Biocatalytic Conversion of Vanillin to 3-Carboxy Muconate

Aaron Gosling

Submitted for Doctorate of Philosophy

2009
Abstract

The organic chemical industry permeates our daily lives, giving us products such as pharmaceuticals, plastics, adhesives and paints. 90 to 95% by mass of all organic chemicals produced are derived from petroleum oil or natural gas [1]. To move towards a more sustainable chemical industry, processes which transform renewable feed stocks need to be discovered and developed.

Plant biomass is renewable, and typically contains 30% (w/w) lignin [2]. Lignin can be dismantled to yield a smaller and simpler compound, vanillin [3]. This project investigated the construction of a synthetic metabolic pathway to transform vanillin into the novel polymer building block 3-carboxy muconate (3CM).

The metabolic pathway which performs this transformation consists of three enzymes; 4-hydroxy benzaldehyde dehydrogenase (HBD), vanillate monooxygenase (VMO) and protocatechuate 3,4-dioxygenase (P34O). These enzymes were cloned from Acinetobacter baylyi, expressed in Escherichia coli, and purified. Characteristics pertinent to the behaviour of these enzymes acting in a pathway, such as kinetics, had been reported for only P34O. This information for HBD and VMO was gained by characterising these enzymes for kinetic behaviour, substrate specificity and stability.

The genes for all three enzymes were then combined within a single recombinant E. coli host. This whole cell biocatalyst transformed 1 mM vanillin into 1 mM 3CM. Evidence was gathered to show the rate limiting factor for the conversion was the expression level of VMO. In vitro characterisation had identified that this enzyme had markedly lower intrinsic activity than the other two enzymes, and required higher relative expression.

The usefulness of 3CM as a polymer building block was explored. The butadiene system of 3CM was chemically isomerised during conversion to a trimethyl ester form. This trimethyl ester was found to copolymerise with styrene. The composition of the copolymer could be varied by varying the concentrations of monomers in the feed.
This project increased the body of knowledge of the enzymology of the vanillin to 3CM pathway, demonstrated that 3CM can be made by biocatalytic transformation of the renewable compound vanillin, and that 3CM is a useful polymer building block.
Acknowledgements

I would like to acknowledge my supervisors, Melissa, Tony and Mike. You all gave me different things, and all of those things were very important to my education, and to the submission of this thesis. Melissa, you were important for praise and encouragement. Tony, the writing in this document would not be nearly as fluent if not for your input. Mike, the idea for this project was yours, and your continued involvement from the other side of the world meant more to me than I can say here.

I would also like to acknowledge the technical assistance of Jane Fowler. I would never have gotten the synthetic pathway into one host without you, and I also found you really fun to work with.

I would also like to acknowledge Angela Ziebell for leading the way, showing me it is actually possible to complete a PhD and mostly for giving me an excellent friend to bounce ideas off.

I would like to thank the former Biocatalysis group of Molecular and Health Technologies, CSIRO. There has been turmoil, and change over the past 3 and a half years, but let me see if I can list all of you who were important to me. Primarily, Geoff, Linda and Andrew, your friendship, advice and help has been invaluable. May CSIRO finally come to recognise the core of ability it has in you! To Adam, thank you for the chromatography wisdom, the jokes, and the shared enemy. To Jess, Tara, Lauren, Natalie, Lindus, Lia, Mohammed, Glenn, Usha, Jacqui, Brooke, Tom, Asoka, Carolyn, Mira and Carol, you made these labs like home to me, and I hope the tides that have swept you away have also taken you towards your goals.

To Mike O’Shea, Florain Graichen, Ben Leita and Andrew Warden, thank you for sharing your chemistry, and plenty of laughs, with me.

To Jo Cosgriff, Roger Mulder, Carl Braybrook and Ming Cheng, thank you for maintaining the instruments I relied on! To Jo and Roger, thank you thank you for your patience in teaching me a little of what you know.
To Molecular and Health Technologies, CSIRO at large, thank you for the use of well equipped labs, and some amazing lunchtime conversation.

I would like acknowledge my family. Mum, Dad and Annie, Jan and Ian, Sarah and Nat, your continued interest in my work has kept me going on more than one occasion. This work would not have been completed if you didn't keep asking about it.

Finally and most importantly, Kirsti, Bella and Jack, thank you for your patience with my absence, your love, cuddles, interest and support. Everything I do, I do with you foremost in my mind.
Declaration

This thesis does not contain material which has been accepted for the award of any other degree or diploma.

To the best of my knowledge, this thesis contains no material previously published or written by another person.

I was given technical assistance with the cloning described in Chapter 6 by Jane Fowler, and this contribution is acknowledged gratefully. The conception of that work, and the design of experiments was mine.

Aaron Gosling
Table of Contents

Title page .......................................................................................................................... i
Abstract ........................................................................................................................... ii
Acknowledgements ......................................................................................................... iv
Declaration ....................................................................................................................... vi
Table of Contents ........................................................................................................... vii
Table of Figures .............................................................................................................. xiv
Table of Tables ............................................................................................................... xvii
List of Abbreviations ..................................................................................................... xix
1 Introduction .................................................................................................................. 1

1.1 The Context of the Project ..................................................................................... 1
  1.1.1 The need for renewable feed stocks in chemical manufacture .................... 1
  1.1.2 Vanillin as a renewable chemical platform ..................................................... 1

1.2 Biocatalysis ............................................................................................................. 3
  1.2.1 Green chemistry .............................................................................................. 3
  1.2.2 Whole cell multi-enzyme biocatalysts ............................................................... 4

1.3 The Vanillin to 3-Carboxy Muconate Pathway ..................................................... 6
  1.3.1 Metabolism of vanillin .................................................................................... 6
  1.3.2 The choice of 3CM as the target product ......................................................... 8
  1.3.3 The catabolic pathway containing the vanillin to 3CM reactions .................... 8
  1.3.4 4-Hydroxy benzaldehyde dehydrogenase ....................................................... 11
  1.3.5 Vanillate monooxygenase ............................................................................... 13
  1.3.6 Protocatechuate 3,4-dioxygenase .................................................................. 16

1.4 The Chemistry of 3-Carboxy Muconate ............................................................... 20
  1.4.1 Prior synthesis of 3CM .................................................................................. 20
  1.4.2 The problem with 3CM .................................................................................. 21
  1.4.3 Renewable monomers and polymers ............................................................... 22
  1.4.4 3-Carboxy muconate’s potential as a polymer building block ....................... 24
1.5 Aims, scope and approach of the project

2 Materials and Methods

2.1 The Genetic Source of Enzymes

2.2 General Materials and Methods for Cloning

2.2.1 Polymerase chain reactions

2.2.2 Restriction cleavage, ligation and transformation

2.2.3 Microbial growth for cloning experiments

2.2.4 DNA analysis

2.3 General Materials and Methods for Production of Cell Lysates

2.3.1 Recombinant protein expression

2.3.2 Cell lysis

2.4 General Materials and Methods for Protein Purification

2.4.1 Chromatography

2.5 General Material and Methods for Protein Analysis

2.5.1 SDS-PAGE

2.5.2 Protein concentration assays

2.6 General Materials and Methods for Enzyme Analysis

2.6.1 Substrates for enzymatic reactions

2.6.2 Continuous steady state enzyme activity assays

2.6.3 Discontinuous steady state enzyme activity assays

2.6.4 Estimating Michaelis – Menten constants

2.6.5 Substrate specificity assays

2.7 General Material and Methods for Chemical Analysis

3 4-Hydroxy Benzaldehyde Dehydrogenase

3.1 Introduction
6.1.1 Literature review and background for biotransformation of vanillin to 3CM
6.1.2 Aims and approach for the biotransformation of vanillin to 3CM

6.2 Material and Methods ................................................................. 114
6.2.1 Cloning to create the synthetic pathway............................... 114
6.2.2 Microbial growth and recombinant protein expression to generate biocatalyst ................................................................. 117
6.2.3 Whole cell biotransformations ............................................. 119
6.2.4 LC MS analyses of biotransformations ............................... 119

6.3 Results and Discussion ............................................................... 120
6.3.1 LC MS analyses method development ............................... 120
6.3.2 Toxicity of expressing the recombinant enzymes .......... 122
6.3.3 Performance of individual components of the synthetic pathway 124
6.3.4 Time course of the vanillin to 3CM pathway ............... 130
6.3.5 The amount of each enzyme present ............................. 133

6.4 Summary and Conclusions ..................................................... 136
6.5 Further work ............................................................................ 137

7 Polymerisation of 3-Carboxy Muconate ...................................... 139
7.1 Introduction ............................................................................. 139
7.1.1 Radically initiated polymerisation reactions in solution .... 139
7.1.2 UV initiated radical reaction in the solid state .................. 143
7.1.3 Aims and approach for polymerisation of 3CM .......... 144

7.2 Material and Methods ................................................................. 145
7.2.1 Synthesis of 3CM ................................................................. 145
7.2.2 Synthesis of trimethyl 3-carboxy E,E muconate .......... 146
7.2.3 Radical reaction of 3CM-Me₃ in solution ....................... 147
9.6  $^{13}$C-NMR Spectra of Copoly Styrene and 3CM-Me$_3$ .......................200
9.7  Comparison of the FTIR Spectra of Copoly Styrene 3CM-Me$_3$ and Homopolystyrene .................................................................201
9.8  Average Molecular Weight of Copoly 3CM-Styrene and Dimethyl Styrene with Varied Comonomer Composition............................................202
9.9  Spectral Characteristics of Compounds Assessed for UV Reactivity203
9.10 Change in the FTIR Spectrum of Dinaphthyl Methyl Ammonium E,E-Hexadienedioate Due to UV Exposure........................................208
9.11 No Change in the FTIR Spectrum of Dibenzyl Ammonium E,E-Hexadienedioate Due to UV Exposure.........................................................209

10  References .............................................................................................................210
### Table of Figures

| Figure 1: | Enzymatic reactions employing vanillin as a substrate | 7 |
| Figure 2: | Structure of 3-carboxy-muconate | 8 |
| Figure 3: | Biochemical pathway for the catabolism of ferulate | 10 |
| Figure 4: | The reaction catalysed by HBD | 11 |
| Figure 5: | The reaction catalysed by VMO | 13 |
| Figure 6: | The reaction catalysed by P34O | 16 |
| Figure 7: | Oxidative aromatic ring cleavage of vanillin to yield 3-carboxy muconate-1-methyl ester | 20 |
| Figure 8: | Mechanism of spontaneous isomerisation of 3CM | 22 |
| Figure 9: | Potential polymerisation reactions of 3CM | 25 |
| Figure 10: | Structures of 3CM and analogues previously polymerised by UV exposure in the solid state | 26 |
| Figure 11: | Map of the recombinant plasmid pET-HBD | 42 |
| Figure 12: | SDS-PAGE gel showing the expression and IMAC purification of HBD | 46 |
| Figure 13: | Effect of ionic conditions on the activity of HBD | 49 |
| Figure 14: | Stability of HBD after incubation at 37 and 45 °C | 51 |
| Figure 15: | Non substrates for HBD from *A. baylyi* | 52 |
| Figure 16: | Substrates for HBD from *A. baylyi* | 53 |
| Figure 17: | Proposed catalytic cycle for VMO | 60 |
| Figure 18: | Map of the recombinant plasmid pET-VMO | 64 |
| Figure 19: | Exponential decay function used in regression analysis of the stabilisation of VMO | 65 |
| Figure 20: | Half life’s calculated for VMO incubated on ice with various additives | 70 |
| Figure 21: | SDS-PAGE analysis of VMO purification | 76 |
Figure 22: Structure of the VanA / VanB operon from A. baylyi ..................... 78
Figure 23: Inhibition and activation of VMO .................................................. 81
Figure 24: VMO activity on vanillate, 3,4-dihydroxy benzoate and syringate 83
Figure 25: Map of the recombinant plasmid pET-P34O .............................. 91
Figure 26: The difference between P34O levels in whole cells and soluble lysate after induction at 37°C ................................................................. 96
Figure 27: Effect of temperature and IPTG concentration on recombinant expression of soluble active P34O .............................................................. 98
Figure 28: Silver stained SDS PAGE gel showing the progress of the purification of P34O ................................................................. 102
Figure 29: Regression fit of Michaelis – Menten equation to experimentally measured rates of P34O .......................................................... 106
Figure 30: The cloning strategy for constructing pACYC-HBD ..................... 115
Figure 31: The cloning strategy for constructing pET-VMO-P34O .............. 117
Figure 32: Example chromatogram of LC MS analysis of the synthetic pathway metabolites ................................................................. 121
Figure 33: The impact of recombinant protein expression on the final optical density of recombinant hosts ......................................................... 123
Figure 34: Metabolites detected in the reactions containing biocatalysts with the vanillin to 3CM pathway after overnight incubation with various substrates ................................................................. 130
Figure 35: Time course of vanillin to 3CM biotransformation ..................... 132
Figure 36: SDS-PAGE analysis of E. coli HBD-VMO-P34O before and after induction to express the vanillin to 3CM pathway enzymes ...................... 134
Figure 37: Some of the reactions possible during initiation of radical polymerisation ................................................................. 140
Figure 38: Some of the reactions possible during propagation of radical polymerisation ................................................................. 141
Figure 39: Some of the reactions possible during termination of radical polymerisation.

Figure 40: Two step synthesis of 3CM.

Figure 41: Esterification of 3CM-Na3.

Figure 42: Reaction vessel used to study radical reactions of 3CM in solution.

Figure 43: Structures of compounds used in UV reactivity experiments.

Figure 44: UV reactor used in solid state polymerisation studies.

Figure 45: The reactivity towards radicals displayed by the isomeric series maleic anhydride, fumarate esters, and maleate esters.

Figure 46: Structure of compounds used in solution free radical experiments.

Figure 47: 1H-NMR of homostyrene and copoly styrene – 3CM containing 33 mol% 3CM in the reactants.

Figure 48: 1, 2 and 1, 4 addition reactions using a 3CM-Me3 example.

Figure 49: The composition of 3CM-Me3 and dimethyl fumarate styrene copolymers as a function of the concentration of comonomers in feed.

Figure 50: Comparison of planarity of structural models of 3CM and E,E hexadienedioate.

Figure 51: Non-planar citrate molecules packing in close proximity in crystals of brucinium dihydrogen citrate trihydrate.

Figure 52: Convergent pathway strategy for broadening the feed stocks for 3CM production.
## Table of Tables

Table 1: Dehydrogenases known to oxidise vanillin to vanillate ................... 12
Table 2: Enzymes reported to convert vanillate to 3,4-dihydroxy benzoate . 15
Table 3: Summary of some kinetic properties of P34O from a variety of organisms ................................................................. 17
Table 4: Oligonucleotides used for PCR amplification .............................. 30
Table 5: Oligonucleotides used in DNA sequencing .................................. 33
Table 6: Summary of a typical IMAC purification of 4-hydroxy benzaldehyde dehydrogenase ............................................................... 47
Table 7: Apparent kinetic constants $K_{app}^m$ and $k_{app}^{cat}$ for 4-hydroxy benzaldehyde dehydrogenase....................................................... 55
Table 8: Results from one preparative VMO purification ........................... 75
Table 9: Apparent kinetic constants $K_{app}^m$ and $V_{app}^{max}$ for vanillate monooxygenase ............................................................................. 85
Table 10: Data for a typical purification process for P34O ......................... 101
Table 11: Comparison of kinetic constants estimated here and in literature studies 104
Table 12: Strains containing two plasmids to assess the performance of the enzymes of the vanillin to 3CM pathway in *E. coli* ........................................ 118
Table 13: Performance of LC MS analyses for detecting vanillin to 3CM pathway metabolites ............................................................................. 120
Table 14: The mean concentrations of vanillyl alcohol and vanillin measured after biotransformation of 1 mM vanillin with strains possessing no recombinant vanillin oxidising activity ............................................ 126
Table 15: The mean concentrations of vanillyl alcohol, vanillin and vanillate measured after biotransformation of 1 mM vanillin with strains recombinant HBD 127
Table 16: The mean concentrations of vanillyl alcohol, vanillin, vanillate and 3,4-dihydroxy benzoate measured after biotransformation of 1mM substrate with strains possessing recombinant VMO ..............................................128

Table 17: The mean concentrations of vanillate, 3,4-dihydroxy benzoate and 3CM measured after biotransformation of 1 mM substrate with strains possessing recombinant P34O .................................................................129

Table 18: Comparison of the spectra of copolystyrene and 3CM-Me₃ with homopolystyrene ........................................................................................................158

Table 19: Comparison of the spectral characteristics of dibenzyl ammonium (E,E) - muconate (A₂-HDA) and dinaphthylammonium (E,E) - muconate (I₂-HDA) measured here and reported previously .............................................................165

Table 20: Change in FTIR signals from positive and negative control salts due to UV exposure .........................................................................................................167
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3CM</td>
<td>3-carboxy muconate</td>
</tr>
<tr>
<td>3CM-Me₃</td>
<td>Trimethyl-3-carboxy muconate</td>
</tr>
<tr>
<td>A280</td>
<td>Ultra Violet absorbance at 280 nm</td>
</tr>
<tr>
<td>AIBN</td>
<td>Azo iso butyl nitrile</td>
</tr>
<tr>
<td>AmSO₄</td>
<td>Ammonium sulphate</td>
</tr>
<tr>
<td>ATCC</td>
<td>American Type Culture Collection</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>CoA</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>CoA-SH</td>
<td>Thiol form of coenzyme A</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribose nucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreotol</td>
</tr>
<tr>
<td>ε</td>
<td>Molar extinction coefficient</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylene diamine tetraacetic</td>
</tr>
<tr>
<td>FAD</td>
<td>Flavin adenine dinucleotide</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier Transform Infra Red</td>
</tr>
<tr>
<td>GC-MS</td>
<td>Gas Chromatography – Mass Spectrometry</td>
</tr>
<tr>
<td>GPC</td>
<td>Gel permeation chromatography</td>
</tr>
<tr>
<td>HBD</td>
<td>4-hydroxy benzaldehyde dehydrogenase</td>
</tr>
<tr>
<td>HDA</td>
<td>((E,E)) – hexadienedioic acid</td>
</tr>
<tr>
<td>HIC</td>
<td>Hydrophobic interaction chromatography</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>IEF</td>
<td>Isoelectric focus electrophoresis</td>
</tr>
<tr>
<td>IEX</td>
<td>Ion exchange chromatography</td>
</tr>
<tr>
<td>IMAC</td>
<td>Immobilised metal ion chromatography</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl thio galactoside</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>$k_{\text{app}}_{\text{cat}}$</td>
<td>Apparent catalytic constant</td>
</tr>
<tr>
<td>$K_{\text{app}}_{\text{m}}$</td>
<td>Apparent Michaelis constant</td>
</tr>
<tr>
<td>$k_{\text{cat}}$</td>
<td>Catalytic constant</td>
</tr>
<tr>
<td>$K_{m}$</td>
<td>Michaelis constant</td>
</tr>
<tr>
<td>LB</td>
<td>Luria broth</td>
</tr>
<tr>
<td>LC MS</td>
<td>Liquid chromatography – mass spectroscopy</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass to charge ratio</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>NAD(P)+</td>
<td>Oxidised nicotinamide adenine dinucleotide (phosphate)</td>
</tr>
<tr>
<td>NAD(P)H</td>
<td>Reduced nicotinamide adenine dinucleotide (phosphate)</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance spectroscopy</td>
</tr>
<tr>
<td>OD$_{600}$</td>
<td>Optical density at 600 nm</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>P34O</td>
<td>Protocatechuate 3,4-dioxygenase</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PMSF</td>
<td>Phenyl methane sulphonyl fluoride</td>
</tr>
<tr>
<td>RBS</td>
<td>Ribosome binding site</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate polyacrylamide electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>Tris-HCl</td>
<td>2-Amino-2-(hydroxymethyl)-1,3-propanediol hydrochloride</td>
</tr>
<tr>
<td>UF</td>
<td>Ultrafiltration</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra violet</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>$V_{\text{max}}$</td>
<td>Maximum velocity of enzyme reaction at constant enzyme concentration</td>
</tr>
<tr>
<td>VMO</td>
<td>Vanillate monooxygenase</td>
</tr>
<tr>
<td>v/v</td>
<td>Volume / volume</td>
</tr>
<tr>
<td>w/w</td>
<td>Weight / weight</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 The Context of the Project

1.1.1 The need for renewable feed stocks in chemical manufacture

The organic chemicals which give us materials, fuels and organic chemicals are not currently produced sustainably. 90 to 95% by mass of all organic chemicals produced in 2004 were derived from petroleum oil or natural gas [1]. Fossil fuels are not a renewable resource, and their usefulness is finite [4-6]. This was poignantly expressed by Sheik Ahmad Zaki Yamani, Saudi Arabia's former oil minister, who was quoted as saying ‘The Stone Age came to an end not for a lack of stones, and the oil age will end, but not for a lack of oil’ [7].

To move towards a sustainable chemical manufacturing industry, processes which use renewable feed stocks need to be discovered and developed. This is currently best exemplified in the enormous effort to manufacture liquid fuels from carbohydrates (see [8] for review).

1.1.2 Vanillin as a renewable chemical platform

Plant biomass is often put forward as at least a partial replacement for petroleum feed stocks, as it is renewable, and the amount of this material is massive [9, 10]. It has been estimated that 18.7 billion tonnes of plant biomass was harvested in 2000 [11].

The composition of plant biomass varies depending on the plant, but some generalisations can be made. Most of plant biomass material exists as polymeric material. The three major components of plant biomass are the carbohydrate based polymers cellulose and hemicellulose, and the aromatic based polymer lignin. Depending on the plant, lignin typically constitutes 25 to 30% (weight / weight (w/w)) of the dry mass [2, 12], and is the second most abundant form of plant biomass behind cellulose [13, 14]. It is a structurally complex polymer, synthesised from a number of different phenyl proprionoid monomers, joined in a number of ways, producing a highly branched and heterogenous structure [14-16].
The lignins currently produced at the largest volumes in industry are by-products from paper making [17], and are known as lignosulphates or kraft lignin, depending on the process which produced them. It is apparent that any future commercial scale conversion of plant biomass carbohydrates to liquid fuels will also produce a high volume of lignin by-product [18]. Although there are applications for this lignin waste, greater value could be found if it was depolymerised to release useful aromatic compounds [19]. These compounds are currently sourced directly from crude oil, or synthesised from ethylene produced from crude oil [1]. Vanillin (4-hydroxy-3-methoxy benzaldehyde) has been shown to be a major product of lignin depolymerisation [2, 20, 21], and is already commercially produced renewably from kraft lignin [3].

In addition to being renewable, vanillin is also a useful and valuable chemical commodity in its own right [22]. It finds application in flavour and fragrance industries, as well as being used as a synthon for pharmaceuticals [23].

In 2001, 12,000 tons of vanillin was produced worldwide [24]. In the same year, a yearly production capacity of $5 \times 10^5$ tons of kraft lignin was reported [17]. Given typical yields of around 10% (w/w) for processes which oxidise kraft lignin in alkaline solution to produce vanillin [25], it can be seen that there is potential for producing four times more vanillin from the kraft lignin waste generated in papermaking alone. This example is given solely to illustrate the feasibility of producing renewable vanillin at large scale, and it is not suggested that this route of producing renewable vanillin is optimal. There are many problems with the production of vanillin from kraft lignin. The production of kraft lignin uses chlorine, and there have been many environmental concerns raised about the organochlorines produced as wastes (e.g. [26]). Similarly, the alkaline oxidation of kraft lignin to make vanillin generates a high volume of waste [23]. However, it is clear that there is enough lignin to theoretically increase the amount of renewable vanillin produced. It is hoped that the development of plant biomass to liquid fuel technologies will encourage development of “cleaner” processes for manufacturing renewable vanillin, which could be used as a platform starting material for the synthesis of other molecules.
1.2 Biocatalysis

1.2.1 Green chemistry

In addition to the challenges of discovering and developing processes in chemical manufacturing that use renewable starting material, the chemical manufacturing industry, like all industries, faces challenges in reducing the waste it produces (reviewed in [27]). Many traditional chemical manufacturing processes involve high temperatures and pressures, which waste energy, and generate more mass of material waste than mass of product.

The economic and environmental benefits of reducing the energy consumption and waste production in chemical manufacture have been long recognised [28, 29]. These benefits were implicit in the Twelves Principles of Green Chemistry [30], which ISI Web of Knowledge (on 2/12/08) reported to have been cited 951 times since its publishing in 1998. These principles give goals for chemists designing new processes, and can be distilled down to the aims of utilising renewable starting material, reducing wastes and avoiding the use of toxic and / or hazardous substances.

An area of technology that meets many of these goals by its nature is biocatalysis [31, 32]. Biocatalysis is the use of biological entities to perform chemical transformations. These entities are mostly enzymes, and can be used either inside or outside cells. Biocatalysts are most often used at ambient temperatures, atmospheric pressures, and in aqueous solution. This means they usually require less energy input than high temperature, high pressure chemical processes. They are non-toxic, unlike most metallic catalysts, and do not pose the risk of explosion that highly reactive reagents do. Biocatalysts can also catalyse multiple reactions sequentially, removing the need to purify intermediates, and reducing the number of steps involved in chemical synthesis [31].

It should be noted that biocatalysis is not inherently good at meeting all of the goals of Green Chemistry. There are examples of biocatalytic processes using renewable starting materials being found to be less “clean” than traditional
chemical processes using petrochemical starting materials, primarily due to pesticide and fertiliser use in generating the renewable starting material [33].

It can however be generalised that the features of biocatalysis are usually compatible with the goals of Green Chemistry, and this project was conceived within the context of these goals.

1.2.2 Whole cell multi-enzyme biocatalysts

Biocatalytic systems can be assembled in many different ways. They can be assembled both in vivo and in vitro. They can involve one or many enzymic reactions. Here, examples of genetically manipulated whole cell multi-enzyme systems will be concentrated on, as these are the most relevant to this work. These systems can be classified as being altered by synthetic biology or metabolically engineered.

“Synthetic biology” [34] involves creating biological systems which are not found in Nature. This term is applied to biocatalytic production of chemicals where there is an imparting of new metabolism to a whole cell, such that the biotransformation is, to some extent, parallel to host metabolism [35], rather than interwoven into it. An example of this is the transformation of the alkaloid norlaudanosoline to the pharmacologically important (S)-tetrahydroberberine using six enzymes [36]. This transformation was intricate, involving highly regio- and stereospecific chemistry, none of which was endogenous to the Saccharomyces cerevisiae host used. It was performed by heterologous expression of genes of both plant and human origin. Although whole cell biocatalysts can cause waste of starting material through unwanted side reactions, this was not significant in this example. There are also examples where interference of host metabolism has been minimised by deleting the genes encoding the interfering activities [37], or outcompeted by over-expression of the heterologous genes [38]. The use of whole cells as biocatalysts for multi-enzyme reactions does however suffer the caveat that technology for fine control over the levels of heterologous expression of recombinant enzymes are not yet well developed [39].
“Metabolic engineering” [40] is typically an iterative process of optimising cellular properties. In the context of biocatalytic production of chemicals, it involves creating genetic alterations to optimise the flux of carbon from host metabolism towards a desired product. The metabolic fluxes through the system are then quantitated, modelled, and genetic alterations for further optimisation are identified [41]. Synthetic biology and metabolic engineering clearly overlap, but can be distinguished from one another as metabolic engineering operates on a systemic level, using analytical techniques to enable understanding of metabolic flux [42]. It also looks wider than heterologous expression of enzymes to optimise the distribution or magnitude of this flux. There are examples of gene silencing used to change flux distribution. One such example was the use of an *E. coli* gene deletion mutant as a recombinant host. This organism was blocked in the metabolism of glucose to aromatic amino acids, which enabled metabolic flux from this aromatic anabolic pathway to be channelled by heterologous enzymes towards the product, catechol [43]. Gene silencing has also been applied on the transcriptional level, in an example involving the removal of expression of butyrate kinase activity by the use of antisense RNA. This stopped butyrate production and enhanced production of acetone and butanol in *Clostridium acetobutylicum* [44]. Metabolic flux can also be limited by the transport of substrate into the cell. An example of this involved a *Saccharomyces cerevisiae*, metabolically engineered to ferment arabinose to ethanol. Wild type *S. cerevisiae* do not metabolise arabinose, and heterologous expression of bacterial enzymes for the catabolism of this sugar imparted this ability to the recombinant organism. However, it was expression of a galactose permease which transported arabinose that increased ethanol productivity significantly [45].
1.3 The Vanillin to 3-Carboxy Muconate Pathway

1.3.1 Metabolism of vanillin

The number of known instances where vanillin is a metabolic intermediate is surprisingly restricted. Vanillin is intermediate in the anabolic production of glucovanillin in *Vanilla planifolia*, the orchid which grows the vanilla pod [46]. It is a catabolite intermediate in the degradation of lignin depolymerisation products in many microbes [22].

The Kyoto Encyclopaedia of Genes and Genomes [47] and BRENDA [48] databases listed nine enzymatic reactions where vanillin was the substrate on 16/9/2008 (Figure 1).

Of these nine reactions, three involve the condensation of an additional group with the phenol functional group, to form a vanillyl sugar or sulphate. These metabolites are not known to be further metabolised.

The remaining six reactions involve transformation of the aldehyde functionality. Three of these reactions are hypothetical, as they are the reverse reactions of those physiologically observed, and there is no report of the reversibility or irreversibility of these reactions. Two of these reactions, operating in the physiological direction, are involved in the catabolism of lignin depolymerisation products. The first of these is the formation of a carbon to carbon bond with C2 of the acetyl group of acetyl-CoA [49]. The second is dimerisation of two vanillin units, eliminating O2 [50]. The third hypothetical transformation is the reverse of the conversion of vanillylamine to vanillin [51]. In addition to these hypothetically reversible reactions, the aldehyde functionality has been shown to be reversibly converted to the cyanohydrin [52]. The reduction and oxidation of vanillin’s benzaldehyde group to form the benzyl alcohol [53] and benzoic acid [54], respectively have also been demonstrated.

The oxidation of vanillin to vanillate is the metabolic reaction relevant here.
Figure 1: Enzymatic reactions employing vanillin as a substrate
1.3.2 The choice of 3CM as the target product

The target product of this project was chosen by examining the known metabolites of vanillin for their likely usefulness. The target selected was 3-carboxy muconate (3CM) ((1E,3Z)-buta-1,3-diene-1,2,4-tricarboxylic acid, Figure 2).

![Structure of 3-carboxy-muconate](image)

*Figure 2: Structure of 3-carboxy-muconate*

*Adapted from [55]*

3CM was chosen as the target of this project primarily due to its high functional group density, with three carboxyl groups and two alkene groups on only seven carbon atoms. These structural features made it seem likely to be a useful polymer building block, as both multiple carboxylic acid and alkene functionalities can undergo polymerisation reactions. This will be further discussed in Section 1.4.4.

The biocatalytic transformation of vanillin to 3CM seemed plausible because it required only three sequential enzymatic reactions. There was also potential for the redox cofactor required by one enzyme to be recycled by another, such that the conversion of vanillin to 3CM resulted in no net change in host redox state (Figure 3B).

1.3.3 The catabolic pathway containing the vanillin to 3CM reactions

Generally, the products of lignin depolymerisation are catabolised by convergent pathways which funnel a wide diversity of small aromatic compounds to a small number of dihydroxy substituted aromatic metabolites. The aromatic rings of these central dihydroxy metabolites are then oxidatively...
cleaved open by dioxygenases. The ring cleavage products are then trafficked into central metabolism by conversion into common metabolites (e.g. pyruvate, succinate, acetate etc) [56].

The conversion of vanillin to 3CM requires three enzymic reactions which are involved in the catabolism of the lignin depolymerisation product ferulate (Figure 3). This catabolism has been well studied in a number of bacterial species (e.g. *Pseudomonas putida* HR199 [57] and *Acinetobacter baylyi* [58]). In the first reaction, 4-hydroxyl benzaldehyde dehydrogenase (HBD) oxidises the benzaldehyde functional group of vanillin to give vanillate, with the concomitant reduction of a nicotinamide dinucleotide redox cofactor (NAD(P)⁺) [57]. The methoxyl ring substituent of vanillate is then transformed to a hydroxyl substituent by vanillate monooxygenase (VMO) to give 3,4-dihydroxy benzoate, consuming O₂, and producing formaldehyde. This reaction also oxidises NAD(P)H, regenerating the cofactor reduced in the HBD reaction [59]. The aromatic ring of 3,4-dihydroxy benzoate is then oxidatively cleaved by protocatechuate 3,4-dioxygenase (P34O) to give 3CM [60].
Figure 3: Biochemical pathway for the catabolism of ferulate
A is adapted from [57], B is adapted from [58], C is adapted from [61]
The enzymes relevant to this project, 4-hydroxy benzaldehyde dehydrogenase, vanillate monooxygenase and protocatechuate 3,4-dioxygenase are abbreviated to HBD, VMO and P34O, respectively.
1.3.4 4-Hydroxy benzaldehyde dehydrogenase

It has long been known that microbes can oxidise vanillin to vanillate. In 1955, it was observed that soil fungi grew on vanillin, and catabolised this carbon source through a vanillate intermediate [62]. There have been many reports since of microbes which consume vanillin and accumulate vanillate [49, 63-70].

The first observation of vanillin oxidising activity specifically assignable to a dehydrogenase was in 1970, when the crude cell lysates of *Pseudomonas acidovorans* were shown to oxidise vanillin on addition of NAD$^+$ (Figure 4) [71]. A later study observed that HBD activity in lysate was enhanced by addition of NAD$^+$, but not NADP$^+$ [49]. As these observations were made on crude lysate, they may be confounded by transhydrogenase activity converting NAD$^+$ to NADP$^+$, and required confirmation.

![Figure 4: The reaction catalysed by HBD](image)

The gene sequence for a HBD was reported in 1997 [54]. The authors of this report demonstrated the enzyme’s activity by expressing the gene in *E. coli*, imparting the ability to oxidise vanillin to vanillate to these whole cells. Since this time, the gene for the enzyme has been experimentally demonstrated to exist in a number of important soil microbes, including *Pseudomonads* [72-74], *Acinetobacter baylyi* [58], and *Sphingomonas* sp.14DN61 [75].

Although the gene and reaction catalysed by HBD were known, this enzyme had not been purified or characterised.
There are other enzymes which also catalyse the oxidation of vanillin to vanillate. These include an aryl aldehyde oxidase [76]. This reaction differs from one catalysed by a dehydrogenase in that an oxidase utilises $\text{H}_2\text{O}_2$ as the source of electrons, where a dehydrogenase utilises NAD(P) (Figure 4).

There are also a number of dehydrogenases, which are not HBD, and are known to oxidise vanillin (Table 1). Some of these are broad specificity enzymes whose primary substrate is not vanillin. For example, Overhage et al. (1999) reported the purification and amino terminal amino acid sequence of a vanillin oxidising dehydrogenase with no other characterisation [77]. This amino acid sequence is consistent with coniferyl aldehyde dehydrogenase (NCBI reference number AJ006231) [78]. Baré et al. reported the purification and characterisation of a vanillin: NAD$^+$ oxidoreductase which was constitutively expressed by a Pseudomonad [79]. This paper gave the preferred reaction conditions, substrate specificity, molecular characteristics such as native molecular mass and UV spectra, and inhibition of the enzyme by thiol blocking reagents, but not kinetic constants. Results presented in Sections 3.3.4 and 3.3.5 demonstrate that this enzyme is different to the HBD expressed as part of the ferulate catabolism pathway of Acinetobacter baylyi in its substrate specificity and isoelectric point.

Table 1: Dehydrogenases known to oxidise vanillin to vanillate

<table>
<thead>
<tr>
<th>Enzyme Name</th>
<th>E.C. Number</th>
<th>Designated Natural Substrate</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vanillin Dehydrogenase</td>
<td>1.2.1.67</td>
<td>Vanillin</td>
<td>[80]</td>
</tr>
<tr>
<td>Vanillin: NAD$^+$ oxidoreductase</td>
<td></td>
<td>Vanillin</td>
<td>[79]</td>
</tr>
<tr>
<td>Coniferyl Aldehyde Dehydrogenase</td>
<td>1.2.1.68</td>
<td>Coniferyl aldehyde</td>
<td>[77]</td>
</tr>
<tr>
<td>Aldehyde Dehydrogenase (NAD$^+$)</td>
<td>1.2.1.3</td>
<td>Acetaldehyde</td>
<td>[75]</td>
</tr>
<tr>
<td>4-Hydroxy Benzaldehyde Dehydrogenase</td>
<td>1.2.1.64</td>
<td>4-Hydroxy benzaldehyde</td>
<td>[81]</td>
</tr>
<tr>
<td>Alkan-1-ol Dehydrogenase</td>
<td>1.1.20.99</td>
<td>Primary alcohols</td>
<td>[82]</td>
</tr>
</tbody>
</table>
1.3.5 Vanillate monooxygenase

Vanillate monooxygenase (VMO) catalyses the second reaction in the pathway from vanillin to 3CM, which is the conversion of the 3 methoxy group of vanillate to a hydroxyl group (Figure 5). This reaction involves three substrates, vanillate, NAD(P)H, and O₂, and yields three products, 3,4-dihydroxy benzoate, NAD(P)⁺, and formaldehyde. It was first described in microbes in 1967 [59].

![Diagram of vanillate monooxygenase reaction]

**Figure 5: The reaction catalysed by VMO**

VMO is a structurally complex enzyme system consisting of two components, an oxygenase and a reductase, encoded by the genes designated *vanA* and *vanB*, respectively, in both *Pseudomonas* sp. strain HR199 [83], and *Acinetobacter baylyi* [84].

The molecular biology of VMO from a number of microbes has been studied, with reports of the DNA sequence and functional cloning [54, 83, 85-87], as well as the regulation of expression [88-90] of the enzyme from a number of species published.

There has been only one report of the purification of a VMO [90]. This study cloned each subunit of the enzyme from *Comamonas testosteroni* separately with hexahistidine affinity tags added as translational fusions to the N-terminus of each protein. The proteins were expressed and purified by immobilised metal ion affinity chromatography. The two components were then mixed for activity studies. Little characterisation of the purified enzyme was performed. The
molecular masses of the subunits were estimated using SDS-PAGE, and the activity of the enzyme towards vanillate and 3,4-dimethoxy benzoate was compared, without estimation of kinetic constants.

VMO has been documented as having broad substrate specificity. It can be generalised that benzoates with a methyl or methoxy group in the 3 position on the ring are usually substrates [86, 91], although a 2-methoxy group (but not 4-methoxy) has also been shown to be acted on at reduced rate [59]. There have been no reports of quantitative substrate specificity studies using kinetic parameters to compare substrates. As such, there are gaps in the understanding of the substrate specificity of this enzyme.

Qualitative information about the preference for the nucleotide redox cofactor has been reported [90-92]. Buswell and Ribbons reported that NADH and NADPH elicit similar activity from VMO in the crude extracts of several different Pseudomonas species [91]. The authors qualified their observation by stating that transhydrogenation activity in these crude preparations could account for the observed activity with both NADH and NAD(P)H. Hibi et al. performed a study to examine cofactor preference in vivo [92]. This involved comparing heterologously expressed VMO activity between the standard E. coli K-12 host strain, and two mutant K-12 strains. Activity was lower in the mutant host with a limited intracellular pool of NADPH, while VMO activity in a mutant with a raised intracellular pool of NADPH was the same as in the wild type host, suggesting VMO has a preference for NADPH. However, this approach could not control for interfering intracellular processes, and again does not give definitive information about VMO. Providenti et al. stated that equivalent activity was measured when NADH or NADPH was included in the reaction mixture with purified VMO, without quantifying this observation [90]. Quantitative kinetic data would help clarify the uncertainty around the cofactor requirement for this enzyme.

There have been a number of reports stating that VMO is unstable [54, 59, 83, 86, 91, 93-96]. The enzyme from Pseudomonas species was readily and irreversibly inactivated by exposure to atmospheric oxygen [59, 94]. A. baylyi VMO was also described as labile to oxygen in cell lysates [86]. There has been no exploration of the stabilisation of VMO activity, or the mechanism of this
inactivation since Cartwright et al. observed in 1967 that fractions maintained under N₂ remained active for longer than those exposed to atmospheric O₂.

The activity of VMO has also been reported to be sensitive to dilution [59, 94]. As the two components of the enzyme are readily separated during size exclusion chromatography in mild conditions [94], it seemed that the subunits are not covalently joined. The protein to protein interaction required to gain the proximity for electron transfer from the reductase to the oxygenase occur may well be dependant on the concentration of the components.

There are two enzyme classes, additional to VMO, that catalyse the conversion of vanillate to 3,4-dihydroxy benzoate (Table 2). These enzymes differ from VMO in their reaction mechanism and substrate specificity.

**Table 2: Enzymes reported to convert vanillate to 3,4-dihydroxy benzoate**

<table>
<thead>
<tr>
<th>Enzyme Name</th>
<th>E.C. Number</th>
<th>Designated Natural Substrate</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vanillate Monooxygenase</td>
<td>1.14.13.82</td>
<td>Vanillate</td>
<td>[83]</td>
</tr>
<tr>
<td>Vanillate O-Demethylase</td>
<td>1.2.3-</td>
<td>Vanillate</td>
<td>[97]</td>
</tr>
<tr>
<td>4-Methoxybenzoate Monooxygenase</td>
<td>1.14.99.15</td>
<td>4-Methoxy benzoate</td>
<td>[98]</td>
</tr>
</tbody>
</table>

Vanillate O-demethylase has been isolated from an anaerobic microbe and characterised [97]. This enzyme is more structurally complex than VMO, and has three separate protein components. Similarly to VMO this enzyme was sensitive to oxygen. The enzyme removes the methyl group from the methoxy moiety of vanillate, and passes it to a folate cofactor. The folate cofactor requires a further three enzymes and coenzyme A to regenerate it.

4-methoxybenzoate monooxygenase converts vanillate to 3,4-dihydroxy benzoate at 6% of the turnover rate of its natural substrate [98]. These enzymes are similar to VMO in that they are also two component enzyme systems, and unstable when exposed to atmospheric O₂.
1.3.6 Protocatechuate 3,4-dioxygenase

Protocatechuate 3,4-dioxygenase (P34O), E.C. 1.13.11.3 is the only enzyme class known to catalyse the final reaction in the vanillin to 3CM pathway. This reaction incorporates both atoms of O₂ into the aromatic ring of 3,4-dihydroxy benzoate causing its oxidative scission and formation of 3CM (Figure 6).

Figure 6: The reaction catalysed by P34O

The enzyme was first described in 1947 [99] and first isolated in 1954 [60]. Since this time it has been extensively studied.

The P34O reaction is irreversible [60]. P34O’s are reasonably specific for the substrate 3,4-dihydroxy benzoate [100], but have been documented to act on other substituted catechols such as 4-methyl or 4-sulfo catechol, at very low rates [101, 102]. They have also been reported to be inhibited by structural analogues of 3,4-dihydroxy benzoate [103], and divalent metal ions such as Ni²⁺ and Pb²⁺ [104].

The kinetic parameters of the enzyme from a variety of sources as listed on the BRENDA Enzyme Database [48] were generally around the same order of magnitude (Table 3). The Kₘ for P34O from a broad range of microbes for both 3 4-dihydroxy benzoate and O₂ was in the tens of micromolar range.
Table 3: Summary of some kinetic properties of P34O from a variety of organisms

Adapted from [48]

<table>
<thead>
<tr>
<th>Organism</th>
<th>Strain</th>
<th>$K_m$ (mM)</th>
<th>Substrate</th>
<th>Inhibitors relevant to this system</th>
<th>$K_i$</th>
<th>Comment</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Azotobacter vinelandii</em></td>
<td>ATCC 13705</td>
<td>0.0275</td>
<td>O$_2$</td>
<td>Protocatechualdehyde</td>
<td>0.001</td>
<td></td>
<td>[105]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.018</td>
<td>Protocatechuate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.043</td>
<td>O$_2$</td>
<td>4-Hydroxybenzoate</td>
<td>0.1</td>
<td></td>
<td>[101]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.03</td>
<td>Protocatechuate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.07</td>
<td>O$_2$</td>
<td></td>
<td></td>
<td>Spectrophotometric assay</td>
<td>[106]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.002</td>
<td>Protocatechuate</td>
<td></td>
<td></td>
<td>Oxygen electrode assay</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.09</td>
<td>O$_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.003</td>
<td>Protocatechuate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pseudomonas putida</em></td>
<td>ATCC 23975</td>
<td>0.04</td>
<td>O$_2$</td>
<td></td>
<td></td>
<td>Native enzyme</td>
<td>[107]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.03</td>
<td>Protocatechuate</td>
<td></td>
<td></td>
<td>Recombinant enzyme</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.056</td>
<td>O$_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.029</td>
<td>Protocatechuate</td>
<td></td>
<td></td>
<td>4-Hydroxybenzoate</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[108]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4-Hydroxybenzoate</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td>[109]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Protocatechualdehyde</td>
<td>0.014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>vanillate</td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>protocatechuic acid methyl ester</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism Strain</td>
<td>km (mM)</td>
<td>Substrate</td>
<td>Inhibitors relevant to this system</td>
<td>Ki</td>
<td>Comment</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------</td>
<td>-----------</td>
<td>-----------------------------------</td>
<td>----</td>
<td>---------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>Acinetobacter calcoaceticus ATCC 14987</td>
<td>0.0588</td>
<td>O₂</td>
<td>Catechol</td>
<td>0.38</td>
<td></td>
<td>[103]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0714</td>
<td>Protocatechuate</td>
<td>Protocatechualdehyde</td>
<td>0.073</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brevibacterium fuscum ATCC 15993</td>
<td>0.8</td>
<td>O₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.125</td>
<td>Protocatechuate</td>
<td></td>
<td></td>
<td></td>
<td>[110]</td>
<td></td>
</tr>
<tr>
<td>Rhizobium trifolii TA1*</td>
<td>0.0175</td>
<td>Protocatechuate</td>
<td></td>
<td></td>
<td></td>
<td>[111]</td>
<td></td>
</tr>
<tr>
<td>Agrobacterium radiobacter DSM 5681</td>
<td>0.054</td>
<td>Protocatechuate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.205</td>
<td>Protocatechuate</td>
<td>Protocatechuate</td>
<td>0.102</td>
<td>P-3,4-O type II</td>
<td>[102]</td>
<td></td>
</tr>
<tr>
<td>Hydrogenophaga pallieroni DSM 5680</td>
<td>0.058</td>
<td>Protocatechuate</td>
<td>Protocatechuate</td>
<td>0.066</td>
<td></td>
<td>[102]</td>
<td></td>
</tr>
<tr>
<td>Nocardia erythropolis S-1*</td>
<td>0.1</td>
<td>Protocatechuate</td>
<td></td>
<td></td>
<td></td>
<td>[104]</td>
<td></td>
</tr>
<tr>
<td>Tecoma stans *</td>
<td>0.33</td>
<td>Protocatechuate</td>
<td></td>
<td></td>
<td></td>
<td>[112]</td>
<td></td>
</tr>
</tbody>
</table>

* Designates environmental isolates that do not appear to have been deposited
Structural data on the protein has been generated from crystal structures of the enzyme from *Acinetobacter baylyi* (formerly classified as *Acinetobacter calcoaceticus* strain BD413) [113], and a *Pseudomonas putida* strain (formerly classified as *P. aeruginosa*) [114]. The structure and mechanism of the enzyme was recently reviewed [115]. The enzyme in both species was heterodimeric, and the separately encoded proteins are known as the α and β chains. One Fe$^{3+}$ atom was present for each heterodimeric unit. This mononuclear, non-heme Fe$^{3+}$ is bound by four residues of the β chain, two histidine and two tyrosine residues, at the interface of α and β chains, so the interface of the two subunits was assumed to form the substrate binding site. This region has many basic residues, helping to provide charge balance for the anionic substrate 3,4-dihydroxy benzoate.

Much interest in the enzyme has focused on the mechanism of the catalysed reaction. It has been postulated that Fe$^{3+}$ binds both hydroxyl groups of 3,4-dihydroxy benzoate during catalysis, followed by the sequential addition of dioxygen to the C3 and C4 of protocatechuate [113]. A more detailed mechanism has been proposed [116, 117]. This mechanism involved activation of 3,4-dihydroxy benzoate for electrophilic attack by molecular oxygen. The O$_2$ molecule forms an iron-alkylperoxo intermediate, before a rearrangement yields a cyclic anhydride intermediate. This was postulated to be hydrolysed to yield the product 3CM.
1.4 The Chemistry of 3-Carboxy Muconate

1.4.1 Prior synthesis of 3CM

The target product for this project, 3CM (Figure 2), was a known metabolite of aromatic catabolism, and had been previously synthesised. A small number of methods for the preparation of 3CM have been reported [118-120]. All of these methods have their problems, and it could be argued none is adequate for economically and environmentally sustainable production of 3CM.

The first report of the synthesis of 3CM used a purified enzyme preparation of protocatechuate 3,4-dioxygenase (P34O) from Pseudomonas fluorescens to oxidatively cleave protocatechuate, yielding 3CM [118]. This enzymatic synthesis provides a “green” method for the preparation of 3CM, but the cost of 3,4-dihydroxy benzoate ($482 / 100 g from Sigma, 19/09/08, >97%) is approximately ten fold higher than vanillin ($38 / 100 g from Sigma, 19/09/08, >99%), giving impetus for developing a system that uses vanillin as a starting material.

The oxidative cleavage of the aromatic ring of vanillin with sodium chlorite was reported a year after the first report of enzymatic 3CM synthesis [119] (Figure 7). This reaction yielded 3-carboxy muconic acid-1-methyl ester in 28% yield, which has been saponified and protonated to give the tricarboxylic acid. This method suffers from low yield. It also uses or forms toxic compounds (H₂SO₄, NaOCl, toxic ClO₂ formed as a by-product).

![Figure 7: Oxidative aromatic ring cleavage of vanillin to yield 3-carboxy muconate-1-methyl ester](image)
Another synthetic route to a different stereoisomer of 3CM, 3-carboxy-cis, trans-
muconate from 3,4-dihydroxy benzoate has been reported [120], although a
later report demonstrated that this synthesis did not produce the muconate
product claimed [55].

Review of the literature for reports of the synthesis of structural analogues of
3CM did not uncover a superior chemical method. There is a large volume of
literature on the oxidative cleavage of aromatic rings to form substituted
muconates from aromatics with vicinal dihydroxy substituent's (see [121] for
review). These methods could be used to cleave 3,4-dihydroxy benzoate to
yield 3CM. However, similarly to the enzymatic synthesis of 3CM, these
methods suffer the disadvantage of using an expensive starting material.

1.4.2 The problem with 3CM

A caveat to the usefulness of 3CM was apparent from the literature. Muconates
substituted on C3, including 3CM, have been shown to be unstable in solution
[55, 118, 122-124]. This instability is manifested in isomerisation from \((E,Z)\)-
3CM rapidly to the stereoisomeric \((E,E)\)-3CM, and more slowly to a
constitutionally isomeric muconolactone. This isomerisation is accelerated by
low pH, and increased temperature [123]. A tentative mechanism was proposed
for this isomerisation by Ainsworth and Kirby [55] (Figure 8). Jaroszewski and
Ettlinger supported this mechanism [124].
It seems likely that the formation of the $E,E$ isomers is kinetically favoured and formation of the muconolactone isomer is thermodynamically favoured. If the metabolically active ($E,Z$) isomer is treated with conditions that favour isomerisation (heat and acid), ($E,E$) isomer forms rapidly, but transiently, while the lactone forms slowly, but accumulates until the conversion of 3CM into the lactone is almost quantitative [125].

The mechanism of the isomerisation requires a free acid functionality on C1. When C1 is derivatised to the methyl ester, this isomerisation is not observed [55].

1.4.3 Renewable monomers and polymers

The vast majority, by mass, of polymer building blocks are currently sourced from petrochemicals [1]. Consequently, there is much work underway to find both methods of synthesising currently used polymer building blocks from...
renewable feed stocks and also new polymer building blocks which are renewable (reviewed in [126]).

Adipic acid is a building block for Nylon 6,6, which is currently manufactured from the petrochemicals cyclohexane or phenol [127], and is an example of a currently used polymer building block for which a renewable route has been found. Metabolic engineering created an E. coli that converted glucose to hexadienedioic acid which was subsequently chemically hydrogenated to yield adipic acid [128]. Highlighting the challenges facing the development of renewable manufacture of polymer building blocks, this metabolic engineering route has been determined to be economically uncompetitive with manufacture from cyclohexane, primarily due to the relative costs of crude oil and glucose [129]. It seems likely that this type of technology will come into large scale practise if crude oil prices continue to rise, and biomass conversion technologies make cheaper starting materials, like glucose or vanillin, available.

A more commercially successful example is found in the manufacture of 1,3 propanediol. Traditional organic syntheses of this monomer were not commercially viable, but it was known that a polyester of terephthalic acid and 1,3 propanediol had very appealing material properties [130]. Metabolic engineering created an organism which transformed glucose to 1,3 propanediol, and this is now the building block of the high performance terephthalic acid and 1,3 propanediol polyester marketed as Sorona® [131].

Recently, a collaboration between Cargill and Novozymes was announced to produce 3-hydroxy proprionic acid from renewable material using biocatalysis [132]. This molecule can be used as a polyester building block, or can be thermally dehydrated to form the high volume monomer acrylic acid [133]. This type of investment by large corporations like Cargill signals that renewable manufacture of currently used monomers is likely to become economically feasible in the future.

There are also examples of research on renewable materials to develop them into replacements for currently manufactured polymers. Perhaps the best known examples of this are polyesters of biologically produced hydroxyl
carboxylic acids. There has been extensive work on developing polyesters based on lactic acid. Lactic acid is renewably produced on a large scale by fermentation of hexose sugars [134], and the poly lactide polymers synthesised from this building block are commercially available as Natureworks PLA [135].

There has also been extensive research on polyhydroxyalkanoates (PHA’s), which are thermoplastic polymers synthesised as energy and carbon storage molecules in microbes (see [136] for review). These polymers can be produced by fermentation on a multitude of renewable feed stocks (reviewed in [137]), and also in metabolically engineered plants from CO₂ (e.g. [138]).

Both PLA and PHA have low glass transition temperatures, making their processing difficult, and also tend to be brittle [135, 137]. These deficiencies in material properties limit their applications. It seems that renewable materials which may have novel applications as polymer building blocks need to demonstrate equivalent or superior material properties in the resultant polymer to be viable replacements to petrochemically derived monomers.

1.4.4 3-Carboxymuconate's potential as a polymer building block

The structure of 3CM suggests that it could be a useful polymer building block. Such polymerisation could occur through reactions of either the alkene or carboxylic acid functionalities of 3CM (Figure 9).
Figure 9: Potential polymerisation reactions of 3CM

The previous polymerisation of 3CM has not been reported, so the literature was surveyed for polymers synthesised from compounds structurally related to 3CM.

There were a number of reports of polymerisation of compounds structurally similar to 3CM through radical reactions of the butadiene bonds. \(E,Z\)-hexadienedioate (Figure 10) has been reportedly polymerised by free radical reaction in solution, both as the free acid [139], and as the diethyl ester [139, 140]. Therefore free radical reaction in solution offered an experimental system for polymerising 3CM.

Another interesting polymerisation system reported extensively in the literature had successfully polymerised a number of structural analogues of 3CM (Figure 10). This system involved the ultraviolet light catalysed polymerisation of isomers and derivatives of hexadienedioic and hexadieneoic acids in the solid state [141]. This solid state reaction system was discovered fortuitously. While purifying diethyl \(E,Z\)-hexadienedioate with the aim of using it to synthesise a polymer, these authors observed that some crystalline material became insoluble during repeated recrystallisation [142]. Investigation revealed that radiation (visible light, UV, x-ray or \(\gamma\)) had caused the conjugated double bonds to polymerise.
Exposing the 1-naphthylmethyl ammonium salts of structural analogues to UV irradiation resulted in more of these compounds polymerising, making the system more generally applicable [143]. The authors postulated that the aromatic moieties propensity to stack one on top of the other forced the acid component to take on a columnar arrangement in a solid crystal, with the butadiene double bonds close enough to react [144]. These authors have further stated that monomers which stacked close to 5Å from one another in the crystal appear to be those that polymerise [145]. This suggested that exposure of 3CM to UV in the solid state would also be an experimental system worth exploring.
1.5 Aims, scope and approach of the project

No work prior to this project had been performed on constructing a synthetic pathway system for transforming vanillin to 3CM. This project required the construction of a synthetic pathway based on research carried out on the constitute enzymes, and it also needed to be established if 3CM was a useful target molecule. This formed the three broad aims set for this project.

1) It was aimed that this project ascertained if a synthetic pathway system for transforming vanillin into 3CM was theoretically possible. To establish this, HBD and VMO required characterising for their kinetic behaviour, so that the theoretical requirements to make the system function could be understood. Further detailed characterisation of these enzymes, such as elucidating reaction mechanisms were seen as beyond the scope of this study.

The strategy to address this aim involved cloning, expressing and purifying HBD and VMO for in vitro characterisation. It should be acknowledged that there are intrinsic limitations with this approach. The behaviour of enzymes varies depending on their environment, so the behaviour of purified enzymes may not be reproduced inside a cell. Similarly, the behaviour of purified enzymes in one set of in vitro conditions (e.g. buffer, temperature, ionic strength, level of purity etc) may not be reproduced in any other set of conditions. As such, the characterisation of enzymes under one set of conditions can act only as a guide for how they will behave in any other set of conditions.

2) The second aim of this project was to construct the first prototype of the synthetic pathway using data from the enzyme characterisation to guide decisions about how best to construct the system. The focus was to create a functional system, and identify its weaknesses, without undertaking work on optimisation.

The strategy to address this aim involved subcloning the genes encoding the enzymes of the vanillin to 3CM pathway on plasmids such that they could all be transformed into a single recombinant host. This host could then be cultured, recombinant protein expression induced, and resting cells used as biocatalysts. This biocatalyst could then be characterised by transformations it performed.
3) A third aim in this project was to determine if 3CM was a useful polymer building block.

The strategy to address this involved studying the behaviour of 3CM in two reaction systems; free radical reaction in solution and UV initiated reaction in the solid phase.
2 Materials and Methods

2.1 The Genetic Source of Enzymes

A genetic source for the three enzymes of the vanillin to 3CM pathway was required. It was seen as important to understand the behaviour of each of the three enzymes independently of each other before combining them, and would have been desirable to choose from a suite of well characterised enzymes, picking out those with desirable characteristics. This was not possible, as there was little characterisation of HBD and VMO reported in the literature. There was, however, extensive literature on the enzymology of P34O. As the organisms that contained P34O also contained HBD and VMO, the organism that had a P34O with the most favourable characteristics was chosen as the genetic source for all of the enzymes.

The kinetic constants estimated for P34O are given in Table 3, Section 1.3.6. As the enzyme was to be used in a pathway, and the concentration of its substrate would be determined by the activity of the enzyme preceding it (VMO), a low $K_m$ was seen as a more favourable characteristic than a high $k_{cat}$. The $K_m$ values for 3,4-dihydroxy benzoate reported for the same strain of $P. putida$ confusingly vary from $\sim0.0025$ mM [106] to $\sim0.025$ mM [101, 107]. The value 0.0025 mM will be considered an outlier as the value 0.025 mM was corroborated by two independent reports. The three P34O’s from Azotobacter vinelandii, Pseudomonas putida, and Acinetobacter calcoaceticus had $K_m$ values for 3,4-dihydroxy benzoate which were within the same order of magnitude ($\sim0.03$ mM). The $K_m$ for $O_2$ in enzymes from Azotobacter vinelandii, Pseudomonas putida, and Acinetobacter calcoaceticus was variable, but all within the same order of magnitude ($\sim0.05$M). This constant is either higher, or unreported for the other species in this table. By considering the $K_m$ for the aromatic substrate primarily, with reference also to the $K_m$ for $O_2$, the field of possible donors for P34O to be used here was narrowed to these three microbes.

The enzyme from $P. putida$ was reported to be inhibited by vanillate [109], while this metabolite was reported as having no effect on the kinetics of the enzyme.
from *A. calcoaceticus* [103]. The affects of vanillin and vanillate have not been reported for the *A. vinelandii* enzyme. Having knowledge that a metabolite which would be present in this system was known to have no inhibitory effect on the P34O from *Acinetobacter calcoaceticus* BD413 (American Type Culture Collection, 33305) made this enzyme seem like the optimal choice. This organism has been reclassified as *Acinetobacter baylyi* [146], and was used as a donor of the genes for all enzymes.

### 2.2 General Materials and Methods for Cloning

Standard molecular cloning methods were used [147], unless otherwise stated. Enzymes, buffers and deoxynucleotide triphosphates, unless otherwise stated, were obtained from New England Biolabs.

#### 2.2.1 Polymerase chain reactions

*A. baylyi* genomic DNA was used as a template for polymerase chain reactions (PCR), after isolation using a DNeasy Blood and Tissue kit (Qiagen). Oligonucleotides used for PCR primers were synthesised by Micromon (Victoria, Australia), and the sequence of these is given in Table 4. These primers were designed using the genomic sequence of *A. baylyi* [148].

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR <em>hcaB</em> forward</td>
<td>CAG GAT CCG CAA CAT GTA CAG TTA CTT ATT G</td>
</tr>
<tr>
<td>PCR <em>hcaB</em> reverse</td>
<td>TAA AAC CAT AGC GGC CGC TTA AAT TGG G</td>
</tr>
<tr>
<td>PCR <em>vanAB</em> forward</td>
<td>TCA GAT CTT TTT ATT AAA AAT GCC TGG</td>
</tr>
<tr>
<td>PCR <em>vanAB</em> reverse</td>
<td>ACT CTC GAG TTA AGC AAG ATC AAT GAC</td>
</tr>
<tr>
<td>PCR <em>pcaHG</em> forward</td>
<td>TAA GAT CTT TCT CAA ATT ATT TGG GGA G</td>
</tr>
<tr>
<td>PCR <em>pcaHG</em> reverse</td>
<td>TCC TCG AGT TAG ATA TCG AAA AAT ACC</td>
</tr>
</tbody>
</table>

PCR was performed in an iCycler (BioRad). PCR reactions were performed with Phusion DNA polymerase in the “HF” buffer provided, 40 ng/mL *A. baylyi*
genomic DNA, 200 μM deoxynucleotide triphosphates, 200 nM of each primer, and 0.02 U / μL polymerase. The concentration of MgCl₂ and thermocycling program used varied depending on the gene being amplified and is given in the Materials and Methods section of the chapter specific to each enzyme.

2.2.2 Restriction cleavage, ligation and transformation

The vectors used in this study were pETDuet-1 and pACYCDuet-1 (Novagen). Both plasmid and PCR generated DNA were isolated and purified using kits supplied by Qiagen (Australia). Restriction enzymes and buffers for these were different for each enzyme, and are given in the Materials and Methods section of the chapter specific to each enzyme. After restriction cleavage, DNA was gel purified.

DNA concentration of gel purified linear DNA was measured using UV absorbance at 260 nm with a NanoDrop 1000 spectrometer (Thermo Scientific). Ligation reactions were performed with T4 DNA ligase and buffer at 16°C for 24 hours.

The recombinant host used was *Escherichia coli* BL21-Gold (DE3) (Stratagene). These cells were made electrocompetent by following a published method [149], and stored at -80°C. Transformations were performed by electroporation using 2 mm gap electroporation cuvettes (Molecular BioProducts, product #5520) in a Biorad Micropulser with 2.5 kV for 6.3 ms.

2.2.3 Microbial growth for cloning experiments

Luria Broth (LB) [150] components (10 g/L tryptone, 5 g/L yeast and 10 g/L NaCl) and bacteriological agar #1 (used at 1% w/v) were sourced from Oxoid. Antibiotics were sourced from Sigma Aldrich.

Cultures were grown at 37°C, unless otherwise stated. Microbes were grown overnight to generate both genomic and plasmid DNA for cloning experiments. Media was sterilised by autoclaving for 20 minutes. Antibiotics were sterilised by filtering through a 0.22 μm filter directly into media after its temperature had fallen below 55°C. Carbenicillin was used at 100 μg/mL, and chloramphenicol was used at 34 μg/mL, unless otherwise stated.
Small scale liquid cultures (<10mL) were grown in 50 mL plastic screw top tubes, with agitation at 250rpm after inoculation with a single colony picked from solid media.

2.2.4 DNA analysis

Agarose electrophoresis was performed using analytical grade agarose (Promega). This was run using a PowerPack 1000 with Mini Sub Cell GT (BioRad). The running buffer was TAE [150], with DNA solutions mixed in loading buffer II [150] before loading and visualised with ethidium bromide (Sigma E1510). Molecular mass estimations of DNA were made by comparison to Lambda phage DNA digested by HindIII (Finnzymes).

Colony PCR was used to screen transformants for recombinant plasmids. In this analysis, single colonies were picked with sterile pipette tips from the plates and resuspended in 20 μL of PCR solution. This technique was performed as described in Section 2.2.1, with deviations in the composition of reaction solutions. Taq DNA Polymerase (New England Biolabs) and Taq DNA Polymerase Buffer (New England Biolabs) were used in place of Phusion DNA polymerase and HF buffer. Template DNA solution was omitted. Cells suspended in this solution provided template DNA.

Plasmid DNA was also analysed by restriction enzyme digestion before the molecular mass of the fragments was estimated by agarose gel electrophoresis.

The DNA sequence of recombinant plasmids was analysed by the Monash University Micromon facility. This facility uses the Applied Biosystems PRISM BigDye Terminator Mix, version 3.1 chemistry and an Applied Biosystems 3730S Genetic Analyser. Sequencing primers were synthesised by Micromon (Table 5).
<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novagen pET upstream sequencing primer</td>
<td>ATG CGT CCG GCG TAG A</td>
</tr>
<tr>
<td>Novagen T7 terminator sequencing primer</td>
<td>GCT AGT TAT TGC TCA GCG G</td>
</tr>
<tr>
<td>Novagen DuetDOWN1 sequencing primer</td>
<td>GAT TAT GCG GCC GTG TACAA</td>
</tr>
<tr>
<td>Novagen DuetUP2 sequencing primer</td>
<td>TTGTACACGGCCGCATAATC</td>
</tr>
<tr>
<td>Novagen ACYCDuetUP-1 sequencing primer</td>
<td>GGA TCT CGA CGC TCT CCC T</td>
</tr>
<tr>
<td>vanB sequencing primer</td>
<td>GTC AAT ATG TGA GCC TGC G</td>
</tr>
<tr>
<td>pcaG sequencing primer</td>
<td>GGA TGC GTT GAC CTT GCG</td>
</tr>
<tr>
<td>PVP Up sequencing primer</td>
<td>GCA GGC ATA GCA ATG GTC TTT C</td>
</tr>
<tr>
<td>PVP Down sequencing primer</td>
<td>ATT CTG GCA AAT GGC GTT AC</td>
</tr>
</tbody>
</table>
2.3 General Materials and Methods for Production of Cell Lysates

2.3.1 Recombinant protein expression

Seed cultures were grown at 30°C overnight in LB with the appropriate antibiotic. The cells were harvested by centrifugation (20 minutes at 10 x 10^3 g and 4°C), and the medium decanted. The cells were resuspended in ice cold sterile filtered Phosphate Buffered Saline (PBS), composed of 10 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl, and 2.7 mM KCl.

Expression cultures were seeded to an OD_{600} of 0.1 with this PBS cell suspension and grown in LB with the appropriate antibiotic at the concentration given in Section 2.2.3. Growth was carried out at 37°C in unbaffled flasks with 20% liquid volume incubated in shaking incubators at 200 rpm, unless otherwise stated. Growth was monitored by measuring the optical density of the culture at 600 nm (OD_{600}). Recombinant protein expression was induced by the addition of isopropyl thiogalactosidase (IPTG) (Progen Biosciences) sterile filtered into the induced culture medium to achieve a final IPTG concentration of 1.0 mM. The cultures were harvested when the OD_{600} did not change over a 45 minute period. The cells harvested by centrifugation (20 minutes at 10 x 10^3 g and 4°C), washed with ice cold sterile filtered PBS, and again pelleted by centrifugation. The pellets were stored at -20°C.

2.3.2 Cell lysis

Lysate was generated on the day of the experiment. During lysis, all steps were performed at 4°C. The buffer used varied depending on the enzyme, and the composition of the lysis buffer is given in the Materials and Methods section of the chapter specific to each enzyme. Cell pellets were resuspended in 10 mL of buffer and subjected to four discrete passes through an Avestin EmulsiFlex-C5 homogeniser at a pressure of approximately 600 KPa. This crude lysate was clarified by centrifugation at 10x 10^3 g for 20 minutes, and then 0.22 μm filtered.
2.4 General Materials and Methods for Protein Purification

2.4.1 Chromatography

Most of the materials and methods used for protein purification were specific to each enzyme, and are given in the Materials and Methods section of the chapter specific to each enzyme. However, some generalisations can be made. Chromatography was performed using an ACTAPrime system, monitoring the absorbance at 280nm. Unless otherwise stated, chromatography was performed at 4°C. Unless otherwise stated, chromatographic media was packed into XK 16 columns (GE Healthcare). All columns were equilibrated with 10 column volumes of the loading buffer before loading the protein.

2.5 General Material and Methods for Protein Analysis

2.5.1 SDS-PAGE

All electrophoresis equipment and reagents were sourced from Invitrogen. SDS-PAGE was performed using 10% Bis-Tris gels run in MOPS buffer using an Xcell Surelock mini vertical gel tank against SeeBlue2 molecular weight standards. Protein was detected by using SimplyBlue Safe Stain, or by silver staining [151]. Images of gels were taken using a GelDoc Image Station (BioRad). Image analysis was performed using QuantityOne software (BioRad).

2.5.2 Protein concentration assays

Protein concentrations were determined by the Bradford dye binding assay [152] using a reagent supplied by Bio-Rad.

2.6 General Materials and Methods for Enzyme Analysis

2.6.1 Substrates for enzymatic reactions

Vanillin (99%), vanillic acid (97%), 3,4-dihydroxy benzoate (97%), NAD$^+$ (≥95%), NADH (98%), and NADPH (≥90%) were all sourced from Sigma Aldrich.
2.6.2 Continuous steady state enzyme activity assays

Continuous enzyme activity assays were performed using a Shimadzu 1601 dual beam spectrophotometer. Reaction solutions were 1 mL, contained in quartz cuvettes with a 1 cm path length. The cuvette in the front light beam held a complete reaction mixture with all components. The cuvette in the rear light beam held an incomplete reaction mixture, with all components except the aromatic substrate. This compensated for endogenous activities and non-enzymatic reactions which caused a change in absorbance that was not attributable to the activity of the enzyme on the aromatic substrate. The time interval over which activity was measured was varied to ensure that the reaction rate was measured over the linear portion of the reaction time course.

2.6.3 Discontinuous steady state enzyme activity assays

Discontinuous enzyme activity assays measured the amount of product present in quenched reaction samples taken at a number of time points. This assay was used where a continuous assay was not possible. The primary reason excluding a continuous activity assay was interfering UV absorbance by a species in the reaction mixture. For example, when measuring HBD activity, vanillin absorbed strongly at the wavelength being monitored to follow NADH production. A discontinuous assay was also used where highly variable results were obtained by the continuous assay. For example, in crude lysates, variable results were obtained for VMO activity by measuring the decrease of absorbance due to NADH oxidation, as interfering NADH oxidising activities were very high compared to VMO activities.

The discontinuous assays involved sampling aliquots of reaction solutions at measured time intervals, which were added to a solution of methanol/glacial acetic acid (90:10 v/v) to stop the reaction. Before reactions were started by the addition of enzyme, a sample of reaction solution was taken diluted in methanol/glacial acetic acid (90:10 v/v). An appropriate amount of enzyme was then added to this “time zero” sample, so that it contained all reaction components, but no reaction occurred, due to enzyme denaturation in the low pH organic solvent. This “time zero” sample was analysed immediately prior to analysis of
time point samples, to compensate for any minor activity which occurred in the methanol/acetic acid solution. For steady state kinetic experiments, aliquots were taken at three time intervals to ensure that the rate measured was linear with time.

Stopped reaction samples were analysed by High Performance Liquid Chromatography (HPLC). A Shimadzu Prominence system was employed, equipped with an Alltima C18 column (150 x 4.6 mm, 5µM particle size) preceded by a 7.5 mm guard column of the same specifications. Chromatography was performed at 30°C. Detection was performed spectrophotometrically at 220 nm, and quantitation was made against external standards. Mobile phases were isocratic, and contained 0.1% (volume / volume (v/v)) trifluoroacetic acid, although the composition was varied between 10 and 25% (v/v) methanol depending on the analytes.

2.6.4 Estimating Michaelis – Menten constants

Before performing analyses to estimate Michaelis - Menten constants, experiments were performed to ascertain that the initial rate of the reaction was linearly proportional to the amount of enzyme in the reaction.

The reactions of the enzymes of this pathway all had more than one substrate. The apparent kinetic constants of these enzymes in response to substrate concentrations were estimated by varying the concentration of one substrate while holding the concentrations of all other substrates constant. Apparent kinetic constants apply only to the conditions in which they are measured, and are not accurate when the concentrations of the other substrates are different. As such, these constants can act only as a guide. However, they are much more convenient to estimate than the “true” constants. The apparent constants of an enzyme with two substrates can be reliably estimated from triplicate velocity measurements at six primary substrate concentrations and one saturating concentration of the secondary substrate [153]. Estimation of the “true” constants requires triplicate measurements at 36 concentrations (six primary substrate concentrations at each of six different concentrations of the secondary substrate) [153]. Even with this volume of data, the true constants
may not be more representative of the enzyme’s behaviour than apparent constants, as the variation witnessed in these data sets is typically high [153].

Unless otherwise stated, substrate concentrations were varied from 0.25 to 10 times the approximate value of the $K_m$. Steady state kinetic data from these activity assays were fitted to the Michaelis-Menten equation for one substrate using the software XLfit 4.1 (ID Business Solutions, Guildford, U.K.). This software regression fit the hyperbolic curve described by the Michaelis – Menten equation, and estimated $K_m$ and $V_{\text{max}}$ values, along with standard errors for these values.

### 2.6.5 Substrate specificity assays

For substrate specificity assays, substrates were incubated with the enzyme in parallel with a known substrate for comparison. Stopped reaction samples were analysed by liquid chromatography mass spectrometry (LC MS).

LC MS was performed with the same instrument as HPLC (Section 2.6.3), equipped with a LCMS-2010EV mass spectrometer (MS) detector. The mobile phase composition for LC MS was varied depending on analytes, but always consisted of acetonitrile and H$_2$O with 0.1% glacial acetic acid (v/v). 10% of the eluent was diverted onto the MS. Ionisation was performed with an electrospray ionisation probe. The negative spectrum was scanned between m/z 50 to 250 with a scan time of 0.05s. The detector, interface, CDL and Q-array voltages were 1.5 kV, 4.5 kV, -20 V and 150 V respectively. The CDL and block temperatures were 250 and 200°C respectively. Drying and nebulising gases (N$_2$) were set to 0.1 MPa and 0.5 L/min respectively.

### 2.7 General Material and Methods for Chemical Analysis

$^1$H-NMR and $^{13}$C-NMR was carried out using either a 200 MHz Bruker Av200, or 400 MHz Bruker Av400 spectrometer. These instruments were maintained by Dr Roger Mulder and Dr Jo Cosgriff of CSIRO Molecular and Health Technologies.

FTIR spectra were measured on an Elmer Perkin Spectrum 2000 spectrometer.
UV spectra were measured on a Shimadzu UV1601 spectrometer.

The purity of 3CM-Na$_3$ synthesised chemically was assessed by HPLC, using a published ion pairing method for separating tricarboxylic acids [154]. The instrument and instrumental parameters were as described in Section 2.6.3.

Gas chromatography- mass spectra (GC-MS) were obtained with a ThermoQuest (TRACE DSQ) GC mass spectrometer in the positive ion mode with an ionization energy of 70 eV. The gas chromatography was performed with a SGE BPX5 column using a temperature program of 40°C for 2 minutes, then heating at 40°/min to 300°C where the temperature was held for 2 minutes with a split ratio of 20, an injector temperature of 250°C and the transfer line was set to 250°C. High-purity helium was used as carrier gas with a flow rate of 0.8 ml/min. This instrument was maintained by Carl Braybrook and Dr Jo Cosgriff of CSIRO Molecular and Health Technologies.

Gel permeation chromatography (GPC) was performed on a Waters 515 HPLC pump and Waters 717 Plus Autosampler equipped with Waters 2414 refractive index detector and three Mixed C and one mixed E PLgel columns (each 7.5mm×300 mm) from Polymer Laboratories. Tetrahydrofuran (flow rate 1.0 mL/min) was used as eluent. The columns were calibrated with narrow polydispersity polystyrene standards (Polymer Laboratories). This instrument was maintained by Dr Ming Chen of CSIRO Molecular and Health Technologies.
3 4-Hydroxy Benzaldehyde Dehydrogenase

3.1 Introduction

3.1.1 Literature review and background for HBD

The first enzyme in the vanillin to 3CM pathway, 4-hydroxy benzaldehyde dehydrogenase (HBD), oxidised the benzaldehyde group of vanillin to the benzoate group of vanillate, with the concomitant reduction of NAD(P)$^+$ to NAD(P)H (Figure 4, Section 1.3.4).

As discussed in Section 1.3.4, this enzyme was known before this study, but little of its character had been described. The gene sequence and activity of HBD had been reported [54], as had the existence of this enzyme in a number of important soil microbes, including Pseudomonads [72-74], Acinetobacter baylyi [58], and Sphingomonas sp.14DN61 [75]. Observations had been made on the enzyme’s preference for NAD$^+$ over NADP$^+$ in crude lysate [49, 71], but these observations may be confounded by transhydrogenase activity converting NAD$^+$ to NADP$^+$, and required confirmation. The substrate range of HBD was shown to include other 4-hydroxy benzaldehydes in whole cells of A. baylyi [58] and lysates of Escherichia coli heterologously expressing this dehydrogenase from Sphingomonas paucimobilis, but no quantitation of substrate preference had been made.

Many of the characteristics which had not been adequately reported in the literature were relevant to the understanding of HBD’s behaviour as part of the synthetic vanillin to 3CM pathway. The enzyme’s preference for NAD$^+$ or NADP$^+$ was pertinent to being able to predict if cofactor recycling of the vanillin to 3CM could be expected to occur in practise. The reversibility of the reaction catalysed by HBD was relevant to its behaviour in the synthetic pathway. It was known that E. coli possessed an endogenous vanillin reducing activity [155]. This activity would compete with HBD for the substrate vanillin, and the reversibility of the HBD catalysed reaction would be significant in determining the outcome of this competition. The kinetic characteristics of HBD were important to the behaviour of the synthetic pathway, as HBD activity supplied
the rest of the pathway with substrate. At the time at which HBD was being characterised, the stability of HBD was relevant. It was not certain at this point whether the artificial pathway would be best constructed in vitro or in vivo. While an in vitro system would have been less complex, it would have required that all three enzymes were sufficiently stable in vitro. The preferred reaction conditions of HBD were also relevant characteristics to the assembly of an in vitro system which had not been adequately described.

In addition to the characteristics which were relevant to the assembly of the synthetic vanillin to 3CM pathway, other characteristics, such as substrate specificity and molecular characteristics were pertinent to the broader understanding of the enzyme. HBD can be classified within the superfamily of enzymes, aldehyde dehydrogenases [156]. Within this superfamily, benzaldehyde dehydrogenases such as HBD have been classified by sequence homology to the same branch as the well studied Class 3 Aldehyde Dehydrogenases [157]. Class 3 aldehyde dehydrogenases generally use NAD\(^+\) instead of NADP\(^+\), exist as homodimers in cellular cytosols and show high specificity for their substrate [157, 158], although aromatic aldehyde dehydrogenases which do not conform to the generalisations about substrate specificity have been well documented [159, 160]. It had not been experimentally demonstrated that HBD displayed these characteristics.

3.1.2 Aim and approach for HBD

In order to better understand the behaviour of this enzyme, and its place in the pathway studied here, it was aimed to purify and characterise HBD cloned from A. baylyi. This allowed testing of some of the characteristics predicted by sequence similarities to Class 3 Aldehyde Dehydrogenases, confirmation of conclusions drawn from previously performed experiments with whole cells or lysate, and the reporting of fundamental characteristics measurable only with purified enzymes. This work has been published elsewhere [161], and is included here in more detail.
3.2 Materials and Methods

3.2.1 Cloning the \textit{hcaB} gene encoding HBD

The gene encoding HBD in \textit{A. baylyi}, termed \textit{hcaB}, and the corresponding flanking DNA have previously been identified and published [58]. This sequence was used to design primers for amplification of \textit{hcaB}, incorporating recognition sites for the restriction enzymes \textit{BamH}I and \textit{NotI} (PCR \textit{hcaB} forward and PCR \textit{hcaB} reverse, Table 4, Section 2.2.1). The PCR conditions used were: “hot start” at 98°C for 30 s; 30 cycles of 98°C for 10 s, 44°C for 30 s, and 72°C for 50 s; final extension at 72°C for 500 s. The reaction contained 1.5 mM MgCl$_2$.

The PCR product was cloned into pET-Duet1 between the \textit{BamH}I and \textit{NotI} restriction sites using standard techniques [147]. The recombinant plasmid was called pET-HBD (Figure 11).

Figure 11: Map of the recombinant plasmid pET-HBD
The sequence of pET-HBD was determined using the primers Novagen pET upstream sequencing primer and Novagen T7 terminator sequencing primer (Table 5, Section 2.2.4). The sequence of the coding DNA was confirmed to be as expected (Appendix section 9.1). A hexahistidine affinity purification motif (his-tag) was fused to the amino terminal of the gene encoding HBD. The coding sequence following the his-tag coding region had 100% homology with the sequence reported in the genome sequence of *A. baylyi* [148] (Appendix 9.1).

### 3.2.2 Heterologous expression of HBD

Seed cultures of *E. coli* BL21(DE3) (Novagen) carrying recombinant plasmid were grown overnight at 37°C in 100 mL of Luria Broth (LB) in an unbaffled 1 L shake flask containing 100 μg/mL carbenicillin, and used to inoculate 1.5 L of LB containing 25 μg/mL carbenicillin in a 2 L Sartorius Biostat B fermenter. When the OD$_{600}$ of the culture reached 1.0, recombinant protein expression was induced by addition of IPTG to a final concentration of 1 mM. The culture was harvested by centrifugation when the OD$_{600}$ was no longer increasing.

### 3.2.3 Purification of HBD

All buffers for purification, unless stated otherwise, were pH 7.0, and contained 500 mM NaCl. Cell pellets of approximately 1 g (wet weight) were thawed on ice, and resuspended in 10 mL of 20 mM imidazole supplemented with 0.1 mg/mL deoxyribonuclease (Sigma, DN25) and 1.0 mg/mL MgCl$_2$. The suspensions were lysed and the lysate clarified as described in Section 2.3.2.

Chromatography was operated at 1 mL/min, with a 1.0 mL His-Trap FF IMAC column (GE Healthcare) at 22°C. The method was based on a general protocol [162]. 2.0 mL of lysate was loaded onto the column and washed through the column using 10 mL of 20 mM imidazole. Two intermediate wash steps of 10 mL each were included (55 mM and 110 mM imidazole respectively) before the target protein was eluted with 300 mM imidazole.

Fractions containing HBD activity against vanillin were diafiltered using an ultrafiltration device (Pall) with a 3 kDa molecular weight cutoff, to change the
buffer to 50 mM Tris-HCl, pH 8.5, 100 mM NaCl and 20% (v/v) glycerol. These enzyme solutions were stored at -80°C.

3.2.4 HBD steady state activity assay

All substrates were obtained from Sigma Aldrich. Vanillin (Sigma, 99%), NAD\(^+\) (Sigma, disodium salt from *Saccharomyces cerevisiae*, ≥95%), NADP\(^+\) (BioChemica, ≥94%), FAD (Sigma, ≥95%) and NADH (Sigma, ≥98%) were used. A discontinuous enzyme activity assay was used (Section 2.6.3). The reaction conditions employed for this assay were, unless otherwise stated, 25 mM glycine, pH 9.3, 100 mM NaCl, 25 mM KCl, 1 mM vanillin, 5 mM NAD\(^+\) in a total volume of 440 μL at 21 °C. One unit of activity was defined as the amount of enzyme which produced 1 μmol of vanillate per minute in these conditions. The amount of enzyme added to the reaction mixture was varied from 1 x 10\(^{-3}\) to 1.8 x 10\(^{-2}\) U. In this range of enzyme amounts, the initial reaction velocity was found to give a linear response to increasing enzyme amounts.

For substrate specificity screens, substrates were incubated with the enzyme for a minimum of five minutes before the reaction was stopped.

3.2.5 Isoelectric focusing electrophoresis

All electrophoresis equipment and reagents were sourced from Invitrogen. Isoelectric focusing electrophoresis (IEF) was performed using pH 3 to 10 gels using IEF Marker 3-10 standards. The Coomassie stain, gel running and imaging apparatus described in the Section 2.5.1 were used.

3.2.6 Analytical size exclusion chromatography

Size exclusion chromatography was performed with Sephadex G200 resin packed in an XK16 column with a 16 mm diameter (GE Healthcare) to a bed height of 70mm, calibrated with native molecular size standards from Sigma (MWND500).
3.3 Results and Discussion

3.3.1 Expression and purification of HBD

Soluble 4-hydroxy benzaldehyde dehydrogenase was expressed to a high level
in *E. coli* BL21(DE3). The expected molecular mass of this protein calculated
from the amino acid sequence was 55 kDa. Densitometry analysis of SDS
PAGE of clarified lysate (lane 1, Figure 2) revealed a band of the expected
molecular mass composing around 10% of the total optical density of that lane.
A corresponding band was not present in SDS-PAGE analysis of the lysate of a
host grown under identical conditions, carrying only the empty vector pETDuet.
SDS PAGE analysis of whole cells showed no further increase in this band after
two hours induction.

The recombinant enzyme was purified using immobilised metal ion
chromatography (IMAC). Elution with 300 mM imidazole gave strong HBD
activity, and SDS PAGE analysis showed a single band with an apparent
molecular size of 55 kDa (lanes 5 and 7, Fig.2). This indicated that purification
was achieved, and HBD was electrophoretically homogenous (>95%) as
detected by Coomassie Brilliant Blue staining after a single purification step.
**Figure 12: SDS-PAGE gel showing the expression and IMAC purification of HBD**

Lanes 1, 2, 5 and 7 were loaded with 6.8 μg of total protein. The gel was run in reducing conditions.
The purification also achieved a high recovery of 4-hydroxy benzaldehyde activity. 94% of the activity measured in the clarified lysate was found in the purified material (Table 1).

**Table 6: Summary of a typical IMAC purification of HBD**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Protein concentration (mg/mL)</th>
<th>Specific activity (U / mg protein)</th>
<th>Fold purification</th>
<th>Total activity (U)</th>
<th>Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarified lysate</td>
<td>5.77</td>
<td>3.2</td>
<td>Not applicable</td>
<td>37</td>
<td>100</td>
</tr>
<tr>
<td>IMAC eluted HBD</td>
<td>0.27</td>
<td>22</td>
<td>6.9</td>
<td>35</td>
<td>94</td>
</tr>
</tbody>
</table>

This was the first purification of HBD reported in the literature. Two other enzymes displaying dehydrogenase activity against vanillin have been purified previously [77, 79]. However, these enzymes differed significantly from the enzyme described here. The enzymes previously described both showed low activity and were constitutively expressed, rather than the primary catabolite enzyme expressed on induction of the pathways for catabolising cinnamic acids. The N-terminal amino acid sequence of the enzyme purified by Overhage et al [77] matched that of an enzyme which has been purified and identified as coniferyl aldehyde dehydrogenase (NCBI reference number AJ006231) [78], not the enzyme characterised in this work. The enzyme purified by Baré et al [79] showed different substrate specificity, in addition to exhibiting markedly lower activity, indicating a significant difference from the subject of this chapter.

### 3.3.2 Molecular characteristics of HBD

Size Exclusion Chromatography was used to estimate the molecular size of the native protein. Interpolation from a calibration curve of native protein standards calculated that HBD in solution was 116 kDa, compared to the 55 kDa molecular mass estimated by SDS-PAGE. This suggested the enzyme may exist in its native state as a dimer. This was consistent with what has been reported for Class 3 rat liver aldehyde dehydrogenase [158], which shows amino acid sequence homology to HBD.
When SDS-PAGE was run under non-reducing conditions, the purified enzyme displayed the same electrophoretic profile as seen under reducing conditions. This gave evidence that disulphide bonds between subunits were not present and suggested that the two subunits were possibly held together by non-covalent intermolecular forces that were disrupted with SDS and boiling.

Isoelectric focusing electrophoresis estimated the protein’s isoelectric point to be 5.6, consistent with the value of 5.8 predicted from the amino acid sequence.

3.3.3 Characterisation of preferred reaction conditions for HBD

During the process of optimising the reaction conditions for the enzyme assay, the enzyme was found to be affected by the ionic conditions (Figure 13). HBD was 1.5 fold more active in a reaction containing 125 mM NaCl compared to a reaction containing 0 mM NaCl. As the concentration range which showed this effect was significantly higher than the concentration ranges typically observed for specific effects [163], these results suggested a non-specific ionic strength effect enhanced activity.

HBD was also observed to be activated by K⁺ ions. A further 1.5 fold increase in activity was observed when the reaction solution contained 100 mM NaCl plus 25 mM KCl compared to 125 mM NaCl with 0 mM KCl. The effect of K⁺ did not appear to be concentration dependant, as the same magnitude of increase was seen with 125 mM KCl plus no NaCl. This suggested that K⁺ was having some specific effect which enhanced activity. Activation by K⁺ has been observed in another benzaldehyde dehydrogenase [160] and homoisocitrate dehydrogenase [164] and this activation has been proposed to operate through promoting conformational change in yeast aldehyde dehydrogenase [165].

The optimal pH for the enzyme activity was found to be pH 9.3. The activities measured at pH 6.85 and 10.7 were approximately 30% of the activity measured at pH 9.3. Other aldehyde dehydrogenases are known to have catalytically important thiolate groups from cysteine residues in the active site [166], so a higher than neutral pH optimum was not unexpected.
Figure 13: Effect of ionic conditions on the activity of HBD

Reactions were performed in 50 mM Tris-HCl at pH 8.5, with 1 U of HBD
Values are means of triplicate reactions, and error bars depict standard deviations
The recombinant 4-hydroxy benzaldehyde dehydrogenase from *A. baylyi* was also characterised in its response to temperature. The Arrhenius plot used to determine this value was linear between 21 °C and 54 °C, and the activation energy of the enzymic reaction was determined to be 14.5 kJ/mol in this range.

HBD was observed to be reasonably stable at the upper range of these temperatures during these experiments. This was explored by pre-incubating the enzyme at 37 °C and 45 °C in the assay buffer containing the substrate vanillin without NAD⁺ for 9 hours (Figure 14). The enzyme displayed 90% and 75% (respectively) of the initial activity after this time period, indicating good thermal stability under the conditions studied. This was greater tolerance for elevated temperatures than would usually be expected for an enzyme from a non-thermophilic microbe. For example, the benzaldehyde dehydrogenase I and II found in another *Acinetobacter* species, *A. calcoaceticus* NCIB 8250, display lower stability with half life's of 520 and 30 minutes at 37°C, respectively [167]. The effect of having the substrate vanillin present during incubation at these elevated temperatures was not studied, so no comment can be made about stability of the enzyme in the absence of substrate, or in the presence of NAD⁺ rather than vanillin.
Figure 14: Stability of HBD after incubation at 37 and 45 °C

Enzyme solutions (6 U/mL) were incubated in a waterbath for measured time intervals before aliquots were withdrawn, and used to initiate standard HBD assay reactions containing 0.6 U HBD

3.3.4 Characterisation of substrate specificity of HBD

Substrate specificity of the enzyme was assessed using LC MS. Compounds determined to be substrates were identified by a reduction of the peak area of the substrate with a measurable appearance of a product, after incubation with the enzyme. The detection limit for the amount of enzyme and incubation time used was estimated to be an activity of ten fold lower than that shown against vanillin. However, it should be noted that this estimation assumed that the reaction conditions used, which were optimised for activity against vanillin, would also be the optimal conditions for HBD activity against other substrates.

The majority of the compounds tested failed to satisfy these criteria and therefore were not considered to be good substrates for this enzyme (Figure 15). The compounds found not to be substrates included benzyl alcohol and acetophenone, indicating that HBD did not act on these representative alcohol and ketones in the conditions tested. No activity was observed against 2-phenyl
acetaldehyde, cinnamaldehyde or coniferyl aldehyde, suggesting that the aldehyde group acted on needed to be a benzaldehyde group. Cyclohexane carbaldehyde is a cyclic, non-aromatic analogue to benzaldehyde, which was not a substrate, indicating that an aromatic ring was important for HBD activity. The finding that HBD did not act on syringaldehyde was consistent with the low activity against this substrate of the HBD from *S. paucimobilis* [80], and suggested that more than one methoxy group on the ring of a compound inhibits the action of this enzyme.

*Figure 15: Non substrates for HBD from A. baylyi*
Some of the compounds found to be non-substrates (Figure 15) did not fit with trends in substitution patterns of those compounds found to be substrates (Figure 16). For example, benzaldehyde was found to be a substrate, demonstrating that there was no essential requirement for substitution on the aromatic ring. 2, 4-Dihydroxy benzaldehyde was also acted on, demonstrating that a 2-hydroxyl substituent was tolerated. However, 2-hydroxyl benzaldehyde was not found to be a substrate, which was not consistent with the two former observations, making it seem likely that this compound would be a substrate acted on at less than 10% of the rate of vanillin.

The compounds which were found to be substrates were benzaldehyde, 3- and 4-hydroxy benzaldehydes, 3-methoxy benzaldehyde, 3,4- and 2, 4- dihydroxy benzaldehydes, vanillin, and isovanillin (3-hydroxy-4-methoxy benzaldehyde) (Figure 16). HBD acted only on benzaldehydes and acted on more substrates with substitutions at the 3 or 4 ring positions compared to the 2 position. Hydroxy substitution was preferred to methoxy substitution. The products of the enzymatic reaction for each benzaldehyde substrate gave the same LC retention time, UV and mass spectra as an authentic standard of the corresponding benzoates.

![Chemical structures of substrates](image)

Figure 16: Substrates for HBD from A. baylyi

Qualitatively, the enzyme’s substrate specificity pattern was consistent with 3-hydroxy benzaldehyde: NAD$^+$ oxidoreductase (1.2.1.64), with activity against
benzaldehyde, benzaldehydes with 3 or 4 hydroxyl substituents and vanillin. However, 3-hydroxy benzaldehyde: NAD\(^+\) oxidoreductase was also reported to show low activity against 2-hydroxy benzaldehyde [168] while activity against this substrate was not detected in this study. When comparing HBD with benzaldehyde: NAD\(^+\) oxidoreductases (1.2.1.28) from A. calcoaceticus, it was noted that the latter were reported to have broader substrate specificity, with activity evident against cyclic non-aromatic substrates (such as perillaldehyde) and halogenated benzaldehydes [20]. No HBD activity was detected against 4-chloro benzaldehyde or cyclohexane carboxaldehyde. These comparisons indicated that the enzyme’s substrate specificity is more comparable to 3-hydroxy benzaldehyde: NAD\(^+\) oxidoreductase (1.2.1.64) than benzaldehyde: NAD\(^+\) oxidoreductases (1.2.1.28).

To assess the reversibility of the reaction, 4-hydroxy, 3,4-dihydroxy, 3-methoxy-4-hydroxy or unsubstituted benzoates (1 mM) were incubated for 15 minutes with 5 mM NADH and 1.8 \times 10^{-2} \text{ units of 4-hydroxy benzaldehyde dehydrogenase}. No change was observed in the concentrations of these acids, nor was production of the corresponding aldehyde detected, indicating the enzymatic reaction was irreversible against these substrates under the conditions tested. The irreversible nature of the reaction was consistent with studies on other aldehyde dehydrogenases [169].

To examine the cofactor preference of the enzyme, 5 mM of NAD\(^+\), NADP\(^+\), or FAD was incubated for 5 minutes with 0.1 mM vanillin and 1.8 \times 10^{-2} \text{ U HBD}. Activity was detected only in the reaction containing NAD\(^+\). When the enzyme concentration was increased ten fold to 0.18 U, enzymic oxidation of vanillin in samples containing NADP\(^+\) was also detected, at a rate approximately 200 fold lower than observed with NAD\(^+\). FAD was not observed to elicit any activity, as reactions containing this cofactor showed no decrease in substrate nor formation of product.

3.3.5 Characterisation of kinetics of HBD

The apparent kinetic constants (k\(^{\text{app}}\)\(_{\text{cat}}\) and K\(^{\text{app}}\)\(_{m}\)) for some substrates were estimated. In these studies, the enzyme was observed to be substrate inhibited
by concentrations of vanillin above 1 mM. Substrate inhibition has been a common observation in benzaldehyde dehydrogenases [159, 160]. This placed a limitation on the substrate concentrations that could be used to estimate the apparent kinetic constants. However, data at substrate concentrations high enough to approach $V_{\text{max}}$ (eg 10 x $K_{\text{app},m}$) were still consistent with the Michaelis Menten equation, allowing for direct regression fitting to be used for data processing [153].

The estimates of kinetic constants were used to quantitate the substrate preferences of HBD. The 4-hydroxy substituted substrates chosen were considered “natural” substrates, as they would be found as metabolites of 4-hydroxy cinnamic acid catabolism in A. baylyi [58]. Unsubstituted benzaldehyde was included to give an indication of the importance of the 4-hydroxy substitution (Table 2).

**Table 7: Apparent kinetic constants $K_{\text{app},m}$ and $k_{\text{app},\text{cat}}$ for HBD**

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_{\text{app},m}$ (µM)</th>
<th>$k_{\text{app},\text{cat}}$ (s⁻¹)</th>
<th>$k_{\text{app},\text{cat}} / K_{\text{app},m}$ (s⁻¹.M⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzaldehyde</td>
<td>29 (1.5)</td>
<td>4.6 (0.1)</td>
<td>1.6 x 10⁵</td>
</tr>
<tr>
<td>vanillin</td>
<td>44 (5.6)</td>
<td>18 (0.8)</td>
<td>4.1 x 10⁵</td>
</tr>
<tr>
<td>4 - hydroxy benzaldehyde</td>
<td>6 (1.8)</td>
<td>4.0 (0.4)</td>
<td>6.6 x 10⁵</td>
</tr>
<tr>
<td>3,4 – dihydroxy benzaldehyde</td>
<td>37 (2.1)</td>
<td>25 (0.1)</td>
<td>6.8 x 10⁵</td>
</tr>
<tr>
<td>NAD⁺</td>
<td>600 (160)</td>
<td>14 (1.0)</td>
<td>2.3 x 10⁵</td>
</tr>
</tbody>
</table>

a The values for the benzaldehydes were measured at 5 mM NAD⁺, and those for NAD⁺ were measured at 100 µM vanillin. The standard error of these values is given in parenthesis.

The value of $k_{\text{app},\text{cat}} / K_{\text{app},m}$ is generally accepted as a “specificity constant” for comparing the efficiency of catalysis against different substrates [170]. A higher
specificity constant indicates higher efficiency against that substrate. The $k_{\text{app}}^\text{cat}$/$K_{\text{app}}$ value for all substrates tested were within the same order of magnitude. However, the $k_{\text{app}}^\text{cat}$/$K_{\text{app}}$ values showed that catalysis was more efficient with the 4-hydroxy substituted benzaldehydes compared to the unsubstituted benzaldehyde. The data also showed that 4-hydroxy and 3,4-dihydroxy benzaldehyde were preferred substrates over vanillin.

The order of the substrate preference for HBD may have reflected the ease with which *A. baylyi* could catabolise the product of each reaction. The highest $k_{\text{app}}^\text{cat}$/$K_{\text{app}}$ value was found for 3,4-dihydroxy benzaldehyde ($6.8 \times 10^5$ s$^{-1}$·M$^{-1}$). This substrate was converted directly to the ring cleavage substrate, 3,4–dihydroxy benzoate. This was higher than for 4-hydroxy benzaldehyde ($6.6 \times 10^5$ s$^{-1}$·M$^{-1}$) and 4-hydroxy 3-methoxy benzaldehyde ($4.1 \times 10^5$ s$^{-1}$·M$^{-1}$), the HBD products of which would both require one further enzymatic transformation to form the ring cleavage substrate. Benzaldehyde had the lowest $k_{\text{app}}^\text{cat}$/$K_{\text{app}}$ ($1.6 \times 10^5$ s$^{-1}$·M$^{-1}$). The benzoate produced from this substrate would require two further metabolic steps to be transformed to the ring cleavage substrate. This pattern may reflect a preference evolved in the enzyme to help channel flux through the metabolites which require the least protein synthesis to be catabolised.

It can also be seen that HBD had lower $K_{\text{app}}$ values for all benzaldehyde substrates examined than for NAD$^+$. This may reflect the physiological need of the bacteria to maintain the metabolic pool of benzaldehydes at a lower level than that required for NAD$^+$, as has been noted in other reports [159]. Indeed, all of the values of the $k_{\text{app}}^\text{cat}$/$K_{\text{app}}$ were reasonably high. This high efficiency could be expected of an enzyme acting on a group of substrates known to exhibit toxicity, such as benzaldehydes [171].

The nomenclature of HBD was not clear in the literature. The enzyme had been called variously vanillin dehydrogenase [10], 4-hydroxy benzaldehyde dehydrogenase (NCBI accession number AAP78946), and benzaldehyde dehydrogenase (NCBI accession number YP046389). As the enzyme preferred 4-hydroxy benzaldehyde to either benzaldehyde or vanillin, the name 4-hydroxy benzaldehyde dehydrogenase has been employed here. The substrate specificity pattern of HBD (Section 3.3.4) was also consistent with enzymes of
the class “4-hydroxy benzaldehyde dehydrogenase”, E.C. 1.2.1.64. However, differences between HBD and the enzymes classified as E.C. 1.2.1.64 existed. 4-hydroxy benzaldehyde dehydrogenase from *P. putida* [172] showed only 37% amino acid sequence identity to HBD from *A. baylyi* and also preferred NADP⁺ over NAD⁺. The enzyme classified as E.C. 1.2.1.64 from *Pseudomonas mendocina* had an absolute requirement for NADP⁺ [173]. These differences suggest that the enzyme referred to as HBD here would be best classified separately from E.C. 1.2.1.64.

### 3.4 Summary and Conclusions

HBD was cloned from *A. baylyi*, expressed with a hexahistidine affinity tag and purified to electrophoretic homogeneity by IMAC. It was found to be homodimeric. Its optimum pH was 9.3 and it was activated by K⁺. It preferred NAD⁺ to NADP⁺, and acted on a number of benzaldehydes. Quantitation of substrate specificity found that the enzyme had a preference for 4-hydroxy and 3,4-dihydroxy benzaldehyde over vanillin and benzaldehyde.

Class 3 Aldehyde Dehydrogenase family members generally exist as homodimers and prefer NAD⁺ to NADP⁺ [157] and are activated by monovalent cations [20]. The enzyme cloned from *A. baylyi* showed broader substrate specificity than is generally typical of Class 3 aldehyde dehydrogenases [157], but this broad substrate specificity has been reported for benzaldehyde dehydrogenase I and II from *A. calcoaceticus* [19, 20].

It is also worth noting that this enzyme had characteristics that would prove useful in industrial application. It was a highly active enzyme, with reasonably high efficiency, indicated by high $k_{\text{cat}}^{\text{app}} / K_m^{\text{app}}$ for several substrates. HBD could find application in the conversion of a number of benzaldehydes, as substrate specificity studies showed it acted on variously substituted benzaldehydes. Another useful characteristic of this enzyme was its observed thermostability. Remarkably, only 25% of the initial activity was lost after 9 hours at 45 °C.

The high activity and stability of HBD were also positive characteristics for its employment in a synthetic pathway. Perhaps more important than these factors
were the low $K_{\text{app}}$ measured for vanillin (44 μM), and the irreversibility of the reaction. The irreversibility of the reaction meant that controlling the relative concentrations of substrates and products was not necessary for the pathway to be theoretically feasible.

HBD was observed to be substrate inhibited by vanillin concentrations greater than 1 mM. This was a negative characteristic for the use of this enzyme in a synthetic pathway, if high productivities are desired in a batch process. This problem of high substrate concentrations inhibiting activity and lower productivities could perhaps be engineered around, if the biotransformation was performed in continuous culture.

### 3.5 Further work

Further work is required in the characterising HBD to better understand its behaviour as part of the synthetic vanillin to 3CM pathway. It would be prudent to estimate the kinetic constants of the enzyme under conditions more comparable to physiological conditions (eg pH 7.5). It would also be desirable to examine if there was any inhibitory effects caused by the other metabolites in the pathway.

Much further research could be carried out on HBD for academic purposes. This could include further investigation of the thermal stability of the enzyme, and studying the reaction mechanism.
4 Vanillate Monooxygenase

4.1 Introduction

4.1.1 Literature review and background to VMO

Vanillate monooxygenase (VMO) catalyses the conversion of the methoxy group of vanillate to a hydroxyl group, yielding 3,4-dihydroxy benzoate (Figure 5, Section 1.3.5). It also employs NAD(P)H and O$_2$ as substrates, and gives NAD(P)$^+$ and formaldehyde as products. VMO is composed of two components, VanA and VanB.

The VanB subunit is a reductase. Analysis of its amino acid sequence using the Interpro Database [174] showed homology to known motifs including a NAD(P)H binding site, an enzyme bound flavin mononucleotide moiety and a [2Fe-2S] ferrodoxin centre. These structural features suggested that this subunit takes electrons from the nicotinamide nucleotide cofactor, and transfers them along an electron transport chain, before passing them to the oxygenase subunit (Figure 17).

The VanA encoded component is an oxygenase. Homology to known motifs was again found using the Interpro Database [174]. These motifs include a Rieske [2Fe-2S] domain, and a non-heme Fe$^{2+}$ binding site. These structural features suggested that the Rieske cluster accepts and transfers electrons from the reductase to the non-heme iron centre, which can then activate the O$_2$ towards attack on the ring system (Figure 17).
Figure 17: Proposed catalytic cycle for VMO.
Based on [175]

These structural elements make VMO similar to other non-heme iron aromatic ring hydroxylating enzymes, and VMO has been classified to fall within the phthalate family of this enzyme class [175]. The crystal structure of the oxygenase subunit of a member the phthalate family, 2-oxoquinoline 8-monooxygenase, has been solved [176]. This study showed that the enzyme’s conformation changed depending on whether the Rieske cluster was reduced or oxidised. The oxoquinoline substrate bound directly over the Fe(II), and sterically filled the active site such that there was no path for O₂ to enter. Upon reduction of the Rieske centre, the active site opened, allowing O₂ to move into the active site. This mechanism allowed control over the order of binding of substrates. It was postulated that this was necessary to limit the formation of destructive reactive oxygen species.

As discussed in (Section 1.3.5) VMO has been cloned from many microbes [54, 83, 85-87]. The genes encoding the components of VMO in A. baylyi are transcribed as an operon, with the coding regions for VanA and VanB separated by only two nucleotides. These genes have been previously functionally expressed in E. coli as a transcript from a single promoter [86], so it was assumed that the appropriate RNA structure for translation of the second open reading frame (ORF) is recognised by this host. While the stoichiometry of the reductase to the oxygenase of other non heme aromatic ring hydroxylating oxygenases is typically 1:1, other arrangements are also known [177], so that one reductase molecule may optimally provide electrons for multiple
oxygenases molecules. The stoichiometry of VanB reductase to VanA oxygenase in the VMO holoenzyme had not been reported. Therefore it was not known if expressing both subunits as the \textit{vanAB} cistron in \textit{E. coli} would give appropriate levels of expression of both subunits to achieve optimal activity.

It has been repeatedly stated that the enzyme was inactivated by exposure to atmospheric O$_2$ [54, 59, 83, 86, 91, 93-96]. In 1967 it was observed that VMO maintained under N$_2$ remained active for longer than those exposed to atmospheric O$_2$. No report since has described study of this inactivation. It was an aim of the work performed on VMO that its stability be quantitated, and various buffer conditions assessed for stabilising effects. It was uncertain before studying this enzyme whether the synthetic pathway would be best assembled \textit{in vitro} or \textit{in vivo}. Combining purified enzymes in a tube would certainly have offered a less complex model system for observing the behaviour of this synthetic pathway, if VMO could be stabilised. It was also important to understand and perhaps extend the stability of VMO for its purification and characterisation.

Although the purification of similar multi component oxygenase systems has been previously report (eg [181]), there has been only one report of the purification of a VMO [90]. This study expressed the two components with hexahistidine affinity tags in separate recombinant hosts, and purified them separately, before recombining the two components to form active enzyme. This study did not report the characterisation of many of the aspects of VMO significant to its use in a synthetic pathway, such as kinetic constants, or the reversibility of the reaction.

As will be discussed in Section 4.3.2, the strategy of purifying the different components of VMO separately which was employed in this literature study was not employed here. Instead, the purification strategy used for VMO was to simultaneously purify the two components without affinity tags. This strategy has been used in many other studies (eg [182]), and systems have been developed for discovering unknown proteins based on the co-purification of all subunits of multi component [183]. As such, the strategy of purifying VMO as an intact multi component enzyme system was not unprecedented. Indeed, this
approach had the advantage of simplicity over the strategy of separate purification of the different components, which would have involved generating multiple expression constructs, developing multiple activity assays and chromatography methods.

The protein to protein interaction required to gain the proximity for electron transfer from the reductase to the oxygenase occur may well be dependant on the concentration of the two components [94].

No report was found which described the reversibility of the reaction catalysed by VMO.

The specificity of VMO for its aromatic substrate has been reported as reasonably broad [86, 91]. It has been generalised that VMO’s from *Pseudomonas* species act on benzoates with a methyl or methoxy group in the 3 position on the ring [91]. The substrate range of VMO from *A. baylyi* was studied in whole cell systems of resting cells of both *A. baylyi* and a recombinant *E. coli* expressing the enzyme [86]. This range was generally consistent with what had been reported for the *Pseudomonas* species. However, activity against syringaldehyde was not found and this was not consistent with substrate structural patterns. As this report used whole cell systems for examining substrate range, assuming that both substrate and product can cross the cell membrane, the finding that syringaldehyde was not a substrate needed confirmation.

Both NADH and NADPH serve as cofactors for known VMOs [90-92]. Buswell and Ribbons reported that NADH and NADPH elicit similar activity from VMO in the crude extracts of several different *Pseudomonas* species [91]. The authors qualified their observation by stating that transhydrogenation activity in these crude preparations could account for the observed activity with both NADH and NAD(P)H. Hibi *et al.* performed a study to examine cofactor preference in vivo [92]. This involved comparing heterologously expressed VMO activity between the standard *E. coli* K-12 host strain, and two mutant K-12 strains. Activity was lower in the mutant host with a limited intracellular pool of NADPH, while VMO activity in a mutant with a raised intracellular pool of NADPH was the same as
in the wild type host. This evidence suggested that VMO had a preference for NADPH. Similarly to the report of the substrate range examined in a whole cell system, this approach could not control for interfering intracellular processes, and again did not give definitive information about VMO. Providenti et al stated that purified VMO gave equivalent activity against NADH and NADPH, without quantifying this observation [90]. More quantitative data would be useful in demonstrating which, if any, redox cofactor is preferentially acted on. This information would be significant for understanding VMO's behaviour in the synthetic pathway. If VMO acted on the same form of redox cofactor used by HBD (e.g., non-phosphorylated NAD), there would be stoichiometric cofactor regeneration within the pathway.

4.1.2 Aims and approach for VMO

To enable rational employment of VMO within a synthetic vanillin to 3CM pathway, more characterisation of VMO was required. Most importantly, estimation of kinetic constants for the enzyme was required. To estimate these parameters, it was necessary to purify VMO, which required better understanding of the stability of the enzyme. Studying the stability of VMO was also seen as important in the characterisation of the enzyme, as knowledge gained on the nature of VMO’s stability could be useful in employing VMO as part of a synthetic pathway.
4.2 Materials and Methods

4.2.1 Cloning the vanAB genes encoding VMO

As was the case for HBD, the genome sequence of *A. baylyi* [148] was used to design primers for amplification of vanAB, incorporating recognition sites for the restriction enzymes *BglII* and *XhoI* (PCR vanAB forward and PCR vanAB reverse, Table 4, Section 2.2.1). The PCR conditions used were: “hot start” at 98°C for 30 s; 30 cycles of 98°C for 10s, 52°C for 30s, and 72°C for 60s; final extension at 72°C for 500s. The reaction contained 1.5 mM MgCl₂. The PCR product was cloned into pET-Duet1 between the *BglII* and *XhoI* restriction sites using standard techniques [147]. The recombinant plasmid was called pET-VMO (*Figure 18*).

**Figure 18: Map of the recombinant plasmid pET-VMO**

The sequence of pET-VMO was determined using the primers Novagen pET upstream sequencing primer, Novagen T7 terminator sequencing primer and *vanB* sequencing primer (*Table 5*, Section 2.2.4). It was found to be as expected (Appendix 9.2).
4.2.2 Stabilisation of VMO

Cells were lysed as per the general method (Section 2.3.2) into buffer containing 50 mM Tris-HCl, pH 7.5, 10% glycerol, and 1 mM dithiothreitol (DTT). To these lysates were added aliquots of concentrates of putative stabilising compounds. Unless otherwise stated, the final concentration of additive was 1 mM. Lysates were incubated in an ice bath, sampled over time, and assayed for VMO activity.

A curve described by the exponential decay equation (Figure 19) was regression fit to the activity / incubation time data, using the curve fitting software XLfit 4.1 (ID Business Solutions, Guildford, U.K.). The half life of the enzyme was then calculated dividing the natural logarithm of 2 by the decay constant estimated by the fitting software.

\[ y = A \cdot e^{-B \cdot x} \]

Where:
- \( A \) = Initial activity
- \( e \) = Base of the natural logarithm
- \( B \) = Decay constant

*Figure 19: Exponential decay function used in regression analysis of the stabilisation of VMO*

4.2.3 Purification of VMO

All chromatographic media and columns were obtained from GE Life Sciences (Rydalmere, NSW). Chromatography was performed at 5 (±1) °C.

4.2.3.1 Preparation of cell extracts for VMO

Cell lysate was prepared by expressing VMO, and lysing cells in 20 mM Tris-HCl, pH 7.5, 1 mM DTT, 10% (v/v) glycerol, 1 mg/mL MgSO\(_4\).7H\(_2\)O, and 0.1 mg/mL deoxyribose nuclease as per the general method (Section 2.3.2).
4.2.3.2 Ion exchange chromatography

For method development of ion exchange chromatography (IEX), pre-packed Q Sepharose Fast Flow 1 mL columns were used. For preparative purification, Sepharose Q Fast Flow media was packed into XK 16 columns with 40 mm bed height. All steps were performed at 75 cm/hour. Unbound protein was washed through the column with “VMO IEX Buffer” (20mM Tris-HCl, pH 7.5, 1 mM DTT, 10% (v/v) glycerol). A wash step was included with mobile phase of VMO IEX Buffer plus 100mM NaCl, before eluting VMO activity with VMO IEX Buffer plus 250mM NaCl.

4.2.3.3 Hydrophobic interaction chromatography

The hydrophobic interaction chromatography (HIC) column was an XK 16 column, packed with Phenyl-Sepharose (high substitution) to 10 mm bed height. All steps were performed at 50 cm/hour. Unbound protein was washed through the column with “VMO HIC Buffer” (50mM Tris-HCl, pH 7.5, 1 mM DTT, 10% (v/v) glycerol, 250 mM NaCl) plus 250mM AmSO4 ((NH4)2SO4). A wash step was included with mobile phase of VMO HIC Buffer plus 150mM AmSO4, before eluting VMO activity with VMO HIC Buffer.

4.2.3.4 Ultra filtration

Ultra Filtration (UF) was performed with 30 kDa molecular weight cut-off units (Pall Nanosep, product number OD030C33).

4.2.4 VMO steady state activity assays

Where possible, continuous activity assays (Section 2.6.2) were used to assay VMO activity. This assay monitored the change in absorbance at 340 nm, measuring the oxidation of NADH over time [91]. The reaction solution contained 50 mM Tris-HCl, pH 7.5 and 25 mM NaCl in 1 mL total volume. Unless otherwise stated, the substrates vanillate and NADH were included at concentrations of 1 and 0.15 mM, respectively. Activity was measured at 22 (±0.5)°C. The rate measured was calculated using an extinction coefficient for NADH of 6.22 mM−1.cm−1 [178]. One unit of VMO activity was defined as the amount of enzyme which oxidised 1 μmol of NADH per minute in these
conditions. Steady state continuous assays used for estimating kinetic constants contained $2 \times 10^{-4}$ units of VMO.

Discontinuous HPLC assays (Section 2.6.3) were used when the continuous spectrophotometric method was interfered with. The reaction conditions for this assay were the same as for the continuous assay. One unit of VMO activity was defined as the amount of enzyme which produced 1 μmol of 3,4-dihydroxy benzoate per minute in these conditions. The discontinuous assay was used in experiments where Fe(II) was present. Fe(II) autoxidatises to Fe(III) [179]. Fe(III) absorbs more strongly at 340 nm than Fe(II), so the autoxidation process interfered with measurement of VMO activity by NADH oxidation.

Discontinuous assays were also used for studying inhibition with $2 \times 10^{-4}$ units of VMO. Potential inhibitors were added to a final concentration of 1 mM, unless otherwise stated.

Reactions for examining VMO substrate specificity were monitored by the discontinuous assay, performed with aromatic substrates and cofactors at 1mM and $2 \times 10^{-3}$ units of VMO. When examining the reversibility of the reaction the oxidised form of NAD$^+$, and analytical grade formaldehyde (Rhône Poulenc) were used. These solutions were incubated for 90 minutes at 22 (±0.5) °C before the reactions were stopped and the mixtures were analysed by LC MS, as per the general methods (Section 2.6.5).
4.3 Results and Discussion

4.3.1 Stabilisation of VMO

Preliminary experiments to investigate VMO stability involved generating lysate in a phosphate based buffer. As will be discussed more fully below, it became apparent that phosphate may have been reducing the half life of active VMO. Both phosphate [59, 91] and Tris-HCl [93, 94] buffers have been previously used to assay the activity of VMO, so a Tris-HCl based buffer was adopted for lysis when it appeared phosphate was not innocuous. The stabilisation of VMO was assessed by generating lysate in this Tris-HCl based buffer at pH 7.5, adding various additives, and assaying for activity over time. The activity measured plotted against the incubation time gave half life of the enzyme in those conditions (Figure 20).

The untreated sample had a half life of 5 hours 20 minutes. In comparison, the purified VMO from Comamonas testosterone reported no loss of activity over 6 hours, but complete inactivation after 24 hours at pH 8.0 in phosphate buffer [90]. Crude VMO from a Pseudomonas fluorescens species lost only 46% of its activity after 24 hours at pH 7.8 in phosphate [59]. This indicated that A. baylyi VMO was less stable in vitro than these two other VMO’s.

The comparison of stability of VMO from different species was made between enzymes in different buffer conditions. The literature reports of stability of VMO from the Comamonas and Pseudomonas species were made on enzymes in phosphate buffer, which destabilised A. baylyi VMO. The sample to which phosphate was added had a half life of 1 hour, 14 minutes (Figure 20). This was approximately only 25% of the length of untreated sample half life, demonstrating that phosphate accelerated the inactivation of VMO. If the VMO from the Comamonas and Pseudomonas species were also detrimentally affected by phosphate, they could be further stabilised and made more convenient to work with by simply using a different buffering agent.

The half life of VMO from A. baylyi in buffers to be used during purification was also assessed. 250 mM NaCl plus 250 mM AmSO₄ added to the lysate gave a
VMO half life of 5 hours 28 minutes, which was similar to the untreated sample. The sample with 250 mM NaCl had a half life of 7 hours 23 minutes, significantly higher than the untreated sample. The significance of these results was that they suggested that the half life of VMO was long enough that active VMO could be recovered from a purification process using these conditions. However, they were not long enough to prevent significant VMO inactivation during purification.

Other additives trialled for stabilising effect on VMO activity included vanillate and FeSO$_4$, and Na$_2$SO$_4$ (Figure 20). Interestingly, in experiments performed prior to identifying the detrimental effect of phosphate on VMO stability, adding vanillate plus FeSO$_4$ to VMO in a phosphate based buffer had a stabilising effect. While this stabilisation in phosphate buffer could have been advantageous, they were negated by problems these conditions caused. Fe(II) was not soluble in phosphate, and the precipitate formed could not be loaded into chromatography systems. The presence of substrate (eg vanillate) interfered with activity assays. However, in Tris-HCl based buffer, these additives combined gave a significantly lower half life than untreated VMO, but no significant difference when vanillate or FeSO$_4$ were added separately. Na$_2$SO$_4$ was included in the experiment to control for any effect conveyed by the anion SO$_4^{2-}$. The Na$_2$SO$_4$ sample did not give a significantly different half life to FeSO$_4$, indicating that any effect was due to the cation.

Finally, the additives 5 mM DTT and the serine protease inhibitor phenyl methane sulphonyl fluoride (PMSF) were also examined. Increased concentration of the reducing agent DTT did not increase the half life significantly, indicating that use of mild reducing agents did not stop oxidative inactivation. Similarly, including PMSF did not significantly alter the half life, indicating that proteolysis by serine proteases did not contribute to inactivation.
Figure 20: **Half life’s calculated for VMO incubated on ice with various additives.**
Additives with a final concentration of 1 mM unless stated otherwise were combined with lysate containing 1.5 U/mL of VMO in 50 mM Tris-HCl, pH 7.5, 10% glycerol, and 1mM DTT. Aliquots of lysate were withdrawn at measured times, and assayed for remaining VMO activity by HPLC. Values are means of triplicate reactions, and error bars depict standard deviations.
It was concluded that the enzyme had limited stability, and this meant that a heterogeneous preparation of VMO containing a mixture of active and inactive enzyme would result from purification. However, the length of the half life in solution was sufficient for activity to be recovered from a purification process. Work on purification was undertaken, as there were more advantages seen to studying a purified enzyme preparation than disadvantages envisaged in using a heterogeneous preparation of VMO. However, the short half life of VMO also meant that protein concentration could not be used to estimate enzyme concentration. In studying the kinetics of purified enzymes, the $K_{cat}$ is commonly calculated by dividing the $V_{max}$ observed by the molar concentration of the enzyme. This concentration is typically measured by assaying an enzyme preparation of known purity for protein concentration. The estimation of the $k_{cat}$ could not be accurately made by this means, as the relative amounts of active and inactive enzyme could not be accurately estimated.

The short half life of VMO also meant that any synthetic vanillin to 3CM pathway assembled *in vitro* would have a rapidly decreasing concentration of VMO. The major advantage to assembling the synthetic pathway *in vitro* would be simplicity, but VMO’s inactivation would complicate the behaviour of the pathway, and negate this advantage.

### 4.3.2 Method development for purification of VMO

A number of ways that VMO activity could be lost during purification were reported in the literature [91]. These included oxidative inactivation, inactivation due to dilution and the two components becoming separated during chromatography due to different behaviour. It was therefore necessary to develop a purification strategy which was rapid, maintained as high a concentration of VMO as possible, using conditions that did not separate the subunits.

The use of affinity tags was avoided as they could have caused complications later in the co-expression of all three enzymes within a whole cell host. For example, if a hexahistidine affinity tag was added to both subunits, as had been previous reported [90], and these two hexahistidine tagged VMO subunits then
co-expressed with the hexahistidine tagged HBD, the host cell would have to generate 18 extra of histidine units for affinity tags in addition to the 47 coded for in the native sequence, which is 38% more histidine residues. This drain on the metabolism of the host was seen as potentially problematic. Other affinity tags are larger fusion domains than the six amino acid hexahistidine fusion, and would potentially also place a large metabolic drain on the host cell. As such, the expression of VMO with affinity fusion tags was seen as more potentially detrimental to the progress of the project than developing a purification strategy which was not based on affinity tags.

Ion exchange (IEX) was developed as the first step in the purification process, using 1 mL pre-packed columns. VMO activity bound effectively to the strong cation exchanger Sepharose Q in mild conditions (20mM Tris-HCl, pH 7.5, 1 mM DTT, 10% (v/v) glycerol). When lysate containing 53mg of protein was loaded onto a 7 x 25 mm column containing 1mL of media at 80 cm / hour, enzyme activity was not detected in the eluent.

IEX elution was initially explored using a gradient of increasing NaCl concentration. When VMO was loaded at 0 M NaCl, and a linear gradient was applied over 20 column volumes to 1 M NaCl, only 10% of activity was recovered. The fraction which contained the most concentrated VMO activity had 0.05 U / mL VMO activity, compared to 0.67 U / mL in the lysate. It appeared that dilution of VMO during chromatography was limiting recoveries. Gradient elution was abandoned, though this strategy can achieve high resolution [180].

A strategy of step elution was then explored. This allowed greater VMO concentration in the eluent and was more rapid than gradient elution. Both of these factors helped limit VMO inactivation. During development of a step elution method, recoveries of over 70% of the total activity loaded were achieved. It was also notable that when VMO was loaded at 0 mM NaCl and eluted with 250mM NaCl, the eluent had the same concentration of enzyme activity as was found in the cell lysate (0.94 U / mL).
However, the performance of this purification step was not sufficient to stand alone, with only approximately 3 fold increases in specific activity typically achieved. Hydrophobic interaction chromatography (HIC) was investigated as a second chromatography step again using step elution.

To optimise the performance of the HIC, experiments were performed to ascertain the lowest ionic strength that allowed VMO binding to the HIC resin used. Adding AmSO₄ to a concentration of 250 mM in the buffer used to elute VMO from the IEX column (250mM NaCl, 20mM Tris-HCl, pH 7.5, 10% glycerol, 1 mM DTT), no VMO activity was detected in the material that flowed unretained through the column. VMO activity eluted with 0 mM AmSO₄ in the above buffer, and not at the next higher concentration tested with step elution, 50 mM AmSO₄.

Ultrafiltration (UF) was used as the final step in the purification of VMO. This technique allowed rapid sample concentration, and also size based separation to remove low molecular weight contaminants. It was found that all of the VMO activity loaded was retained by 5 and 30 kDa molecular weight cut off (MWCO) units, and that 20 (±3) % of VMO activity loaded was retained in a unit with a 100 kDa MWCO. The 30K MWCO units were adopted.

4.3.3 Purification of VMO

The method developed for purification of VMO was scaled up for preparing VMO for characterisation. The instability of VMO meant that the metrics typically used to follow enzyme purification, such as fold purification and percentage recovery were not useful for measuring the efficacy of the purification. Specific activity values decreased with the steps after the IEX in the purification (
Table 8), despite SDS-PAGE analysis showing that VMO was being enriched by these steps (Figure 21). It was concluded that this trend of specific activity decreasing with purification steps reflected inactivation of the enzyme during purification.
Table 8: Results from one preparative VMO purification

<table>
<thead>
<tr>
<th>Sample</th>
<th>Total activity (U)</th>
<th>Recovery (%)</th>
<th>[Protein] (mg/mL)</th>
<th>Specific activity (U / mg)</th>
<th>Fold purification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysate</td>
<td>16</td>
<td>100</td>
<td>19.8</td>
<td>0.078</td>
<td>1</td>
</tr>
<tr>
<td>IEX Target</td>
<td>13</td>
<td>83.4</td>
<td>0.81</td>
<td>0.46</td>
<td>5.8</td>
</tr>
<tr>
<td>HIC Target</td>
<td>1.3</td>
<td>8.5</td>
<td>0.20</td>
<td>0.39</td>
<td>4.9</td>
</tr>
<tr>
<td>30K UF Retentate</td>
<td>0.47</td>
<td>3.0</td>
<td>1.4</td>
<td>0.16</td>
<td>2.1</td>
</tr>
</tbody>
</table>

The specific activity of the partially purified VMO (0.16 U / mg) was low. It can be calculated that the specific activity would have still been low (0.43 U / mg), had VMO been purified to homogeneity with no further loss of activity. The specific activity of purified HBD (22 U / mg) (Table 7, Section 3.3.5) and purified P34O (55 U / mg) (Table 9, Section 5.3.5) were orders of magnitude higher. The low specific activity of VMO may be typical of non-heme monooxygenases of the phthalate dioxygenase family. The specific activity of purified 2-oxoquinoline 8-monooxygenase from *P. putida* (0.97 U / mg) [181] was within an order of magnitude of that of VMO. 2-Oxoquinoline 8-monooxygenase also lost activity during purification, so its measured specific activity would have been lowered by this inactivation.

Due to loss of VMO activity, densitometry of Coomassie stained SDS-PAGE gels was seen as a more representative measure of the efficacy of the purification process than enzyme activity measurements. VMO bands on SDS-PAGE gels (Figure 21) were identified by using specific activity data from chromatography experiments. Bands in the range of expected molecular masses which were more intense in samples that had higher VMO activity were assigned to VanA and VanB. These bands were estimated to have molecular
masses of 43 and 38 kDa, respectively, in reasonable agreement with the calculated values of 41 and 36 kDa, respectively.

**Figure 21:** SDS-PAGE analysis of VMO purification

4 μg of total protein loaded into all lanes except HIC Wash, which contained 0.4 μg of total protein

VMO on Coomassie Blue stained SDS-PAGE gels composed 37% of total protein after ultrafiltration (Figure 21). VMO was not purified to homogeneity. However, it was substantially enriched, with enough activity remaining to allow for the characterisation of the enzyme.

Proceeding with the characterisation of VMO with a preparation which was only partially purified made it likely that reproducibility of the results of the characterisation would be difficult. A key reason for studying enzymes that have been purified to homogeneity is that these preparations are the most highly
defined. Therefore, their behaviour could be expected to be the most highly reproducible. It would be difficult to reproduce the level of purity, and enzyme inactivation, of the partially purified enzyme preparation used here, making it difficult to reproduce the findings of its characterisation. While it would certainly have been desirable to purify VMO to homogeneity, its instability limited the number and length of purification steps possible. It was therefore concluded that further development of the purification process would give marginal benefit to the project, and effort should be focused on the characterisation of the partially purified enzyme.

There was evidence from SDS-PAGE that the oxygenase and reductase components were enriched differently. The ratio of the intensity of the VanA band to the VanB band was approximately 2.5 after partial purification (Lane labelled “UF Retentate”, Figure 21). Lysate samples differed, with a ratio of approximately 5 (Lane labelled “Lysate”, Figure 21), suggesting there was a higher proportion of oxygenase subunit present before chromatography than after. The components of protein complexes often remain intact during purification (eg [182, 183]). Assuming that functional VMO holoenzyme purified as a single unit, and excess components were dissociated from the holoenzyme and separated during chromatography, the difference in component abundance suggested that the amount of VanB, the reductase, expressed was insufficient to fully employ the amount of VanA which had been expressed. If this was true, greater VMO activity could be obtained by increasing the relative expression of VanB.

This hypothesis was supported by the structure of the gene used to express VanB. The open reading frames encoding each component for VMO in A. baylyi were transcribed as an operon, with the coding regions separated by only two nucleotides. As these genes have been previously cloned as a transcriptional fusion, and functionally expressed in E. coli [86], it seemed likely that the DNA structures required for transcription and translation of the second open reading frame were recognised by this host, and functional expression could be achieved by treating the operon as a single gene. Examination of the DNA sequence (Figure 22) of the genes encoding VMO revealed that the region 7 to 10 base pairs upstream of the ATG start codon of VanB did not contain a
sequence with homology to a Shine Dalgarno ribosomal binding sequence, the consensus sequence of which is AGGAGG [184]. The presence and location of a Shine Dalgarno sequence to bind to the ribosome and initiate translation is an important factor in the expression level of a protein [185].

![Diagram of VanA/VanB operon from A. baylyi](image)

**Figure 22: Structure of the VanA / VanB operon from A. baylyi**

Clearly, other factors also impact on the expression levels of activity enzymes including correct folding and cofactor assembly.

### 4.3.4 Preferred reaction conditions for VMO

VMO was characterised with respect to its preferred reaction conditions. Its pH optimum was found to be 7.5 in Tris-HCl buffer, in fair agreement with pH 7.8 found for a *Pseudomonas* VMO [59] and pH 7.5 for 2-oxoquinoline 8-monooxygenase from *P. putida* [181]. NaCl concentrations ranging from 25 mM to 1 M elicited the same VMO activity, which was approximately 2 fold higher than that measured in 20mM Tris-HCl, pH 7.5, without additional NaCl.

The inhibition or activation of the partially purified VMO by various compounds was also studied. Compounds which had some effect on the stability of VMO, such as ferric and ferrous iron, and phosphate were examined. Ethylene diamine tetraacetic (EDTA) was also examined, and was expected to inhibit the activity of this metalloenzyme.
These experiments found that VMO was substantially inhibited by Fe(III) (Figure 23). Reactions containing FeCl₃ were measured to have only 20% of the activity measured in the standard reaction with no inhibitors. Control reactions, containing NaCl with the same concentration of Cl⁻ as the FeCl₃ sample, also gave a lower activity than the untreated sample, but this difference was not significant. There was, however, a significant difference between the reactions containing FeCl₃ and NaCl, with the former displaying only 13% of the activity of control reactions containing NaCl. This demonstrated that the iron species was the cause of the observed inhibition, rather than the chloride anion.

Reactions containing EDTA, phosphate, EDTA plus FeSO₄ and Na₂SO₄ did not show significantly different activity to the sample in the standard reaction solution.

Reactions containing FeSO₄ and phosphate plus FeSO₄ displayed higher activity than the reactions in standard reaction mixture, or control reactions containing the same concentration of SO₄²⁻ as Na₂SO₄, or containing phosphate on its own. Comparison with the latter two control reactions demonstrated that the Fe(II) was responsible for the activation. To ascertain if the enhanced reaction rate was due to chemical demethylation of vanillate catalysed by Fe(II), reaction solutions incubated with no enzyme were analysed. LC MS detected no significant consumption of vanillate, or production of 3,4-dihydroxy benzoate when VMO was omitted. It was concluded that FeSO₄ and phosphate plus FeSO₄ activated VMO.

It was interesting that the effect on VMO activity of Fe depended on its oxidation state. The 2-oxoquinoline 8-monooxygenase from P. putida was also activated by Fe²⁺ [181]. Although there have been no reports of experimental demonstration of the presence of Fe²⁺ in VMO, the presence of this ion within the active site of VMO has been inferred from sequence homology to other non-heme iron aromatic ring hydroxylating enzymes [54, 86, 87]. The reaction mechanism of other non-heme oxygenases involves a cycling of the redox state of the active site Fe²⁺ oxidising to Fe³⁺ during the reaction and then reducing back to Fe²⁺ after the reaction has occurred, before being ready to catalyse another reaction [186]. It was possible that having an excess of Fe²⁺ present in
the reaction buffer facilitated more rapid regeneration of the active site redox state to an Fe$^{2+}$, resulting in a net increase of reaction rate. Inversely, it was possible that inclusion of an excess of Fe$^{3+}$ in the reaction buffer slowed this regeneration process, resulting in a net decrease of reaction rate.
Figure 23: Inhibition and activation of VMO

Values are means of triplicate reactions, and error bars depict standard deviations.
4.3.5 Substrate specificity of VMO

The reversibility of the reaction catalysed by VMO had not been previously reported. This aspect of substrate specificity was important to understanding the behaviour of the enzyme in a pathway.

“Reverse” reaction solutions were prepared containing the products of the forward reaction as substrates (3,4-dihydroxy benzoate, formaldehyde, and NAD\(^+\)). Control reactions containing the substrates of the “forward” reaction were incubated in parallel. In the conditions used, the “forward” reaction proceeded to 82% of completion (Figure 24). In the “reverse” reactions, no consumption of 3,4-dihydroxy benzoate, or production of vanillate was observed. This demonstrated that the reaction was irreversible under these conditions.

The substrate specificity of VMO from *A. baylyi* had previously been examined in a whole cell system [86]. One finding of this study particularly was a little surprising, with 3, 5-dimethoxy-4-hydroxy benzoate (syringate) being found to not serve as a substrate while the structurally similar but sterically bulkier 3,4, 5-trimethoxy benzoate was found to be a substrate. To verify these findings, reactions were prepared containing syringate. Control reactions containing vanillate were incubated in parallel, and the conditions used again resulted in 82% consumption of vanillate, with the accompanying formation of 3,4-dihydroxy benzoate. A similar amount of syringate was measured to be consumed (86%) (Figure 24). If the reaction catalysed by VMO activity on syringate was the conversion of the 3-methoxy group to a 3-hydroxy group, the product would be expected to be 5-methoxy-3,4-dihydroxy benzoate. The product detected by LC MS of the VMO activity on syringate had the correct molecular mass (m/z 183) to be this compound. No authentic standard was available to also match retention time and UV spectrum to confirm that this was the product. These results demonstrated that syringate was a substrate for VMO, and was acted on with similar activity to vanillate. It was possible that syringate could not enter the whole cells to access the enzyme in the whole cell
system reported in the literature [86], or that the product could not leave the cells.

![Chemical structures]

*Figure 24: VMO activity on vanillate, 3,4-dihydroxy benzoate and syringate*

### 4.3.6 Kinetic characterisation of VMO

The apparent Michaelis - Menten kinetic constants for various substrates were estimated. These characteristics for the substrate vanillate were important for understanding the behaviour of VMO in the synthetic vanillin to 3CM pathway.
The substrate specificity of the enzyme was quantified by estimating these constants for 3-methoxy benzoate, vanillate, and syringate. As VMO was not purified to homogeneity, and there was no means for estimating the fraction of active VMO in the purified VMO preparation, the apparent $k_{cat}$ could not be calculated. The ratio of $k_{cat}^{app}$ / $K_{m}^{app}$, (catalytic efficiency) usually used to compare substrate preference for an enzyme, was substituted here by $V_{max}^{app} / K_{m}^{app}$.

The data showed that of the three substrates tested, vanillate gave the highest catalytic efficiency of 7.8 s$^{-1}$ (Table 9), so was the preferred substrate in the conditions used. 3-Methoxy benzoate and syringate had catalytic efficiency values of 1.1 and 0.33 s$^{-1}$, respectively.

The $K_{m}^{app}$ for all these substrates was in the μM range. The $K_{m}^{app}$ for vanillate was 25 μM. A low $K_{m}^{app}$ was a beneficial quality for VMO to have in a synthetic pathway, as the enzyme would perform efficient catalysis at low vanillate concentration. Related monooxygenases also showed low $K_{m}$ values for their aromatic substrates. The phthalate dioxygenase from *Burkholderia cepacia* had a $K_{m}$ of 0.2 μM [187], while 4-hydroxy benzoate 3-monooxygenase from *Acinetobacter calcoaceticus* had a $K_{m}$ of 41 μM for 4-hydroxy benzoate [188].

The data for adenine nucleotide cofactors showed that the enzyme effected comparable catalysis regardless of whether the cofactor was NADH or NADPH, with catalytic efficiency values of 21 and 55 s$^{-1}$, respectively. The $K_{m}^{app}$ for NADPH (3.1 ± 1.5 μM) was approaching the lower limit for estimation with these experimental conditions, and this was reflected in the magnitude of the standard error relative to this value. The uncertainty in the $K_{m}^{app}$ for NADPH meant that the difference between the catalytic efficiency for NADH and NADPH could not be stated to be significant. It was concluded that VMO had no clear preference for either of the cofactors examined in these conditions. The lack of preference of the enzyme for NADH or NADPH found here is in contrast to the suggestion that NADPH was the preferred cofactor *in vivo* [92]. However, the evidence presented by Hibi *et al* for this suggestion was primarily circumstantial, as discussed in Section 4.1.
The low micromolar $K_{app}^{m}$ for both cofactors was a beneficial trait for VMO’s use in a multi-enzyme system. The intracellular NADPH concentration for growing *E. coli* has been estimated at 0.56 mM [189], and 0.18 mM for resting cells [190]. The intracellular NADH concentration for growing *E. coli* has been estimated as ranging between 0.3 and 0.42 mM [191] and estimates of 1.0 mM [192] and 0.54 mM [190] have been reported for resting cells. Although these values are clearly highly dependant on the growth conditions and physiological state of the cells, they are orders of magnitude higher than the $K_{app}^{m}$ estimated for either cofactor. This indicated that low substrate concentration should not limit the reaction catalysed by VMO in a whole cell biocatalyst.

**Table 9: Apparent kinetic constants $K_{app}^{m}$ and $V_{app}^{max}$ for VMO**

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_{app}^{m}$ (μM)</th>
<th>$V_{app}^{max}$ (M.s$^{-1}$)</th>
<th>$V_{app}^{max} / K_{app}^{m}$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 - methoxy 4 – hydroxy benzoate (vanillate)</td>
<td>25 (5.3)</td>
<td>$2.0 \times 10^{-4}$ ($7.9 \times 10^{-6}$)</td>
<td>7.8</td>
</tr>
<tr>
<td>3 - methoxy benzoate</td>
<td>103 (39)</td>
<td>$1.1 \times 10^{-4}$ ($2.0 \times 10^{-5}$)</td>
<td>1.1</td>
</tr>
<tr>
<td>3, 5 – dimethoxy – 4-hydroxyl benzoate (syringate)</td>
<td>263 (83)</td>
<td>$7.9 \times 10^{-5}$ ($2 \times 10^{-5}$)</td>
<td>0.33</td>
</tr>
<tr>
<td>NADH</td>
<td>10 (2.4)</td>
<td>$2.2 \times 10^{-4}$ ($5.1 \times 10^{-6}$)</td>
<td>21</td>
</tr>
<tr>
<td>NADPH</td>
<td>3.1 (1.5)</td>
<td>$1.7 \times 10^{-4}$ ($1.8 \times 10^{-6}$)</td>
<td>55</td>
</tr>
</tbody>
</table>

*The values for the benzoates were measured at 150 μM NADH. The values for the adenine nucleotide cofactors were measured at 1 mM vanillate. The standard error of these values is given in parenthesis.*
4.4 Summary and Conclusions

VMO was cloned from *A. baylyi* and expressed. VMO activity in the crude lysate of these recombinant hosts had a half life of approximately 5 hours in a variety of Tris-HCl based buffers. VMO was partially purified to 37% purity by ion exchange, hydrophobic interaction chromatography and ultrafiltration. The partially purified enzyme had an optimum pH of 7.5, was activated by Fe$^{2+}$ and inhibited by Fe$^{3+}$. No significant preference for NADH or NADPH was found, while the enzyme had a preference for vanillate over 3-methoxy benzoate or syringate.

Many of these characteristics were shared by other monooxygenases. 2-Oxoquinoline 8-monooxygenase was particularly analogous to VMO, displaying similar inactivation, pH optimum, activation by Fe$^{2+}$, and specific activity.

It must be noted that this enzyme was difficult to work with. Its inactivation by exposure to atmospheric O$_2$, dilution and separation of the components illustrated that the enzyme was not robust. Its use in an *in vitro* synthetic pathway would be problematic. It also had relatively low intrinsic activity. It was estimated that the specific activity of this enzyme, had it been purified to homogeneity, would have been approximately 50 to 100 fold lower than HBD and P34O. This meant that it would require significantly higher expression than HBD and P34O to achieve balanced activities in a synthetic pathway.

VMO did however have some characteristics which were advantageous for its use in a synthetic pathway. It had a low $K_{app}$ for vanillate (25 μM), NADH (10 μM) and NADPH (3.1 μM). Its lack of preference for NADH or NADPH meant that it would be expected to utilise both cofactors in a whole cell system. Its activity on NADH would theoretically partially balance the HBD activity on NAD$^+$. Its activity on NADPH could cause an imbalance between the oxidised and reduced form of this cofactor, but the use of NADH was expected to buffer some of this imbalance. The reaction was also found to be irreversible. HBD and P34O also catalysed irreversible reactions. With the knowledge that the synthetic pathway consisted of three irreversible reactions, and at least partial
cofactor recycling, made it possible to conclude that the biotransformation was theoretically likely to occur when the system was assembled.

4.5 Further work

The VMO studied here sufficed for use in the synthetic pathway. However, it is not optimal, and undertaking further work on the enzyme should take this into consideration.

If short term optimisation of the synthetic pathway was required, then it would be prudent to undertake some further characterisation of the VMO from A. baylyi described in this chapter to better understand its behaviour as part of the synthetic vanillin to 3CM pathway. Future work should develop and utilise assays for measuring the two components of VMO separately, to further elucidate the separate roles VanA and VanB play in VMO activity.

No kinetic constants were estimated for the substrate O2. The instrumentation (eg an oxygen electrode) for performing these measurements was not present in the laboratories where the work was performed. No attempts were made to use this instrumentation, as achieving control over the concentration of O2 inside a whole cell was beyond the scope of this project. However, to develop a deeper understanding of the behaviour of the system, it would be beneficial to know at what concentrations O2 became limiting to the velocity of the VMO reaction. It would also be desirable to examine if there was any inhibitory effects caused by the other metabolites in the pathway.

If longer term work is to be carried out on the synthetic pathway, it may be prudent to employ a different VMO. It appeared from comparison of data on the stability of VMO from A. baylyi, a Pseudomonas species and Comamonas testosterone that the enzyme currently used was the least stable of these three. To replace it would require cloning, expression, purification and characterisation of another VMO. While this would be time consuming, it may allow for the more stable enzyme to be purified to homogeneity. It may also be prudent to test the stability of these VMO’s in vivo, after expression in the recombinant E coli system.
Much further research could be carried out on VMO. The electron transport components of VMO could be studied by spectroscopic techniques such as electron paramagnetic resonance (eg [193]). It would be interesting to investigate if these structural features are destroyed during oxidative inactivation.
5 Protocatechuate 3,4-dioxygenase

5.1 Introduction

5.1.1 Literature review and background for P34O

The final reaction in the vanillin to 3CM pathway is catalysed by protocatechuate 3,4-dioxygenase (P34O). This reaction incorporates both atoms of O$_2$ into the aromatic ring of 3,4-dihydroxy benzoate causing its oxidative scission and formation of 3CM (Figure 6, Section 1.3.6).

As discussed in Section 1.3.6, there has been extensive description of P34O in the literature. Crystal structures of P34O from A. baylyi [113], and P. putida [114] have been solved. The enzyme in both species is heterodimeric, and the proteins, known as the α and β chains, are encoded by pcaG and pcaH, respectively. The molecular mass of these proteins from A. baylyi are 23 and 27 kDa, respectively [194].

The mechanism of the enzyme was recently reviewed [115]. The P34O reaction is irreversible [60]. P34O’s are reasonably specific for the substrate 3,4-dihydroxy benzoate [100]. The pH optima of P34O’s from different organisms range from pH 10 for P. putida [106] to pH 5.2 for the enzyme from the flowering shrub Tecoma stans [112]. The enzyme from A. baylyi has a pH optimum of 8.5 [103], which is more representative of the pH optima of P34O from a range of organisms listed on the BRENDA Database [48].

The $K_{\text{app}}^\text{m}$ and $k_{\text{app}}^\text{cat}$ for 3,4-dihydroxy benzoate have been reported as 78 µM and 120s$^{-1}$, respectively, for P34O from A. baylyi expressed in a recombinant system [100]. This was consistent with the $K_{\text{app}}^\text{m}$ reported for 3,4-dihydroxy benzoate (71.4 µM) for a different strain of A. calcoaceticus purified from the wild type organism [103]. The P34O from this strain of A. calcoaceticus was also estimated to have a $K_{\text{app}}^\text{m}$ for O$_2$ of 58.8 µM in the same study. The specific activity for purified P34O was 34 U / mg for the A. baylyi enzyme [195] and 20 U /mg for the A. calcoaceticus enzyme.
The irreversibility, $K_m$ for both substrates and $k_{\text{cat}}$ were all favourable for this P34O being an efficient enzyme in the synthetic vanillin to 3CM pathway.

5.1.2 Aim and approach for P34O

The established knowledge about the properties of the enzyme was sufficient for this project, and there was no need to add to this characterisation. However, work was performed to verify that the enzyme cloned from *A. baylyi* in this project behaved equivalently to the other *Acinetobacter* P34O’s which had been previously characterised. To achieve this, it was necessary to clone the recombinant P34O from *A. baylyi*, express it in *E. coli*, purify and compare the kinetic constants to other well characterised P34O’s.

Purification of P34O from *A. baylyi* has been reported [103, 195]. Of these two reports, the more recent ([195]) achieved higher purity as indicated by specific activity, so this method was used as a template for purification performed here. This protocol involved an ammonium sulphate (AmSO$_4$) precipitation, hydrophobic interaction chromatography (HIC), ion exchange chromatography (IEX) and size exclusion chromatography (SEC).
5.2 Materials and Methods

5.2.1 Cloning the pcaHG genes encoding P34O

Primers for amplification of pcaHG incorporated recognition sites for the restriction enzymes BglII and XhoI (PCR pcaHG forward and PCR pcaHG reverse, Table 4, Section 2.2.1). The PCR conditions used were: “hot start” at 98°C for 30 s; 30 cycles of 98°C for 10 s, 48°C for 30 s, and 72°C for 50 s; final extension at 72°C for 500 s. The reaction contained 2.5 mM MgCl₂. The PCR product was cloned into pET-Duet1 between the BglII and XhoI restriction sites after digestion in NEBuffer 3, using general techniques. The recombinant plasmid was called pET-P34O (Figure 25).

![Diagram of pET-P34O](image)

**Figure 25:** Map of the recombinant plasmid pET-P34O

The sequence of pET-P34O was determined using the primers Novagen pET upstream sequencing primer, Novagen T7 terminator sequencing primer and pcaG sequencing primer (Table 5, Section 2.2.4). It was found to be as expected (Appendix 9.3).
5.2.2 P34O activity assay

To measure P34O activity, a published continuous method was used [103]. This assay measures the decrease in absorbance of the sample solution at 290 nm. As the substrate for this enzyme (3,4-dihydroxy benzoate) absorbs at this wavelength, and the product, 3CM, does not, this decrease in absorbance was used to calculate the rate at which the substrate was consumed, using the millimolar extinction coefficient of 2.31 L . mmol⁻¹ . cm⁻¹ [60]. The cuvette in the front cell of the spectrophotometer contained 50 mM Tris-acetate, pH 7.5, 0.4 mM 3,4-dihydroxy benzoate and enzyme. The cuvette in the rear cell of the spectrophotometer contained all components present in the front cuvette, but no 3,4-dihydroxy benzoate. The temperature, unless otherwise stated was 22 (±0.5) °C. One unit of P34O activity was defined as the amount of enzyme required to consume 1 μmol of 3,4-dihydroxy benzoate per minute, in these conditions.

5.2.3 Purification of P34O

P34O was expressed as per the general method (Section 2.3.1), except where otherwise stated. Lysate was generated in Buffer A (30 mM Tris-HCl, pH 7.3 and 2.5 mM DTT) supplemented to 1.0 mg/mL MgSO₄ and 0.1 mg/mL deoxyribonuclease as per the general method (Section 2.3.2).

All of the four steps of the purification process were performed at 4°C.

Saturated AmSO₄ was added dropwise while mixing using a magnetic stirring bar. Lysate was taken to 30% saturation over 6 hours. The precipitated material was pelleted by centrifugation at 10x 10³g for 20 minutes, and the supernatant recovered. The supernatant was taken to 50% saturation over 2 hours and pelleted by centrifugation at 10x 10³g for 20 minutes. The pellet was recovered and stored overnight at 4°C.

The HIC column was 50 mm long x 16 mm diameter, packed with Phenyl-Sepharose, high substitution (GE Healthcare). Chromatography was performed at 75 cm/hour. The material which precipitated between 30 and 50% saturation with ammonium sulphate was resuspended in a minimal volume of Buffer A plus
1 M AmSO₄, and loaded onto the column. Some method development was performed on this step (Section 5.3.3). The final method involved washing unbound protein through the column with Buffer A plus 1 M AmSO₄. After the baseline returned to zero, the mobile phase was changed to Buffer A plus 0.5 M AmSO₄. After the baseline again returned to zero, a linear gradient from Buffer A plus 0.5 M AmSO₄ to Buffer A with 0 M AmSO₄ was performed over a 50 mL volume (5 column volumes). Fractions containing high P34O specific activity were pooled.

The IEX column was 21 mm long x 16 mm diameter, packed with DEAE-Sepharose (GE Healthcare). Chromatography was performed at 90 cm/hour. The developed method involved performing two passes through this column. Before each, P34O was diluted with Buffer A until the conductivity of this solution measured conductivity equivalent to 0.1M Tris-HCl at pH 7.3 and 4°C. In the first pass, the enzyme was loaded and unbound protein was washed through the column with Buffer A. After the baseline returned to zero, the mobile phase was changed to Buffer A plus 0.15 M NaCl to elute P34O. Fractions containing P34O activity were pooled, and loaded back onto the IEX column with Buffer A. The mobile phase was changed to Buffer A plus 0.1 M NaCl. After the baseline again returned to zero, a linear gradient from Buffer A plus 0.1 M NaCl to Buffer A plus 0.2 M NaCl was performed over 42 mL (10 column volumes) at 20 cm/hour. Fractions containing high P34O specific activity were pooled.

The P34O from IEX was concentrated by ultrafiltration (UF), performed with 100 kDa molecular weight cut-off units (Pall Nanosep, product number OD100C34) to reduce the sample volume below 0.6 mL. This concentrate was loaded onto the SEC column. This column was 60 mm long x 16 mm diameter, packed with Sephadex G-200 (GE Healthcare). Chromatography with this column was performed at 15 cm/hour, using Buffer A as the mobile phase. The P34O activity was collected in the void volume.

P34O from the SEC was stored at -80°C.
5.3 Results and Discussion

5.3.1 Cloning P34O

The two genes which encode the two subunits of P34O (pcaHG) from *A. baylyi* had been reportedly cloned as a 11 kb restriction fragment inserted into pUC18 and expressed in *E. coli* [196]. This strategy was not appropriate for this project because this fragment contained multiple other *A. baylyi* genes which were not required for the vanillin to 3CM pathway. However, the fact that P34O activity was expressed in this construct, implied that messenger RNA (mRNA) was transcribed from the pUC18 lac promoter, and *E. coli* ribosomes recognised the ribosomal binding sites native to *A. baylyi* and translated both of the pcaHG transcripts [194].

The two genes pcaHG coding for the two proteins of the heterodimeric enzyme were amplified by PCR from *A. baylyi* genomic DNA to give a single discistronic DNA molecule. This PCR product was inserted into multiple cloning site II (MCS II) of the pETDuet-1 parent plasmid (Figure 25). This placed the two open reading frames (ORFs) coding for P34O under the control of a single promoter to be transcribed as a discistronic mRNA. This was expected to yield a transcriptional fusion mRNA of the genes encoding P34O (pcaHG), which would then be translated into the two protein products. Seven bases upstream from the start AUG of the second transcribed ORF, pcaG, was the sequence TGGAGT, which was 66% homologous to the consensus Shine Dalgarno sequence AGGAGG, for ribosomal binding of the mRNA in *E. coli* [184], allowing expression within the recombinant *E. coli* BL21(DE3) system.

The DNA of the insert in pET-P34O was sequenced, and confirmed to be as expected (Appendix 9.3). The ORF encoding the β chain, pcaH, had an N terminal addition of the amino acids ADL after the starting M, followed by 100% homology with the sequence reported in the genome sequence of *A. baylyi* [148] due to coding sequence within MCS II. This was expected to not change the enzyme significantly.
5.3.2 Expression of P34O

P34O activity was found in the lysate of *E. coli* BL21 (DE3) expressing P34O, but not of the empty vector control *E. coli* BL21 (DE3) / pETDuet-1. This demonstrated that expression of activity had been achieved.

However, this expression was found to occur at a low level. The specific activity of clarified lysate was measured to be 0.04 U / mg, compared to 0.14 U / mg reported for the enzyme cloned from the same strain of *Acinetobacter* expressed in *E. coli* [196]. Despite low specific activity, there was evidence to suggest that strong expression of P34O was being achieved. Bands corresponding to the calculated molecular masses of the subunits of P34O (23 and 27 kDa) were evident in SDS-PAGE analysis of whole cells after induction of recombinant protein expression, but not before (Figure 26). These bands were not evident in SDS-PAGE analysis of the soluble fraction of the cell free lysate. This suggested that P34O was being expressed in an insoluble form using the standard expression protocol.
Figure 26: The difference between P34O levels in whole cells and soluble lysate after induction at 37°C

To optimise expression of soluble, active P34O, expression was performed in six cultures, with variables of two incubation temperatures after induction (20°C and 37°C) and three concentrations of IPTG (0.05, 0.2 and 1mM) examined in a matrix format. A lower temperature during induction is thought to slow the rate of protein synthesis more than the rate of protein folding, thereby allowing more time for each protein molecule to assume its correctly folded state, reducing the probability of aggregation [197, 198]. Similarly, lowering the concentration of the
inducer, IPTG, can reduce the amount of mRNA, slowing the rate of protein synthesis [197, 199]. The cells for all six cultures were harvested and lysed in a volume of buffer adjusted such that all the suspensions had equal cell densities. The P34O activity and protein concentration of the lysates were measured.

These analyses found that the incubation temperature after induction was more influential than IPTG concentration in achieving greater P34O activity in lysates (Figure 27). The three cultures induced at 20°C, regardless of IPTG concentration, yielded higher specific activity lysates than the three induced at 37°C. Different IPTG concentrations did not give a significant difference between the P34O specific activity in the lysate of the three cultures induced at 20°C. The highest average specific activity measured in this experiment, in the lysate from the cultures incubated at 20°C after induction and induced with 1mM IPTG (0.75 U / mg), was significantly higher than the value of 0.14 U / mg reported for a recombinant P34O cloned from the same strain of Acinetobacter and expressed in E. coli [196]. The specific activity observed from this experiment was considered sufficient, and no further optimisation of expression was performed.
Figure 27: Effect of temperature and IPTG concentration on recombinant expression of soluble active P34O. Cultures were grown as per the standard expression conditions until they reached OD600 of 1.0. IPTG was then added, and incubation continued at the temperature indicated for 3 hours. Cells were harvested and lysed as per the standard method (Section 2.3). Values are means of triplicate activity assays, and error bars depict standard deviations.

5.3.3 Method Development for Purification of P34O

The most recently reported method for purifying P34O from *A. baylyi* was a brief description of a four step process involving; ammonium sulphate (AmSO4) precipitation, hydrophobic interaction chromatography (HIC), ion exchange chromatography (IEX) and size exclusion chromatography (SEC) [195]. The HIC and IEX steps from this report were optimised.

The HIC step was performed by Vetting *et al.* [195] by binding P34O to a Phenyl-Sepharose stationary phase in 1M AmSO4, and eluting it with a linear gradient to 0M AmSO4. Here, this wide range of AmSO4 concentrations did not achieve satisfactory separation, typically giving around 4 fold increase in
specific activity over the material loaded. An improvement was made in this linear gradient protocol by decreasing the concentration of AmSO₄ to 0.5M after unretained material had flowed through the column. Detectable P34O activity was not eluted at this AmSO₄ concentration while the UV detector monitoring absorbance at 280 nm (A280) returned to zero. Starting a linear gradient to 0 M AmSO₄ from this point typically gave an 8 fold increase in specific activity over the material loaded.

Subsequent to the HIC step, the IEX step was performed by Vetting *et al.* [195] by diluting P34O, binding it to a DEAE-Sepharose stationary phase in 0 M NaCl, and eluting with a linear gradient to 0.25M NaCl. Here, this method typically yielded approximately 12 fold increase in specific activity over the material loaded. It was found that the recoveries from this step were high and that P34O could be passed through this column twice. The method developed involved a first pass in which the enzyme was eluted by a step change in NaCl concentration of the eluent, and then after rediluting and reloading, eluted by NaCl gradient of narrower concentrations than the literature method. The first pass typically achieved 20 fold increase in specific activity, with close to complete recovery of P34O activity, while the second pass typically achieved around 4 fold increase in specific activity with around 50% recovery of activity. These indicators were a substantial improvement on the method as reported.

### 5.3.4 Purification of P34O

The first step in the process for purifying P34O was AmSO₄ precipitation. Initial experiments with AmSO₄ precipitation often did not increase P34O specific activity, and this step was also the source of losses of activity. A typical AmSO₄ precipitation yielded 52% of the enzyme found in the soluble lysate, and achieved between 0.8 and 2 fold purification. The data presented for this step in
Table 10 was typical. This step did, however, concentrate the enzyme solution for loading onto the first column, so it was maintained in the process.

The HIC step achieved some purification, but was also the cause of substantial enzyme loss. HIC separation of AmSO₄ precipitated material typically gave 6 fold purification over soluble lysate, with a 20% recovery of the activity found in the lysate (Table 10). The IEX step was clearly the highest performing step of the purification process for P34O. The enzyme was enriched 92 fold by the two IEX steps combined, with around 50% recovery of the P34O activity loaded onto the column typical (Table 10).

SEC was performed on a stationary phase with a high molecular mass exclusion limit (200KDa), exploiting the unusually large size of soluble P34O (612 kDa [113]). The enzyme eluted in the void volume unretained, while smaller contaminating host proteins were retained on the column. SEC typically requires small volumes of sample compared to the volume of the column. SEC eluent from the IEX step was too dilute to loaded directly onto the SEC column, so it was concentrated by ultrafiltration (UF) using a unit with a 100 kDa nominal molecular weight cut-off. This gave size based separation, introducing redundancy into the purification. SDS-PAGE analysis suggested that most of the purification achieved in the SEC step was in fact achieved by the UF sample preparation (Figure 28). The purification achieved by this step was typically a 2 fold increase in specific activity. Recoveries were usually in the range of 50% of the P34O activity found after the IEX step (Table 10). The entire purification of P34O process achieved a specific activity of 55 U/mg (Table 10).
Table 10: Data for a typical purification process for P34O

<table>
<thead>
<tr>
<th>Step</th>
<th>[Protein] (mg/mL)</th>
<th>Specific activity (U / mg)</th>
<th>Fold purification</th>
<th>Total activity (U)</th>
<th>Recovery (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysate</td>
<td>19</td>
<td>0.086</td>
<td>1</td>
<td>16.6</td>
<td>100</td>
</tr>
<tr>
<td>AmSO₄</td>
<td>16</td>
<td>0.068</td>
<td>0.8</td>
<td>8.5</td>
<td>52</td>
</tr>
<tr>
<td>HIC</td>
<td>0.22</td>
<td>0.56</td>
<td>6.5</td>
<td>3.8</td>
<td>23</td>
</tr>
<tr>
<td>IEX pass 1</td>
<td>0.007</td>
<td>12</td>
<td>140</td>
<td>3.4</td>
<td>21</td>
</tr>
<tr>
<td>IEX pass 2</td>
<td>0.0017</td>
<td>52</td>
<td>601</td>
<td>1.9</td>
<td>11</td>
</tr>
<tr>
<td>SEC</td>
<td>0.011</td>
<td>55</td>
<td>639</td>
<td>1.0</td>
<td>6</td>
</tr>
</tbody>
</table>

P34O bands on SDS-PAGE gels were identified as those in the molecular mass range expected for the P34O dimer, and were more intense in samples that had high P34O activity. These bands were estimated to have molecular masses of 28 and 27 kDa, in good agreement with the expected values, calculated from the amino acid sequence of 27 and 23 kDa, respectively. Densitometry of a silver stained SDS-PAGE gel indicated the protein was 70% pure (Figure 28). The densitometry data for the intensity for the bands of P34O, where they were detected, agreed with the fold purification data calculated from increases in specific activity.
Figure 28: Silver stained SDS PAGE gel showing the progress of the purification of P34O
0.10 μg of protein was loaded into all lanes except IEX 2, which had 0.03 μg

The specific activity achieved (55 U / mg) was higher than that reported in the literature for the P34O from Acinetobacters (20 U/mg [103] and 34 U/mg [195]). The specific activity data presented in this study suggested that great purity was achieved. The purification of Hou et al. achieved electrophoretic homogeneity with no other contaminating proteins detected with the stain Amido Black [103]. The purified P34O in this study was not found homogeneous by electrophoresis. While this seems contrary to the specific activity data, visualisation of proteins on electrophoresis gels using Amido Black has been estimated as a little under 10 fold less sensitive than Coomassie R-250 [200]. Coomassie R-250, has in turn been estimated as 100 fold less sensitive than silver staining [201]. It seemed likely that the low abundance contaminating host proteins detected in the final preparation here by silver staining would be
rendered undetected if Amido Black stain was used, thus decreasing the sensitivity by 1000 fold. No measure of purity, other than specific activity, was offered by Vetting et al. [195].

Both literature studies [103, 195] subsequently crystallised P34O from their purified preparations, but this was not considered a measure of purity. The apparent different states of purity of these preparations, as measured by specific activity, indicated that the level of purity did not detrimentally impact the crystallisation process. Indeed, it has been observed that purification to homogeneity is not essential for crystallisation of many proteins [202]. The fact that the P34O from *Pseudomonas* species was purified by crystallising the enzyme without chromatography [101], indicated that P34O is an enzyme which is particularly amenable to crystallisation.

### 5.3.5 Kinetics of recombinant P34O

The apparent $K_m$ and $k_{cat}$ for 3,4-dihydroxy benzoate were estimated for P34O. The aim of this analysis was to allow comparison of enzyme in this system with both wild type and recombinant P34O from *Acinetobacter* species.

The $K_{app,m}$ was found to be 7.5 µM (± 0.05 µM) and the $k_{app,cat}$ was found to be 27 s$^{-1}$ (± 2 s$^{-1}$) using the most highly purified preparation of P34O. The values obtained were significantly different from the values of the same parameters reported in the literature (Table 11). The $K_{app,m}$ and $k_{app,cat}$ for P34O from *A. baylyi* has been reported as 78 µM and 120s$^{-1}$, respectively, for an enzyme expressed in a recombinant system [100]. The $K_{app,m}$ has also been reported as 71 µM for *A. calcoaceticus* expressed in the wild type organism [103]. The differences between the values estimated here and those reported in the literature became less significant when comparing the value of the $k_{app,cat}$ divided by the $K_{app,m}$ (catalytic efficiency). The $k_{app,cat} / K_{app,m}$ measured was 4 x 10$^6$, while the literature value was 2 x 10$^6$. 

Table 11: Comparison of P34O kinetic constants estimated here and in literature studies
(Standard errors of the values given in parenthesis)

<table>
<thead>
<tr>
<th>P34O Source</th>
<th>Organism</th>
<th>$K_{\text{app}}$ (µM)</th>
<th>$k_{\text{app cat}}$ (s$^{-1}$)</th>
<th>$k_{\text{app cat}} / K_{\text{app m}}$</th>
<th>% Purity of preparation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A. baylyi</td>
<td>7.5 (0.05)</td>
<td>27 (2)</td>
<td>$4 \times 10^6$</td>
<td>70</td>
<td>This work</td>
</tr>
<tr>
<td></td>
<td>A. baylyi $^b$</td>
<td>14.0 (0.9)</td>
<td>15 (0.3)</td>
<td>$1 \times 10^6$</td>
<td>11</td>
<td>This work</td>
</tr>
<tr>
<td></td>
<td>A. baylyi</td>
<td>10.7 (0.9)</td>
<td>7.4 (0.2)</td>
<td>$7 \times 10^5$</td>
<td>14</td>
<td>This work</td>
</tr>
<tr>
<td></td>
<td>A. baylyi</td>
<td>14.3 (0.7)</td>
<td>Not calculated</td>
<td>Not applicable</td>
<td>Lysate</td>
<td>This work</td>
</tr>
<tr>
<td></td>
<td>A. baylyi</td>
<td>13.5 (0.8)</td>
<td>Not calculated</td>
<td>Not applicable</td>
<td>Lysate</td>
<td>This work</td>
</tr>
<tr>
<td></td>
<td>A. baylyi</td>
<td>78</td>
<td>120</td>
<td>$2 \times 10^6$</td>
<td>Not reported</td>
<td>[100]</td>
</tr>
<tr>
<td></td>
<td>A. calcoaceticus</td>
<td>71</td>
<td>Not reported</td>
<td>NA</td>
<td>100</td>
<td>[103]</td>
</tr>
</tbody>
</table>

a Estimated as percentage of total optical intensity in the lane of a silver stained SDS-PAGE gel

b Reaction performed at 24°C

The kinetic constants estimated in this study on different occasions were more consistent with one another than with the values reported in the literature (}
Table 11). This was the case regardless of the purity of the enzyme used. For example, the three $K_{\text{app}}^m$ values estimated independently with partially purified preparations (7.5, 14.0 and 10.7 µM) were closer to the values obtained in crude lysate (14.3 and 13.5 µM) than to the values reported for purified preparations in the literature (78 and 71 µM). Although most of the parameters of the assay (eg pH, buffer composition etc) were replicated, the temperature was not accurately matched. Hou et al. [103] report the temperature of the reaction to be 24°C, while reactions were performed at 22°C here. One experimental run was performed at this temperature to ascertain if this was the source of the discrepancy using partially purified P34O. This run estimated values for $K_{\text{app}}^m$ and $k_{\text{app}}^{\text{cat}}$ of 14.0 (± 0.9) µM and 15 (± 0.3) s⁻¹, which were still significantly different from 78 µM and 120 s⁻¹.

A possible explanation for the differences in kinetic constants estimated in this study and reported by Hou et al [103] and Brown et al [100] involves differences in the methods used for the estimation of the kinetic constants. The constants estimated here were calculated from direct regression fitting of the Michaelis – Menten equation for one substrate to the experimental data (Figure 29), as has been advocated in the literature [153, 203, 204]. The constants from the literature were estimated using the double reciprocal “Lineweaver Burk” plot. The reliability of this method was questioned from as early as 1932 when Hanes noted that regression fitting of a reciprocal equation gave “the minimum value for the sum of the squares of the deviations between calculated and observed values of $l/v$ (not of $v$)” [205]. More rigorous statistical analyses of double reciprocal plots suggested that valid regression fitting of the line to the data required different weightings for the velocity data measured at different substrate concentrations, and that a confidence interval of the kinetic constants could be estimated from this fitting [206, 207]. It appears that the general consensus was reached that double reciprocal plots were not ideal for estimating kinetic constants [208-211], but recent papers still describe problems arising from their common usage [212-214]. No confidence interval, nor reference to statistical weighting methods, was reported for the literature kinetic constants. If the literature constants were estimated from data with high variation, the confidence interval of the constants would be large, and those
constants may not have been significantly different from those estimated here. It should however be noted that using a double reciprocal plot of the data measured in this study, without appropriate weighting, did not significantly change the kinetic constants estimated. This suggested that this methodology difference may not explain the differences in kinetic constants.

![Graph showing the relationship between the mol of substrate converted per mol of active site (s⁻¹) and 3, 4-dihydroxy benzoate (mM).](image)

**Figure 29: Regression fit of Michaelis – Menten equation to experimentally measured rates of P34O**

*Values are means of triplicate reactions, and error bars depict standard deviations*

A further explanation for the differences in kinetic constants may involve the difference in amino acid sequence in this recombinant form of the enzyme, which had an additional three non-native amino acids to the N-terminal of the β chain of the enzyme. This may have been sufficient to change the behaviour of the enzyme. However, this region of the enzyme was far from the active site [100]. In addition, the crystal structure of P34O from *A. baylyi* did not detect the amino acids serine and glutamine coded for on the N terminal [113]. This suggested that these amino acids were cleaved during post-translational processing and that the N-terminal of the β chain did not contribute significantly to catalysis, making this explanation seem less plausible.
While no definitive explanation can be offered for the differences between the kinetic constants reported for P34O in the literature and those estimated here, if these differences were significant, they were not seen as detrimental to the overall progress of the project. 3,4-dihydroxy benzoate was known to inhibit metalloenzymes by chelating the metal ions bound to the enzyme [215-217], so any traits in P34O which could minimise the intracellular concentration of its substrate were seen as beneficial. A lower $K_{\text{app}}$ for 3,4-dihydroxy benzoate was estimated here, suggested greater catalytic efficacy at reduced substrate concentration. This would result in a lower metabolic pool of 3,4-dihydroxy benzoate in the recombinant biocatalyst with this P34O, compared to that reported.
5.4 Summary and Conclusions

P34O was cloned from *A. baylyi*, and expressed in *E. coli* BL21 (DE3). Greater expression was achieved by lowering the temperature of induction. P34O was purified 639 fold with a 6% recovery by a four step protocol involving AmSO₄ precipitation, HIC, IEX and SEC to a specific activity of 55 U / mg. Silver staining of SDS PAGE estimated this preparation to be 70% pure.

The $K_{\text{app}}^\text{m}$ and $k_{\text{app}}^\text{cat}$ of this P34O were estimated to be 7.5 µM (± 0.05 µM) and 27 (±2 s⁻¹), respectively. These values were not consistent with those reported in the literature. While conclusions can not be drawn on the reasons for this difference, the lower $K_{\text{app}}^\text{m}$ of the P34O purified here was seen as a beneficial trait for an enzyme in a synthetic pathway.

5.5 Further Work

As was the case for VMO (Section 4.5), the kinetic constants for P34O in response to the concentration of O₂ were not estimated. The instrumentation for performing these measurements was not present in the laboratories where the work was performed. No attempts were made to obtain this instrumentation, as achieving control over the concentration of O₂ inside a whole cell seemed unlikely. However, a complete picture for comparison to literature constants would be beneficial to have and future estimation of the kinetic constants for O₂ is recommended.

It would also be interesting to further explore the discrepancy between the kinetic constants reported in the literature, and those estimated here. P34O could be cloned such that the authentic amino acid sequence was expressed, and this P34O could be purified using the method described here, and the kinetic constants estimated.
6 Biotransformation of Vanillin to 3-Carboxy Muconate

6.1 Introduction

6.1.1 Literature review and background for biotransformation of vanillin to 3CM

The characteristics of the enzymes of the synthetic vanillin to 3CM pathway relevant to technical feasibility of the synthetic pathway have been examined. This specific information was integrated with generalisations about assembling synthetic pathways reported in the literature, to help guide the construction of the synthetic vanillin to 3CM pathway.

The \textit{in vitro} characterisation identified that VMO was the least active of the three enzymes. VMO was also found to be the least stable of the enzymes \textit{in vitro}. Although enzymes are typically more stable in a cellular environment than in an isolated soluble form \cite{218, 219}, there are still examples of whole cell biotransformations in which enzyme stability was the limiting factor \cite{190, 220}. These two factors made it desirable to express more molecules of VMO than P34O or HBD to achieve balanced activity.

The Duet plasmid system has been commonly used for introducing synthetic metabolic pathways into \textit{E. coli} \cite{221-223}. It appears that these literature studies have found this expression system convenient for assembling synthetic pathways, but prediction of the relative expression of individual enzymes within the synthetic pathway does not appear to be intuitive. For example, two variations of a synthetic pathway expressing six heterologous enzymes were compared \cite{223}. The only difference in these systems was the copy number of the plasmid which carried the gene for a flavonoid hydroxylase. Greater flavonoid hydroxylase activity was found in the system which expressed this enzyme from a lower copy number plasmid. It is more commonly expected that an enzyme expressed from a higher copy number would be expressed to a higher activity \cite{224-226}.
The yield of synthetic pathway biotransformations can be impacted by thermodynamics, and an example of this can be found in the use of the Duet plasmid system to create a synthetic pathway for creating H₂ in *E. coli* [222]. The yields in this system were closely linked to the directionality of reactions within the pathway. The enzymes of the vanillin to 3CM pathway were all found to catalyse irreversible reactions *in vitro*. This indicated the thermodynamic favourability of the reaction. The presence of substrates, and all three enzymes in an active state in a cell was expected to allow the transformation to proceed.

The yield of whole cell biotransformations is commonly lowered by the metabolism of the recombinant host drawing metabolites away from the pathway of interest. There are many examples in metabolic engineering literature of product yields being increased by removing the interfering host metabolism (e.g. [227-229]). For example, lycopene, a valuable terpene, was produced in *E. coli* by adding a synthetic pathway for its synthesis [230]. It was found that an acetate forming pathway was drawing substrates away from the synthetic pathway, and that the lycopene yield was increased from 1.4 mg/L to 2.1 mg/L by inactivating genes from this acetate production pathway. *E. coli* XL1-Blue possesses an endogenous vanillin reducing activity, which converted the aldehyde group of vanillin to an alcohol to yield vanillyl alcohol [155]. This activity could lower the yield of 3CM by making a percentage of the vanillin added as a substrate unavailable to the synthetic pathway. It was uncertain if the *E. coli* strain used here, BL21(DE3), also possessed this activity, or any other activity which limit the yield by drawing vanillate, 3,4-dihydroxy benzoate or 3CM away from the synthetic pathway.

Endogenous host activity could also have an effect on the rate of the biotransformation, as well as the yield. For example, if it was found that the recombinant host did possess the vanillin reducing activity, the relative rates of vanillin entering the cell, and its reduction to vanillyl alcohol would determine the concentration of vanillin available to HBD. HBD’s $K_{\text{app}}$ for vanillin was 44 μM (Table 7, Section 3.3.5). In addition to vanillin, all three enzymes of the vanillin to 3CM pathway used a substrate which was known to also be employed by host cellular reactions (NAD⁺, NADH, NADPH and O₂). The concentration of such substrates for a synthetic pathway has been reported to be a rate limiting
factor due to endogenous host activity [231-234]. One estimate for the concentration of NAD$^+$ in resting *E. coli* cells is 310 μM [190], which was lower than the $K_{\text{app}}$ value of HBD for NAD$^+$ (600 μM, Table 7, Section 3.3.5). This suggested that HBD’s rate could be limited by the availability of this cofactor. There was however potential for some level of cofactor regeneration within the vanillin to 3CM pathway. HBD was found to use NAD$^+$ almost exclusively (Section 3.3.4). VMO used both NADH and NADPH without significant difference (Section 4.3.5). As such, there was a possibility that the NAD$^+$ reduced by HBD would be regenerated by oxidation by VMO. It was also possible VMO activity would exhaust cellular NADPH, unless host metabolism regenerated it.

There are examples in the literature of substrate limitation arising through low permeability of the cell membrane. In one extreme case, intact whole cells could not be used as biocatalysts because the substrate maleate did not traverse the membrane to any extent [235]. The rate of substrate traversing the membrane has been found to be the rate limiting step in other systems [236, 237], and there has even been strains of *E. coli* engineered to have more permeable membranes for use as whole cell biocatalysts [238]. Vanillin biotransformations performed with *E. coli* had already been reported in the literature (eg [54]), so it appeared that vanillin could traverse the membrane of this microbe. It was however, unknown if 3CM could exit an *E. coli* cell.

Toxicity of synthetic pathway metabolites and / or heterologously expressed enzymes can also be problematic [239-242]. This toxicity can lead to poor growth of the host, or selection for mutations that deactivate the synthetic genes [240]. For example, an *E. coli* host was used to produce vanillin via a synthetic pathway, and the rate of vanillin production in this system decreased as vanillin accumulated [241]. This decrease in productivity was attributed to toxicity on the host of the accumulating vanillin. Vanillin was the substrate here, and not the product, so its concentration was expected to decrease rather than increase. However, aldehydes such as vanillin are known to have toxic effects on cells [169], so there was potential for toxicity of vanillin to impact on the biotransformation. Of the other metabolites 3,4-dihydroxy benzoate was also expected to cause some toxicity to the cell, as it is known to inhibit
metalloenzymes [215-217]. Accumulation of 3CM was also known to be toxic to A. baylyi [243]. No literature on the effect of 3CM against E. coli was found.

6.1.2 Aims and approach for the biotransformation of vanillin to 3CM

An aim of the project was to assemble a functional vanillin to 3CM pathway in a recombinant host. In addition to assembling the system, a further aim of the project was to assess the system in terms of yield and rate.

While planning the assembly of the synthetic pathway, it was apparent that more molecules of VMO would be required for activity equivalent to HBD and P34O. To address this, the construction of the synthetic pathway was planned so that the gene for VMO was placed on the higher number, pET-based plasmid. To further coordinate the levels of activity of the three enzymes, it was also planned to obtain a lower number of either HBD or P34O molecules per cell by placing the relevant gene on the lower copy number, pACYC-based plasmid. Although P34O and HBD had a similar $k_{cat}$ for their aromatic substrate (27 and 18 s$^{-1}$, respectively), HBD was expressed to a higher percentage of total soluble protein than P34O from a pET-based plasmid. It was therefore planned to put HBD on the lower copy number plasmid, pACYCDuet.
Factors which were expected to affect the yield of the biotransformation included endogenous host activity. It was of particular interest to determine if the host strain BL21(DE3) had the known endogenous activity which reduced vanillin to vanillyl alcohol, and if so, whether HBD was able to out-compete it. It was also important to determine if the *E. coli* host could act on vanillate, 3,4-dihydroxy benzoate or 3CM. The yield of the reaction could also be impacted by the ability of 3CM to leave the cell, and it was important to determine this.

Measuring the rate of the biotransformation catalysed by the synthetic pathway was anticipated to identify the rate limiting step. From this, further investigation of the factors leading to one reaction being rate limiting, such as substrate concentration and enzyme concentration, and toxicity of enzymes and/or metabolites, could be explored. To minimise toxic effects of metabolites, it was planned to use resting cells. It was planned that the grow behaviour of the *E. coli* hosts be measured after recombinant protein expression was induced to detect toxicity of enzyme expression.

To assess the yield and rate of the pathway, it was planned that a quantitative LC MS method be developed. This tool could then be used to measure the amount of each metabolite present when different combinations of enzymes were expressed. It could also follow the reaction over time to identify any accumulation of a particular substrate.
6.2 Material and Methods

6.2.1 Cloning to create the synthetic pathway

The cloning work to create the synthetic pathway described below in this section was performed with the diligent and meticulous technical assistance of Jane Fowler.

*E. coli* BL21(DE3) [244] and plasmids of the Duet series [199] were the host / vector system used to assemble the system. The Duet plasmids have two multiple cloning sites to allow for the expression of two cistrons on each plasmid. There are numerous plasmids in this series, with compatible origins of replication, and different antibiotic resistance markers. These features allow multiple plasmids encoding multiple proteins to exist within the one host. The two parent plasmids used here, pETDuet-1 and pACYCDuet-1, exist at approximately 40 and 12 copies per cell, respectively.

HBD was found to have relatively high specific activity (*Table* 6, Section 3.3.1). To help balance this high activity with the activity of the other two enzymes, the gene for HBD was placed on the lower copy number plasmid, pACYCDuet. To do this, the gene encoding HBD was excised from pET-HBD using *BamH*I and *NotI* and inserted into the same sites on pACYCDuet-1 (Figure 30). The resulting plasmid was called pACYC-HBD, and had similar features to pET-HBD.

The sequence of pACYC-HBD (Section 9.4) was determined using the primers Novagen DuetDOWN1 sequencing primer and Novagen pACYCDuetUP-1 sequencing primer (*Table* 5, Section 2.2.4). This sequence data was as expected.
Figure 30: The cloning strategy for constructing pACYC-HBD
To create a plasmid encoding both VMO and P34O, DNA coding for P34O was excised from pET-P34O using BglII and XhoI and inserted into the BamHI and SalI sites of the plasmid pET-VMO (Figure 31). The recombinant plasmid was called pET-P34O-VMO.

The sequence of pET-P34O-VMO (Section 9.5) was determined as per general methods using the primers Novagen pET upstream sequencing primer, Novagen T7 terminator sequencing primer, Novagen DuetDOWN1 sequencing primer, Novagen DuetUP2 sequencing primer, PVP Up sequencing primer and PVP Down sequencing primer (Table 5, Section 2.2.4). The sequence data showed that the regions coding for P34O and VMO were identical to those of the parent plasmids.
6.2.2 Microbial growth and recombinant protein expression to generate biocatalyst

There were eight combinations in which the three enzymes could be combined in single recombinant host bearing two plasmids encoding pET based ampicillin resistance and pACYC based chloramphenicol resistance (Table 12). This set
of hosts was created by co-transforming two plasmids, one pET based, and one pACYC based, into *E. coli* BL21(DE3). This gave eight recombinant hosts with either zero, one, two or three of the three enzymes of the synthetic pathway, all carrying two plasmids. These strains were created so that any metabolic activities endogenous to *E. coli* against the vanillin to 3CM pathway metabolites could be detected. *E. coli* BL21 (DE3) was co-transformed with two plasmids by electroporation as per the general method (Section 2.2.2).

**Table 12:** The strains created which contained two plasmids for assessing the performance of the enzymes of the vanillin to 3CM pathway in *E. coli*

<table>
<thead>
<tr>
<th>Strain name</th>
<th>pET based vector</th>
<th>pACYC based vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty vector</td>
<td>pETDuet-1</td>
<td>pACYCDuet-1</td>
</tr>
<tr>
<td>HBD</td>
<td>pETDuet-1</td>
<td>pACYC-HBD</td>
</tr>
<tr>
<td>VMO</td>
<td>pET-VMO</td>
<td>pACYCDuet-1</td>
</tr>
<tr>
<td>P34O</td>
<td>pET-P34O</td>
<td>pACYCDuet-1</td>
</tr>
<tr>
<td>HBD-VMO</td>
<td>pET-VMO</td>
<td>pACYC-HBD</td>
</tr>
<tr>
<td>HBD-P34O</td>
<td>pET-P34O</td>
<td>pACYC-HBD</td>
</tr>
<tr>
<td>VMO-P34O</td>
<td>pET-P34O-VMO</td>
<td>pACYCDuet-1</td>
</tr>
<tr>
<td>HBD-VMO-P34O</td>
<td>pET-P34O-VMO</td>
<td>pACYC-HBD</td>
</tr>
</tbody>
</table>

Whole cell biocatalysts were generated fresh prior to each experiment by culturing and expressing recombinant protein as per general methods (Section 2.3.1), with the exception that the antibiotic concentration was halved where
hosts contained both a pET and a pACYC based plasmid (as recommended by the vendor of the plasmids when hosts carrying multiple plasmids were used). Cells were grown to stationary phase so that the physiological state of all strains would be comparable. These cells were harvested by centrifugation and resuspended in phosphate buffer saline (PBS, Section 2.3.1).

6.2.3 Whole cell biotransformations

All manipulations involving cells were performed with sterile technique.

Aliquots of cell suspension in PBS were added to microfuge tubes and the cells pelleted by centrifugation at \(13 \times 10^3\) g for 5 minutes at 22\(^\circ\)C. The buffer supernatant was discarded, and the cell pellets were resuspended in reaction solutions consisting of PBS buffer and substrates at a concentration of 1 mM, unless stated otherwise. The cell density in each reaction was equivalent to an \(\text{OD}_{600}\) of 14. Reactions were incubated at 22 (±0.5) \(^\circ\)C on an orbital mixer to keep the cells suspended. Unless otherwise stated, reactions were incubated overnight.

6.2.4 LC MS analyses of biotransformations

LC MS was used to measure the progress of biotransformations. Samples were withdrawn from reaction solutions, cells were pelleted by centrifugation, and the supernatant diluted into the mobile phase used for LC MS (30% acetonitrile: 80% H2O: 0.1% glacial acetic acid (v/v)). Samples containing 3CM were incubated overnight at 4\(^\circ\)C before LC MS analysis.

LC MS was performed as per substrate specificity assays (Section 2.6.5).
6.3 Results and Discussion

6.3.1 LC MS analyses method development

In order to measure the concentrations of all metabolites simultaneously, a LC MS method was developed. This method involved removing the cells from a reaction mixture by centrifugation, and diluting an aliquot of the supernatant into mobile phase (30% acetonitrile, 0.1% acetic acid). The amount of each of the metabolites of the vanillin to 3CM pathway (vanillin, vanillate, 3,4-dihydroxy benzoate and 3CM) was then quantitated by LC MS.

The analytes were separated by isocratic elution from a standard reverse phase C18 column (Figure 32). The retention times of vanillin, vanillate and 3,4-dihydroxy benzoate showed high reproducibility, as did the peak area per unit concentration of standards of these compounds quantitated by UV detection at 220 nm (Table 13). The analyses of 3CM were more problematic. 3CM was not retained on the column, eluting with the void volume, along with UV absorbing interfering substances in the reaction mixtures. However, it was possible to detect and quantitate 3CM by MS amongst these interfering substances by measuring ions at a mass to charge ratio (m/z) of a singly charged 3CM molecule (185 in the negative spectrum).

Table 13: Performance of LC MS analyses for detecting vanillin to 3CM pathway metabolites

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Retention time a</th>
<th>Peak area per μM a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vanillin</td>
<td>5.56 (0.005)</td>
<td>14900 (835)</td>
</tr>
<tr>
<td>Vanillate</td>
<td>3.83 (0.006)</td>
<td>22600 (889)</td>
</tr>
<tr>
<td>3,4-dihydroxy benzoate</td>
<td>2.59 (0.007)</td>
<td>19900 (1154)</td>
</tr>
<tr>
<td>3CM</td>
<td>2.08 (0.067)</td>
<td>111000 (16666)</td>
</tr>
</tbody>
</table>

a Means of 9 analyses, triplicate injections at 10, 30 and 100 μM. Standard deviation (SD) given in parenthesis.
During method development, it was observed that the intensity of the MS response at m/z 185 was poorly repeatable when samples containing 3CM were analysed soon after preparation. Different peak areas were obtained for the same sample, injected at different times. The cause of this variability was found to be variability in the length of time that had passed since preparation of the sample. 3CM was known to isomerise in acidic conditions [55, 118, 122], and samples were prepared by diluting reaction mixture into mobile phase containing 0.1 % (v/v) acetic acid. 3CM, and the two isomers produced by acid catalysed isomerisation, all had a singly charged negative ion at m/z of 185, and were highly polar, eluting in the void volume. As such, all three compounds were detected as a single peak. The amount of each isomer present in the samples changed over time as the isomerisation reaction progressed. The area of the peak at 2.08 minutes detected at m/z 185 increased when the same sample was injected over time. This made it seem likely that each isomer had a different potential for ionisation in the mass spectrometer, thereby eliciting a

Figure 32: Example chromatogram of LC MS analysis of the synthetic pathway metabolites
different intensity of detector response. This variability was decreased by incubating samples containing 3CM at 4°C for 24 hours before analysis, allowing the isomerisation of 3CM to proceed to completion. More variability was observed in the peak areas of 3CM standard samples than for the other analytes despite this incubation. This could be due to higher variability in the intensity of detection by mass spectroscopy compared to spectrophotometry.

Each biotransformation reaction was replicated with no added substrate to detect co-eluting compounds from buffers or the cells. These “blank” analyses did not yield peaks detected at 220 nm with the same retention for vanillin, 3,4-dihydroxy benzoate, or peaks detected at m/z 185 with the same retention time as 3CM. This meant that the quantitation of concentrations of these metabolites could be performed without compensating for interference. Analyses of “blank” reaction did however always show small peaks detected at 220 nm with the same retention time as vanillate. These typically gave a peak area equivalent to less than 5 μM vanillate. The mean peak area of the triplicate of the “blank” analyses was subtracted from the mean peak area of the triplicate of the biotransformation analyses for quantitating vanillate.

6.3.2 Toxicity of expressing the recombinant enzymes

To compare the relative impact on the host of expressing the enzymes of the vanillin to 3CM pathway, the growth of the strains during expression of different combinations of recombinant proteins (Table 12) was measured.

The growth for all eight strains before induction was similar (Figure 33). The cultures behaved differently after induction, and the final OD_{600} provided a measure of the impact of the expression of recombinant proteins.
Figure 33: The impact of recombinant protein expression on the final optical density of recombinant hosts

Cultures grown as per standard method (Section 2.3.1), with IPTG added at OD$_{600}$ 0.8

The strains all carried one of two pACYC based plasmids; pACYCDuet or pACYC-HBD. The identity of the pACYC based plasmid present did not appear to have an impact on the growth of the hosts. For example, the host carrying pACYC-HBD and pET-P34O-VMO reached a mean OD$_{600}$ of 4.3, while the strain carrying the same pACYC based plasmid together with a different pET based plasmid, pETDuet, reached a much lower mean final OD$_{600}$ of 1.2.

It was the identity of the pET based vector which appeared to have a distinct effect on the final OD$_{600}$. Surprisingly, the two lowest cell yields were found in the two strains carrying the pETDuet empty vector. This suggested some toxic effect of expression of the peptide encoded by the multiple cloning site of the empty vector. It has previously been observed that empty pET vectors drastically lowered the viability of host cells in the presence of IPTG [245]. While interesting, it should be noted that this toxic effect of the pETDuet vector was not significant for the behaviour of the synthetic pathway, which did not require an empty vector. The two strains carrying pET-VMO also gave lower cell yields. The only difference between pET-VMO and pET-P34O VMO was the identity of the recombinant enzyme as they were cloned into the same site of pETDuet-1.
as P34O. The strain carrying pACYCDuet and pET-VMO reached a final OD$_{600}$ of 1.6, while the strain carrying pACYCDuet and pET-P34O reached a final OD$_{600}$ of 3.6. This difference was also present between the strains carrying pACYC-HBD and pET-VMO (final OD$_{600}$ = 1.8) and the strain carrying pACYC-HBD and pET-P34O (final OD$_{600}$ = 3.2). This suggested that VMO had some toxicity to the cell.

Interestingly, the highest cell yield came from the host expressing all three enzymes of the vanillin to 3CM pathway (final OD$_{600}$ = 4.3). This suggested that hosts suffered less toxicity from the expression from all three enzymes than expressing two, one or no enzymes from the vanillin to 3CM pathway. This may have been due to expression of lower levels of enzymes with some toxicity (e.g., VMO) in the host expressing all three enzymes compared to those expressing only VMO.

6.3.3 Performance of individual components of the synthetic pathway

The behaviour of each of the eight strains containing different combinations (Table 12) of the plasmids encoding the enzymes of the synthetic pathway was examined to assess the performance of each of the enzymes individually. The LC MS analytical method was used to measure the concentration of all metabolites present. Comparing the amount of metabolites detected with the amount of substrate added allowed examination of the endogenous ability of the eight different strains E. coli BL21(DE3) endogenous ability to act on metabolites of the synthetic pathway.

The results presented in this section can be preceded by two observations which apply to all strains examined. All metabolites of the synthetic pathway were detected in the supernatant of biotransformation experiments performed with whole cell biocatalysts. This demonstrated that all metabolites of the pathway were able to access the enzymes of the synthetic pathway. As no evidence of cell lysis, such as increased viscosity of the reaction medium, or foaming, was observed, it seems likely that the metabolites were able to enter and leave the cells. This was important as it was uncertain if 3CM would be able to traverse the cell membrane and be recovered. These experiments also
demonstrated that all enzymes were expressed actively from all of the four recombinant plasmids used (pACYC-HBD, pET-VMO, pET-P34O, pET-VMO-P34O).

6.3.3.1 Endogenous metabolism acting on vanillin

It had been reported that *E. coli* XL1-Blue possessed the metabolic ability to reduce vanillin to vanillyl alcohol [155]. This was also observed with the BL21 (DE3) strain used during these experiments. Four of the strains constructed had the empty vector pACYCDuet (Table 12), and therefore no recombinant metabolic ability to oxidise vanillin to vanillate. When these four strains were presented with 1 mM vanillin, all produced large amounts of a compound with the same retention time, UV spectrum and electrospray mass spectrum as vanillyl alcohol. Between 46 and 100% of the vanillin initially present was transformed to vanillyl alcohol, depending on the strain (Table 14). This was significant for the overall project as it identified that this previously reported endogenous vanillin reducing reaction would be competing for the use of vanillin in this strain. The level of vanillin reducing activity appeared to be lowered by the expression of recombinant enzymes, with 100% of the vanillin reduced to vanillyl alcohol in the strain with two empty plasmids, and only 46% of the vanillin initially present reduced in the strain simultaneously expressing two recombinant enzymes, VMO and P34O. This difference could be due to either lowered expression of the endogenous enzyme(s) responsible for reducing vanillin, or the intracellular concentration of any cofactor required for the reaction.

The mass balance of the reactions with all four strains accounted for 100 (±10) % of the 1 mM of starting material. This indicated at least 90% of the vanillin present at the start of the reaction stayed unconverted, or was reduced to vanillyl alcohol. If any other endogenous strain metabolism acted on vanillin, it consumed only ≤ 10%. This suggested that no other endogenous strain activity converted vanillin, or further converted vanillyl alcohol. This was significant as it showed that yields would only be impacted by the vanillin reducing activity.
Table 14: The mean concentrations of vanillyl alcohol and vanillin measured after biotransformation of 1 mM vanillin with strains possessing no recombinant vanillin oxidising activity

<table>
<thead>
<tr>
<th>pACYC plasmid</th>
<th>pET plasmid</th>
<th>Vanillyl alcohol (µM) a</th>
<th>Vanillin (µM) a</th>
<th>% of moles accounted for</th>
</tr>
</thead>
<tbody>
<tr>
<td>pACYCDuet</td>
<td>pETDuet</td>
<td>1067 (43)</td>
<td>25 (2)</td>
<td>109</td>
</tr>
<tr>
<td>pACYCDuet</td>
<td>pET-VMO</td>
<td>889 (32)</td>
<td>148 (12)</td>
<td>104</td>
</tr>
<tr>
<td>pACYCDuet</td>
<td>pET-P34O</td>
<td>721 (40)</td>
<td>293 (17)</td>
<td>101</td>
</tr>
<tr>
<td>pACYCDuet</td>
<td>pET-VMO-P34O</td>
<td>460 (9)</td>
<td>432 (12)</td>
<td>89</td>
</tr>
</tbody>
</table>

a Standard deviations of the triplicate measurements given in parenthesis

6.3.3.2 Biotransformations forming vanillate

Strains containing pACYC-HBD, but no VMO to act on the vanillate formed, were set up in reactions with vanillin as a substrate. These strains transformed all of the initial vanillin (Table 15). No significant amount of vanillyl alcohol was detected after the reaction. This indicated that HBD was able to completely out-compete the vanillin reducing activity, making it unlikely that the vanillin reducing activity would lower the yield of the reaction. Both the irreversibility of the HBD reaction and the high activity of HBD were probably important in achieving the quantitative transformation of vanillin to vanillate.

The mass balance indicated that no endogenous host metabolism acted on the vanillate formed. 99 and 103% of the 1 mM of substrate initially present was measured as either vanillyl alcohol or vanillate in the two strains that possessed HBD but no recombinant activity to act on vanillate.
Table 15: The mean concentrations of vanillyl alcohol, vanillin and vanillate measured after biotransformation of 1 mM vanillin with strains recombinant HBD

<table>
<thead>
<tr>
<th>plasmid</th>
<th>Vanillyl alcohol (µM) a</th>
<th>Vanillin (µM) a</th>
<th>Vanillate (µM) a</th>
<th>% of moles accounted for</th>
</tr>
</thead>
<tbody>
<tr>
<td>pACYC-HBD</td>
<td>13 (36)</td>
<td>0 (0)</td>
<td>980 (9)</td>
<td>99</td>
</tr>
<tr>
<td>pACYC-HBD</td>
<td>9 (43)</td>
<td>0 (0)</td>
<td>1015 (58)</td>
<td>103</td>
</tr>
</tbody>
</table>

a SDs of the triplicate measurements given in parenthesis

6.3.3.3 Biotransformations forming 3,4-dihydroxy benzoate

Strains containing pET-VMO, but no P34O to act on the 3,4-dihydroxy benzoate formed, were set up in reactions (Table 16). Reactions performed with the strain carrying pACYC-HBD and pET-VMO produced 993 µM of 3,4-dihydroxy benzoate from 1 mM vanillin. Reactions performed with the strain carrying pACYCDuet and pET-VMO produced 916 µM of 3,4-dihydroxy benzoate from 1 mM of vanillate.

The mass balance for the reaction with the strain expressing HBD and VMO (91%) was artifactually low due to negative concentrations calculated for the metabolites vanillyl alcohol and vanillate. This artefact was due to subtraction of the area of peaks in chromatograms of “blank” controls as described in Section 6.3.1. These interferences meant that there was lower certainty in the significance of the data for these reactions. Despite this uncertainty, it appeared that 3,4-dihydroxy benzoate was not acted on by host metabolism in this case of the HBD and VMO expressing strain, with 993 µM of the initial 1 mM vanillin substrate accounted for.

Only 916 µM of the initial 1 mM of vanillate substrate was accounted for in the strain expressing VMO alone. The balance of material may have been present as unconverted substrate, as low vanillate concentrations were subject to
uncertainty. It was concluded that any metabolic activity against 3,4-dihydroxy benzoate possessed by *E. coli* BL(DE3) was only able to divert small amounts of this metabolite from the synthetic pathway.

**Table 16: The mean concentrations of vanillyl alcohol, vanillin, vanillate and 3,4-dihydroxy benzoate measured after biotransformation of 1mM substrate with strains possessing recombinant VMO**

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Vanillyl alcohol (µM) a</th>
<th>Vanillin (µM) a</th>
<th>Vanillate (µM) a</th>
<th>3,4-dihydroxy benzoate (µM) a</th>
<th>% of moles accounted for</th>
</tr>
</thead>
<tbody>
<tr>
<td>pACYC-HBD</td>
<td>Vanillin</td>
<td>-49 (0)</td>
<td>0 (0)</td>
<td>-27 (20)</td>
<td>993 (17)</td>
</tr>
<tr>
<td>pACYC-Duet</td>
<td>Vanillate</td>
<td>-5 (2)</td>
<td>0 (0)</td>
<td>37 (16)</td>
<td>916 (22)</td>
</tr>
</tbody>
</table>

a SDs of the triplicate measurements given in parenthesis

### 6.3.3.4 Biotransformations forming 3CM

Two strains containing pACYC-Duet, and either pET-P34O or pET-VMO-P34O where set up in reactions, with 1 mM of the substrates 3,4-dihydroxy benzoate and vanillate, respectively. Greater than 1 mM of 3CM product was detected for both of these reactions. This artefact was probably due to uncertainty in the molar response of the mass spectrometer response to this analyte, as discussed in Section 6.3.1. More 3CM was detected than should have been possible to form, so it seemed likely that endogenous host metabolism was not consuming 3CM.
Table 17: The mean concentrations of vanillate, 3,4-dihydroxy benzoate and 3CM measured after biotransformation of 1 mM substrate with strains possessing recombinant P34O

<table>
<thead>
<tr>
<th>pACYC plasmid</th>
<th>pET plasmid</th>
<th>Substrate</th>
<th>Vanillate (µM) a</th>
<th>3,4-dihydroxy benzoate (µM) a</th>
<th>3CM (µM) a</th>
<th>% of moles accounted for</th>
</tr>
</thead>
<tbody>
<tr>
<td>pACYCDuet</td>
<td>pET-P34O</td>
<td>Vanillate</td>
<td>14 (25)</td>
<td>1 (0)</td>
<td>1138 (14)</td>
<td>113</td>
</tr>
<tr>
<td>pACYCDuet</td>
<td>pET-VMO-P34O</td>
<td>3,4-dihydroxy benzoate</td>
<td>29 (35)</td>
<td>1 (0)</td>
<td>1059 (44)</td>
<td>112</td>
</tr>
</tbody>
</table>

a SDs of the triplicate measurements given in parenthesis

The strain containing all three enzymes in the pathway was incubated with 1 mM vanillin, and 735 µM of 3CM was produced (Figure 34). This demonstrated that it was possible to transform vanillin into 3CM using the whole cell biocatalyst constructed in this study, which was one of the aims of the project. 3CM was also produced when vanillate or 3,4-dihydroxy benzoate were the substrate, with 782 and 1036 µM produced, respectively.

In agreement with experiments involving one or two of the enzymes of the synthetic pathway in whole cell biocatalysts, the concentration of material added was accounted for, ± 10%. With vanillin, vanillate or 3,4-dihydroxy benzoate as the substrate, 93, 98 or 104%, respectively, of the initial 1 mM of substrate was accounted for. This demonstrated that host metabolism was not diverting significant amounts of material from the synthetic pathway, and that 100% yields should be possible for the biotransformation. A mean of 8 µM of vanillyl alcohol was detected when vanillin was the substrate, but this concentration was not significantly different to zero.

Vanillate was still present after biotransformations were incubated overnight with vanillin or vanillate as the substrate. When the substrate was vanillin, 182 µM of vanillate accumulated and remained unconverted. A similar amount (210 µM) remained unconverted when 1 mM of vanillate was provided as the substrate. In contrast, 1036 µM of 3CM was produced when 3,4-dihydroxy benzoate was provided as the substrate. The fact that P34O was able to biotransform 100% of substrate while VMO biotransformed approximately 80%
of substrate suggested that the VMO reaction was the slowest of the pathway. VMO converted a similar amount of vanillate regardless of whether that substrate was provided \textit{in situ} by HBD, or as externally added substrate. Each mole of vanillate formed \textit{in situ} by HBD also provided one mole of NADH to the whole cell system, while each mole of externally added vanillate did not provide any reduced cofactor. These two supplies of substrate gave a similar amount of vanillate conversion, suggesting that the redox cofactor regeneration which existed within the synthetic pathway was not a factor which impacted on the yield of this reaction.

![Graph of metabolites detected in reactions containing biocatalysts](image)

**Figure 34:** Metabolites detected \textit{in the reactions containing biocatalysts (cell OD\textsubscript{600} 14)} with the vanillin to 3CM pathway after overnight incubation with various substrates

Values are means of triplicate reactions, and error bars depict standard deviations

### 6.3.4 Time course of the vanillin to 3CM pathway

One of the aims of this project was met by successfully biotransforming vanillin to 3CM. The results of the experiments presented in Section 6.3.3 made it seem likely that 100\% conversion of vanillin to 3CM could be achieved, but it had not been demonstrated. These experiments also did not describe the rate of the biotransformation, as they analysed only a single reaction time point. It was important to demonstrate full conversion of vanillin to 3CM, and to measure the
rate at which this occurred. It was also a further aim of the project to identify the rate limiting factor of this system. It appeared that the reaction catalysed by VMO was the rate limiting reaction of the vanillin to 3CM pathway (Section 6.3.3.4). To examine whether the biotransformation went to completion, the rate of conversion, and the rate limiting reaction biotransformations using vanillin as the substrate were sampled over time.

1 mM of vanillin provided as a substrate for these reactions was converted to 1 mM 3CM (Figure 35). This took 39 hours. This reaction rate was quite slow, but none of the parameters of the biotransformation had been optimised. The cell densities employed for the reaction, OD\textsubscript{600} of 14, were quite modest. It should be facile to increase the reaction rate by increasing the cell density. The effect on rate of many parameters could impact on the reaction kinetics (eg temperature, concentration of the substrates that can be added externally (eg vanillin and O\textsubscript{2}), physiological state of the cells. As these and other aspects of the biotransformation were unoptimised, much scope for improving the reaction rate exists.

It was found that vanillate was the only intermediate which accumulated (Figure 35). Vanillin was rapidly and quantitatively converted to vanillate within 23 minutes of starting the reaction. The vanillate formed was slowly consumed over the next 39 hours. The overall shape of the curve describing the amount of vanillate present during the biotransformation approximated a hyperbola. This suggested that the rate of vanillate conversion was decreasing over time. 3,4-Dihydroxy benzoate did not accumulate to a concentration higher than 20 μM at any point, indicating that the P34O reaction was converting this metabolite to 3CM as rapidly as the VMO reaction was creating it.
The time course data (Figure 35) suggested that there was a sufficiently high concentration of vanillate for VMO to act at an optimum rate. Within 23 minutes, there was 1 mM of vanillate present, well above VMO’s $K_{\text{app}}$ for that substrate, 25 μM (Table 9, Section 4.3.6). Similarly, the reaction of HBD to produce the 1 mM of vanillate within 23 minutes would have produced 1 mM of NADH. It seems likely that a concentration of NADH higher than VMO’s $K_{\text{app}}$ for this substrate, 10 μM, would have been available to VMO for at least the duration of time that HBD was converting vanillin to vanillate. A complex network of host metabolism could be expected to act on this NADH, so that the 1mM produced by HBD was not entirely available as a substrate for VMO. As such, it can be seen that while stoichiometric cofactor regeneration existed within the system, kinetic cofactor regeneration did not. It cannot be ruled out that low NAD(P)H concentration lowered VMO’s rate. However, VMO’s $K_{\text{app}}$ for these redox cofactors was low, with 10 μM for NADH and 3.1 μM for NADPH. As discussed in Section 4.3.6, it seemed likely that these cofactors would exist in vivo at higher concentrations than this.

---

**Figure 35:** Time course of vanillin to 3CM biotransformation

Reactions performed as per Section 6.2.3.

Values are means of triplicate reactions, and error bars depict standard deviations.
It appeared that the velocity of VMO was decreasing over time, and this would be expected if the amount of active enzyme was decreasing over time. VMO was known to be unstable. However, the half life of VMO activity in lysate was approximately 5 hours 30 minutes (Figure 20, Section 4.3.1). The VMO activity was substantially more stable than this in this whole cell system. The rate of vanillate formation, and 3,4-dihydroxy benzoate formation between 12 and 26 hours was approximately constant, suggesting that the amount of active enzyme present during this 14 hour period was not significantly changing. VMO inactivation was probably not a significant factor in VMO being the rate limiting reaction. The VMO reaction was slower than the P34O reaction even early in the biotransformation when VMO inactivation had not proceeded to a high degree.

6.3.5 The amount of each enzyme present

To determine the levels of active expression achieved for each of the enzymes of the synthetic pathway, a cell sample of the strain used as the biocatalyst was lysed, and the activity of the three enzymes in the lysate was measured and compared.

VMO activity was the lowest of the three activities by several orders of magnitude, with 0.0045 (± 0.0003) U / mL of lysate. The activity of P34O and HBD was 0.51 (± 0.003) and 0.94 (± 0.01) U / mL of lysate, respectively. This clearly identified low VMO expression as an important factor limiting the rate of the synthetic pathway.

With the enzyme activity data alone, it could not be concluded whether one or both components of VMO were poorly expressed. To address this, SDS-PAGE analysis was performed on the whole cells used in the time course biotransformation experiment (Figure 36). These cells were sampled before and after induction. The samples taken after induction showed increases in the intensity of bands at the molecular masses expected for P34O, HBD and the oxygenase component of VMO. The reductase component of VMO, expected at 39 kDa, was not detected. This suggested that low VMO activity was caused by low expression of the reductase component.
Figure 36: SDS-PAGE analysis of E. coli HBD-VMO-P34O before and after induction to express the vanillin to 3CM pathway enzymes.

Equivalent numbers of cells loaded to each lane

Lane 1: Pre-induction
Lane 2: 1 hour post induction
Lane 3: 2 hours post induction
Lane 4: 3 hours post induction
Although the stoichiometry of the reductase to the oxygenase components in VMO was not experimentally demonstrated, SDS PAGE analysis of partially purified VMO suggested there was approximately 2 oxygenase units for every reductase unit (Section 4.3.3). It also seemed likely that low expression of the VMO reductase relative to the oxygenase component was due to the lack of an efficient *E. coli* ribosome binding site upstream of the gene encoding this protein (Figure 22, Section 4.3.3).

Another noteworthy result in the expression level analyses of the whole cell biocatalysts was the inconsistency between the expression level of P34O measured by enzyme activity assays and SDS PAGE analysis. Bands corresponding to the molecular mass of each P34O subunit were much more intense than that of HBD in the SDS-PAGE analysis (Figure 36), suggesting that P34O was expressed to a higher level than HBD. However, the specific activity of HBD (0.94 U / mL) was higher than P34O (0.51 U / mL). The activity of the enzymes per mole of active site was similar. The $k_{\text{cat}}^{\text{app}}$ for HBD was 18 s$^{-1}$ (Table 7, Section 3.3.5) and 27 s$^{-1}$ for P34O (
Table 11, Section 5.3.5), so HBD was not an intrinsically more active enzyme. It seemed likely that P34O had been expressed in an inactive form. During the initial cloning and expression of P34O, the expression of soluble active P34O from pET-P34O was found to be increased by inducing at 20°C compared to 37°C (Figure 27, Section 5.3.2). The biocatalysts generated here were induced for recombinant protein expression at 37°C. It seemed likely that a proportion of the intensity of the SDS PAGE bands at the expected molecular mass for P34O subunits was an inactive form. This result highlighted the complexities in balancing the expression of multiple recombinant proteins. It has been suggested that the simplest way to balance the expression of numerous enzymes in a synthetic pathway is to exploit plasmid copy number [224]. The plasmids used here had the same promoter and ribosome binding sites upstream of the start codon, so that transcriptional and translational control should have been equivalent and plasmid copy number should have theoretically been the most significant factor in determining relative expression levels. However, HBD was expressed from the lower copy number plasmid (the pACYC based plasmid), and had higher activity in the lysate than P34O.

6.4 Summary and Conclusions

Eight recombinant *E. coli* BL21(DE3) strains were constructed. All of these strains carried a pET and a pACYC based plasmid, and encoded zero, one, two or three of the enzymes of the vanillin to 3CM pathway. The growth behaviour of these strains after induction of recombinant protein expression suggested that the expression of all three enzymes of the synthetic pathway within a single strain was less toxic to the cells than expression of individual enzymes, or empty vectors in the presence of IPTG.

A LC MS method was developed to quantitate the four metabolites of the vanillin to 3CM pathway. This was used to examine the behaviour of biotransformations performed with the eight recombinant strains. 100 (±10)% of the material added to these biotransformations was accounted for, suggesting that the only endogenous host metabolic activity on the intermediates of the pathway was the reduction of vanillin to vanillyl alcohol. This activity was out-competed by HBD activity, which quantitatively converted 1 mM of vanillin to
vanillate. This lack of background host activity on the intermediates of the pathway meant that 100% yields of the reaction were theoretically possible.

Biotransformations performed with strains encoding all of the vanillin to 3CM pathway enzymes converted 1 mM vanillin to 3CM with a 100% yield after 39 hours.

The rate limiting reaction in this conversion was that catalysed by VMO. It appeared that the most important factor in this reaction being rate limiting was low expression of the reductase component, relative to expression levels of the oxygenase component, as well as HBD and P34O.

6.5 Further work

The “proof of principle” was shown for the biotransformation of vanillin to 3CM on an analytical scale, in an unoptimised system. To advance this technology, the productivity of the system needs to be improved. It seems likely that increasing the expression of the VMO reductase component would increase the conversion rate of the synthetic pathway. It seems likely that stronger expression of this protein could be achieved simply, by adding an efficient ribosome binding site upstream of the gene.

The system also needs to be made complete, by developing a method to recover 3CM from the reaction. A yield of 100% in the conversion is excellent, but would mean little if only a small fraction of 3CM could be recovered. 3CM is unstable in acidic conditions, and requires protonation for solvent extraction to be effective. As such, it is recommended that anion exchange be examined for 3CM recovery. This could be performed at a high pH to stabilise the product. 3CM is a tricarboxylic acid, so could be expected to bind to an anion exchanger more tightly than the majority of other organic compounds found in the supernatant of a whole cell biotransformation.

In the longer term, it is recommended that the expression of the enzymes of the synthetic pathway is optimised. The system constructed in this work did not achieve balanced expression levels of activity of the three enzymes. This was not detrimental to successfully converting vanillin to 3CM. However, it may
become important for future optimisation of this system. The initial strategy was to place the enzyme with the highest \textit{in vitro} activity on the lowest copy number plasmid to help balance the expression of activity. It appeared that expression level of the enzymes was not proportional to the copy number of the plasmid they were expressed from. This may have been due to the use of a plasmid system designed for over-expression of proteins, which typically leads to formation of inclusion bodies of insoluble protein. It may be beneficial for this application to use plasmids which express recombinant proteins to a lower level. This would allow more rational design to achieve balanced enzyme expression. It would also mean that the cellular resources channelled into recombinant protein synthesis are used economically and not wasted on production of inactive enzyme.

After optimising the expression of the enzymes of the synthetic pathway, the parameters affecting the biotransformation could be examined. As discussed in Section 6.3.4, parameters such as the density of cells used in the reaction, temperature, concentration of the substrates, physiological state of the cells, and buffer conditions could be investigated to optimise the productivity of the synthetic pathway.

It is also recommended that a continuous production system be investigated. This could be achieved by retaining the biocatalyst cells behind a tangential cross flow membrane while pumping substrate solution through the cell suspension.
7    Polymerisation of 3-Carboxy Muconate

7.1    Introduction

7.1.1    Radically initiated polymerisation reactions in solution

Radically initiated polymerisation reactions in solution had been shown to polymerise structural analogues of 3CM (Section 1.4.4) [139, 140]. Many of the fundamental aspects of these polymerisation reactions in solution are well understood [246]. Although radical reactions are complex, and involve many competing processes, a widely accepted and simplified model for how polymers are formed involves three processes.

In the initiation process, radicals are generated (Figure 37). In the work described here, the radical initiator azo bis isobutyl nitrile (AIBN) was used. This species decomposes into N₂ and two cyanoisopropyl radicals. These radical species can then undergo a number of reactions, depending on the reactant monomers, solvent and temperature [247]. One of these reactions is the addition to the double bond of a monomer species. This generates a covalent bond between the initiator and the monomer, concomitantly generating a new radical centre. There are many reactions possible between initiator and monomer (Figure 37). It is generally expected that the initiator will add to the least substituted alkene carbon, or the tail, with an accompanying formation of a new radical centre on the more highly substituted alkene carbon, or the head [248].
The propagation process involves newly formed radicals undergoing addition reactions with double bonds. This lengthens the polymer chain by one monomer unit and forms a new radical centre on the growing polymer. As during initiation, many reactions are possible (Figure 38). A radical centre can be on either the head or the tail of growing polymer chain, and can add to either the head or the tail of an alkene monomer. Typically all reactions are observed to some extent. It is known that the chemical nature of both the monomers and the groups surrounding the radical centre, as well as reaction conditions such as solvent, influences propagation which reactions predominate [246, 249].
While propagation is occurring, radical species are also being extinguished by the process of termination. Two main types of reaction can be observed in termination; combination and disproportionation (Figure 39). In combination, two radical centres combine to form a sp3 bond. No new radical centre is formed, and the resultant molecule has a molecular mass which is the sum of the two radicals. In disproportionation, one radical abstracts a hydrogen from another, leading to the formation of two molecules with no radicals; one with an sp3 carbon on the end of the chain, and one with an sp2 carbon on the end of the chain [246].

*Figure 38: Some of the reactions possible during propagation of radical polymerisation*
Figure 39: Some of the reactions possible during termination of radical polymerisation

The complexity of these processes leads to a population of hugely heterogeneous molecules. When the monomer contains more than one double bond, as 3CM does, the complexity, and the heterogeneity of products, is further exacerbated. The structural possibilities for polymer products are multiplied by the fact that these processes may be occurring at either or both double bonds.

Much recent development in free radical polymerisation has involved controlling these processes to give less heterogeneous polymer products (see [250] for review). This control allows desirable molecular characteristics such as high molecular mass or a given functionality on the end groups of the polymer chains to be achieved. However, it was beyond the scope of this study to optimise molecular characteristics of any polymer formed by 3CM. Variables that could be altered for such optimisation, such as temperature and solvent, or addition of co-ordinating agents or nitroxides, were not explored in this work. A single set of standard conditions were applied to all syntheses [251].

Radical polymerisation can be performed with a single monomer (homopolymerisation), or with multiple monomers (copolymerisation). While homopolymerisation has the advantage of simplicity, copolymerisation allows the “tuning” of the material properties of the resultant polymer [252]. This tuning is possible as some of the important properties related to a given chemical functionality can be viewed as additive in polymers. For example, if monomer A yields a homopolymer with a low melting temperature and a monomer B yields
a homopolymer with a high melting temperature, the copolymer AB is likely to have an intermediate melting temperature which can be altered by the relative amounts of A and B present.

Copolymerisation also allows for a greater number of monomers to be used, as some alkenes do not homopolymerise. For example, molecules that contain an alkene conjugated to diacid derivatives, such as dimethyl fumarate (Figure 46), do not homopolymerise under radical initiation, but readily copolymerise with styrene [253].

Copolymerisation behaviour also opens possibilities of utilising both the tricarboxylic acid and alkene functionalities of 3CM in a cross-linked unsaturated polyester system. Unsaturated diacids, such as fumarate derivatives, are currently commonly used as cross-linking agents in unsaturated polyesters [254, 255]. These monomers are incorporated in polyesters at low percentages of the total moles of monomer (mol%) through their multiple carboxylic acid functionalities. Radical copolymerisation with compounds like styrene then causes the polyester chains to be covalent bound at the unsaturated unit, forming a rigid three dimensional matrix. These materials are useful in applications as fibre glass systems, where a solution of unsaturated polyester and styrene are filled in a mould, and then radical copolymerisation initiated to form a solid cast of the mould’s shape. 3CM has three carboxylic acid groups, so its incorporation into a polyester chain is likely to result in a branched, rather than linear chain [256, 257]. This branching would be expected to impart material properties which are different from those found in materials synthesised from linear unsaturated polyesters.

7.1.2 UV initiated radical reaction in the solid state

In addition to examining 3CM’s reactivity towards radical polymerisation in solution, the reactivity of 3CM towards radical polymerisation in the solid state was also examined. A system for generating polymers from structural analogues of 3CM through this reaction system was found in the literature (Section 1.4.4) [141].
Radical polymerisation in the solid phase involves processes analogous to those which occur during radical polymerisation in solution. However, the monomers are in the solid state when the reaction occurs, so their conformation and orientation is restricted, and made regular by this. This regularity is carried through to the macromolecular product, so that the polymers produced by this system have very desirable qualities. They have high stereoregularity, which is difficult to achieve in traditional liquid phase polymerisation systems [258]. The average molecular weight (MW) of poly-diethyl (E,Z)-hexadienedioate formed in the solid phase has been reported to be as high as $4 \times 10^8$ Da [259]. The same monomer polymerised in solution had an average MW of $3 \times 10^5$ Da [260]. The favourable qualities of polymers synthesised from solid state monomers, such as high MW and stereoregularity, have resulted in high value uses for these materials [261, 262]. They have the disadvantage that they are often highly insoluble [142], making them difficult to process.

Solid state polymerisation also has several environmentally friendly advantages over solution polymerisation methods. It uses no solvents or toxic metal catalysts. This reaction also has low energy requirement, as no heating is required. Indeed, the reaction has been reported to proceed by placing an appropriate reactant in sunlight [263].

7.1.3 Aims and approach for polymerisation of 3CM

One of the aims of this project was to demonstrate an application for 3CM. Literature reports of polymers formed from structural analogues of 3CM suggested that reaction of the alkene groups would be more likely to yield a polymer than reaction of its carboxylic acid groups (Section 1.4.4). Two reaction systems were the most likely to yield polymers through reactions of these alkene groups; radically initiated polymerisation reactions in solution and UV initiated radical reaction in the solid state. It was aimed to ascertain if a polymer could be synthesised using 3CM in these two reaction systems.
7.2 Material and Methods

7.2.1 Synthesis of 3CM

3CM was synthesised by a process combining published methods [55, 118, 119] (Figure 40).

![Figure 40: Two step synthesise of 3CM]

Husband et al described the synthesis of 1-methyl ester of 3CM [119]. The protocol called for 132 mM vanillin in aqueous H₂SO₄ at pH 0.5. The aromatic ring of vanillin (≥99%, Sigma) was oxidatively cleaved by adding four equivalents of NaClO₂ (technical grade, Ajax Fine Chemicals), as a 1.32 M aqueous solution. The reaction proceeded for 45 minutes. Sparging with N₂ removed any chlorine gases from the solution so that it was safer to handle. The solution was then saturated with NaCl. The reaction mixture was extracted with diethyl ether. Viscous yellow oil was left when the solvent was removed. The 1-methyl ester of 3CM was recovered from this oil by recrystallisation from ethyl acetate and petroleum spirit (40-60°C). Yields were typically 23% of the theoretical molar yield. This was slightly lower than the 28% yield reported by Husband et al.

The 1-methyl ester group of this species protected it from isomerisation. After this group was removed, all manipulations were performed close to 0°C to counter this instability, unless otherwise stated.

The 1-methyl ester of 3CM was saponified to give the trisodium salt of 3CM as described by Ainsworth and Kirby [55]. This reaction was performed by adding 3 equivalents of NaOH, as an aqueous 0.387 M solution, to 3CM-Me. The
reaction proceeded at room temperature for 20 minutes. Five volumes of methanol were then added. A further six volumes of 2-propanol were added over three hours. This was left for a further hour before the white solid product was recovered by filtration. Molar yields were typically between 95 and 100%.

This 3CM-\(\text{Na}_3\) was then purified as per McDonald et al [118]. Five volumes of methanol were added to 119 mM aqueous 3CM-\(\text{Na}_3\) solution. This caused coloured impurities to precipitate, which was removed by centrifugation. To the supernatant, 8.5 volumes of 2-propanol was added over 4 hours, and allowed to stand overnight at 0\(^\circ\)C. The resulting solid product was again recovered by filtration. Yields of this purification based on mass were typically over 90%.

7.2.2 Synthesis of trimethyl 3-carboxy \(E,E\) muconate

To generate monomer for the study of solution radical reaction of 3CM, \(E,Z\)-3CM-\(\text{Na}_3\) was converted to \(E,E\)-3CM-\(\text{Me}_3\) (Figure 41).

![Figure 41: Esterification of 3CM-\(\text{Na}_3\)](image)

This was achieved by refluxing a 20 mg/mL methanolic suspension of 3CM-\(\text{Na}_3\) with 1.6 molar equivalents concentrated \(\text{H}_2\text{SO}_4\) overnight. The product was recovered by diluting the reaction 1:3 mixture with \(\text{H}_2\text{O}\) and extracting with dichloromethane. Molar yields were typically 30%, as judged by \(^1\text{H-NMR}\).

The product was purified by column chromatography on silica gel 60 (Merck, 1.09385.1000) with chloroform containing 5% (v/v) methanol as the mobile phase. This gave separation of the \(E,E\) isomer from the \(Z, E\) and constitutional
isomeric lactones of 3CM. Only material found to be homogeneous by $^1$H-NMR and GC-MS was used for reactions.

7.2.3 Radical reaction of 3CM-Me$_3$ in solution

Styrene (≥ 99%, Sigma) was prepared for reaction by passing it through an aluminium oxide (90 active, neutral, Merck #101077) column to remove 4-t-butyl catechol present as a radical scavenger. The radical initiator AIBN (98%, Aldrich) was prepared as a solution in toluene before adding to the reaction vessel. Monomers and toluene were added to reaction ampoules (Figure 42) to make a solution with the final concentration of total monomer species of 1.5M. These glass ampoules had fittings allowing them to be evacuated by high vacuum pump. Oxygen can extinguish radicals, so its concentration in the reaction mixture was reduced using freeze/thawing cycles under vacuum. This involved sealing the ampoule, immersing it in liquid N$_2$ for 5 minutes, opening the ampoule to a high vacuum (~3 x 10$^{-3}$ mbar) for 2 minutes, before sealing, and thawing completed the cycle. This cycle was performed 3 times. The ampoule was then immersed in a water bath controlled at 60°C (± 2°C), to initiate the formation of radical species. Reactions were performed for 18 hours, unless otherwise stated.
To recover the product of the reaction, the contents of the ampoules were transferred to a 100 fold larger volume of methanol, with vigorous stirring. This caused polymeric species to precipitate. Solid products were recovered by filtration, and dried in a vacuum oven before further analyses. The methanol supernatant was dried on a rotavap, and the non-volatile material left by this was also analysed for unreacted monomers, and polymer / oligomer products which were methanol soluble.
7.2.4 Synthesis of ammonium 3CM salts for solid state reactivity experiments

Ammonium salts of 3CM were synthesised for assessing their reactivity towards polymerisation in the solid phase. This synthesis involved passing through a protonated 3CM intermediate. The isomerisation of \((E,Z)-3CM\) to \((E,E)-3CM\) discussed in the Literature Review section (Section 1.4.2) was avoided by performing manipulations with the protonated 3CM rapidly at 0°C.

Aqueous 3CM (0.4 M) was protonated with three equivalents of 0.2 M HCl. This solution was rapidly extracted into five volumes of organic solvent. The solvent was placed on dry ice to solidify any traces of water or aqueous HCl, and then filtered. The organic solution was further dried over MgSO\(_4\) and then filtered. Three equivalents of each amine (Figure 46) were added to form the salt. The product salt precipitated and was recovered by filtration. The solvent used was either diethyl ether, or ethyl acetate. The choice between these solvents was made based on which best solubilised the amine to be used, to maximise separation between unprotonated amine which remained in solution, and the precipitated ammonium salt product.
All salts were purified by reprecipitation from methanol with diethyl ether or ethyl acetate. These salts were examined for correct ammonium to acid stoichiometry, purity and correct 3CM stereochemistry by $^1$H-NMR.

Salts of a dicarboxylic analogue of 3CM, $(E,E)$-hexadienedioic acid (98%, Aldrich), were also synthesised as previously reported [264]. This method involved dissolving the acid in methanol, adding two equivalents of the amine, and then adding excess diethyl ether to precipitate the product. This was recovered by filtration, and recrystallised from methanol.

### 7.2.5 Assessment of solid state photoreactivity

A UV reactor designed by Dr Angela Ziebell [265] was used in assessing the photoreactivity of 3CM in the solid state. This reactor had three parallel Phillips TL-K 40W lamps, spaced at 10 cm, mounted inside a sealable stainless steel box (Figure 44). The temperature in the UV reactor was 40°C. Samples were placed at 10 cm path length from the centre lamp for eight hours. Controls
samples were included in the reactor, wrapped in aluminium foil, so that they were not exposed to UV radiation. FTIR spectroscopy was used to determine if the structure of the samples exposed and unexposed to UV were different. KBr disks were typically prepared with 1 mg of sample per 50 mg of KBr.

Figure 44: UV reactor used in solid state polymerisation studies
7.3 Results and Discussion

7.3.1 Monomers used in radical reactions of 3CM in solution

Before attempting polymerisation, a monomer needed to be synthesised. The trisodium salt of the $E,Z$ isomer of 3CM created by the biocatalytic system was not suitable for use in reactions performed in solution. This form of 3CM was not soluble in organic solvents. The $E,Z$ isomer of 3CM was not stable in solution. This was problematic because different alkene isomers have different radical reactivity. For example the reactivity towards radicals of maleic anhydride is greater than fumarate esters which have greater reactivity than maleate esters [266] (Figure 45). The isomerisation of 3CM’s double bond to either the $E,E$ or lactone form (Figure 8, Section 1.4.2) could be expected to change the reactivity of its alkene groups. In this scenario, the interaction between isomerisation and polymerisation reactions would interact, and this complexity could have confounded interpretation of the reaction behaviour of 3CM. To ensure that no isomerisation during polymerisation, the trimethyl ester of the more stable $E,E$ isomer of 3CM (Figure 46) was employed in all reactions in solution.

![Figure 45: The reactivity towards radicals displayed by the isomeric series maleic anhydride, fumarate esters, and maleate esters Adapted from [266]](image)

The monomers styrene and dimethyl fumarate were used as controls in solution radical reaction experiments. Styrene was chosen as it was a well characterised polymer synthon. Dimethyl fumarate was chosen as it was an aliphatic unsaturated polyacid, with carboxyl groups trans to each other across a double
bond, and therefore structurally related to the isomer of 3CM used in these studies (Figure 46).

![Figure 46: Structure of compounds used in solution free radical experiments](image)

**7.3.2 Radical homopolymerisation of 3CM-Me₃ in solution**

The homopolymerisation of styrene and dimethyl fumarate have been previous studied in the reaction conditions used in this work [253]. Styrene was known to form homopolymer product while dimethyl fumarate was known to have no propensity to homopolymerise, so these compounds were included as positive and negative controls, respectively.

After reaction (as per Section 7.2.3), the solutions were fractionated to separate any polymer products. The reactions containing styrene had products which were methanol insoluble, had $^1$H-NMR spectra concordant with a literature spectrum of polystyrene [267], and a high average MW was estimated by GPC ($3.5 \times 10^5$ Da), indicating that polystyrene had been synthesised. This showed that the reaction conditions used were appropriate for radical polymerisation of a reactive monomer, and that the analytical techniques used were appropriate for determining that polymerisation had occurred.

The product of the reactions containing dimethyl fumarate were methanol soluble, had $^1$H-NMR spectra concordant with a literature spectrum of unreacted dimethyl fumarate [268], and gave an average MW estimated to be below 500 Da by GPC. No polymer was expected from this negative control reaction, and the analytical observations supported this.
The product of the reactions containing 3CM-Me$_3$ was methanol soluble. The $^1$H-NMR spectra of this methanol soluble material as not different from the spectra of starting material, and GPC detected no polymer product. This indicated that 3CM-Me$_3$ did not homopolymerise under these conditions.

To assess if the amount of radical initiator was not appropriate for the homopolymerisation of 3CM-Me$_3$ to occur, reactions containing double and half the usual mol % of AIBN radical initiator were trialled. These reactions also yielded no polymeric material for either of the 3CM-Me$_3$ or dimethyl fumarate, while polystyrene was produced at all concentrations of AIBN.

It was concluded that 3CM had little, or no, propensity to homopolymerise under the conditions employed in this study. Structural analogues (free acids and methyl esters of fumarate and maleate as well as maleic anhydride) with alkene groups conjugated to two carbonyl functionalities have little or no propensity to homopolymerise in standard conditions under radical initiation [269]. The more structurally related diethyl hexadienedioate has also previously been shown to have no propensity to polymerise under radical initiation by AIBN in benzene [139]. The lack of reactivity of $E,E$-3CM-Me$_3$ towards homopolymerisation can be explained by the behaviour of these analogues. It was however important that this lack of reactivity be demonstrated experimentally, because $E,E$-3CM did have structural differences to these analogues.
7.3.3 Radical copolymerisation of 3CM-Me$_3$ with styrene in solution

As 3CM-Me$_3$ did not display any homopolymerisation behaviour, copolymerisation was examined. The monomers dimethyl fumarate and styrene were again used as controls, this time as a positive control comonomer pair. It was known that dimethyl fumarate will copolymerise with styrene under AIBN initiation, and that in these conditions, styrene was more readily incorporated into the growing polymer chain than dimethyl fumarate [253].

7.3.3.1 $^1$H-NMR and GPC evidence of styrene 3CM copolymerisation

Initial copolymerisation reactions performed with styrene and 3CM-Me$_3$ at 66:33 mol%, respectively. This reaction mixture yielded methanol insoluble products. A comparison of $^1$H-NMR of homopolystyrene to this methanol insoluble product showed a significant difference (Figure 47). The spectrum of the styrene 3CM-Me$_3$ product had strong signals in the region of 4.0 to 2.8ppm, which were absent from the homopolystyrene spectrum. Peaks in this region are typical of methyl esters. The broadness of these peaks indicated high heterogeneity in the electronic environment surrounding the methyl ester groups which is typical in polymers synthesised by radical initiation in solution. The $^1$H-NMR of copolystyrene dimethyl fumarate also had similar methyl ester signals. This gave evidence that 3CM-Me$_3$ had copolymerised with styrene.
Figure 47: $^1$H-NMR of homostyrene and copoly styrene – 3CM containing 33 mol% 3CM in the reactants
The peaks labelled with a letter (e.g. A-D) are assigned to a proton from a styrene moiety, labelled with the corresponding letter in the structure, while those labelled with a number (e.g. 1 or 2) are assigned to a proton from a 3CM moiety, labelled with the corresponding number in the structure.
The methanol insoluble product of the initial styrene 3CM-Me₃ copolymerisation had an average molecular weight of \(1.1 \times 10^5\) Da, as measured by GPC, showing that the material containing both 3CM-Me3 and styrene was indeed polymeric.

7.3.3.2 Spectral characterisation of the styrene 3CM copolymer

This copoly styrene 3CM-Me₃ was further characterised by \(^{13}\)C-NMR and FTIR (Sections 0 and 1.1, respectively) in addition to \(^1\)H-NMR and GPC (Table 18). All spectra carried distinctive signals typical of methyl ester functionality, with peaks at 51.5 ppm in the \(^{13}\)C-NMR, and \(\sim 1700\) cm\(^{-1}\) in the FTIR. These signals were not present in homopolystyrene. This further confirmed the incorporation of 3CM-Me₃ into the copolymer.
Table 18: Comparison of the spectra of copolystyrene and 3CM-Me<sub>3</sub> with homopolystyrene

<table>
<thead>
<tr>
<th>Spectrum</th>
<th>Analyte</th>
<th>Signal</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Copoly styrene</strong> 3CM-Me&lt;sub&gt;3&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>1&lt;sup&gt;1&lt;/sup&gt;H-NMR&lt;sup&gt;a&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.4 to 6.8</td>
<td>Styrene aromatic H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.8 to 5.6</td>
<td>3CM alkene H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0 to 2.8</td>
<td>3CM methyl ester H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.4 to 0.8</td>
<td>Styrene and 3CM-Me&lt;sub&gt;3&lt;/sub&gt; alkane H</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>7.4 to 6.8</td>
<td>Styrene aromatic H</td>
</tr>
<tr>
<td></td>
<td>Homopoly styrene</td>
<td>2.4 to 0.8</td>
<td>Styrene and 3CM-Me&lt;sub&gt;3&lt;/sub&gt; alkane H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>175 to 166</td>
<td>3CM carbonyl C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>145</td>
<td>Styrene tertiary aromatic C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>140</td>
<td>3CM alkene C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>128 to 125</td>
<td>Styrene aromatic C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51.5</td>
<td>3CM methyl ester C</td>
</tr>
<tr>
<td></td>
<td><strong>1&lt;sup&gt;3&lt;/sup&gt;C-NMR&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>145</td>
<td>Styrene tertiary aromatic C</td>
</tr>
<tr>
<td></td>
<td>Homopoly styrene</td>
<td>128 to 125</td>
<td>Styrene aromatic C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>47 to 40</td>
<td>Styrene and 3CM-Me&lt;sub&gt;3&lt;/sub&gt; alkane C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1734</td>
<td>3CM C=O stretch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1600</td>
<td>Styrene aryl C-C stretch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1492</td>
<td>Styrene aryl C-H bend</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1452</td>
<td>Styrene aryl C-H bend</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1156</td>
<td>3CM C-O stretch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>755</td>
<td>Styrene aryl C-H wag</td>
</tr>
<tr>
<td></td>
<td></td>
<td>697</td>
<td>Styrene aryl ring bend</td>
</tr>
<tr>
<td></td>
<td><strong>FTIR&lt;sup&gt;c&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Homopoly styrene</td>
<td>1601</td>
<td>Styrene aryl C-C stretch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1492</td>
<td>Styrene aryl C-H bend</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1452</td>
<td>Styrene aryl C-H bend</td>
</tr>
<tr>
<td></td>
<td></td>
<td>755</td>
<td>Styrene aryl C-H wag</td>
</tr>
<tr>
<td></td>
<td></td>
<td>697</td>
<td>Styrene aryl ring bend</td>
</tr>
</tbody>
</table>

a Measured at 500MHz in CDCl<sub>3</sub>
b Measured at 500MHz in CDCl<sub>3</sub>
c Measured as 2 % (w/w) KBr disk
Interestingly, there was evidence that aliphatic alkene groups were present in copoly styrene 3CM-Me$_3$. The $^1$H-NMR showed signals between 6.8 and 5.6 ppm, and the $^{13}$C-NMR showed a signal at 145 ppm. This suggested that radicals were not adding to both double bonds in the butadiene system of 3CM-Me$_3$. There were two likely scenarios which would leave alkene functionality in the copolymer (Figure 48). In the first scenario, radicals would add to only one double bond of 3CM-Me$_3$ (1, 2 addition [270]), leaving the other double bond unreacted, and probably inaccessible to radicals due to steric hindrance. Alternatively, radicals could be adding to one double bond, with the newly created radical centre forming on the other double bond through resonance of the butadiene system of 3CM-Me$_3$ (1, 4 addition). [270] This mechanism forms a double bond where a single bond existed before reaction. Both mechanisms are observed under certain situations for butadienes [270-272]. Knowledge of the mechanism of the addition reaction would be useful in understanding the reactivity of the pendant ester groups and the material properties of the copolymer.

![Figure 48: 1, 2 and 1, 4 addition reactions using a 3CM-Me$_3$ example](image)

The effect of varying the relative amounts of 3CM-Me$_3$ and styrene in the feed was studied. In parallel to this, a series of dimethyl fumarate and styrene copolymers with varying amounts of each monomer in the feed were also
synthesised. The resultant polymers were analysed by GPC, \textsuperscript{1}H-NMR and FTIR.

7.3.3.3 Varying the composition of the styrene 3CM copolymer

As would be expected, a higher mol % of both ester monomers in the feed yielded a higher mol % of both ester units within the copolymer products. Evidence for this came from \textsuperscript{1}H-NMR spectroscopy. The ratio of the intensity of peaks assignable to ester units to the intensity of peaks assignable to styrene units increased as a function of the mol % of the ester monomer in the feed (Figure 49). This trend was also observed in FTIR data, where the ratio of the areas of the peaks around 1700cm\textsuperscript{-1}, assignable to C=O groups, to those around 700cm\textsuperscript{-1}, assignable to styrene. These results indicated that the composition of a 3CM-Me\textsubscript{3} styrene copolymer could be controlled by manipulation of the feed conditions, allowing tuning of the properties of the material.
Figure 49: The composition of 3CM-Me3 and dimethyl fumarate styrene copolymers as a function of the concentration of comonomers in feed

1H-NMR signals at 3.5ppm were assigned to methyl ester protons from ester comonomers and signals at 7.0ppm were assigned to aromatic protons from styrene.

Values are means of triplicate reactions, and error bars depict standard deviations

The 3CM-Me3 styrene copolymer gave a higher ratio of the signal of the ester protons to the signal of the aromatic protons than for dimethyl fumarate (Figure 49). This indicated that the numbers of 3CM-Me3 units present per styrene unit were equivalent to the numbers of dimethyl fumarate units per styrene unit. 3CM-Me3 had three methyl ester groups per unit while dimethyl fumarate had two, so a methyl ester signal from 3CM-Me3 would be 1.5 fold higher per mole. As such, it appears that the two ester monomers have behaved similarly over the feed concentrations examined.

The composition of a copolymer as a function of the concentrations of each monomer in the feed is typically expressed as constants termed reactivity ratios [273]. This allows quantitative comparison of the reactivity of each monomer with itself and with the other monomer present. This calculation could not be
validly made with this data, due to limitations in the analytical techniques. The area under the peaks in $^{1}H$-NMR could not confidently be entirely assigned to one monomer. For example, the peak around 7.0 ppm was probably mostly assignable to the aromatic protons from styrene. However, the alkene protons from 3CM-Me$_3$ also gave signals in this region. If unreacted double bonds remained in the polymer chain, then signals from these groups would erroneously be counted signals from styrene units. Although there was less interference of this sort in the FTIR spectra, a molar extinction coefficient would be necessary to calculate the number of units of each monomer present in the polymer chain using FTIR. This required synthesising a homopoly 3CM-Me$_3$ model compound. As it had already been ascertained that 3CM did not show any propensity to homopolymerise, quantitation of reactivity ratios was not further pursued.

The molecular mass characteristics of the copolymers of varying composition were measured by GPC. The average MW’s were between 10 and 30 kDa. No trend was discernable as a function of varying mol % of ester monomer in the feed. It seems likely that other factors impacting on the molecular mass characteristics of the copolymer confounded this data, making any trend that exists difficult to detect (Section 9.8).

7.3.3.4 Applications for styrene 3CM copolymers

These experiments demonstrated that 3CM could be incorporated into a copolymer with styrene, and that the composition of this copolymer could be varied. This gave the “proof of principle” that 3CM could be used as a building block for copolymers synthesised by radical initiation. This finding provides polymer chemists with a renewable building block with pendant carboxylic acid groups that can be copolymerised with styrene.

The molecules which already fulfil this function, such as maleic, fumaric and acrylic acids, although non-renewable, are very cheap. The most commonly used form of maleic acid in commercial polymer synthesis (maleic anhydride) costs approximately US$1600 per metric tonne (two independent quotes obtained from Dalian Chem. Imp. & Exp. Group Co., Ltd. and Henan Harvest
Chem Co., Ltd, 23/9/08). 3CM does have the advantage of renewability to compete with such feed stocks. However, it is still important that 3CM polymer products have a high value if they are to compete with inexpensive monomers such as maleic anhydride.

This high value may come from the use of both of 3CM’s polymerisable functionalities. If 3CM can be incorporated into a polyester chain through reaction of its acid groups, its double bonds could be used as points to cross-link different polyester chains. A polyester synthesised with 3CM units could be expected to be more highly branched than a polyester synthesised with a diacid. The unique functionality that 3CM has may confer properties onto such a material which could not be achieved by the monomers currently used in this application. This would give 3CM value. The results of this study showed that the double bonds of 3CM copolymerise with styrene under radical initiation, giving this application potential.

7.3.4 Solid state reaction of 3CM with UV exposure

7.3.4.1 Monomers used in solid state reactions

To explore an application for 3CM which would give it value, the solid phase polymerisation behaviour of 3CM was examined. This system has been shown to generate polymers with high value applications [261, 262].

Thirteen salts of 3CM were synthesised, along with two salts of (E,E)-hexadienedioate and the components of these salts will be referred to by the letter code given in parenthesis under each structure in Figure 43.

The ammonium counter-ions for the 3CM salts were chosen for a variety of reasons. Literature reported primary aromatic amines to be the counter-ions which most regularly yielded solid state UV reactivity for a range of carboxyl substituted butadiene salts [145]. Benzyl (A) and naphthyl methyl (I) amines had been shown to give reactive salts with other carboxyl substituted butadienes [145]. The hydrocarbon chain length between the amine and the aromatic ring was varied, by using phenyl ethyl amine (B) and phenyl butyl amine (C), but not phenyl propyl amine, due to its toxicity. Substituents on the aromatic moiety
were varied by using \( p \)-tolyl-ethyl amine (E), 2-(4-methoxy-phenyl) amine (F), and 3,4-dimethoxy benzyl amine (G). As naphthyl methyl amine has been reported to give the most generally reactive salts [143], other multiple ring systems were included, namely 1, 2, 3,4-tetrahydronaphthyl amine (J), 1, 2-diphenyl ethyl amine (K), and 3, 3- diphenyl propyl amine (L). The heterocycle piperonyl amine (H) was included in the study to give diversity, as was the highly non-polar but non-aromatic adamantyl amine (D). 3CM-Na\(_3\) was included as it is the form of 3CM produced by the biocatalytic system.

Dibenzyln ammonium \((E,E)\) – hexadenedioate (A\(_2\)-HDA) and dinapthyl methyl ammonium \((E,E)\) - hexadenedioate (I\(_2\)-HDA) were synthesised as negative and positive control materials for UV reactivity experiments, respectively. These compounds have been shown to be unreactive and reactive respectively towards polymerisation in UV light [264].

The salts synthesised for this work were all characterised by \(^1\text{H}\)-NMR, FTIR and UV spectroscopy (Appendix 4). \(^1\text{H}\)-NMR and FTIR were used to ensure that no isomerisation of 3CM had occurred. \(^1\text{H}\)-NMR was also used to ensure that the correct stoichiometry of amine to acid was present in the salt.

Literature spectral data was available for the salts synthesised as positive and negative controls for solid state UV reactivity experiments. The measured spectral properties of A\(_2\)-HDA agreed with those reported [264], with small differences between \(^1\text{H}\)-NMR and FTIR data accounted for by instrumental differences (Table 19). The measured UV molar extinction co-efficient was higher by 18% (3200 L. mol\(^{-1}\). cm\(^{-1}\)) than reported [264]. Although no impurities were detected in the \(^1\text{H}\)-NMR and FTIR, this difference may have been due to the presence of a low concentration of a highly absorbing impurity. Purification by reprecipitation did not improve the discrepancy in this \(\varepsilon\), and A\(_2\)-HDA behaved as expected in the solid state reaction system, so this discrepancy was not further pursued. The measured spectral properties of I\(_2\)-HDA agreed well with those reported [141] (Table 19). The concordance of the spectroscopic characterisations showed that the same salts had been synthesised here as reported in the literature.
Table 19: Comparison of the spectral characteristics of dibenzyl ammonium (E,E) - muconate (A₂-HDA) and dinaphthylammonium (E,E) - muconate (I₂-HDA) measured here and reported previously

<table>
<thead>
<tr>
<th>Compound</th>
<th>Spectral feature</th>
<th>Literature value [141, 264]</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>¹H-NMR shifts (ppm)</td>
<td>4.05, 6.02, 6.87, 6.33</td>
<td>4.20, 6.15, 7.00, 7.45</td>
</tr>
<tr>
<td>A₂-HDA</td>
<td>FTIR C= C= C= C stretching (cm⁻¹)</td>
<td>1610</td>
<td>1620</td>
</tr>
<tr>
<td></td>
<td>FTIR C=O stretching (cm⁻¹)</td>
<td>1515</td>
<td>1515</td>
</tr>
<tr>
<td></td>
<td>UV λmax (nm)</td>
<td>262</td>
<td>260</td>
</tr>
<tr>
<td></td>
<td>UV ε (L. mol⁻¹. cm⁻¹)</td>
<td>17400</td>
<td>20600</td>
</tr>
<tr>
<td>I₂-HDA</td>
<td>¹H-NMR shifts (ppm)</td>
<td>8.10, 7.94, 7.4-7.6, 7.10, 6.14, 4.55</td>
<td>8.10, 7.93, 7.57, 7.11, 6.15, 4.53</td>
</tr>
<tr>
<td></td>
<td>FTIR C= C= C= C stretching (cm⁻¹)</td>
<td>1615</td>
<td>1615</td>
</tr>
<tr>
<td></td>
<td>FTIR C=O stretching (cm⁻¹)</td>
<td>1509</td>
<td>1510</td>
</tr>
<tr>
<td></td>
<td>UV λmax (nm)</td>
<td>261</td>
<td>260</td>
</tr>
<tr>
<td></td>
<td>UV ε (L. mol⁻¹. cm⁻¹)</td>
<td>25100</td>
<td>23500</td>
</tr>
</tbody>
</table>

Although spectral details of 3CM have been published [55, 118, 124, 274] no information was available for the ammonium 3CM salts. The data measured here for the trisodium salt and 1-methyl ester of 3CM agreed well with the literature spectral data (Section 9.9).
7.3.4.2 Photoreactivity of 3CM in the solid state

The monomeric salts were assessed for reactivity due to UV exposure by comparing the FTIR spectra of samples which either exposed to the UV radiation, or wrapped in aluminium foil in the UV reactor. A limitation of using FTIR in these experiments was that it could not show an increase in molecular mass, so it could not directly demonstrate polymerisation. However, FTIR could show that a change in chemical structure due to reaction. As such, these experiments determined photoreactivity, rather than photopolymerisation. GPC is typically used to determine if polymerisation has occurred, but was not suitable here as many of the polymers previously synthesised by this system were highly insoluble [264].

It had been previous reported that I₂-HDA polymerised on treatment with UV, and that this reaction caused the band assignable to C=O bonds at 1509cm⁻¹ in the monomeric species to shift to 1531cm⁻¹ after polymerisation [141]. The system used here reproduced this change in FTIR spectra of the positive control I₂-HDA (Table 20 and Section 9.10). The sample shielded from UV radiation had a band assignable to C=O bonds at 1510cm⁻¹ which shifted to 1531cm⁻¹ in the sample exposed to UV radiation.

A₂-HDA had been shown to have no UV photoreactivity in these conditions [141]. This salt was employed in these experiments as a negative control. The FTIR spectra of UV treated and untreated A₂-HDA were not different (Table 20, Section 9.11). This demonstrated that this experimental system appropriately reported no change to compounds which were not UV reactive in these conditions.
Table 20: Change in FTIR signals from positive and negative control salts due to UV exposure

<table>
<thead>
<tr>
<th>Sample</th>
<th>Exposed to UV?</th>
<th>Functional group</th>
<th>FTIR Peak (cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive control</td>
<td></td>
<td>C=O</td>
<td>1509</td>
</tr>
<tr>
<td>I₂-HDA</td>
<td>No</td>
<td>C=C</td>
<td>1615 and 1630</td>
</tr>
<tr>
<td>Positive control</td>
<td>Yes</td>
<td>C=O</td>
<td>1531</td>
</tr>
<tr>
<td>I₂-HDA</td>
<td></td>
<td>C=C</td>
<td>1638 and 693</td>
</tr>
<tr>
<td>Negative control</td>
<td></td>
<td>C=O</td>
<td>1517</td>
</tr>
<tr>
<td>A₂-HDA</td>
<td>No</td>
<td>C=C</td>
<td>1621</td>
</tr>
<tr>
<td>Negative control</td>
<td></td>
<td>C=O</td>
<td>1517</td>
</tr>
<tr>
<td>A₂-HDA</td>
<td>Yes</td>
<td>C=C</td>
<td>1620</td>
</tr>
</tbody>
</table>

Ammonium 3CM salts (A to L) were showed no difference in the position or intensity of a FTIR peak between UV treated and untreated samples. This indicated that reaction due to UV exposure did not occur in the 3CM salts assessed. Therefore, photopolymerisation could not have occurred. This lack of reactivity could have been due to a number of factors, discussed in Section 7.3.4.3.

7.3.4.3 Discussion of 3CM’s lack of UV reactivity in the solid state

One characteristic of 3CM derivatives that seemed most likely to prevent photoreactivity was the structural differences between 3CM and the other acid butadienes which had been previously polymerised. 3CM had a carboxylic acid substitution on C3 (Figure 10, Section 1.4.4). This C3 carboxylate group could cause steric strain by interacting with C6. This would force a non-planar conformation on the conjugated butadiene double bonds. None of the previously polymerised compounds had a C3 substituent, and all would be expected to be planar. Modelling with ChemDraw 3D [275] followed by energy minimisation of 3CM and E,E- hexadienedioate illustrates this difference in planarity (Figure 50). Spectral data have been reported that give experimental evidence that 3CM had a non-planar conformation [55, 124]. The magnitude of
coupling between the hydrogen on C2 and C5 observed in the $^1$H-NMR spectra was lower than typically observed for butadienes, and the authors of this study attributed this observation to non-planarity of the butadiene backbone [55]. Other authors have concluded that weaker UV absorption by 3-methyl muconic acid, compared to muconic acids unsubstituted at C3 was evidence of non-planarity, and further hypothesised that the stereochemistry of other C3 substituted muconates would be similar, due to steric crowding between the C3 substituent and the C6 carboxyl group [124].

![Comparison of planarity of structural models of 3CM and E,E hexadienedioate](image)

*Figure 50: Comparison of planarity of structural models of 3CM and E,E hexadienedioate*

*The top view shows each molecule aligned with the plane of the page.*

*The lower views shows the molecule aligned with the plane perpendicular to the page.*

This disparity in stereochemistry with previously polymerised compounds did not preclude the possibility of 3CM molecules polymerising in the solid state. It has been stated that solid state polymerisation requires monomers to stack within 5Å of one another [145]. It appears that planarity is not required to obtain this proximity. Diethyl E,Z- muconate polymerised upon UV exposure [276], and the ethyl ester chains on this molecule were not planar. Another example of non-planar molecules packing within 5Å was found in the x-ray crystal structure
of brucinium dihydrogen citrate trihydrate [277]. Citrate is a tricarboxylic acid, with no C=C bonds, and therefore no planarity in its carbon backbone. The modelling program Mercury 1.4.1 [278] measured C2 of citrate to be 4.9Å from C4 of an adjacent molecule in this crystal structure (Figure 51).

Figure 51: Non-planar citrate molecules packing in close proximity in crystals of brucinium dihydrogen citrate trihydrate

Other possibilities for this lack of reactivity exist. It could be possible that impurities in the crystal lattices of the 3CM derivatives inhibited the photoreaction. No impurities were detected by 1H-NMR in the salts before they were examined for UV reactivity. This technique will not have detected trace impurities due to its detection limits; neither with it have detected impurities without hydrogen atoms. It may be that impurities were present in the salts containing 3CM, and these inhibited photoreactivity.

Another possibility was that 3CM derivatives required more UV energy to cause the 3CM derivatives to react than the positive control material. To assess this, I3-3CM was subjected to UV exposure for five days (compared to the 8 hours exposure used for the positive control I2-HDA). This did not result in a difference between the FTIR spectra of UV treated and untreated samples. It was concluded that the 3CM derivatives tested would not react in the solid state after exposure to practical amounts of UV radiation.
It could also have been possible that the method of crystallisation meant that the solids formed were not suitable for photoreaction to occur. Assembly in the solid state is known to vary depending on the crystallisation conditions (e.g. temperature, solvent etc). However, the method of forming the crystal did not appear to be crucial for reaction of diethyl $E,Z$ muconate, which polymerised when it was converted to the solid phase by a range of different methods (precipitated, crystallised etc.) [259]. In this study, the monomer reacted with similar yields independent of the crystallisation conditions. The effect of crystallisation method on photoreactivity was not examined here, where the solubilised monomer was precipitated by adding a less polar solvent. This method has been shown to produce reactive crystals [259].

### 7.4 Summary and Conclusions

To enable the study of polymerisation behaviour of 3CM in solution, it was necessary to convert it to the trimethyl ester of the $E,E$ isomer. This removed the possible complexities of interactions between polymerisation reactions, and isomerisation reactions due to 3CM’s instability. This conversion also suffered poor yields. The instability of 3CM made it seem likely that loss of material would accompany most processes undertaken with it, making it difficult to work with, particularly in a commercial setting.

The $E,E$ isomer of 3CM-Me$_3$ did not homopolymerise under the conditions used. It was concluded that standard reaction conditions would be unlikely to generate a homopolymer of this species, as chemically analogous monomers also show little propensity to homopolymerise.

The $E,E$ isomer of 3CM-Me$_3$ did however, copolymerise with styrene. $^1$H-NMR data suggested that the copolymer product had alkene functionality remaining. 1, 2, or 1, 4 addition reactions of the butadiene group would explain this. The composition of this copolymer could be varied by varying the relative concentrations of the two comonomers in the feed. This will allow the rational optimisation of the material properties of these copolymers.

Solid aryl ammonium salts of 3CM did not react after being exposed to UV light. It seems likely that this was due to unfavourable solid phase packing. The non-
planar conformation of the 3CM species could have caused this unfavourable packing.

Thus it is concluded that 3CM has been shown to have use as a polymer building block. However, it was not demonstrated that these uses are valuable.

7.5 Further work

Given the difficulties in working with 3CM, plausible applications for it should be of relatively high value. High value could be realised if 3CM can perform a function which is not currently fulfilled by renewable chemicals. Molecular and Health Technologies, CSIRO is currently undertaking work to assess if 3CM can be used as an unsaturated unit in a polyester chain to allow cross-linking with styrene. This chemical function is currently being fulfilled by non-renewable dicarboxylic acids. 3CM may find value in this application through its renewability, and its unique chemical functionality.

3CM could also obtain high value as a building block in a copolymer, if these copolymers have superior properties. The synthesis and material property testing of styrene and 3CM copolymers is recommended. However, the testing of material properties usually requires a large amount of material. It is recommended that this work be undertaken after the optimisation and scaling up of the vanillin to 3CM biotransformation.
8 General Discussion and Conclusions

8.1 Summary

Vanillin was identified as a renewable starting material that could be used for the sustainable production of a potential feedstock chemical. 3CM was selected as a compound which could be made from vanillin using biocatalysis, and has a structure that suggested that it could be useful as a polymer building block. The metabolic pathway from vanillin to 3CM from *Acinetobacter baylyi* was chosen to perform this biotransformation.

*In vitro* characterisation and literature studies on HBD, VMO and P34O, the three enzymes of this pathway, were performed. It was an aim that the enzymes be combined in a synthetic pathway, and that the behaviour of this pathway be understood. It was a further aim to determine if the product of the synthetic pathway, 3CM, could be used as a polymer building block.

The enzymes of this pathway were individually cloned and expressed in *E. coli*.

The first enzyme of the pathway, 4-hydroxy benzoic aldehyde dehydrogenase (HBD) was expressed with a hexahistidine affinity tag, purified to greater than 95% purity. HBD had not been previously characterised. HBD was found to retain 90% of its activity after 9 hours at 37°C, and when purified had a moderate specific activity of 22 μmol/min/mg. It was found to prefer NAD⁺ to NADP⁺. The $K_{app}$ for NAD⁺ and vanillin were found to be 600 and 44 μM, respectively. The reaction it catalysed was irreversible.

Vanillate monooxygenase (VMO) catalysed the second reaction of the pathway. The stability of this enzyme had not been previously quantitated, or studied in a wide variety of conditions. VMO was found to have a half-life of approximately 5 hours in the lysate of recombinant *E. coli* in Tris-HCl buffers. This stability was sufficient to achieve partial purification to 37% purity. The partially purified VMO catalysed an irreversible reaction, and used NADH or NADPH without preference. The $K_{app}$ for NADH, NADPH and vanillate were 10, 3 and 25 μM, respectively.
The final enzyme in the pathway, protocatechuate 3,4-dioxygenase (P34O) had been extensively described in the literature. The recombinant enzyme used in this study was purified using a published method, and its kinetic parameters were measured for a comparison with those previously reported. It was found that the recombinant enzyme used in this study had a lower $K_{\text{app}}^m$ and $k_{\text{app}}^{\text{cat}}$ for 3,4-dihydroxy benzoate (7.5 µM and 27 s$^{-1}$, respectively) to those reported in the literature (78 µM and 120s$^{-1}$, respectively) [103].

The genes cloned from *A. baylyi* were combined within a single recombinant *E. coli* host. This was achieved by combining the genes encoding P34O and VMO on the higher copy parent plasmid pETDuet, and the gene encoding HBD on the lower copy number parent plasmid pACYCDuet, and co-transforming these plasmids into a single *E. coli* BL21(DE3) host. Resting cells of this host converted 1 mM vanillin to 1 mM 3CM in 39 hours. The rate limiting reaction of the pathway using this biocatalyst was the VMO catalysed step. Low levels of expression of the reductase component of VMO contributed to this step being rate limiting. No other reports describing a biocatalytic method of converting vanillin to 3CM were known.

3CM was examined for its usefulness as a polymer building block. The trisodium salt of the *E,Z* isomer was chemically converted to the trimethyl ester of the *E,E* isomer. This species did not homopolymerise, but did co-polymerise with styrene. No other report of the incorporation of 3CM into a polymer had been found. The amount of the muconate monomer incorporated into the polymer was proportional to the amount provided in the starting material up to 65 mol% 3CM. Salts of structural analogues of 3CM had been reported in the literature as forming high value polymers in the solid state on UV exposure [143]. Salts of the *E,Z* isomer of 3CM were trialled for their ability to polymerise in these conditions, and were found to be unreactive.
8.2 General Discussion

8.2.1 Enzymological feasibility of the system

Theoretical predictions were made about how the synthetic pathway could be expected to behave. It is common to use equilibrium constants, kinetic constants, and enzyme concentrations for these predictions [279-283]. It was not possible to use the *in vitro* characterisation of HBD, VMO and P34O to make quantitative predictions of the rates of the synthetic pathway, as the active enzyme concentration achievable in the whole cell biocatalyst was unknown. There have also been examples demonstrating that using *in vitro* enzyme kinetics data to make quantitative predictions of *in vivo* metabolite concentration can be prone to error [284]. However, qualitative predictions about the theoretical feasibility of the pathway were made based on the reversibility and kinetic constants for the individual enzymic reactions, as well as the net cofactor requirement for the synthetic pathway.

The reactions catalysed by the enzymes were all irreversible, indicating that the equilibrium constants for all reactions thermodynamically favoured oxidation of vanillin along the pathway towards 3CM. This was important in being able to predict that the pathway would theoretically operate, because it removed the need to control the concentrations of metabolites to control the direction of the reaction. Gaining control over the concentrations of metabolites shared between the synthetic pathway and host metabolism (the nicotinamide dinucleotide cofactors and O$_2$) would be complex.

The apparent Michaelis constants also suggested that stringent control of metabolite concentrations would be unnecessary. The $K_{\text{app}}^m$ for all of the substrates of all of the enzymes was estimated, with the exception of O$_2$, and were found to be in the μM range. It had previously been shown a recombinant *E. coli* host expressing *Pseudomonas* HBD transformed 14 mM of vanillin to 14 mM of vanillate in one hour [54]. It therefore seemed likely that it would be possible to provide the enzymes with a concentration of substrate which would effect efficient catalysis.
Balancing the amount of redox cofactors is often a difficult and complex task in creating synthetic pathways [285]. This difficulty could be avoided with this pathway, as stoichiometric regeneration of the cofactors used by the pathway was possible, because VMO would act on NADH and HBD would act on NAD$^+$. This meant that the pathway theoretically did not require endogenous metabolism to regenerate either NAD$^+$ or NADH for the pathway to continue operating. This factor simplified the operation of the pathway substantially. However, VMO also employed NADPH, with no significance difference between the catalytic efficiency estimated for the two cofactors. VMO could have been expected to oxidise any NADPH present in the cell, and this would not have been regenerated through reduction by HBD. The impact of this in growing cells could be significant, but resting cells were used, and a low metabolic pool of available NADPH was not expected to be detrimental to the system.

After studying the characteristics of the three enzymes of the vanillin to 3CM pathway, it could be predicted that appropriate assembly of the pathway should result in successful conversion of vanillin to 3CM.

### 8.2.2 Construction of the first embodiment of the system

The characteristics of the enzymes of the synthetic vanillin to 3CM pathway suggested that vanillin should be converted to 3CM when an appropriate recombinant whole cell biocatalyst was assembled. Although expression of recombinant proteins is a mature technique, achieving active expression of multiple enzymes composing a synthetic pathway has been problematic for many studies reported in the literature [286, 287]. Many of these technical difficulties arise from the difference between the organism which the enzyme is taken from and the organism which is the recombinant host. For example, the synthesis of oxygen functionalised terpenes by bacterial systems involved difficulties in heterologous expression of plant cytochrome P450’s in an *E. coli* host [288]. These enzymes typically exist embedded in the membranes of the endoplasmic reticulum [289], and bacteria do not possess these organelles. To avoid this complexity, the enzymes constituting the vanillin to 3CM pathway were taken from *A. baylyi*, an organism similar to the recombinant host, *E. coli*. Both of these bacteria belong to the *Gammaproteobacteria* class. The
expression of the enzymes of the vanillin to 3CM pathway from *A. baylyi* in *E. coli* has been achieved without adding *E. coli* transcription and translation elements to the *A. baylyi* gene [58, 86, 196].

However, there were difficulties in achieving coordinated expression. The low expression of the VMO reductase component caused the VMO reaction to be rate limiting (Section 6.3.5). Some low level expression of the reductase component must have occurred, as VMO activity was observed. That the reductase was not detected by SDS-PAGE while the other enzymes were was attributed to the coding sequence gene lacking the Shine-Dalgarno translational initiation sequence that the other oxygenase coding region. In addition, there were other examples of how the complexity of a whole cell recombinant protein expression system required an empirical approach to find suitable conditions for balanced expression of activity. When the enzymes were expressed individually from the pET based plasmid, from the same promoter, with the same ribosomal binding site (for the first transcribed ORF), different conditions were required for optimal expression of the three recombinant enzymes. P34O activity was expressed to a low level when induction occurred at 37°C, while both HBD and VMO activities were expressed to a high level with induction at this temperature. This highlighted that each protein was unique, and generalisations could not be applied to all. The pET plasmids are typically designed for overexpression of a single protein. This can lead to the formation of inclusion bodies of insoluble protein [290], as appeared to have occurred with P34O. A biocatalyst requires expression of active enzymes, with the minimum drain on the host metabolism [39]. A biocatalyst also requires coordination of expression, so that the relative amounts of enzyme are optimal [291]. Achieving this coordination in the host vector system employed here would seem to require an empirical approach, as did expression of the enzymes individually.

It was highly important for the whole cell biocatalyst to be functional that vanillin could enter the *E. coli* host, and that 3CM could leave. It had previously been shown that vanillin entered an *E. coli* strain [54]. However, it was not certain that 3CM would traverse the cell membrane. 3CM is a highly polar small molecule, probably existing as a tri-anion at physiological pH. As such, it seemed unlikely to have diffused unassisted across the membrane, and a transporter protein
was likely to be required for 3CM to be exported from the cell. Although no such transporter had been identified in \textit{E. coli}, there has been a report of a mutant strain of \textit{Pseudomonas} which had an apparently inducible system for transporting 3CM into the cell [292]. Unfortunately, this report comes from 1972, and no sequence data for this transport system can be used to search for homologues in \textit{E. coli}.

Experiments performed to ascertain the toxicity of expression of the recombinant proteins found that the cells expressing all three of the enzymes grew to higher density than cells expressing two, one or none of the recombinant enzymes, boding well for optimising the expression of the synthetic pathway. Metabolites of synthetic pathways can also be toxic to recombinant hosts [293]. Experiments performed to assess if accumulations of any of the metabolites of the synthetic pathway had a detrimental effect on the resting cell biocatalysts also suggested that the synthetic pathway was reasonably benign.

In assessing the performance of this whole cell biocatalyst, the yield of the reaction was found to be 100%. This indicated no metabolites were lost from the pathway due to endogenous activities. The \textit{E. coli} host was observed to reduce vanillin to vanillyl alcohol in many control experiments in which the host did not express HBD. This had been previously reported [155]. This vanillin reducing activity endogenous to \textit{E. coli} could have diverted material away from the pathway. However, HBD out-competed this endogenous vanillin reducing activity. No other endogenous activity was detected against any metabolites. This, along with the irreversibility of the reactions, meant that 100% conversion was expected, provided active enzyme was present and able to access the substrates.

The estimated kinetic constants of the individual enzymes allowed identification of low VMO reductase expression level as the rate limiting factor in the biocatalyst. It is common to compare the $K_m$ of an enzyme to the measured metabolite concentration to conclude that the enzyme velocity is or is not limited by low substrate concentration [294-296]. In this study, the concentration of the substrates for VMO could be compared to the $K_{\text{app}}$ for the substrates as evidence that VMO’s rate was not limited by low substrate concentration,
assuming that the intracellular concentration of vanillate was equivalent to the extracellular concentration of vanillate. Having estimates of kinetic constants was also important in suggesting that VMO’s expression level was rate limiting. HBD and P34O were found to be more active per molecule of enzyme \textit{in vitro} than VMO. The $k_{\text{cat}}$ could not be estimated for VMO due to its instability and lack of purity (Section 4.3.3). However the level of activity of the three enzymes could be compared by dividing the specific activity by the percentage purity of the most purified preparation. These values gave the activity of the enzyme in “ideal” conditions (optimal pH, saturating substrate concentrations etc), so will not necessarily be the same as they would be in cellular conditions. They also were not precise, as they involved three experimentally measured numbers, all with associated uncertainty. However, the difference between this value for VMO and the two other enzymes was large. These data showed that VMO was orders of magnitude less active, with a rate of 0.43 U per mg of the enzyme. HBD and P34O catalysed their reactions at a rate of 22 and 78 U per mg of enzyme. This made it likely that high levels of VMO expression would be required to stop it being rate limiting. These levels of expression were not achieved, so as predicted, the amount of VMO present was the rate limiting factor.

The further development of the recombinant biocatalyst constructed here should be convenient to achieve. The commercial plasmid system used for co-expression of multiple enzymes has four plasmids, each with two multiple cloning sites. Only three of a potential eight multiple cloning sites were used in this embodiment of the biocatalyst. This gives extra capacity for exploring possibilities for optimising the relative levels of the enzymes by manipulating their expression levels through the copy number of the plasmid they are encoded on. It also leaves scopes for adding additional enzymes.

\subsection*{8.2.3 The potential value of 3CM}

3CM displayed behaviour that would make it a useful chemical, but more work is required to demonstrate that it has value.
Copolymerisation of the trimethyl ester of $E,E$ isomer 3CM with styrene demonstrated that the alkene functionalities of this form of 3CM were reactive towards radicals. This, in itself, makes 3CM useful, as it gives another renewable polymer building block for polymer chemists to explore. The carboxyl groups remaining after incorporation in a polymer chain could be useful for attaching pendant groups, chelating metal ions, or simply providing a negative charge. However, it still remains to be tested if a 3CM - styrene copolymer has advantageous properties, illustrating that the next step in developing this polymer building block needs to be a demonstration that its usefulness can provide value.

A caveat on the value of 3CM was that it was difficult to work with. Its propensity to isomerise in acidic conditions meant that material was lost when these conditions had to be used. The biocatalytic system generated 3CM in the form of the trisodium salt, which was not soluble in organic solvents. As many polymer synthesising reactions are conducted in organic solvents, 3CM needed to be converted from this salt to another form to be widely useful as a polymer building block. In this project, this conversion involved acidic conditions, so isomerisation, and therefore losses occurred.
8.3 Conclusions

It can be concluded from this research that the enzymes of the vanillin to 3CM pathway had the characteristics necessary to make a biocatalytic system for this conversion possible. A whole cell biocatalyst was successfully constructed and shown to quantitatively convert vanillin into 3CM. The rate of this conversion was limited by the reaction catalysed by VMO. 3CM displayed useful reactivity, copolymerising with styrene.
8.4 Further work

There are many aspects of this work that could be usefully extended. Further work which applies specifically to areas covered in each chapter is given in each chapter. The recommendations made here are of a more general nature, and are given under the two headings; short term, and long term. The former is further work which should be, or is being, undertaken as a priority, as it has higher significance, or would give benefit without requiring extensive effort. The latter are future directions which would require more extensive effort.

8.4.1 Short term

The most pressing area of further work in this project is to demonstrate that the product of this biocatalytic system has value. As discussed in Section 7.3.3.4, one application which may see 3CM prove its value is as a building block for unsaturated polyesters. These polymer chains can be further polymerised by reaction with styrene, which provides cross-linking between chains. This forms a fibre glass system with many applications. By demonstrating copolymerisation with styrene, this study showed that 3CM could provide this point of cross-linking, if it can be incorporated into a polyester chain. Polyesters synthesised using a tri-acid such as 3CM would be branched. This branching could provide greater entanglement of the polyester chains, or a more rigid network after reaction with styrene. This functionality and its renewable nature give 3CM a narrow chemical niche which may prove valuable. Work is underway to examine this application at CSIRO Molecular and Health Technologies, studying the synthesis of polyesters which contain 3CM. These polyesters will then be cross-linked with styrene. The material properties of these cross-linked polymers will be tested at Deakin University.

The research on the synthesis and material property testing of cross-linked unsaturated polyesters of 3CM requires it be synthesised on a tens of grams scale. The chemical synthesis of 3CM is highly exothermic, and generates chlorine gases, making it difficult to scale up. It would be more expedient to scale up the synthesis of 3CM using the biocatalyst constructed here. To generate material for polymer studies, the biocatalyst needs to be optimised to
increase VMO activity, and a method for product recovery needs to be developed as discussed in Section 6.5.

The synthesis of 3CM on a preparative scale, and the studies into the value of cross-linked unsaturated polyester containing 3CM should conclusively determine if 3CM has value, and if it is worthwhile performing further work on this product.

If 3CM does not prove valuable, the whole cell biocatalyst developed here may still prove useful. A next step could be to determine the usefulness of the next metabolite along the vanillin catabolic pathway. This compound is a chiral lactone, 3-carboxy muconolactone, as shown in Figure 3, Section 1.3.3. This molecule may prove to have high inherent value through its chirality. It is also stable. Like 3CM, it has a high density of functionality. It contains two carboxyl groups and one double bond, so could be used as a polymer building block in a similar way to 3CM. The biocatalyst constructed here could be modified to generate this lactone product by the addition of a single gene, pcaB from A. baylyi, encoding 3-carboxymuconate lactonising enzyme [297].

8.4.2 Long term

If this system can be shown to produce a valuable product, there will be interest in developing it further than could be achieved by the suggestions offered in each chapter and in Section 8.4.1. The options that seem most obvious for this optimisation are modifying the system for the use of cheaper starting material, better understanding the impact of substrate concentration on the conversion rate and increasing the host cells robustness by exploring alternative options for expression of the recombinant enzymes.

The depolymerisation of lignin generates an enormously heterogeneous mixture of compounds. The metabolic strategy for the catabolism of such diverse mixtures in the environment is to employ convergent pathways. Enzymes with broad substrate specificity are often found in these convergent pathways. HBD had this feature, and this could be used in mimicking the convergent pathway strategy. For example, adding two additional enzymes could broaden the number of compounds which could be converted to 3CM from three to ten
(Figure 52). If the system were altered so that 3-carboxy muconolactone was the product, six recombinant enzymes would be required in total. The strategy of creating a biocatalyst with convergent pathways could allow greater yields of product from a low value crude starting material, such as depolymerised lignin.
Figure 52: Convergent pathway strategy for broadening the feed stocks for 3CM production
Substrate concentrations, as well as the amount of each enzyme present, affect the observed reaction rate of multi-enzyme systems. In this system, the concentration of the vanillin available to HBD was controlled by the rate that it enters the cell. Future work should endeavour to understand whether this is achieved by passive or facilitated means, and whether the flux through the pathway can be enhanced by manipulation of this. The concentration of all aromatic metabolites downstream of vanillin are controlled by the differential between the rates of the reactions producing and consuming it. The concentration of these intermediate metabolites would probably be best controlled by relative enzyme expression levels. The concentrations of the substrates which are shared with the host cell, nicotinamide dinucleotide cofactors and O₂, may be difficult to control, and even more difficult to alter. Any longer term work undertaken on optimising this system should aim to develop conditions which give reproducible concentrations of these substrates (eg chemostat), determine what the concentrations of these metabolites are, and then decide if the effort required to alter these concentrations would give a proportional benefit.

Alternative strategies for expression of recombinant enzymes may optimise this system by improving the robustness of the host organism. The commercial plasmid system used in this project offers convenient routes to further development of this system, but it was designed for overexpression of recombinant proteins. High levels of recombinant protein synthesis are not necessarily appropriate for a biocatalytic system. Simply increasing expression of all enzymes in a pathway does not necessarily achieve large increases of the flux through that pathway [298]. Overexpression of proteins in a recombinant host carries a large metabolic price, depleting energy, along with pools of metabolites associated with transcription and translation [39]. Optimal enzyme expression in a whole cell biocatalyst should not perturb the host excessively, and should aim to balance the activity of the constituent enzymes [299]. As such, longer term work on developing this system should weigh the convenience offered by the plasmid based system here against the disadvantages of high metabolic load and the requirement of costly antibiotics as selection markers. It may be that integrating the genes for the synthetic
pathway into the genome of the host may be preferable. It may also be that using a plasmid based system which offers expression which is linearly proportional to the amount of inducing agent would be preferable to allow easier optimisation of expression [300].
9 Appendices

9.1 DNA and Amino Acid Sequence for the Recombinant *hcaB*

Sequence with 100% homology with the reported sequence of *hcaB* from *A. calcoaceticus* BD413 is underlined (NCBI accession number gi|32306886|gb|L05770.5|ACCPCAO [301] and NCBI accession number gi|49529273|emb|CR543861.1| [148]).

ATGGGCAGCAGCCATACCATCACATCACACACCACAGCCAGGATCCGCAACATGTACAGTTACTT
M G S S H H H H Q D P Q H V Q L

ATTGATGGTCAGTCTGTGATGTGCTGCAATCAAGCCACCTTTGAACGCATTAGCCCAATAT
I D G Q S V D A A N Q A T F E R I S P I

GACGGTCATGTGGTCAAGTGTTGCTGCTGCAACACCTTTGAAGGCATGCTGACGTCGACTTT
D G H V A S V A A A A T L E D V D R A L

GAATCGGGTACGCAGGCTCATTTCAAAATCTGTGAAAATTTTCGCGACTGAACATCGTTTA
E S A S R A F Q I W S K V S P T E R R L

CGTTTGCCTCAAAGCCGACGATCTGATGGAATCAAATACTCGCAATGGTAAATGCTGAATCTGGT
R L L K A A D L M Q N T E K F I E I G

ATCGGTGAAACACGGTCAACAGCCACATGGTATGGCTTTAATGTACATCTGGCAGCCAAC
M R E T G S T A T W Y G F N V H L A N

ATGCCTACGCAGGCGCCGCCGATGACCACTCAATATGACGGAAGTTTATCGCCGCTCGTGT
M L R E A A M T Q I D G S L I P S D

GTACCGGGCAATCTGGCAATGGGAATTCGGGTGCCATGTGGTGTTGTAGTCGGGATCGCG
V P G N L A M G I R V P C G V V G I A

CCTTGGGAATGCACCTGTATTTTTGCAACCCGAGCAGCTTGCAATCTGGTTTATGCTGAAC
P W N A P V I L P T R A L A M P L A C G

AATACGGTAGTACTAAAAGCCTCCGAAGCTTGGCAGCTGCGCAACCCATCAGCTTGATCTGGC
N T V V L K A S E A C P A T H R L I G A

ATTTTAATGGAAGCTGTTTTGGAAGGTTGTTTAAATGTCATTACACATGACCTCAAA
I L N E A G L G E G V V N V I T H A P Q

GATGCAACCAACAGTCGTTGAGCTGTAGTTAGACGACAACCTGCGAGCATTAAACGCA
D A P Q V V E R L I E H P A V K R I N F

ACCGGTTCGACCAAAGTTGGTAAAAAATATTGCGCGGAACGCAAGGAGCATCTAAACCT
T G S T K V G I I A E T A A K H L K P
**DNA Sequence:**

```
GTTTTATTAGAACCTTGGTGTAAGGCCTGCTGTGGTATTAAAATGATGCCGATCTGGAT
V L L E L G K A P V V I L N D A D L D

GAAGCTGTGAATGCGGTTGCATTTGGTGCATTTTTCAATCAAGGCCAGATCTGTATGTCG
E A V N A V F G A F F N Q G Q I C M S

ACAGAAGGCTGTTTGGTATCCAGAGCATCAGCGATCGGTTTATGGAAGCTCATTCAA
T E R V L V Q D I A D R F I E K L I Q

AAAACAGCCTCACTTAAAGCAGGTAATCCTAAGCGAGCAAGCGAGCTATGTTGTTGTTACTT
K T A S L K A G N P T Q Q G S M L G V L

GAAAGCCGACGCCGCAGCAGCGAATTCCGACTTACTTACTTGAAGATTGCACAGCAAAAAAGGT
E S R R A A E R I Q H L L E D A Q Q K G

GCCAACCTTCGTTGGGATCCAGATTCAAGATACCTCTTCGCAACCCGACACTCTGACTG
A N L P L G I Q I O D T L M Q P T L V L

GATATTCAACCCGAATGTTGTTATGCGAAGAATCAATTTCGGGCCAGTTGTTACAGTA
D I Q P E M L L Y R E E S F G P V C T V

CAACGTTTTTGCTAGCATTGAAGACGGTGTCGCTGCTGGCAATGACAGATGTTGTTGTTTA
Q R F A S I E D G V R L A N D S E F G L

TCTTCGGCTGTATTAGCTAAAACTTTTTGGGCTGGGATGGATGTTGCCAAAACAGATGAC
S S A V F S Q N F G L A M D V A K Q I D

TCCGGGATTTGCCATATCAGGTGGAACCGTGACACGATGAAGCAAGCAAGATGCGCTTCGG
S G I C H I N G A T V H D E A Q M P F G

GGCACTAAAATCGAGTGGTATGGCAGTACAGATGAAAGCAGAGATGACCCCTTCGGT
G T K S S G Y G R F G S K A S I A E F T

GAGCTTCGCTGGATTACCCTACAACACAGCGCCAGCCATTCAAATTTAA
E L R W I T V Q T Q P R H Y P I *
```
9.2 DNA and Amino Acid Sequence for the Recombinant vanAB

Sequence with 100% homology with the reported sequence of vanAB from *A. calcoaceticus* BD413 is underlined (NCBI accession number gi|2271494|gb|AF009672.1|AF009672 [89] and NCBI accession number gi|49529273|emb|CR543861.1 [148]).

```
atggcagatctTTTTATTAAAAATGCCTGGTATGTCGCTTGCCGCCCAGAAGAAATTCAA
M A D L F I K N A W Y V A C R P E E I Q
GATAAAACCATTAGCCGTACGATTTTGTTGGAATAATCGTATTTTATCGTGGAAGAA
D K P L G R T I C G E K I V F Y R G K E
AATAAAGTGCAAGGCTGAGGATAGTTCTGTCTATCGTGCGCGCCATTGCTTTTGTT
N K V A A V E F C P H R G A P L S L G
TATGTGAAAGATGTCATCTAGTTTGGCTGTTATCATGTTTTAGTTATGGGATGTGAAGGA
Y V E D G H L C G Y H G L V M G E G
AAGACCATTGCTATGCGCTGCAAGCGTGTTAGGCGGCGGCTTCCCCTGCAATAAAGCATATGCT
K T I A M P A Q R V G G F P C N K A Y A
TTGGTCGAAAATACGGATTTTATTTGCGTTGGGCTGAAATCTCTGGCAAAATGAA
V V E K Y G F I W W P G E K S L A N E
```

```
GCAGACTTACCACACTCGAGATGGCGATCACATCCAGAAGTGAGTTATGCGGTTGGCTG
A D L P T L E W A D H P E W S Y G G L
TTTCATATCCAGTGTGATATTGTGCTCTCGTTGATGAACTCTGATGGAATTTGACTCATGAA
F H I Q C D Y R L M I D N L M D L T H E
ACTTATGTTCATTCCAGGAGATATCTGGCAAAAGAGAAATTTGATGAAACGATTGCGGCTGAA
T Y V H S S S I Q K E I D E A L P V T
AAAGTCGATGGACACATGTGGAAATCTCAGATCAGATCCAGATGGAATTTATGGCGCCCA
K V D G D H V V T E R Y M E N I I A P P
TTCTGGCAAAATGGCGTTACGTGGAATTTAGCTGATGATGTGCTGTGATGCGCTATG
F W Q M A L R G N H L A D D V P V D R W
CAACGGTGTTCATCTTTCGTACATGGAATAGTATTGGCAGATGGAATTTGGGCTGCGTCAT
Q R C H F F A P S N V H I E V G V A H A
GGTACATGGGTGTTATAACGCAACAAAAAGACAAAAAAGTGGCGTCTGCTGGTGCTCGATT
G H G G Y N A P D K K V A S V V V V D F
ATTACGCCTGAACAAGAAACGTCTCATGTGATATTTTGGGGCATGCGACATGAAATTCCAG
I T P E T E T S H W Y F W G M A R N F Q
```
GGAACACCAGATCATCGTGATGTATTTATGACAGATGAGGAACATGCTTTAAATGATCAG
G T P D H R D V F M T D E E H A L N D Q

TTTACACCGTGGTGGTGGCGGTGGCGAGATCTAAAATTTTAGTCATTGATCTTGCTTTAA
F T P C C S R A K S K I L V I D L A *
9.3 DNA and Amino Acid Sequence for the Recombinant pcaHG

Sequence with 100% homology with the reported sequence of pcaHG from A. calcoaceticus BD413 is underlined (NCBI accession number gi|49529273|emb|CR534861.1 [148]). The sequence also had 99% homology with another published sequence (NCBI accession number gi|32306886|gb|L05770.5|ACCPCAOP [302]).

ATGGCAGATTTTCTAAATATTGCTTGGGAGCTTACGCTCAATGCCGAAGATCAT
MADLSQIQIWFYARTNDH

CCGCCAGCTTATGCACACAGGTATATAAAAACAGTGTTTACGCTCCCTAAGATATGCCACCT
PPAYKTSTVLRSPKNA

ATTTCTATTGCAGAAACTTTAAGTGAAATGCTACCTGACCACATTTTAGGTATGATAAAATT
ISIAETLSTVAPHFSADKF

GGTCCAAAAAGACAATGACTTTAAGTGAACTATGCAACATATAGATGCTGATAAAATT
GPKDNILINYAKDGLPIGE

CGTGTCATGTGTCATGTGTTAGTTAGTCGACCTGTAAATGCAAGATT
RVIHGYVRDQFGRPVKNAL

GTTGAAATATGCAAGCAATGCACTCGTCGTTATGCATCCAAAATGATCAATATATC
VEVWQANASGRYRHPRNDQYI

GGTGCCATGGATCTTAATTTCGGTGGTTGCGGCATGAGTTGACTGATAAATGTTAT
GAMDPNGCGCRMLTDNHY

TATGTATTCGCGATAGTATGAGCAATCCGCTATACCCATGGCGAAGCGTATCAATGATGG
YVFRTIKGPYPWRNRINE

CGTGCTCTCAGATTTCTGTTCTTCTCTTTAATTTGAGATGTTGCTGGCACAACGTTAATTTCG
RPASHIFSLIDAGWWWQRRLIS

CAGTTCTATTTTGAAGGGCGATACATTGATGTATGTTTCTCTGGCACAACGTTAATTTCG
QFYFEGDTLIDSCPILTIP

TCTGACAAACAACGTGTGGCGCTAATGTTGAGATTAGAAAGACAAGCAATTTATTCATTAGCT
SEQQLRALKIADVSNFIEA

GACAGGCGGTGTTATCGGTTGCTACGTGGTGTCGCGCCCGGACTTTACTCGGA
DSRCYFDITLGRRRATYFEN

AATGACTTAATGATGAGTAAAAC
NDLT *
ATGAATGGTTGGAATTTTCAGGAACTCAAAGAAACCCCATCTCAAACAGGTGGTCCATAT
M N G W N F Q E L K E T P S Q T G G P Y

GTCCATATTGGTCTTTTGGCACCACAAGGCAATATCGAAGTATTATTGGAACACAATTTGGAT
V H I G L P K Q A N I E V F E H N L D

AACAAACTTGTACAGGACAATACGCAAGTGCAACCGATCCGATTGAAATCTGGCAAGCAGATACCAAT
N N L V Q D N T Q G Q R I R L E G Q V F

GATGGGCTTGGTTTACCCTTACGTGACGTACTGATTGAAATCTGGCAACGAGATACCAAT
D G L G L P L R D V L I E I W Q A D T N

GGCGTTTACCCAAAGTCAGTGATACCTGAAGTAAACAAGTGATCCGAACATTGGTGTTTTG
G V Y P S Q A D T Q G K Q V D P N F L G

TGGGGGCGTGACAGGTGCAGATTTTGGTACAGGCTTCTGGAGTTTTAATACCATTAAGCCA
W G R T G A D F G T G F W S F N T I K P

GGTGCTGTTCACCCAGCCGTAAGGGTCAACTCAAGCACCACCACATATTTCACGTGATTATTTT
G A V P G R K G S T Q A P H I S L I F

GCCTGCTGGTATCAACATGGTCTTTCACACTCGTGCTCTATTTTGATGACGAAGCAGACAGCA
A R G I N I G L H T R V Y F D D E A E A

AATGCAAAAGATCCTGTTTTAAACAGTATCGAGTGGGCGACAGTGTCAAAACTGGTT
N A K D P V L N S I E W A T R R Q T L V

GCCAAACGTGAAAGGCTGTGATTGATTTCTGGTTTATCGATATTCTGTATTTCAAGGT
A K R E E R D G E V V Y R F D I R I Q G

GAAAACGAAAGGCTATTTTTTCGATATCTAA
E N E T V F F D I *
9.4 DNA and Amino Acid Sequence for Multiple Cloning Site 1 of Recombinant Plasmid pACYC-HBD

ATGGCGAGCCAGCCTACCACCATCACCACATCACCACAGCCAGATCCGCAACATGTACGTTACTT
MGSSHHHHHHSQDPQHVQLL

ATGTGATGGTCAGTCTGCTGATGCTGCAAAACATCGAGATCTCTGAGATCTCAGTTTCTT
MGSSHHHHHHHHHSHHSQATFERISPI

GACGGTCATGTTGCAAGTGGTGCTGCTGCAACACTTGGAGATGTGATCGTGACTTT
DGHVASVAATLEEDVDRAL

GAATCCGCTAGCCGCTGCAATTCCAAATCTGGTCAAAGATTTCGCCGACTGAACGTCGTTTA
ESASRAFAQIWSKVSPTRRL

CGCTTGCTCAAAGCGGCAAGTATGAGGATCAAATCCGAAAGTTATCGGAAATGCGGTT
RLLKADLMQNTEKFIEIG

ATGCGGAAACAGTGGTCAACACGCAATAGGTATGGCTTTAATGTACACATCTGCGCCAGCAAC
MRETGSTAATWYGFNVHLAAN

ATGCTACGCAGCCGCGCCGCTGACGCAACACTCAAAATCGACGAAATACATCGCTGAT
MLREAAMMTQIDGSLPDS

GTACCGGCCAATCTGGCAATGGGAATTCCGGGTGCTGCTGCTGATGCTGCGGCAATGCGG
VPGNLAMGIRVPCGVVVGIAG

CCTTGGGAAATGCACCCTGTTATTTTTGCGCAACCCCGAGACCTTGGCAATTTGCTGCTG
PWNAVPVILPTRALAMPACG

AATACGGTGTAGTACTCAAAGCTCGGAAAGTTGGGCAACCCGCACATCGCTTGATAGGTGCA
NTVVKLKASEACPATHRLIGA

ATTTAAAAATGAGCTGTTTGGGTTAGGTTGTTGTTAATGTTGATTACATGATCGACCTCAA
ILNEAGLGEGVNVNITHAPQ

GATGCACCAACAAAGTCTGTTAGGCTGCTGATGAAACACCCCTGCAATTAAACGCGATTAATT
DAPQVVVERIHEHPAVKINF

ACCGGTTGCAAAAAGTGGTGTTAAATATTGCGGAAAACGCGAAGCATCTTAAACCT
TGSTKVGKIIIAETAAKHLKP

GTTTATTAGAAACTTGGTGGTAAAGCGCGCCTGTTGGTGATTTTAAAATGATGCGGCAATG
VLLLGGKAPVVIILNDADLD

GAAGCTGATGATGGGCTGATTGGCATTGATTGGCATGTTTTGAAATGATGCGGCAATG
EAVNAVAFGFAFNQGQICMS

ACAGAAGCTGTTTGGTCAAGAGCGGCATTGCAATCGGTTTTATGGAAAGCTCATTTCA
TERVLQDGIADRFIEKLIQ

AAAAACACCCCTACCTAAAAGAGCTATCACCACGCGCAAGCGGAGTGATGTTGGGTACTTT
KTSALKAGNPQTQGSMGLGVL
9.5 DNA and Amino Acid Sequence for Multiple Cloning Sites 1 and 2 of Recombinant Plasmid pET-VMO-P34O

Coding sequence for P34O

ATGGGCGACGCCATCCACCACATCATCACACCACGAGCCAGATCTTTCTCAAATTATTGGGGA
M G S S H H H H H H H S Q D L S Q I I W G

GCTTACGTCAGCTGTAATACCGAAGATCATCCGACGCTTATGCACCGTTATATAAAACC
A Y A Q R N T E D H P P A Y A P G Y K T

AGTGTGTACAGCTGCTCTAAGAAATGACTCTCATTCTATTCGAGAAACTTTAAGTGAGTC
S V L R S P K N A L I S I A E T L S E V

ACTGCACCATTTTATATGCTGTTTAAATTTGTCACAAAGCAGATATGACTTAAATCTTGAC
T A P H F S A D K F G P K D N L I N

TATGCCAAGATGTTTGTCCGACAGTGTTGTTGGATCGTACATTGTCATTTATCTGACG
Y A K D G L P I G E R V I V H G Y V R D

CAGTTTGTCGACCTCTGAATATCCGACTTTGTGAAGTATGCAAGCCATGCTCTGGT
Q F G R P V K N A L V E V W Q A N A S G

CGTTATCGCTACTCAAATGATCATATATACGATGGCATCTAATTTCTGGAGTTGT
R Y R H P N D I G A M D P N F G G C

GGCCGTATGTTGACGTATGATAATATGTTTATTATGATTATCCCGTACGTTAAACCGAGTTCCA
G R M L T D D N G Y Y V F R I K P G P

TACCCATGGGCTAAACGTATCAATGAATGCGCTCCGCTCATTGACTTTCTCTTAAATT
Y P W R N R I N E W R P A H I H F S L I

GCAGATGTGTGGGTCCACAGTTTATATATCGAGTCTATTTGAAAGCGATACATGATT
A D G W A Q R L S Q F Y F E G D T L I

GATTCTTGGCAATCTCTGAACACCATTTCTTCTCAGCCAACAGCTGTCGTGCCCTATTTGCA
D S C P I L K T I P S E Q Q R R A L I A

TCTAGAGAACAAAGACCAATTTTCTTCTGAAGCTGACACCGTATTGTTCATCTCAGCACATTACG
L E D K S N F I E A D S R C Y R F D I T

CTTCGGTCGTGCGCCGCGACTTTCTCCCTTACTTTATGACTTTATGATGAGTAAAAA
L R G R R A T Y F E N D L T *

AC

ATGAATGTGTTGGAATTTTCAGGAACTCAAATTAAACCCATCTCCACAAAAAGGTTGCTCCCATAT
M N G W N F Q E L K E T P S Q T G G P Y

GTCATATTGCTTCTTTGCTACAAACACAGCCAATATCGAAGTTATTTGCAACACAATTTGGAT
V H I G L L P K Q A N I E V F E H N L D
AACAACCTTGTACAGGACAATACGCAAGGTCAACGCATCCGATTAGAAGGTCAAGTGTTT
N N L V Q D N T Q G Q R I R L E G Q V F
GATGGGCTTGTTTACCTTCCTACGTGACGTACTGATTGAAATCTGGCAAGCAGATATCAAT
D G L G L P L R D V L I E I W Q A D T N
GGCGTTTACCAAAATGCTAAGGTGATATCTCAAGGTGAATGAAATCTGGCAAGCTCAAC
G V Y P S Q A D T Q G K Q V D P N F L G
TGGGCGCTACAGGTCGATTTTTGATACGGCTTCTGGAGGTTTTAATACCCATTAAAGCCA
W G R T G A D F G T G F W S F N T I K P
GGTGCCTGTCCAGCCGTAAGGTCGATATCTCAAGGTGAATGAAATCTGGCAAGCTCAAC
G A V P G R K G S T Q A P H I S L I I F
GGGACGTGATCAAGCATTTGCTTCACACTCGTGCTTATATTGATGACGAAGCAGAAGCA
A R G I N I G L H T R V Y F D D E A E A
AATGCAAACATCTCTGTTTTTTAACACGTACGTACGTGCCAGAAGCTCAACACTTGTT
N A K D P V L N S I E W A T R Q T L V
GCCAAACGGTGAAAGGGTGATTTGTGAAGGTTTTATCGTTTTATCGATATCTCGATTTGAGTT
A K R E E R D G E V V Y R F D I R I Q G
GAAAACGAAAACGGTATTTTTCTCGATATCA
E N E T V F F D I *

Sequence between the two ORFs
CTCGACAAGCTTGCGGCCGCATAATGCTTAAGTCGAAACAGAAATATCGATTGACAC
G G A D L F I K N A W Y V A C R P E E I Q
GATAAACCATTAGGCCTGATTTGTTGAGGTTTTATCTGGCAAGCATTTATGCTCCAGAAGAA
D K P L G R T I C G E K I V F Y R G K E
AATAAAAGTGCTAGGCATTTGTTAGGCTTTATCTGGCAAGCATTTATGCTCCAGAAGAA
Y V E D G H L V C G Y H G L V M G C E G
AAGGCGCTGATGCTGCGCAGCTGAGTGTTGCTTGCATATCTGCTGCTGATGCTGCTGCTGCTG
K T I A M P A Q R V G G F P C N K A Y A
GTTGTCGAACAATACGGATTTATTTGGGTTTGGCCAGGTGAAAAATCTCTGGCAAATGAA
V V E K Y G F I W V W P G E K S L A N E
GCAGACCTTACCACACACTCGAATGGGCAGATCATCCAGAGTTTGGGCTGGGCTTG
A D L P T L E W A D H P E W S Y G G L
TTTATATCCAGTGTGATTATCGTCTCATGATTGATAACCTGATGGAATTTGACCTGAA
F H I Q C D Y R L M I D N L M D L T H E
ACTTATGTTCATTCGACGATATCAGTCCAAAAAGAAATGTGAAAGCATATTGCCAGTAACC
T Y V H S S S I G Q KE I E A L P V T
AAAGTCGATGGAGACCAGATTGGAATCTGAACGCTACATGGAAAAATTATTATGCGGCAAC
K V D G D H V T E R Y M E N I I A P P
TTCTGGCAAATGGCGTTACGTGGAAATCATTTAGCTGATGTGGCTGTGAGTCTGG
F W Q M A L R G N H L A D D V P V D R W
CAACGGTGTCACTTTCTTTGCACCTTAGTAAATGCTCCATATTGAAATGCGGGTG
G C T G A C T G E G G T G A T G T G G T G G C T A T G C G G
Q R C H F F A P S N V H I E V G V A H A
GGTCATGGTGTGTTATAACGCACCAAAAAAGCAAAAGATCGCTCTCGATGTGGTGCTGATT
G H G G Y N A P D K K V A S V V D F
ATTAGCCCTGAAACAGAAACGTCTCAGCTGATTTTTTGGGCAAGCGTAAATTTCCAG
I T P E T E T S H W Y F W G M A R N F Q
CCAGAACAAGCAAACATCGAGCGAGGAGATGTGAAAGGACAAGATTTAAAGCCGAA
P E N Q Q L T D Q I R E G Q G K I F T E
GATTTAGAAATGCTGGAACAGCAGCAGCAGAAAAATATTTCTACGTAACCCACAGCGC
D L E M L E Q Q C Q Q N I L R N P Q R Q L
TTGATGCTGAAATATCGATGCTGCGGGGTTCTCAATCGCAAGATTATTTGACCTGTTACT
L M L N I D A G G V Q S R K I I D R L L
GCAAAAGAAAATAGCCGGCTCACCTCAAGATACACAACAGGAAATTTCCCAATATTTGCTGTA
A K E N S P S P Q D T Q R K F P N I R V
ATTTAA
I *
GG
ATGATAAAACATGGACGTGATCATTCACAAAAATTTACCGCTTTAATTCCAGTCACTG
M I N M D V I I H K I H Q L T P S I R A
TTTGAGTTGATCTCTGCAATGGCAGCGCAGCTTTACCGCTTTTGAGCGGATGCTCAGATATT
F E L I S A N G S D L P A F D A G S H I
GACGTACATCTCAAAAATGGAATGGACTCGGCAAATATTGACCTTTCTAATCTGTTGAGTGA
A D V H L K N G L T R Q Y S L S N C C S E
CAACATCGCTATGTGATTGGTGTATTGCACGATGAAAATTCAAGAGGTGGTTCACGTTGT
Q H R Y V I G V L H D E N S R G G S R C
ATTCACTCAGGATTATCCTGTAAGGCGGATCATTAAAATATTGTAACCAGCGCAATTTATTTT
I H Q D Y R E G D H L N I G T P R N L F
GAAATTCAATTCTAAAAACTCAAGAAAGCAGTTTTATTTATGCGAGCGGATTTGAGATTACGCCT
E I H S K T V K F A G G I G T P
ATTCTCTCATTGGCTTATCCTCTAAAACATCGAACAATTCCATTTGAACATACATATTATTTT
I L S M A Y R L K H Q Q I P E L H Y F
GTACGTAGTCATGAAATGATTGATTTATGCGAACCCTGACTGAACCTCCAGAGCAGA
V R S H E M I A F Y G N L T E H F P E Q
ATTCTTTTTCATCTTCCAATCATGCTCTGAAACACAATGTGAATGTGAAATGGCTGAA
I H F H I Q N Q S E T Q C E M S K V L E
GAAGTAGCTCCAGATCGACATTTATATGCTGTGCTGCTGAGCTGTTTTATGCAATTAGTG
E V A P D R H L Y V C G P A G F M Q F V
ATGGAGTTCACAGAAGCAGGATTTGTCGATGAAAGCTCTACATGAAAGAATGGCTGTAAT
M D S A Q Q A G W S D E Q L H Q E H F V
GCACGGCAATACGAAATGCTGATGAAAGCTCTTATACATTGAAGTGCTGGTGCTGAGA
A P Q I D Q S Q N E A F T I E V L G S D
CGTAAAATTGAAGTTCTTCTGCCCTACAGCAGGGCACTCAAGCGCTTGCTGGAACATG
R K I E V S A H Q T A T Q A L L E H G F
GACGTGGCTTTATGTGAAACAGGATTTTGCCACCGCTATTACGCGCGTGTTTCT
D V P V S C E Q G I C G T C I T R V V S
GGAAACCAGATACATGCTGATGATTTATGACAGATGGAAACATGCTTTTAATGATCAG
G T P D H R V M F T D E H A L N D Q
TTTACACCGTGTTGTTTCCGGGGCGAAGTCTCAAATTTTTAGTCATTGATCTTGA
F T P C C S R A K S K I L V I D L A *
9.6  $^{13}$C-NMR Spectra of Copoly Styrene and 3CM-Me$_3$

This material was synthesised with 33 mol% 3CM-Me$_3$ in the feed
9.7 Comparison of the FTIR Spectra of Copoly Styrene 3CM-Me<sub>3</sub> and Homopolystyrene

This material was synthesised with 10 mol% 3CM-Me<sub>3</sub> in the feed and homopolystyrene
9.8 Average Molecular Weight of Copoly 3CM-Styrene and Dimethyl Styrene with Varied Comonomer Composition

![Graph showing the average molecular weight of copoly 3CM-Styrene and Dimethyl Styrene with varied comonomer composition. The graph plots the average molecular weight (kDa) against the mol% ester comonomer in feed. Two types of comonomers are indicated: 3CM-Me₃ styrene and dimethyl fumarate styrene.](image-url)
## 9.9 Spectral Characteristics of Compounds Assessed for UV Reactivity

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>$^1$H-NMR Shifts (ppm)</th>
<th>FTIR Bands between 2000 and 1000 cm$^{-1}$ (cm$^{-1}$)</th>
<th>UV Spectral Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dibenzyl ammonium $E,E$-muconate</td>
<td>3-A$_2$</td>
<td>D$_2$O 4.20 (s, CH$_2$NH$_3^+$, 4H), 6.15 (dd, J = 12Hz, CH=CHCO$_2^-$, 2H), 7.00 (dd, J = 12Hz, CH=CHCO$_2^-$, 2H), 7.45 (s, Ar, 10H)</td>
<td>1620 (m), 1573 (m), 1540 (m), 1454 (w), 1364 (s), 1229 (w), 1189 (w), 1172 (w), 1103 (w), 1073 (w)</td>
<td>10mM phosphate, pH 7.0 $\lambda_{max}$ = 260nm $\varepsilon$ = 20600 L. mol$^{-1}$. cm$^{-1}$</td>
</tr>
<tr>
<td>Dinaphthyl methyl ammonium $E,E$-muconate</td>
<td>3-I$_2$</td>
<td>CD$_3$OD 4.53 (s, CH$_2$, 4H), 6.15 (dd, J = 12Hz, CH=CHCO$_2^-$, 2H), 7.11 (dd, J = 12Hz, CH=CHCO$_2^-$, 2H), 7.57 (m, Ar, 8H), 7.93 (t, J = 9 Hz, Ar, 4H), 8.10 (d, J = 9Hz, Ar, 2H)</td>
<td>1630 (m), 1615 (m), 1546 (m), 1510 (s), 1443 (w), 1381 (s), 1364 (s), 1297 (w), 1266 (w), 1192 (w), 1154 (w), 1073 (w)</td>
<td>CH$<em>3$OH $\lambda</em>{max}$ = 260nm $\varepsilon$ = 23500 L. mol$^{-1}$. cm$^{-1}$</td>
</tr>
<tr>
<td>3-carboxy $E,Z$-muconic acid-1-methyl ester</td>
<td>2</td>
<td>D$_2$O 3.78 (s, -OCH$_3$, 3H), 6.21 (d, J = 12Hz, CH=CHCO$_2^-$, 1H), 6.78 (s, -C(CO$_2^-$)=CHCO$_2^-$, 1H), 7.22 (d, J = 12Hz, CH=CHCO$_2^-$, 1H)</td>
<td>1721 (m), 1705 (s), 1688 (m), 1648 (m), 1607 (m), 1442 (m), 1418 (m), 1380 (w), 1346 (m), 1273 (s), 1230 (m), 1197 (m), 1186 (w), 1099 (w), 1011 (w)</td>
<td>10mM phosphate, pH 7.0 260nm (shoulder) $\varepsilon$ = 9870 L. mol$^{-1}$. cm$^{-1}$</td>
</tr>
<tr>
<td>Trisodium 3-carboxy $E,Z$-muconate</td>
<td>1-Na$_3$</td>
<td>D$_2$O 5.95 (d, J = 12Hz, CH=CHCO$_2^-$, 1H), 6.4 (s, -C(CO$_2^-$)=CHCO$_2^-$, 1H), 6.6 (d, J = 12Hz, CH=CHCO$_2^-$, 1H)</td>
<td>1669 (w), 1650 (w), 1597 (s), 1568 (m), 1539 (m), 1457 (w), 1410 (m), 1380 (m), 1342 (w), 1185 (w), 1083 (w)</td>
<td>10mM phosphate, pH 7.0 $\lambda_{max}$ = 253nm $\varepsilon$ = 8340 L. mol$^{-1}$. cm$^{-1}$</td>
</tr>
<tr>
<td>Compound</td>
<td>Structure</td>
<td>$^1$H-NMR Shifts (ppm)</td>
<td>FTIR Bands between 2000 and 1000 cm$^{-1}$ (cm$^{-1}$)</td>
<td>UV Spectral Characteristics</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>------------------------</td>
<td>-------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Tribenzyl ammonium 3-carboxy E,Z-muconate</td>
<td>1-A$_3$</td>
<td>D$_2$O 4.18 (s, CH$_2$NH$_3^+$, 6H), 5.93 (d, $J$ = 12Hz, CH=CHCO$_2^-$, 1H), 6.39 (s, -C(CO$_2$)=CHCO$_2^-$, 1H), 6.60 (d, $J$ = 12Hz, CH=CHCO$_2^-$, 1H), 7.48 (s, Ar, 15H)</td>
<td>1626 (m), 1606 (m), 1535 (s), 1508 (s), 1453 (w), 1422 (w), 1397 (w), 1355 (s), 1292 (w), 1217 (w), 1159 (w), 1121 (vw), 1106 (vw), 1089 (vw), 1073 (vw), 1010 (vw)</td>
<td>10mM phosphate, pH 7.0 λ$_{max}$ = 255nm ε = 6600 L. mol$^{-1}$. cm$^{-1}$ $R^2$ = 0.951</td>
</tr>
<tr>
<td>Triphenyl ethyl ammonium 3-carboxy E,Z-muconate</td>
<td>1-B$_3$</td>
<td>D$_2$O 3.04 (t, $J$ = 7Hz, CH$_2$CH$_2$NH$_3^+$, 6H), 3.30 (t, $J$ = 7Hz, CH$_2$CH$_2$NH$_3^+$, 6H), 5.97 (d, $J$ = 12Hz, CH=CHCO$_2^-$, 1H), 6.42 (s, -C(CO$_2$)=CHCO$_2^-$, 1H), 6.64 (d, $J$ = 12Hz, CH=CHCO$_2^-$, 1H), 7.38 (m, Ar, 15H)</td>
<td>1639 (m), 1605 (m), 1532 (s), 1498 (m), 1455 (w), 1420 (w), 1397 (w), 1355 (s), 1289 (w), 1210 (w), 1159 (w), 1116 (vw), 1032 (vw)</td>
<td>10mM phosphate, pH 7.0 λ$_{max}$ = 255nm ε = 9070 L. mol$^{-1}$. cm$^{-1}$ $R^2$ = 0.954</td>
</tr>
<tr>
<td>Triphenyl butyl ammonium 3-carboxy E,Z-muconate</td>
<td>1-C$_3$</td>
<td>D$_2$O 1.67 (m, -CH$_2$(CH$_2$)$_2$CH$_2$NH$_3^+$, 12H), 2.69 (t, $J$ = 7Hz, CH$_2$NH$_3^+$, 6H), 2.88-3.00 (t, $J$ = 7Hz, ArCH$_2$CH$_2$, 6H), 5.96 (d, $J$ = 12Hz, CH=CHCO$_2^-$, 1H), 6.41 (s, -C(CO$_2$)=CHCO$_2^-$, 1H), 6.63 (d, $J$ = 12Hz, CH=CHCO$_2^-$, 1H), 7.33 (m, Ar, 15H)</td>
<td>1636 (m), 1561 (s), 1497 (m), 1454 (w), 1419 (w), 1388 (m), 1360 (s), 1290 (w), 1212 (w), 1125 (w), 1028 (vw)</td>
<td>10mM phosphate, pH 7.0 λ$_{max}$ = 255nm ε = 10300 L. mol$^{-1}$. cm$^{-1}$ $R^2$ = 0.950</td>
</tr>
<tr>
<td>Compound</td>
<td>Structure</td>
<td>$^1$H-NMR Shifts (ppm)</td>
<td>FTIR Bands between 2000 and 1000 cm$^{-1}$ (cm$^{-1}$)</td>
<td>UV Spectral Characteristics</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>------------------------</td>
<td>-----------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Triadamantyl ammonium 3-carboxy $E$,$Z$-muconate</td>
<td>1-D$_3$</td>
<td>CD$_3$OD 1.72 – 1.84 (m, CH$_2$, 36H), 2.12 (s, CH, 9H), 5.94 (d, $J = 12$Hz, CH=CHCO$_2$-, 1H), 6.56 (s, -C(CO$_2$-)=CHCO$_2$-, 1H), 6.65 (d, $J = 12$Hz, CH=CHCO$_2$-, 1H)</td>
<td>1614 (m), 1601 (m), 1554 (s), 1421 (w), 1392 (w), 1364 (s), 1316 (w), 1290 (w), 1211 (w), 1174 (vw), 1128 (w), 1092 (w), 1042 (w)</td>
<td>CH$_3$OH 260nm (shoulder) ε = 3080 L. mol$^{-1}$. cm$^{-1}$ R$^2$ = 0.950</td>
</tr>
<tr>
<td>Tritolyl ethyl ammonium 3-carboxy $E$,$Z$-muconate</td>
<td>1-E$_3$</td>
<td>D$_2$O 2.20 (s, -CH$_3$, 9H), 2.85 (t, $J = 2$Hz, -CH$_2$CH$_2$NH$_3^+$), 3.15 (t, $J = 2$Hz, -CH$_2$CH$_2$NH$_3^+$), 5.80 (d, $J = 8$Hz, CH=CHCO$_2$-, 1H), 6.3 (s, -C(CO$_2$-)=CHCO$_2$-, 1H), 6.5 (d, $J = 8$Hz, CH=CHCO$_2$-, 1H), 7.15 (m, Ar, 12H)</td>
<td>1640 (m), 1607 (m), 1529 (s), 1459 (w), 1421 (w), 1353 (s), 1290 (w), 1213 (w), 1164 (vw), 1121 (vw), 1036 (vw)</td>
<td>10mM phosphate, pH 7.0 $\lambda_{max} = 256$nm ε = 7560 L. mol$^{-1}$. cm$^{-1}$ R$^2$ = 0.988</td>
</tr>
<tr>
<td>Tri((4-methoxy phenyl) ethyl ammonium) 3-carboxy $E$,$Z$-muconate</td>
<td>1-F$_3$</td>
<td>D$_2$O 2.94 (t, $J = 7$Hz, CH$_2$CH$_2$NH$_3^+$, 6H), 3.23 (t, $J = 7$Hz, CH$_2$CH$_2$NH$_3^+$, 6H), 3.84 (s, OCH$_3$, 9H), 5.95 (d, $J = 12$Hz, CH=CHCO$_2$-, 1H), 6.41 (s, -C(CO$_2$-)=CHCO$_2$-, 1H), 6.63 (d, $J = 12$Hz, CH=CHCO$_2$-, 1H)</td>
<td>1641 (m), 1612 (s), 1583 (m), 1514 (vs), 1464 (w), 1442 (w), 1420 (w), 1398 (w), 1355 (s), 1301 (w), 1249 (s), 1177 (w), 1151 (vw), 1118 (w)</td>
<td>10mM phosphate, pH 7.0 $\lambda_{max} = 270$nm ε = 6110 L. mol$^{-1}$. cm$^{-1}$ R$^2$ = 0.987</td>
</tr>
<tr>
<td>Tri(3,4-methoxy benzyl ammonium) 3-carboxy $E$,$Z$-muconate</td>
<td>1-G$_3$</td>
<td>D$_2$O 3.85 (s, -OCH$_3$, 18H), 4.20 (s, -CH$_2$NH$_3^+$), 5.95 (d, $J = 12$Hz, CH=CHCO$_2$-, 1H), 6.4 (s, -C(CO$_2$-)=CHCO$_2$-, 1H), 6.65 (d, $J = 12$Hz, CH=CHCO$_2$-, 1H), 7.05 (m, Ar, 12H)</td>
<td>1667 (m), 1592 (m), 1516 (s), 1464 (m), 1418 (m), 1404 (m), 1375 (w), 1356 (w), 1263 (s), 1237 (m), 1159 (m), 1140 (m), 1026 (m)</td>
<td>10mM phosphate, pH 7.0 $\lambda_{max} = 272$nm ε = 10300 L. mol$^{-1}$. cm$^{-1}$ R$^2$ = 0.953</td>
</tr>
<tr>
<td>Compound</td>
<td>Structure</td>
<td>(^1)H-NMR Shifts (ppm)</td>
<td>FTIR Bands between 2000 and 1000cm(^{-1}) (cm(^{-1}))</td>
<td>UV Spectral Characteristics</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>-------------------------</td>
<td>-------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Tripiperonyl ammonium 3-carboxy E,Z-muconate</td>
<td>1-H(_3)</td>
<td>D(_2)O 1-H(_3) CD 3OD</td>
<td>1597 (m), 1562 (s), 1531 (s), 1501 (m), 1491 (m), 1445 (m), 1423 (vw), 1387 (m), 1355 (s), 1293 (w), 1275 (vw), 1249 (s), 1216 (vw), 1200 (vw), 1122 (vw), 1099 (w), 1036 (m)</td>
<td>10mM phosphate, pH 7.0 (\lambda_{\text{max}} = 279\text{nm} ) (\varepsilon = 10700\ \text{L. mol}^{-1}.\text{cm}^{-1}) (R^2 = 0.971)</td>
</tr>
<tr>
<td>Trinaphthyl methyl ammonium 3-carboxy E,Z-muconate</td>
<td>1-I(_3)</td>
<td>CD(_3)OD</td>
<td>1625 (m), 1601 (m), 1534 (s), 1505 (s), 1421 (w), 1392 (w), 1352 (s), 1292 (w), 1265 (vw), 1218 (vw), 1173 (vw), 1146 (vw), 1079 (vw), 1055 (vw), 1024 (vw)</td>
<td>CH(<em>3)OH (\lambda</em>{\text{max}} = 260\text{nm} ) (\varepsilon = 8100\ \text{L. mol}^{-1}.\text{cm}^{-1}) (R^2 = 0.951)</td>
</tr>
<tr>
<td>Tri(1,2,3,4-tetrahydro naphthyl(^1)-ammonium) 3-carboxy E,Z-muconate</td>
<td>1-J(_3)</td>
<td>CD(_3)OD</td>
<td>1714 (w), 1643 (s), 1561 (m), 1557 (m), 1548 (s), 1491 (m), 1450 (m), 1440 (w), 1434 (w), 1365 (m), 1275 (w), 1163 (w), 1096 (vw), 1037 (vw)</td>
<td>CH(<em>3)OH (\lambda</em>{\text{max}} = 257\text{nm} ) (\varepsilon = 6920\ \text{L.mol}^{-1}.\text{cm}^{-1}) (R^2 = 0.924)</td>
</tr>
<tr>
<td>Compound</td>
<td>Structure</td>
<td>$^1$H-NMR Shifts (ppm)</td>
<td>FTIR Bands between 2000 and 1000 cm⁻¹ (cm⁻¹)</td>
<td>UV Spectral Characteristics</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-----------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>Tri(3, 3-diphenyl propyl ammonium) 3-carboxy $E,Z$-muconate</td>
<td>1-K₃</td>
<td>CD₃OD 3.1 (m, CH₂CHNH₃⁺, 6H), 4.4 (t, CH₂CHNH₃⁺, 3H), 5.9 (d, J = 12Hz, CH=CHCO₂⁻, 1H), 6.6 (s, -C(CO₂⁻)=CHCO₂⁻, 1H), 6.8 (d, J = 12Hz, CH=CHCO₂⁻, 1H), 7.2 (m, Ar, 30H)</td>
<td>1665 (m), 1643 (m), 1597 (m), 1558 (s), 1550 (m), 1493 (s), 1450 (m), 1406 (m), 1362 (m), 1249 (w), 1154 (vw), 1116 (vw), 1074 (vw), 1030 (w)</td>
<td>CH₃OH $\lambda_{\text{max}}$ = 256nm $\varepsilon = 5950$ L.mol⁻¹.cm⁻¹ $R^2 = 0.973$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-L₃</td>
<td>CD₃OD 2.35 (m, CH₂CH₂NH₃⁺, 6H), 2.75 (m, CH₂CH₂NH₃⁺, 6H), 4.00 (t, J = blahbalh CHCH₂CH₂NH₃⁺, 3H), 5.87 (d, J = 12Hz, CH=CHCO₂⁻, 1H), 6.55 (s, -C(CO₂⁻)=CHCO₂⁻, 1H), 6.65 (d, J = 12Hz, CH=CHCO₂⁻, 1H), 6.95 (m, Ar, 30H)</td>
<td>1754 (m), 1710 (m), 1643 (s), 1601 (m), 1566 (s), 1549 (s), 1495 (s), 1453 (m), 1391 (m), 1275 (w), 1227 (vw), 1188 (vw), 1160 (w), 1074 (w), 1029 (w)</td>
<td>CH₃OH $\lambda_{\text{max}}$ = 252nm $\varepsilon = 3830$ L.mol⁻¹.cm⁻¹ $R^2 = 0.954$</td>
</tr>
</tbody>
</table>
9.10 Change in the FTIR Spectrum of Dinaphthyl Methyl Ammonium $E,E$-Hexadienedioate Due to UV Exposure.

The spectrum in green was measured for UV untreated I$_2$-HDA, and the spectrum in blue was measured for UV treated I$_2$-HDA.
9.11 No Change in the FTIR Spectrum of Dibenzyl Ammonium $E,E$-Hexadienedioate Due to UV Exposure.

The spectrum in black was measured for UV untreated $A_2$-HDA, and the spectrum in blue was measured for UV treated $A_2$-HDA.
10 References


75. Peng, X., et al., Characterization of Sphingomonas aldehyde dehydrogenase catalyzing the conversion of various aromatic aldehydes to their carboxylic acids. Applied Microbiology and Biotechnology, 2005.


133. Zacher, A.H., et al., *Conversion of beta-hydroxy carbonyl compounds to form conversion products including, e.g. acrylic acid used as feedstock and/or end-use chemicals by converting material to reaction products having alpha-beta-unsaturated carbonyl compound.* 2007, Battelle Memorial Inst (Batt) Cargill Inc (Crgi).


144. Matsumoto, A., S. Nagahama, and T. Odani, Molecular design and polymer structure control based on polymer crystal engineering. Topochemical polymerisation of 1, 3-diene mono- and dicarboxylic acid derivatives bearing a naphthylmethylammonium group as the countercation. Journal of the American Chemical Society, 2000. 122: p. 9109.


176. Martins, B.M., T. Svetlitchnaia, and H. Dobbek, 2-Oxoquinoline 8-
Monooxygenase Oxygenase Component: Active Site Modulation by Rieske-[2Fe-
177. Tarasev, M., et al., Similar enzymes, different structures: Phthalate dioxygenase
is an α3α3 stacked hexamer, not an α3β3 trimer like "normal" Rieske
180. Scopes, R., Separation by Adsorption, in Protein Purification. Principles and
181. Rosche, B., et al., 2-Oxo-1,2-dihydroquinoline 8-monooxygenase, a two-
component enzyme system from Pseudomonas putida 86. Journal of Biological
182. Rucker, P., F.M. Torti, and S.V. Torti, Recombinant ferritin: modulation of
subunit stoichiometry in bacterial expression systems. Protein Eng., 1997. 10(8):
p. 967-973.
183. Rigaut, G., et al., A generic protein purification method for protein complex
1030-1032.
184. Shine, J. and L. Dalgarno, Determinant of cistron specificity in bacterial
185. Ma, J., A. Campbell, and S. Karlin, Correlations between Shine-Dalgarno
Sequences and Gene Features Such as Predicted Expression Levels and Operon
186. Wackett, L.P., Mechanism and applications of Rieske non-heme iron
187. Pinto, A., M. Tarasev, and D.P. Ballou, Substitutions of the bridging aspartate
178 result in profound changes in the reactivity of the Rieske center of phthalate
188. Fernandez, J., et al., Purification and characterization of Acinetobacter
calcoaceticus 4-hydroxybenzoate 3-hydroxylase after its overexpression in
189. Sundararaj, S., et al., The CyberCell Database (CCDB): a comprehensive, self-
updating, relational database to coordinate and facilitate in silico modeling of
D293.D295.
190. Walton, A.Z. and J.D. Stewart, Understanding and Improving NADPH-
Dependent Reactions by Nongrowing Escherichia coli Cells. Biotechnology
191. Sánchez, A.M., G.N. Bennett, and K.-Y. San, Effect of different levels of NADH
availability on metabolic fluxes of Escherichia coli chemostat cultures in defined
192. Penfound, T. and J.W. Foster, NAD-Dependent DNA-Binding Activity of the
Bifunctional NadR Regulator of Salmonella typhimurium. Journal of
Dehydrogenase. Spectral Deconvolution of the Neutral and Anionic Flavin


251. O'Shea, M.S., Bulk Copolymerisation of Multifunctional Siloxane Macomers with Styrene, in Department of Chemistry. 1991, University of Queensland: Brisbane.


