Microbial catabolic pathways for complex aromatics

Ph.D. Thesis

Shannu Palamuru
This thesis is dedicated to my parents

Malan Bee and Ramulu and my wife Keerthi

For their endless love, support and encouragement
A thesis submitted for the degree of Doctor of Philosophy of
The Australian National University

Microbial catabolic pathways for complex aromatics

Shannu Palamuru

Research School of Chemistry, Australian National University
Canberra, Australia

and

Commonwealth Scientific and Industrial Research Organization
Land and Water flagship
Canberra, Australia

October 2015
Declaration

I Shannu Palamuru hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma at the Australian National University or any other educational institution, except where due acknowledgement is made in the thesis.

Name: Shannu Palamuru

Sign:

Date: 21.10.2015
Acknowledgments

First and foremost I would like to thank my wonderful parents, wife, brothers and friends for supporting me through my university years.

It is a pleasure for me to express my particular gratitude to my supervisor Dr John Oakeshott for his guidance and encouragement throughout my PhD studies. Without his guidance and support this dissertation would not have been possible. He has also looked out for me more broadly through my time in Canberra.

I am also very grateful to my other CSIRO supervisor Dr Gunjan Pandey, and Prof. Jim Spain of Georgia Institute of Technology, for all their support and guidance. Likewise my ANU supervisor Professor Chris Easton for all his input, particularly in respect of the reaction chemistry of my enzymes.

Additionally I thank Dr Nikki Dellas (CSIRO) for all her advice and practical assistance in my lignin dehydrogenase work. Specifically she cloned and purified some of the dehydrogenases (2550, 1498, 2705, 3344, 474, and 724) from the SG61-1L bacterium and led the phylogenetic analysis of the dehydrogenases.

I am most grateful to Dr Tekle Tafese Fida and Professor Jim Spain (Georgia Institute of Technology) who provided me with the 2,4-dinitroanisole degrading bacterium Nocardioides sp. strain JS1661 that they had previously isolated and who carried out the growth and cell free extract experiments on this bacterium.

I am also most grateful to: Professor Eiji Masai (Nagaoka University of Technology, Japan) for providing the guaiacylglycerol-β-guaiacyl ether (GGE) stereoisomers and the Sphingomonas paucimobilis SYK-6 bacterium; Dr Andrew Warden (CSIRO) for helpful discussions and the NMR analysis of GGE compounds; Dr
Jian-Wei Liu and Dr Chris Coppin (CSIRO) for technical advice throughout my PhD; Dr Matt Taylor for assistance with analytical techniques; Dr Lars Jermiin (CSIRO) for assistance with the generation of protein sequence alignments and phylogenetic tree construction; Dr Peter Campbell (CSIRO) for assistance with tryptic digestion and protein identification; Stuart Kewish and Catherine Reid at the Australian Paper Mill for providing sewage samples; Dr Jung-Woo Park (Australian National University) for the IC-PMS analysis of metal in the DNAN O-demethylase enzyme; Dr Stephen Pearce for assisting with bioinformatics; and Alison Scott for supporting me with ANU administration.

I have also found many fantastic colleagues and friends at CSIRO and wish to thank them for their friendship here. I particularly acknowledge Robyn Russell, Peter Hart, Sue Dorrian, Chris Coppin, Jian-Wei Liu, Ree Murnane and Eva Zinkovsky.
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABTS</td>
<td>2,2’-Azinobis-(3-ethylbenzthiazoline-6-sulfonate)</td>
</tr>
<tr>
<td>ACN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>ADH</td>
<td>Alcohol dehydrogenase</td>
</tr>
<tr>
<td>AMS</td>
<td>Ammonium sulphate</td>
</tr>
<tr>
<td>ANAN</td>
<td>2-Amino-4-nitroanisole</td>
</tr>
<tr>
<td>ANU</td>
<td>Australian National University</td>
</tr>
<tr>
<td>BCA</td>
<td>Bicinchoninic acid assay</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>BTEX</td>
<td>Benzene, toluene, ethylbenzene and xylene</td>
</tr>
<tr>
<td>ChE</td>
<td>Cholinesterase</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethylsulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNAN</td>
<td>2,4-Dinitroanisole</td>
</tr>
<tr>
<td>dnH1</td>
<td>DNAN hydrolase α-subunit</td>
</tr>
<tr>
<td>dnH2</td>
<td>DNAN hydrolase β-subunit</td>
</tr>
<tr>
<td>DNP</td>
<td>2,4-Dinitrophenol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FAD</td>
<td>Flavin adenine dinucleotide</td>
</tr>
<tr>
<td>FPLC</td>
<td>Fast protein liquid chromatography</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>GGE</td>
<td>Guaiacylglycerol-β-guaiacyl Ether</td>
</tr>
<tr>
<td>GH</td>
<td>Glycoside hydrolase</td>
</tr>
<tr>
<td>GS-HPV</td>
<td>α-Glutathionyl-β-hydroxypropiovanillone</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-transferases</td>
</tr>
<tr>
<td>HAA</td>
<td>3-Hydroxyanthranilic acid</td>
</tr>
<tr>
<td>HBT</td>
<td>1-Hydroxybenzotriazole</td>
</tr>
<tr>
<td>HEPES</td>
<td>2-[4-(2-hydroxyethyl)piperazin-1-yl] Ethanesulfonic acid</td>
</tr>
<tr>
<td>HLADH</td>
<td>Horse liver alcohol dehydrogenase</td>
</tr>
<tr>
<td>HMX</td>
<td>Cyclotetramethylenetetranitramine</td>
</tr>
<tr>
<td>HNB</td>
<td>Hexanitrostilbene</td>
</tr>
<tr>
<td>HPLC</td>
<td>High pressure liquid chromatography</td>
</tr>
<tr>
<td>HPV</td>
<td>Hydroxypropiovanillone</td>
</tr>
<tr>
<td>ICP-MS</td>
<td>Inductively coupled plasma mass spectrometry</td>
</tr>
<tr>
<td>IPA</td>
<td>Isopropylamine</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>$k_{cat}$</td>
<td>Turnover number</td>
</tr>
<tr>
<td>$k_{cat}/K_M$</td>
<td>Catalytic efficiency</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilo Dalton</td>
</tr>
<tr>
<td>$K_M$</td>
<td>Michaelis constant</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>LC/MS</td>
<td>Liquid chromatography/mass spectrometry</td>
</tr>
<tr>
<td>LiP</td>
<td>Lignin peroxidase</td>
</tr>
<tr>
<td>M/Z</td>
<td>Molecular Ion</td>
</tr>
<tr>
<td>MeOH</td>
<td>Methanol</td>
</tr>
<tr>
<td>MM</td>
<td>Minimal Medium</td>
</tr>
<tr>
<td>MMA</td>
<td>4-Methoxymandelic Acid</td>
</tr>
<tr>
<td>MnP</td>
<td>Manganese peroxidases</td>
</tr>
<tr>
<td>MPHPV</td>
<td>α-(2-methoxyphenoxy)-β-Hydroxypropiovanillone</td>
</tr>
<tr>
<td>NA</td>
<td>Nutrient agar</td>
</tr>
<tr>
<td>NAC</td>
<td>Nitroaromatic compounds</td>
</tr>
<tr>
<td>NAD</td>
<td>Nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>NADP</td>
<td>Nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>NB</td>
<td>Nutrient broth</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>OD&lt;sub&gt;600&lt;/sub&gt;</td>
<td>Optical density at 600 nm</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PAH</td>
<td>Polyaromatic hydrocarbons</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PMSF</td>
<td>Phenyl methane sulfonyl fluoride</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>POP</td>
<td>Persistent organic pollutant</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinylidene fluoride</td>
</tr>
<tr>
<td>RAST</td>
<td>Rapid Annotation using Subsystem Technology</td>
</tr>
<tr>
<td>RDX</td>
<td>Cyclotrimethylenetrinitramine</td>
</tr>
<tr>
<td>RT</td>
<td>Retention time</td>
</tr>
<tr>
<td>SADH</td>
<td>Saccharomyces cerevisiae alcohol dehydrogenase</td>
</tr>
<tr>
<td>SDR</td>
<td>Short-chain alcohol dehydrogenases/reductases</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>TFA</td>
<td>Trifluoroacetic acid</td>
</tr>
<tr>
<td>THF</td>
<td>Tetrahydrofolate</td>
</tr>
<tr>
<td>TNT</td>
<td>2,4,6-trinitrotoluene</td>
</tr>
<tr>
<td>TOF</td>
<td>Agilent 1200 Infinity Series Time Of Flight</td>
</tr>
<tr>
<td>TSB</td>
<td>Trypticase soy broth</td>
</tr>
<tr>
<td>VA</td>
<td>Veratryl alcohol</td>
</tr>
<tr>
<td>Ni-NTA</td>
<td>Nickel-nitrilotriacetic acid</td>
</tr>
</tbody>
</table>
# Table of contents

Declaration .......................................................................................................................... I

Acknowledgments ............................................................................................................... II

List of Abbreviations ....................................................................................................... IV

Table of contents .............................................................................................................. VIII

Abstract ............................................................................................................................ 1

## Chapter 1  Literature review ......................................................................................... 3

1.1  Foreword .................................................................................................................. 3

1.2  Aromatic compounds and aromaticity ................................................................. 6

1.2.1  Natural and synthetic aromatic compounds ................................................... 8

1.3  Lignin .................................................................................................................... 10

1.3.1  Current physicochemical processing of lignin ................................................... 14

1.3.2  Lignin degradation by bacterial enzymes ......................................................... 24

1.4  Nitroaromatic compounds ................................................................................. 31

1.4.1  Chemistry of the nitroaromatics .................................................................... 32

1.4.2  Synthetic and natural nitroaromatics .............................................................. 33

1.4.3  Biodegradation of nitroaromatics ................................................................... 35

1.5  Thesis outline ....................................................................................................... 37

## Chapter 2  Phylogenetic and kinetic characterization of a suite of dehydrogenases from a newly isolated bacterium that catalyse the turnover of guaiacylglycerol-β-guaiacyl ether stereoisomers ....................................................................................... 39

2.1  Introduction .......................................................................................................... 39

2.2  Materials and Methods ....................................................................................... 41

2.2.1  Materials ........................................................................................................... 41

2.2.2  Media and growth conditions ......................................................................... 42

2.2.3  Instrumentation and analytical methods ......................................................... 42

2.2.4  Isolation of GGE-degrading bacteria ............................................................... 45

2.2.5  Dye decolorizing experiments ....................................................................... 46

2.2.6  Growth and resting cell experiments .............................................................. 47

2.2.7  Genome sequencing, nucleotide sequences and accession numbers .......... 48

2.2.8  Cloning and overexpression of dehydrogenase genes .................................. 49
Chapter 3  Aerobic biodegradation of 2,4-dinitroanisole (DNAN) by a novel hydrolase from \textit{Nocardioides} sp. JS1661 ........................................................................ 76

3.1  Introduction ...................................................................................................... 76

3.2  Experimental .................................................................................................... 77

3.2.1  Materials ........................................................................................................ 77

3.2.2  Isolation of DNAN-degrading bacteria ............................................................ 78

3.2.3  Growth of DNAN-degrading isolates ............................................................... 78

3.2.4  Enzyme assays ................................................................................................ 79

3.2.5  Analytical methods .......................................................................................... 81

3.2.6  Nucleotide sequence accession numbers ....................................................... 82

3.2.7  Purification of DNAN hydrolase .................................................................... 82

3.2.8  DNAN hydrolase mass estimation and gene identification .............................. 83

3.2.9  Determination of optimum temperature and pH ............................................. 85

3.2.10  Determination of kinetic constants .............................................................. 86

3.3  Results .............................................................................................................. 86

3.3.1  Isolation and identification of DNAN-degrading strains .................................. 86

3.3.2  Growth of JS1661 .......................................................................................... 87

3.3.3  Enzyme assays ................................................................................................ 88

3.3.4  Genes encoding DNAN hydrolase ................................................................. 91

3.3.5  Genes encoding 2,4-DNP catabolic pathways ............................................... 95

3.3.6  DNAN hydrolase characterization .................................................................. 97

3.3.7  Substrate specificity and kinetic analysis ....................................................... 98

3.3.8  Discussion ...................................................................................................... 99

Chapter 4  General discussion ............................................................................. 102
Abstract

This thesis concerns the microbial degradation of recalcitrant aromatic compounds, and the enzymes responsible for that degradation. The two foci are guaiacylglycerol-β-guaiacyl ether (GGE), which is a model compound representing a key intermediate in the deconstruction of lignin, and 2,4-dinitroanisole (DNAN), which is a widely used component of explosives. The degradation of these compounds is of interest both because of the insights it provides into the evolution of catabolic pathways and because of the potential uses of the microbes and enzymes involved in chemical manufacturing and environmental bioremediation.

Chapter 1 reviews the existing literature on the microbial degradation of lignin and nitroaromatic explosives such as DNAN, including consideration of their potential industrial applications.

Chapter 2 describes the isolation and characterization of the GGE degrading Erythrobacter sp. SG61-1L strain and compares it with the GGE degradation activity of the previously isolated Sphingomonas sp. SYK-6. SG61-1L was found to degrade GGE significantly more rapidly than SYK-6, although the same pathway appeared to operate in both strains. The first step in GGE degradation by SYK-6 had previously been shown to be catalysed by four stereospecific dehydrogenases and I found seven homologous genes in the SG61-1L genome. These thirteen genes were heterologously expressed in E. coli and the activities of their products were studied with the four stereoisomers of GGE. Two of the SG61-1L enzymes in particular had higher activities for the four GGE isomers than any of the four SYK-6 enzymes. A phylogenetic analysis showed that the GGE dehydrogenase activity of the two strains was mainly confined to a particular subclade of classical short chain dehydrogenases but two enzymes from other clades of these dehydrogenases were also found to have significant GGE activity.

Chapter 3 describes the mineralization of DNAN by Nocardioides sp. strain JS1661 and the identification and characterization of a complex demethylase enzyme in this bacterium which catalyses the cleavage of the methyl group on DNAN. The hydrolytic removal of the methyl group forms methanol and 2,4-dinitrophenol, which is
mineralized via a well-established catabolic pathway. The demethylase enzyme was purified to more than 90% purity, revealing two distinctive bands on SDS-PAGE. Tryptic digestion and peptide sequencing of these bands identified two overlapping genes in the genome of this bacterium. The two genes, named *dnH1* and *dnH2*, had 40.5% nucleotide and 21.3% amino acid sequence similarity to one another. Their closest known relatives were beta lactamase proteins which had 37% and 55% amino acid sequence similarity to *dnH1* and *dnH2* respectively. Only one set of homologues was found, from *Nocardia testacea*, which are co-translated via cognate stop-start codons in the same way as *dnH1* and *dnH2*.

Chapter 4 is a general discussion of the findings described in the two previous chapters. The consideration of evolutionary issues focuses on the functional and phylogenetic relationships of the two sets of enzymes characterized. While the dehydrogenation of the β-aryl ether linkages in GGE by the GGE dehydrogenases seems to have evolved in three clades of classical short chain dehydrogenases, the *O*-demethylation of the DNAN by the *dnH1/dnH2* enzyme appears to be a new class of activity which is functionally distinct from previous classes of *O*-demethylase enzymes and only very distantly related to other enzymes, the closest known being beta lactamases. The GGE dehydrogenases may be useful in chemical syntheses requiring various forms of stereospecificity, including but not limited to the use of biomass feedstocks. The DNAN demethylase may prove useful in various bioremediation and biosensors strategies, although challenges will first need to be overcome in respect of its heterologous expression.
Chapter 1  Literature review

1.1  Foreword

Biocatalysts, be they microbes or isolated enzymes, hold great potential for deployment in both sustainable chemical manufacturing processes and the bioremediation of recalcitrant residues of synthetic chemicals in the environment. This is because they can replace energy-intensive physicochemical catalysts, carry out the reactions with much greater specificity, and often efficiency, and open up the possibility of many new transformations previously beyond the capabilities of these industries.

The potential for biocatalysts to reduce the carbon footprint of chemical manufacturing can be seen through the effects of a single example, the production of ammonia, which is currently carried out using the Haber process. This process relies on high temperature and pressure catalysis that alone is estimated to contribute 3 to 5% of total global CO₂ emissions annually [139, 284]. An effective biocatalytic method for reacting hydrogen and nitrogen to produce ammonia could therefore alone reduce global CO₂ emissions significantly [179, 221]. Biocatalysts also have the potential to replace petrochemical feedstocks with renewable resources that have previously been beyond the scope of the physicochemical processes currently available for chemical manufacturing. Lignocellulosic biomass is a case in point – it is estimated that 50 million tons of such biomass could be used in place of petrochemical feedstocks if only commercially viable ways could be developed to systematically deconstruct it into usable subunits [68, 87, 127].
Both microbial and enzymatic products and processes are now being developed for the bioremediation of toxic compounds in the environment. Microbial bioremediation is finding applications in the clean-up of contaminated soils [136, 216]. Both biostimulation (the stimulation of endogenous microbial activity by the addition of nutrients [50, 216, 259, 279]) and bioaugmentation (inoculation with live microbial communities optimised for the degradation of particular compounds of concern [50, 259, 279]) prove more cost-effective than traditional physicochemical methods in many situations. Enzymatic bioremediation, which relies on enzymes specifically developed to degrade particular compounds [120], is particularly suited to the clean-up of contaminated waters or wettable surfaces (eg. commodities), where there is a premium on rapid remediation and/or where the use of live microbes might be considered inappropriate.

Aromatic compounds are major targets for biocatalytic processing in both chemical manufacturing and bioremediation. Plants, animals and microbes all produce a wide array of aromatic secondary compounds, the lignin mentioned above being one very abundant example of a complex plant polyaromatic compound. Highly substituted aromatics are also a major component of the chemical manufacturing business, with applications ranging from bulk commodity compounds like many pesticides (eg. organophosphates, pyrethroids) [180] through to higher value fine chemicals (eg. vanillin, syringaldehyde and phenolic oligomers) [101] and pharmaceuticals (eg. phenacetin, acetanilide) [114]. The fundamental benzene ring building blocks and a number of difficult linkages between the rings make many natural and synthetic aromatics problematic to break down, either by biological or by physicochemical means. In the case of the naturally occurring aromatics, various lineages of microbes have had
vast evolutionary timescales in which to develop pathways that would break abundant aromatics down into useful sources of carbon and energy. However effective breakdown of lignin apparently remains beyond all but a few soil microbes (notably white rot fungi and certain Actinomycetes and Rhodococcus). Many synthetic aromatics and polyaromatics may be inherently more susceptible to break down than lignin at least, but soil microbes have had much less time to evolve effective pathways for this, and many of the compounds therefore remain formally classified as persistent organic pollutants (POPs) in the environment [93, 115].

The goals of this thesis are to elucidate the pathways which certain soil bacteria have evolved to deconstruct and utilise two particular sets of aromatics. One set involves a naturally occurring polyaromatic whose deconstruction would provide a new feedstock for the chemical industry. The second set involves a relatively recently developed synthetic aromatic which is becoming an increasing priority for bioremediation. More specifically, in the first case the goal is to identify and characterise microorganisms and the enzymes within them which may have potential for the systematic deconstruction of lignin into components that could be used by the chemical industry. The particular focus is on the most abundant inter-subunit β-O-4 (β-aryl ether) linkages between the aromatic rings in lignin. In the second case, the goal is to find and characterise microbes for the breakdown of a nitroaromatic compound, 2,4-dinitroanisole (DNAN), which is becoming the explosive of choice in many resource extraction and military circumstances. The particular challenge here is the combination of ortho and para nitro groups and a methoxy group attached to the benzene ring. The nitro groups are highly deactivating groups, and the elimination of the methoxy group
is thermodynamically unfavourable, so DNAN has proven relatively resistant to metabolically useful microbial attack up to now.

The remainder of this chapter will review the chemistry of the two sets of target compounds and the microbiological and enzymatic issues surrounding their degradation. The second chapter describes my isolation and characterisation of a bacterium that can degrade the β-O-4 linkages in model lignin dimers, and a detailed investigation of the enzymes involved in the first step in this pathway. The third chapter then describes the isolation and characterisation of a bacterium and its enzyme that is responsible for the initial hydrolytic breakdown of DNAN. The general discussion in chapter 4 considers the findings described in the two previous chapters in both an evolutionary context and in respect of the potential applications in chemical manufacturing and bioremediation.

### 1.2 Aromatic compounds and aromaticity

Aromatic compounds, also known as arenes, are chemical compounds that contain conjugated planar ring systems with delocalized π-electron clouds instead of discrete alternating single and double bonds. One typical aromatic compound is benzene [237]. Benzene has high thermodynamic and chemical stability, which are due to its delocalization of π-electrons and consequent resonance stabilization. Benzene has two resonance structures which are represented, together with a resonance hybrid, in Figure 1-1A [31, 237].

Benzene has six electrons in its resonance structures, and all its carbons are sp² hybridized, so the benzene ring is planar with six partially overlapping p-orbitals forming
a large loop orbital, which provides the unusual stabilization to the molecule (Figure 1-1B). This key feature of aromatic molecules is known as aromaticity and the energy required to break its resonance stabilization is known as resonance energy. The unusual resonance stability of benzene and substituted benzene allows them to undergo nucleophilic addition reactions, but generally not electrophilic substitutions [31, 237]. The electrophilic substitution reaction requires the $\pi$-electrons to split and form two $\sigma$-bonds. Therefore, it is chemically and thermodynamically difficult to break the aromaticity or resonance. However, some microbes have otherwise unusual high redox potential enzymes which can break the aromatic compounds and allow the microbes to use them as sources of nutrient and energy [31, 237].

Figure 1-1. The benzene structure and its aromaticity. (a) Benzene resonance structures and resonance hybrid structure. (b) Benzene planar structure with $p$-orbitals overlapping [36].
1.2.1 **Natural and synthetic aromatic compounds**

Aromatic compounds play important roles in the biochemistry of living organisms. They include diverse complex molecules such as lignin, tannins and several other recalcitrant secondary metabolites, various essential amino acids, and nucleotides. Many of the more complex aromatics are primarily formed from three of the amino acids, phenylalanine, tyrosine and tryptophan. Plants and animals have limited capacity to metabolize many aromatic compounds [57, 117] but many microbes are able to degrade a range of aromatics and their derivatives. Over millions of years they have evolved complex operons for the efficient expression of pathways of enzymes and other proteins involved in the metabolism of various naturally occurring aromatics.

Aromatic molecules with one hetero atom (i.e. other than carbon) in the ring are identified as hetero-aromatics. Some hetero-aromatics are produced naturally as pigments, nucleotides, antibiotics and vitamins [117]. Synthetic hetero-aromatics include various plastics, drugs, dyes and pesticides. They include thiophene, furan, 4,5-dihydrofuran, pyrrole and pyridine [117].

Polyaromatic hydrocarbons (PAHs) contain multiple aromatic rings fused together. They have generally been produced from fossil hydrocarbons under low oxygen concentrations or through to the incomplete combustion of organic matter. Example of natural PAHs include, phenanthrene, 3-nitrobiphenyl and 1- and 2-nitronaphthalene. Synthetic PAHs are also used in dyes, plastics, pesticides and drugs [69, 117].
A set of four aromatics known as BTEX (benzene, toluene, ethylbenzene and xylene) is very widely used as starting material by many chemical manufacturing industries. About 35 million tonnes of BTEX are used worldwide every year, much of it for the production of derivatives such as substituted, heterocyclic and polyaromatic hydrocarbons with diverse uses [41, 117, 238]. Until now, BTEX has been largely sourced from fossil hydrocarbons.

Nitroaromatics are an important class of substituted aromatics which are widely used as pesticides, explosives and antibiotics. Some nitroaromatics are produced by plants and fungi but most are synthesized by the chemical industry. Examples of synthetic nitro-aromatics are carbofuran, 2,4,6-trinitrotoluene (TNT) and 2,4-dinitroanisole (DNAN).

Many synthetic aromatics have become persistent organic pollutants when released into the environment. A few microbes have been reported to slowly degrade some of these synthetic xenobiotics, often those which are relatively similar to natural organic compounds, through adaptations of previously developed catabolic pathways for related natural aromatics. Often the xenobiotic compounds are recalcitrant to individual microbes but a community of microbes may be able to degrade them collectively. The microbes may use oxic and/or anoxic strategies to assimilate or breakdown these compounds but in either case the process generally involves transformation of the compound into simpler intermediates, followed by ring fission to produce chemicals that can be incorporated into one or other of the central metabolic pathways. Often these initial transformation reactions and subsequent ring fissions are referred to as upstream and downstream pathways respectively [78].
1.3 Lignin

Lignocellulosic biomass could contribute significantly to climate change mitigation by providing a renewable source of transport fuel, electrical energy and commodity chemicals while also recycling atmospheric carbon dioxide. Worldwide, lignocellulosic biomass is the largest source of renewable carbon. The European Union already obtains 66 % of its renewable energy from biomass and the United States Department of Energy aims to have 20 % of its transport fuels and 25 % of its commodity chemicals derived from biomass by 2030 [193]. Geosciences Australia has estimated that as much as 500,000 km² of arable land in Australia could be suited to the sustainable production of various forms of ‘second generation’ (non-food) plant biomass [71, 80, 147, 193, 251, 273].

Lignocelluloses are comprised of cellulose, hemicellulose and lignin. Cellulose and hemicellulose are sugars and can be processed by fermentation with microbes into first generation biofuels, such as bioethanol, biodiesel and biogas, which have gained some market shares in countries such as Brazil, the US and China [273]. However, the production of these first generation biofuels has to compete with food and forage production and biodiversity conservation. Therefore, much interest has turned to the development of second generation plant biomass. Lignin could be the critical resource for the latter because, unlike the cellulose and hemicellulose, it has little food and forage value. Indeed it is a waste stream in many primary industry production chains [66, 193, 273].

Lignin is a complex aromatic polymer [34, 263] which does not have a precise structure and its composition can vary from plant to plant. Typical lignin percentages
range from 10 to 30% of the biomass in vascular plants [263, 278]. Lignin helps maintain the mechanical strength of plant structures, acts as a barrier against microbial attack, and helps in water transport by protecting the soluble cellulose and hemicellulose polymers [26, 89, 263]. It is largely derived from three aromatic alcohols (coumaryl, syringyl and coniferyl alcohol) which are formed through a well-established phenylalanine pathway, with the phenylalanine generated through the shikimate biosynthetic pathway [70, 89, 263]. The three primary aromatic alcohols undergo a dehydrogenative polymerization to form a complex in which they become the corresponding p-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) monolignol structures (Figure 1-2). This polymerization step is catalysed by enzymes such as laccases, peroxidases and several other oxidases [116, 263, 277]. Gymnosperms (softwoods) have more G and fewer H units whereas angiosperms (hardwoods) have more G and S units and fewer H units [70, 150, 263]. Monocots such as grasses have relatively more H units and fewer G and S units [91, 263].
Figure 1-2. Monolignols and their aromatic residues in the lignin polymer [277]

The dehydrogenative polymerization above actually occurs through several steps, the first of which involves electron shuffling of the aromatic alcohols to generate radical monomers which have electron deficient β-carbons. Radical coupling reactions amongst these, and with other downstream molecules, then form various lignin dimers, trimers and tetramers. The linkages among the component units may be phenylpropane β-aryl ether (β-O-4), α-aryl ether (α-O-4), phenylcoumaran (β-5), biphenyl (5-5), dibenzodioxocin and β-β-linked structures [116, 261, 263] (Figures 1-3 and 1-4). A typical lignin comprises ~70-85 % β-O-4 linkages but its monomer composition is influenced by several environmental factors such as light, soil and temperature, as well as its location in the cell wall and the developmental stages of the tissue with which it is associated. The structure of the lignin is made more complex by the chiral nature of several of the linkages [98, 205]. The paper and pulp industries use the cellulose from plant biomass for making paper and separate the lignin as a waste byproduct [116, 263].
Figure 1-3. Examples of lignin structures, highlighted with key linkages [281]

Figure 1-4. Some major dimer structures [102].

Note: Adapted with permission (see Appendix 1) from Zakzeski J, Bruijnincx PCA, Jongerius AL, Weckhuysen BM: The catalytic valorization of lignin for the production of renewable chemicals. Chemical Reviews 2010, 110(6):3552-3599.
1.3.1 Current physicochemical processing of lignin

Plant biomass is one of the few renewable resources produced in amounts sufficient to substitute for worthwhile proportions of the fossil carbon used in industrial processes. Globally, about 50 million tons of lignin waste is generated per annum from pulping and other processes [61]. Currently however only about 1.1 million tons of this is used in chemical manufacturing because of the problems of cost efficient and systematic decomposition of the lignin [61]. About 1 million tons of the 1.1 tons is subject to acid sulphite processing and the remaining 100,000 tons is subject to the Kraft process [30, 61, 282]. The high temperature and reactive chemicals (NaHSO₃, Mg(HSO₃)₂, CaCO₃) in these processes convert the lignin to soluble forms which are then separated from the cellulose and hemicellulose.

![Diagram of sulphonated lignin and Kraft lignin](image)

**Figure 1-5.** (A) An example of sulphonated lignin produced by acid sulphite processing. (B) An example of Kraft lignin showing the addition of various carboxylic and thiol groups produced by the Kraft process.

Note: Adapted with permission (see Appendix 1) from Zakzeski J, Bruijnincx PCA, Jongerius AL, Weckhuysen BM: The catalytic valorization of lignin for the production of renewable chemicals. Chemical Reviews 2010, 110(6):3552-3599.
In the sulphite process, the side chains of the lignin are sulphonated by sulphurous salts such as sulphites and bisulphites at high temperature (140-170 °C) in digesters, which makes the lignin hydrophilic (Figure 1-5A) [281]. This process not only dissolves the lignin but also hydrolyses the hemicelluloses, retaining the cellulose for use in pulp and paper making. The sulphonation mostly occurs at α-carbons in the lignin, cleaving most of the C-O-C linkages [13, 53, 131, 281]. The non-cellulose fraction, usually known as brown liquor, is then treated with metal or ammonium salts to separate the hemicellulose and sulphonated lignin from unreacted sulphites. The highly sulphonated lignin obtained is soluble in a wide range of solvents. It is generally used for products of relatively low commercial value such as dye dispersants, binders, animal feed pellets and concrete admixtures [281].

The Kraft process yields better quality pulp than sulphite processing and also has advantages in the treatment of resin-rich pine wood. In Kraft processing, an aqueous solution of sodium hydroxide and sodium sulphide (white liquor) is used to cook wood chips in digesters at high temperature (170 °C) for approximately 2 hrs [40], which depolymerizes a proportion of the lignin to smaller, more soluble subunits (Figure 1-5B). Subsequent concentration of this black liquor produces a solid called Kraft lignin, most of which is burnt to generate electricity and a little of which is used for producing commodities such as dyes, agrochemicals, asphalt emulsifiers and autoxidizing agents [281].

Lignin degradation can also be achieved by several other thermochemical techniques such as pyrolysis, gasification, oxidation and hydrogenolysis [193]. However, the combination of high temperatures, pressures, solvents and catalysts used in these
processes present economic and ecological problems and they are not currently used at significant scales commercially [193].

In pyrolysis, the lignin biomass is treated at moderate temperatures (200-760 °C) in the absence of oxygen and with or without catalysts in large chambers to convert the lignin biomass to biofuels and other industrial products. The pyrolysis treatment initiates several radical reactions, rearrangements and eliminations in the lignin molecule, breaking it into several smaller subunits [10, 281].

In gasification, the lignin biomass is treated at extremely high temperatures (450 – 1650 °C) under limited oxygen and steam in huge chambers to generate useful gaseous products such as hydrogen and carbon monoxide. However, some other gases such as carbon dioxide, methane, sulphur and nitrogen are also produced in the process, which are then treated with special catalysts to convert them into more useful products. The resultant syngas can be condensed and liquefied to transport fuels or industrial precursors and can also be used to generate electricity [10, 104, 281].

In the hydrogenolysis process, the lignin biomass is treated with hydrogen gas in the presence of chemical catalysts to lyse the C-C and C-O-C bonds through hydrogenation. The resulting breakdown products can subsequently be transformed to valuable products. The hydrogenation also forms gases such as H₂S, which is desulphurised and H₂ gas is harvested.

In the oxidation process the lignin biomass is treated with oxidising agents such as oxygen, nitrobenzene and hydrogen peroxide in the presence of catalysts to convert
the lignin to useful renewable resources such as vanillin, phenols, syngas products and various other hydrocarbons [130, 254, 281].

As noted earlier, all of the above processes are essentially non-systematic, in so much as they break lignin non-specifically into a wide range of products, each with relatively low product yield and only a few of which are useful in chemical manufacturing industries. Some of the products can also be at least as recalcitrant to further chemical processing as lignin [10, 88, 104, 148]. By contrast, biocatalytic degradation of lignin could potentially be an inexpensive, energy-efficient and well controlled process that breaks down the lignin systematically and efficiently to a specified set of more useful products. Several fungi and bacteria with potential in this respect have indeed been identified and, in a few cases, their enzymes were characterized for their potential uses in industrial lignin degradation [51, 145, 278].

A key prerequisite for lignin degradation is substrate accessibility/availability to the microbial (extra or intracellular) enzymes. Intracellular enzymes are less likely to be involved in preliminary lignin degradation, but extracellular enzymes have been identified as primary invaders of the lignin polymer in previous studies [44, 59]. The intracellular enzymes are generally very specific for particular lignin-derived monomeric and dimeric compounds [44, 59].

Lignin degrading enzymes are widely distributed in various species of fungi, including aerobic ascomycetes (Trichoderma reesei), basidiomycetes species (Phanerochaete chrysosporium), and a few anaerobes (Orpinomyces sp). However, by comparison, only a few well characterized lignin degrading enzymes have been identified in bacteria (Rhodococcus sp., Streptomyces sp.). Identifying efficient bacterial
enzymes might be key to the effective transformation of lignocellulosic biomass to commodities. Most of the lignin degrading enzymes discovered and studied so far have been from aerobic microbes, which use oxic conditions to attack the lignin [29, 33, 44].

Lignin has also been reported to be degraded under low oxic and anoxic conditions with various anaerobic and microaerophilic microbes. Some examples include anaerobic incubations of swamp sediments with \(^{14}\text{C}\) lignocellulose softwood derived from the herbaceous plant \textit{Spartina alterniflora}, hardwood from \textit{Rhizophora mangle} and \(^{14}\text{C}\) synthetic lignin, where slow degradation of up to 16 % of the lignocellulose to \(^{14}\text{C}\text{CO}_2\) and \(^{14}\text{C}\text{CH}_4\) has been found after up to 294 days. Similarly, plant biomass incubated in a closed bioreactor with a micro-oxic to anaoxic gradient for a year at 30° C resulted in partial lignin degradation, but no individual microbes potentially responsible for this were identified or characterized further [18, 262].

On other hand several individual anoxic bacterial isolates have been isolated, albeit the relevant enzymes not yet characterized, using lignin model substrates such as benzoic acid, ferulic acid, gallate and other lignin monomeric substrates. Some examples include the isolation of \textit{Eubacterium oxidoreducens} sp. nov. from the rumen of a hay-fed steer which was able to degrade gallate, pyrogallol, phloroglucinol and quercetin in the presence of electron acceptors such as H\(_2\) or formate [132]. Likewise, cell-free extracts of \textit{Syntrophococcus sucromultans} were able to metabolize several lignin derivatives in the presence of formate or vanillate, although again, none of the enzymes responsible were characterized [20, 60].

The research in this thesis is concerned with oxidative lignin degradation so the remainder of this literature review focuses on oxic fungal and bacterial enzymes.
1.3.1.1 Lignin degradation by fungal enzymes

Certain so-called white rot fungi, in particular *Trametes versicolor*, *Heterobasidium annosum* and *Irpex lacteus*, have been found to degrade lignin efficiently via a combination of extracellular peroxidases and laccases with high redox potential. These fungi preferentially degrade the lignin in the plant cell wall, retaining the enriched cellulose as white mottled regions called white rots. The so called brown rot fungi by comparison prefer degrading cellulose and hemicellulose, thereby retaining the brownish lignin [100, 163, 190, 278].

The extracellular peroxidases produced by white rot fungi have been classified into three families of class II peroxidases, namely lignin peroxidases (LiPs; EC 1.11.1.14), manganese peroxidases (MnPs; EC 1.11.1.13) and versatile peroxidases (VPs; EC 1.11.1.16). The enzymes in all these families exist in several isoforms. Several low molecular weight compounds such as veratryl alcohol, Mn$^{3+}$, oxalate, 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) and 4-hydroxybenzoic acid also act as redox mediators for the peroxidases, getting oxidised at the enzyme active site and then diffusing into the polymer to break the substrate. Apart from the extracellular peroxidases, several intracellular enzymes such as glyoxal oxidase, veratryl alcohol oxidase, quinone reductase, cellobiose dehydrogenase and pyranose 2-oxidase can also play a role in lignin degradation [112, 190, 278].

1.3.1.2 Lignin peroxidase

Lignin peroxidases (LiPs) are 38-46 kDa metalloenzymes, specifically haem-glycoproteins, with a very low pH optimum of 3.2-4.0 and relatively high redox potentials, up to 1.4 V [74]. The first LiP identified was from *Phanerochaete*
*chrysosporium*, a basidiomycete white rot fungus [100, 112, 163, 190, 278]. LiPs catalyse
the \( \text{H}_2\text{O}_2 \) and veratryl alcohol dependent oxidation of various electron rich aromatic
substrates, including phenolics. The veratryl alcohol (VA) is a natural substrate for LiPs,
acting as a redox mediator between the enzyme and the lignin. LiPs oxidise VA to a VA-
cation radical which then diffuses into the lignin polymer and oxidises it. A typical
reaction can be represented in simplified form as

\[
\text{1,2-bis (3,4-dimethoxyphenyl)propane-1,3-diol} + \text{H}_2\text{O}_2 \rightarrow \text{3,4-dimethoxybenzaldehyde} + 1-(3,4-
dimethoxyphenyl)ethane-1,2diol + \text{H}_2\text{O}.
\]

The ability of LiPs to oxidise phenolic as well as
non-phenolic substrates is due to their high redox potentials [100, 112, 163, 190, 278].

As with other haem peroxidases, the catalytic cycle of LiPs is actually
considerably more complex than the simple representation above. As shown below, the
first step is an oxidation involving two electrons (one from Fe (III) and another from the
porphyrin \( \pi \)-electron system of the haem group) of native ferric ions with \( \text{H}_2\text{O}_2 \) to
produce \( \text{H}_2\text{O} \) and compound I, which is a ferric porphyrin radical cation ([Fe(IV)=O⁺⁺]).

Compound I is then reduced to native ferric ion by two successive one electron
reductions where the electrons are gained from veratryl alcohol (VA). In the first
reduction, compound I is reduced by VA to compound II, which is an oxyferryl complex,
[Fe (IV) =O]. Compound II is then further reduced by another reduction step [100, 163,
274, 278] (Figure 1-6A). The overall reaction is represented as

\[
\text{Native ferric enzyme} + \text{H}_2\text{O}_2 \rightarrow \text{Compound I ([Fe (IV)=O⁺⁺])} + \text{H}_2\text{O}
\]

\[
\text{Compound I ([Fe (IV)=O⁺⁺])} + \text{VA} \rightarrow \text{Compound II ([Fe (IV)=O])} + \text{VA}^{**}
\]

\[
\text{Compound II ([Fe (IV)=O])} + \text{VA} \rightarrow \text{Native ferric enzyme} + \text{VA}^{**}.
\]
(Note that in the absence of donor substrate and the presence of excess H$_2$O$_2$, compound II can be further oxidised to compound III, a ferric-superoxide complex [Fe(III)O•⁻] which is spontaneously reduced to native enzyme by VA [163, 274, 278].

The LiPs have very broad substrate specificity, because the phenolic radicals (donor substrates) generated during the reduction reactions diffuse into the lignin polymers and attack them non-specifically, producing various ring opening, rearrangement and other intramolecular reactions. These reactions have been studied with model lignin compounds, in particular β-1′ (eg. 1-(3′,4′-diethoxyphenyl)-1,3-dihydroxy-2-(4″-methoxyxyphenyl)-propane) and β-O-4 compounds (eg. guaiacylglycerol-β-guaiacyl ether), where the oxidation of VA to veratraldehyde can be measured spectrophotometrically by an increase in absorbance at 310 nm [100, 163, 278].

### 1.3.1.3 Manganese and versatile peroxidases

Manganese peroxidases (MnP) are also haem-glycoproteins, in this case of 38-62.5 kDa, again with a very low pH optimum (3.0-4.0) but with a slightly lower redox potential of 0.8-0.9 V. A MnP was first identified in *P. chrysosporium* but others have subsequently been found in species belonging to the Meruliaceae, Coriolaceae, Polyporaceae, Strophariaceae and Tricholomataceae [100, 163]. Eleven different isoforms of MnP have been identified in *Ceriporiopsis subvervispora* [100, 163]. The catalytic cycle of the MnPs is similar to that of the LiPs, except that Mn$^{2+}$ acts as the redox mediator instead of veratryl alcohol. The principle MnP reaction is oxidation of Mn$^{2+}$ to Mn$^{3+}$, which then diffuses and oxidises monomeric and dimeric phenolic substrates through the formation of phenoxy radicals. The highly reactive Mn$^{3+}$ also chelates with organic acids (which can act as Mn$^{3+}$ stabilizers) such as oxalate, gluconate,
malaonate and 2-hydroxybutyrate to efficiently oxidise those substrates (Figure 1-6B).

The overall redox reaction is represented as

\[
\text{Native MnP} + \text{H}_2\text{O}_2 \rightarrow \text{MnP-I} + \text{H}_2\text{O}
\]

\[
\text{MnP} + \text{Mn}^{+2} \rightarrow \text{MnP-II} + \text{Mn}^{+3}
\]

\[
\text{MnP II} + \text{Mn}^{+2} \rightarrow \text{Native MnP} + \text{Mn}^{+3}.
\]

MnPs cannot degrade as many non-phenolic compounds as LiPs, but they are able to oxidize some, such as veratryl alcohol, anisyl alcohol, and benzyl alcohol, in the presence of thiols. The Mn\textsuperscript{+3} oxidises thiols to thiyl radicals which can then extract protons from such substrates. Interestingly there appears to be some overlap in the regulation of MnP and LiP production; MnP production in *Bjerkandera* sp. increases in the presence of diffuser oxidants, and excess MnP production also inhibits VA production, which in turn diminishes LiP production [100, 163, 190, 278].

**Figure 1-6.** Comparison of the catalytic mechanisms of lignin peroxidase (A) and manganese peroxidase (B). Both mechanisms begin with an initial 2e\textsuperscript{-} oxidation of native Fe III to compound I, which is then reduced to compound II and compound III in the presence of mediators such as VA and Mn\textsuperscript{+2}. The oxidised mediators then diffuse into the lignin polymer to attack it non-specifically. VA acts as a mediator for LiP and Mn\textsuperscript{+2} for MnP. See the text for further details [55].
Another type of peroxidase called versatile peroxidase (VP) has been identified in the genera *Pleurotus*, *Bjerkandera* and *Lepista*. VP has a hybrid structure with various oxidising sites connected to a haem co-factor and it has a mix of LiP and MnP characteristics, including efficient oxidation of a broad range of phenolic and non-phenolic compounds in the presence of mediators [190, 278]. The hybrid nature of versatile peroxide was demonstrated in a series of site-directed mutagenesis experiments to create a MnP activity in the LiP of *P. chrysosporium*, and vice versa. For example, introduction of a mutation at Ser168-Trp mutation in MnP made it capable of degrading a wide range of LiP substrates whilst retaining its natural MnP activities [100, 112, 278].

1.3.1.4 Laccases

Laccases are another class of enzymes reported to catalase lignin degradation. Laccases are 50 to 110 kDa multinuclear copper (Cu) oxidase enzymes with redox potentials of 0.6 to 0.9 V [278]. Examples include ascorbic oxidase (EC 1.10.3.3), nitrite reductase (EC 1.7.2.1) and ceruloplasmin (EC 1.16.3.1). They are found in plants, insects and microorganisms. In plants laccases play a role in lignin polymerization whereas in bacteria and fungi they are involved in lignin degradation [167]. Laccases have four copper atoms and they carry out one electron oxidations of four reducing substrates, with the four electrons stored in a fully reduced state of the enzyme. The four electrons in the reduced state of the enzyme can then reduce dioxygen to water. The redox reaction is represented as

\[
4RH + O_2 \rightarrow 4R + 2H_2O \quad [84, 112, 167].
\]
The four copper ions (T1Cu, T2Cu and two T3Cu) in the enzyme are distributed in three domains. The T1Cu site is an electron abstracting site, which process also gives rise to the blue color of the enzyme. Electrons abstracted at that site are then transferred to the T2 and T3Cu sites, where they are stored in a fully reduced state capable of reducing dioxygen to H2O [278]. The principle reaction of laccases is the oxidation of low molecular weight compounds to form the respective radicals, which then act as redox mediators and initiate rearrangement and ring opening reactions on molecules such as lignin [278]. For example the fungus *Pycnoporus cinnabarinus* is capable of degrading phenolic and non-phenolic compounds very efficiently by laccase-mediated redox mediators such as 2,2'-azinobis-(3-ethylbenzthiazoline-6-sulfonate) (ABTS), 1-hydroxybenzotriazole (HBT), and 3-hyroxyanthranilic acid (HAA) [112], which are oxidised to radical cations to cleave various C-C and C-O-C bonds in the substrates [278].

1.3.2 Lignin degradation by bacterial enzymes

Lignin degradation has been much less studied in bacteria than fungi and the work to date suggests bacteria generally have less activity against lignin [34]. Nevertheless several bacteria, generally Actinomycetes, have been shown to grow on different types of lignin, including Kraft lignin (*Novosphingobium* sp. B-7), 14C-labelled lignin (*Nocardia* and *Streptomyces* species) and native lignin (*Pseudomonas* and *Flavobacterium* species and *Streptomyces viridosporus* sp. T7A). Several other bacteria have also been isolated on the basis of their growth on model lignin β-O-4 compounds. These bacteria have been obtained from plantation soils (*Bacillus* sp., *Ochrobactrum* sp. and *Leucobacter* sp.), paper mill sewage soils (*Sphingomonas paucimobilis* SYK-6), and
xenobiotic contaminated soils (*Rhodococcus jostii* sp. RHA1). The latter two bacteria have been functionally characterized quite thoroughly with respect to their metabolism of the model β-O-4 compound guaiacylglycerol-β-guaiacyl ether (GGE) [4, 33]. Otherwise however, there has been little biochemical characterization of the lignin degrading enzymes themselves. On the other hand several convenient phenotypic and colorimetric assays have now been developed which could facilitate such work. Moreover recent advances in DNA sequencing technology now make whole genome sequencing of lignin degrading bacteria relatively straightforward and candidate oxidative genes found in the genome can then be expressed heterologously and tested for lignin activity using the assays above. Such an approach has recently been applied successfully to the lignin degrading Actinomycete *Amycolatopsis* sp. 75iv2, which expresses many peroxidases ([29, 34, 44, 48, 49, 161, 203, 236, 256] and see below).

### 1.3.2.1 Bacterial peroxidases

The first bacterial lignin peroxidase (ALiP-P3) to be reported was from *Streptomyces viridosporus* T7A [29, 34, 49, 206]. This Actinomycete secretes ALiP-P3, which can degrade wheat straw lignin [49] into smaller aromatic subunits, with acid-precipitable, polyphenolic, polymeric lignin (APPL) also accumulating as an intermediate [282]. The enzyme was found to have activity against several model β-1 (diarylpropane) and β-O-4 (GGE) compounds, but was not as efficient in these reactions as the fungal LiPs above. Little is known about the structure or reaction mechanism of ALiP-P3 [206], but similar oxidation reactions of these substrates have been reported for four stereospecific dehydrogenase enzymes isolated from *Sphingomonas paucimobilis* SYK-6 ([29, 34] and see below).
The first detailed characterization of a bacterial lignin degrading enzyme involved *Rhodococcus jostii* RHA1, a gram positive bacterium isolated from polychlorinated biphenyl contaminated soil (Figure 1-7). A comparison of the *R. jostii* RHA1 genome with other lignin degrading microbes revealed the existence of two previously unannotated dye-decolorizing type peroxidases (Dyps) which were then annotated as DypA and DypB peroxidases [212]. The DypA has 41 % and 36 % similarity with the dye-decolorizing peroxidase of *Bacteroides thetaiotaomicron* (strain VPI-5482) (BtDyp) and the encapsulin protein of the fungus *Thermotoga maritime*, respectively. DypB has 27-35 % similarity to the iron acquisition proteins EfeB, EfeO and EfeU of *E. coli* K-12 [212]. Heterologous expression of DypA and DypB showed both to be active against classical peroxidase substrates such as Mn$^{+2}$, pyrogallol and anthraquinone dyes, ABTS and Reactive Black [29, 212]. DypA has relatively higher specificity for ABTS pyrogallol and Reactive Black 4, and DypB is relatively more active on Mn$^{+2}$ and ABTS. DypB has been further characterized in detailed kinetic and crystallographic studies [212], with several lines of evidence showing that it behaves like the fungal peroxidases. It produces the same products, including the intermediate compounds I and II (Figure 1-7), as do the fungal peroxidases, but with less efficiency.
Figure 1-7. Catalytic mechanism of DypB peroxidase enzyme from *R. jostii* RHA1, showing rate constants for β-aryl ether turnover, the formation of vanillin and the reaction with Mn$^{2+}$ [4].

Note: Adapted with permission (see Appendix 2) from Ahmad M, Roberts JN, Hardiman EM, Singh R, Eltis LD, Bugg TDH: Identification of DypB from *Rhodococcus jostii* RHA1 as a lignin peroxidase. Biochemistry 2011, 50(23):5096-5107.

Recent bioinformatic analysis of the genome of the biomass degrader *Amycolatopsis* sp. 75iv2 mentioned above has also revealed the presence of three Dyps (Dyp1 to 3) [29]. Dyp1 and Dyp2 belong to the C-clade of Dyps which is more closely related to the fungal peroxidases, while Dyp3 is grouped with other bacterial enzymes, such as the lignin degrading enzymes from *R. jostii* RHA1 [29, 212, 282]. Dyp2 is characterized as an unusual C-type dye-decolorising peroxidase in the haem peroxidase family [212]. Heterologously expressed Dyp2 is active against Mn$^{2+}$, Reactive Black 5 and ABTS dyes and it is highly active against phenolic substrates as measured in 2,4-dichlorophenol peroxidase assays [212]. In the absence of mediators, Dyp2 readily breaks model β-O-4 phenolic compounds but does not react with non-phenolic...
compounds, perhaps because it has relatively low redox potential sites. The high $k_{\text{cat}}/K_M$ values of Dyp2 against its good substrates are close to those of the MnP from *Phanerochaete chrysosporium* and the versatile peroxidase of *Pleurotus eryngii* for those substrates [212]. Dyp2 can also catalyse the oxidative decarboxylation of 4-methoxymandelic acid (MMA) in the presence of both Mn$^{2+}$ and O$_2$, reminiscent of the MMA oxidation seen for the LiP of *P. chrysosporium* in the presence of veratryl alcohol and H$_2$O$_2$ [49, 282]. Given the extent of its activities, further studies on Dyp2 would clearly be worthwhile.

1.3.2.2 Bacterial laccases

Laccases have been tentatively identified in several bacterial genera, predominantly *Bacillus*, but none have yet been thoroughly characterized at a functional level. The most promising evidence comes from *Bacillus atrophaeus* and *Bacillus pumilus*, which were originally isolated in a screen for growth on the laccase substrate ABTS. *B. atrophaeus* was also found to grow on guaiacylglycerol-β-guaiacyl ether (GGE) and Kraft lignin, albeit neither of these compounds were completely mineralised by this bacterium. Further studies will be required to confirm the laccase activity in these two strains [107, 205].

1.3.2.3 Non-peroxidase bacterial enzymes

Several intracellular bacterial enzymes with lignin degrading activity have been identified from various genera. Of these, the enzymes from the gram negative bacterium *Sphingomonas paucimobilis* SYK-6 are the most extensively characterized. This bacterium was isolated from Kraft pulp effluent [160, 206], and found to have activity against several dimeric lignin compounds such as GGE (Figure 1-8) via an oxido-
reductive mechanism. The pathway for this has been elucidated and is shown in Figure 1-9. Initially, GGE is oxidised by NAD dependent, short chain Cα-dehydrogenase enzymes (LigD, LigL, LigN and LigO) to α-(2-methoxyphenoxy)-β-hydroxypropiovanillone (MPHPV) [218], which in turn is a carbonyl substrate for glutathione S-transferases (GSTs) (ligF and ligE and ligP) [158, 212]. The MPHPV is then reductively cleaved at C-O-C linkages (≈ 80 % lignin has C-O-C linkages) by β-etherases (ligF and ligE and ligP), which adds a hydrogen to generate α-glutathionyl-β-hydroxypropiovanillone (GS-HPV) and guaiacol. The glutathione attached to the HPV is then removed by a glutathione lyase enzyme (ligG) [4, 158]. Both the dehydrogenases and the GSTs show stereoselectivity among the four GGE and two MPHPV enantiomers respectively (Figure 1-8). The stereoselectivities were elucidated with an end-point assay using heterologously expressed E. coli crude cell extracts [158, 164, 218, 248].

![Figure 1-8. Stereoisomers of GGE and MPHPV. GGE exists in two diastereomers (Threo and Erythro) and each diastereomer exists in two enantiomers (-)-(αS,βR)-GGE, (+)-(αR,βS)-GGE (Threo) and (-)-(αS,βR)-GGE, (+)-(αS,βS)-GGE (Erythro). MPHPV exists in two enantiomers, (+)-(βS)-MPHPV and (+)-(βR)-MPHPV [218].](image)
Figure 1-9. Catabolic pathway for GGE in *Sphingobium* sp. strain SYK-6. The enzymes ligD, ligL, ligN and ligO are dehydrogenases which convert GGE to MPHPV. The MPHPV is then cleaved by the ligE, ligF and ligP GST enzymes to give GS-HPV and guaiacol. The glutathione attached to HPV is cleaved by ligG, a GST lyase enzyme. Abbreviations: GS−, reduced glutathione; GSSG, oxidized glutathione; GS-HPV, α-glutathionyl-β-hydroxypropiovanillone; HPV, β-hydroxypropiovanillone [218].

Interestingly the fungal strain *Chaetomium sp.* 2BW-1 is also known to have a GST enzyme which is involved in the GGE degradation, but it does not act on MPHPV as above [177].

*Streptomyces* sp. SirexAA-E is a cellulolytic bacterium known to have an aromatic ring dioxygenase (SACTE_2871) fused with a carbohydrate binding domain (CBM 5/12). SACTE_2871 binds to lignin compounds through the CBM 5/12 domain and catalyses the O₂ dependent intra-diol cleavage of various non-methoxylated subunits of lignin and other compounds [29, 205].

Several other enzymes such as dehydrogenases, hydrolases, demethylases and peroxidases of bacteria and fungi are known to synergistically enhance lignin degradation [59]. For example a cellulobiose dehydrogenase from the fungus *Schizophyllum commune* enhances lignin degradation in the presence of the MnP of *Phanerochaete chrysosporium* [85]. The fungi *Phlebia radiata*, *Dichomitus squalens* and *Ceriporiopsis subvermispora* are also known to produce lignin degrading peroxidases.
(LIPs and MnPs) and laccases and when they are co-cultivated they show enhanced rates of enzyme production and lignin degradation [59]. The fungus Rigidoporus lignosus secretes a laccase and a MnP which do not solubilize $^{14}$C lignin individually but do so when combined together [85].

### 1.4 Nitroaromatic compounds

Nitroaromatic compounds (NACs) are organic compounds containing one or more nitro (-NO$_2$) groups on a benzene ring (Figure 1-10). They are comparatively rare in nature (some of them being antibiotics) but many synthetic NACs are used in the production of explosives, pesticides, herbicides, pharmaceuticals and several speciality chemicals [11, 90, 156, 239]. NACs are synthesized via nitration reactions with nitric and sulphuric acids, the nitronium ions (NO$_2^+$) generated being added to the benzene ring via electrophilic substitution reactions. The simplest NACs synthesized are nitrobenzene, nitrotoluene and various nitrophenols, which are precursors for a wide range of products [238]. The nitration reactions can be controlled to allow substitution of the nitro group onto the ortho, para and meta positions of the benzene ring, as well as for replacing the halogens in halogenated aromatics (Zinke nitration) [99]. Multiple nitro group compounds such as 1,3,5-trinitrobenzene, 2,4,6-trinitrotoluene and 1,3,5-trinitrophenol can also be synthesized by Wolffenstein-Boters reactions (concurrent oxidation and nitration of simple aromatics to nitrophenols with higher oxides such as nitric acids in the presence of catalysts such as mercury salt) [99]. The unusual substitution or addition reactions to the benzene ring make NACs recalcitrant to biological degradation.
1.4.1 Chemistry of the nitroaromatics

The distinctive chemistry of the NACs gives them distinctive redox properties. The nitrogen bonded to two oxygen atoms, which have more electronegativity than the nitrogen, and the resultant nitrogen-oxygen bond polarization (electronegativity difference 0.5) cause the nitrogen to carry a partial positive charge, which therefore acts as a strong electrophile. This and the presence of the oxygen make the nitro group a strong oxidising agent. Furthermore, once this group is reduced the ring becomes activated and susceptible to oxidation. Overall, three functions combine to make some nitroaromatics explosives; the oxidising potential of the nitro groups, the abundant availability of oxidisable carbon in the activated ring, and the fact that the latter reaction produces both heat and gaseous nitrogen and carbon monoxide [11, 90, 156, 239, 240].

\[ \text{Nitrobenzene} \]

\[ \text{Nitrotoluene} \]

\[ \text{4-nitrophenol} \]

\[ \text{1,3,5-trinitrobenzene} \]

\[ \text{2,4,6-trinitrotoluene} \]

\[ \text{Picric acid} \]

Figure 1-10. Examples of nitroaromatics.
1.4.2 Synthetic and natural nitroaromatics

Synthetic NACs such as picric acid (2,4,6-trinitrophenol), 2,4,6-trinitrotoluene (TNT), cyclotrimethylenetrinitramine (RDX), cyclotetramethylenetetranitramine (HMX) and 2,4-dinitroanisole (DNAN) have been commonly used as the explosive ingredients in munitions [156, 238] (Figures 1-10 and 1-11). Picric acid was the first to be used for this purpose but its reaction with the metal in casting shells forms picric salts, making it more shock sensitive and susceptible to incomplete detonation [99, 204, 241, 268]. The TNT that was developed next is more stable and less sensitive to inadvertent detonation, but also less powerful compared to picric acid [137, 241]. However TNT also had other uses, as a starting material for the manufacture of several other industrially important heterocyclic aromatic compounds such as hexanitrostilbene (HNB) and 1,3,5-trinitrobenzene (TNB) [239]. Problems with its presence in the environment and the recent availability of other NACs which are more powerful have seen TNT use in munitions decline, although it is still used as a minor component of some munitions together with newer NACs such as RDX and HMX. The latter are bigger molecules with more nitrogen bonds and they therefore release more energy when exploded [5, 47, 137, 241, 257]. DNAN is now finding increasing use, its greater shock friction and thermal stability giving it important advantages in terms of safety during manufacture, transport and handling [62, 73, 84, 146, 192].

As noted, in addition to their use as explosives, many NACs are also used in making various herbicides, insecticides, ovicides and fungicides [11, 90, 240]. Derivatives of nitrobenzenes, nitrotoluenes and nitrophenols are used as the starting materials for many of these. For example mono-nitro compounds are used in the
synthesis of carbofuran, parathion, fluorodifen, nitrofen, and bifenox [90, 238], while di-nitro compounds are used to make 2,5-dinitro-o-cresol, dinoseb, and binapacryl. Reduced NACs form amines such as aniline and some major pharmaceuticals such as paracetamol are also made from NAC feedstocks [156, 238].

![1,3,5,7-tetranitro-1,3,5,7-tetrazocane (HMX)](image)

![1,3,5-trinitro-1,3,5-triazinan (RDX)](image)  ![2,4-dinitroanisole (DNAN)](image)

**Figure 1-11.** Further examples of nitroaromatic explosives [117].

Some complex NACs are also made by biotic or abiotic processes in the environment. Some of the abiotic processes start from incomplete combustion of fossil fuels, generating hydrocarbons which can react with atmospheric nitrogen dioxide to form simple nitroaromatic compounds such as nitrobenzenes and nitrotoluenes, as well as various nitro-polyaromatic hydrocarbons (nitro-PAHs) such as nitronaphthalenes and nitrobianaphs [117, 240]. In aqueous environments, sunlight can also catalyse the nitration of dissolved hydrocarbons [117, 240]. Plants, fungi and bacteria can also produce a limited range of nitroaromatic compounds. In particular *Streptomyces* species are well known to produce various nitroaromatic antibiotics [90, 156]. Chloramphenicol
from *Streptomyces venezuelae* and aureothin from *Streptomyces thioluteus* are perhaps the best known but several others such as neoaurithinin, thaxtomin, azomycin, rufomycins and dioxapyrrolomycin are also produced by various *Streptomyces*. Gram negative bacteria from the genera *Pseudomonas*, *Enterobacter*, *Coralloccocus* and *Burkholderia* are also known to produce the antifungal compound pyrrolnitrin [117], while pyrrolomycins A, B, and E, which are anti-fungal and anti-bacterial, are produced by the gram positive *Actinosporangium vitaminophilum* [117]. Several other bioactive NACs such as the siderophores (iron chelating compounds) are also produced by *Streptomyces* species [117, 156, 210, 238, 239].

### 1.4.3 Biodegradation of nitroaromatics

Many of the properties of NACs which make them useful also make them hazardous to human health and the environment. Many are toxic, mutagenic and carcinogenic. Moreover many are very stable and persistent in the environment. Accordingly several are listed in the U.S. Environmental Protection Agency’s list of priority pollutants [117, 156, 240].

Details of their mutagenic and carcinogenic effects are influenced by the type of nitroaromatic and the attached functional groups [146, 182, 186, 199, 201]. However, in general terms, studies on *Salmonella*, *E. coli* and mammalian cell lines have shown that mono-, poly- and, heterocyclic NACs can all cause transversions and frameshift mutations in genetic material [54, 117, 156]. Various oxidation and reduction products and other derivatives of NACs can also interfere with DNA synthesis. NACs are also known to cause methaemoglobinemia, neurotoxicity, nephrotoxicity and several other reproductive and respiratory disorders in humans and animals [156]. NACs are released
into the environment through fossil fuel (petrol and diesel) spills, agricultural uses of many pesticides, insecticides and fungicides, and several munitions-related activities, mostly through the use and disposal of explosives and in training operations [209]. Industrial accidents have also released high concentrations of NACs into the environment; for example, an explosion in a China National Petroleum plant in Jilin City in 2005 released approximately 100 tons of benzene and nitrobenzene into nearby rivers and eventually the Pacific Ocean [135].

Thermochemical treatments are available to remediate many NAC contaminated environments relatively efficiently but they are energy-intensive and expensive and not suitable for treating broad scale contamination. There is therefore considerable interest in developing biological methods for their decontamination [156]. There are practical challenges here, such as bioavailability, and, given that many NACs have been relatively rare in nature, microorganisms have generally had little reason to develop catabolic pathways for them. Some microbes can degrade some of the simpler NACs such as nitrobenzoate and nitrophenols via well-established oxidative and reductive pathways, but many of the more complex NACs such as TNT and 2,4-dinitroanisole have proven relatively refractory to microbial degradation [94, 156].

To date, several bacteria belonging to the genera *Pseudomonas, Burkholderia, Comamonas, Arthrobacter, Acidovorax and Rhodococcus* have been isolated for their ability to degrade a range of relatively simple NACs such as nitrobenzenes, nitrotoluenes and nitrophenols [8, 239]. In aerobic conditions the oxidation of such NACs is initiated by mono- or dioxygenase enzymes, which involves introduction of hydroxyl groups by spontaneous elimination of nitro groups. This type of reaction was identified in the
degradation of 2-, 3- and 4-nitrophenols. In *Pseudomonas putida* B2 the 2-nitrophenol is oxidised by a NADH-dependent monooxygenases enzyme while a dioxygenase-catalysed reaction was observed in *Alcaligenes eutrophus* JMP222 against 2,6-dinitrophenols. However some NACs can be catalysed via reduction, due to the electrophilic nature of the nitro groups. In *P. putida* TW3, the metabolism of 3-nitrobenzoate is catalysed by a NADPH-dependent reductase to 4-hydroxylaminobenzoate, which is further metabolized by subsequent deamination reactions. In some *Rhodococcus* and *Nocardioides* species the NACs are reduced by forming a hydride-Meisenheimer complex with a NADPH-dependent reductase also containing cofactor F₄₂₀ [65, 117].

### 1.5 Thesis outline

This thesis describes the degradation of two very different complex aromatics, lignin and 2,4-dinitroanisole (DNAN), by newly isolated bacteria, and key enzymes therein. While the breakdown of both is chemically challenging, the lignin is potentially an abundant source of carbon to support microbial growth and the DNAN can provide both carbon and nitrogen. Unlike the lignin, the DNAN is also toxic unless broken down, putting further selective pressure on the evolution of effective catabolic pathways. In the case of lignin, evolution has had vast timeframes in which to develop useful pathways but for DNAN the timeframe has been relatively short, even for organisms like bacteria and archaean which have comparatively shorter generation times, large population sizes and considerable capacity for horizontal gene transfer. My goals are to explore the biochemical options that have evolved under the two scenarios and to assess their potential as the foundations for biotechnologies that might utilize the lignin
effectively in chemical manufacturing or decontaminate toxic NACs effectively in the environment.

Chapter 2 describes a bacterium isolated from paper mill samples which can oxidize the major lignin substructure GGE. This includes the annotation of candidate dehydrogenase genes identified in its genome sequence. The purification of the cognate enzymes and their biochemical characterization are then described, along with a phylogenetic consideration of their functions.

Chapter 3 describes another bacterium isolated from an activated sludge sample from the Holston Army Ammunition Plant in the US on the basis of its ability to mineralize DNAN. The first step of DNAN mineralization is shown to be hydrolytic removal of a methyl group to methanol and 2,4-dinitrophenol (2,4-DNP). The latter is mineralized through a well-established catabolic pathway [73]. The purification and characterization of the novel demethylase are then described.

Chapter 4 is a general discussion of the findings described in the two preceding chapters in the contexts of both the broad evolutionary questions raised above and potential biotechnological applications in chemical manufacturing and bioremediation.
Chapter 2  Phylogenetic and kinetic characterization of a suite of dehydrogenases from a newly isolated bacterium that catalyse the turnover of guaiacylglycerol-β-guaiacyl ether stereoisomers

2.1 Introduction

The lignin polymer, representing 15-40 % of plant-derived biomass, is a potential renewable alternative to petrochemical feedstocks for chemical manufacturing industries, particularly with regard to aromatic compounds [202]. One of the major hurdles for the utilization of lignin in this way is its recalcitrance; the polymeric structure of lignin coupled with the nature of its chemical bonds renders it highly resistant to chemical or biological degradation. Chemical degradation of lignin has been utilized by some industries but it usually involves non-selective destruction of lignin via burning or treatment under highly alkaline conditions, which renders it less useful or of lower value to many downstream applications [101]. The biological depolymerization of lignin should be a more selective and energy-efficient process, and therefore potentially a cost effective and environmentally sustainable alternative to the chemical processes currently employed [101]. However, to date, the identification of suitable biocatalysts involved in lignin depolymerization has proven difficult.

Lignin has a polyaromatic structure with more than five different types of intermonomer linkages [275]; the β-O-4 linkage (also known as the β-aryl ether linkage) represents 45-70 % of these linkages [33, 159]. The first step in the biological
degradation of the lignin polymer occurs non-specifically through the action of laccases and extracellular peroxidase enzymes (manganese peroxidase, versatile peroxidase, and lignin peroxidase) via radical ion mechanisms [142, 155, 159]. While fungi are thought to be the main contributors to biological lignin degradation, recent reports suggest that certain bacterial strains may also play a role in lignin polymer or kraft lignin degradation [4, 28, 29, 34, 39, 149, 151, 165, 215, 229, 230, 232, 242, 250, 272]. While the early steps involved in the biological degradation of the lignin polymer are non-specific, a few bacteria have shown potential in directly and specifically degrading smaller units of the lignin polymer into industrially useful chemicals [266, 285].

In the 1980s, dehydrogenase activities from Pseudomonas spp. were found which could catalyze Cα-alcohol oxidation of β-aryl ether- or diarylpropane-linked lignin dimers [92, 188]. In this case, the specific proteins responsible were not identified and the degradation pathways were not further explored. Subsequently however, specific dehydrogenases catalyzing such reactions have been characterized from another bacterium, Sphingomonas paucimobilis SYK-6. SYK-6 has been reported to degrade several different types of model lignin dimers [159], including a molecule with a β-O-4 linkage known as guaiacylglycerol-β-guaiacyl ether (GGE). The proposed pathway for the degradation of GGE in this organism begins with the oxidation of GGE via one of several Cα-dehydrogenases (LigD, LigO, LigL, or LigN) [218], followed by ether bond cleavage via one of several glutathione S-transferases (GSTs; LigE, LigF, LigP) [158, 248], and then glutathione removal via one of several GSTs (LigG and probably others yet to be identified) [81, 158] (see Figure 2-3A below). These reactions generate the end products guaiacol and β-hydroxypropiovanillone (HPV), the latter of which is eventually fed into the protocatechuate (PCA) 4,5-cleavage pathway [33, 159]. In SYK-6, the biological
conversion of GGE to HPV and guaiacol is performed by a series of stereoselective enzymes [81, 82, 218]; there are two stereocentres in the GGE molecule and therefore four possible stereoisomers. Expression of the Cα-dehydrogenases LigD, LigO, LigL and LigN in *Escherichia coli* and gene disruption experiments in SYK-6 showed that these enzymes are responsible for the first step in the pathway, collectively oxidizing all GGE stereoisomers but individually exhibiting preferences for one or two of them [218].

Here I report the isolation of a bacterium (SG61-1L) that is able to degrade the GGE model lignin dimer at a significantly faster rate than SYK-6. These results led to my identification of individual enzymes from this bacterium with higher activities and a broader range of stereoisomer specificities than the SYK-6 dehydrogenase above.

### 2.2 Materials and Methods

#### 2.2.1 Materials

(+)/(-)-GGE (1-(4-hydroxy-3-methoxyphenyl)-2-(2-methoxyphenoxy) propane-1,3-diol)) was custom synthesized (at > 95 % purity) by Albany Molecular Research Inc. (Albany, NY, USA). The percent composition of each stereoisomer in this GGE was as follows: 39.5 % (αS,βR)-GGE, 39.5 % (αR,βS)-GGE, 10.5 % (αS,βS)-GGE, and 10.5 % (αR,βR)-GGE. The four individual GGE stereoisomers, (αR,βS)-, (αS,βR)-, (αS,βS)- and (αR,βR)-GGE, were a gift from Dr Eiji Masai (at 100 %, 88.8 %, 96 %, and 93 % purity, respectively) [98]. Vanillin, vanillic acid, guaiacol and all other analytical grade chemicals were purchased from Sigma Aldrich (St. Louis, MO, USA) unless otherwise specified. *Sphingobium* sp. strain SYK-6 was purchased from the Biological Resource Centre (NBRC) culture collection (Kisarazu, Chiba, Japan). *Pseudomonas putida* KT2440 was purchased from DSMZ (Braunschweig, Germany).
2.2.2 Media and growth conditions

Minimal medium (MM) was prepared by adding 1.36 g KH₂PO₄, 1.78 g Na₂HPO₄ 2H₂O, 0.50 g MgSO₄ 7H₂O, and 0.50 g NH₄Cl to 900 ml of double distilled water. The pH was adjusted to 7.2 with a 1M NaOH solution, and 1 ml of trace element solution (prepared as a 1:1000 stock in one litre of distilled water consisting of 0.10 g Al(OH)₃, 0.05 g SnCl₂ 2H₂O, 0.05 g KI, 0.05 g LiCl, 0.08 g MgSO₄, 0.05 g H₃BO₃, 0.10 g ZnSO₄ 7H₂O, 0.01 g CoCl₂, 0.01 g NiSO₄ 6H₂O, 0.05 g BaCl₂, and 0.05 g (NH₄)₆Mo₇O₂₄ 4H₂O) was added. No vitamins were added to the MM. The solution was filter-sterilized with 0.22 µm Fast Cap-60 filters (Pall Life Sciences, Port Washington, NY) and diluted to a final volume of 1 litre. The nutrient agar (NA), nutrient broth (NB) and Luria broth (LB) used as rich media for bacterial growth were purchased from Becton, Dickinson and Company (Franklin Lakes, NJ, USA).

2.2.3 Instrumentation and analytical methods

Samples from growth and resting cell experiments were analyzed with an Agilent 1200 Infinity series time of flight (TOF) liquid chromatography/mass spectrometry (LC/MS) instrument (Agilent Technologies, Santa Clara, CA). Metabolites analyzed by TOF LC/MS were quantified through the use of a diode array detector (DAD) with an extracted wavelength range of 275-280 nm using the Applied Biosystems Analyst® QS 1.1 software package (Life Technologies, Carlsbad, CA). Standard curves were constructed for GGE, guaiacol, vanillin, and vanillic acid and the identities of these metabolites throughout growth and resting cell experiments were confirmed by comparison to the retention time in LC and TOF MS spectral data (see Table 2-2 below).
A standard curve for NADH (Figure 2-1) was constructed by adding varying amounts of NADH to wells otherwise simulating the kinetic assay reaction conditions (50 mM Tris pH 8.0 at 28 °C). The MPHPV standard curve was constructed following its biosynthesis from GGE as follows: a GGE dehydrogenase reaction was set up (see below) to convert the GGE substrate to MPHPV. The peak corresponding to MPHPV was extracted from the LC-MS and any remaining GGE was quantified. The amount of biosynthetic MPHPV was quantified by subtracting the remaining GGE from the amount of GGE added to the initial reaction. The MPHPV fraction from the LC-MS was then evaporated to dryness and resuspended in 50 mM Tris pH 8.0. The standard curve for MPHPV (Figure 2-1) was then constructed as for the NADH standard curve above. The identity of HPV (for which a standard was not available) was assigned based on TOF MS spectral data reported previously for the SYK-6 bacterium [154].

Runs were performed with either the Grace Alltima C18 (4.6 mm x 250 mm) column (W.R. Grace & Co, Columbia, MD) or the CHIRALPAK IE (4.6 X 250 mm) chiral column (Daicel Corporation, Tokyo, Japan). An isocratic method was developed for use with the C18 column with a mobile phase of acetonitrile-water-formic acid (55:45:0.1) and a flow rate of 0.8 ml min$^{-1}$ for 15 min. All substrates and intermediates were monitored at 275-280 nm and quantified by peak area. A separate isocratic method was developed for use with the chiral column, involving a mobile phase of water-acetonitrile-ethyl acetate (85:15:2) containing 0.1 % formic acid and using a flow rate of 0.8 ml min$^{-1}$ for 35 min. All stereoisomers were quantified by peak area from DAD detection at the extracted wavelength range of 265-300 nm using the Agilent time-of-flight (TOF) software (version A.01.00) (Agilent Technologies). The separation of the four stereoisomers achieved is shown in Figure 2-2.
Figure 2-1. Construction of NADH and MPHPV standard curves.
Figure 2-2. Chiral column separation of the four GGE stereoisomers. The GGE mixture (top panel) and each individual stereoisomer (bottom four panels) were run separately through the chiral column on the TOF LC-MS (see section 2.2.3).

2.2.4 Isolation of GGE-degrading bacteria

Bacteria were isolated from environmental samples provided by the Australian Paper Maryvale Mill (Victoria, Australia, March 2012). The samples were collected from several sewage sludge waste sites and from the membrane bioreactor at the paper mill.
In the first round of enrichments, 1 g of the combined samples was inoculated into 100 ml of MM containing 157.5 µM GGE as a sole carbon source. Enrichments were incubated in 500 ml baffle flasks at 28 °C on a shaker at 200 rpm and monitored for loss of substrate every 24 hours by TOF LC/MS with the Grace Alltima C18 column using the isocratic method as described above. Once the substrate concentration had dropped below 63 µM, 1 ml of the culture was transferred to the next generation of enrichments and the process above repeated 11 times. Finally, the enrichments were plated onto quarter-strength NB agar plates and incubated at 28 °C. After one week, morphologically different bacteria were isolated and individually tested for GGE degrading activity on the TOF LC/MS. The GGE degrading bacteria were identified from their full-length 16S rRNA sequences as obtained from genome sequencing (described below). The most promising bacterial degrader of GGE isolated from these experiments was an alpha-proteobacterium named SG61-1L.

2.2.5 Dye decolorizing experiments

Isolated bacteria were streaked onto NB-agar plates containing either Azure B (Sigma Aldrich) or Reactive Black 5 (Sigma Aldrich) at a final concentration of 0.4 mM. To prepare the plates, an aqueous dye solution was sterile filtered and then added to autoclaved quarter strength NB-agar. Plates were incubated at 28 °C and observed for clearing of the dye accompanied by bacterial growth.
2.2.6 Growth and resting cell experiments

SG61-1L and SYK-6 were grown in NB media at 28 °C and harvested at mid log phase (at an OD$_{600}$ of 0.7) by centrifugation at 4,648 x g for 15 min at 20 °C. The cells were washed twice with MM and resuspended (to a final OD$_{600}$ of 0.7) in MM or 20 mM phosphate buffer (pH 7.2) for growth culture and resting cell experiments, respectively.

In the first set of growth culture experiments, washed cells were inoculated at 1 % w/v into MM (pH 7.2) containing either 200 µM GGE or 330 µM vanillin as a sole carbon source and shaken at 200 rpm at 28 °C. SG61-1L cells were pelleted again following the disappearance of GGE and resuspended in the same volume of fresh MM supplemented with 200 µM GGE, and both degradation and growth were again monitored over time at 28 °C. Samples were collected at various time points, filtered, and analyzed by TOF LC/MS using the Grace Alltima C18 column.

To determine relative bacterial degradation rates for each GGE stereoisomer, a second set of growth culture experiments was carried out as described above in MM supplemented with 157.5 µM GGE. The relative levels of each stereoisomer present in the culture were monitored over time and analyzed via chiral separation of the four different GGE stereoisomers using the chiral column as described above. The four stereoisomers were identified based on the correlation with elution profiles and mass spectral data from authentic standards [98].

As a negative control, a third bacterium, Pseudomonas putida KT2440 was separately tested in growth cell experiments with GGE as a sole carbon source according to the protocols described above.
In resting cell experiments, 35 ml of cells were grown in NB at 28 °C and harvested at log phase (OD$_{600}$ values were 0.63 and 0.67 for SG61-1L and SYK-6, respectively), washed twice with phosphate buffer and resuspended in 100 ml MM containing 157.5 µM GGE as the only carbon source. Samples were collected and analyzed by TOF LC/MS using the C18 column as described for the growth culture experiments.

All the above experiments were performed in triplicate.

2.2.7 Genome sequencing, nucleotide sequences and accession numbers

Genomic DNA was isolated from a pure culture of SG61-1L using Qiagen Genomic DNA buffers and 500/G genomic tips (Qiagen, Hilden, Germany). Short (500 bp) and long (2000 bp) insert libraries were sequenced with MiSeq Illumina technology by the Micromon DNA sequencing facility (Monash University, Melbourne, Australia). The raw reads were then assembled with the SPAdes assembler version 3.5.0 [14] which performs error correction, assembly, scaffolding and mismatch correction. The final assembly was 3,771,711 bp long with a G/C content of 63 %, and was comprised of three scaffolds 3.6 Mb, 94 kb and 10 kb in length. Annotation of this assembly was performed by the NCBI Prokaryotic Genome Annotation Pipeline and identified 3,209 protein coding sequences, 49 tRNA sequences and two rRNA clusters. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under accession no. JXQC00000000. The version used in this thesis is the first version, JXQC01000000.
#### 2.2.8 Cloning and overexpression of dehydrogenase genes

Thirteen candidate dehydrogenase genes were selected from the SG61-1L genome sequence based on two sets of criteria: (1) percent inferred amino acid identity to the four SYK-6 dehydrogenases, and (2) genomic context: i.e., proximity to genes that have potential to function elsewhere in the GGE degradation pathway.

Eleven of these thirteen candidates from SG61-1L in addition to the four genes originally characterized from SYK-6 (LigD, ligO, LigL, LigN) [218], were PCR-amplified using Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) with Gateway® primers containing attB sites (Table 2-1) and an N-terminal 6x histidine tag. AttB-flanked PCR products were subcloned into the pDNR201 vector using BP-colonese II and subsequently cloned into the pDEST14 destination vector using LR-clonase II according to the manufacturer’s protocol (Life Technologies, California, USA). The remaining two gene candidates from SG61-1L, 724 and 3329, were PCR amplified (primer sequences are in Table 2-1) and subsequently cloned into pETcc2, a pET14b-derived vector (Novagen, Madison, WI) containing an N-terminal 6x histidine tag and thrombin cleavage site.

All 17 expression plasmids were subsequently transformed into *E. coli* BL21(DE3) pLysS cells (Life Technologies). One colony of each was inoculated into LB media and grown overnight at 37°C. This culture was then transferred into a one litre culture of LB media and grown at 37 °C to an OD$_{600}$ of 0.8-1.0, at which point 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added for induction. The culture was then transferred to 28 °C and shaken overnight. Induced cultures were harvested at 6,000 x $g$ for 20 min at 4 °C.
<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5'→3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG61_474_N6xH_for</td>
<td>CCATGCATACACATACACATACACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_474_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61-724_xho1_rev</td>
<td>TATCTCGAGTCACTGACATCAGCAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61-724_ndel_for</td>
<td>ATTACATATGAAAGATTTTGCGGACCAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_1498_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_1498_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_2549_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_2549_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_2550_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_2705_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_2705_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_2706_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_2706_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_3175_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_3175_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_3329_ndel_for</td>
<td>ATTACATATGAAAGATTTTGCGGACCAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_3329_bamhl_rev</td>
<td>TATTTGACATGCTAGTCACTGACATGCTAGGATGGGATGGGACATTC</td>
</tr>
<tr>
<td>SG61_3344_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_3344_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_3726_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_3726_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_3730_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>SG61_3730_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligo-syk6_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligo-syk6_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligoN-syk6_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligoN-syk6_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligoD-syk6_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligoD-syk6_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligoL-syk6_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligoL-syk6_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligoO-syk6_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligoO-syk6_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligoN-syk6_N6xH_for</td>
<td>CCATGCATACACATACACATACATGAAAGACACTCGACAGGATTCGACATTTC</td>
</tr>
<tr>
<td>ligoN-syk6_N6xH_rev</td>
<td>CACCTCTGACAAAGATGCTCTGATGACAGCATTCGAGGATTCGACATTTC</td>
</tr>
</tbody>
</table>
2.2.9 Protein purification

The harvested pellet from each induced culture was resuspended in lysis buffer (10 % glycerol, 50 mM Tris pH 8, 100 mM NaCl and 20 mM imidazole) and disrupted at 15,000 psi on a Microfluidics M-110P homogenizer (Microfluidics, Westwood, MA) at 4 °C. The lysate was then centrifuged at 18000 x g for 30 min at 4 °C. The supernatant was applied to a gravity column containing nickel-nitrilotriacetic acid (Ni-NTA) resin (Qiagen), washed with lysis buffer, and eluted in elution buffer (10 % glycerol, 50 mM Tris pH 8, 100 mM NaCl, 250 mM imidazole). The protein samples were dialyzed overnight at 4 °C in dialysis buffer (50 mM Tris pH 8, 100 mM NaCl) and protein concentrations were measured with a NanoDrop ND-1000 UV-Vis Spectrophotometer (Thermo Scientific, Waltham, MA); NanoDrop values were corrected based on theoretical extinction coefficients.

2.2.10 Enzymatic transformation of GGE to MPHPV

LC-MS monitoring of enzymatic transformation from GGE to MPHPV was performed using the following reaction conditions: 50 μM of the purified enzyme, 20 mM Tris pH 7.5, 5 mM NAD+, and 100 μM GGE in a final volume of 100 µl. The reaction was incubated at 28 °C for 12 h and quenched by addition of acetonitrile prior to LC-MS analysis. The LC-MS method employed was as described above using the Grace Alltima C18 column.
2.2.11 Enzyme kinetics

All enzymes were assayed against each of the four GGE stereoisomers on a Spectromax-M3 plate reader (Molecular Devices, Sunnyvale, CA). Reactions were performed at 28 °C in a final volume of 50 µl containing 2 mM of either NAD⁺ or NADP⁺, 20 mM Tris pH 7.5, and concentrations of substrate ranging from 0.1 µM to 2 mM. The concentration of each enzyme was optimized for each reaction and ranged from 0.013 µM to 4.5 µM. All reactions were set up in a 384-well plate and monitored by the increase in absorbance at 340 nm over time, consistent with reduction of the cofactor NAD(P)⁺ to NAD(P)H. Initial velocities at each substrate concentration were corrected based on standard curves constructed for the two products, MPHPV and NADH (Figure 2-1). All kinetic parameters were calculated using GraphPad Prism (GraphPad Software, CA, USA).

2.2.12 Phylogenetic Analysis

A phylogenetic tree was constructed from an alignment of dehydrogenase protein sequences from SYK-6, SG61-1L, one closely related bacterium that was unable to transform GGE (Pseudomonas putida KT2440, data not shown), and two other bacteria (Novosphingobium aromaticivorans DSM12444 and Sphingobium chlorophenolicum L-1) that may have potential to degrade GGE based on NCBI-nr blastp hits to GGE dehydrogenases from SYK-6. Deduced dehydrogenase sequences present in these five genomes that shared 20 % or higher amino acid identity with any of the four previously characterized SYK-6 enzymes (LigD, LigO, LigL, LigN) or the seven most closely related enzymes from SG61-1L (named 724, 2550, 2549, 2705, 2706, 1498, 3329) were included in the alignment. Also included in the alignment were several functionally
verified dehydrogenases from various organisms that met the same criteria for percent identity as described above.

The alignment was constructed with MAFFT using the L-INS-i algorithm [122] and consisted of a total of 166 sequences (Appendix 3). The program AliStat was used to remove poorly aligned positions within the alignment with a mask value of 98 (software available from http://www.csiro.au/Outcomes/ICT-and-Services/Software/AliStat.aspx). Symtest using Bowker’s matched pairs test for symmetry was used to confirm that the amino acid sequences evolved under stationary, reversible, and homogenous (SRH) conditions (software available from http://www.csiro.au/Outcomes/ICT-and-Services/Software/SymTest.aspx). A maximum-likelihood phylogenetic tree was then constructed with the IQtree server using an ultrafast bootstrap approximation approach with 10,000 replicates [166, 173].

2.3 Results

2.3.1 Isolation of a novel GGE-degrading bacterium

GGE-degrading bacteria from a paper mill waste site were identified through a series of enrichments as described in the Materials and Methods section. After 12 generations of enrichment, cultures were plated and individual colonies were isolated and tested for their ability to degrade GGE. Out of the fifty-five bacterial colonies tested, only two were able to use GGE as a sole carbon source. One of these strains degraded GGE at a very slow rate; after 800 h, 7 % of GGE remained in growth cultures (data not shown). The other strain, named SG61-1L, was pursued further due to its ability to completely degrade 200 µM GGE in 75 h (Figure 2-3C).
The sequences of two identical full length copies of 16S ribosomal genes (1481 bp) identified from the assembled genome (locus tags SZ64_16285 and SZ64_16315) of SG61-1L revealed that this alpha-proteobacterium belongs to the family Erythrobacteraceae, in the same order as the lignin-degrading strain Sphingobium sp. SYK-6 (Sphingomonadales). The full length 16S ribosomal genes from SG61-1L and SYK-6 share 94% identity. The 16S genes of SG61-1L are also 97-98% identical to several type strains, including Altererythrobacter troitsensis strain KMM 6042(T), Croceicoccus marinus strain E4A9T, Altererythrobacter dongtanensis strain JM27(T), Altererythrobacter epoxidivorans strain JCS350(T), and Altererythrobacter xinjiangensis strain S3-63(T).

The Rapid Annotation using Subsystem Technology (RAST) (NMPDR Bioinformatics Resource Center) [12, 27, 181] annotation of SG61-1L genome has revealed its genomic features and various metabolic capabilities of this bacterium such as aromatic compounds, nitrogen and protein metabolism, which are further presented in appendix 3.
Figure 2-3. Bacterial growth and degradation of GGE. (A) The pathway for transformation of GGE to HPV in SYK-6 and SG61-1L. (B) Metabolite formation and disappearance over the course of the GGE growth cell experiment in SG61-1L (red) and SYK-6 (blue). Each graph depicts a different metabolite (labeled in the upper right corner). (C) GGE degradation (left y-axis) and bacterial growth (right y-axis) over time for SYK-6 (left) and SG61-1L (right) during...
growth cell experiments with GGE as the sole carbon source as compared with no carbon source control experiments. The red arrows indicate the time at which cultures of SG61-1L were pelleted and resuspended in fresh MM supplemented with GGE, as described in the section 2.2.6. Abbreviations: GGE, guaiacylglycerol-β-guaiacyl ether; MPHPV, α-(2-methoxyphenoxy)-β-hydroxypropiovanillone; GS-HPV, α-glutathionyl-β-hydroxypropiovanillone; HPV, β-hydroxypropiovanillone.

2.3.2 Dye decolorization screening

Dye decolorization is widely used as a preliminary screen for the identification of microbes and/or enzymes that can degrade lignin via radical ion mechanisms [6, 15, 28, 35, 165]. The fifty-five bacteria isolated in this work, along with SYK-6, were tested for their ability to decolorize two dyes, Azure B and Reactive Black 5. Fourteen of these bacteria could decolorize Azure B while two others could decolorize Reactive Black 5 (data not shown). Neither of the two GGE degrading bacteria (SG61-1L and SYK-6) could decolorize either of the two dyes. These results indicate that dye screening may not necessarily be a reliable method for identifying bacteria that are able to degrade certain lignin substructures, especially those employing degradation routes that do not generate radical ions.

2.3.3 Comparison of GGE degradation rates in SG61-1L and SYK-6

Growth and GGE degradation rates for SG61-1L and SYK-6 were measured by growing the cells in MM supplemented with the GGE stereoisomer mixture as the sole carbon source. Both SG61-1L and SYK-6 showed an increase in OD$_{600}$ and a progressive loss of GGE over time (Figure 2-3 C), but SG61-1L grew on GGE much faster than SYK-6 (Figure 2-3). At 74 h, when GGE was completely degraded by SG61-1L, the cells (OD$_{600}$ 0.05) were removed by centrifugation and resuspended in a fresh lot of MM containing GGE. The second lot of GGE was completely degraded in the next 60 h while the OD$_{600}$
of the culture rose to 0.09. The negative controls with no added carbon showed no detectable growth of either bacterium (Figure 2-3 C). Bacterial growth under conditions where vanillin (a catabolic intermediate of GGE in both strains) was used as a sole carbon source were comparable (per mole of carbon) with those observed for growth on GGE as a sole carbon source (Figure 2-4), indicating that bacterial growth rates with GGE were real and not an artifact of growth from cellular reserves. With the exception of HPV (which did not appear in SYK-6 growth cultures over the measured time period), the same metabolites were identified in both bacteria and correlated well with mass spectral data, these also being consistent with previously published spectra for the metabolites produced by SYK-6 [161] (Table 2-2).

**Figure 2-4.** Growth of SG61 (panel A) and SYK-6 (panel B) on monoaryl compounds over time. For each graph, the left y-axis shows the concentration of metabolite and the right y-axis shows the OD_{600} values for either bacteria. Both bacteria were monitored for growth using vanillin as a sole carbon source. The experiments were monitored for disappearance of vanillin and formation/disappearance of the transient metabolite, vanillic acid.

**Table 2-2.** Retention times and m/z values for metabolites observed during GGE degradation
Consistent with their respective growth and GGE degradation rates, the metabolites also appeared earlier in the SG61-1L culture supernatant compared to the SYK-6 culture supernatant. Complete transformation of GGE to its oxidized product, MPHPV, occurred within 75 h in SG61-1L but required over 160 h in SYK-6 (Figure 2-3 C). MPHPV did not build up in the SG61-1L growth culture supernatant and was continuously transformed to downstream metabolites, consistent with the appearance of guaiacol (a product of the MPHPV etherase reaction) after only a few hours of growth (Figure 2-3). By contrast, MPHPV accumulated in the SYK-6 growth culture supernatant to levels well above that observed for SG61-1L and detectable levels of guaiacol did not appear until approximately 135 h (Figure 2-3 B).

A second set of growth experiments was then carried out for both bacteria on the same GGE mixture but this time it was analyzed on the chiral column to determine the rates at which individual stereoisomers were degraded in the presence of a mixture

<table>
<thead>
<tr>
<th>Compound</th>
<th>Molecular weight</th>
<th>Authentic standard</th>
<th>SYK-6</th>
<th>SG61-1L</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rt (min)</td>
<td>m/z [molecular ion]</td>
<td>Rt (min)</td>
<td>m/z [molecular ion]</td>
</tr>
<tr>
<td>GGE</td>
<td>320.34</td>
<td>303.12 [M-H\textsubscript{2}O+H\textsuperscript{+}]</td>
<td>5.12</td>
<td>303.12 [M-H\textsubscript{2}O+H\textsuperscript{+}]</td>
</tr>
<tr>
<td>MPHPV</td>
<td>318.32</td>
<td>-</td>
<td>6.78</td>
<td>319.10 [M+H\textsuperscript{+}]</td>
</tr>
<tr>
<td>HPV</td>
<td>196.20</td>
<td>-</td>
<td>3.78</td>
<td>197.06 [M+H\textsuperscript{+}]</td>
</tr>
<tr>
<td>Vanillin</td>
<td>152.15</td>
<td>153.05 [M+H\textsuperscript{+}]</td>
<td>5.53</td>
<td>153.05 [M+H\textsuperscript{+}]</td>
</tr>
<tr>
<td>Vanillic acid</td>
<td>168.15</td>
<td>169.05 [M+H\textsuperscript{+}]</td>
<td>4.16</td>
<td>169.08 [M+H\textsuperscript{+}]</td>
</tr>
<tr>
<td>Guaiacol</td>
<td>124.14</td>
<td>n/a</td>
<td>7.50</td>
<td>n/a</td>
</tr>
</tbody>
</table>
of GGE stereoisomers. Consistent with the first set of experiments discussed above, these experiments also demonstrated that the GGE substrate disappears more slowly in SYK-6 compared to SG61-1L (Figure 2-5) (the absolute rates of GGE disappearance analyzed via the chiral column were found to be slightly different from the rates obtained in the first set of growth experiments, for reasons that could not be determined). Both bacteria degraded the pair of erythro enantiomers ((αR,βS)-GGE and (αS,βR)-GGE) at an overall faster rate than the threo enantiomers ((αR,βR)-GGE and (αS,βS)-GGE). However, the rates for the two members of each enantiomer pair differed in SG61-1L but could not be distinguished in SYK-6. In SG61-1L, the erythro isomer (αR,βS)-GGE disappeared slightly earlier than its enantiomer, (αS,βR)-GGE, and the threo isomer (αS,βS)-GGE also disappeared slightly earlier than its enantiomer, (αR,βR)-GGE.
Resting cell experiments were performed in order to more rigorously decipher differences in degradation rates independent of cell growth for the two bacteria. The same metabolites appeared in these studies, which further validates that the same functional catabolic pathway is operating in both strains (data not shown). Due to limited cell growth, metabolites derived from HPV degradation (not observed during growth experiments) were transiently observed here and included vanillic acid and
vanillin (Figure 2-3). For SG61-1L, vanillic acid and vanillin were observed as early as 10 min, while for SYK-6, these metabolites were only observed after 200 h (data not shown). Furthermore, GGE quickly disappeared in SG61-1L (0.5 h) as compared to SYK-6 (30 h) (data not shown). Collectively, the resting cell and growth results demonstrate that SG61-1L transforms and degrades GGE and subsequent metabolites at a significantly faster rate than SYK-6.

### 2.3.4 Identification and characterization of GGE dehydrogenase candidates

Thirteen candidate GGE dehydrogenase genes identified from the assembled SG61-1L genome were selected for expression and further characterization. These 13 genes were chosen from among over 100 putative dehydrogenases identified in the genome on the basis of their relatively high percent amino acid identity to LigD, LigO, LigL, or LigN in SYK-6, and/or proximity to other genes in the genome that may be involved in lignin degradation (Table 2-3). The 13 sequences selected showed 20-67 % amino acid identity with at least one of the SYK-6 dehydrogenases and demonstrated varying levels of genomic context. For example, SG61-1L 724 was not surrounded by genes with annotated roles in lignin degradation but shared high identity with LigO (39 %), while SG61-1L 3344 was adjacent to two glutathione transferases and other genes potentially involved in lignin degradation but showed only 17-22 % percent identity to the four SYK-6 GGE dehydrogenases.
<table>
<thead>
<tr>
<th>Gene name</th>
<th>locus_tag</th>
<th>%ID to SYK-6 GGE dehydrogenases$^1$</th>
<th>Genomic context</th>
</tr>
</thead>
<tbody>
<tr>
<td>474</td>
<td>SZ64_16365</td>
<td>15-20 (D)</td>
<td>Immediately downstream of two 2,4'-dihydroxyacetophenone dioxygenase genes, and upstream of one benzyl alcohol dehydrogenase gene</td>
</tr>
<tr>
<td>3726</td>
<td>SZ64_14315</td>
<td>16-20 (O)</td>
<td>3.2 kb upstream of 3730</td>
</tr>
<tr>
<td>3191</td>
<td>SZ64_15940</td>
<td>18-21 (N)</td>
<td>3.2 kb upstream of vanillate demethylase oxygenase (on opposite strand)</td>
</tr>
<tr>
<td>3730</td>
<td>SZ64_14335</td>
<td>17-22 (L)</td>
<td>3.2 kb downstream of 3726</td>
</tr>
<tr>
<td>3344</td>
<td>SZ64_12435</td>
<td>17-22 (O)</td>
<td>Downstream of two GSTs (6 kb) and upstream of an alcohol dehydrogenase (1 kb), and ligM (5 kb)$^2$</td>
</tr>
<tr>
<td>3175</td>
<td>SZ64_16025</td>
<td>19-22 (L)</td>
<td>2.6 kb downstream of vanilate demethylase oxygenase</td>
</tr>
<tr>
<td>3329</td>
<td>SZ64_12360</td>
<td>28-31 (O)</td>
<td>3.5 kb downstream of genes in the Protocatechuate 4,5-Cleavage Pathway</td>
</tr>
<tr>
<td>2549</td>
<td>SZ64_00035</td>
<td>27-34 (O)</td>
<td>Immediately downstream of 2550</td>
</tr>
<tr>
<td>2706</td>
<td>SZ64_01650</td>
<td>30-37 (L)</td>
<td>Immediately downstream of 2705</td>
</tr>
<tr>
<td>724</td>
<td>SZ64_15290</td>
<td>30-39 (O)</td>
<td>none</td>
</tr>
<tr>
<td>2705</td>
<td>SZ64_01645</td>
<td>34-42 (N)</td>
<td>Immediately upstream of 2706</td>
</tr>
<tr>
<td>2550</td>
<td>SZ64_00030</td>
<td>30-47 (L)</td>
<td>Immediately upstream of 2549</td>
</tr>
<tr>
<td>1498</td>
<td>SZ64_05225</td>
<td>30-67 (N)</td>
<td>1.1 kb downstream of an aldehyde dehydrogenase</td>
</tr>
</tbody>
</table>

1. Percent amino acid identity to all four SYK-6 dehydrogenases (LigD, LigO, LigL, LigN) are written as a range. The SYK-6 dehydrogenase with the highest percent amino acid identity to the candidate dehydrogenase is in parentheses.

2. LigM, tetrahydrofolate-dependent O-demethylase, is an enzyme responsible for degradation of vanillate and syringate in SYK-6 [1].
All 13 dehydrogenase genes from SG61-1L along with the four originally characterized GGE dehydrogenases from SYK-6 were cloned, overexpressed in *E. coli*, and where possible, the corresponding enzymes purified. Three of the SG61-1L enzymes were insoluble (2549, 3191, and 2706) and therefore unable to be assayed or purified, while three others from this bacterium (3730, 3175, 3726) were soluble but did not demonstrate any detectable dehydrogenase activity with the GGE substrate. An SDS-PAGE gel of the ten soluble purified SG61-1L enzymes is shown in Figure 2-6. The corresponding gel for the four SYK-6 enzymes showed similar levels of purity (data not shown). The kinetic constants for the seven soluble and active SG61-1L enzymes and the four from SYK-6 are reported in Table 2-4. The enzymatic reaction was verified by LC-MS to confirm *in vitro* transformation of GGE to MPHPV (Figure 2-7).

![Figure 2-6. SDS-PAGE gel of purified GGE dehydrogenase enzymes from SG61-1L: lane 1. Precision Plus Protein™ Prestained Standards; lane 2. SG61-1L 474; lane 3. SG61-1L 724; lane 4. SG61-1L 1498; lane 5. SG61-1L 2705; lane 6. SG61-1L 2550; lane 7. SG61-1L 3175; lane 8. SG61-1L 3329; lane 9. SG61-1L 3344; lane 10. SG61-1L 3726; lane 11. SG61-1L 3730.](image)
Table 2-4. Kinetic constants for short-chain dehydrogenases from SG61-1L and SYK-6 on all GGE stereoisomers (ND – no detectable level of activity)

<table>
<thead>
<tr>
<th>Name</th>
<th>$\alpha_S,\beta_R$ (µM)</th>
<th>$\alpha_R,\beta_S$ (µM)</th>
<th>$\alpha_R,\beta_R$ (µM)</th>
<th>$\alpha_S,\beta_S$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LigD</td>
<td>0.39 ± 0.08</td>
<td>15 ± 13</td>
<td>25 ± 10</td>
<td>12.2 ± 2.3</td>
</tr>
<tr>
<td>LigL</td>
<td>1.6 ± 0.3</td>
<td>39 ± 5</td>
<td>3.0 ± 0.5</td>
<td>61 ± 16</td>
</tr>
<tr>
<td>LigN</td>
<td>51 ± 14</td>
<td>15.3 ± 2.8</td>
<td>1.7 ± 0.3</td>
<td>35.0 ± 7.2</td>
</tr>
<tr>
<td>LigO</td>
<td>118 ± 29</td>
<td>1.2 ± 0.4</td>
<td>51 ± 12</td>
<td>40.3 ± 9.6</td>
</tr>
</tbody>
</table>

| 2550 | 3.6 ± 0.6               | 53 ± 9                  | 2.2 ± 0.8               | 0.8 ± 0.2               |
| 1498 | 88 ± 15                 | ND                      | ND                      | 57 ± 12                 |
| 2705 | 357 ± 47                | 235 ± 41                | ND                      | 354 ± 44                |
| 3329 | ND                      | ND                      | 49 ± 16                 | 35 ± 10                 |
| 3344 | 66 ± 12                 | 202 ± 16                | 19.0 ± 3.0              | 162 ± 24                |
| 474  | 109 ± 18                | ND                      | ND                      | 163 ± 22                |
| 724  | 77 ± 18                 | 2.0 ± 0.4               | 4.2 ± 0.8               | 33.7 ± 8.8              |

<table>
<thead>
<tr>
<th>Name</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$k_{cat}/K_{M,GGE}$ (µM$^{-1}$s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LigD</td>
<td>4.9 ± 0.2</td>
<td>46.1 ± 2.3</td>
</tr>
<tr>
<td>LigL</td>
<td>83.4 ± 0.1</td>
<td>2.3 ± 0.1</td>
</tr>
<tr>
<td>LigN</td>
<td>50.6 ± 5.1</td>
<td>0.6 ± 0.1</td>
</tr>
<tr>
<td>LigO</td>
<td>7.0 ± 0.7</td>
<td>0.3 ± 0.0</td>
</tr>
</tbody>
</table>

64
<table>
<thead>
<tr>
<th></th>
<th>αS,βR</th>
<th>αR,βS</th>
<th>αR,βR</th>
<th>αS,βS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LigD</td>
<td>12.5 ± 2.7</td>
<td>1.14 ± 1.06</td>
<td>0.92 ± 0.4</td>
<td>0.68 ± 0.13</td>
</tr>
<tr>
<td>LigL</td>
<td>53.3 ± 9.0</td>
<td>1.9 ± 0.3</td>
<td>11.3 ± 2.1</td>
<td>6.3 ± 1.7</td>
</tr>
<tr>
<td>LigN</td>
<td>1.0 ± 0.3</td>
<td>4.2 ± 0.8</td>
<td>15.1 ± 2.9</td>
<td>2.0 ± 0.4</td>
</tr>
<tr>
<td>LigO</td>
<td>0.06 ± 0.01</td>
<td>2.4 ± 0.8</td>
<td>0.06 ± 0.01</td>
<td>0.39 ± 0.10</td>
</tr>
</tbody>
</table>

|     |     |     |     |     |
|-----|-----|-----|-----|
| 2550| 12.9 ± 2.4 | 2.2 ± 0.4 | 7.4 ± 2.7 | 95.1 ± 20 |
| 1498| 0.04 ± 0.01| ND     | ND     | 0.05 ± 0.01|
| 2705| 0.14 ± 0.02| 0.012 ± 0.002| ND    | 0.18 ± 0.03|
| 3329| ND     | ND     | 0.012 ± 0.004| 0.003 ± 0.001|
| 3344| 0.03 ± 0.01| 0.011 ± 0.001| 0.015 ± 0.003| 1.2 ± 0.2 |
| 474 | 0.013 ± 0.002| ND     | ND     | 0.08 ± 0.01 |
| 724 | 12.4 ± 3.1 | 360 ± 74 | 368 ± 69 | 21.4 ± 6.0 |

65
Figure 2-7. Enzymatic conversion of GGE to MPHPV. (A) LC-MS chromatogram of an enzymatic reaction using SG-61-1L 724 enzyme and the GGE substrate showing conversion of GGE to MPHPV (B) The mass spectrum for the peak at a retention time of 5.12 min (GGE). (C) The mass spectrum for the peak at a retention time of 6.82 min (MPHPV).
All four SYK-6 GGE dehydrogenases were able to degrade all four GGE stereoisomers. In general, LigN and LigL displayed higher turnover values than LigD and LigO for all four stereoisomers (Table 2-4). The $k_{cat}$ values for LigD and LigN correlate with previous data [215] to the extent that they demonstrate fastest turnover of $(\alpha R,\beta R)$-/(\alpha R,\beta S)-GGE and $(\alpha S,\beta S)$-GGE, respectively. However other aspects of the experimental results do not correlate well with previous data from SYK-6. For example, according to the $k_{cat}$ values herein, LigO displays the fastest turnover rate for $(\alpha S,\beta S)$-GGE and not $(\alpha R,\beta R)$- or $(\alpha R,\beta S)$-GGE as reported previously [218]. Additionally, the majority of the $K_M$ and $k_{cat}/K_M$ values for each enzyme do not correlate with the previously reported stereoisomer preferences, assuming that the preferred stereoisomer would generate a lower $K_M$ value and a higher $k_{cat}/K_M$ value compared to other stereoisomers for a given enzyme. These discrepancies can be explained by the fact that very different methods of analysis were employed in the two sets of experiments. While the previous report determined enzyme stereoselectivity through genetic knockouts (for LigD) or enzyme performance in crude cell extracts in the presence of two stereoisomers (for LigD, LigN, and LigL) [215], all of the experiments herein were performed in vitro on purified enzymes in the presence of a single stereoisomer. Thus, differences in enzyme concentrations, for example, would have influenced the earlier results but not those reported here.

Two of the SG61-1L enzymes, 724 and 2550, turned over all four stereoisomers very efficiently, with $k_{cat}$ values up to 550-fold higher than those for the SYK-6 enzymes (Table 2-4). SG61-1L 724 had the highest $k_{cat}$ values for all stereoisomers and the highest specificity constant estimates for two stereoisomers (360 $\mu$M$^{-1}$ s$^{-1}$ and
368 μM⁻¹ s⁻¹ for (αR,βS)-GGE and (αR,βR)-GGE, respectively). SG61-1L 2550 had significantly lower $K_M$ values (3.6 and 0.8 μM, respectively) for the other two stereoisomers, (αS,βR)-GGE and (αS,βS)-GGE, yielding specificity constant estimates of 12.9 μM⁻¹ s⁻¹ and 95.1 μM⁻¹ s⁻¹, respectively. SG61-1L 724 and 2550 both showed relatively high sequence identity to the SYK-6 GGE dehydrogenases (30-39 % and 30-47 %, respectively, Table 2-3). Three other SG61-1L enzymes, 3329, 2705, and 1498, also shared high sequence identity with SYK enzymes (28 %-31 %, 34-42 %, and 30-67 %, respectively) but performed relatively poorly in kinetic assays. Interestingly, the two remaining enzymes, 3344 and 474, performed similarly or better than these latter three enzymes despite their low levels of sequence identity to the SYK-6 enzymes (17-22 % and 15-19 %, respectively). Aside from SG61-1L 3344 and 3329, which used the cofactor NADP⁺, all other SG61-1L enzymes functioned with the cofactor NAD⁺. The four SYK-6 enzymes also preferred the cofactor NAD⁺, as in the previous experiments [218].

There was no obvious relationship between the kinetic parameters of any individual SG61-1L enzyme for the different stereoisomers and the relative degradation rates of those isomers in the growth experiments. This is perhaps unsurprising given that: (1) several of the enzymes have relatively good kinetics for one or more of the isomers; (2) we do not know the relative importance of $k_{cat}$ and $K_M$ to degradation rates under the conditions of the growth experiment; (3) we do not know the relative abundances of the enzymes in vivo; and (4) there may be other dehydrogenases in the organism that are presently uncharacterized but also contribute to GGE oxidation.
2.3.5 Phylogenetic relationships between dehydrogenases from various organisms

A phylogenetic tree was constructed in an attempt to articulate relationships between sequence, function and stereoselectivity among the dehydrogenases characterized in this study. For contextual purposes, 136 additional sequences showing 20% amino acid identity or higher to at least one of the characterized SYK-6 and SG61-1L dehydrogenases were included in the phylogeny. These sequences were from three bacteria: one that was tested (along with SYK-6 and SG61-1L) and unable to degrade GGE (*Pseudomonas putida* KT2440; data not shown) and two that were not known to degrade GGE (*Novosphingobium aromaticivorans* DSM12444 and *Sphingobium Chlorophenolicum* L-1). Notably, *N. aromaticivorans* DSM12444 may have the ability to degrade GGE given that it encodes active MPHPV β-etherase GSTs [82] as well as several putative Cα-dehydrogenases with relatively high sequence identity (37-76%) to one or more of the SYK-6 enzymes. Nineteen sequences from other bacteria that encode enzymes with experimentally verified activities were also included in the tree; all of these sequences shared 20% amino acid identity or higher to at least one of the characterized SYK-6 and SG61-1L dehydrogenases as described in section 2.2.12.

As seen in Figure 2-8, the four SYK-6 GGE dehydrogenases and seven of the thirteen SG61-1L dehydrogenases chosen in this study cluster together in the same clade. Moreover, this clade does not include any amino acid sequences from the non-GGE degrading bacterium, *P. putida*, but does include five amino acid sequences from *N. aromaticivorans* and one from *S. chlorophenolicum*. With the exception of SG61-1L 2706 and 2549, which were completely insoluble, all remaining SG61-1L dehydrogenases in this clade displayed activity on the GGE stereoisomers, and this clade
was therefore named “the GGE-dehydrogenase clade”. SG61-1L 3344 and 474 lay outside of the GGE-dehydrogenase clade despite the fact that they displayed measurable activities on two or more GGE stereoisomers.

**Figure 2-8.** Phylogenetic tree of GGE dehydrogenase proteins and their closest relatives from five different organisms including SG61-1L (SG61), SKY-6 (SLG), *Pseudomonas putida* KT2440 (PP), *Novosphingobium aromaticivorans* DSM 12444 (Naro), and *Sphingobium chlorophenolicum* L-1 (Sphch). Also included in the alignment were deduced amino acid sequences of genes from a variety of bacteria whose annotations have been experimentally verified (names are highlighted in various colours, abbreviations are listed below, and references are listed in Appendix 4). The clades most closely related to each verified amino acid sequence are colored in the same shade. The purple clade represents the “GGE dehydrogenase clade” identified in this work. The names of the SG61-1L and SYK-6 enzymes characterized herein are shaded and outlined if they exhibited GGE dehydrogenase activity and are otherwise outlined if they did not (insoluble enzymes are represented with the symbol “?” while enzymes that displayed
no activity are represented with the symbol “x”). Accession numbers are defined in Appendix 4. of the supplemental material for all gene names (corresponding to the representative amino acid sequences) in the tree that are not represented by gene identification (GI) number (Naro and Sphch) or locus ID (SLG and PP). Abbreviations: ydfG_Ecoli, 3-hydroxy acid dehydrogenase from E. coli (light pink); ADH_Syanoikuyae, ‘bulky-bulky’ ketone dehydrogenase from Sphingobium yanoikuyae DSM6900 (magenta); kduD_Ecoli, 2-dehydro-3-deoxy-D-glucuronate dehydrogenase from E. coli (fuschia); (S)-PED_Aaro, (S)-1-phenylethanol dehydrogenase from Aromatoleum aromaticum EbN1 (plum); phaB_Burkholderia Acetoacetyl CoA reductase from Burkholderia sp. Strain RPE75 (brown); phaB_Synechocystis, Acetoacetyl CoA reductase from Synechocystis sp. Strain PCC6803 (brown); fabG_Ecoli, β-ketoacyl-[acyl carrier protein] reductase from E. coli (maroon); badH_Rpalustris, 2-hydroxycyclohexanecarboxyl-CoA dehydrogenase from Rhodopseudomonas palustris CGA009 (red); Ga5DH_Ssuis, glucuronate 5-dehydrogenase from Steptococcus suis (red); ADH_Ralstonia, ‘bulky-bulky alcohol dehydrogenase from Ralstonia sp. DSM 6428 (orange); 2,5-DDOL_Spaucimobilis, 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase from Sphingomonas paucimobilis UT26 (yellow); (R)-ADH_Lkefiri, (R)-specific alcohol dehydrogenase from Lactobacillus kefiri (yellow); CPNA_Comamonas, cyclopentanol dehydrogenase from Comamonas sp. Strain NCIMB 9872 (light green); steA_Ctestosteroni, 7α,12α-dihydroxyandrosta-1,4-diene-3,17-dione dehydrogenase from Comamonas testosteroni TA441 (dark green); Gluc-DH_Bsubtilis, glucose 1-dehydrogenase from Bacillus subtilis (turbquoise); LVR_Caquaticum, levodione reductase from Corynebacterium aquaticum M-13 (light blue); bdh_Pputida, 3-hydroxybutyrate dehydrogenase from Pseudomonas putida Strain ZIMET 10947 (dark blue); bdh1_ and bdh2_Rpickettii, 3-hydroxybutyrate dehydrogenase from Ralstonia pickettii T1 (dark blue).

2.4 Discussion

Herein I report the isolation of SG61-1L, a bacterium that is able to degrade the GGE model lignin dimer at a much faster rate than the previously characterized GGE-degrading bacterium, SYK-6. In order to understand the molecular basis for its activity and investigate the biotechnological potential of the enzymes involved, a suite of its dehydrogenases that are prime candidates for the first step in β-aryl ether degradation, the NAD(P)⁺ dependent oxidation of GGE to MPHPV have been characterized.
The analyses suggest that the four SYK-6 enzymes are kinetically inferior to two of the SG61-1L enzymes, 724 and 2550, as catalysts for all four stereoisomers. This result may in part explain the faster rate of GGE transformation observed for SG61-1L.

Most of the SG61-1L and SYK-6 enzymes with GGE dehydrogenase activity characterized herein fall into a single clade, and it is likely that this clade is in general a good predictor of GGE dehydrogenase activity. It is also apparent that the sequences of GGE dehydrogenases sitting in this clade have diverged significantly from other functionally annotated dehydrogenases. Three additional uncharacterized SYK-6 genes are also present in the GGE dehydrogenase clade, but it seems unlikely that these genes play a larger role in GGE oxidation in their respective organisms than those already characterized. This notion is based on a previous experiment with a mutant SYK-6 bacterium containing knockouts of three GGE dehydrogenase genes (LigD, LigN, and LigL) that results in nearly complete loss of the ability to oxidize GGE compared to wild type [218].

Under the selection pressures imposed on bacteria in an environment that is rich in lignin-derived substructures, it is possible that dehydrogenases outside of the GGE dehydrogenase clade and native to other metabolic pathways have evolved some ability to contribute to GGE oxidation. For example, 3344 and 3730 localize to a larger clade of the phylogenetic tree that contains a functionally verified levodione reductase from Corynebacterium aquaticum M-13; 3344 and 3730 show 38 % and 43 % amino acid identity to this levodione reductase, respectively, indicating that they may have evolved from a levodione reductase. This idea that biological degradation occurs through multiple unrelated dehydrogenases native to different pathways is reminiscent of a recent report on the oxidation of dehydrodiconiferyl aldehyde in SYK-6, which is
proposed to occur through the action of several very distinct aldehyde dehydrogenases [246]. SYK-6 and SG61-1L were both isolated from pulp and paper mill waste sites [121], but unlike SYK-6, SG61-1L was specifically selected through rounds of enrichment culturing for its ability to utilize GGE as the sole carbon source. This difference is perhaps reflected in the poorer kinetic parameters for the SYK-6 GGE dehydrogenases as compared to the two highest performing SG61-1L GGE dehydrogenases.

It is also relevant to note here an apparently more divergent Cα-dehydrogenase has been previously purified from *Pseudomonas sp.* GU5 [185]. These enzymes have $K_M$ values for GGE (represented by a mixture of stereoisomers) as low as 11-12 µM ($k_{cat}$ values were not reported) [92, 188], these values being lower than most obtained herein. The genes encoding these proteins were not identified but the monomeric molar mass of the Cα-dehydrogenase from *Pseudomonas sp.* GU5 was estimated to be approximately 52,000 kDa [188], which is quite different from the average monomeric molar mass range for all characterized proteins in this work (approximately 30,000 kDa), suggesting that another yet unidentified type of Cα-dehydrogenase can oxidize GGE in this organism.

There is no obvious correlation between phylogeny and stereoselectivity within the GGE dehydrogenase clade. For example, 1498 from SG61-1L and LigN from SYK-6 share the highest sequence identity (67 %) of any pair in the GGE dehydrogenase clade, yet 1498 was only able to oxidize the (αS,βR)- and (αS,βS)-GGE stereoisomers while LigN could oxidize all four. Similarly, 2550 from SG61-1L and LigL from SYK-6 also cluster together in this clade and were both able to oxidize all four GGE stereoisomers, but each enzyme displayed a different pattern of stereospecificity. Empirical structural data will

73
likely be necessary to identify the sequence motifs guiding stereoselectivity preferences among these GGE dehydrogenases.

To the best of my knowledge, SG61-1L is one of the fastest known bacterial degraders of a lignin substructure; previously reported rates of bacterial degradation for model lignin dimers are quite slow [108, 133], albeit various conditions were used and the results may therefore not be directly comparable. I find here that SG61-1L can degrade GGE much faster than SYK-6 under the same experimental conditions and that in SYK-6, transformation of MPHPV to $\alpha$-glutathionyl-$\beta$-hydroxypropiovanillone (GS-HPV) appears to be a rate-limiting step.

Following GGE oxidation to MPHPV, the enzymes that perform the next two reactions (the $\beta$-etherase reaction to generate GS-HPV followed by glutathione removal to generate achiral HPV) are GSTs. They have been identified as LigF/LigE/LigP and LigG, respectively in SYK-6, and homologues have been functionally verified in closely related Sphingomonads [81, 82, 158, 248] including *Novosphingobium aromaticivorans* DSM12444, which also encodes genes present in the GGE dehydrogenase clade. There are several uncharacterized GSTs in the SG61-1L genome, two of which have the highest sequence identity to LigP (63-79 %) and five of which have highest sequence identity to LigF (34 %-59 %). Curiously, none of the SG61-1L encoded GSTs have high sequence identity to LigG (below 17 %). Thus there are several candidates in SG61-1L for the first GST step, but the second apparently involves an enzyme not closely related to LigG. Given that the LigG reaction in SYK-6 is specific to only one of the two GS-HPV stereoisomers [81], it has also been suggested that there is at least one other, as yet uncharacterized, GST with activity for the other GS-HPV stereoisomer in SYK-6. Future work will involve characterization of these later steps in the pathway in order to
articulate mechanisms by which SG61-1L proceeds through GGE degradation at a relatively rapid rate.

There is an emerging interest in the development of lignin degrading enzymes for industrial purposes, especially for integration into processes generating replacements for petroleum-based commodities. Selective depolymerization of lignin or lignin components could generate high value, complex aromatics including catechols, resorcinols, keto acids, and polyhydroxy aromatics [101, 264]. Vanillin, for example, has been produced from lignocelluloses in a *Rhodococcus jostii* RHA-1 mutant [213]. Vanillin also has potential to be produced on a large scale via biodegradation of lignin substructures (such as β-aryl ether linked lignin dimers) by engineered versions of bacteria such as SG61-1L and SYK-6 or by free enzymes driving the relevant reactions derived from them [134, 159]. Thus far, a group of dehydrogenases from SG61-1L have been characterized that are involved in the first step of a degradation pathway for a major lignin subunit and could be used to produce a high-value chemical such as vanillin. The priorities on a high $k_{cat}$ vs low $K_M$ and broad vs narrow stereoselectivity for a given enzyme will vary depending on the process, but the suite of dehydrogenases characterized here from SG61-1L provides a wide range of options across these criteria. Any further protein engineering required to make promising enzymes more fit for purpose can also now be guided by the findings reported here on the phylogenetic distribution of GGE dehydrogenase activity. SG61-1L is also a very promising source of catalysts for downstream steps in the GGE degradation pathway, particularly the β-etherase step which results in cleavage of the highly stable aryl-aryl ether bond of MPH PV to produce GS-HPV.
Chapter 3 Aerobic biodegradation of 2,4-dinitroanisole (DNAN) by a novel hydrolase from *Nocardioides* sp. JS1661

3.1 Introduction

2,4-Dinitroanisole (DNAN) is one of the insensitive nitroaromatic ingredients increasingly used as a replacement for 2,4,6-trinitrotoluene (TNT) in munitions. DNAN or its degradation products can be toxic to earthworms, bacteria, algae and plants [58, 146]. Therefore, the release of DNAN to the environment could pose ecological and environmental risks. There is little information about the environmental behaviour of DNAN [58] and no bacteria capable of complete biodegradation have previously been reported. The initial reaction in the biotransformation of DNAN by aerobic bacteria, and in abiotic transformation with zero valent iron, is the reduction of the nitro group in the ortho position to yield 2-amino-4-nitroanisole (2-ANAN) [94, 178, 191]. The best characterized case of aerobic bacterial biotransformation involves a *Bacillus* strain which can transform DNAN slowly to 2-ANAN as a dead-end product [191]. Under anoxic conditions DNAN is biotransformed by soil microflora to toxic metabolites such as diaminoanisole (DAAN) [7, 94, 174, 178, 194]. Biotransformation of DNAN to 2,4-dinitrophenol (2,4-DNP) has been reported in mammals [106] but not as yet in bacteria. During alkaline hydrolysis, DNAN is converted to 2,4-dinitrophenolate via an unstable hydride-Meisenheimer complex [214]. Photo-transformation of DNAN results in 2-hydroxy-4-nitroanisole and 2,4-DNP as major and minor products respectively [94, 207]. The pathway of 2,4-DNP biodegradation is well known and the genes involved have
been characterized from *Rhodococcus erythropolis* and *Nocardia simplex* [96, 269]. Another *Rhodococcus* sp. has been reported to degrade 4-nitroanisole by a pathway involving removal of the methyl group and subsequent degradation of the resulting 4-nitrophenol via 4-nitrocatechol and 1,2,4-trihydroxybenzene [219].

This chapter reports collaborative work with Dr Tekle Tafese Fida and Professor Jim Spain to isolate and characterize a bacterium capable of growing on DNAN under aerobic conditions and then to characterize an O-demethylase enzyme which catalyses the first step in the pathway. The initial isolation and characterization of the bacterium was carried out by Dr Fida and Professor Spain and the enzymology and gene discovery by myself.

### 3.2 Experimental

#### 3.2.1 Materials

DNAN was obtained from Alfa Aesar (Ward Hill, MA, USA) and 2,4-DNP, formaldehyde and methanol from Sigma-Aldrich (St, Louis, MO, USA). The hydride-Meisenheimer complex of 2,4-DNP was chemically synthesized as described by Behrend and Wagner [17]. All other chemicals were obtained by Sigma Aldrich unless otherwise stated.

Minimal medium (MM) was prepared as described in section 2.2.2. Trypticase soybroth (TSB) was prepared by adding 8 g tryptone and 1.25 g of glucose to 1 litre of MM.
3.2.2 **Isolation of DNAN-degrading bacteria**

An activated sludge sample from Holston Army Ammunition Plant (Tennessee, USA) was inoculated (20 % v/v) into 1/4 strength MM pH 6.5 containing 100 μM DNAN and the suspension was incubated at 30 °C with shaking. Following the disappearance of DNAN as monitored by high-performance liquid chromatography (HPLC) (see below), samples (20 %, v/v) were repeatedly transferred into fresh medium and then spread on MM agar (1.5 %) plates containing 100 μM DNAN. Individual colonies that appeared after 4 days of incubation were tested for the ability to degrade DNAN in carbon and nitrogen free MM.

Two isolates that used DNAN as their sole source of carbon, nitrogen and energy were selected for further study. 16S rRNA gene sequence analysis was used to identify these strains from draft genome sequences obtained using Illumina sequencing technologies (Oregon State University, OR). The genomes were assembled using the A5 pipeline [255] and annotated by RAST [12, 27, 181]. The size of assembled genome for the preferred strain was 5.742 Mb, distributed across 276 contigs, with 71.2 % G/C content (data not shown).

3.2.3 **Growth of DNAN-degrading isolates**

Starter cultures of the two strains isolated above were grown with shaking at 30 °C in MM containing 1.9 mM ammonium sulphate (AMS). When large biomass was required for enzyme purification work they were then grown in MM containing 100 μM DNAN and 10 mM sodium acetate.
To determine growth yields, the MM was supplemented with either 200 μM DNAN or 200 μM 2,4-DNP and inoculated to an OD_{600} of 0.004 with cultures pre-grown on 100 μM DNAN. Biomass was quantified by measuring protein 6 h after disappearance of 2,4-DNP. Controls with DNAN or 2,4-DNP were inoculated with autoclaved cells. One ml samples were centrifuged and the cell pellets suspended in 0.1 N NaOH, lysed by heating at 95 °C for 10 min and assayed for protein as described below.

Specific activities of cells grown on DNAN or 2,4-DNP were determined using DNAN-induced and non-induced cells. Induced cells were grown in MM supplemented with 100 μM DNAN or 2,4-DNP as the sole carbon and energy source. Non-induced cells were grown in MM supplemented with 10 mM sodium acetate. Cells were harvested during logarithmic growth, washed twice with MM and suspended to an OD_{600} of 0.004 (corresponding to a protein concentration of about 6 or 8.5 μg ml⁻¹ for acetate and DNAN or 2,4-DNP grown cells, respectively) in MM supplemented with 100 μM DNAN or 100 μM 2,4-DNP and incubated at 30 °C. Disappearance of DNAN, or 2,4-DNP was monitored at appropriate intervals by high performance liquid chromatography (HPLC) as described below.

### 3.2.4 Enzyme assays

Cells were grown for 48 h in MM containing 100 μM of DNAN or 2,4-DNP supplemented with 10 mM sodium acetate. An additional 100 μM of DNAN or 2,4-DNP was added after complete degradation of the initial DNAN or 2,4-DNP. Immediately following disappearance of the second addition of DNAN or 2,4-DNP, cells were harvested by centrifugation, washed twice with 20 mM phosphate buffer pH 7.0, and suspended in the same buffer. Cells were lysed by two passages through a French press.
pressure cell at 10,000 psi. The lysates were clarified by centrifugation at 100,000 \( x g \) for 30 min at 4 °C. For some of the experiments, the resulting cell lysates were either ultrafiltered using Microcon centrifugal filters (10 kDa) (Millipore, MA, USA) or dialyzed overnight using Slide-A Lyzer dialysis membranes (10 kDa) (Pierce, IL, USA) against 20 mM potassium phosphate buffer pH 7.0 at 4 °C.

DNAN hydrolase assays were performed in 3 ml reaction mixtures in 20 mM potassium phosphate buffer pH 7.0 containing 100 \( \mu \)M DNAN and cell extract (0.3 to 0.5 mg of protein). At appropriate intervals, samples were removed and the reactions were stopped by addition of 0.5 % trifluoroacetic acid (TFA). Precipitated proteins were removed by centrifugation and the supernatant was analyzed by HPLC as indicated below. Boiled lysates were used in controls. Concentrations of methanol were determined in 1 ml reaction mixtures by an indirect method. Briefly, after complete transformation of DNAN to 2,4-DNP by dialyzed cells in 2 ml sealed HPLC vials, methanol was converted to formaldehyde by incubation with 0.5 U alcohol oxidase for 15 min as described by Klavons and Bennett [128]. The reactions were stopped by addition of 0.5 % TFA to avoid further conversion of formaldehyde to formic acid by alcohol oxidase. Formaldehyde was then derivitised with 2,4-pentanedione for 15 min at 58 °C in the presence of ammonium acetate and glacial acetic acid as described by Summers [244]. The derivative was analyzed by HPLC as described below.

2,4-DNP depletion by hydride transferase activity was assayed in 3 ml reaction mixtures as described by Behrend and Wagner [17]. Briefly, the reaction mixtures contained 20 mM potassium phosphate pH 7.0, 100 \( \mu \)M 2,4-DNP, 200 \( \mu \)M NADPH, and cell extract (0.3 to 0.5 mg of protein). The reaction mixtures were incubated at 30 °C. At
appropriate time intervals, samples were removed and the reactions were stopped by addition of 0.5 % TFA. The acidified reaction mixtures were clarified by centrifugation and depletion of 2,4-DNP analyzed by HPLC.

3.2.5 Analytical methods

Concentrations of DNAN, 2,4-DNP and the hydride-Meisenheimer complex of 2,4-DNP were determined using an Agilent 1100 HPLC system with a Merck Chromolith C-18 reverse phase column (4.6 mm by 100 mm, 5 μm). The mobile phase for DNAN or 2,4-DNP consisted of 95 % water and 5 % acetonitrile with 0.1 % TFA, delivered at a flow rate of 1.5 ml min⁻¹ over a period of 8 min. The hydride-Meisenheimer complex of 2,4-DNP was analyzed as described by Blasco et al. [22] using the same HPLC column as above but with a mobile phase of 98 % 20 mM phosphate buffer pH 7.0 without TFA, and 2 % acetonitrile with 0.015 % TFA, delivered at a flow rate of 1.5 ml min⁻¹. The formaldehyde derivative (3,5-diacetyl-1,4-dihydrolutidine) was analyzed on a Zorbax ODS C-18 reverse phase column (4.6 mm by 100 mm, 5 μm) with a mobile phase of 60 % water and 40 % acetonitrile with 0.1 % TFA, at a flow rate of 0.5 ml min⁻¹. DNAN was monitored at 298 nm (retention time [RT] 5.7 min in phosphate buffer), 2,4-DNP at 260 nm (RT 4 min in water-acetonitrile and 3.1 min in phosphate buffer), the hydride-Meisenheimer complex at 400 nm (RT 1.5 min in phosphate buffer) and 3,5-diacetyl-1,4-dihydrolutidine at 410 nm (RT 2.1 min in phosphate buffer). Nitrite concentrations were determined colorimetrically as described previously [200]. Protein assays were done using a bicinchoninic acid (BCA) assay kit (Rockford, IL, USA). Samples were centrifuged for 5 min at 16,000 x g after addition of BCA reagents and then analyzed spectrophotometrically. All assays were performed in duplicate.
3.2.6 Nucleotide sequence accession numbers

The 16S rRNA gene sequences of Nocardioides sp. JS1661 and JS1660 were deposited in GenBank with accession numbers KM026539 and KM026540, respectively. Catabolic gene clusters involved in DNAN hydrolase and 2,4-DNP degradation were deposited in GenBank with accession numbers KM213001 and KM189438 respectively.

3.2.7 Purification of DNAN hydrolase

JS1661 cells were grown to a log phase (OD$_{600}$ 3.1) in 10 L of TSB. The cell pellet from a 1000 x $g$ centrifugation was washed twice and resuspended in 50 mM phosphate buffer pH 8.5. 25 µg ml$^{-1}$ of DNase (Invitrogen, Carlsbad, CA) was added and the suspension incubated on a magnetic stirrer on ice (0 °C) for 1 h to reduce the viscosity. The cells were then lysed using a bench top homogeniser (Micro Fluidics M-11OP) at 20,000 psi for 20 min (there was no activity in culture supernatants suggesting the enzyme was intracellular; data not shown). The lysate was centrifuged at ~48,000 x $g$ for 30 min at 0 °C and the supernatant collected for enzyme purification.

The supernatant was initially brought to 20 % saturation with finely grounded AMS, stirred for 30 min, and the precipitated proteins removed by centrifugation at ~18,600 x $g$ for 30 min. The 20 % saturated AMS supernatant was then brought up to 40 % saturation and stirred for 6 h, and the precipitated proteins removed by centrifugation at ~18,600 x $g$ for 30 min. The whole AMS precipitation step was carried out at 0 °C. The precipitated proteins were then dissolved in 50 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer pH 8.5, loaded onto a hydrophobic interaction column (HiTrap, bead volume 65 ml), (GE Healthcare, Buckinghamshire, UK), eluted with a 3-0 M gradient of AMS and 10 ml fractions were...
collected. The pooled active fractions were concentrated using a 100 kDa cut-off Amicon Ultra-15 Centrifugal Filter (Millipore) and 5 ml of concentrated protein in 50 mM HEPES buffer pH 8.5 was manually loaded onto a gel filtration column (Superdex 200 Prep Grade, bead volume 130.69 ml; GE Healthcare). The proteins were eluted with a 0-1 M NaCl gradient as 5 ml fractions. The active fractions were pooled and loaded onto a strong anion exchange column (15Q, bed volume 9 ml; GE Healthcare) and eluted with 0-1 M NaCl in 50 mM HEPES pH 8.5, in 3 ml fractions. The active fractions from this column were in turn loaded onto a strong hydrophobic interaction column (15Q, bed volume 9 ml; GE Healthcare) and eluted with a 3-0 M gradient of AMS in 3 ml fractions. The pooled active fractions from this final column had 0.78 mg ml\(^{-1}\) of protein as measured by Nano drop (Qiagen ND 1000 Spectrophotometer). All chromatography steps were carried out at 4 °C on an automated AKTA fast protein liquid chromatography apparatus (FPLC, GE Healthcare). DNAN hydrolase activity during purification was quantitatively measured by the formation of 2,4-DNP at 400 nm on a Spectromax-M3 spectrophotometer (Molecular Devices, Sunnyvale, CA).

### 3.2.8 DNAN hydrolase mass estimation and gene identification

The purified fractions above were analysed using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and native polyacrylamide gel electrophoresis (Native PAGE) using 4-20 % precast precise polyacrylamide gels (Thermo Scientific, Waltham, MA).

For the SDS-PAGE, the protein solutions were treated with SDS loading buffer at 95 °C for 15 min and subjected to electrophoresis (90 volts for 1 h) in Tris-HEPES-SDS (12:23:1 g L\(^{-1}\)) at 4 °C. After electrophoresis the proteins in the gel were visualized by
staining with Coomassie Brilliant Blue. Precision Plus protein standards (Bio-Rad Laboratories, Inc. California, USA) were used to estimate the molecular weight of the relevant proteins.

Native PAGE was performed with HEPES-Tris buffer (pH 8.2) at 4 °C, 70 volts for 2 h with high molecular weight protein standards (GE Healthcare).

For tryptic digested internal peptide sequence analysis, the protein bands from an SDS-PAGE gel as above were diced in approximately 1 mm cubes and placed in a ZipPlate (Millipore). The gel cubes were then washed thrice for 30 min each. The first wash was with 100 µl of 5 % acetonitrile (ACN), 95 % 25 mM ammonium bicarbonate (NH₄HCO₃) solution, and the last two were with 100 µl of 50 % ACN, 50 % 25 mM NH₄HCO₃ solution. After the third wash the gel cubes were treated with 100 % ACN for 15 min followed by rehydration and protein digestion with 25 mM NH₄HCO₃ solution containing 160 ng µl⁻¹ of sequencing grade trypsin (Promega, Wisconsin, USA). Formic acid (130 µl 1 % v/v) was then added to elute the peptides from the gel cubes and bind them onto the C₁₈ membrane in the plate.

The bound peptides were then eluted with 1 % formic acid in 70 % methanol and 30 % water and loaded on an Agilent Zorbax SB-C18 5 µm 150 X 0.5 mm column. They were eluted from the column with 5 % ACN, 95 % water and 0.1 % formic acid at a flow rate of 20 µl min⁻¹ for 1 min, followed by an increasing gradient of ACN concentrations (firstly to 20 % over 1 min at 5 µl min⁻¹, then to 50 % over 28 min, and finally to 95 % over 1 min). The peptide sequencing was performed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) in an Agilent 1100 capillary liquid chromatography system configured with an Agilent XCT ion trap mass spectrometer.
Mass spectral data resulting from the tryptic digests were matched with a custom sequence database of translations of all open reading frames (ORFs) >150 bp in the genome of the same strain using Agilent’s Spectrum Mill software (revision A.03.02.060).

For N-terminal sequencing, the two peptides recovered from the SDS-PAGE bands were transferred onto a polyvinylidene fluoride (PVDF) membrane by the electro-blotting method, and the membrane stained with amido-black dye for 2 min, followed by destaining several times with cold milli-Q water. The bands were then excised from the PVDF membrane and subjected to 6 cycles of automated Edman degradation on an Applied Bio-systems 494 Precise Protein Sequencing system with β-lactoglobulin as a positive control. All the N-terminal sequencing was performed by the Australian Proteome Analysis Facility (Melbourne, Australia).

3.2.9 Determination of optimum temperature and pH

The optimum temperature of the purified hydrolase enzyme was determined by measuring the initial rates of hydrolysis of DNAN in 50 mM HEPES pH 8.5 at different temperatures (0-60 °C) for 10 min. This involved quantitatively measuring the formation of 2,4-DNP spectrophotometrically as described in section 3.2.7. Similarly, the optimum pH was determined by assaying the enzyme with DNAN in 50 mM HEPES buffer at pHs ranging from 6 to 9.5. All these experiments were done in triplicate.

The metal associated with the DNAN hydrolase enzyme was identified using Inductively Coupled Plasma Mass Spectrometry (ICP-MS) at the Australian National University (ANU), Canberra. Approximately 11 µM purified protein in 50 mM HEPES pH
8.5 was freeze dried in an FD355DMP apparatus (FTS Systems, SP Scientific, Warminster, PA) and redissolved in 2ml of 2 % HNO₃ in RO water. The resultant acid denatured protein was analysed with ICP-MS using Rhodium (Rh) and HNO₃ only as controls.

3.2.10 Determination of kinetic constants

The DNAN hydrolase enzyme was assayed against DNAN and an analogue, 2,4-dinitrophenetole, on a Spectromax-M3 plate reader (Molecular Devices). The reactions were carried out at 28 °C in a final volume of 300 µl containing 53 ng of enzyme in 50 mM HEPES pH 8.5 and concentrations of substrate ranging from 10 µM to 500 µM. All reactions were set up in a 96-well plate and quantitatively measured by 2,4-DNP formation at 400 nm over time. Kinetic parameters were calculated using GraphPad Prism (GraphPad Software, CA, USA).

3.3 Results

3.3.1 Isolation and identification of DNAN-degrading strains

Enrichment culturing with DNAN yielded two bacterial isolates which were able to grow on DNAN as the sole source of carbon, nitrogen and energy. 16S rRNA gene sequence analysis indicated that one of the isolates, designated *Nocardioides* sp. strain JS1661, was most closely related (98 % sequence identity) to *Nocardioides nitrophenolicus* [280]. The second isolate, designated *Nocardioides* sp. strain JS1660, was most closely related (98 % sequence identity) to *Nocardioides oleivorans* [220]. The two strains behaved similarly in growth experiments except that JS1660 was more sensitive to high (>200 µM) DNAN concentrations (data not shown). Therefore subsequent experiments were conducted with *Nocardioides* sp. strain JS1661.
3.3.2 Growth of JS1661

The growth of JS1661 on MM supplemented with 100 μM DNAN was associated with the transient accumulation of 2,4-DNP and the disappearance of the latter was in turn accompanied by release of nitrite (Figure 3-1). About 90% of the theoretically expected nitrite accumulated in the culture fluid. Stoichiometric accumulation of methanol was observed after the depletion of the DNAN (data not shown). The growth yield was 18 ± 2 g of protein mol⁻¹ of DNAN.

Parallel experiments on MM supplemented with 100 μM 2,4-DNP yielded 17 ± 0.5 g of protein mol⁻¹ of 2,4-DNP. Modest yields were expected because most previous reports of 2,4-DNP biodegradation either involved use of supplemental carbon sources or provided no quantitative results of growth yields [96]. JS1661 did not grow on 4-nitroanisole, 2-nitroanisole or methanol (up to 5 mM) (data not shown).

Experiments were then carried out to determine whether the enzymes involved in DNAN and 2,4-DNP metabolism in JS1661 were inducible or constitutive. Cells pre-grown in either DNAN or acetate and transferred to MM supplemented with DNAN.
exhibited no lag period prior to DNAN transformation. The initial specific activity was $153 \pm 22 \text{ nmol min}^{-1} \text{ mg}^{-1} \text{ protein}$ for DNAN grown cells and $141 \pm 8 \text{ nmol min}^{-1} \text{ mg}^{-1} \text{ protein}$ for acetate grown cells. However, cells grown on acetate did not transform 2,4-DNP initially at a detectable rate, whereas 2,4-DNP-grown cells degraded it immediately ($86 \pm 1 \text{ nmol min}^{-1} \text{ mg}^{-1} \text{ protein}$). 2,4-DNP degradation activity only became evident after 30 h for the acetate grown cells (Appendix 5). These results indicate that the enzymes involved in 2,4-DNP biodegradation are inducible whereas the hydrolase activity is constitutive.

### 3.3.3 Enzyme assays

Soluble enzymes in dialyzed extracts of DNAN-grown cells of *Nocardioides* sp. JS1661 catalyzed stoichiometric transformation of DNAN to 2,4-DNP and methanol with an initial specific activity of $170 \pm 2 \text{ nmol min}^{-1} \text{ mg}^{-1} \text{ protein}$ (Figure 3-2). No attempt was made to optimize the assay conditions. Transformation of DNAN ($101 \pm 2 \text{ μM}$) to 2,4-DNP ($103 \pm 7 \text{ μM}$) was accompanied by stoichiometric release of methanol ($104 \pm 4 \text{ μM}$). Formaldehyde was detected in the presence of alcohol oxidase but not in its absence. Addition of cofactors such as NADPH, NADH, or tetrahydrofolate (THF) did not stimulate the transformation of DNAN to 2,4-DNP and oxygen consumption was not required for the reaction (data not shown). No activity was detected with boiled crude lysate (data not shown). These results indicate that the initial reaction in the biodegradation of DNAN is hydrolytic removal of the methyl group to yield 2,4-DNP without a requirement for added cofactors.
2,4-DNP hydride transferase specific activity in extracts prepared from cells grown on 2,4-DNP was 35 nmol min⁻¹ mg⁻¹ protein in the presence of 200 µM NADH. Disappearance of 2,4-DNP was accompanied by formation of the hydride-Meisenheimer complex of 2,4-DNP [17, 64], the identity of which was verified by HPLC comparison with the chemically synthesized standard (Figure 3-3). The hydride-Meisenheimer complex of 2,4-DNP is not commercially available and the synthetic standard was not sufficiently pure to allow quantification, as has been noted previously [17, 64, 211]. The hydride-Meisenheimer complex disappeared from the reaction mixture after a few minutes, which indicated that it was unstable or further transformed [17, 64, 211]. No release of nitrite or accumulation of other metabolites was detected. 2,4-DNP was not transformed when NADPH was substituted for NADH (data not shown). The results indicate that the initial reactions in 2,4-DNP biodegradation involve the transfer of hydride ions to form hydride-Meisenheimer complexes of 2,4-DNP, as established previously in Rhodococcus sp. [141, 211] and Nocardia sp. [17, 64, 65].

Figure 3-2. Biodegradation of DNAN by soluble, dialysed cell extracts of Nocardioiodes sp. JS1661.
Given the results above, the following DNAN biodegradation pathway is proposed (see also Figure 3-4 below). Initially the DNAN is hydrolysed by a cofactor independent hydrolase enzyme to 2,4-DNP and methanol, and the 2,4-DNP in turn is a substrate for an inducible NADPH-dependent reductase, which metabolizes it through a well-established hydride-Meisenheimer complex, releasing the nitrite.

![Image of HPLC analysis]

**Figure 3-3.** HPLC analysis of (A) *Nocardioides* sp. JS1661 culture fluid analyzed after incubation with 2,4-DNP for 2 min, and (B) the hydride-Meisenheimer complex synthetic standard.
Figure 3-4. Proposed pathway of DNAN biodegradation by *Nocardioides* sp. JS1661.

### 3.3.4 Genes encoding DNAN hydrolase

DNAN hydrolase was purified from JS1661 cells in order to determine the protein sequence and identify the gene(s) responsible. The purification described in section 3.2.7 achieved > 90 % pure protein after five purification steps, with 33.5 fold purification and 3 % yield (Table 3-1). The specific activity of the purified enzyme was 103.9 µmoles min⁻¹ mg⁻¹ protein.

<table>
<thead>
<tr>
<th>Purification step</th>
<th>Total protein (mg)</th>
<th>Specific activity (min⁻¹ mg⁻¹ protein)</th>
<th>Total activity (nmoles min⁻¹)</th>
<th>Fold of purification</th>
<th>Yield %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell-free crude enzyme</td>
<td>2615.4</td>
<td>3E+03</td>
<td>4E+06</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Ammonium sulphate precipitation (28-40 %)</td>
<td>800.0</td>
<td>8E+03</td>
<td>3E+06</td>
<td>2.5</td>
<td>77</td>
</tr>
<tr>
<td>Weak hydrophobic interaction chromatography</td>
<td>71.1</td>
<td>4E+04</td>
<td>2E+06</td>
<td>13.1</td>
<td>36</td>
</tr>
<tr>
<td>Gel filtration chromatography</td>
<td>25.9</td>
<td>5E+04</td>
<td>6E+05</td>
<td>14.1</td>
<td>14</td>
</tr>
<tr>
<td>Anion exchange chromatography</td>
<td>12.5</td>
<td>6E+04</td>
<td>4E+05</td>
<td>18.5</td>
<td>9</td>
</tr>
<tr>
<td>Strong hydrophobic interaction chromatography</td>
<td>2.3</td>
<td>1E+05</td>
<td>1E+05</td>
<td>33.5</td>
<td>3</td>
</tr>
</tbody>
</table>
Electrophoresis of the purified protein revealed a single active band on a native gel with an approximate molecular weight of 234 kDa, and two approximately equimolar bands of approximately 39 and 37 kDa on SDS PAGE (Figure 3-5). These results suggest that the DNAN hydrolase may be a heteropolymer of two subunits in equal proportions.

**Figure 3-5.** Native gel (A) and denatured SDS-PAGE gel (B) of the purified DNAN hydrolase. The upper band of the denatured SDS-PAGE (36 kDa) represents dnH1 and the lower band (34 kDa) represents dnH2. (C) SDS-PAGE gel electro blotted on PVDF membrane for N-terminal sequencing.

Screening of the tryptic and N terminal sequences (Table 3-2) of the two bands against the inferred gene products in the JS1661 genome identified two matches, to peptides designated dnH1 and dnH2 respectively. The matched sequences comprised 37 % (122/328 residues) and 55 % (178/319 residues) of the peptides respectively (Figure 3-6). A search against the protein family (Pfam) database found dnH1 had 22 % identity to a metallo-beta-lactamase protein (PF00753) from *Bacillus cereus* [37] while no Pfam identity was identified for dnH2.

The *dnH1* and *dnH2* genes were found to overlap slightly with one another on the same strand of a 5 kb contig in the draft JS1661 genome (Figure 3-7). The overlap
involves four nucleotides, GTGA, which encompass the stop and start codons for the 
dnH1 and dnH2 genes respectively. Thus the dnH2 start methionine is encoded by a 
rarer GTG codon, rather than the more common ATG codon. The two genes would be 
co-translated through a single ribosomal binding site 5´ to dnH1. The nearest upstream 
ORF encodes a hypothetical protein and the downstream ORF is most closely related (25 
% amino acid identity) to a gene encoding a beta phytoene dehydrogenase from 
_Myxococcus xanthus_ which appears to have no functional connection with DNAN 
metabolism (Figure 3-8). The upstream and downstream genes were not in the same 
operon as _dnH1_ and _dnH2_.

**Figure 3-6.** Inferred amino acid sequences of the dnH1 (A) and dnH1 (B) proteins 
showing matches to the tryptic and internal peptides sequences (red and yellow 
highlight respectively). Assuming N-terminal threonines, it appears that the first nine 
amino acids of dnH1 (marked in blue) were post-translationally removed by an unknown 
mechanism.
Given the primary sequences of dnH1 and dnH2, it appears that the DNAN hydrolase may be heterohexamer comprising three copies each of the two subunits.

Table 3-2. N-terminal sequences of the DNAN hydrolase from Edman N-terminal sequencing

<table>
<thead>
<tr>
<th>Cycle #</th>
<th>Major sequence</th>
<th>Minor sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T, S, G</td>
<td>D, A</td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td>S, E, Y</td>
</tr>
<tr>
<td>3</td>
<td>(S)</td>
<td>N, A</td>
</tr>
<tr>
<td>4</td>
<td>A, T</td>
<td>I</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>V</td>
</tr>
<tr>
<td>6</td>
<td>V, D</td>
<td>N, L</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cycle #</th>
<th>Major sequence</th>
<th>Minor sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T</td>
<td>D, G, A</td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>E, N, D</td>
</tr>
<tr>
<td>5</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>T</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3-7. DNA sequence of the DNAN hydrolase genes indicating the single ribosomal binding site (RBS) and putative start/stop codons. See Appendix 6 for the full sequence.
3.3.5 Genes encoding 2,4-DNP catabolic pathways

The *Nocardioides* sp. JS1661 genome also contains genes very similar in sequence and organization to the genes involved in picric acid and 2,4-DNP degradation in *Rhodococcus opacus* [96, 269], except for the insertion of several genes in JS1661 that appear to be mainly involved in transport functions (Figure 3-8 and Table 3-3). The 2,4-DNP catabolic and the DNAN hydrolase genes were not located on the same contig in the draft genome.

**Figure 3-8.** Organization of genes involved in DNAN demethylase (A) and 2,4-DNP degradation (B) in *Nocardioides* sp. JS1661. The 2,4-DNP catabolic genes are also compared with their homologues from *Rhodococcus opacus* [96] (C). Protein coding genes represent: orfA, gene encoding hypothetical protein; dnH1, upstream hydrolase; dnH2, downstream hydrolase; carB, gene encoding putative phytoene dehydrogenase; orfB, L-carnitine dehydratase; npdC, hydride transferase I; orfD, aldehyde dehydrogenase; orfE, acetyl-CoA synthetase; orf1, ABC transporter; orf2, amino acid transport system permease; orf3, ABC transporter permease; orf4, ABC transporter ATP-binding; orf5, amino acid transport ATP-binding protein; orff, Lyase; npdR, transcriptional regulator; npdG, NADPH-dependent F420 oxidoreductase; npdH, protein converting 2H−PA to product X; npdI, hydride transferase II; orfJ, enoyl-CoA hydratase; orfK, acyl-CoA dehydrogenase.
Table 3-3. Genes encoding DNAN hydrolase and 2,4-DNP catabolic enzymes in comparison with the closest amino acid matches from the NCBI database.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Enzyme or protein size (no. of aa)</th>
<th>Proposed/confirmed function</th>
<th>Identity (%)</th>
<th>Closest organism</th>
<th>Accession no. of closest match</th>
</tr>
</thead>
<tbody>
<tr>
<td>orfA</td>
<td>107</td>
<td>Hypothetical protein</td>
<td>59</td>
<td>Achromobacter xylosoxidans</td>
<td>WP_006384465</td>
</tr>
<tr>
<td>dnhA</td>
<td>328</td>
<td>DNAN hydrolase α-subunit (metallo-beta-lactamase)</td>
<td>22</td>
<td>Bacillus cereus</td>
<td>Gl-157836766</td>
</tr>
<tr>
<td>dnhB</td>
<td>318</td>
<td>DNAN hydrolase β-subunit (hypothetical protein)</td>
<td>26</td>
<td>Marinobacter jannaschii</td>
<td>WP_027858191</td>
</tr>
<tr>
<td>carB</td>
<td>541</td>
<td>Phytoene dehydrogenase</td>
<td>26</td>
<td>Myxococcus xanthus</td>
<td>WP_011551016</td>
</tr>
<tr>
<td>orfB</td>
<td>394</td>
<td>L-Carnitine dehydratase</td>
<td>73</td>
<td>Rhodococcus opacus</td>
<td>WP_005257508</td>
</tr>
<tr>
<td>npdC</td>
<td>295</td>
<td>Hydride transferase</td>
<td>85</td>
<td>Rhodococcus opacus</td>
<td>AAK29142</td>
</tr>
<tr>
<td>orfD</td>
<td>483</td>
<td>Aldehyde dehydrogenase</td>
<td>80</td>
<td>Rhodococcus opacus</td>
<td>WP_005257514</td>
</tr>
<tr>
<td>orfE</td>
<td>532</td>
<td>Acetyl-CoA synthetase</td>
<td>65</td>
<td>Rhodococcus opacus</td>
<td>AAK38099</td>
</tr>
<tr>
<td>orf1</td>
<td>386</td>
<td>ABC transporter</td>
<td>35</td>
<td>Sphaerobacter thermophilus</td>
<td>WP_012873153</td>
</tr>
<tr>
<td>orf2</td>
<td>287</td>
<td>Amino acid transport system permease</td>
<td>40</td>
<td>Actinoteca ferrariae CFS-4</td>
<td>EYR64100</td>
</tr>
<tr>
<td>orf3</td>
<td>355</td>
<td>ABC transporter permease</td>
<td>35</td>
<td>Bradyrhizobium sp. strain ORS 375</td>
<td>WP_009030206</td>
</tr>
<tr>
<td>orf4</td>
<td>222</td>
<td>ABC transporter ATP-binding protein</td>
<td>43</td>
<td>Geobacillus thermocatenulatus</td>
<td>WP_025950112</td>
</tr>
<tr>
<td>orf5</td>
<td>231</td>
<td>Amino acid transport ATP-binding protein</td>
<td>45</td>
<td>Amphimedon queenslandica</td>
<td>XP_003390336</td>
</tr>
<tr>
<td>orfF</td>
<td>160</td>
<td>Lyase</td>
<td>74</td>
<td>Rhodococcus opacus</td>
<td>AAK38100</td>
</tr>
<tr>
<td>npdR</td>
<td>268</td>
<td>Transcriptional regulator</td>
<td>77</td>
<td>Rhodococcus opacus</td>
<td>AAK38101</td>
</tr>
<tr>
<td>npdG</td>
<td>220</td>
<td>NADPH-dependent F420 oxidoreductase</td>
<td>72</td>
<td>Rhodococcus opacus</td>
<td>AAK29135</td>
</tr>
<tr>
<td>npdH</td>
<td>95</td>
<td>Protein converting 2H−PA to product X</td>
<td>54</td>
<td>Rhodococcus opacus</td>
<td>WP_005257523</td>
</tr>
<tr>
<td>npdI</td>
<td>349</td>
<td>Hydride transferase II</td>
<td>86</td>
<td>Rhodococcus opacus</td>
<td>AAK29131</td>
</tr>
<tr>
<td>orfJ</td>
<td>232</td>
<td>Enoyl-CoA hydratase</td>
<td>61</td>
<td>Rhodococcus opacus</td>
<td>AAK38105</td>
</tr>
<tr>
<td>orfK</td>
<td>384</td>
<td>Acyl-CoA dehydratase</td>
<td>87</td>
<td>Rhodococcus opacus</td>
<td>WP_00525752</td>
</tr>
</tbody>
</table>

a The first 4 genes in the table are on a separate contig from the genes below.
b aa, amino acids, c Functions established in this study, d Biochemically characterized previously.
3.3.6 DNAN hydrolase characterization

The activity of the purified DNAN hydrolase enzyme was tested across a temperature range from 20 °C to 60 °C, showing highest activity at 20 °C and essentially no activity at 50 °C or above (Figure 3-9 A). Its pH optimum was found to be around pH 9, suggesting a reaction mechanism involving base-catalysed hydrolysis [73] (Figure 3-9 B).

The ICP-MS analysis showed that the DNAN hydrolase contains Zn, which is a characteristic feature of metallo-beta-lactamases (Figure 3-10). The data suggest a molar ratio of metal:enzyme of almost 1:1, however protein crystallization and empirical analyses of the reaction chemistry would be needed to confirm this.

Figure 3-9. Effect of temperature (A) and pH (B) on the activity of the DNAN hydrolase
Figure 3-10. Metal analysis of the DNAN hydrolase on ICP-MS, with Rh used as a positive control.

### 3.3.7 Substrate specificity and kinetic analysis

The activity of the DNAN hydrolase enzyme was tested against various di-nitro (DNAN, 2,4-dinitrophenetole and 1,5-dimethoxy-2,4-dinitrobenezene) and mono-nitro (2-nitroanisole and 4-nitroanisole) compounds. The enzyme was found to catalyse the transformation of the three di-nitro compounds (Table 3-4 and data not shown) but no activity was observed for the mono-nitro compounds (data not shown). DNAN was the preferred substrate among the two di-nitro compounds for which kinetic constants could be calculated, because of its higher $K_{cat}$ and $K_{cat}/k_M$ and lower $k_M$. Kinetic constants could not be determined against 1,5-dimethoxy-2,4-dinitrobenezene because of this compound’s insolubility.
Table 3-4. Kinetic constants for DNAN hydrolase from *Nocardioides* sp. strain JS1661 on DNAN and 2,4-dinitrophenetole

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_M$ (µM)</th>
<th>$k_{cat}$ (nmoles sec$^{-1}$ mg$^{-1}$)</th>
<th>$k_{cat}/K_M$ (µM$^{-1}$ sec$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNAN</td>
<td>30.32 ± 3.96</td>
<td>1927 ± 53.84</td>
<td>63.55 ± 13.59</td>
</tr>
<tr>
<td>2,4-dinitrophenetole</td>
<td>142 ± 19.5</td>
<td>540.4 ± 26.88</td>
<td>3.80 ± 1.37</td>
</tr>
</tbody>
</table>

3.3.8 Discussion

I have found that the first step in DNAN degradation in the DNAN mineralizing bacterium *Nocardioides* sp. strain JS1661 involves $O$-demethylation, which is achieved by an enzyme from the metallo-beta-lactamase (MBL) family. Notably this family has not previously been associated with $O$-demethylation activity. The enzyme is in fact a complex heteromeric protein with no close relatives, even within the MBL family. The two gene products contributing to the heteromeric enzyme are sequence-related to one another, albeit relatively distantly (21.3%), and are co-transcribed and co-translated from overlapping genes. This suggests that they have co-evolved over significant evolutionary times. However DNAN itself is a synthetic molecule that would have only been introduced into the environment relatively recently and it may be that the enzyme has only recently utilized DNAN as a substrate. The possible evolutionary origins of the enzyme are considered in more detail below and in chapter 4.

Related to this issue, however, is the fact that four sequence-unrelated classes of bacterial enzymes have been implicated in $O$-demethylase reactions for methoxy substituted aromatics in the past. These respectively involve the Rieske non-heme iron oxygenases [32, 46, 243] and cytochrome P-450 dependent $O$-demethylases [234, 276],
both of which require oxygen and NAD(P)H, the THF-dependent O-demethylases [162], and the atrazine hydrolase (trzN) from Arthrobacter [227, 228]. The latter has been reported to catalyze cofactor independent cleavage of the methoxy group of the heterocyclic aromatic atratone, albeit at a relatively slow rate [225, 228]. The question as to why none of these four types of enzymes is involved in DNAN demethylation in JS1661 is also addressed in some detail in chapter 4 below.

The biodegradation pathway of 2,4-DNP is well established for Gram-positive bacteria such as Rhodococcus erythropolis [141, 211], Nocardioides simplex FJ2-1A [64, 65], and Nocardioides sp. strain CB 22-2 [17]. The pathway is initiated by the formation of the hydride-Meisenheimer complex of 2,4-DNP by the action of an NADPH-dependent reductase and hydride transferases [64]. In contrast, it has been suggested that the initial step of 2,4-DNP biodegradation in Gram-negative bacteria such as Burkholderia sp. strain KU-46, involves removal of a nitrite group to form 4-nitrophenol [110]. Neither nitrite nor 4-nitrophenol was detected during the transformation of 2,4-DNP by cell extracts from Nocardioides sp. JS1661 and the strain did not grow on 4-nitrophenol. Taken together with the presence of the genes encoding the gram-positive pathways for 2,4-DNP degradation in JS1661, the results indicate that the pathway of 2,4-DNP biodegradation in Nocardioides JS1661 is closely related to the pathway previously established for the Rhodococcus and Nocardia species.

The presence of genes involved in 2,4-DNP biodegradation, the constitutive hydrolase activity and the inducible 2,4-DNP degradation all suggest that the hydrolase genes have been acquired relatively recently by the parental 2,4-DNP-degrading strain to enable it to grow on DNAN. Bacteria are known to evolve by recruitment of genes
encoding catabolic enzymes to extend existing pathways [38, 113, 129, 223] and it is common for the early stage of pathway assembly to involve loss of regulatory functions that results in constitutive expression of key enzymes [38]. Based on these characteristics and the unique reaction chemistry of the DNAN-O-demethylation, the novel DNAN O-demethylase enzyme described here has been classified with a new Enzyme Commission number (EC 3.3.2.14).

The robust biodegradation of DNAN by JS1661 also suggests that the isolate might be a good candidate for potential applications in waste treatment and biodegradation. The fact that it was isolated from a DNAN manufacturing plant suggests that it may be involved in DNAN degradation in the waste treatment system at the site. We are currently investigating its activity and distribution in other ecosystems, and characterizing the demethylase enzyme and its evolution. The potential applications of the DNAN hydrolase are considered in more detail in chapter 4.
Chapter 4  General discussion

4.1  The lignin dehydrogenases

This section provides a general discussion of the findings described in Chapter 2 on the stereospecific Cα-dehydrogenases from *Erythrobacter* sp. SG61-1L (SG61-1L) and *Sphingomonas* sp. SYK-6 (SYK-6) that catalyse the first step in the cleavage of the C-O-C bonds in certain lignin dimers.

As noted earlier, lignin is a complex molecule with an indefinite structure whose subunits are tightly bound together in various ways and proportions. Because of this, lignin has proven to be relatively recalcitrant to degradation in nature. It is also difficult to monitor its biodegradation through many chromatographic techniques. However, as a potentially rich source of aromatic alcohols, it is an attractive target for systematic deconstruction by biocatalytic means [150, 193].

Both SG61-1L and SYK-6 were isolated from Kraft paper mill samples, but SYK-6 was isolated from Japan in 1989 and SG61-1L from Australia in 2014. The two strains have qualitatively comparable catabolic activities with the model racemic GGE substrate but SG61-1L degraded it faster than SYK-6 in my growth experiments. Some of the individual dehydrogenases I isolated and characterized from SG61-1L also proved to be more effective catalysts of the first step in the degradation than those previously isolated from SYK-6. In particular the 2550 and 724 enzymes from SG61-1L showed relatively high catalytic efficiencies for all four isomers. The SG-61-1L dehydrogenases, which showed 20-67 % identity with the SYK-6 dehydrogenases (Table 4-1), may have been subject to more intense or sustained selection pressures to degrade lignin than
those in SYK-6. Notably, while lignin has been a part of the natural environment over geological time scales [233], Kraft lignin is a product of relatively recent industrial processes [233] and such effective selection for the degradation and utilization of GGE may be a relatively recent and geographically localized phenomenon.

**Table 4-1.** Percentage identities of the Lig dehydrogenase amino acid sequences of SYK-6 (columns) and SG61-1L (rows).

<table>
<thead>
<tr>
<th>Genes</th>
<th>LigD</th>
<th>LigL</th>
<th>LigN</th>
<th>LigO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lig474</td>
<td>21 %</td>
<td>21 %</td>
<td>17 %</td>
<td>19 %</td>
</tr>
<tr>
<td>Lig724</td>
<td>33 %</td>
<td>33 %</td>
<td>28 %</td>
<td>35 %</td>
</tr>
<tr>
<td>Lig1498</td>
<td>32 %</td>
<td>26 %</td>
<td>67 %</td>
<td>36 %</td>
</tr>
<tr>
<td>Lig2549</td>
<td>25 %</td>
<td>27 %</td>
<td>28 %</td>
<td>32 %</td>
</tr>
<tr>
<td>Lig2550</td>
<td>29 %</td>
<td>37 %</td>
<td>26 %</td>
<td>31 %</td>
</tr>
<tr>
<td>Lig2705</td>
<td>34 %</td>
<td>35 %</td>
<td>36 %</td>
<td>34 %</td>
</tr>
<tr>
<td>Lig2706</td>
<td>29 %</td>
<td>36 %</td>
<td>29 %</td>
<td>33 %</td>
</tr>
<tr>
<td>Lig3329</td>
<td>26 %</td>
<td>27 %</td>
<td>27 %</td>
<td>29 %</td>
</tr>
<tr>
<td>Lig3344</td>
<td>17 %</td>
<td>20 %</td>
<td>16 %</td>
<td>20 %</td>
</tr>
</tbody>
</table>

The Lig enzymes are alcohol dehydrogenases (ADHs). As a group, ADHs are very widely distributed in both prokaryotes and eukaryotes [3, 208] and generally quite numerous within individual organisms (often > five per bacterial genome [184]). They catalyse the oxidation of alcohols to aldehydes or ketones using various electron acceptors, commonly NAD⁺ and NADP⁺. The alcohols may be linear, branched, cyclic, primary and secondary alcohols [3, 67, 208]. Generally NADP⁺ is utilized by ADHs involved in anabolic processes while NAD⁺ is used more by ADHs involved in catabolic processes [3]. ADHs are divided into superfamilies, families and subfamilies as described below.
At the highest level the ADHs are categorized into three groups based on the type of electron acceptor utilized: (1) NAD⁺ or NADP⁺ dependent ADHs, (2) pyrroloquinoline quinone (PQQ), haem-group and F420 dependent enzymes, and (3) flavin adenine dinucleotide (FAD⁺) dependent enzymes [208]. The NAD⁺ or NADP⁺ dependent ADHs, which are the most extensively characterized, are further categorized into three superfamilies: (1) the zinc (Zn) dependent ADHs, (2) metal independent short chain ADHs and (3) iron activated ADHs [52, 105, 119, 208]. These three superfamilies have different structural folds and show no significant sequence similarities to one another.

The Lig dehydrogenases belong to the metal-independent short chain ADH (SDR) superfamily, which contains about 25 % of all identified dehydrogenases to date [118, 119] and encodes a great diversity of biological functions. ADHs in this superfamily are characteristically about 250-350 residues long and share definitive domains, even though they may have as little as 15-30 % amino acid sequence similarities to one another [118, 119, 208]. Their most conserved feature is the Rossmann fold, in which a twisted β-sheet is flanked by 2 – 3 α-helices on each side to create a distinctive cofactor binding site in the N-terminal part of the protein. The C-terminal part of the molecule contains the substrate binding site [118, 125, 208]. The two major families of SDRs are categorized as classical (~250 residues) and extended (~350 residues) SDRs, although three much smaller families, denoted as intermediate, divergent and complex ADHs, are also recognized [118, 119, 208]. Most but not all of the classical SDRs prefer NADP⁺, and most but not all extended SDRs prefer NAD⁺ as co-factors [95, 118, 119, 208].
The classical SDRs are further divided into seven subfamilies according to their coenzyme preference and the molecular basis for it. Four of the subfamilies (cD1d, cD1e, cD2, cD3) bind NAD$^+$ while the other three (cP1, cP2 and cP3) bind NADP$^+$. Enzymes belonging to the NAD$^+$-binding subfamilies contain an acidic amino acid at or near the C-terminal end of the second $\beta$-strand in the Rossmann fold, while those displaying a preference for NADP$^+$ have a basic amino acid around the same position. The precise locations of these acidic or basic residues are determinants for which subfamily they belong to. For example, of the NAD$^+$ preferring subfamilies, cd1d and cd1e include enzymes that have an aspartate or glutamate at the end of the second $\beta$-strand, respectively, while cd2 and cd3 include enzymes that have an acidic residue in the first or second position after the second $\beta$-strand, respectively. Of the NADP$^+$ preferring subfamilies, cp2 includes enzymes that contain a basic residue in the first position after the second $\beta$-strand, cp1 includes enzymes containing a basic residue in the gly-rich motif (located upstream of the second $\beta$-strand), and cp3 contains a basic residue at each of the above mentioned positions [95, 118, 119, 208].

While fewer members of the extended SDR family have been structurally characterized and less is known about the exact determinants for their coenzyme preference, this family has also been classified into three subfamilies based on cofactor preference and its molecular basis. Subfamilies eD1 and eD2 include enzymes that in general prefer NAD$^+$ and contain an acidic amino acid either at the end of the second $\beta$-strand or in the second position after the strand, respectively. Subfamily eP1 includes enzymes that in general prefer NADP$^+$ and contain a basic residue in the first position after the second $\beta$-strand [95, 118, 119, 208].
A conserved domain database search showed the Lig dehydrogenases have higher similarities to classical than extended SDRs [154]. Preliminary crystallographic analyses of the 3344, 474 and 1498 enzymes by Dr. N. Dellas have identified a Rossmann fold in each, with glycine at the key \( \beta \)-strand position in 3344 and aspartic acid at that position in the 474 and 1498 enzymes. This correlates with the classical SDR classification as the 3344 enzyme utilizes NADP\(^+\) and the 474 and 1498 enzymes utilize NAD\(^+\).

The reaction chemistry of ADHs involves the transfer of a hydride group from a substrate (such as GGE) to the oxidised form of a cofactor (such as NAD\(^+\)) to form the corresponding aldehyde or (in the case of GGE) ketone [138]. In the case of GGE, both the substrates and the reduced cofactor product (NADH) can exist as stereoisomers (4 and 2 respectively; see Figures 2-2 and 4-1) and, in many other SDR ADHs at least, the hydride transfer between substrate and cofactor occurs in an interdependent stereospecific way [189]. The nature of the stereospecificity of the enzyme for its substrate is usually largely determined by the relative bulks of the groups attached to the carbonyl carbon of the substrate [189]. The aromatic groups in each GGE subunit and the relative chirality of the two hydride groups that are candidates to be transferred will mean that some of the isomers would potentially occupy quite different spaces in the enzyme’s active sites. It is therefore unsurprising that we see a range of GGE stereospecificities across the different Lig enzymes, with preferences for \( \alpha S, \beta S \) and \( \alpha S, \beta R \) usually correlated. Conversely it is notable that SG61-1L 2550 and 724 have unusually high activities for all four isomers [184, 189].
Interestingly, another NAD$^+$ dependent dehydrogenase (from *Pseudomonas* sp GU5) with activity against GGE has been identified and characterized [188]. It has a sub unit molecular mass of 52 kDa, which is much larger than the Lig dehydrogenase under discussion here, and indeed most other classical SDR dehydrogenases. Further characterization of this enzyme (such as its metal dependence and amino acid residues at the key sites in the Rossmann fold) has not been reported but, based on its molecular weight, it could be an extended SDR type of dehydrogenase.

A search of the non-redundant NCBI database for Ca-dehydrogenases with at least 20 % amino acid identity to any of the Lig dehydrogenases from SYK-6 or SG61-1L (Table 4-1) identified several from *Novosphingobium* species (chapter 2 and data not shown) [198]. Three of these species (*Novosphingobium* sp. strain PP1Y, *Novosphingobium aromaticivorans* strain DSM12444 and *Novosphingobium* sp. B7) are of particular interest because they also have GST genes which have been shown to encode enzymes downstream of the Lig dehydrogenases in the degradation pathway of GGE [82]. Functional characterization of the dehydrogenases from these species...
showing greatest sequence similarity to the Lig dehydrogenases would now be worthwhile.

The phylogenetic analysis of the Lig dehydrogenases from SYK-6 and SG61-1L and selected other sequences showed that many but not all of the enzymes known to be active against GGE cluster together in a ‘GGE dehydrogenase clade’ (Figure 2-8). The origins of this clade are unknown. The same is true for the two active enzymes that lie outside this clade, although in the latter cases we do have at least some clues as to their origins. Thus one of them lies in a clade containing a functionally characterized levodione reductase from *Corynebacterium aquaticum* and the other is NADP⁺ rather than NAD⁺ dependent, suggesting it may have evolved from biosynthetic rather than catabolic dehydrogenases. The fact that active GGE dehydrogenases have been found in three different clades is itself noteworthy, suggesting possible multiple origins for enzymes that have evolved activity against GGE.

Interestingly there is considerable heterogeneity in the stereospecificities of the GGE dehydrogenases even within the GGE dehydrogenase clade. For example 1498 and LigN are 67% similar, yet 1498 was only able to oxidize the (αS,βR)- and (αS,βS)-GGE while LigN could oxidize all four stereoisomers. This could be because of differences in C-terminal sequences responsible for substrate binding. However empirical structures and substrate docking results would be needed to identify the specific sequence differences responsible. Such work is currently under way in a collaboration between Dr N. Dellas, my supervisors Drs Gunjan Pandey and John Oakeshott, and Drs Tom Peat and Janet Newman in CSIRO Materials Science and Engineering. Currently Dr Dellas has
obtained crystal structures for the 3344, 474 and 1498 dehydrogenases of SG61-1L but docking studies are still in progress.

The range of stereospecificities found across the GGE dehydrogenases I have characterized is of particular interest in respect of potential applications in the chemical manufacturing industry. As noted earlier, stereospecific enzymes are of great interest as potential biocatalysts for the production of pharmaceuticals and fine chemicals whose functionality depends on particular chirality [45, 109, 231, 252]. One example of an enantioselective dehydrogenase being used to synthesize a high value drug involves the enantioselective reduction of acetophenone to the antimicrobial agent \(R^+\)-phenylethanol by a NADPH dependent alcohol dehydrogenase isolated from *Lactobacillus kefir* [109]. Another example involves the conversion of cyclohexanol to cyclohexanone by a cyclohexanol dehydrogenase from *Acinetobacter NCIMB 9871* which is then used in the production of chiral lactones [252]. Similarly, one or more of the lignin dehydrogenases could have potential in the production of various stereospecific ketones from alcoholic substrates. Modern protein engineering could be used to make these enzymes fit for a variety of such purposes [45]. These efforts would be greatly aided by the structural work currently under way as described above.

In respect of biomass utilization, I note that there is increasing use of enzymes such as xylanases, lipases and primary lignin degrading enzymes (laccases) in the paper and pulp industries. These enzymes are used to increase the paper quality by removing or degrading the wood pitch, lignin and hemicelluloses; the lignin in particular makes paper brittle and reduces its longevity [45, 267]. Cellulases and laccases are also used in recycling waste paper and to reduce the chemicals and waste in these industries [40,
Enzymes can also be applied to plant biomass to transform it into biofuel; for example glycoside hydrolases (GHs) from the bacterium *Caldanaerobius polysaccharolyticus* are used synergistically to break β-1,4 glucosidic and β-1,4 mannosidic linkages to transform plant cell wall polysaccharides to fermentable sugars, which can then be further converted to ethanol and other molecules [21]. Likewise cellobionic acid phosphorylase (CBAP) from the cellulolytic bacterium *Saccharophagus degradans* catalyses the breakdown of cellobionic acid to produce fermentable sugars or biofuel molecules [171]. As described in chapter 1, the lignin peroxidases from organisms such as *Rhodococcus jostii* RHA1, *Streptomyces viridosporus* T7A and *Amycolatopsis* sp. 75iv2 could potentially be used to degrade lignin into smaller molecules and the lignin dehydrogenases identified here may have value in converting some of the latter into biofuel and commodity chemicals. It is helpful in this respect that the lignin dehydrogenases characterized here have relatively low *Km* values (e.g. 0.39 ± 0.08 µM for LigD against αS,βR) and high *kcat/Km* values (e.g. 360 ± 74 µM⁻¹ s⁻¹ for 724 against αR,βS).

### 4.2 DNAN demethylase

This section provides a general discussion of the findings described in chapter 3 on the DNAN demethylase enzyme, its relationship to other demethylases, its evolutionary origins and its potential future biotechnological applications.

The DNAN demethylase enzyme which I have found is responsible for transforming DNAN to 2,4-dinitrophenol in *Nocardioides* sp. JS1661 is a heterohexameric product of two genes, *dnH1* and *dnH2*. Bioinformatic analysis shows that these two genes are positioned on a single 5 kb contig and are expressed through a
coupled transcription-translation mechanism. The two genes share four nucleotides (GTGA) which act as termination and reinitiating codons for \textit{dnH1} and \textit{dnH2} respectively, although another GTG located four codons downstream could also act as an alternative start codon for \textit{dnH2}. Transcripts from the two genes are co-translated from a single ribosomal binding site. The \textit{dnH1} and \textit{dnH2} genes show 21% amino acid identity to each other. A homologue search found the nearest relative of both which has an annotated function has only 22% and 23% amino acid identity to them respectively (Table 3-3). It is a metallo-beta-lactamase protein (PF00753) from \textit{Bacillus cereus} [37]. The nearest upstream ORF encodes a hypothetical protein and the downstream ORF is most closely related (25% amino acid identity) to a beta phytoene dehydrogenase from \textit{Myxococcus xanthus} which appears to have no functional relationship to the DNAN demethylase genes. Neither of these flanking genes is in the same operon as \textit{dnH1} and \textit{dnH2}.

2,4-dinitrophenol (2,4-DNP) is an intermediate in the degradation pathway for DNAN by JS1661 and genes inferred to encode enzymes in its 2,4-DNP degradation pathway were also found in the genome of this bacterium (Figure 3-8). They are arranged in an operon with an inducible mechanism, as described earlier, which suggests that this pathway is quite highly evolved for 2,4-DNP. By contrast, the DNAN demethylase operon is located between unrelated genes and expressed constitutively, which suggests that JS1661 has acquired it more recently. The early stages of pathway assembly often involve the loss of specific regulatory functions, resulting in constitutive expression of the genes in question [73, 76, 113].
A homologue search against the NCBI non-redundant database found the closest matched genes to \textit{dnH1} and \textit{dnH2} encode two hypothetical proteins from another \textit{Nocardia} strain, \textit{Nocardia testacea}, which was isolated from the sputum of a Japanese male patient suffering from a non-tuberculosis mycobacterial infection. Interestingly \textit{N. testacea} and \textit{Nocardioides} sp. JS1661 are isolated from different countries and sources yet the \textit{N. testacea} proteins give the best match for the DNAN degrading genes so far.

The two \textit{N. testacea} genes are also located on a single contig, as per \textit{dnH1} and \textit{dnH2} in JS1661. The \textit{dnH1} amino acid sequence has 31 and 63 % identity to WP\_039827747 and WP\_039827746 respectively and the corresponding figures for \textit{dnH2} are 41 and 22 % (Table 4-2). WP\_039827747.1 and WP\_039827746.1 were 31.7 % similar to each other. Pfam categorization of WP\_039827747.1 clustered it with the metallo-beta-lactamase protein (PF00753) of \textit{Bacillus cereus} as per the two JS1661 proteins, but no Pfam grouping was identified for WP\_039827746.1 [37].

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4-2.png}
\caption{DNA sequences of parts of the \textit{Nocardia testacea} sp. homologues showing the putative start and stop codons between the two genes. This arrangement is precisely similar to that which I have found between the two DNAN demethylase genes of \textit{Nocardioides} sp. JS1661.}
\end{figure}
The DNAN demethylase nucleotide sequences of JS1661 and their homologues in *N. testacea* are 27-54% identical to each other, as shown in Table 4-3 and Fig 4-3. The amino acid sequences of the two peptides in JS1661 were 329 and 319 residues long respectively, corresponding to molecular masses of 39 and 37 kDa respectively. The homologues in *N. testacea* were 327 (*WP_039827746.1*) and 213 (*WP_039827747.1*) residues long respectively, giving peptide masses of 35 and 28 kDa respectively. Interestingly, the two genes in *N. testacea* also appear to be co-transcribed and translated, with four overlapping nucleotides, similar to the arrangement of the homologues *Nocardioides* sp. JS1661 genes (Figures 4-2 and 3-7). The nearest upstream ORFs in *N. testacea* encode a hypothetical protein and a histidine kinase with a two component regulatory system and the nearest downstream genes encode three mobile element proteins. None of these upstream or downstream ORFs are part of the operon containing the beta lactamase genes. The presence of the mobile element genes suggests that the bacterium may have acquired the beta-lactamase genes by some form of horizontal gene transfer. While *N. testacea* has not yet been found in soil, it would be worthwhile testing it for the ability to metabolize DNAN and structurally related compounds. This could help to elucidate the evolutionary origin of the DNAN demethylase enzyme [73, 76].
Figure 4-3. DNA sequences of the \( dhn1 \) and \( dhn2 \) genes from *Nocardoides* sp J1661 and their homologues from *Nocardoides* testacea.
Table 4-3. Array of DNA sequence identities among the two Nocardioides sp JS1661 and two N. testaceus genes:

<table>
<thead>
<tr>
<th>Gene</th>
<th>N. testacea dnH1</th>
<th>N. testacea dnH2</th>
<th>JS1661 dnH1</th>
<th>JS1661 dnH1</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. testacea dnH1</td>
<td></td>
<td>27.5 %</td>
<td>41.2 %</td>
<td>28.3 %</td>
</tr>
<tr>
<td>N. testacea dnH2</td>
<td></td>
<td></td>
<td>42.8 %</td>
<td>54 %</td>
</tr>
<tr>
<td>JS1661 dnH1</td>
<td></td>
<td></td>
<td></td>
<td>40.5 %</td>
</tr>
</tbody>
</table>

Given their similarities to dnH1 and dnH2, it is also relevant to the question of their evolutionary origins to consider the function of the beta-lactamases. These enzymes were originally identified as a cause of bacterial resistance towards beta lactam antibiotics such as penicillin [56]. They have now been identified in diverse genera, ranging across Bacillus, Nocardia, Actinomycetes, Enterobacteriaceae and Pseudomonades [185, 271]. Beta-lactamases are divided into four classes based on their structure and function. Classes A, C and D all have a diagnostic serine in their active site which hydrolyses the lactam ring by forming an acyl intermediate. By contrast class B enzymes require a metal cofactor, usually mono- or divalent zinc, in their active site to catalyse the hydrolysis, and they are therefore referred to as metallo-beta-lactamases (MBLs) [185]. Based on differences in their sequences, the MBLs are further classified into the B1, B2 and B3 subclasses. The three-dimensional structures of representatives of all three of these subclasses have indicated monomeric and quaternary structures, except for the L1 and L2 enzymes in subclass B3, which are homo-tetrameric and homodimeric respectively [83, 185]. The A, C and D class beta lactamases have also generally been reported to be monomers (eg. NDM-1 of Bacillus), but some dimers (in the TEM plasmid RP4, Bacillus sp. and Enterobacteriaceae), tetramers (the AmpC beta-lactamase of Pseudomonas aeruginosa and Enterobacteriaceae) and a hexamer (L-ascorbate 6-
phosphate lactonase of *Ecoli, Klebsiella* and *Salmonella* sp.) have also been described [25, 37, 83, 86, 185, 260]. The dnH1 and dnH2 homologues are most similar to class B metallo-beta-lactamase enzymes, as it contains Zn. None of the metallo-beta-lactamases reported so far are hetero-hexamers, as observed in the DNAN demethylase of the JS1661 bacterium, so the latter may represent a new subclass of the class B enzymes.

Nor have any beta-lactamases so far been implicated in the degradation of nitroaromatic compounds or the hydrolysis of O-methyl groups on nitroaromatics as reported here for the DNAN demethylase of *Nocardia* sp. JS1661. However, four sequence-unrelated types of O-demethylase enzymes have been reported previously to catalyse O-demethylase reactions on methoxylated substrates. These are the Rieske non-heme iron oxygenases, cytochrome P-450 monoxygenases, tetrahydrofolate (THF)-dependent O-demethylases and atrazine hydrolases [73, 144, 176, 276].

The Rieske non-heme iron oxygenases and cytochrome P-450 monoxygenases are oxygen and NAD(P)H dependent enzymes. They have been identified in several bacterial genera, including *Rhodococcus, Streptomyces, Pseudomonas*, and *Comamonas* where they have been shown to catalyse the O-demethylation of vanillate, guaiacol and veratrole to protocatechuate (PCA) [42, 176, 196]. For example the vanillate-O-demethylase in *Rhodococcus jostii* RHA1 and *Pseudomonas* sp. HR199 is a two component system comprising both an oxygenase and a reductase encoded by the *vanA* and *vanB* genes respectively. Typically vanA oxygenases include Rieske non-heme iron oxygenases, mononuclear non heme iron oxygenases and cytochrome P450-like
enzymes, and vanB reductases include flavin mononucleotide, NAD and chloroplast ferredoxin-type iron-sulfur enzymes [32, 196].

Homology searches have identified several other genes in Pseudomonadaceae, with 35-52 % amino acid similarity to vanA and vanB of R. jostii RHA1. Vanillate-O-demethylases and cytochrome P450 enzymes from Streptomyces species [175, 176, 196, 197, 245] and Comamonas testosteroni BR6020 [42, 175, 176, 197] have also been shown to catalyse 3-O- or 4-O-demethylation of veratrate substrates (Figure 4-4).

Some other cytochrome P450 monoxygenases are well known to catalyse O-demethylation reactions in the presence of co-factors (NAD(P)H) and reductases (flavin proteins). Examples include a wheat enzyme which catalyses O-demethylation of the triazine ring on the herbicide prosulphuron and a CYP51 enzyme present in animals, plants, bacteria and fungi which catalyses transformation of lanosterol to other sterols [143, 144, 276].

One type of THF O-demethylase is a multi-component methyl transferase system which has been found in anaerobes such as Acetobacterium dehalogenans, Moorella thermoacetica and Acetobacterium woodii. These systems catalyse the transfer of a methyl group from a substrate (vanillate, syringate and guaiacol), first to a corrinoid protein by a methyl transferase I, and then to THF to form a methyl-THF compound by a methyl transferase II enzyme [19, 123, 124, 162, 170]. However in the aerobic Sphingomonas paucimobilis SYK-6 the pathway of methyl transfer appears to be a one step mechanism, which transfers the methyl group from the substrate (syringate) to THF directly to form methyl-THF by a simple syringate-O-demethylase enzyme encoded by the desA gene [124, 162]. This enzyme does not have any detectable sequence
similarities with the anaerobic O-demethylase systems above, but does have sequence similarities to various aminomethyl transferase, dimethylglycine dehydrogenase, sarcosine dehydrogenase and sarcosine oxidase alpha-subunit enzymes, suggesting that it might have evolved from such proteins [157, 160, 162]. Interestingly S. paucimobilis SYK-6 also has a LigM enzyme which catalyses a DesA type of O-demethylation of vanillate and 3-O-methylgallate to PCA and gallate intermediates, respectively in the presence of THF. This enzyme is 49 % (amino acid) similar to DesA [123, 124, 157, 162].

![Vanillic acid](image1.png) ![Veratrate](image2.png) ![Atratone](image3.png)

**Figure 4-4.** Some examples of methylated substrates for O-demethylase enzymes.

Triazines are nitrogen-containing heterocyclic rings which may have various of substituents and some of which are widely used as herbicides. Examples include atrazine (N-isopropyl), simazine (-Cl), ametryn, prometryn (alkyl) and atratone (O-CH₃) [16, 224, 226, 227, 253]. Several gram negative bacteria (Pseudomonas sp., Ralstonia sp., Agrobacterium sp. and Rhizobium sp.) can hydrolyse triazines with –Cl and –F substituents with an AtzA hydrolase enzyme, whereas several gram positive bacteria (Arthrobacter sp. and Nocardioides sp.) can hydrolyse a broad range of s-triazines with
a TrzN hydrolase enzyme [16, 224, 226, 227, 253]. Hydrolysis of atratone is of particular interest in the current context, because TrzN is used to catalyse an O-demethylase reaction, releasing methanol as a by-product. This is directly analogous to the reaction on DNAN catalysed by DNAN demethylase. However there are no sequence similarities between the two enzymes. TrzN is a metallo-hydrolase from amidohydrolase superfamily with a mononuclear Zn centre which is essential for enzyme activity [16, 224, 226, 227, 253].

The question then is why has evolution not utilized one or other of the four other enzyme groups (Rieske non-heme iron oxygenases, cytochrome P-450 monooxygenases, tetrahydrofolate (THF)-dependent O-demethylases and atrazine hydrolases) known to catalyse O-demethylation reactions for the hydrolysis of DNAN. I suggest the answer lies in the different reaction chemistries used by the different enzyme groups, and the unique requirements for the O-demethylation of nitroaromatics. The nitro groups on the benzene ring hold partial positive charges which resist direct substitution reactions. These properties may render O-demethylation energetically too difficult for the other groups of enzymes.

As previously noted, DNAN is a relatively recently introduced munitions ingredient which is now replacing much of the 2,4,6-trinitrotoluene (TNT) previously used. Like TNT, DNAN is a nitro-organic compound, storage or use of which can lead to contamination of soil and water, so decontamination strategies are needed to treat affected sites and water bodies and the demilitarization of obsolete munitions. The Nocardioides strain and/or DNAN-O-demethylase enzyme I have characterized may therefore have value as a biocatalytic decontamination agent, and it may also have value
in various biosensor applications for the detection of DNAN contamination. Below I consider the potential of such applications to detect (biosensors) and decontaminate DNAN spills and contaminated sites using various bioremediation processes [24].

Preliminary studies reported in chapter 3 revealed the ability of the *Nocardioides* sp JS1661 to biodegrade DNAN rapidly over a range of concentrations and conditions. Use of the strain for bioaugmentation may therefore be a feasible option for aerobic bioremediation of contaminated soil. Biostimulation may be less feasible in general because it appears that the ability to hydrolase DNAN is not widely distributed among soil organisms [192]. Unpublished preliminary studies by Prof. Jim Spain and Dr Gunjan Pandey indeed indicate that inoculation of microcosms with strain JS1661 can be a highly effective form of decontamination. Additionally, the DNAN demethylase enzyme itself could be suitable for free enzyme bioremediation, as it has reasonable kinetic properties for such applications, specifically a low *Km* (30.3 µM) and reasonable *kcat /KM* (63.55 µM⁻¹ sec⁻¹). Moreover there is the potential to further improve its catalytic properties and stability in the open environment by modern protein engineering and *in vitro* evolution techniques [222].

The DNAN demethylase gene/enzyme systems could also prove useful in transgenic phytoremediation strategies for the clean-up of DNAN contaminated soils [2]. There are already successful examples of transgenic phytoremediation of other nitrogenous organics such as atrazine and TNT [153, 224, 270]. The atrazine phytoremediation utilizes the atzA enzyme mentioned above which hydrolytically removes the chlorine from atrazine to form the non-toxic intermediate hydroxyatrazine [153, 224, 270]. A modified *atzA* gene, *p-atzA*, has been successfully transformed and
expressed in alfalfa, *Arabidopsis thaliana* and tobacco and these plants have then been shown to degrade atrazine in the soil to hydroxyatrazine. Similarly, the human cytochrome P450 enzymes CYP1A1 and CYP1A2 have been transformed and expressed in tobacco cells (*Nicotiana tabacum* L) to catalyse the *N*-dealkylation of atrazine [23].

A pentaerythritol tetranitrate reductase (PETN reductase) gene/enzyme system from *Enterobacter cloacae* PB2 that can degrade nitroaromatics such as glycerol trinitrate (GTN), pentaerythritol tetranitrate (PETN) and TNT has also been introduced into transgenic tobacco plants which were then found to be able to degrade GTN and TNT [2, 270]. Interestingly, TNT degradation by some bacteria and other plants with some endogenous ability to degrade the compound produce 2-amino-4,6-dinitrotoluene and 2,4-diamino-6-nitrotoluene intermediates which are actually more toxic than TNT [75]. However, the TNT intermediates produced by the tobacco plants expressing PETN reductase lead to much less inhibitory effect on growth than expected, albeit the intermediates generated in this case have yet to be identified [75, 168, 270].

Now that the DNAN gene is available, it may be worthwhile to transform and express it in appropriate plant material to treat DNAN contaminated soils by phytoremediation as above.

Enzymes are attracting increasing attention for their potential as ‘front end’ detection systems in biosensors for a variety of compounds in the medical, pharmaceutical, food and agricultural areas. They are becoming cheaper, portable and often more sensitive alternatives to traditional chromatographic tools (HPLC) for detecting and, increasingly, quantifying the compound of interest [9, 172, 265]. Depending on the detection methods, the signal can be converted to optical,
electrochemical or piezoelectric outputs [217]. In cholinesterase (ChE)-based biosensors for organophosphate and carbamate insecticides, the immobilised ChE gets inhibited by the presence of the pesticide which is measured by optical or electrochemical sensors [9, 172, 217]. Similarly tyrosine-based biosensors catalyse monophenols to O-quinone but are inhibited by the presence of atrazine and carbamate pesticides [9, 172]. Other biosensors utilize peroxidase, glutathione S-transferase, acid phosphatase and organophosphorus hydrolase reactions [9, 140, 172, 217]. The DNAN demethylase could be applied in a similar way for the detection of nitroaromatics in the environment. As noted above, the DNAN enzyme has a low $K_m$ and good $k_{cat}/K_m$. Moreover the corresponding metabolite (2,4-dinitrophenol) has a yellow colour which can be detected at low concentrations. Thus the enzyme may be a potential biosensor tool for detecting even quite low concentrations of DNAN in soil and water bodies.

Further work on the structure biology and reaction chemistry of the DNAN demethylase could provide the basis for protein engineering work to enhance the catalytic efficiencies and robustness of this enzyme for the various potential applications above. It might also be worthwhile to test and analyse the homologue from the Nocardioides testeca sp. against DNAN. Another clear priority for further fundamental and applied work will be to develop effective heterologous expression systems for the enzyme, given its hetero hexameric structure and the overlapping nature of the cognate genes.
References


20. Bernard-Vailhe MA, Besle JM, Dore J: Transformation of 14C-lignin-labeled cell walls of wheat by Syntrophococcus sucromutans, Eubacterium oxidoreducens, and


64. Ebert S, Fischer P, Knackmuss HJ: Converging catabolism of 2,4,6-trinitrophenol (picric acid) and 2,4-dinitrophenol by Nocardioides simplex FJ2-1A. Biodegradation 2001, 12(5):367-376.


106. Hoyt N, Brunell M, Kroeck K, Hable M, Crouse L, O'Neill A, Bannon DI: Biomarkers of oral exposure to 3-nitro-1,2,4-triazol-5-one (NTO) and 2,4-dinitroanisole (DNAN) in blood and urine of rhesus macaques (Macaca mulatta). Biomarkers 2013, 18(7):587-594.


132. Krumholz LR, Bryant MP: *Eubacterium oxidoreducens* sp. nov. requiring H₂ or formate to degrade gallate, pyrogallol, phloroglucinol and quercetin. Archives of Microbiology, 144(1):8-14.


169. Nagata Y, Ohtomo R, Miyauchi K, Fukuda M, Yano K, Takagi M: Cloning and sequencing of a 2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase gene


Appendices

1. Permission for Figure 1-3 and 1-5
2. Permission for Figure 1-7
3. RAST function annotation of SG61-1L genome and subsystem coverage
4. Alignment of dehydrogenase genes used for IQtree
5. Some of the gene names and corresponding accession numbers for genes included in the phylogenetic analysis
6. Biodegradation of 2,4-DNP by *Nocardioides* sp. JS1661 pre-grown on acetate and 2,4-DNP.
7. DNA sequence of the DNAN hydrolase genes
Appendix 1. Permission/license is granted for figure 1-3 and 1-5

Title: The catalytic valorization of lignin for the production of renewable chemicals


Publication: Chemical Reviews
Publisher: American Chemical Society
Date: Jun 1, 2011
Copyright © 2011, American Chemical Society

• This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:
• Permission is granted for your request in both print and electronic formats, and translations.
• If figures and/or tables were requested, they may be adapted or used in part.
• Please print this page for your records and send a copy of it to your publisher/graduate school.

• Figure 1-3 and 1-5 were adapted with permission from (Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, Davison BH, Dixon RA, Gilna P, Keller M, Langan P, Naskar AK, Saddler JN, Tschaplinski TJ, Tuskan GA, Wyman CE: Lignin valorization: improving lignin processing in the biorefinery. Science 2014, 344(6185):709), American Chemical Society.

• One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

• If credit is given to another source for the material you requested, permission must be obtained from that source.
Appendix 2. Permission/license is granted for figure 1-7

Title: Identification of DypB from *Rhodococcus jostii* RHA1 as a Lignin Peroxidase

Author: Ahmad M, Roberts JN, Hardiman EM, Singh R, Eltis LD, Bugg TDH

Publication: Biochemistry

Publisher: American Chemical Society

Date: Jun 1, 2011

Copyright © 2011, American Chemical Society

- This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:
- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Figure 1-7 adapted with permission from (Ahmad M, Roberts JN, Hardiman EM, Singh R, Eltis LD, Bugg TDH: Identification of DypB from *Rhodococcus jostii* RHA1 as a lignin peroxidase. Biochemistry 2011, 50(23):5096-5107), Copyright (2011) American Chemical Society.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.
- If credit is given to another source for the material you requested, permission must be obtained from that source.
Appendix 3. RAST annotation of SG61-1L genome and subsystem coverage showing the SG61-1L capabilities.
<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>PDB ID</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>146275658</td>
<td>Naro/1-269</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334343146</td>
<td>Sphch/1-268 no co</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>87199139</td>
<td>/1-266 no comment</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>146275675</td>
<td>Naro/1-252</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334343486</td>
<td>Sphch/1-273</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>87198897</td>
<td>/1-259</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>146275698</td>
<td>Naro/1-254</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342862</td>
<td>Sphch/1-257</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>87200486</td>
<td>/1-256</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342592</td>
<td>Sphch/1-246</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342590</td>
<td>Sphch/1-248</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342994</td>
<td>Sphch/1-245</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342992</td>
<td>Sphch/1-245</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342962</td>
<td>Sphch/1-244</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342592</td>
<td>Sphch/1-248</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>146275658</td>
<td>Naro/1-269</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334343146</td>
<td>Sphch/1-268 no co</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>87199139</td>
<td>/1-266 no comment</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>146275675</td>
<td>Naro/1-252</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334343486</td>
<td>Sphch/1-273</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>87198897</td>
<td>/1-259</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>146275698</td>
<td>Naro/1-254</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342862</td>
<td>Sphch/1-257</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>87200486</td>
<td>/1-256</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342592</td>
<td>Sphch/1-246</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342590</td>
<td>Sphch/1-248</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342994</td>
<td>Sphch/1-245</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342992</td>
<td>Sphch/1-245</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342962</td>
<td>Sphch/1-244</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342592</td>
<td>Sphch/1-248</td>
<td></td>
</tr>
</tbody>
</table>

**Sequence Alignment**

```
<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>PDB ID</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>146275658</td>
<td>Naro/1-269</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334343146</td>
<td>Sphch/1-268 no co</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>87199139</td>
<td>/1-266 no comment</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>146275675</td>
<td>Naro/1-252</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334343486</td>
<td>Sphch/1-273</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>87198897</td>
<td>/1-259</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>146275698</td>
<td>Naro/1-254</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342862</td>
<td>Sphch/1-257</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>87200486</td>
<td>/1-256</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342592</td>
<td>Sphch/1-246</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342590</td>
<td>Sphch/1-248</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342994</td>
<td>Sphch/1-245</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342992</td>
<td>Sphch/1-245</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342962</td>
<td>Sphch/1-244</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>334342592</td>
<td>Sphch/1-248</td>
<td></td>
</tr>
</tbody>
</table>
```
Appendix 5. Some of the gene names and corresponding accession numbers for genes included in the phylogenetic analysis

<table>
<thead>
<tr>
<th>Gene name</th>
<th>gene identification (GI) number or accession ID</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADH_Syanoikuyae</td>
<td>4BMV (pdb ID)</td>
<td>[152]</td>
</tr>
<tr>
<td>Ga5DH_Ssuis</td>
<td>146318879</td>
<td>[283]</td>
</tr>
<tr>
<td>phaB_Synechocystis</td>
<td>16330475</td>
<td>[249]</td>
</tr>
<tr>
<td>bdh_Pputida</td>
<td>167033792</td>
<td>[72, 183]</td>
</tr>
<tr>
<td>2,5-DDOL_Spaucimobilis</td>
<td>1708836</td>
<td>[169]</td>
</tr>
<tr>
<td>kduD_Ecoli</td>
<td>1730032</td>
<td>[258]</td>
</tr>
<tr>
<td>ydfG_Ecoli</td>
<td>2506152</td>
<td>[79]</td>
</tr>
<tr>
<td>LVR_Caquaticum</td>
<td>31615388</td>
<td>[235]</td>
</tr>
<tr>
<td>(R)-ADH_Lkefiri</td>
<td>33112056</td>
<td>[43]</td>
</tr>
<tr>
<td>fabG_Ecoli</td>
<td>388477173</td>
<td>[195]</td>
</tr>
<tr>
<td>badH_Rpalustris</td>
<td>39647578</td>
<td>[187]</td>
</tr>
<tr>
<td>ADH_Ralstonia</td>
<td>512125550</td>
<td>[152]</td>
</tr>
<tr>
<td>phaB_Burkholderia</td>
<td>518742570</td>
<td>[126]</td>
</tr>
<tr>
<td>CPNA_Comamonas</td>
<td>62286565</td>
<td>[111]</td>
</tr>
<tr>
<td>Gluc-DH_Bsubtilis</td>
<td>649014081</td>
<td>[97]</td>
</tr>
<tr>
<td>steA_Ctestosteroni</td>
<td>729773326</td>
<td>[103]</td>
</tr>
<tr>
<td>(S)-PED_Aaro</td>
<td>81821161</td>
<td>[63]</td>
</tr>
<tr>
<td>bdh1_Rpickettii</td>
<td>84570594</td>
<td>[247]</td>
</tr>
<tr>
<td>bdh2_Rpickettii</td>
<td>84570596</td>
<td>[247]</td>
</tr>
<tr>
<td>SG61_1203</td>
<td>SZ64_06710</td>
<td></td>
</tr>
<tr>
<td>SG61_1215</td>
<td>SZ64_06655</td>
<td></td>
</tr>
<tr>
<td>SG61_1243</td>
<td>SZ64_06525</td>
<td></td>
</tr>
<tr>
<td>SG61_1419</td>
<td>SZ64_05615</td>
<td></td>
</tr>
<tr>
<td>SG61_1498</td>
<td>SZ64_05225</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_2016</td>
<td>SZ64_08220</td>
<td></td>
</tr>
<tr>
<td>SG61_2549</td>
<td>SZ64_00035</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_2550</td>
<td>SZ64_00030</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_2583</td>
<td>SZ64_09815</td>
<td></td>
</tr>
<tr>
<td>SG61_2705</td>
<td>SZ64_01645</td>
<td>this work</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>SG61_2706</td>
<td>SZ64_01650</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_2863</td>
<td>SZ64_11380</td>
<td></td>
</tr>
<tr>
<td>SG61_3175</td>
<td>SZ64_16025</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_3191</td>
<td>SZ64_15940</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_3329</td>
<td>SZ64_12360</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_3344</td>
<td>SZ64_12435</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_354</td>
<td>SZ64_16980</td>
<td></td>
</tr>
<tr>
<td>SG61_3726</td>
<td>SZ64_14315</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_3730</td>
<td>SZ64_14335</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_417</td>
<td>SZ64_16655</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_474</td>
<td>SZ64_16365</td>
<td></td>
</tr>
<tr>
<td>SG61_504</td>
<td>SZ64_17190</td>
<td></td>
</tr>
<tr>
<td>SG61_609</td>
<td>SZ64_02580</td>
<td></td>
</tr>
<tr>
<td>SG61_724</td>
<td>SZ64_15290</td>
<td>this work</td>
</tr>
<tr>
<td>SG61_782</td>
<td>SZ64_15575</td>
<td></td>
</tr>
<tr>
<td>SG61_907</td>
<td>SZ64_14565</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 6. Biodegradation of 2,4-DNP by *Nocardioides* sp. JS1661 pre-grown on acetate and 2,4-DNP.
Appendix 7. DNA sequence of the DNAN hydrolase genes indicating ribosomal binding site (RBS), putative start/stop codons and the positions of the N-terminal peptide sequences. The proteins designated “upper band” in the text corresponds to nucleotides 775-1762 and the “lower band” corresponds to 1759-2715. The matched peptides for the upper and lower bands cover 37 % (122/329 AA’s) and 55 % (178/319 AA’s) of the proteins, respectively. Covered amino acids (shown in red) include 113 from peptide sequencing and 6 from N-terminal sequencing (yellow highlighted) for the upper band and 172 from peptide sequencing and 6 from N-terminal sequencing for the lower band. Based on the N-terminal threonine, the first nine amino acids (marked in blue), seem to have been post-translationally removed by an unknown mechanism from the upper and the first amino acid (methionine, encoded by GTG: marked in blue), seems to have been post-translationally removed by the universal N-terminal methionine excision mechanism from the lower band.