bronchiseptica_RB50]

MGFIAGLKIPTISHYLPVAMRLPTLFSMSGSPMLAVEPGRTPGALAAWSDLGPERPR
AVLVVSPHMGMQGLAVSTRDRQVAWHDFGGFPADLYQLQYPAAEGSPAQLQVVDV
LARAGIQAGNDARRPLDHGAWVPLYRHAPDLPVQLSMDERMGQAYLQGA
LAPLRDDGVLSKVSGSLTHLNLDRVRMPHGAPADDYVAPFQQWYAHELAAAGDVEALL
DWQARAPGAARHPHDDHLMPLYVLAGGGMPARRRNDEVAYGALAMDAYQFGA
Pseudomonas syringae pv. tomato str. DC3000 (ATCC# BAA-871 BSL 1)

>hypothetical_protein_[Pseudomonas_syringae_pv._tomato_str._DC3000]_PSPT

MFPSLFISHGSPTLALEPGESGPALAAALPRPRAIVMVSAHWESHELIVNGNPQPE
TWHDGGFPAEFAMQYPAPGLPGLTRVVEETLTLGPARIDSRRPFDHGVVWVPLSL
MYPQADIPVQVSPQRLQGPAQTRQVGRALAGLREQGVLVGSGSITHNLYDLDWSA
APDRVEPWAEEFRWDWIDKLQSNDEAALHHYRTQAFAAPHRAHPSDEHLLPLYFARGA
GGAFSIAAYQGFTMGALEGMDIYRF

Chromobacterium violaceum ATCC 12472 (ATCC# 12472 BSL 2)

>Hypothetical_protein_CV_3550_[Chromobacterium_violaceum_ATCC 12472]

MTTRQPSLFISHGAPTLPEDIPLPHEQLGRDWPRAIVWSEARALTBNLQP
RPETMDFGGFPPVLRPYPAAAATAAATNQGLDAGLPAATEERPLDHGWM
PLMPLMFQADIPVLMISLRHQAGVAEHYALGQALKPLRDNNILVGSGSITHNLWEL
SPEGSEPWAADFAWMDQALLDNIIIHDIVHWQQRAPENPHSDEHFQPLLVA
LGASDDDAVPVKLHDDDRMGSLSMACWRFD

Bacillus cereus ATCC_10987 (ATCC# 10987 BSL 1)

>oxidoreductase_[Bacillus_cereus_ATCC_10987]

MMPSLFHAGSPMLAIQDFTDSFLKILGETYKPAIVIFTHWSEVLTISSDNEYE
TYYDFGGFPPPELYEIKRYAKGSISHASMLTKFKNGIPVHIMTRPLGDLHGWTLH
MYPAEANIPVQISVNPFLSAKEQFKIGEAALGKLQGEDILVIGSVTVHNLRAKWNQT
TPEQWAIEFDDWIKHMQAADDAFNWEKNAPHAQALVPRAEHFVLPIAMGSGE
NSGEVHYELGTLSDYLYQF

Sulfolobus solfataricus P2 (ATCC # 35092 BSL 1)

Predicted dioxygenase

>Sulfolobus_solfataricus_P2- Hypothetical protein

MRRLPAVAGSFYESDPKCLKMQIEWSFRHNIGPRDIPKQSYEKKKRDNLFFIVPHAGY
IYSGPVAATHYLYASEGKPDVIIIGPNHTGLGYSYVSAWPKEWETPLGSVKVDEE
VMQTVMESEVIDLEEKSHLEYSEVQPLFLQYFFDNFIVPIVMIQTEAEFLA
DAIYKVIQKYSKDIDIVLASSDMNYHDPEITMKKDEEAIKIQQLDYGRLYEVVEG
KDVTLCGYGPIMVSLILAKKLGKAYILKHATSGDTSGPDSVGVGAARFGS

*Sphingomonas paucimobilis* SYK-6

Gallate dioxygenase

MAKIIGGFAVSHTPTIAFAHDANKYDPPWAPIFQGFEPVKQLAEQKDVTFTYVYN
DHMTSFFEYSHFALGVGEEYSPADEGGQRDLPPIGDPELAKHIAECLVAEDFLAYWQGMIGDHGAFPSLSVLLPHEHGWPICRIVPLQCGLQHPIPKARRFWSNGIRSAYAIQSYPRDIKVAIAGTGGLSHQVHERAGFNNEWDMEFMERLANDPESLLGATVTD
LAKKGWEAGEYMMLMRGALSPEVKTLHQSYYFLPSMTAIATMLFEDQGDAAPP
AESDEALRARAKRELAVEEIEGTYPFTIDRAVKGFRINHFLRLEIPDFRKFVEDPE
GLFADSDLTEEKSILRNDRWIGMIHYGVFFMLEKMAVLGIGNIDVYAAFRGLSVP
EFQKTRNAAITYSVAGKQ
Appendix 2 Mass spectra

YgiD-pTOM
Ps-pET15b
cv-pTOM

4700 Linear Spec #1 MZQP = 31000.0, 36964

Mass (m/z)

100
90
80
70
60
50
40
30
20
10
0

10000
20000
30000
40000
50000
60000
70000
80000
90000
100000

% Intensity

5409.4494
3117.5874
3139.7498
1548.9898
1558.4180
3134.4080
63906.0013
33177.6450
20735.4473
1588.3533
1038.3533
46615.7734
2775.4473
1038.3533

10
20
30
40
50
60
70
80
90
100

Mass (m/z)

10000
20000
30000
40000
50000
60000
70000
80000
90000
100000

% Intensity

5409.4494
3117.5874
3139.7498
1548.9898
1558.4180
3134.4080
63906.0013
33177.6450
20735.4473
1588.3533
1038.3533
46615.7734
2775.4473
1038.3533
DesZ
# Appendix 3 Progress for all the proteins studies in this thesis

<table>
<thead>
<tr>
<th>Bacteria Strain</th>
<th>Grown</th>
<th>Frozen Stock</th>
<th>Genome extracted</th>
<th>Extracted Insert</th>
<th>Ligation into Vector</th>
<th>DH5a</th>
<th>BL21</th>
<th>BL21-AI</th>
<th>Sequencing</th>
<th>Soluble Insoluble</th>
<th>Protein Purified</th>
<th>Protein Assayed</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pseudomonas syringae</em> pv. tomato str. DC3000 G1</td>
<td>DNA</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>pTOM</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Chromobacterium violaceum</em> ATCC 12472</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>pET-15b  pTOM</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><em>Sulfolobus solfataricus</em> P2</td>
<td>DNA</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bacillus cereus</em> ATCC_10987</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DesZ</td>
<td>DNA</td>
<td></td>
<td></td>
<td></td>
<td>pET-15b  pTOM</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>DesB</td>
<td>DNA</td>
<td></td>
<td></td>
<td></td>
<td>pET-15b  pTOM</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YgiD</td>
<td>DNA</td>
<td></td>
<td></td>
<td></td>
<td>pTOM</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>
References

17. LINDSAY D. ELTIS JTB (1996) Evolutionary Relationships among Extradiol


42. Serre L, et al. (1999) Crystal structure of Pseudomonas fluorescens


64. Lane BG, *et al.* (1991) Homologies between members of the germin gene family in hexaploid wheat and similarities between these wheat germins and certain *Physarum* spherulins. *Journal of Biological Chemistry* 266(16):10461-10469.


87. Pochetti G & Montanari R (2012) Isothermal titration calorimetry to determine the association constants for a ligand bound simultaneously to two specific protein binding sites with different affinities.


110


135. W.M. H (1930) Rancidity in cheddar cheese. Master (Queen's University, Kinston).

