UNIVERSITY COLLEGE LONDON

Novel chassis for engineered pathways and automating biological experiments in the biotechnology industry

A thesis submitted to University College London for the degree of Doctorate of Engineering in Biochemical Engineering

Aisha Mariam Asra

April 2019

Declaration

‘I, Aisha Mariam Asra confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

Abstract

For the pharmaceutical industry, there has been a movement towards greener chemistry to reduce environmental impact. This movement has sparked interest from the pharmaceutical and chemicals industry in synthetic biology, which involves increasing the use of biological hosts and enzymatic processes to produce pharmaceutical products. This increase in biocatalysis and synthetic biology has sparked more of a need for diverse and versatile enzymes in order to customise the end product as well as diverse and robust host organisms to accommodate the reactions. Synthetic biology also opens opportunities for the discovery of complex molecules which have been inaccessible by traditional synthetic chemistry routes. Using enzymes from the toluene metacleavage pathway found in *Pseudomonas putida* and an omega-transaminase, an engineered metabolic pathway was constructed in *Escherichia coli* and in various *P. putida* strains for the production of novel amines. In this study the conversion of benzoate to 2-hydroxymuconic semialdehyde (2-HMSA) has been investigated in a whole cell biotransformation reaction to compare the feasibility of using *P. putida* to *E. coli* as an alternative host organism for industrial processes. *P. putida* KT2440 showed activity comparable to *E. coli,* indicating that this could be a suitable organism for use in industry. Nineteen omega-transaminases were screened to select a transaminase which was able to convert 2-HMSA to a novel amine. Results showed that *(R)* – selective transaminase appeared to have activity with 2-HMSA, with an arthrobacter mutant transaminase (ArRMut11) showing the highest conversion rates. The conversion of modified starting materials, p-toluic acid, m-toluic acid and 2,3-dimethylbenzoate were investigated with the engineered pathway. Identifying the amine products of the transaminase conversion of 2-HMSA and the modified starting substrates was attempted using a combination of HPLC and mass spectrometry, however the exact structures were not confirmed.

Acknowledgements

Thanks to Professors John Ward and Helen Hailes for being such great supervisors and for their incredible knowledge and support. Thank you also to the past and present members of the Ward and Hailes labs, in particular Benjy Lichman, Jack Jeffries, Maria Bawn, Nadine Tappertzhofen, David Jackson, Fabiana Subrizi, Leona Leipold and Sophie Newgas for creating such a fantastic working environment and helping me to work out why nothing was working. Thanks also to Synthace Ltd, particularly Lily Nematollahi, Steve Kendrew, Chris Grant and Markus Gershater for your support and guidance.

The biggest thank you's to my partner Dhru for your encouragement when I felt most demotivated. And to my parents, Tracey and Ashraf for giving me the faith & determination to follow my interests and for always supporting me through everything unconditionally.

Thanks to the EPSRC and Synthace Ltd for funding this project.

Impact statement

The studies presented here could be impactful in a number of ways which are described below.

Impact in academia

The transaminase screening with the substrate 2-hydroxymuconic semialdehyde (2- HMSA) has laid the groundwork for future studies of transaminases for the conversion of compounds with similar structures. The investigation of the amino acid amine donors which showed glutamate was accepted by the transaminases more readily than alanine or methylbenzylamine may guide future research.

This study has been a proof of concept study to show that a novel molecule can be created using an engineered pathway involving enzymes from different species and expressed in an alternative host organism. This could impact future research by providing some guidance on which *Pseudomonas putida* organisms may have growth and productivity rates which are most comparable with *E. coli.*

Impact in industry

During this research project novel molecules have been created showing a proof of concept that engineering metabolic pathways may be a relevant and promising method for drug discovery. Here *P. putida* have been shown to have comparable efficiency to *E. coli* and literature suggests that solvent tolerability is improved; thus making *P. putida* a potential interesting alternative host for industrial manufacturing processes.

In this project, the usability of Antha software has been fed back to the developing company, Synthace Ltd. and this research has helped to improve

the programming within the software so that it is more flexible and user friendly for scientists performing biological experiments.

Table of contents

Table of Tables

Table of Figures

List of abbreviations

Abbreviation	Definition
2-HMSA	2-hydroxymuconic semialdehyde
ANOVA	Analysis of variance
API	Active pharmaceutical ingredient
AQC	6-aminoquinolyl-N-hydrosysuccinimidyl carbamate
bp	Base pairs
DNA	Dioxyribose nucleic acid
DMSO	Dimethyl sulfoxide
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
C18	18 carbon chain
EDTA	Ethylenediaminetetraacetic
ee	Enantiomeric excess
ESI-TOFMS	Electrospray ionisation time of flight mass spectrometry
GRAS	Generally regarded as safe
HPLC	High performance liquid chromatography
HPSF	High purity salt free
iGEM	International genetically engineered machine competition
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IUPAC	International Union of Pure and Applied Chemistry
Kb	Kilobases
lac	Promoter which is involved in the expression of enzymes for lactose metabolism, originating in *E. coli*
*lac*UV5	Promoter which is involved in the expression of enzymes for lactose metabolism, originating in *E. coli* varying by 2 base pairs with the *lac* promoter
LB	Lysogeny broth
LCMS	Liquid chromatography mass spectrometry
Ltd	Limited
MBA	Methylbenzylamine
NEB	New England Biolabs
NMR	Nuclear magnetic resonance spectroscopy
OD	Optical density
pben	Promoter originating from *P. putida* controlling the metabolism of benzoate
PCR	Polymerase chain reaction
Pfam	Database of protein families
PHA	Polyhydroxyalkanoate
PLP	Pyridoxal 5’ phosphate
Pm	A promoter originating on the TOL plasmid of *P. putida* controlling the metabolic pathway involved in the degradation of toluene
PMP	Pyridoxamine 5’ phosphate

PTE	Polythioester
R&D	Research and development
RPM	Rotations per minute
SEVA	Standard European vector architecture
tac	Promoter which is derived from the *trp* and *lac* promoters
TAm	Transaminase
TBE	Tris Borate EDTA buffer
TFA	Trifluoroacetic acid
TOL	Toluene
tRNA	Transfer ribonucleic acid
trp	Promoter involved in the expression of tryptophan synthesis genes originating from *E. coli*
xylE	catechol-2,3-dioxygenase
xylL	1,2-dihydroxycyclohexa-3,5-diene-carboxylate dehydrogenase
xylS	Transcriptional regulator of the *pm* promoter
xylT	Ferredoxin associated with the TOL *meta*cleavage pathway
xylX	subunit of toluate-1,2-dioxygenase
xylY	subunit of toluate-1,2-dioxygenase
xylZ	subunit of toluate-1,2-dioxygenase
ε	Extinction coefficient
λ_{max}	Lambda max

Chapter 1: Introduction

Chapter 1. **Introduction**

The idea of synthetic biology has been around for a few decades (Benner & Sismour 2005; Hobom 1980), the technology and the interest in this field of research has recently gained momentum. Genetic manipulation is now common place in a laboratory, particularly in microbes. Along with the rapid advances of DNA synthesis technology; genetic engineering is becoming cheaper and faster than ever (Turner 2018). The ability to design a biological system to carry out pre-determined functions is the ultimate goal of synthetic biology. To achieve this; enzymes and metabolic systems must be better characterised and a wider range of chassis or host organisms must also be better understood. This will allow for improved selection of the appropriate enzymes and host organisms to efficiently produce molecules biologically.

There has been a drive to reduce environmental damage caused by all industries. The chemical industries in particular, frequently use non-renewable materials, producing high levels of toxic and chemical waste and using extreme amounts of energy in their processes. The pharmaceutical industry is one of these industries that is under pressure to change manufacturing processes to more green methods of chemical production (Woodley et al. 2013; Wohlgemuth 2010; Tucker 2006). Biological systems work under aqueous conditions and rarely produce any type of toxic waste. This property makes engineering biological systems to produce chemicals and ingredients for the pharmaceutical industry an ideal alternative to current methods of chemical synthesis (Sheldon 2016; Stewart 1997). Limitations to using biological systems however are lower yields produced by enzymes and non-optimised reactions and potentially longer processing times purifying the product from aqueous solution containing cell debris, enzymes and other biological contaminants. To increase the use of biocatalysis in research and development (R&D) and chemical manufacturing, biological processes must be more efficient and more reliable. This is the idea that the fields of synthetic biology and bioengineering aim to achieve.

1.1 Combining biology and engineering

In order to be successful in both designing synthetic pathways for the production of novel molecules and producing high yields of the desired molecule in living micro-organisms it is important to combine the strategic planning and design of engineering with the biochemical knowledge of metabolic pathways. For this to be successful engineering strategies will need to be adapted to suit biological systems (Andrianantoandro et al. 2006).

Engineering biological systems is part of the relatively new discipline of synthetic biology which must be distinguished from molecular biology. Molecular biology has allowed the manipulation of genetic information and has facilitated the study of macromolecules, especially proteins and many cell processes. From molecular biology, tools have been developed so that it is easy and simple to produce any desired protein from a recombinant organism. Synthetic biology uses the tools and knowledge from molecular biology combined with engineering principles to work towards an ultimate goal to be able to standardise, model and predict the biological reactions and processes (Farny 2018). Biology is complicated and unpredictable whereas engineering is made as simple as possible and is standardised. Combining these two disciplines will require adaptation of the normal principles of engineering. These main principles of engineering are standardisation, modularity and decoupling (Endy 2005).

Standardisation is quite self-explanatory; in engineering everything is standardised, all engineers work in the same units and all measurements can be easily understood and translated between any lab in any country. Within biology, however, such strict standards do not exist. Measuring protein concentration varies between labs as it depends on the comparison between the sample and a known protein with known concentration. Measuring the

optical density during cell growth can also vary between labs, some researchers read the absorption at a wavelength of 600 nm, some at 650 nm and some even at 700 nm. In order to progress towards the idea of using synthetic biological components as tools or as a technology, measurements must be standardised so that anyone at any level of understanding of synthetic biology would be able to design a process using a given set of synthetic biology 'tools'. Modularity is essentially simplifying a process. Engineers are able to categorise different levels of complexity and work on a particular level with little or no concern for the rest. Biologists, in general, take more of an interest in the more complex problems and want to explore and solve these; they are viewed as research problems. Another concern for biologists is that the complexities may interfere with the desired result of the research and therefore cannot be simply ignored. The last of the engineering principles that needs to be addressed is decoupling; this is the idea that one complex problem can be resolved by dividing the problem into different aspects and solving these as separate simpler problems. Once these are solved it is possible to combine the separate resolutions for a result to the original complex problem. Similar to abstraction, solving biological problems is often more complex than this. When combining different biological components, it is likely that they will interact in a different way than when observed as separate problems, and therefore the initial problem would not be solved. These issues with bringing engineering principles into biology are discussed further in (Endy 2005).

It is unlikely that it is possible to simply apply the principles of engineering to biology; these principles must be adapted to accommodate the complexity of biological systems. That said, by assuming that the ultimate goal of synthetic biology is to have a set of 'tools' that can be picked out of a catalogue and assembled to produce an object or process with a particular function then biology must become more like engineering, and engineering should not become more like biology. To progress towards this we need a better

understanding of which biological components will work together and which are not compatible. From this idea “BioBrick” has been formed (Knight 2003). BioBricks are a standardised set of genetic elements which can be combined in a mix and match style and have been designed to produce consistent and reproducible results every time. A list of standardised biological parts can be found at the international genetically engineered machine competition’s (iGEM) list of registered parts http://parts.igem.org/Main_Page. Biological parts have been defined here as a nucleic acid sequence that encodes for a defined function which perform as expected without the variability which is often seen in biological experiments (Knight 2003).

Defining and standardising all possible biological parts, however, would be extremely time consuming. New genomes are constantly being sequenced and new functions of genomic DNA are still being uncovered. Improving technologies is speeding up this process, for example it is now possible to produce DNA synthetically whereas in the past cloning DNA could take months and Endy, 2005, states that cloning and producing the correct DNA constructs for an experiment easily takes up to 50% of a molecular biologists time and efforts. It is important to have standardised parts because if synthetic biology is going to become an industrial and globally used technology, then the biological components need to be more predictable. Mistakes in DNA replication are unpredictable and provide one example where biology must be improved to progress towards the success of synthetic biology.

On the other hand, biology is already like engineering as the Barabasi lab, 2004, (Barabási & Oltvai 2004) compare metabolic networks within a cell to man-made networks such as the internet and transport networks. They demonstrate that cell networks are scale-free networks, meaning they have busy hubs which interact with many less busy nodes; in a cell the hubs and nodes are proteins. The hubs are proteins such as the tumour suppressor

phosphoprotein p53 which interacts with hundreds of other proteins but the majority of other proteins have very few interaction partners. In transport links, the hubs are busy airports such as London Heathrow and New York City's JFK which have incoming and outgoing flights from hundreds of other airports but smaller airports such as the East Midlands Airport only interact with a few other airports. By comparing cell networks and the way a cell functions with these man-made networks, the question that arises is: could biology be viewed as an engineering problem? This property of cell networks also shows that there is a possibility that biology could be more predictable than scientists currently think. By bringing engineering strategies to biology we can increase the productivity and economic viability of using microbes to biosynthesise molecules. Characterising pathways and different biological host species will help to progress towards the goal of synthetic biology to make biological experiments more predictable.

1.2 Synthetic biology and metabolic engineering

To progress towards the ideal of synthetic biology and to create microbial factories for the production of useful molecules there must be an increased understanding and ability to engineer metabolic pathways. By adding several genes in sequence within the same vector it is possible to construct a novel biosynthetic pathway, given that the product of one enzyme is a possible substrate of another. Native biosynthetic pathways limit chemical production to the products found in nature, whereas manipulating and engineering these pathways allows the optimisation to produce yields high enough for industrial use and also to alter functional groups and produce novel molecules by introducing enzymes with different functionalities into a pathway. There is huge potential for a variety of molecules, but to access these we must have full usage of the enzymes required; this may mean that *E. coli,* which is the current workhorse of biological research, is not a sufficient host for the

reactions we are interested in. Pfeifer et al., 2001, demonstrated the use of *E. coli* in the production of active pharmaceutical ingredients such as polyketides. One of the most well-known and widely used of these is the antibiotic, erythromycin. Large scale production of erythromycin is currently synthesised by the bacteria *Saccharopolyspora erythraea* (Wu et al. 2011). This is a working example that shows the importance of microbial engineering in the production of industrially relevant chemicals. Here the organism used industrially is the organism in which the working enzymes naturally occur; proving that *E. coli* is not always the most useful host for biosynthesis and that other species can be and are industrially relevant. One reason for this is that organisms have evolved to perform native metabolic reactions in the most energy efficient way (Weeks & Chang 2011). Native genetic sequences have codon biases to accommodate tRNA abundances in their natural host organism to promote correct folding and efficient translation.

Biocatalysis and synthetic biology in industry has recently gained huge momentum, which has resulted in the development of biological based processes for a wide range of products from biofuels to pharmaceutical ingredients. Polyketides are one class of molecules that have been used widely in pharmaceutical ingredients, pesticides and herbicides. Polyketides are synthesised by polyketide synthases, which have recently been investigated in the production of biofuels due to the wide variety of enzymes which provide the ability to tailor the product (Cai & Zhang 2018). Though this method of biofuel production is not yet efficient enough to compete with oil extraction, other methods could be, such as the break down of biomass, such as lignocellulose (Bilal et al. 2018) or sugar beet pulp (Berłowska et al. 2016) to produce bioethanol. Engineering metabolic pathways is not only limited to biofuels, many other high value products have been produced as a result of engineered pathways; one such example is the production of anti-malarial drug, artemisinic acid using engineered yeast. Artemisinic acid is

traditionally extracted from plants; making the drug expensive and difficult to produce in large amounts. The pathway engineering was performed here by combining an existing mevalonate pathway for the formation of farnesyl pyrophosphate (FPP) with enzymes, amorphadiene synthase gene (ADS) from *A. annua* and a novel cytochrome P450 to produce an engineered multi-enzyme pathway for the conversion of simples sugars to artemisinic acid (Ro et al. 2006). It is not only naturally occurring compounds that have been produced using engineered enzymatic cascades. For example, the lengthy and complex manufacturing process for the production of atorvastatin, which is used for controlling cholesterol in the prevention of cardiovascular disease and one of the top 3 prescribed drugs in the United States (ClinCalc 2019), has been replaced by a biocatalytic conversion involving three enzymes. The enzymatic process involved a ketoreductase, a halohydrin dehydrogenase and a glucose dehydrogenase expressed in *E. coli* and used as lyophylisates *in vitro* (Ma et al. 2010).

1.2.1 The TOL *meta*cleavage pathway

The toluene (TOL) pathway, found in *Pseudomonas putida* mt-2 on the pWWO plasmid first isolated by Worsey & Williams in 1975, contains the xylene (xyl) enzymes involved in the catabolic pathway which breaks down aromatic compound such as xylenes, including toluene as shown in Figure 1-1 (Murray et al. 1972). The TOL *meta*cleavage pathway is interesting firstly in terms of degrading pollutants and secondly each enzyme produces a molecule that could be further manipulated into useful molecules (Abril et al. 1989). Examples of this are demonstrated by Reineke and Knackmuss (1979) who utilised enzymes from the TOL *meta*cleavage pathway and created a novel pathway with enzymes that are able to modify halo-aromatics. The resulting host was able to produce Krebs cycle intermediates

and therefore energy from halo-aromatic compounds and thus degrade pollutants from organic chemistry processes.

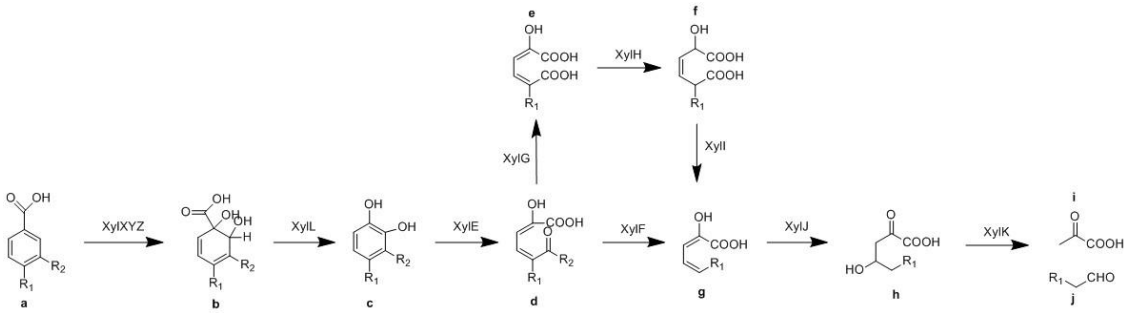


Figure 1-1. The TOL *meta*cleavage pathway is naturally found on the pWWO plasmid in *P. putida.*
The metacleavage pathway is the lower pathway of a large metabolic pathway for the catabolism of aromatic compounds such as toluene into Krebs cycle intermediates. Enzymes: toluate-1,2-dioxygenase is a multisubunit complex encoded by the *xylXYZ* genes; 1,2-dihydroxycyclohexa-3,5-diene-carboxylate dehydrogenase encoded by *xylL*; catechol-2,3-dioxygenase encoded by *xylE*; 2-hydroxymuconic semialdehyde hydrolase encoded by *xylF*; 2-hydroxymuconic semialdehyde dehydrogenase encoded by *xylG*; 4-oxalocrotonate tautomerase encoded by *xylH*; 4-oxalocrotonate decarboxyloase encoded by *xylI*; 2-hydroxypent-2,4-dienoate hydratase encoded by *xylJ*; 4-hydroxy-2-oxovalerate aldolase encoded by *xylK*. Intermediates: a; benzoic acid, b; 1,6-dihydroxycyclohexa-2,4-diene-1-carboxylic acid c; catechol, d; 2-hydroxymuconic semialdehyde, e; 2-hydroxyhexa-2,4-dienedioic acid, f; 2-hydroxyhex-3-enedioic acid, g; 2-hydroxyhexa-2,4-dienoic acid, h; 4-hydroxy-2-oxohexanoic acid, i; pyruvate, j; propanal

Another example of where useful and industrially interesting molecules have been produced by manipulating the TOL *meta*cleavage pathway is in the production of indigo (N. Mermod et al. 1986). Indigo is a blue pigment used in dyeing fabrics and is produced using chemical synthesis methods in industry, however this process produces toxic waste products and projects are on-going with the aim to find an industrially viable process for the production of indigo that is cheaper and more environmentally friendly (Qu et al. 2012). Mermod et al., (1986) used dioxygenases from the TOL *meta*cleavage pathway in their engineered pathway for the production of indigo. This demonstrates the importance and usefulness of engineering synthetic pathways and the usefulness of the TOL pathway enzymes in

particular. Due to the reactive carboxylic and aldehyde groups; many of the intermediates in the pathway have the potential to be pharmaceutically relevant with some modification. The full DNA sequence of this plasmid has allowed the isolation and cloning of single genes so that individual enzymes can be used in engineered pathways to create novel interesting molecules (Horn et al. 1991). Therefore the TOL pathway is an interesting starting point for engineered pathways to create new biosynthetic routes to novel and industrially relevant molecules.

1.2.2 Amines in industry

One particular class of chemicals that are highly relevant in industry is amines. From pharmaceuticals to dyes to fertilisers, amines are used heavily in industry, though many are difficult to synthesise (Thoralf Gross et al. 2002; Ghislieri & Turner 2013). Amines can be particularly reactive and require protecting during synthesis, adding more steps in already laborious reactions (Woodley 2008). Amines are often found in natural compounds and are therefore frequently biologically active, making them particularly relevant and interesting to the pharmaceutical industry. Natural amino acids are highly relevant to industry and the global market size is estimated at around 6.19 million tons in 2013 (Grand View Research Inc 2017). Amino acids are important in human nutrition and feedstock for animals. The amino acid that is used most widely worldwide is L-glutamate, which is used in the food additive monosodium glutamate; approximately half a million tons of L-glutamate are produced per year (Breuer et al. 2004). Abundant amino acids, such as L-cysteine, can be purified from proteins and biological material (Leuchtenberger et al. 2005); however this does not provide the high yields necessary to meet demand; therefore developing enzyme catalysis methods for large scale production would greatly improve the production of such molecules (Breuer et al. 2004). Ghislieri et al., 2013, have demonstrated the use of monoamine oxidases in the deracemisation of amines and the relevance to the pharmaceutical industry in the production of active

pharmaceutical ingredients such as pyrrolidines and isoquinolines. These compounds are found in a wide variety of pharmaceuticals including analgesics, such as morphine, and antifungals (Lichman et al. 2015). Using enzymes in the production of chiral amines is of particular interest due to the highly specific and enantioselective nature of enzymes. In the production of pharmaceuticals, often only one enantiomer of a chiral compound will have the biological activity necessary. An example of this is the production of D-phenylglycine using D-specific acylases which is an important precursor for the antibiotics ampicillin and amoxicillin (Leuchtenberger et al. 2005). The importance of having enantiomerically pure pharmaceuticals is demonstrated by thalidomide (Figure 1-2). The R-enantiomer of Thalidomide is used as a sedative, however the S-enantiomer causes embryonic deformities(Ghislieri & Turner 2013). Enantiomerically pure compounds are therefore much favoured by regulatory authorities and pharmaceutical companies.

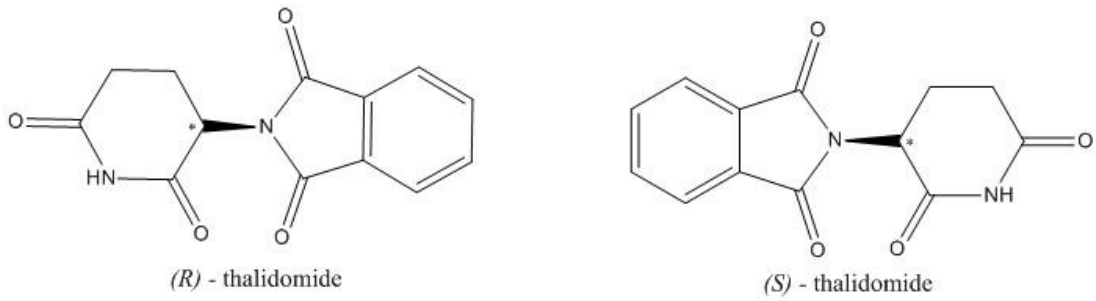


Figure 1-2. Thalidomide.
The (*R*)- enantiomer is a useful pharmaceutical as a sedative; however in the 1960s it was found that the (*S*)-enantiomer caused embryonic deformities.

Producing enantiomerically pure drugs is also preferable as the drugs often have a higher efficacy and lower dosages are required (Ghislieri & Turner 2013). One of the enzymes that has gained interest over the past decade for selective amination in biocatalysis are the transaminases (or aminotransferases) (D. Patil et al. 2018; Gomm & O'Reilly 2018).

1.2.3 Transaminases

Transaminases or aminotransferases are a group of enzymes that are involved in nitrogen *metab*olism within all living organisms and are important in the transfer of an amino group between molecules. Transaminases are found in every organism and therefore there is a huge variety available that have been studied (Guo & Berglund 2017; Höhne et al. 2010). Transaminases are interesting from a biocatalytic point of view as they can be extremely sterio-selective but can also be promiscuous in terms of substrate (Humble & Berglund 2011; Kazlauskas & Bornscheuer 2012). This means that transaminases may be versatile tools in the formation of enantiomerically pure and chiral amine molecules and aldehydes as their promiscuity may allow a range of substrates and therefore a range of product structures. Enantiomerically pure amine molecules are high value products, particularly in the pharmaceutical industry. Another advantage of transaminases is that they only require pyridoxal 5' phosphate (PLP) as a cofactor. PLP is a derivative of vitamin B6 that is inexpensive recycled during the reaction. PLP binds to a lysine on the transaminase and acts as an intermediate in the transfer of the amine group to the aldehyde substrate, becoming pyridoxamine phosphate (PMP). Once the amine has been transferred from PMP to the aldehyde substrate, PLP is reformed, meaning that there is no need to continuously add more cofactor. This is also advantageous as the requirement for expensive cofactors that are often necessary for efficient enzyme function is eliminated. Likewise, expensive organic chemicals and synthetic chemistry processes that would be required for the chemical synthesis of enantiomerically pure amines are also not necessary. Transaminases were initially used to produce amino acids (Stewart 2001) but this now has been expanded so that other amine group containing compounds such as unnatural amino acids, amino sugars and more can be produced via biocatalysis using transaminases (Lee et al. 2005; Malik et al. 2012; Richter et al. 2015; Simon et al. 2013).

Transaminases have great potential to be useful in engineering synthetic pathways and substrates and products are not only limited to amino acids. By taking advantage of the low substrate specificity of transaminases, many new molecules could be produced, examples to demonstrate the variety of possible molecule types range from aminopolyols and sugar derivatives (Figure 1-3) to serinol monoesters (Figure 1-4) (Costa et al. 2017; Subrizi et al. 2019) which are both pharmaceutically relevant as they are pharmacologically active (Ávalos et al. 2008; Gómez & Varela 2009; Teruhiko Ishikawa et al. 2000; Magrioti et al. 2003). New amine molecules and new keto or aldehyde molecules could be produced, the only limitations being that the molecule must always be either an amine donor or amine acceptor in order to be a substrate for the reaction. However this is where building synthetic pathways is useful. Enzymes can be introduced upstream in the pathway in order to produce a suitable amine or keto substrate for the transaminase which is downstream. On the other hand transaminases could be used upstream in the pathway to create amine or keto molecules to feed downstream enzymes. Ager et al. 2001, have demonstrated the use of a novel engineered pathway converting threonine to aminobutyric acid that include transaminases for the production of non-proteinogenic amino acids which are used as pharmaceutical intermediates, thus highlighting the industrial significance of transaminases. In this study *E. coli* was used as the host organism and whilst commercially feasible yields were obtained, the need for a better cell host is also brought to the readers' attention due to the catabolism of some intermediates by background metabolic processes in the cell (Ager et al. 2001). One solution that is given here is to remove or inactivate chromosomal genes from *E. coli* that may cause interfering background reactions. Another more interesting speculation is the increasing usage of other organisms as host cells as genomic and metabolic information for more organisms becomes available. Therefore manipulating a wider variety of organisms in the way that *E. coli* are currently manipulated genetically will become easier and more common. This will allow better

optimisation of biocatalytic processes by providing a wider choice of organisms with varying properties, such as optimal growth at lower or higher temperatures or a lack of certain metabolic background reactions, so that the organism that is most suited to the desired reaction processes can be chosen to reduce background and increase yield of the desired product.

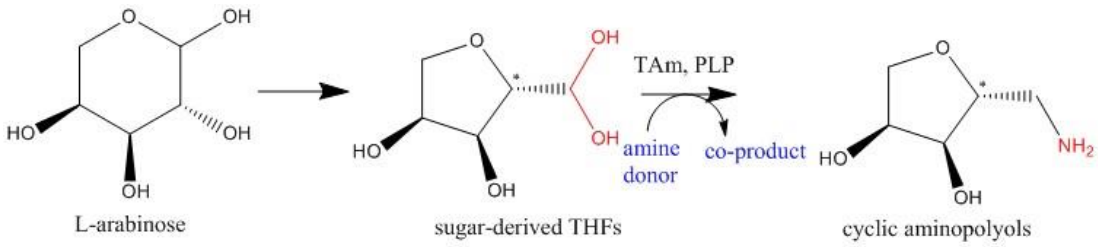


Figure 1-3. Schematic showing the conversion of L-arabinose to cyclic aminopolyols using a transaminase.
TAm: transaminase, PLP: pyridoxal 5' phosphate, THF: tetrahydrofuran(Subrizi et al. 2019)**.**

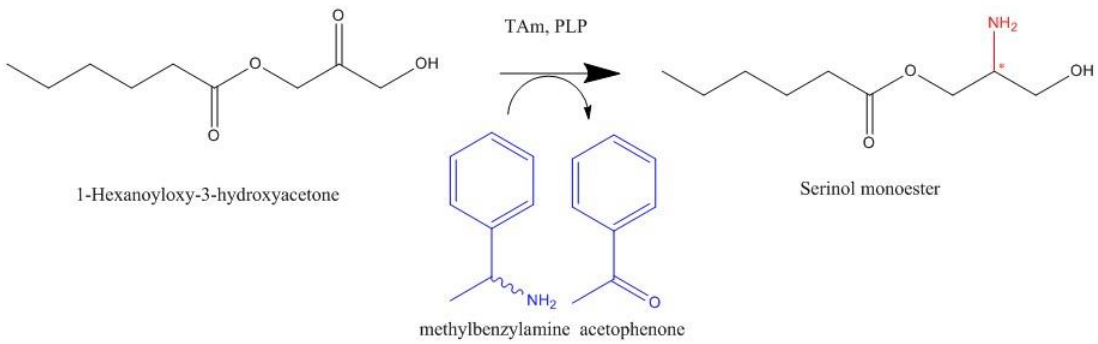


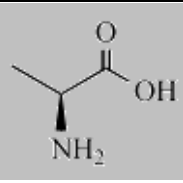
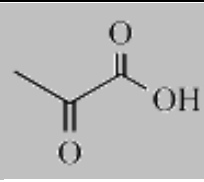
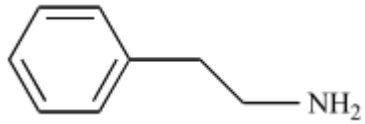
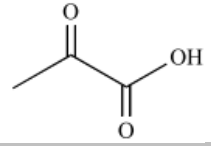
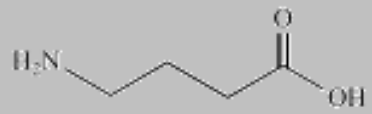
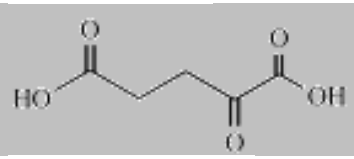
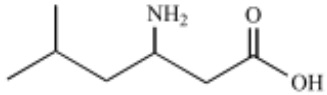
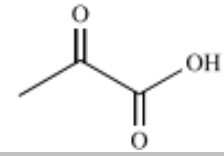
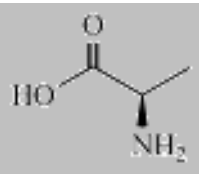
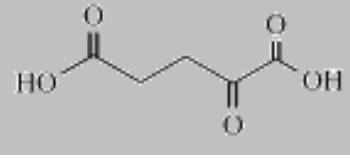
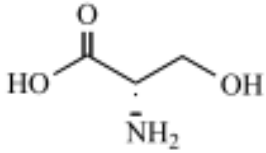
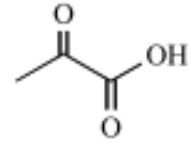
Figure 1-4. Schematic showing the conversion of 1-hexanoyloxy-3-hydrooxyaceone to serinol monoesters using a transaminase
(Costa et al. 2017)

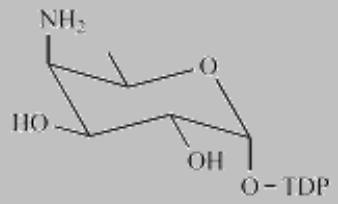
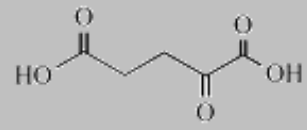
1.2.3.1 Transaminase classes

One property of transaminases is that they are so abundant in nature that it may be possible to find a transaminase that will work with any desired substrate. There are 6 transaminase subgroups, classified by the database of protein families (Pfam), which catalyse the conversion of specific molecule structures (Hwang et al. 2005; Finn et al. 2014). Class I and II includes *L*-

alanine, *L*-aspartate, aromatic and histidol-phosphate aminotransferases. Class III includes omega (ω), ornithine and 4-aminobutyrate transaminases. Class IV includes branched chain amino acid transaminases and *D*-alanine transaminases. Class V includes serine and phosphoserine transaminases and a further class including sugar aminotransferases, sometimes referred to as class VI. Class I, II, IV and V transaminases catalyse conversion when the amine or keto acid group is at the α position and are very substrate specific; class III transaminases catalyse conversion at the ω, β or λ position and have gained a high level of interest recently due to their wider acceptance of substrates (Ward & Wohlgemuth 2010; Hwang et al. 2005). The classes of transaminases are outlined in Table 1-1 with an example of each class and their positional specificity.

Table 1-1 Examples of transaminases in each class and their main donor and acceptors. Adapted from Ward & Wohlgemuth 2010

Class	Positional specificity	Example	Main donor	Main acceptor
I and II	alpha	Alanine TAm		
III	omega	Omega amino acid TAm		
III	gamma	Gamma aminobutyrate TAm		
III	beta	beta-aminocarboxylic acid TAm		
IV	alpha	D-alanine TAm		
V	alpha	Serine TAm		

Class	Positional specificity	Example	Main donor	Main acceptor
VI	Not applicable	TDP-4-amino-4,6-dideoxy- D-glucose TAm		

1.2.3.2 Omega transaminases

Named because of the position of the moiety that they are active upon, ω-transaminases are widely used in a variety of research in the field of biocatalysis. These particular transaminases are interesting not only because of the conversion that they catalyse, but because of their substrate promiscuity. A key limitation of these enzymes when it comes to industrial biocatalysis is that the reaction exists as an equilibrium meaning that completion is never achieved. Figure 1-5 shows the conversion of *(S)*-methylbenzylamine to acetophenone by ω-transaminases.

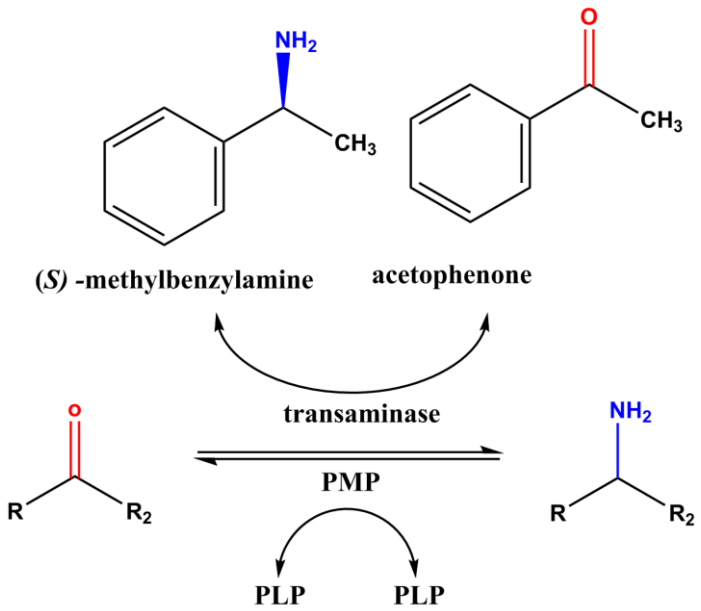


Figure 1-5. Conversion of aldehydes by (*S*)- selective transaminases using methylbenzylamine (MBA) as the amine donor.
Using pyridoxal 5' phosphate (PLP) as a cofactor, the amine moiety is transferred from the amine donor to the aldehyde substrate. Pyridoxamine phosphate (PMP) is generated as the intermediate and once the reaction is complete, PLP is regenerated.

Use of methylbenzylamine (MBA) as an amine donor is common when screening a large number of transaminases for activity against a substrate or set of substrates. There are several examples where ω-transaminases have been used in industry. Such as in the production of the antidiabetic compound, sitagliptin, shown in Figure 1-6.

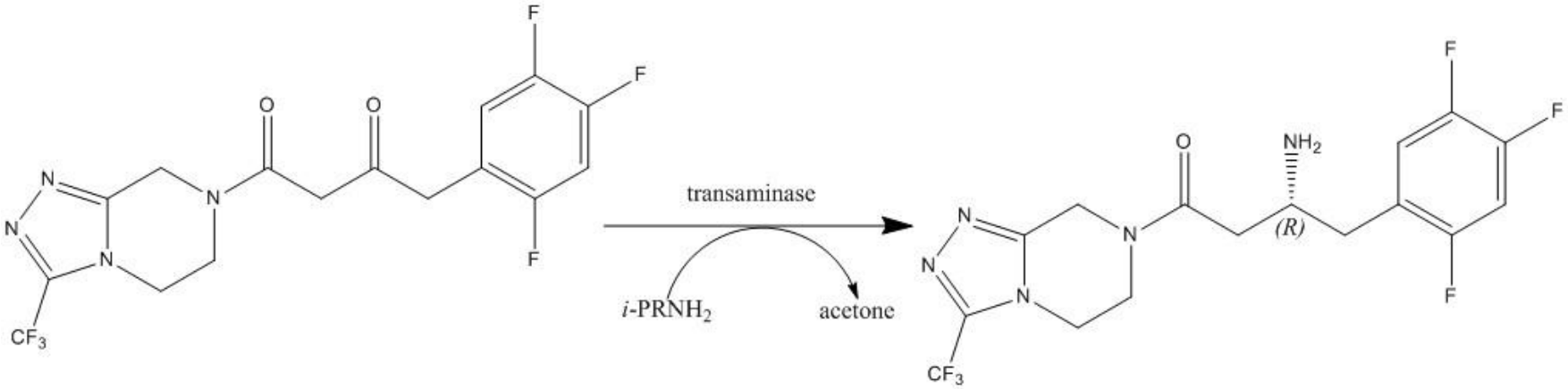


Figure 1-6. The biocatalytic conversion of prositagliptin to enantiopure sitagliptin by an engineered transaminase. The product is then further phosphorylated to the antidiabetic compound, phosphositagliptin as described by Savile, 2010. *i*-$PRNH_2$: isopropylamine

The traditional method of manufacture involves the use of high pressures and a rhodium based chiral catalyst which remains as a trace in the final product (Savile et al. 2010). The use of a stereoselective enzyme eliminates the need for the high pressures and contaminating catalysts and also provides higher enantiopurity of the final product. Other examples of ω-transaminases in industry are in the production of amino acids (Breuer et al. 2004) and 2-aminobutyrate which is used in the production of levetiracetam, the active pharmaceutical ingredient (API) in Keppra™, a therapy used in treatment for epilepsy (Shin & Kim 2009).

1.3 Synthetic biology and alternative host organisms

It is not only the metabolic pathways that are important in creating the desired products; the host organism must also be robust enough to support the genetic manipulation we are imposing. *E. coli* is the current workhorse in bioscience research but we are ever progressing towards a situation where *E. coli* is insufficient to answer the questions that we have left to investigate. Therefore as synthetic biology is developed and becomes increasingly complex; we must also develop a range of host organisms to support the complex pathways that we have built.

E. coli is currently the most popular host as there is an abundance of molecular tools available for this organism. There is a lot of genomic, metabolic and growth information available for *E. coli* as well as a wide range of strains that have been optimised for specific functions and are commercially available such as Rosetta by Novagen which is designed to promote proper folding of eukaryotic proteins and Origami B which improves disulphide bond formation (Baeshen et al. 2015). *E. coli* grows best at 37 °C and a huge variety of growth media are also commercially available. The ease and abundance of information has resulted in *E. coli* becoming the first organism and often only bacteria tested for recombinant protein expression. However not all proteins express well in *E. coli* and as genomic profiling is ever increasing, new proteins and interesting enzymes are being discovered; a growing list of which are not well expressed or are not active when expressed in *E. coli.* Synthetic biology should be extremely versatile; however the most used expression host, *E. coli* has a specific optimum growth temperature of 37 ºC and pH of 6.5-7. These parameters may not be ideal for the desired reactions.

Therefore investigating alternative host organisms are important in the progression of synthetic biology as genomic and metabolic engineering is becoming more diverse and widespread.

1.3.1 Minimal organisms

A synthetic biology host should have the versatility to accommodate any reaction, enzymatic and chemical, this includes reactions that require extreme conditions such as high or low temperatures and various solvents. Many researchers in this area believe that the ideal synthetic biology host would be a minimal organism; one that only has the genetic information absolutely necessary for protein expression or whichever function is required by the scientist (Beites & Mendes 2015). This reasoning behind a minimal host organism is to reduce side reactions and increase the predictability of the system. Many researchers are working towards the development of minimal chasses such as this (Mansy et al. 2008; Noireaux & Libchaber 2004; Kita et al. 2008; Luisi & Stano 2011); but there is still a huge uncertainty surrounding the reality of minimal host organisms experimentally. The host still needs to be robust and support reactions under a wide range of conditions (Ferber 2004). A minimal chassis could also provide the option of customising the host organism and building a chassis tailored to the reaction. It has also been suggested that a range of minimal chassis should be developed to accommodate a range of functionalities (Luisi et al. 2006). These ideas are yet to be tested experimentally.

1.3.2 Molecular biology tools for alternative hosts

Developing tools for alternative organisms, so that other hosts can be used as easily as *E. coli* is a closer reality for the expression of novel pathways and synthetic biological systems. Broad host vectors for the expression of recombinant proteins in alternative hosts do already exist (Fürste et al. 1986; Troeschel et al. 2012; N Mermod et al. 1986) but the molecular tools

available for genetic and metabolic manipulation of other microbes are extremely limited. In order to increase the development of tools for alternative microbes it is important to demonstrate that it is worthwhile to have a wider range of microbial “workhorses”. Although a minimal chassis may provide more predictability to biological systems, in theory, they will also lack the machinery and components that host organisms currently provide for the reaction. Co-factors, molecules and even host cell proteins that are essential for many intermediate structures during a reaction would not be present in a minimal chassis. The most useful properties of alternative host organisms will be different to that of *E. coli* such as high solvent tolerance to accommodate the production of high value aromatic compounds which would otherwise be toxic to the cell (Kusumawardhani et al. 2018) or optimum growth at different temperatures to match the enzymatic or chemical reaction of interest. One such organism is *P. putida;* certain strains of which have been reported to be solvent tolerant and have an optimum growth temperature of 30 °C (Hill & Robinson 1975; Ni et al. 2013). Molecular biology tools are available for *P. putida* and this organism has been used as a host organism in the past. Much genetic engineering of this organism has been genetic engineering of the host genome (De Lorenzo 1994; de Lorenzo et al. 1998). Promoters have been used to create expression vectors for *P. putida* such as the *tac* promoter which consists of a hybrid of the *trp* and *lac* promoters (de Boer et al. 1983; Bagdasarian et al. 1983) and the *Pm* promoter which originates in *P. putida* from the TOL operon and is associated with the TOL *meta*-pathway involved in the breakdown of aromatic compounds such as toluene (Gawin et al. 2017; Worsey & Williams 1975). More recently, the standard European vector architecture (SEVA) system have been developed (Durante-Rodríguez et al. 2014; Silva-Rocha et al. 2013). These vectors are made up of standardized elements and are broad host range plasmids. Promoters used in these vectors are the *trc* promoter, derived from the *trp* and *lac*UV5 promoters (Brosius et al. 1985), and the *xylS*/*Pm* promoter system which consists of the *Pm* promoter and the *xylS* transcriptional regulator gene

(Gawin et al. 2017). Other work that has been done with the SEVA plasmids is the investigation of the strength of the *Pars1* promoter in *Pseudomonas putida* (Durante-Rodríguez et al. 2014). Other promoters have been investigated such as the *Pben* promoter which originate from *P. putida* and controls the *ortho*-pathway involved in the breakdown of benzoate (Pérez-Pantoja et al. 2014). This promoter has been investigated for its use in the context of metabolic control (Pérez-Pantoja et al. 2014). This is progress towards using *P. putida* as an alternative host organism but to compete with *E. coli* as a host organism, molecular biology tools must be as diverse and as easy to use as they are currently with *E. coli*. As the genome sequence has become available it will be increasingly easier to find and design strong promoter sequences specifically optimised for *P. putida* or endogenous strong promoters (Yong et al. 2014) and other molecular biology elements necessary for recombinant protein expression to rival *E. coli* (Martins Dos Santos et al. 2004; Nelson, Weinel, Paulsen, Dodson, Hilbert, V. a P. Martins dos Santos, et al. 2002).

P. putida has already proven to be an industrially relevant host organism for the production of biosynthetic polymers and high value chemicals, these uses are reviewed well by Rehm 2010 and Schmid et al. 2001 respectively. This usefulness of *P. putida* as an alternative host for biocatalytic processes is mainly due to the tolerance to aromatic compounds which are usually toxic to *E. coli* cells (Kusumawardhani et al. 2018; Poblete-Castro et al. 2012; Schmid et al. 2001). Properties that indicate that *P. putida* may be a useful alternative to *E. coli* in metabolic pathway engineering are that *P. putida* is known to be a native host of several interesting metabolic pathways (Denome et al. 1993; Feist & Hegeman 1969; Jimenez et al. 2002). Various *Pseudomonas* species are involved in processes such as nitrogen and carbon fixing, and the *metab*olism of organic solvents (M. M. Bagdasarian et al. 1981). Industrial uses of *Pseudomonas putida* includes the biotransformation of benzene to polyphenylene monomers; benzene *cis*-

glycol, which was produced in the kilograms using this method (Griffin & Magor 1986). Not only were genetically engineered *P. putida* used as a host organism; but the toluene dioxygenase from the TOL pWWO plasmid was also used in combination with one other enzyme to create a short engineered pathway (Ballard et al. 1983; Jenkins et al. 1987). This demonstrates that *P. putida* interesting in terms of both a novel host organism for the expression of engineered pathways as well as in finding new and interesting enzymes that may be incorporated into an engineered pathway (Poblete-Castro et al. 2012).

Other hosts frequently used in industry are antibiotic producers of the *Streptomyces* species (Baltz 2007). In industry antibiotics are commercially produced and extracted from their natural host organisms. This provides precedence for the commercial growth of *Streptomyces* and product purification from these organisms which could be taken advantage of for the production of non-intrinsic molecules (Chater 2006). There is less precedence for cloning in *Streptomyces* species in comparison with gram-negative bacteria such as *E. coli* and *P. putida* meaning that more tools for genetic manipulation will need to be developed to realise the potential of *Streptomyces* as alternative host organisms in synthetic biology (Medema et al. 2011).

1.4 Automation in the biotechnology industry

With the development of new engineered genetic parts and a wider range of engineered host organisms, challenges around the standardisation and reproducibility of these biological parts will become more apparent (Baker 2016). There are a number of factors that play a role in reproducible research. Two of the most simple to identify and overcome are inconsistencies in user handling and inaccurate recording. Due to these it is often unknown whether discrepancies in data are due to biological variance and therefore incorrect conclusions may be made. Both of these can be

overcome by incorporating new liquid handling technologies into biological experiments. By using automated liquid handling technologies consistency is guaranteed. Automated processes also require the user to provide instructions and therefore the input will always be recorded along with any errors that may have been made. The resolves any issues with inaccurate recording of data. Therefore any results from experiments truly show biological variability. Automation also has another major advantage; the ability to perform high throughput experiments in a time and cost effect manor. With the rise in biological based medicines and therapies such as biosynthesised chemicals as well as cell and gene therapies it will be essential to ensure robust and reproducible manufacturing processes. Manufacturing of advanced therapies such as cell and gene therapies is currently expensive and laborious. To aid more cost effective manufacturing process automation is essential (Pollard & Woodley 2007; Sadowski et al. 2016).

Automation in the biological sciences is being used more and more, however equipment and consumables for the liquid handling platforms are expensive and therefore mainly used by industry. For example liquid handling platforms have been developed by Qiagen to conduct high throughput DNA preparations but for a smaller lab this level of high throughput is often unnecessary. The Tecan and Hamilton platforms are the most commonly cited platforms, likely due to their early entry into the liquid handling space in biological laboratories (Roselle et al. 2017; Mora et al. 2017; Wildey et al. 2017). The limitation of the Tecan platforms is the user interface which requires some programming knowledge to use the platform to its full capacity. More recently affordable liquid handling robots are coming to the market (Gupta et al. 2017), and increasing interest is driving competition which in turn will inevitably drive costs of liquid handlers down. However, biological experiments do not involve only liquid handling.

Biological experiments are inherently more complex than processes which currently have a high level of automation. There are many more variables and often many sensitive components which need to be handled in a specific way, for example in transformation of DNA into bacteria cells. Moving cells between ice to 42 °C and plating out onto agar plates all while maintaining sterility and avoiding cross contamination is not a process which typical liquid handling robots are designed for. Therefore automating biological experiments is a major challenge.

1.5 Hypothesis and research question

Amine molecules are of interest to the pharmaceutical industry because they are often biologically active, however complex, chiral amines are difficult to produce using chemical synthesis methods. In addition, there is a need for alternative host organisms as tools for use in synthetic biology based industrial processes. By using a combination of enzymes from TOL pathway, which result in an aldehyde, and a transaminase, this study investigates the hypothesis that metabolic pathways can be manipulated using synthetic biology based methods to produce an amine that would be difficult to produce using synthetic chemistry based methods. The resulting amine is hypothesised to be 2,6-diaminohex-4-enoic acid, which has is a novel amine and has not been created to the researchers knowledge at the time of this study. This study also investigates the possibility that this engineered pathway is functional in *P. putida* as an alternative host organism to *E. coli*.

1.6 Aims and Objectives

The aims and objectives of the studies described in this thesis are outlined below.

1. To compare *P. putida* with *E. coli* as a potential host organism for use in industrial research and for industrial use of biotcatalysis as a more environmentally friendly method to produce pharmaceutical ingredients by comparing conversion of benzoate to 2-hydroxymuconic semialdehyde using a truncate of the TOL *meta*cleavage pathway.
2. To investigate whether it is possible to create an engineered pathway using the TOL *meta*cleavage pathway and a transaminase which could produce a novel molecule.
3. To develop an automated process for performing molecular biology experiments to reduce the lab time required by the researcher. And to test this by developing a cloning system for easier cloning in *Streptomyces* species in order to complete the following final aim;
4. To investigate a *Streptomyces* species as an alternative industrially relevant host organism and compare with *P. putida* and *E. coli* by comparing conversion of the engineered pathway with the results seen in *E. coli* and *P. putida.*

Chapter 2: Materials and Methods

Chapter 2. **Materials and Methods**

2.1 Cloning

The following cloning methods are general methods used throughout the research. Any deviations from these methods are detailed as necessary in the relevant section.

2.2 Plasmids used and constructed

2.2.1 DNA manipulation

Restriction enzymes and Q5 polymerase polymerase chain reaction (PCR) kits used were obtained from New England Biolabs. Quikchange site directed mutagenesis kits were purchased from Agilent Technologies. DNA primers were obtained from MWG Eurofins genomics and delivered as lyophilised powders purified by high purity salt free technology (HPSF). Primers were diluted with ultra-pure water to a concentration of 100 pmol/µL. PCR was performed as per the manufacturer's instructions, typically 20 ng of template DNA was used per 25 µL reaction in a 250 µL microcentrifuge tube unless otherwise stated. Heat cycling was performed using the Prime Thermal Cycler (Cole-Palmer Ltd.) with heated lid. For each new PCR reaction performed , 8 identical reaction tubes were set up and run over a 10 ° C melting temperature gradient to determine the most appropriate melting temperature. DNA concentrations were measured using the Nanodrop 2000 (ThermoScientific).

In-Fusion cloning was performed in cases where traditional restriction enzyme and ligation based cloning was ineffective. In-Fusion cloning plus kit was purchased from Takara Bio and used as per manufacturer's instructions.

2.2.2 Liquid and Solid media culture

Agar plates were made using LB agar from Merck Millipore dissolved as instructed in deionized water and sterilised by autoclaving at 121 °C for 20 minutes. Antibiotics, when appropriate, were added after sterilisation at the following concentrations: ampicillin; 500 µg/ml, kanamycin; 30 µg/ml. For liquid culture LB broth (Miller) from Merck Millipore was used. The powder was also dissolved as instructed and autoclaved to sterilise as described previously. Appropriate antiobiotics were added after sterilization at the concentrations listed above.

2.2.3 Cloning in *E. coli*

E. coli strains used were NEB 10beta (New England Biolabs) for large plasmids over 10 kb, TOP10 (Life Technologies) for plasmids below 10 kb, BL21 DE3* (Life technologies) for transaminase protein expression (all chemically competent). Heat shock transformations were performed as follows; cells were thawed on ice, once thawed, 10-20 ng of purified plasmid DNA was added and incubated on ice for 30 minutes. The mixture was heat shocked at 42 ° C for 30 seconds and put back on ice for 2 minutes. 1 ml of room temperature LB broth was added and incubated for 1 hour at 37 ° C with shaking before 200 ml was spread onto room temperature LB agar plates and incubated overnight at 37 ° C.

2.2.4 Gel electrophoresis, plasmid purification and PCR cleanup

PCR and cloning validation was determined using agarose gel electrophoresis. Agarose was purchased from Bioline and 1% agarose gels were made using TBE buffer purchased from Life Technologies (cat. no. AM9864) and 0.05% ethidium bromide (Sigma-Aldrich). Gel extraction, PCR and plasmid purification kits were purchased from Qiagen and used as per manufacturers instructions.

2.2.5 DNA sequencing

DNA was sequenced using either Source Bioscience or the Eurofins genomics sequencing service and analysed using SnapGene software.

2.2.6 Transformation in *P. putida*

DNA manipulation for *P. putida* vectors was done in *E. coli* and plasmids were transformed using the electroporation method described in Choi et al. 2006. All *P. putida* strains were grown at 30 ° C. Solid and liquid culture used was used as described in 2.2.2.

2.3 *P. putida* and TOL *meta*cleavage pathway assays

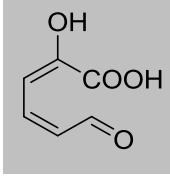
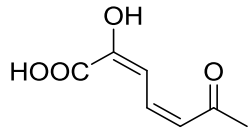
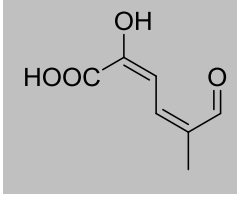
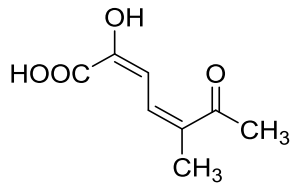
Liquid cultures were incubated with shaking at 200 r.p.m. (1SF-1-W Kuhner) from a starting OD_{600} of 0.2 for 2 hours and induced with 0.1 M IPTG. After one further hour 5 mM substrate was added. The substrates investigated were sodium benzoate, m-toluic acid (3-methylbenzoic acid), p-toluic acid (4-methylbenzoic acid) and 3,4-dimethylbenzoic acid. Growth and absorbance at the wavelength corresponding to the relevant product was measured every hour for a total of 8 hours and then at 24 hours.

2.3.1 Quantification of TOL *meta*cleavage pathway products,

Absorbance of products was measured as specified in Table 2-1 below and the Beer-Lambert law was used to calculate concentration using molar absorption coefficients as described in the literature (Bayly et al. 1966; Kunz & Chapman 1981; Sala-trepat & Evans 1971).

Table 2-1. Absorbance (λ_{max}) and extinction coefficients (ε) used to calculate product concentration

Product	λ_{max} (nm) at pH 7.0	ε (M^{-1})

2-muconic semialdehyde 	375	33 000
2-hydroxy-6-oxohepta-2,4-dienoic acid 	388	13 800
2-hydroxy-5-methyl-6-oxohexa-2,4-dienoic acid 	382	28 100
2-hydroxy-5-methyl-6-oxohepta-2,4-dienoic acid 	390	7 100

2.4 Transaminase assay

The following methods apply specifically to chapter 4 of this thesis.

2.4.1 Initial screening

2.4.1.1 Biotransformation of 2-HMSA

For the production of 2-HMSA to use in the transaminase assays *E. coli* NEB 10beta containing pQR1050 were grown in 50 ml LB broth (Miller, Merck) at

37 °C with shaking at 250 r.p.m. until an OD_{600} of between 0.6 – 0.8 and then inoculated with 0.1 M IPTG and the temperature lowered to 30 °C for 16 hours. Cells were harvested by centrifugation and resuspended in 10 ml 10 mM tris, 5 mM sodium benzoate at pH 8.5. Cells were shaken at 200 rpm at 30 °C for 4 hours. Cells were spun down and supernatant containing 2-HMSA was collected to be used in the transaminase assay or frozen at -80 °C for use at a later date.

2.4.1.2 Transaminase expression

Transaminases were grown in BL21 DE3* at 37 °C, shaking at 250 r.p.m. until reaching an OD_{600} of between 0.6 - 0.8. Cells were induced with 0.1 M IPTG and the temperature lowered to 25 ° C for 16 hours. Cell pellets were resuspended in 25 % well cell pellet w/v 100 mM potassium phosphate buffer pH 8.0, 1 mM PLP. Cells were lysed by sonication for 10 cycles of 10 s on, 10 s off (MSE SoniPrep 150, Sanyo). Insoluble cell components were removed using centrifugation and clarified lysate was frozen at -20 °C for use within 2 weeks.

2.4.1.3 Reaction set up

Transaminase assays were set up in 1 ml reactions containing the following: 30 % v/v transaminase clarified lysate, 5 mM 2-HMSA, 100 mM potassium phosphate pH 8.0, 1 mM PLP, 50 mM amine donor. Reactions were quenched after 1, 3 and 24 hours using 0.2 % trifluoroacetic acid. 5 mM sodium pyruvate was used as the substrate in positive controls. Negative controls were done without the presence of an amine donor.

2.4.2 2-HMSA purification

Two methods were tested for the purification of 2-HMSA from the biotransformation buffer. After 2-HMSA was harvested from the biotransformation as described in 2.4.1.1 attempts to further purify and concentrate the 2-HMSA from this buffer are described as follows.

2.4.2.1 Q-Sepharose Fast Flow resin

Q-Sepharose resin is stored in 20% ethanol, therefore before use this was removed and the resin was equilibrated with 10 mM Tris buffer, pH 8.5. Once equilibrated, 0.5 ml of the resin was added to 0.5 ml of the 2-HMSA solution in a 1.5 ml microcentrifuge tube. The mixture was inverted several times form a homogenous mixture and incubated at room temperature for 15 minutes to allow binding. Concentration of unbound 2-HMSA remaining in the buffer was measured by recording the absorbance at 375 nm. For elution 500 µl of the following buffers were tested: 2 %, 3.5 % and 5 % acetic acid were each investigated in 50 %, 75 % and 100 % methanol. After further 15 minutes incubation at room temperature on a tube roller the concentration of free 2-HMSA in elution buffer was measured.

2.4.2.2 Amberlite IRA-400

Amberlite IRA-400 resin (chloride form, Sigma-Aldrich) was saturated with water as described by Bookser & Zhu 2001. 100 mg of water saturated resin was mixed with 0.5 ml 2-HMSA mixture in a 1.5 ml microcentrifuge tube. This was incubated with shaking in a thermomixer (Eppendorf) at 450 r.p.m. for 16 hours. The resin was then washed with 100 % methanol for 15 minutes. This was repeated twice. Elution of 2-HMSA was tested with the following buffers; 90:10 TFA/water, 90:10 formic acid/water, 95:5 methanol/TFA, 95:5 methanol/formic acid. The concentration of 2-HMSA in the elution buffer was measured.

2.4.3 Analysis

2.4.3.1 HPLC analysis

Samples were centrifuged for 10 minutes at 13 000 rpm to remove any particulates and diluted with water by a factor of 10. Acetophenone production was quantified by HPLC using a C18 column (ACE) and a gradient of 20:80 to 80:20 acetonitrile: water over 20 minutes. Acetophenone was detected at 254 nm. Peaks were also monitored at 210 nm, 375 nm and

245 nm in attempts to identify any new peaks which may correspond to any novel product.

2.4.3.2 Reduction in 2-HMSA

To measure the conversion rates, the reduction in 2-HMSA was measured by taking a 100 µl sample of the reaction mixture which was added to 900 µl of ultra pure water and measuring the absorbance at 375 nm. 2-HMSA has a strong absorption at 375 nm and the concentration is quantifiable using the extinction coefficient as reported in previous literature (Sala-trepat & Evans 1971). Samples were taken at the following times: 0 h, 3 h and 24 h.

2.4.3.3 Derivatisation of 2-HMSA amine products

For the derivatisation of amines the following methods was used, adapted from (Marten & Naguschewski 2011).

50 µl of 1 mg/ml AQC was added to 100 µl of supernatant taken from the transaminase assay or whole cell reaction and 350 µl of 0.2 M borate buffer, pH 8.0 in a 2 ml glass vial. The mixture was left at room temperature for 15 minutes before HPLC analysis.

2.4.3.4 HPLC

After AQC derivatisation, as described in section 2.4.3.3, HPLC analysis was performed using a C18 column (ACE) and a gradient of 20: 80 to 80: 20 140 mM sodium acetate (pH 5.9): acetonitrile respectively over 20 minutes. Peaks were detected at wavelengths of 245 nm, 210 nm and 375 nm.

2.4.3.5 Mass spectrometry

Samples were submitted to the department of Chemistry for mass spectrometry (MS) analysis. Types of MS used were liquid chromatography mass spectrometry (LC-MS) and electrospray ionisation quadrupole time of flight (ESI QTOF MS).

2.4.4 Whole cell conversion

The following methods apply to chapter 5 of this thesis.

2.4.4.1 Testing alternative substrates

P. putida GS1 containing pQR1050 or pQR1062 was grown in 50 ml LB broth at 30 °C, 200 r. p. m. from a starting OD_{600} of 0.2 for 2 hours and induced with 0.1 M IPTG and 1 mM PLP. After one further hour 5 mM substrate was added. Growth and absorbance at the wavelength corresponding to the relevant product (described in Table 2-1) was measured every hour for a total of 8 hours and then at 24 hours.

2.4.4.2 Biotransformation for preparation of samples for analysis

P. putida GS1 containing pQR1050 or pQR1062 were grown in 100 ml LB broth at 30 °C, 200 r.p.m. for 2 hours before being induced with 0.1 mM IPTG. Cells were incubated for a further 16 hours. Cell pellets were resuspended in 10 ml 10 mM Tris, 0.1 mM PLP, 5 mM relevant substrate and incubated at 30 °C, 200 r.p.m. for 24 hours. Cells were pelleted and supernatant was collected for analysis.

Chapter 3: TOL pathway expression in *Pseudomonas putida* and a comparison with *E. coli*

Chapter 3. **TOL pathway expression in *P. putida* and a comparison with *E. coli***

3.1 Background and aims

The TOL *meta*cleavage pathway was well researched during the 1980s as interest in cleaning up environmental pollutants was high. In this study, the TOL pathway is revisited due to the interesting structures of the molecules that are produced rather than the molecules that can be broken down. Finding enzymes and metabolic pathways that may be useful for biocatalysis is essential for the realisation of greener chemistry and more mainstream use of industrial biocatalysis. With an increase in multi-enzyme cascades in biocatalysis, an increase in range of host organisms to accommodate the enzymes will be necessary. To create more environmentally friendly methods for large scale production of molecules via biocatalysis the idea of cell factories is popular (Purcell et al. 2013; Weeks & Chang 2011; Jeandet et al. 2013). This requires a host that can accommodate optimum conditions for the enzymes they are producing as well as high substrate and product tolerances to ensure maximum and economic output. In this chapter *P. putida* is investigated as a potential host organism for the expression of a truncate of the TOL *meta*cleavage pathway (Figure 3-1).

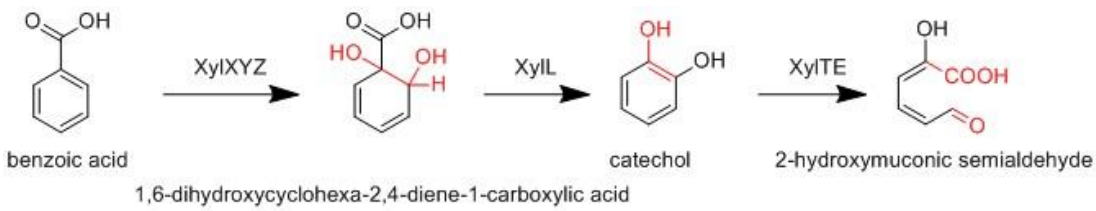


Figure 3-1. The truncate of the TOL *meta*cleavage pathway containing enzymes XylXYZLTE that are investigated throughout this study.
This part of the pathway is responsible for the conversion of benzoate to 2-hydroxymuconic semialdehyde. XylXYZ is a multi subunit enzyme and *xylT* encodes for a ferredoxin which aids in the activity of XylE (Harayama et al. 1991). Changing subgroups are shown in red.

Chapter 3 TOL pathway expression in *P. putida* and a comparison with *E. coli*

P. putida is an interesting alternative to *E. coli* for a number of reasons. For example *P. putida* fast growing with an optimum temperature of 30 °C and non-pathogenic and therefore has the same advantages as *E. coli* in those respects. There are some tools available for cloning and manipulation as *P. putida* is also a gram negative bacteria. There are many strains, of which genomes have been sequenced and have been used in research therefore some ground work is already done. One of those most interesting properties though is the tolerance for solvents, which is particularly high in certain strains (Simon et al. 2015; Mi et al. 2014; Ramos-Gonzalez et al. 2003). Many strains achieve this high solvent tolerance either by breaking down the solvent or by a mechanism eliminating the solvent from the cell. Other research groups have also identified the *P. putida* as a potential industrial research organism (Nikel et al. 2016; Poblete-Castro et al. 2016). In this study a range of *P. putida* strains have been transformed with a plasmid containing a short truncate of the TOL *meta*cleavage pathway. The differences in pathway activity in a whole cell bioconversion were monitored and compared between different strains of *P. putida* and also with *E. coli*.

3.2 Cloning and genetic tools for *P. putida*

Currently there is not a vast range of genetic tools for *P. putida*. There are broad host range plasmids that replicate across a number of gram negative bacteria which can be used in *P. putida*. As it is becoming more recognised that *P. putida* is an industrially relevant organism, more tools are being tested and developed for genetic engineering of this bacteria. Here the pMMB67EH plasmid has been used for expression of a truncate of the TOL *meta*cleavage pathway in both *P. putida* strains and *E. coli.* The pMMB67EH plasmid is a broad host range plasmid with ampicillin resistance and an IPTG controllable *tac* promoter as shown in Figure 3-2.

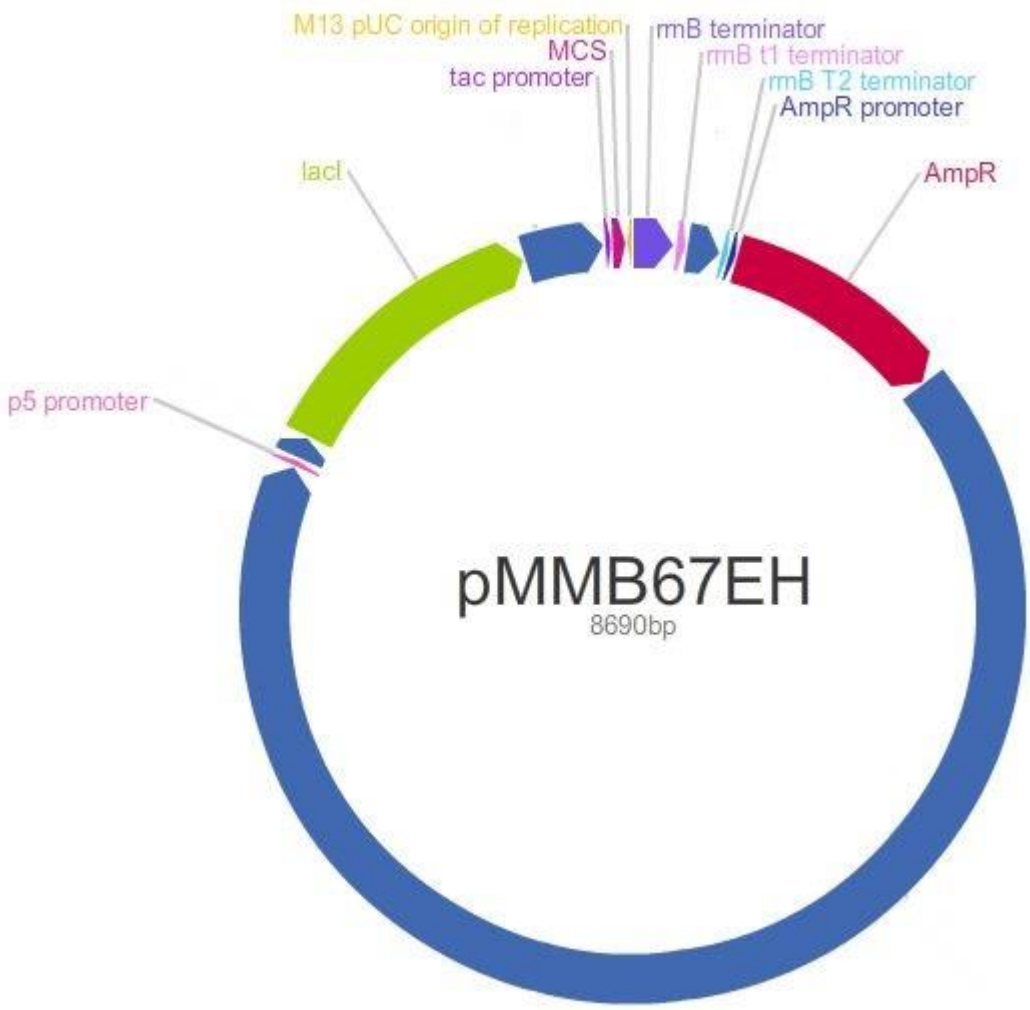


Figure 3-2. The pMMB67EH plasmid.
This plasmid was used as a broad host range plasmid used in the cloning of pQR1050. pMMB67EH was chosen because it can be transformed and expressed in both *E. coli* and *P. putida.*

3.2.1 Cloning the TOL pathway fragment from pQR226 into pMMB67EH

Traditional restriction digestion and ligation techniques were used to clone the TOL fragment from pQR226 (Figure 3-3) into pMMB67EH (methods described in section 2.1). The TOL fragment is approximately 5.5 kbps long and was originally cloned using *Xba*I and *Hin*dIII restriction enzymes to create the pQR226 plasmid used in this study (Jackson 1996). The same enzymes were used to extract the fragment and ligate it into the broad host

vector pMMB67EH as shown in Figure 3-4. Cloning success was confirmed by digesting the DNA isolated from the transformed colony with *Xba*I and *Hin*dIII restriction enzymes and by DNA sequencing as shown in appendix B.

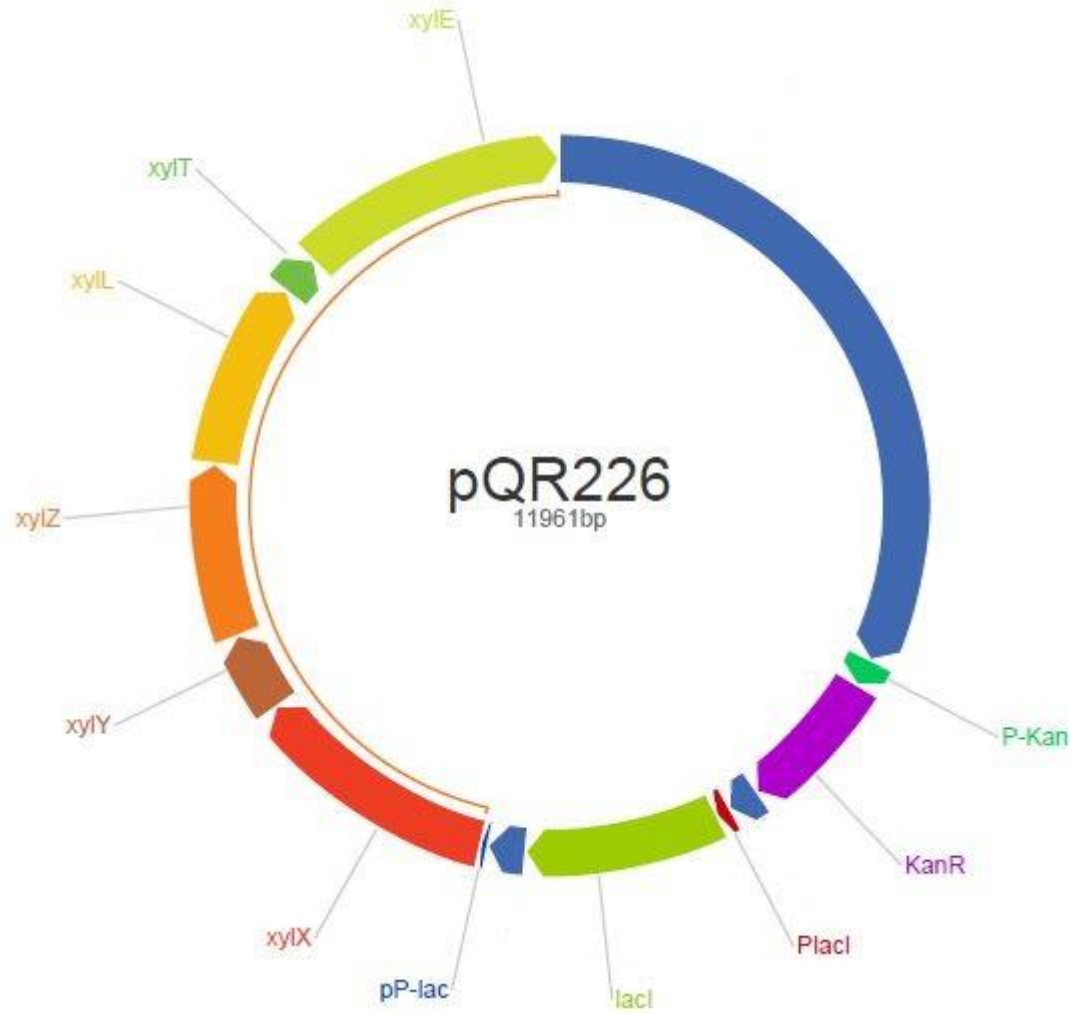


Figure 3-3. The pQR226 plasmid which contains the *xylXYZLTE* genes under lac controlled expression.
pQR226 also contains genes necessary for kanamycin resistance.

3.3 *P. putida* strains

A variety of *P. putida* strains from the Ward lab collection were identified and selected for investigation. Strains were grown on solid LB agar as described in the methods section 2.3. *P. putida* strains were made electrocompetent

and transformed with the pQR1050 plasmid using the method as described in Choi et al. 2006. Not all strains were successfully transformed which may be due to a failure in the method used to make cells electrocompetent due to variations in strain characteristics or due to the large size of the pQR1050 plasmid. *P. putida* strains were transformed with pQR1050 plasmid in order to compare the activity of the TOL pathway enzymes in various strains. Initially *P. putida* KT2440 was used as it is the most documented in the literature and has been investigated as a host organism by other groups (Martínez-García et al. 2014). *P. putida* KT2440 is a derivative of *P. putida* mt-2 and has been widely investigated (Simon et al. 2015; Dammeyer et al. 2013; Durante-Rodríguez et al. 2014; Nelson, Weinel, Paulsen, Dodson, Hilbert, V. a P. Martins dos Santos, et al. 2002; Nelson, Weinel, Paulsen, Dodson, Hilbert, V. A. P. Martins dos Santos, et al. 2002; Nikel et al. 2015; Nikel & de Lorenzo 2014). One of the major advantages of this strain is that it is recognised as GRAS (generally regarded as safe). Once methods had been established with strain KT2440, other *P. putida* strains tested were selected from the Ward lab collection of microorganisms at The Advanced Centre for Biochemical Engineering at University College London. Strains that contained the TOL plasmid or similar genes were avoided so that observed conversion was due to the transformed plasmid only. One strain that was not found in the Ward lab collection was *P. putida* GS1 (DSM 12264). This strain was of particular interest due to reports of high solvent tolerance in the literature (Mi et al. 2014) and was purchased from the DSMZ collection.

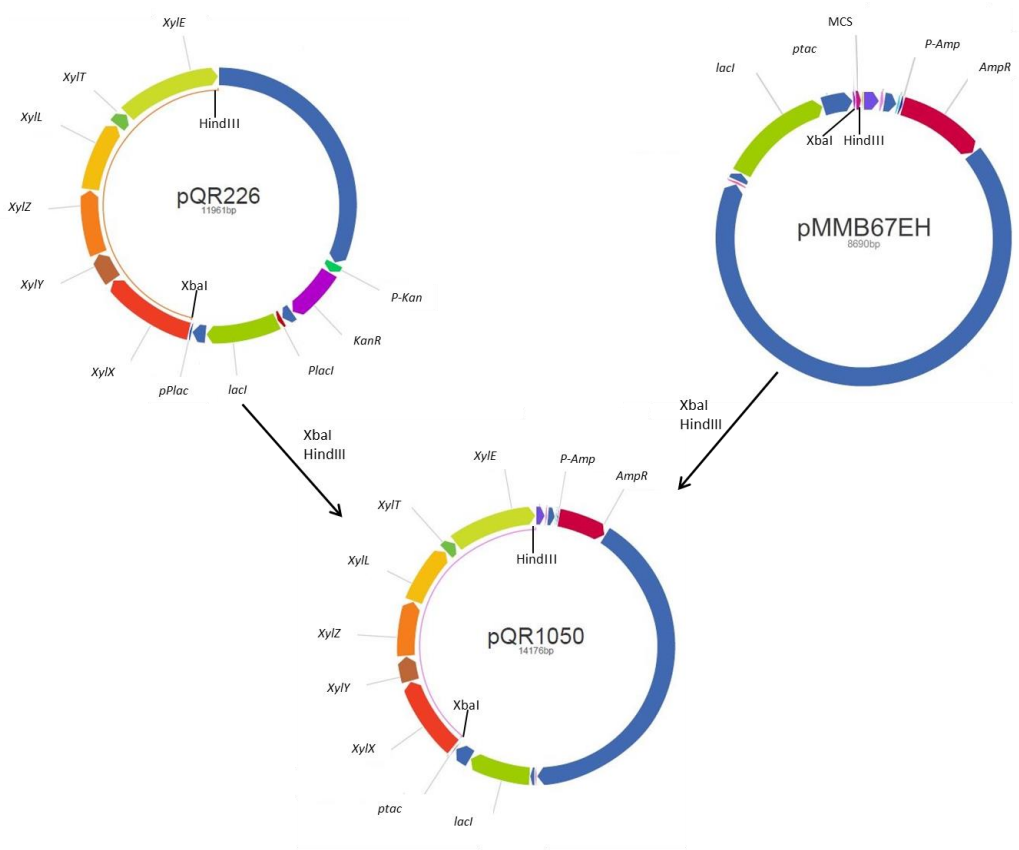


Figure 3-4. Construction of the pQR1050 plasmid.
The 5.5 kb *Xba*I-*Hin*dIII fragment containing the *xylXYZLTE* genes were isolated by restriction enzyme digestion and gel purification before ligation into digested pMMB67EH.

Table 3-1 List of strains that were successfully transformed with pQR1050 and the origin of the strain

***P. putida* strain**	**Source**
KT2440	(M. Bagdasarian et al. 1981)
AC10	(Friello et al. 1976)
PpG277	(Rheinwald et al. 1973; Harder & Kunz 1986)
PpD3	Ward lab collection
PpL3	(Wong & Dunn 1974)
GS1	(Speelmans et al. 1998)
Paw1	(Williams & Murray 1974)

Table 3-2 *P. putida* strains from collection that did not transform with pQR1050

***Pseudomonas* strain**	**Source**
PpS3	(Senior et al. 1976)
PpM3	Ward lab collection
PpK1	Ward lab collection
PpTMC	Ward lab collection
AC105	Ward lab collection
WA1 AjS10 G9	Ward lab collection
PpP1 (PHD)	(Shewan et al. 1960)
NCIMB 10007	Ward lab collection
P. indigofera	(McFadden & Howes 1961)
MT103 PpG1400 Na1	Ward lab collection
MT303	Ward lab collection
AC34	(Gunsalus et al. 1975)
PpN1	Ward lab collection
PpV1	Ward lab collection
AJ3 C1 Bgw	Ward lab collection
AJ2 C1 BgS	Ward lab collection
A6	Ward lab collection
P. acidovorans 9681	(Stanier et al. 1966)
PAO1	(Holloway 1955)
Paw339 PIM2007	Ward lab collection

3.4 Substrate tolerance

The production of 2-hydroxymuconic semialdehyde (2-HMSA) was tested with a range of concentrations of sodium benzoate (2.5 to 20 mM) to determine the optimum concentration for maximum production of 2-HMSA.

Cells were grown in shake flasks in 50 ml of LB broth and gene expression was induced with IPTG after 2 hours. After a further 1 hour the investigative concentration of sodium benzoate was added. Concentration of 2-HMSA and OD_{600} was measured every hour for a further 5 hours. An average of three samples from each of three biological replicates was taken in order to ensure consistent and reliable results. Detailed methods are described in section 2.3. The cell growth and production of 2-HMSA followed similar trends, where higher concentrations of substrate (sodium benzoate, 15 mM and 20 mM) resulted in slower growth and poor conversion to 2-HMSA. Lower concentrations of sodium benzoate (2.5 mM and 5 mM) resulted in roughly double the yield of 2-HMSA compared to the yields from cells exposed to the higher concentrations of sodium benzoate (Figure 3-6). Before induction after 2 hours of growth and addition of substrate at 3 hours growth the growth rates between samples are very similar. It is only after sodium benzoate addition, and therefore the conversion of benzoate to 2-HMSA that the growth rates begin to differ; this indicates that it is the addition of the substrate or an intermediate product in the truncated TOL pathway that is the cause of the reduced growth rate (Figure 3-5).

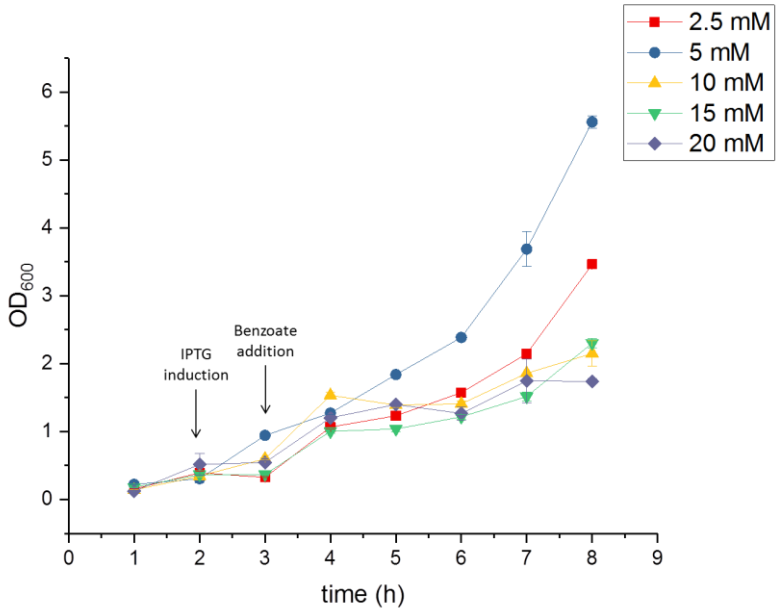


Figure 3-5. Growth of *P. putida* KT2440 containing pQR1050. Expression was induced with IPTG after 2 hours. 2.5 mM, 5 mM, 10 mM, 15 mM or 20 mM sodium benzoate was added after 3 hours.

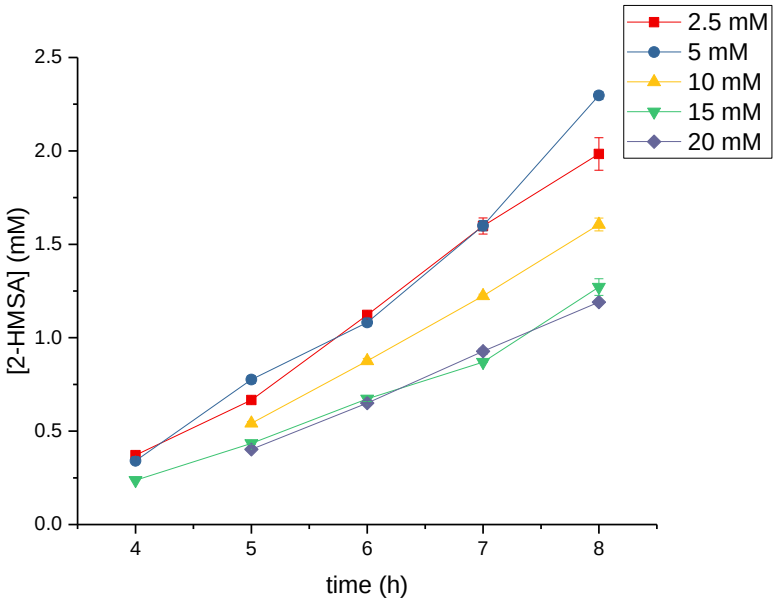


Figure 3-6. Production of 2-HMSA in *P. putida* KT2440 containing pQR1050 after addition of sodium benzoate at various concentrations ranging from 2.5 mM to 20 mM. Errors bars indicate variation arising from three biological replicates per benzoate concentration. Expression was induced with IPTG after 2 hours and sodium benzoate was added after 3 hours.

A one-way ANOVA significance test with a 95% confidence level was performed for the 8 hour time point to determine significant differences between concentrations of 2-HMSA produced. Significance was observed and a post-hoc Tukey test was performed to show relationships between particular groups (Figure 3-7).

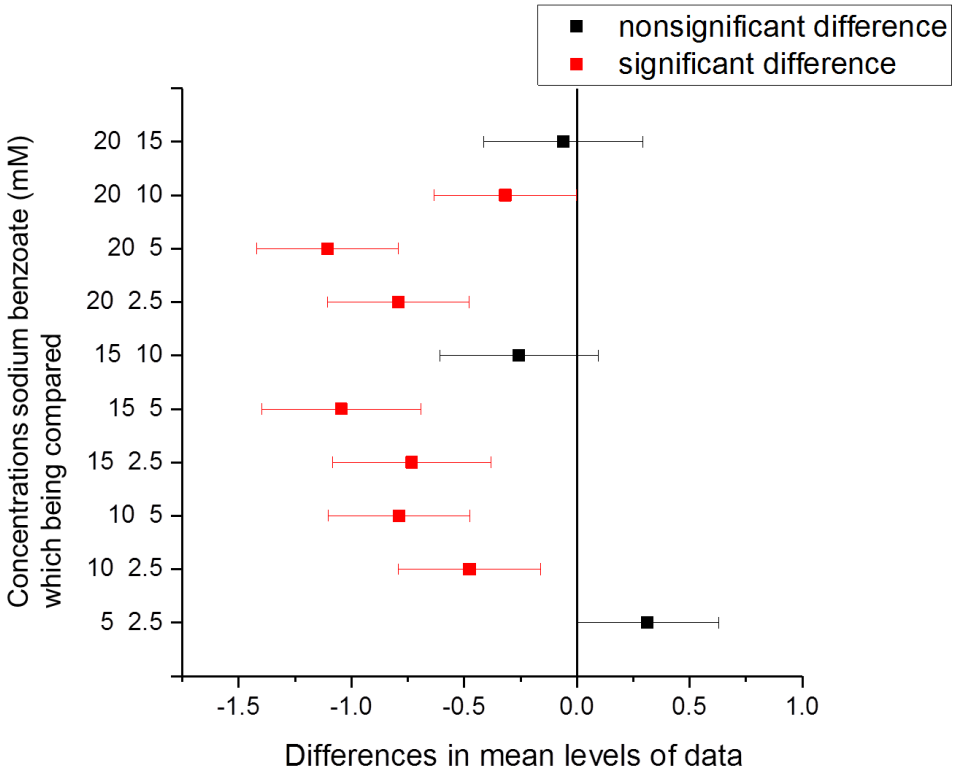


Figure 3-7. Post-hoc analysis showing the relationships between the different substrate concentrations after 8 hours.

Post-hoc analysis shows that at the 8 hours the concentration of 2-HMSA with 10 mM substrate is significantly different to with 2.5 mM and 5 mM, that 15 mM substrate is also significantly different to 2.5 mM and 5 mM and that 20 mM substrate is significantly different to 2.5 mM, 5 mM and 10 mM substrate. The largest differences are seen between 20 mM and 5 mM and between 15 mM and 5 mM. With this information, all following experiments were performed with 5 mM substrate.

3.5 Growth and expression of the TOL pathway in *P. putida*

Several strains of *P. putida* were transformed with pQR1050 and cultured on solid LB agar and in liquid LB broth. For the following experiments to compare the conversion of sodium benzoate to 2-HMSA, expression was induced with IPTG after 2 hours of growth at 30 ° C and sodium benzoate was added after 3 hours (as described in section 2.3). The cell growth and conversion of 5 mM sodium benzoate to 2-HMSA were measured using absorbance at 600 nm and 375 nm respectively. Measurements were taken in triplicate every hour until 5 hours post induction and at 24 hours. In addition to the technical triplicates taken, three biological replicates were performed for each strain and an average was taken. Negative controls were also performed, where absorbance at 375 nm was measured for samples of each strain with IPTG induction but without the addition of the benzoate substrate. The absorbance measurements from the negative control bacteria growth were also taken in triplicate and the average of these was used as a baseline absorbance for each strain. As a result the following figures (Figure 3-8 to Figure 3-15) represent actual conversion rates where baseline absorbance has been subtracted from the average absorbance measurements for each time point. Absorbance measurements of the conversion in *P. putida* strains were compared with *E. coli* DH10β as shown in Figure 3-8 to Figure 3-15. Growth of *E. coli, P. putida* KT2440 and *P. putida* AC10 varied the most between biological replicates which is shown by large error bars in Figure 3-8, Figure 3-9 and Figure 3-10. Large variation in these strains could result from an inherent biological variation in growth. This is unlikely for *E. coli* DH10β due to the commercial development of this strain which will have ensured that this strain produces reliable results (NEB 2019).

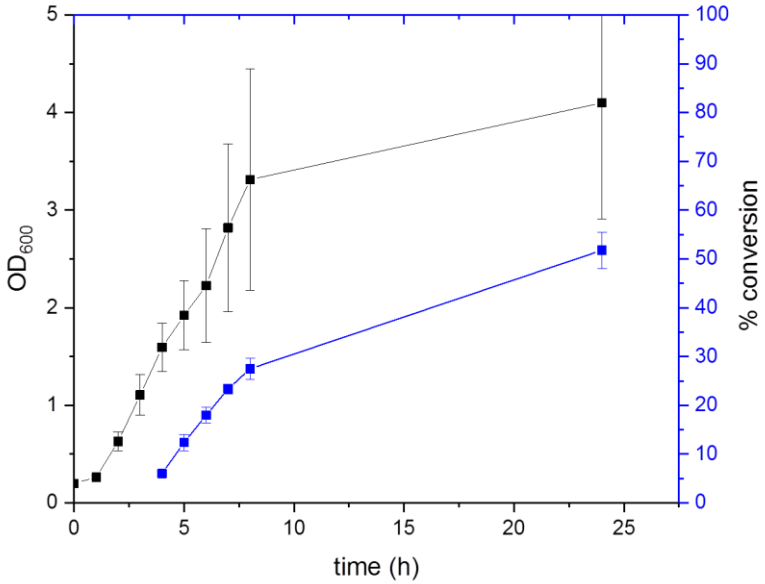


Figure 3-8. Growth and conversion of benzoate to 2-HMSA in *E. coli* DH10β expressing pQR1050.
Expression was induced with IPTG at 2 hours. 5 mM sodium benzoate was added at 3 hours. Errors bars represent variation calculated from biological triplicates.

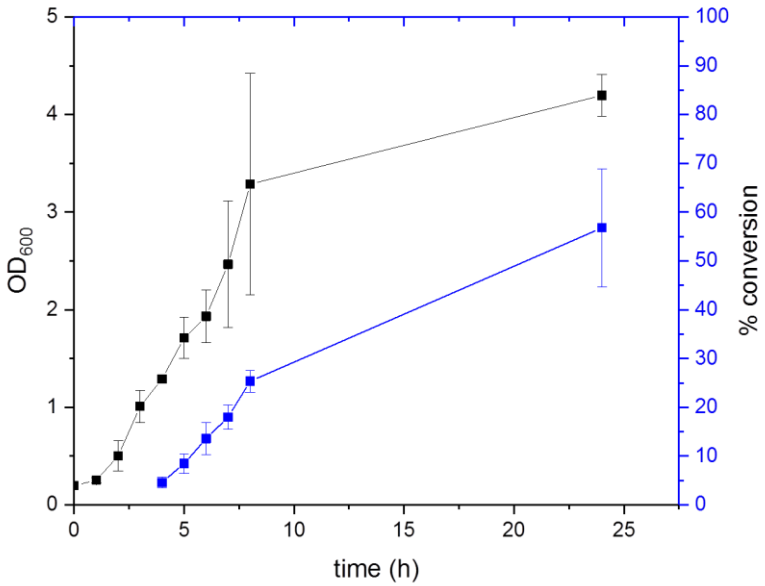


Figure 3-9. Growth and conversion of sodium benzoate to 2-HMSA in *P. putida* KT2440 expressing pQR1050.
Expression was induced with IPTG at 2 hours. 5 mM sodium benzoate was added at 3 hours. Errors bars represent variation calculated from biological triplicates.

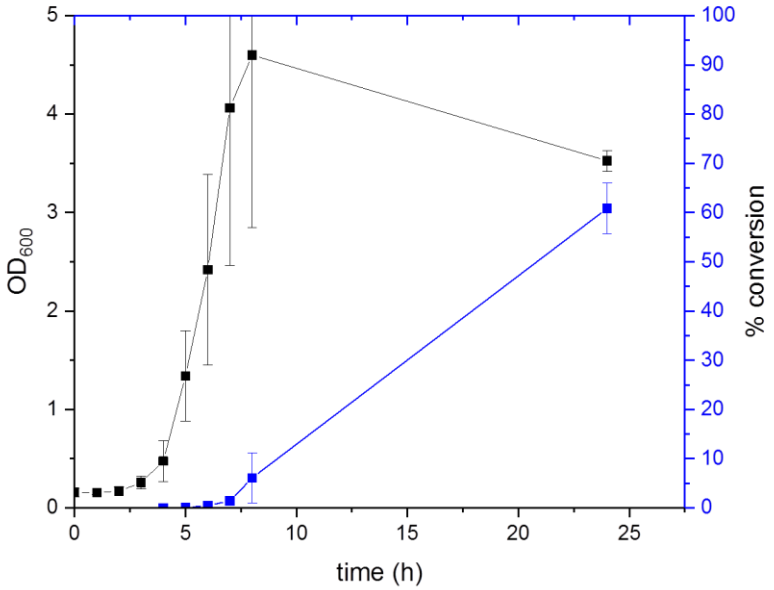


Figure 3-10. Growth and conversion of sodium benzoate to 2-HMSA in *P. putida* AC10 expressing pQR1050.
Expression was induced with IPTG at 2 hours. 5 mM sodium benzoate was added at 3 hours. Errors bars represent variation calculated from biological triplicates.

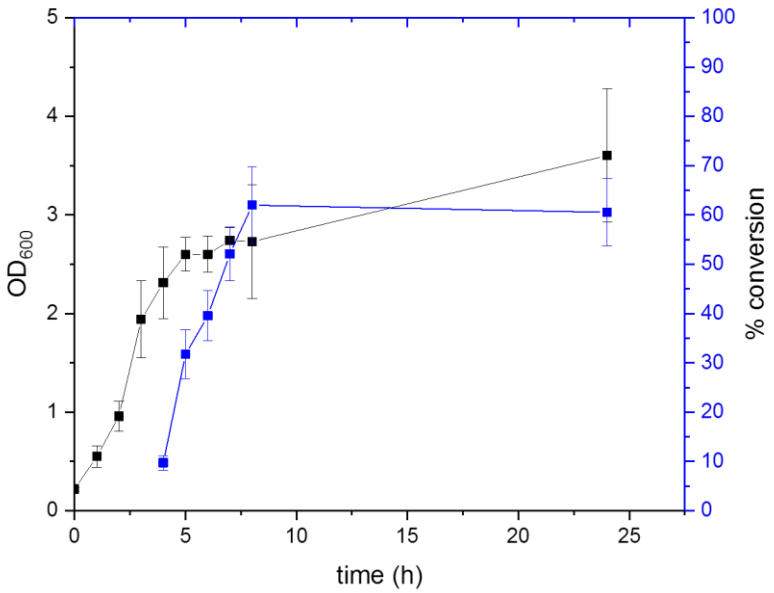


Figure 3-11. Growth and conversion of sodium benzoate to 2-HMSA in *P. putida* GS1 expressing pQR1050.
Expression was induced with IPTG at 2 hours. 5 mM sodium benzoate was added at 3 hours. Errors bars represent variation calculated from biological triplicates.

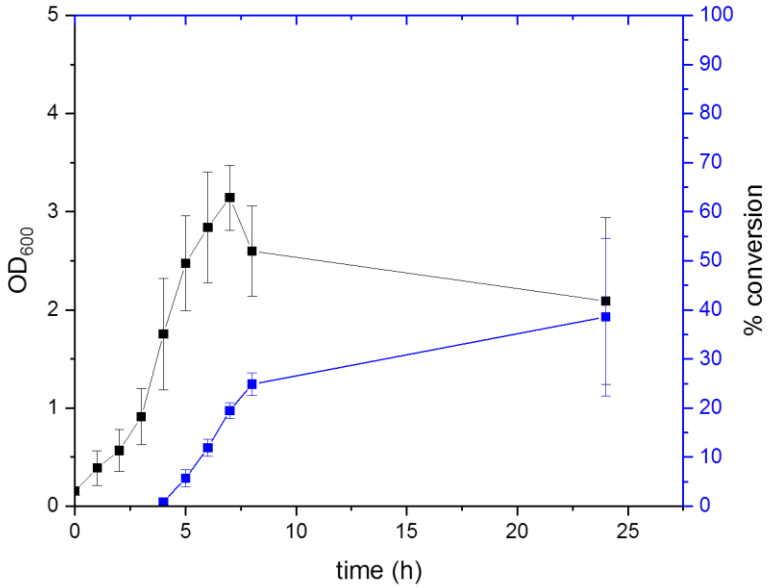


Figure 3-12. Growth and conversion of sodium benzoate to 2-HMSA in *P. putida* G277 expressing pQR1050.
Expression was induced with IPTG at 2 hours. 5 mM sodium benzoate was added at 3 hours. Errors bars represent variation calculated from biological triplicates.

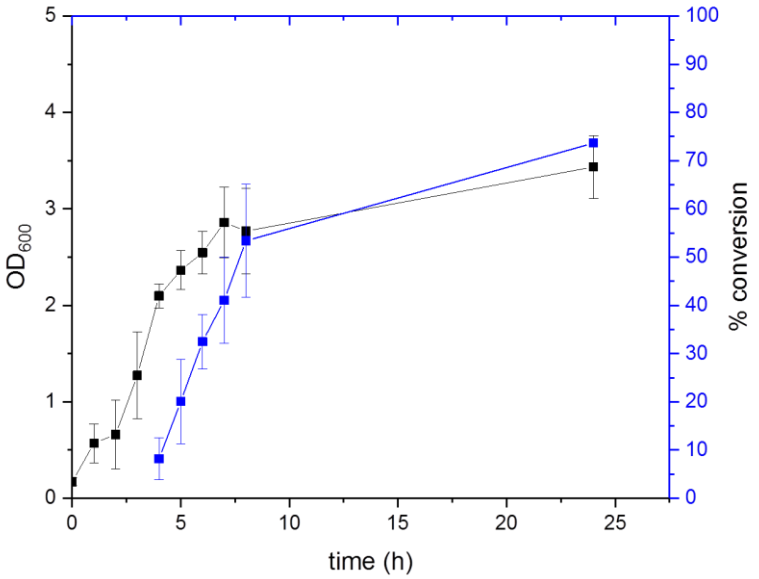


Figure 3-13. Growth and conversion of sodium benzoate to 2-HMSA in *P. putida* ppL3 expressing pQR1050.
Expression was induced with IPTG at 2 hours. 5 mM sodium benzoate was added at 3 hours. Errors bars represent variation calculated from biological triplicates.

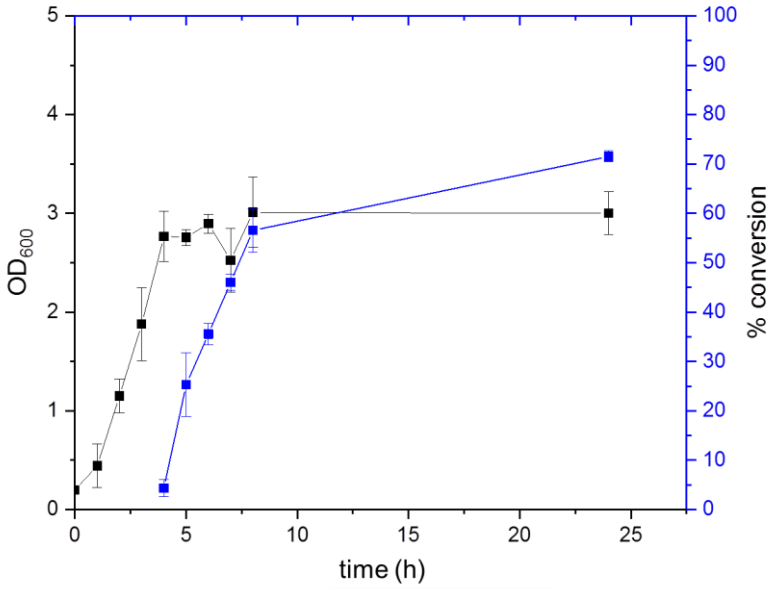


Figure 3-14. Growth and conversion of sodium benzoate to 2-HMSA in *P. putida* ppD3 expressing pQR1050.
Expression was induced with IPTG at 2 hours. 5 mM sodium benzoate was added at 3 hours. Errors bars represent variation calculated from biological triplicates.

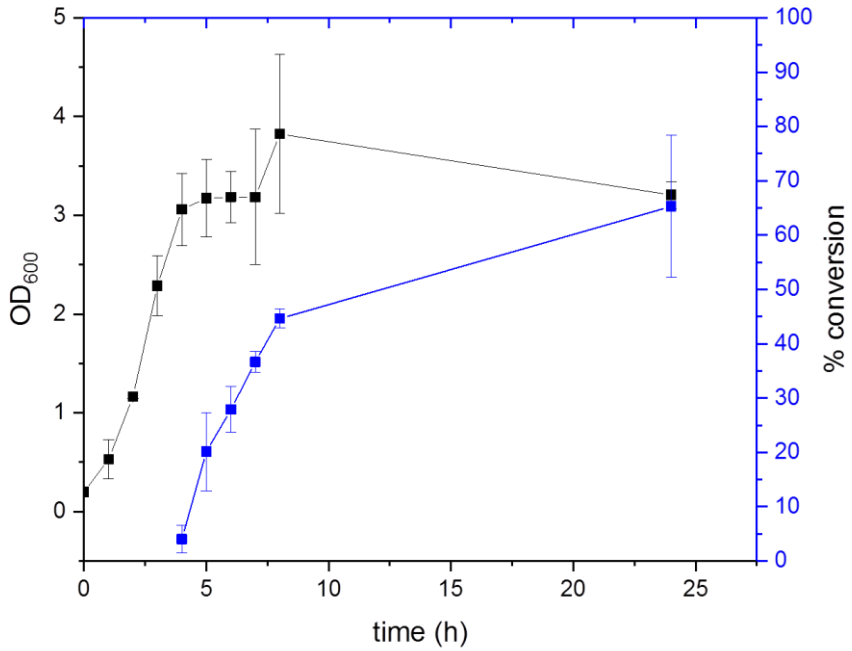


Figure 3-15. Growth andconversion of benzoate to -2HMSA in *P. putida* Paw1 expressing pQR1050.
Expression was induced with IPTG at 2 hours. 5 mM sodium benzoate was added at 3 hours. Errors bars represent variation calculated from biological triplicates.

Table 3-3. Doubling times of *E. coli* and *P. putida* strains investigated in this chapter as observed during the whole cell conversion of sodium benzoate to 2-HMSA

Strain	Doubling time (h)
E. coli DH10β	2.1
P. putida KT2440	2.2
P. putida AC10	2.1
P. putida GS1	1.3
P. putida G277	2.4
P. putida ppL3	2.8
P. putida ppD3	3.0
P. putida Paw1	2.3

A one way ANOVA test to show any differences between concentration of 2-HMSA after 24 hours showed no significant difference between any of the strains including *E. coli.* This shows that *P. putida* is directly comparable to *E. coli* as an organism for expression and activity of recombinant proteins and enzymes. Though at 24 hours there is no significant difference in 2-HMSA concentration between strains, certain strains, such as AC10 (Figure 3-10) have slower rates of production during the initial 8 hours. For AC10 it is clear that this could be explained by the long lag phase that was observed. The optical density is comparable to other strains after 24 hours and the product concentration is also similar, suggesting a similar level of activity. *P. putida* GS1 showed a decrease in product concentration from 8 to 24 hours. This may be due to degradation of the product as observed in the literature (O' Sullivan 2000). Previous literature has reported high solvent tolerance and industrially viable conversion rates observed in *P. putida* GS1 which aligns with the overall aims of this study, therefore this strain was selected for following experiments (Mars et al. 2001).

All strains also show growth to similar optical densities after 8 hours and after overnight growth and similar growth rates (Table 3-3). All replicates were biological replicates and therefore there is variability in the growth. There is less variability in 2-HMSA production for *E. coli, P. putida* KT2440 and *P. putida* G277 (Figure 3-8, Figure 3-9, Figure 3-12 respectively). Other strains

show similar variability in 2-HMSA as with the growth which would be expected. This indicates that the growth variability does not always have a large effect on 2-HMSA production at this small scale volume. The production of 2-HMSA does however appear to follow the same trend as the growth curves, increasing as growth increases.

3.6 Discussion

3.6.1 Many *Pseudomonas* strains did not transform

There are a large number of strains in the Ward lab strain collection with which transformation of plasmid pQR1050 was unsuccessful. This is likely due to the method used to make the cells competent. It was found that concentrations of ampicillin needed to be increased to 1 mg/ml in selective agar due to natural resistance of many *P. putida* strains to the antibiotic. Even at this concentration strains such as PpS3, PpM3, Paw339 PIM2007 and MT103 grew on selective agar providing false positives. With more time, other methods to produce competent cells would be tested with unsuccessful strains. In addition, a construct with kanamycin or other selective gene could be used to overcome the issues encountered with natural resistance to ampicillin. More details of future work can be found in Chapter 7. The strains that do show increased intrinsic resistance may be of interest in industrial processes due to their robust nature so it may be of value to investigate this further in another research project. After further investigation into genotype of the strains, it was discovered that *P. putida* Paw1 did in fact contain the pWWO plasmid encoding for the TOL pathway genes. Therefore the data presented here is not an accurate representation of the conversion as the strain has intrinsic activity. This is interesting in itself as in comparison with other strains, such as D3, L3 and GS1, the Paw1 strain did not show the most efficient conversion. This may be due to the tac promoter used in

pQR1050 which is induced by IPTG and allows for overexpression of the enzymes. In the Paw1 strain we cannot guarantee overexpression.

3.6.2 High substrate concentration reduces conversion

There are several reasons why an increase in substrate concentration could be limiting production of 2-HMSA. One may be that a build up of catechol, which is toxic to cells, slows growth or promotes cell death. Previous studies show that the flux through the TOL pathway is not equal for each step. The catechol 2,3-dioxygenase has slower turnover rates and therefore there is a build up of catechol in the system (Sheridan et al. 1998). This reduces the overall enzyme production and in turn the conversion to 2-HMSA. In addition to this it has also been reported that the inactivation by oxidation of catechol 2,3-dioxygenase is also increased by higher concentration of substrate (Sheridan et al. 1998). However the growth for for 2.5 mM substrate is similar to that of the highest levels and if the catechol toxicity was the limiting factor you would expect an increase in cell growth for 2.5 mM compared with 5 mM which is not what has been observed here. This also does not explain why these is higher conversion between 0-5 hours with 2.5 mM and 5 mM compared with higher substrate concentrations but the growth rates are similar. Another possible reason is that there is substrate inhibition occurring. It is possible that at high concentrations of substrate, the sodium benzoate could be interacting with the active site of downstream enzymes but is not being converted and therefore inhibiting the enzymes. The enzymes have some promiscuity and can accept a range of similar substrates (Murray et al. 1972; Abril et al. 1989), however substrates interacting with downstream enzymes in this pathway has not been reported previously. From an evolutionary perspective it is unlikely that the downstream enzymes of a pathway would accept substrates of the upstream enzymes as this would disrupt the systematic degradation of the initial substrate. In the case of the TOL pathway this would be toxic aromatic compounds and therefore it would not be beneficial to inhibit degradation in this way.

3.6.3 *P. Putida* is comparable to *E. coli.*

The observation that there is no significant difference between *E. coli* and *P. putida* conversion of benzoate to 2-HMSA adds to existing evidence that *P. putida* could be an alternative host organism for use in industrial research and development. As interest in synthetic biology rises there is a need for a range of hosts to withstand high product and substrate concentrations required for industrial manufacturing processes. *E. coli* is most commonly used in industrial research and development because of the wide range of genetic tools available providing easy manipulation of the organism. In addition to this, *E. coli* is fast growing so results are obtained quickly which suits the fast paced industry environment and works well with time pressures imposed by patents. The growth curves of *P. putida* presented here (Figure 3-8 to Figure 3-15) show that *Pseudomonads* are also fast growing bacteria and similar results can be obtained in the same time frame. There are some genetic tools available and more are being developed, for example the SEVA vectors (Silva-Rocha et al. 2013). This means that genetic manipulation can be equally as simple as in *E. coli* and this is only improving as more evidence for *Pseudomonads* as an industrial host organism is built.

P. putida has an optimum growth temperature of 30 ° C which is lower than *E. coli* which is usually grown at 37 ° C. Temperature is often a major factor when optimising processes in both academia and industry. Often issues with recombinant protein expression and solubility are overcome by lowering the temperature for expression. It is thought that this reduces the rate of protein synthesis and therefore reduces degradation, aggregation thus promoting correct folding of the protein (EMBL 2018). The option of using a host organism that grows best at a lower temperature such as *Pseudomonads* at 30 ° C may be beneficial for expression of recombinant proteins that do not express well in *E. coli*. For this particular case, the enzymes originate from *Pseudomonas* and therefore are most likely to perform best at a lower temperature. This is likely to be the reason why the enzymes in this process

performed well in both *Pseudomonas* strains and in *E. coli*. There is an abundance of research uncovering and investigating enzymes from unusual environments. Elucidating the activities and mechanisms of these will require a host organism for recombinant expression. Having the option to use a host organism closest to the originating organism or that at least grows at a similar temperature will be beneficial to researchers. This will help to maximise the output and reduce the failure rates resulting from poorly expressing proteins in *E. coli*.

The variability in growth between biological replicates is seen in both *E. coli* and in *Pseudomonas* strains. This is something that is notoriously common in biological systems and is one of the reasons for the rise in interest in bioengineering (Sadowski et al. 2016). Therefore it is unsurprising that there is high variability in growth, particularly when working with small volumes of 50 ml. Catechol is toxic to cells and therefore may explaining the varying growth rates. This may also explain why the growth appears to slow and cell density even decreases in particular instances (Figure 3-12, Figure 3-14, Figure 3-15). Other explanations for the variability could be differences in mixing due to the position of the flasks in the shaker, flasks in the centre of the shaker will experience differences in throw compared to flasks positioned at the outer edge of the shaker. Varying humidity within the shaking incubator could also affect cell growth. For some strains the variability in growth does not appear to have an effect on the production of 2-HMSA, this may be an indication of the robustness of *P. putida.*

Chapter 4: Engineering the TOL pathway with an omega-transaminase to produce a novel amine

Chapter 4. **Engineering the TOL pathway with an omega-transaminase to produce a novel amine**

4.1 Background and aims

Engineering metabolic pathways is one way to create novel molecules that would otherwise be difficult to produce using traditional chemical synthesis methods and are also likely to have biological activity. One family of molecules that are often biologically relevant and can be difficult to work with in synthetic chemistry are amines. For this reason the transaminase family of enzymes has gained interest over the past decade (Hwang et al. 2005; Nestl et al. 2014; Leuchtenberger et al. 2005). Transaminases facilitate the transfer of an amine group to an aldehyde or ketone e.g. in keto-acids. The product of the truncated TOL pathway studied in chapter 3 is an aldehyde and can be considered a potential transaminase substrate. Using a transaminase to convert the 2-HMSA would in theory create a novel amine molecule which could have potential as a novel synthon. This novel amine product of the engineered pathway would depend on the position of amination and speculative structures of this product are described in more detail in Table 4-4. 2-HMSA itself would be difficult to synthesise and though it is available to purchase from Aurora Fine Chemicals is extremely expensive at a cost of over $1000 minimum order (in 2017). Therefore to create a novel amine using this aldehyde and a transaminase the aldehyde needs to be produced biologically using the TOL enzymes and followed by the addition of the transaminase. This would create an engineered multi-enzyme cascade resulting in a novel amine molecule. The production of 2-HMSA by the TOL pathway enzymes has been previously investigated so the next step is to explore the conversion of the aldehyde by a transaminase.

4.2 Screening omega transaminases for activity with 2-hydroxymuconic semialdehyde using methylbenzylamine as an amine donor

A selection of ω- transaminases were screened for activity with 2-HMSA as the substrate. Transaminases were selected based on previous observations of activity in the literature and within the group. There is an abundance of literature reporting the wide range of substrates accepted by CV2025 (Villegas-Torres et al. 2015; Richter et al. 2015; Kaulmann et al. 2007). Another transaminase, the *Arthrobacter* mutant, ArRMut11, was designed to be robust and accept larger substrates containing aromatic rings (Savile et al. 2010). This and other transaminases such as the *P. putida*, *Klebsiella pneumonia, Vibrio fluvialis* and *Mycobacterium vanbaalenii* have previously shown activity with cyclic and conjugated substrates of a similar size to 2-HMSA (Richter et al. 2015). A full list of the 19 transaminases which were screened is described in Table 4-1.

Transaminases were expressed in *E. coli* BL21(DE3)*. Cells were lysed and cell debris removed, resulting in the use of clarified lysate containing active transaminase enzyme as described in the methods section 2.4.1. 2-HMSA was produced using a whole cell biotransformation in *E. coli* DH10β containing pQR1050 as described in section 2.4.1. 2-HMSA is excreted from the bacteria cells, which were therefore removed and the remaining buffer containing 2-HMSA was used in transaminase screening assays. The concentration of 2-HMSA was calculated by measuring the absorbance at 375 nm and the extinction coefficient as described in the methods section 2.3.1. This method ensured that the 2-HMSA used as the substrate in these assays contained as little other cell material or other contaminants (e.g. from growth media) as possible. Attempts to purify 2-HMSA are described later in this chapter, section 4.3.2.

Table 4-1. List of plasmids containing transaminases used during this study.

Plasmid	Transaminase gene ID	Gene source	Reference
pQR801	CV2025	*Chromobacterium violaceum*	(Kaulmann et al. 2007)
pQR810	PP_5182	*Pseudomonas putida* KT2440	(Sehl et al. 2012)
pQR811	PP_2799	*Pseudomonas putida* KT2440	(Sehl et al. 2012)
pQR813	PAO221	*Pseudomonas aeruginosa* PAO2	(Sehl et al. 2012)
pQR958	PP_3718	*Pseudomonas putida*	(Villegas-Torres et al. 2015; Richter et al. 2015)
pQR959	PP_2180	*Pseudomonas putida*	Unpublished data
pQR960	BSU09260_1971	*Bacillus subtillis*	Unpublished data
pQR961	BSU09260_402	*Bacillus subtillis*	(Lichman et al. 2015)
pQR977	Dgeo_0713	*Deinococcus geothermalis*	(Villegas-Torres et al. 2015)
pQR978	Dgeo_1177	*Deinococcus geothermalis*	Unpublished data
pQR983	Dgeo_2743	*Deinococcus geothermalis*	Unpublished data
pQR986	BLi00767	*Bacillus licheniformis*	Unpublished data
pQR1003	VF_JS17	*Vibrio fluvialis*	(Shin et al. 2003)
pQR1005	KPN_00255	*Klebsiella pneumonia*	(Villegas-Torres et al. 2015)
pQR1006	KPN_00799	*Klebsiella pneumonia*	(Richter et al. 2015)
pQR1010	KPN_01493	*Klebsiella pneumonia*	(Villegas-Torres et al. 2015)
pQR1011	KPN_03745	*Klebsiella pneumonia*	(Villegas-Torres et al. 2015)
pQR1015	Mlut_10360	*Micrococcus luteus*	Unpublished data
pQR1019	Rsph17025_2835	*Rhodobacter sphaeroides*	(Villegas-Torres et al. 2015)
pQR1021	Rsph17029_3177	*Rhodobacter sphaeroides*	(Villegas-Torres et al. 2015)
pQR1048	Mvan4516	*Mycobacterium vanbaalenii*	(Richter et al. 2015)
pQR1049	ArRMut11	*Arthrobacter sp.*	(Savile et al. 2010; Richter et al. 2015)

Chapter 4 Engineering the TOL pathway with an omega-transaminase to produce a novel amine

Transaminases were initially screened using the method with *(S)*- or *(R)*-methylbenzylamine (MBA) as the amine donor which is converted to acetophenone (Figure 1-5) as described in the methods section 2.4.1. Conversion of MBA occurs at a 1:1 ratio with the aldehyde amine recipient therefore the conversion of MBA can be directly associated with the conversion of the aldehyde substrate. The production of acetophenone was quantified using HPLC analysis and percent conversion was calculated (Figure 4-1). A negative control without 2-HMSA was performed to monitor any baseline conversion of MBA to acetophenone. All conversions were normalised against this baseline of acetophenone. 16 out of the 19 transaminases screened showed no conversion. pQR1006 showed a less than 1% conversion. pQR1049 showed 13% conversion and pQR1048 showed 1.5% conversion. Both the *Arthrobacter* (pQR1049) and *Mycobacterium* (pQR1048) transaminases are *(R)* – selective transaminases so it would appear that *(R)* – transaminases are more accepting of 2-HMSA as a substrate.

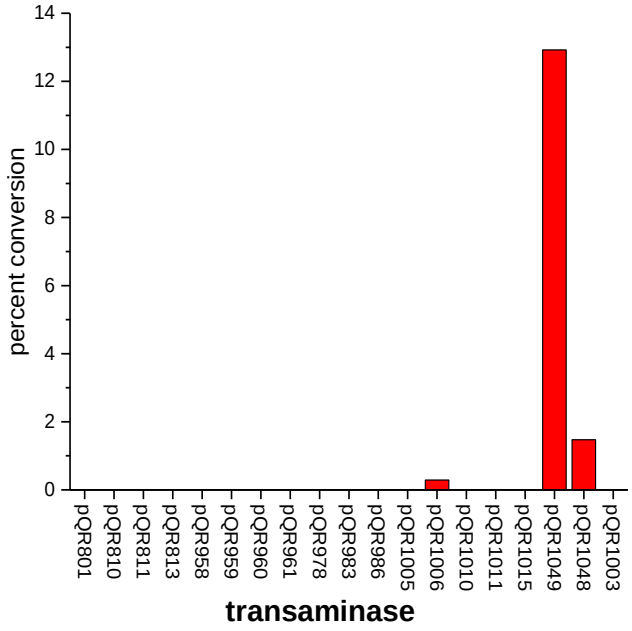


Figure 4-1. Screening of 19 transaminases for activity with 2-HMSA.
***(S)* - MBA as the amine donor with all transaminases except for pQR1049 and pQR1048 for which *(R)* – MBA was used. Conversion was calculated by quantifying the accumulation of acetophenone by HPLC after 24 hours.**

4.2.1 Conversion of 2-HMSA to a novel amine using abundant amino acids as the amine donor

For the use of biotransformation reactions in industry, the reaction must be economically viable; therefore the use of MBA as an amine donor would not be feasible. MBA is a relatively expensive reagent and the cost would not be feasible for a large scale process. One of the aims of this study was also to investigate this engineered pathway in alternative host organisms and therefore develop the process as a whole cell reaction. MBA cannot penetrate the cell wall and therefore would not be suitable for this type of reaction. This led to investigating alternative amine donors with a particular interest in amines that are already abundant within the cell. The most obvious molecules that fit these criteria are amino acids. Abundant amino acids that are continuously produced within the cell are glutamate and

alanine. Serine was also selected as a number of transaminases have been previously successful with serine as the amine donor (Villegas Torres 2014; Deszcz et al. 2015).

To investigate the use of alternative amine donors for the transaminase conversion of 2-HMSA the transaminases with any conversion observed using the MBA/acetophenone assay were investigated alongside additional transaminases expressed in pQR977, pQR1019 and pQR1021 which were identified from previous literature (Villegas Torres 2014). Reactions were set up using 2-HMSA from a whole cell biotransformation in *E. coli* DH10β, active transaminase in the form of clarified lysate from *E. coli* BL21(DE3)* expression, and a 10 fold excess of the relevant amino acid was added as described in methods section 2.4.1. All assays were replicated three times and an average calculated.

To quantify the conversion of 2-HMSA in serine, alanine and glutamate assays the absorbance at 375 nm was recorded and the percent conversion calculated as described in section 2.4.3.2. Degradation of 2-HMSA after 4 hours has been recorded (O' Sullivan 2000) so a negative control without transaminase present was performed to monitor any decrease in 2-HMSA concentration not caused by enzymatic conversion. For alanine and glutamate absorbance was measured after 3 hours only due to the high levels of 2-HMSA degradation observed after a longer period of time (Figure 4-3). Absorbance was measured after 3 hours and after 24 hours for serine (Figure 4-2). Detailed methods for this assay are described in 2.4.1.3.

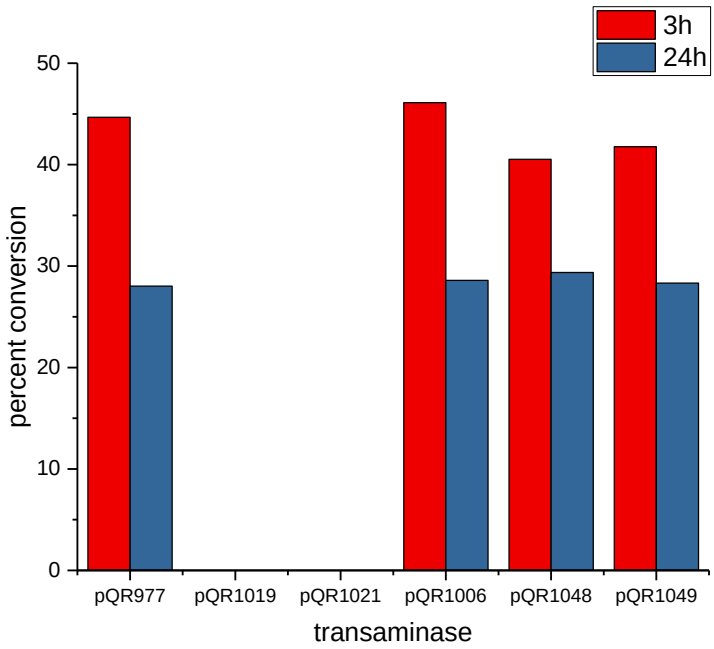


Figure 4-2. Conversion of 2-HMSA after 3 hours and 24 hours with selected transaminases and serine as the amine donor.
Conversions have been corrected using a baseline level of depletion of 2-HMSA concentration seen in negative controls without enzyme present.

From Figure 4-2 conversion after 3 hours is 44% for pQR977, 46% with pQR1006, 40% with pQR1048 and 41% with pQR1049 after correction using the baseline conversion observed in control reactions. After 24 hours there is a much higher depletion of 2-HMSA in the control, this degradation after 4 hours has been observed in previous literature (O' Sullivan 2000) and appears to be a characteristic of 2-HMSA. The apparent conversion rates are as follows; 28% for pQR977, 29% for pQR1006, 30% for pQR1048 and 28% for pQR1049 after adjusting for degradation of 2-HMSA observed in the control sample. The reason for this may be because the transaminase conversion exists as an equilibrium reaction and therefore the conversion has reversed. The reduction in available 2-HMSA due to the degradation as observed in the control reaction may also contribute to a reduced conversion rate. No conversion for pQR1019 or pQR1021 was observed. This may be

due to inactive enzyme or because these transaminases did not accept 2-HMSA as a substrate. Other abundant amino acids tested were glutamate and alanine (Figure 4-3).

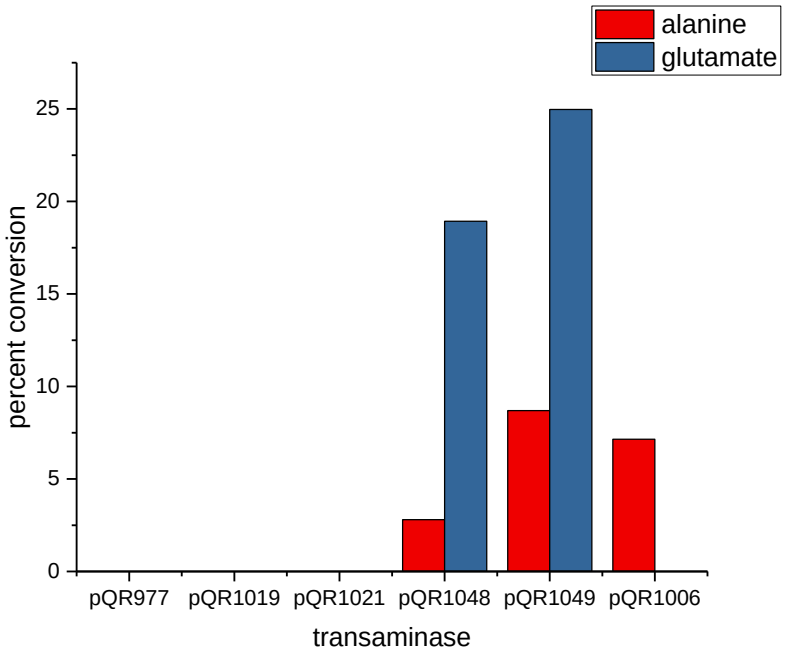


Figure 4-3. Conversion of 2-HMSA after 3 hours with selected transaminases and either alanine or glutamate as the amine donor.
Conversions have been corrected using a baseline level of depletion of 2-HMSA concentration seen in negative controls without enzyme present.

Conversions have been corrected using a baseline level of 2-HMSA depletion to take into account any degradation of 2-HMSA without transaminase as observed in the negative control without transaminase enzyme present. No conversion with pQR977, pQR1019 or pQR1021 was observed. With alanine, 2.8% conversion was observed with pQR1048 and 8.7% with pQR1049. With glutamate conversion rates were better; 19% was observed with pQR1048 and 25% with pQR1049. These assays showed the most consistent conversion with the ArRMut11 transaminase expressed

using the pQR1049 plasmid, therefore this transaminase was used in further reactions to investigate this reaction and the product of the 2-HMSA conversion.

4.3 Structural determination of the novel amine product of the TOL-transaminase engineering pathway

Though evidence of 2-HMSA being consumed were observed, the product formed was unknown. There are several possible molecules that could be formed after amination of 2-HMSA and it is also unknown whether this product may even react further. Further conversion of 2-HMSA with the ArRMut11 transaminase were performed using serine as an amine donor to generate product for further analysis. Analysis of the reaction mixture after quenching was conducted to attempt to determine any new compounds formed.

4.3.1 Initial identification attempts using HPLC and LCMS

Quenched reactions were analysed using HPLC at 0 h, 3 h and 24 h, as described in section 2.4.3.4. Chromatograms were compared between time points to determine appearance of any new peaks formed. Due to the nature of this reaction, including the use of clarified ArRMut11 transaminase lysate expressed in *E. coli* BL21(DE3)* and unpurified 2-HMSA produced during the biotransformation reaction in *E. coli* DH10β there were many peaks present in the HPLC chromatogram and any new peaks were undetectable. In addition to this the amine product of transaminase conversion of 2-HMSA is predicted to be very polar. With an amine and a carboxylic acid, it is likely that the product will elute from a C18 HPLC column immediately and therefore be undetectable by HPLC. LCMS analysis was also performed in an attempt to detect any compounds present corresponding to the mass of the predicted novel amine. Results of the LCMS also showed a high level of

background. Peaks were observed at the expected mass of the product, however due to the high background it was impossible to assume that this was the product. The reaction mixture was not pure enough to use for ESI-TOF MS and therefore presence of the product was not confirmed. To truly determine presence of an amine product and elucidate a possible structure more confidence in mass spectrometry data was required. To obtain this data a more pure reaction mixture or a method for purifying the product from the mixture was required.

4.3.2 Attempts to purify 2-HMSA to reduce background

To reduce background by reducing compounds present in the reaction mixture purification of 2-HMSA was attempted. In theory, 2-HMSA should bind strongly to anionic purification resins. Two methods were trialed; Q Sepharose Fast Flow resin and Amberlite IRA-400 as described in section 2.4.2. The 2-HMSA bound tightly to both resins but elution was difficult. Several elution buffers were tested. Table 4-2 and Table 4-3 provide details concentration of 2-HMSA loaded onto resin compared to the elution concentration for each buffer.

Table 4-2. Concentration of 2-HMSA after elution from Q-Sepharose Fast Flow resin. The concentration of 2-HMSA loaded onto the column was 0.4 mM. Equal volumes for loading and elution were used.

Elution buffer	% recovery of 2-HMSA after elution
2 % acetic acid, 50 % methanol	0.75
2 % acetic acid, 75 % methanol	0.5
2 % acetic acid, 100 % methanol	34.5
3.5 % acetic acid, 50 % methanol	0.75
3.5 % acetic acid, 75 % methanol	1.25
3.5 % acetic acid, 100 % methanol	1
5 % acetic acid, 50 % methanol	1.25
5 % acetic acid, 75 % methanol	0.75
5 % acetic acid, 100 % methanol	1.25

Table 4-3. Concentration of 2-HMSA after elution from Amberlite IRA-400 resin. The concentration of 2-HMSA loaded onto the column was 0.85 mM. Equal volumes for loading and elution were used.

Elution buffer	% recovery after elution
90:10 TFA/water	8.7
90:10 formic acid/water	25.3
95:5 methanol/TFA	10.9
95:5 methanol/formic acid	0

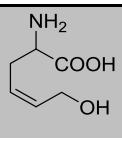
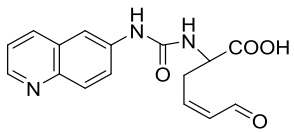
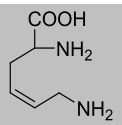
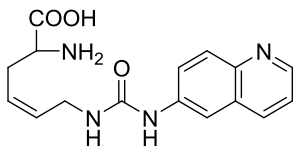
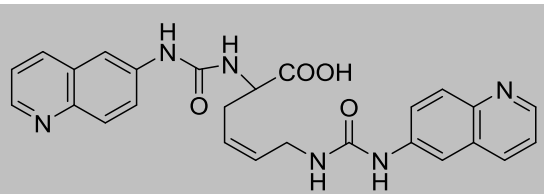
Attempts to elute the 2-HMSA from the resin and maintain a reasonable concentration for subsequent reactions with a transaminase were all unsuccessful.

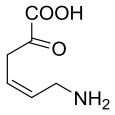
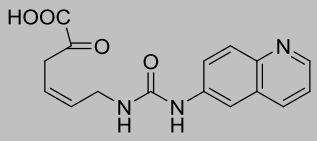
4.3.3 Derivatisation of product for HPLC and LCMS identification

As attempts to begin with a more defined reaction mixture were unsuccessful, the next step was to try to purify the product out of the mixture for identification. As mentioned earlier, it was concluded that the novel amine would be extremely polar and undetectable using HPLC with a standard C18 column. Therefore derivatisation of the amine with 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) was used (methods described in section 2.4.3). This increases the hydrophobicity and facilitates HPLC detection of amines (Marten & Naguschewski 2011). The transaminase assay with ArRMut11 transaminase clarified lysate expressed in *E. coli* BL21(DE3)* and unpurified 2-HMSA produced during the biotransformation reaction in *E. coli* DH10β and using serine as the amine donor was performed as described earlier in section 2.4.1.3. The reaction was quenched at multiple time points, including at 0h, 1h, 2h, 4h, 8h, and after 24h. The AQC derivatisation method was then used with each sample and analysed using HPLC analysis as described in section 2.4.3.4. The HPLC chromatogram at each time point was overlaid on top of one another as shown in Figure 4-4 and Figure 4-5. Using this overlaying technique it was possible to identify which peaks were present before the reaction has begun (at time = 0h) and which peaks corresponded to molecules that were

produced during the reaction. New peaks that arose over the reaction period were investigated further. Though new peaks were observed, they could correspond to a number of molecules within the mixture including free derivatisation agent or excess amine donor derivatised by the agent. The new peaks were isolated by fraction collection and further analysed using ESI Q-TOF MS. Potential structures arising from transaminase conversion of 2-HMSA were theorised, these are detailed in Table 4-4 along with their predicted derivatisation products. These predicted structures were used to identify relevant masses in mass spectrometry data.

Table 4-4. A speculative list of potential products from the 2-HMSA/transaminase conversion with structures after derivatisation and expected mass

Structure number	Structure	Exact mass
1		145.07389
2 (structure 1, after derivatisation)		315.12191
3		144.090
4 (structure 3, after derivatisation)		314.138
5 (structure 3, after double derivatisation)		484.186

Structure number	Structure	Exact mass
6		143.058
7 (structure 6, after derivatisation)		313.106

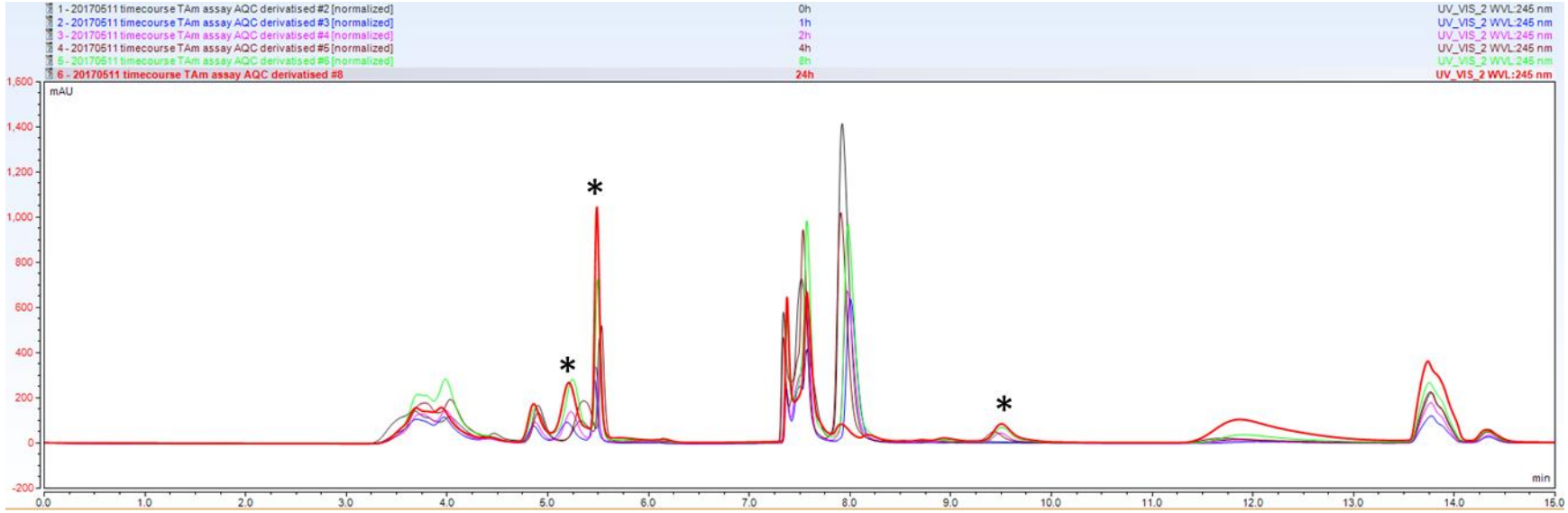


Figure 4-4. The change in HPLC chromatogram over a period of a 24 hour reaction.
Peaks of interest that were further analysed by mass spectrometry are noted with a *. Peaks at retention time 12.0 minutes and 13.8 minutes were not observed consistently during replications and therefore were not investigated further. Coloured curves represent different timepoints at which the reactions were quenched; grey = 0h, blue = 1h, pink = 2h, purple = 4h, green = 8h, red = 24h.

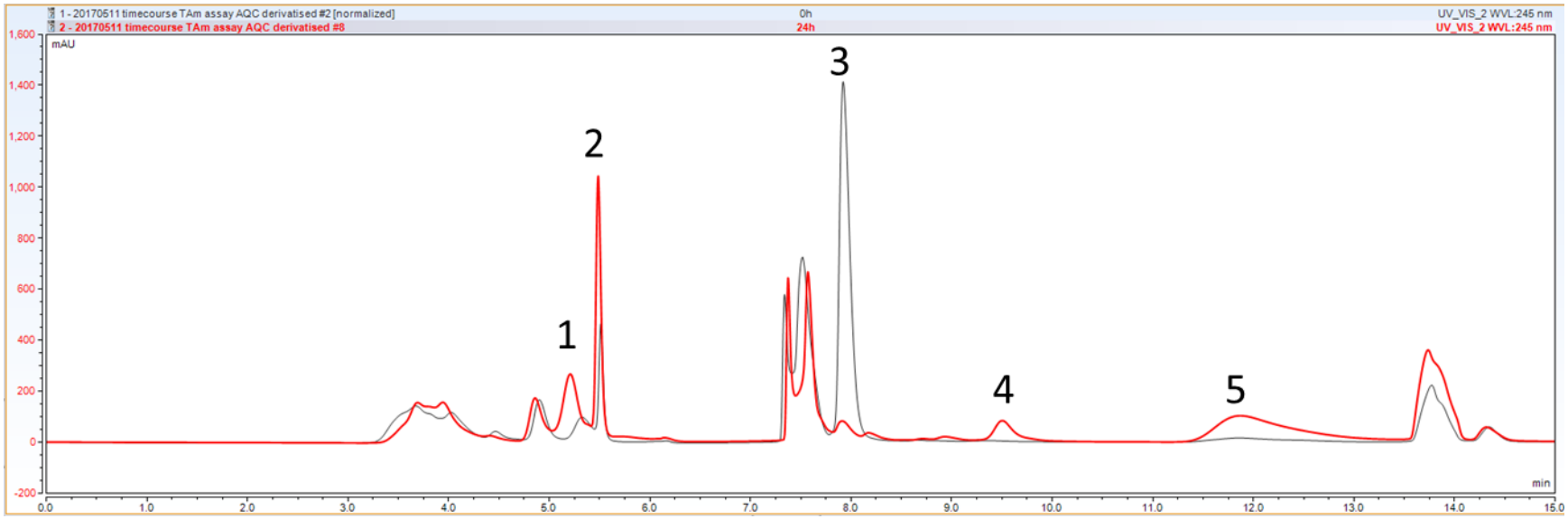


Figure 4-5. HPLC chromatogram showing the change in peaks from t=0h to t=24h.
Peaks at 5.2 (1) and 5.5 (2) minutes increase over time . The peak at 8.0 minutes decreases after 24 hours (3). Two new peaks at 9.5 (4) and 11.9 (5) appear after 24 hours. Peaks 1, 2 and 4 were isolated and collected for further mass spectrometry analysis. Peak 5 was not reproducible over multiple replications of this conversion reaction and therefore was not investigated further. Peaks that showed no or little change in area were of no interest as they most likely correspond to components that are not involved in the reaction. Coloured curves represent different timepoints at which the reactions were quenched; grey = 0h, red = 24h.

Analysis of Figure 4-5 indicated peaks to be isolated for further mass spectrometry analysis. Peak 5 was impossible to collect as it was not reproducible consistently during HPLC runs indicating either degradation of this compound or an artifact arising from impurities in the reaction mixture. Peak 3 was of no interest as it is diminished after 24 hours and therefore is likely to correspond to a component of the reaction that is used such as serine which was used as the amine donor. Peaks at 1, 2 and 4 in Figure 4-5 were isolated and collected. This was repeated for 15 runs. Fractions corresponding to the same peak were pooled and concentrated. The total 24 hour derivatised reaction was analysed using LC-MS. Relevant masses were observed as follows; a mass of 336.2 indicating potential sodium adduct of structure 7 in Table 4-4; mass of 314.18 was also observed indicating the possible presence of structure 4 or a hydrogen adduct of structure 7 in Table 4-4. Purified fractions of peaks 1, 2 and 4 from Figure 4-5 were were submitted for ESI Q-TOF MS which showed the following relevant mass; 144.9296 which would indicate the possible presence of structure 3. Analysis shown in Figure 4-6 is the ESI Q-TOF MS analysis of peak 4 in Figure 4-5 and showed a mass of 337.2717 and a fragmentation pattern which matches what would be expected for product 4 in Table 4-4.

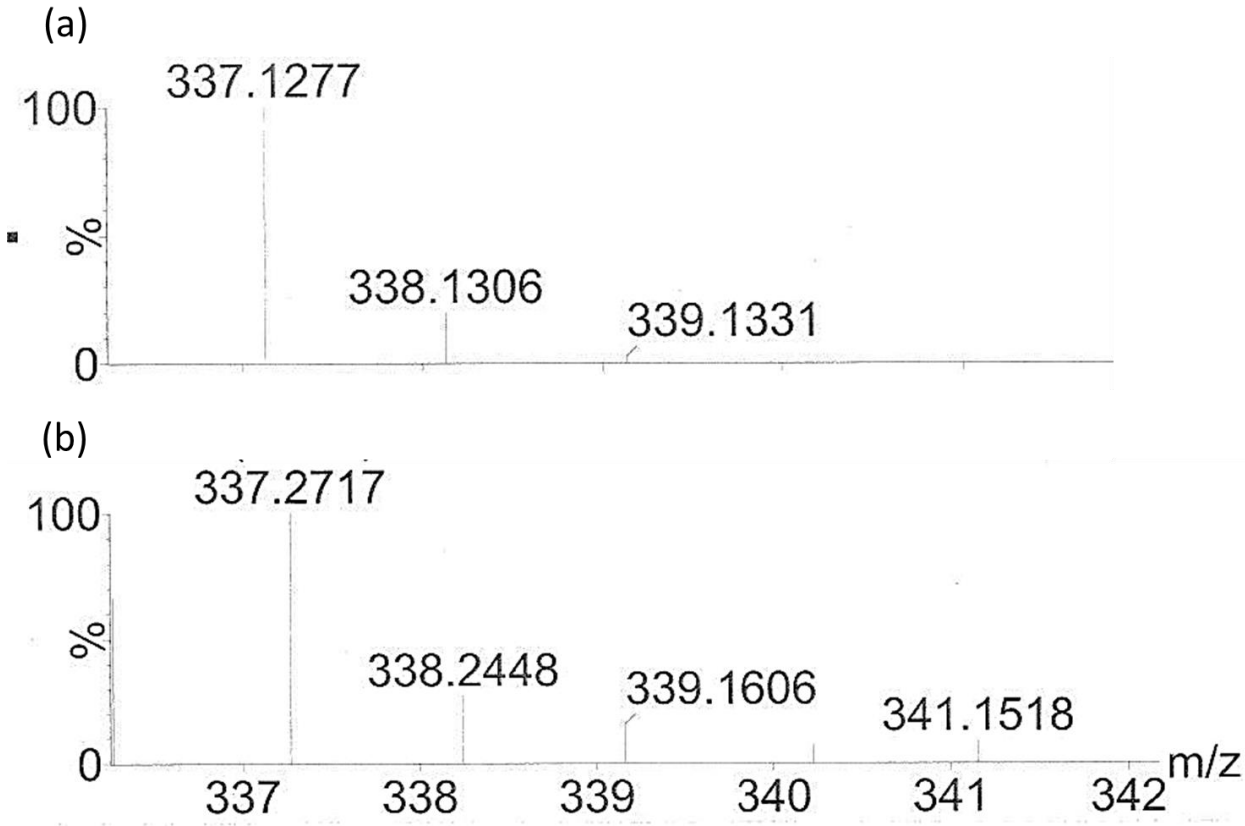


Figure 4-6. ESI Q-TOF MS analysis of the amine product
(a) The expected pattern produced by product 4 in table 4-2 with a sodium adduct as generated using the ESI Q-TOF MS analysis software. (b) The actual observed masses produced by the purified sample corresponding to the peak at 9.5 minutes in Figure 4-5.

There are multiple pieces of evidence to indicate the presence of structure 4 which is the single derivatisation product of structure 3 in Table 4-4. This information leads us to believe that the amine produced by transaminase conversion of 2-HMSA is structure 3 in Table 4-4. Further confirmation of this structure would have to be determined using nuclear magnetic resonance (NMR) techniques, however the concentration of the amine product was not high enough in this instance for NMR techniques. Further work to confirm the structure of this product is outlined in Chapter 7.

4.4 Discussion

4.4.1 Transaminase and amine donor screening

From the data shown in Figure 4-1 it appears that 2-HMSA is a better substrate for *(R)*- selective transaminases even though 2-HMSA is achiral. The activity observed with pQR1049 and pQR1048 is consistent with previous studies where these transaminases were active with other aromatic compounds (Nina Richter et al. 2014).

With serine as an amine donor where the reaction was observed over 24 hours, there appears to be a decrease in conversion compared with after 3 hours. This is due to transaminases having bidirectional activity therefore the reaction will eventually reach equilibrium. This in addition to the known degradation of 2-HMSA explains the lower rate of conversion. It is possible that the actual rate of conversion of 2-HMSA is higher than it appears. This is due to the transaminase conversion possibly occurring faster than the degradation of 2-HMSA. The only way to verify the actual conversion rate would be to quantify the product concentration. Due to the product being novel and the possibility of multiple products this has not yet been possible.

For conversions with alanine and glutamate, activity with pQR977, pQR1019 and pQR1021 should not be expected as these are labeled as serine transaminases in protein family databases. These were chosen due to good activity reported previously with serine (Villegas Torres 2014). However, we are now beginning to discover that there are often large discrepancies between the automatic annotations in the databases which are based on overall protein/protein homologies, and the actual substrate specificity of enzymes (Schnoes et al. 2009; Cozzetto & Jones 2017; Poux et al. 2014); particularly for transaminases which have a diverse and complex evolutionary history (Muratore et al. 2013). Transaminases pQR1019 and pQR1021 showed no activity with 2-HMSA as the substrate. The

Arthrobacter mutant transaminase, pQR1049, showed the most consistent activity with 2-HMSA regardless of amine donor. This aligns with literature that it is very robust and promiscuous (Richter et al. 2015; N. Richter et al. 2014; Savile et al. 2010). The next experimental aim at this point was to clone the selected transaminase into the pQR1050 vector so that the transaminase is expressed with the upstream TOL enzymes as an engineered pathway and to test the engineered pathway with a range of substrates as described in chapter 5.

4.4.2 Structural determination

The mass spectrometry data shown gives an indication of the structure of the product formed. Further work to purify the product further and conduct further structural determination work such as NMR analysis would be preferable. Purification of the product would also allow quantification and further analysis into properties. The current purification method by HPLC results in impurities as seen by mass spectra obtained. It is certain that structure 5 (Table 4-4) was not observed therefore there was no double derivatisation occurring. A collective analysis of mass data gives a strong indication that structure 3 is the amine produced by transamination of 2-hydroxymuconic semialdehyde.

The aims of this study were to show as a proof of concept that it is possible to create novel interesting compounds that would be difficult to produce by traditional chemical synthesis methods. By using an engineered pathway including a transaminase, a novel amine has been created in this study. Using IUPAC nomenclature this amine is named 2,6-diaminohex-4-enoic acid. The properties of this molecule are unknown so further work would be to elucidate these. Speculatively speaking, it is possible that the molecule could undergo polymerisation (Figure 4-9). Previous literature has described a process with 3-chloro-2-HMSA and ammonium chloride to produce a picolinic acid derivative as shown in Figure 4-7 (Riegert et al. 1998).

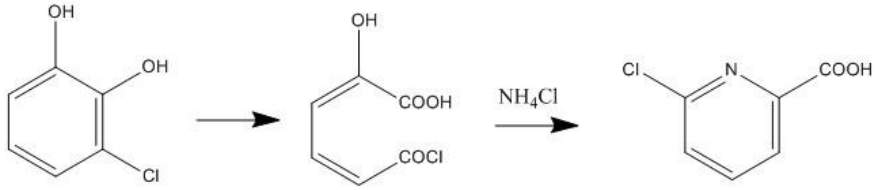


Figure 4-7. The conversion of chloro-catechol to 3-chloropicolinic acid using catechol 2,3-dioxygenase (XylT) and ammonium chloride in a 2 step reaction.

The novel amine produced here could also potentially undergo cyclisation to produce picolinic acid, a precursor to picolinic acid or a derivative. Picolinic acid (Figure 4-8) is of industrial interest as it assists the absorption of zinc ions in the small intestine amongst other functions (Grant et al. 2009).The novel amine produced has a very similar structure to 2-aminomuconic semialdehyde which is a *metab*olite of tryptophan and an intermediate in the biosynthesis of picolinic acid. There is some precedence for this as *metab*olites similar to 2-HMSA have been used to create picolonic acid using high temperatures, pressures and the use of ammonia (Riegert et al. 1998). Chapter 7 describes further work that could be done to investigate this.

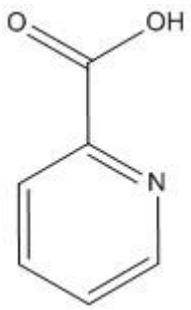
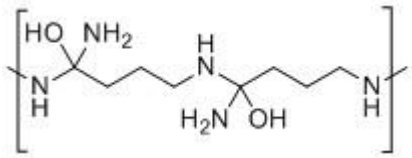


Figure 4-8. Picolinic acid
Picolinic acid could be produced by cyclisation or further reaction of the novel amine produced by transamination of 2-HMSA.

a)



b)

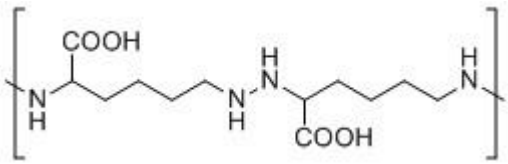


Figure 4-9. Speculation of polymers that may be produced by polymerisation of the novel amine produced by transamination of 2-HMSA.
(a) shows a polymer produced by condensation polymerisation of the carboxylic acid and amine. (b) shows the polymer that may be produced by the dehydrogenation of amines at each end of the molecule to produce a nitrogen-nitrogen bond.

To test the amine product in any further reaction the amine must be purified from the reaction mixture. In addition to purification, the conversion must also somehow be scaled up to produce enough amine for further reactions. Difficulties in purification and scale up of the aldehyde to amine conversion were encountered as described earlier in this chapter. These are challenges that would need to be overcome in any process that is relevant to industry. In order to promote the use of bio-conversions and biotransformations in industry, these processes must be scaleable and be amenable to further downstream processing.

Chapter 5: Whole cell conversion of a range of substrates for the engineered TOL pathway with a transaminase in *Pseudomonas putida* as the host organism

Chapter 5. **Whole cell conversion of a range of substrates for the engineered TOL pathway with a transaminase in *Pseudomonas putida* as the host organism**

5.1 Background

Using the ArRMut11 transaminase selected after the transaminase screening a vector was constructed to incorporate the transaminase gene into the truncate of the TOL *meta*cleavage plasmid (Figure 5-1). The aims of this study were to investigate this engineered pathway as a whole cell reaction. Performing the reaction in whole cells has both benefits and downsides when it comes to commercial relevance. The positive attributes of performing reactions in whole cells is the provision of energy and maintenance of the intracellular conditions that these enzymes would be normally working in, as long as the host cell is appropriate. As well as this, whole cells thrive in aqueous conditions and at ambient temperatures as mentioned previously, thus minimising heating costs and potential environmental pollutants from toxic solvent waste. There are also negatives to using whole cells as a reaction vessel. Reaction conditions that may be necessary for the reaction but are not natural to a cells natural state are more difficult to maintain and monitor. Many molecules, substrates or cofactors for example, are not cell membrane permeable; this could mean that cofactors or substrates cannot get into the cell or could mean that product cannot get out of the cell. Product build up within a cell could become inhibiting and limit conversion or could even be toxic to the cell. On top of this the desired product needs to be extracted and isolated from all the other cellular material which can be difficult and time consuming as discovered in chapter 4. In addition, the overexpressed enzymes compete for energy and resources with other cellular processes. Limiting cell growth or the overproduction of the recombinant enzyme may limit overall conversion to the desired reaction product. It would appear from this list that the negative arguments against using a live organism for chemical production processes may be stronger

than the positive, but the only way to change this is to continue studying reactions using whole cells in order to understand them. From a better understanding it will be possible to optimise reactions and produce systems that are feasible for commercial industrial processes.

The reaction was investigated in *P. putida* GS1 and various starting materials were tested as the TOL pathway is known to accept a range of substrates (Murray et al. 1972; Abril et al. 1989). This versatility of enzymes and enzymatic pathways provides options to tailor the molecules that are created to specific needs for active pharmaceutical ingredients or other chemicals with value. This flexibility also gives us the option to design molecules and "mix and match" enzymes to create novel molecules. The aim of this study was to investigate the promiscuity of the TOL *meta*cleavage pathway which has been investigated previously (Murray et al. 1972; Williams & Murray 1974), and of the *Arthrobacter* mutant transaminase which has also been shown to have a range of substrate specificity with other types of molecules such as steroids (Nina Richter et al. 2014). Using different materials as starting substrates for this engineered pathway could lead to different novel amines being created, as long as the enzymes in the engineered pathway accept the modified intermediates. The rationale behind this proof of principle study is that different amines could be biologically active in different ways or could act as intermediates for other industrially relevant chemicals.

5.2 Construction of a vector containing an engineered pathway with a transaminase

In this study various methylbenzoates were investigated with the truncate of the TOL *meta*cleavage pathway and with the engineered pathway with the ArRMut11 transaminase. A vector was constructed with the transaminase to create a single vector with the engineered pathway (Figure 5-1).

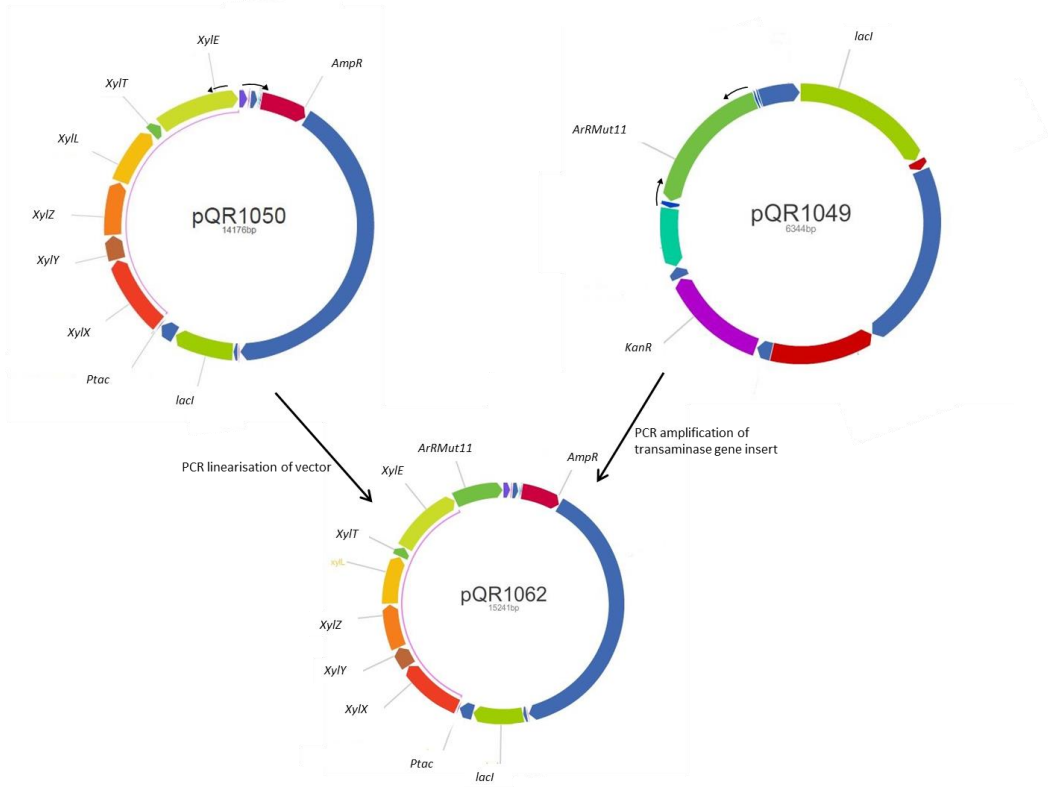


Figure 5-1. Construction of pQR1062 containing the *xylXYZLTE* genes and the ArRMut11 transaminase.
pQR1050 was linearised using PCR amplification with primers containing 5' overhangs overlapping with the ArRMut11 transaminase insert. Primers for ArRMut11 amplification had 5' overlaps with the insertion point of pQR1050. In-fusion cloning was performed as described in section 2.1. Sequence was verified by Sanger sequencing. Black curved arrows depict primer position and direction.

To investigate the conversion of various methylbenzoates to both the semaldehyde and to an amine after transamination,conversion of the methylbenzoate substrates to the aldehyde product of the TOL pathway were performed with pQR1050 and pQR1062 (without and with transaminase respectively). The pathway in pQR1050 ends in the production of the aldehyde and therefore this product should accumulate. With cells expressing pQR1062 the pathway ends in the transaminase which should accept the aldehyde as a substrate and therefore concentration of the aldehyde should be reduced compared with cells expressing the TOL pathway enzymes only. The growth and substrate conversion methods were kept consistent with the methods previously used in *P. putida* comparison studies (described in section 2.3.) except for the addition of 1mM PLP required for transaminase activity. Expression was induced after 2 hours and substrates were added after 3 hours. The concentration of the aldehydes (after enzymatic conversion with XylTE) are measured by recording the absorbance as described earlier in section 2.3.1. Biological triplicates and two sample t-tests were performed to assess statistical significance between concentrations of the aldehyde produced.

5.3 Conversion of benzoic acid

The conversion of benzoic acid to 2-HMSA and to the amine has been confirmed in chapters 3 and 4. The reaction schematic can be seen in Figure 3-1. During previous transaminase activity studies, transaminase was expressed independently of the TOL plasmid and active enzyme in clarified cell lysate was used. The following study investigates the transaminase activity with 2-HMSA in a whole cell reaction where the TOL pathway truncate and the transaminase are expressed using one plasmid, pQR1062.

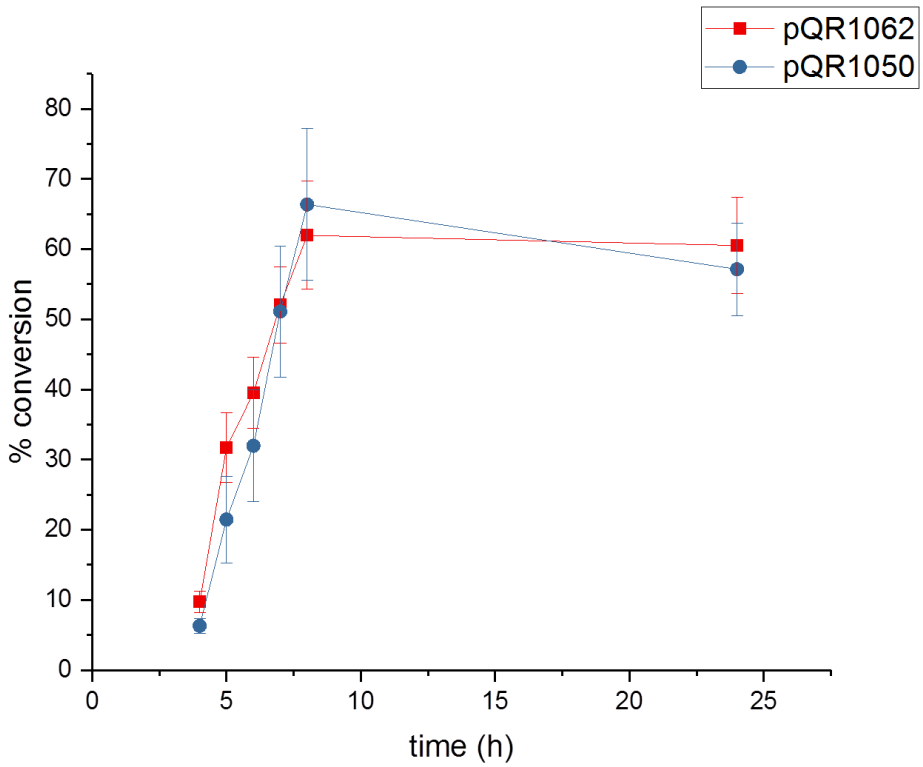


Figure 5-2. A comparison of the conversion of benzoate to 2-HMSA with and without transaminase.
5 mM sodium benzoate was added 1 hour post induction as described in section 2.3. pQR1050 contains the TOL *meta*cleavage pathway enzymes *xylXYZLTE*. pQR1062 contains the full engineered pathway including the TOL *meta*cleavage pathway enzymes and the transaminase. Conversion was calculated by measuring the absorbance at 375 nm.

No significant difference in 2-HMSA concentration was observed between cells expressing pQR1050 and pQR1062 even though transaminase activity has been previously observed with 2-HMSA. This is clear from Figure 5-2 as there is virtually no difference between 2-HMSA concentration at any point between 4 and 8 hours or at 24 hours. This is also seen in the growth rates which appeared to be very similar for cells expressing both plasmids as shown in Figure 5-3.

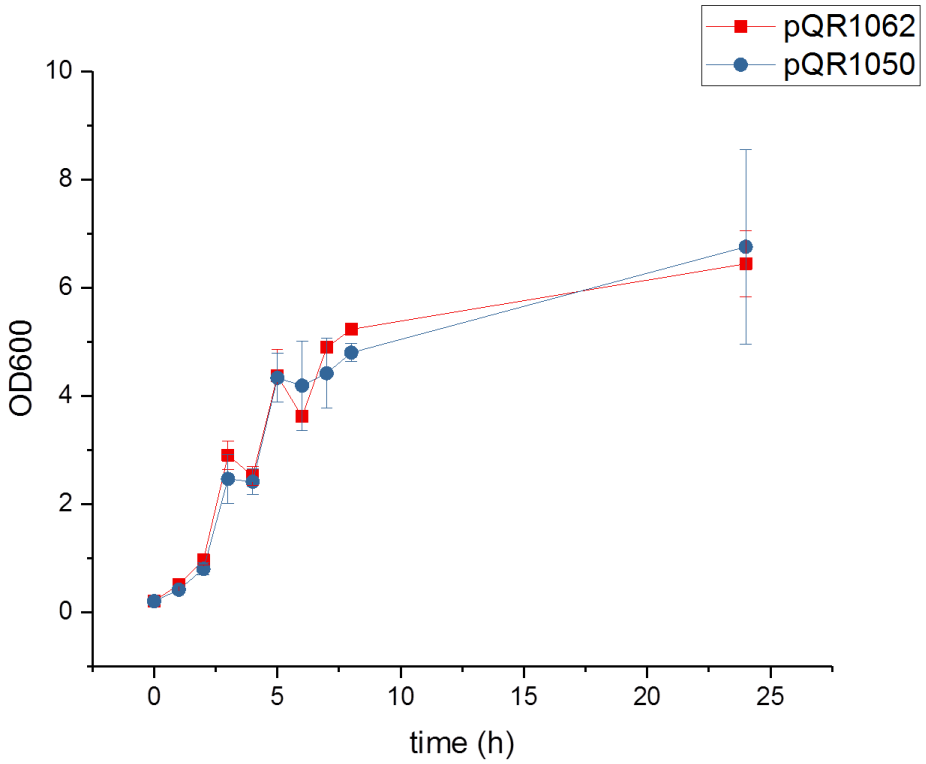


Figure 5-3 A comparison of the growth rates of *P. putida* GS1 expressing pQR1062 and pQR1050.
Expression was induced after 2 hours and sodium benzoate was added after 3 hours.

5.4 Conversion of m-toluic acid

The conversion of modified benzoates were investigated with the TOL pathway to determine whether the TOL pathway enzymes would accept these substrates and whether the transaminase would also accept a range of substrates, thus indicating the flexibility of engineered pathways for producing a desired molecule. M-toluic acid (3-methylbenzoic acid) was one substrate investigated. This substrate was chosen as it has been previously reported to be converted by the TOL enzymes and the resulting aldehyde is also close in structure as 2-HMSA, meaning that it is also likely to be accepted by the transaminase (Murray et al. 1972).

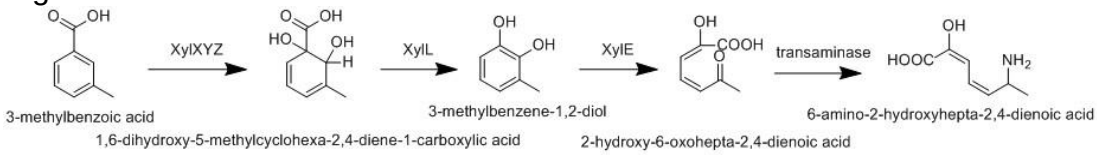


Figure 5-4. Reaction of 3-methylbenzoic acid through to the predicted structure of the amine after transamination of the aldehyde.

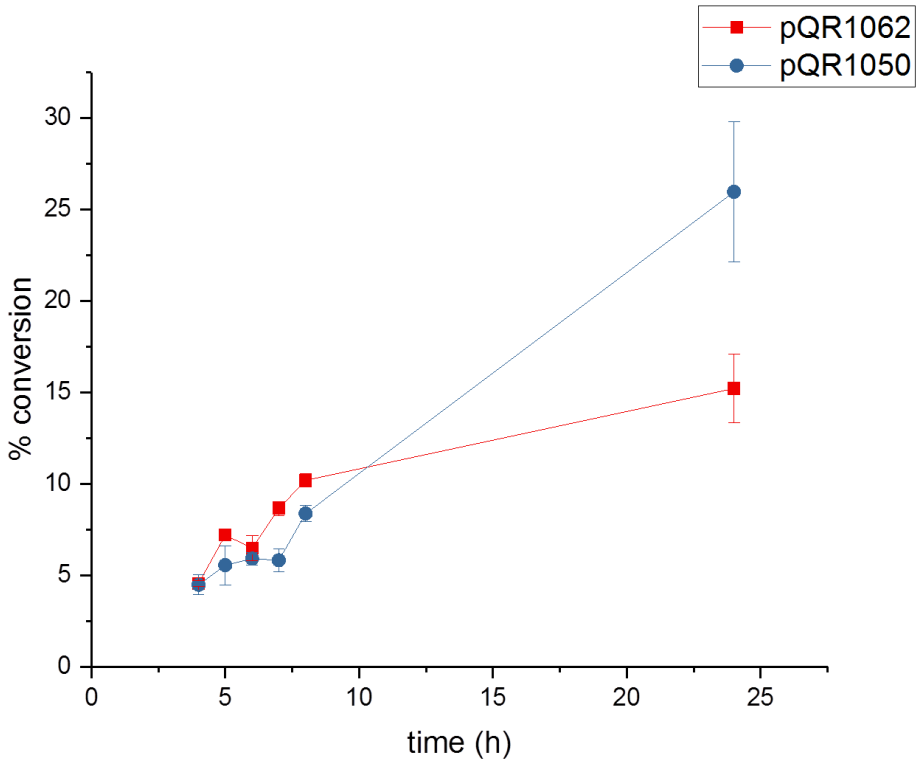


Figure 5-5. Conversion of 3-methylbenzoic acid and conversion with and without transaminase.
5 mM 3-methylbenzoic acid was added 1 hour post induction as described in 2.4.4. pQR1050 contains the TOL *meta*cleavage pathway enzymes *xylXYZLTE*. pQR1062 contains the full engineered pathway including the TOL *meta*cleavage pathway enzymes and the transaminase. Conversion was calculated by measuring absorbance of the aldehyde 2-hydroxy-6-oxohepta-2,4-dienoic acid at 388 nm.

Over the initial 8 hours of expression there does not appear to be much difference between the concentrations of the aldehyde (2-hydroxy-4-methyl-6-oxohexa-2,4-dienoic acid). After 24 hours there is clearly a lower concentration of aldehyde with pQR1062. Statistical analysis confirms that the difference in concentrations is significant. There appears to be an

approximately 40% conversion of the aldehyde by the transaminase after 24 hours based on the average percent conversion (Figure 5-5). The presence of the transaminase is the only difference in the two reactions and therefore it is assumed that the conversion of the aldehyde to a novel amine by the transaminase is the cause of the difference in concentration observed. Growth rates for these conversions are shown in Figure 5-6 and cultures expressing both plasmids grow at a similar rates and diverge after 24 hours. Growth curves show an interesting pattern; both cultures appear to shown a reduction in OD_{600} at 7 hours with growth picking up again after 8 hours. All experiments were performed in triplicate and this trend appeared to occur consistently. After 24 hours the OD_{600} for cells expressing pQR1050 is considerably lower, which could be due to catechol build up inside the cells causing a toxic effect.

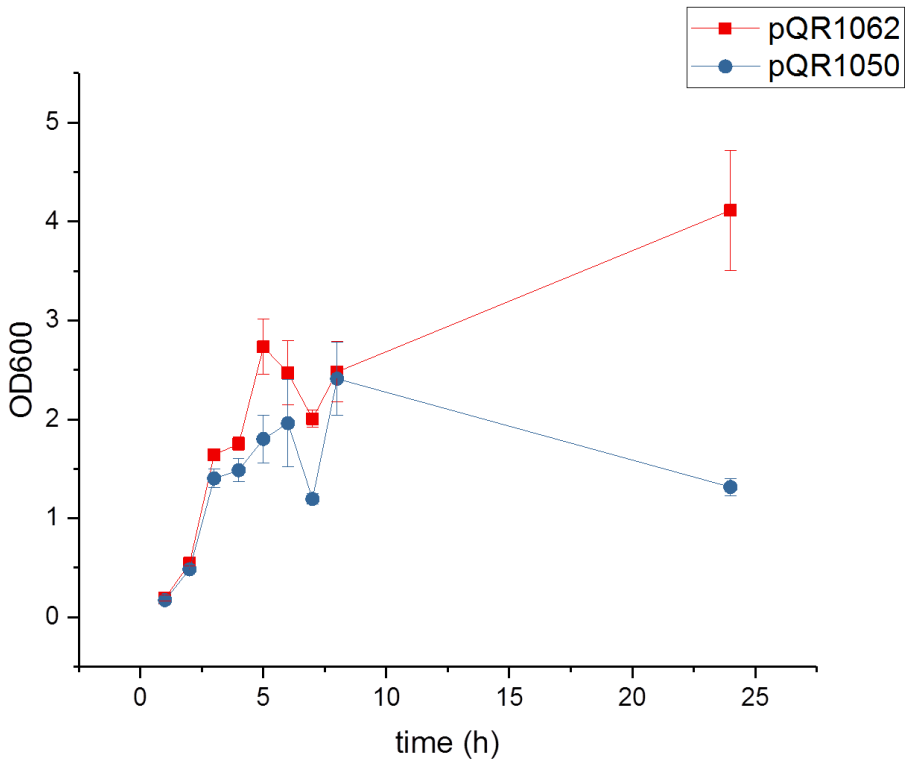


Figure 5-6 Growth rates of *P. putida* GS1 expressing pQR1050 and pQR1062 were measured over time.
Expression was induced after 2 hours and 3-methylbenzoic acid was added after 3 hours.

The whole cell conversion of 3-methylbenzoic acid was also analysed by HPLC. In order to reduce background and unknown compounds present a biotransformation reaction was performed as detailed in section 2.4.4.2. The reaction was stopped after 24 hours and supernatant was derivatised using the AQC method in section 2.4.3.3. The HPLC chromatograms were overlaid to compare which peaks appeared in the presence and absence of the transaminase at shown in Figure 5-7.

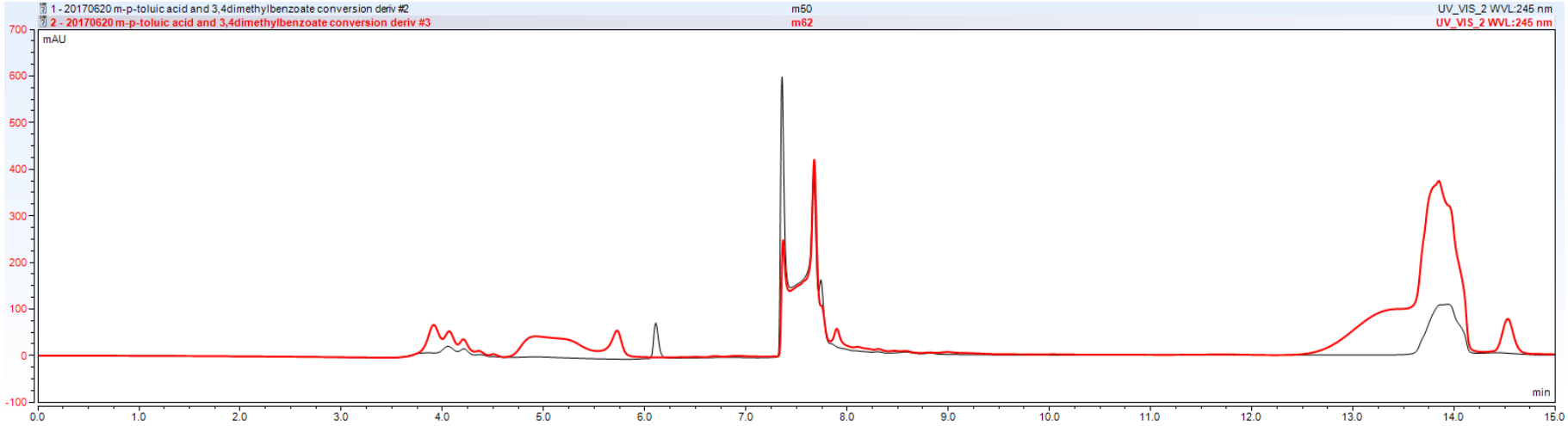


Figure 5-7. HPLC chromatograms of the biotransformation of 3-methylbenzoic acid in cells containing enzymes expressed from pQR1050 (grey) and pQR1062 (red).
pQR1062 contains the ArRMut11 transaminase and therefore is expected to result in an aminated product. pQR1050 does not contain the transaminase and this pathway results in an aldehyde. The identity of the peaks are unknown.

From HPLC analysis the presence of new peaks at retention times of 5 minutes, 5.727 minutes and 14.53 minutes indicates that there was transaminase conversion, however no further analysis was performed so the identity of these has not been confirmed. The new peaks also follow a similar pattern to new peaks observed during the conversion of 2-HMSA to the amine as shown in Figure 4-5. The peaks that appear between 7-8 minutes also follow a similar pattern as observed previously with the conversion of 2-HMSA and are likely to be components of the reaction mixture that do not change such as PLP, however this has not been investigated further.

5.5 Conversion of p-toluic acid

P-toluic acid (4-methylbenzoic acid) was also investigated with the truncated TOL pathway and the engineered pathway to determine whether 4-mehtylbenzoic acid would be converted to an aldehyde and to a novel amine.

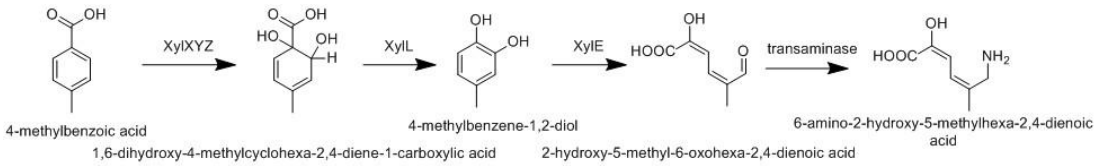


Figure 5-8. Reaction of 4-methylbenzoic acid through to the predicted structure of the amine after transamination of the aldehyde.

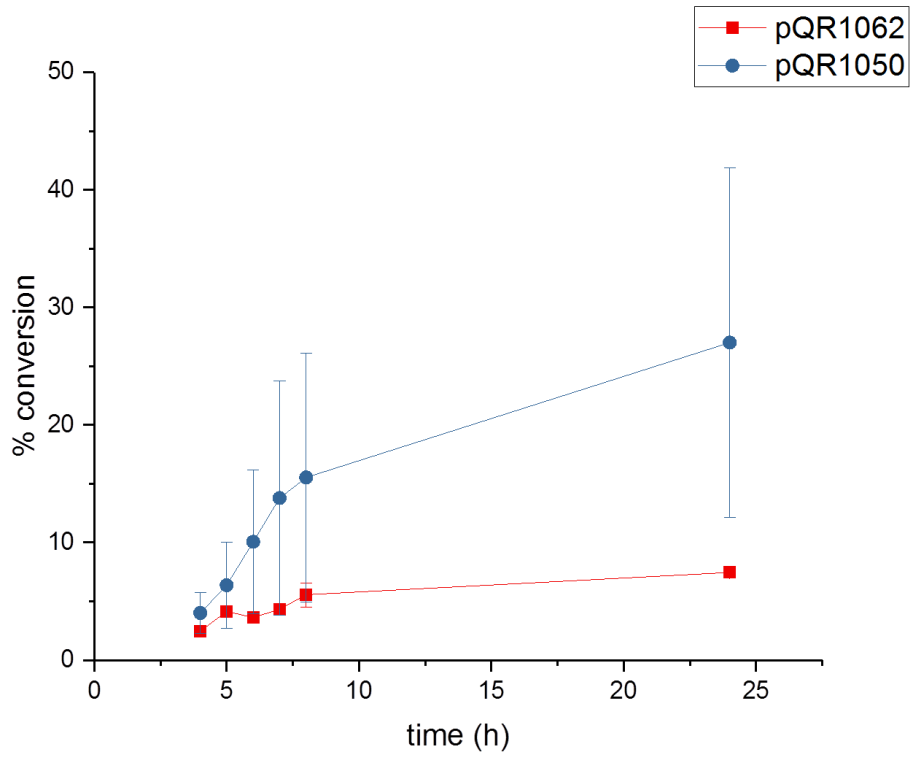


Figure 5-9. Conversion of 4-methylbenzoic acid with and without transaminase. 5 mM 4-methylbenzoic acid was added 1 hour post induction as described in section 2.4.4.2. pQR1050 contains the TOL metacleavage pathway enzymes *xylXYZLTE*. pQR1062 contains the full engineered pathway including the TOL metacleavage pathway enzymes and the transaminase. Concentration was calculated by measuring absorbance at 382 nm.

Concentration of aldehyde (2-hydroxy-5-methyl-6-oxohexa-2,4-dienoic acid) appears to be higher with cells expressing pQR1050 from hour 4, one hour from the addition of substrate although for cells expressing pQR1050 a large error margin was seen as shown by error bars in Figure 5-9. This means that there is no statistical significant difference between concentrations up to 8 hours. There is much lower error observed with cells expressing pQR1062. A significant difference between concentrations was only seen after 24 hours. Based on the average percent conversion there appears to be approximately 74% conversion to the amine after 24 hours though due to the large error observed for pQR1050 this may not be very accurate. The large error bars for pQR1050 have risen from a wide variation in the conversion from

experiment to experiment. The reason for this is unknown. A comparison of the growth curves also shows a wide variation in the growth of cells expressing pQR1050 from experiment to experiment, which could explain the variation in conversion (Figure 5-10). The growth of these cells also show a dip in OD600 at 7 hours, consistent with cells fed with m-toluic acid as a substrate. For the first 8 hours, cells follow a similar growth curve and therefore conversion rates from p-toluic acid to the aldehyde are likely to be comparable. This indicates that the reduction in aldehyde present with pQR1062 could be due to the transaminase converting this aldehyde further.

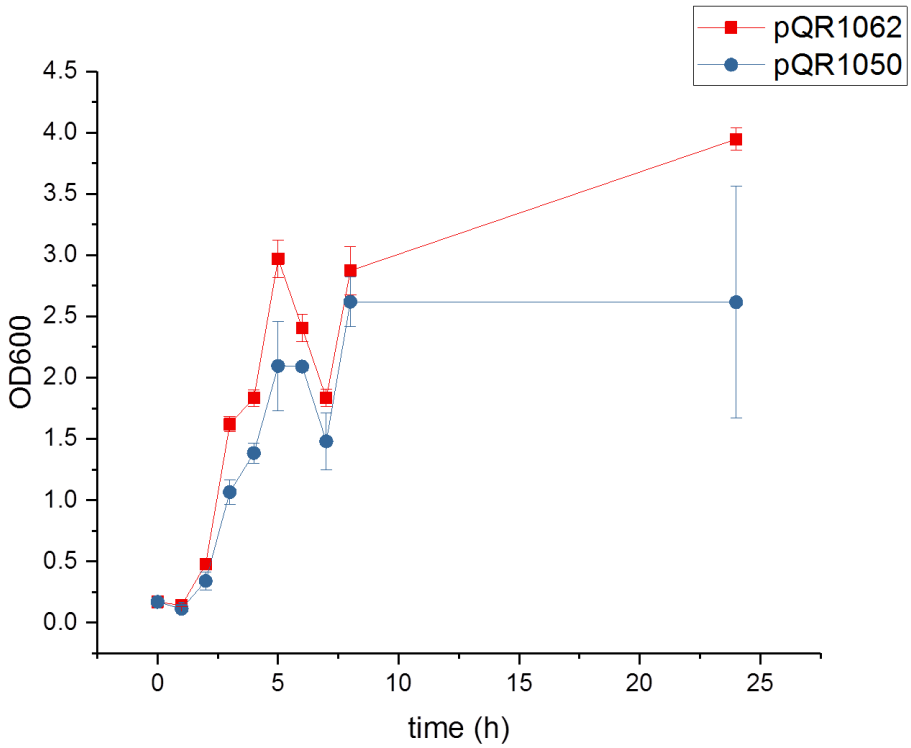


Figure 5-10 Comparison of the growth of *P. putida* GS1 expressing either pQR1050 or pQR1062.
Expression was induced after 2 hours and substrate was added after 3 hours.

Samples were derivatised using AQC and analysed by HPLC as described in sections 2.4.3.3 and 2.4.3.4 respectively. Results of each pQR1050 and

pQR1062 were overlaid to compare which, if any, peaks appear in the presence of the transaminase. The chromatograms are shown in Figure 5-11.

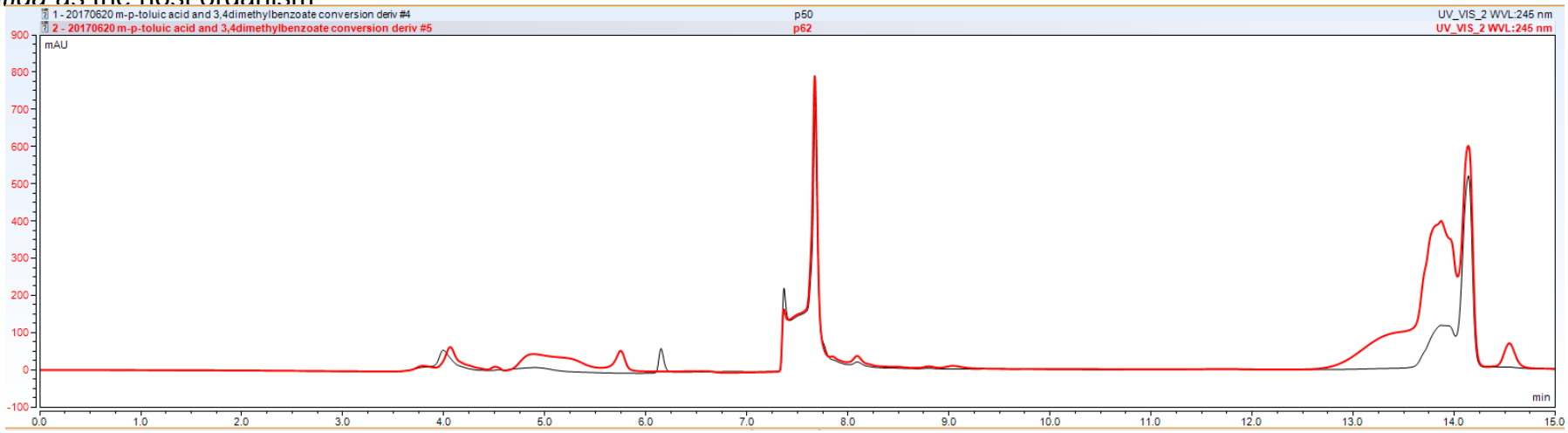


Figure 5-11. HPLC chromatograms of the biotransformation of 4-methylbenzoic acid in cells containing enzymes expressed from pQR1050 (grey) and pQR1062 (red).
pQR1062 contains the ArRMut11 transaminase and therefore is expected to result in an aminated product. pQR1050 does not contain the transaminase and this pathway results in an aldehyde. The identity of the peaks shown here are unknown.

New peaks were observed at 5 minutes, 5.75 minutes and 14.547 minutes, again showing a similar pattern to the conversion of benzoate and 3-methylbenzoic acid.

5.6 Conversion of 3,4-dimethylbenzoic acid

The final modified benzoate to be investigated with the truncated TOL pathway and the engineered pathway was 3,4-dimethylbenzoic acid. This was the largest of the substrates tested, as seen in the reaction scheme below (Figure 5-12).

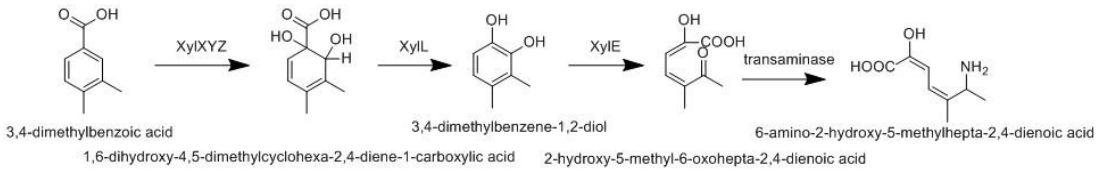


Figure 5-12. Reaction of 3,4-dimethylbenzoic acid through to the predicted structure of the amine after transamination of the aldehyde.

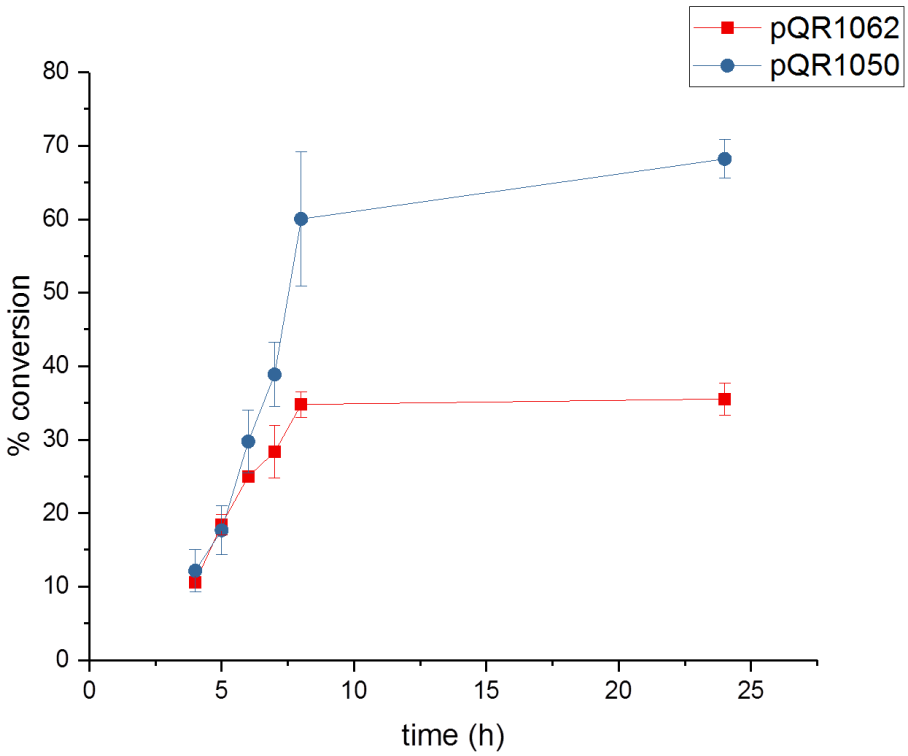


Figure 5-13. Conversion of 3,4-dimethylbenzoic acid and conversion with and without transaminase.
5 mM 3,4-dimethylbenzoic acid was added 1 hour post induction as described in section 2.4.4. pQR1050 contains the TOL *meta*cleavage pathway enzymes *xylXYZLTE*. pQR1062 contains the full engineered pathway including the TOL *meta*cleavage pathway enzymes and the transaminase. Conversion was calculated by measuring absorbance at 390 nm.

With 3,4-dimethylbenzoic acid as the starting material there is a clear difference between apparent percent conversion of the resulting aldehyde (2,4-heptadienoic acid) between pQR1050 and pQR1062 (Figure 5-13). Conversion rates of the aldehyde between hours 4 to 6 are similar, and there is no statistical difference between points during this time. After hour 6 of growth, which corresponds to 3 hours after substrate addition the concentration of aldehyde increases more in cells containing pQR1050 compared with cells containing pQR1062. After 7 hours there is a significant difference between the concentrations of aldehyde. Results show an approximately 48% conversion of the aldehyde by the transaminase to the

aminated product after 24 hours based on the difference between the average percent conversions to the aldehyde. The growth of both cultures were very similar as shown in Figure 5-14. This indicates that conversion rates should also be similar and any difference in the concentration of the aldehyde is due to transaminase conversion.

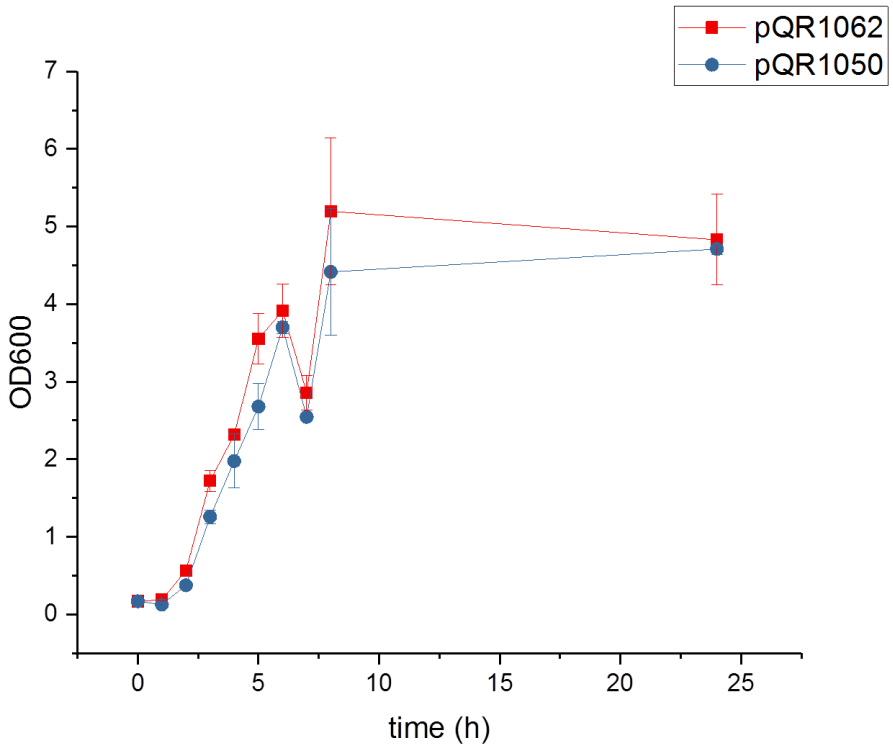


Figure 5-14 Comparison of the growth of *P. putida* GS1 expressing pQR1062 and pQR1050.
Expression was induced after 2 hours and 3,4-dimethylbenzoic acid was added at 3 hours.

The reaction was analysed by HPLC after AQC derivatisation (described in section 2.4.3.3), results for each reaction were overlaid as shown in Figure 5-15.

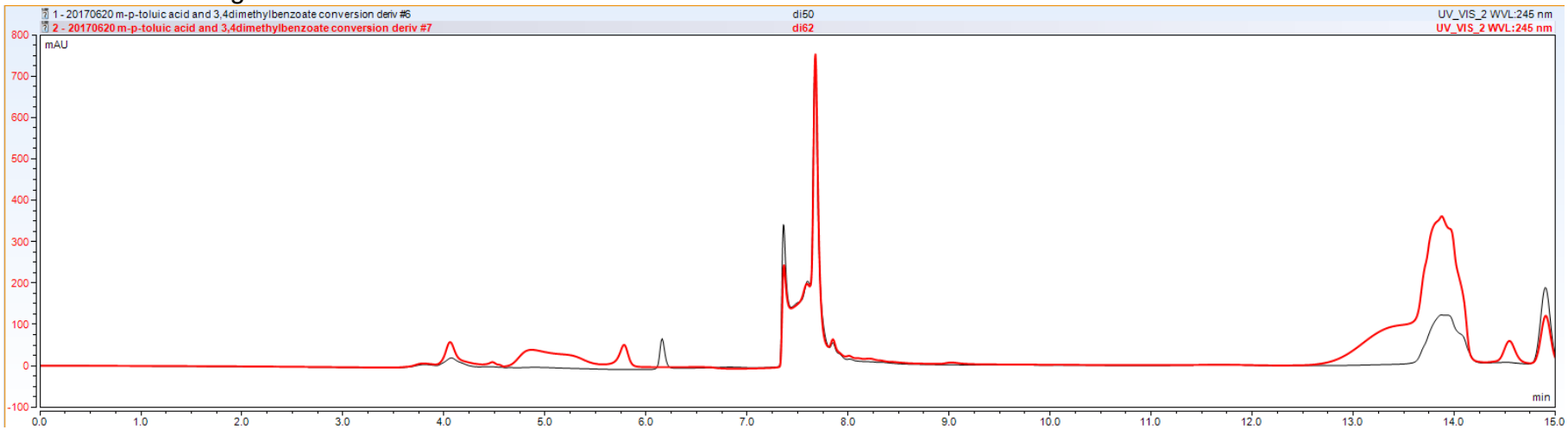


Figure 5-15. HPLC chromatograms of the biotransformation of 3,4-dimethylbenzoic acid in cells containing enzymes expressed from pQR1050 (grey) and pQR1062 (red).
pQR1062 contains the ArRMut11 transaminase and therefore is expected to result in an aminated product. pQR1050 does not contain the transaminase and this pathway results in an aldehyde. Identity of the peaks were not determined.

New peaks were observed at 5 minutes, 5.783 minutes and 14.547 minutes. This also displays the same peak pattern as previous HPLC chromatograms (Figure 4-5, Figure 5-7 and Figure 5-11). The peak at approximately 5 minutes is a particularly broad peak, possibly made of two broad peaks that have merged. This is also consistent with other HPLC traces in figures Figure 5-7 and Figure 5-11.

5.7 Discussion

In chapter 4 we observed the conversion of 2-HMSA to a novel amine however in Figure 5-2 there does not appear to be a difference in concentration of 2-HMSA with and without transaminase which on its own would indicate no transaminase conversion of the 2-HMSA. A possible explanation for this may be poor transaminase expression or activity however transaminase activity is seen for modified substrates in sections 5.4, 5.5 and 5.6 and therefore this can be eliminated as a possible reason for no apparent conversion. Another possible explanation is that the 2-HMSA is degrading, as observed previously (section 4.2.1) and reported in the literature (Asano et al. 2014; Vanier 1998). The instability of the 2-HMSA has resulted in a decrease in concentration even without the transaminase therefore no difference can be seen between the two biotransformations.

There are clear differences between the concentrations of aldehydes produced by the modified substrates. This indicates that these aldehydes are more stable than 2-HMSA. In Figure 5-2, Figure 5-9 and Figure 5-13 the concentration of aldehyde continues to increase until 24 hours when the last sample was analysed whereas 2-HMSA produced by cells containing pQR1050 in Figure 5-5 decreases between 8 and 24 hours. This provides

further confirmation of the instability of 2-HMSA and in comparison, increased stability of the other aldehydes.

HPLC analyses for all substrates show the same peak pattern. This is expected as the substrates are extremely similar in structure and the resulting amine is also expected to be very similar. There are three new peaks that appear which also matches with new peaks observed during the structural analysis of the amine in section 4.3. The peak identified as the novel amine was the peak at a retention time of approximately 9.5 minutes, therefore it would be sensible to assume that the novel peaks observed at a similar retention time for the modified substrates would also be the novel amine produced by the transaminase. The next steps to confirm the structure of the novel molecules produced would require purification of this peak and mass spectrometry analysis as performed in chapter 4.

Searches for the predicted structures of the amines using SciFinder® (scifinder.cas.org) indicates that these molecules have not been published previously in the literature confirming the novelty of the molecules and of the engineered pathway. The properties of the novel amines presented here are unknown. Due to time constraints and the lack of purified highly concentrated product, NMR was not possible to confirm the identities of the products in this chapter. Further work to confirm the structures of these products is described in Chapter 7.

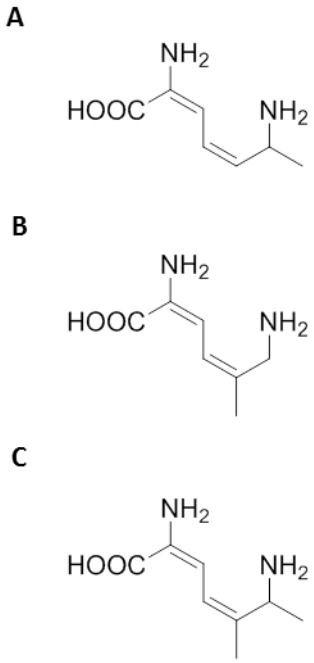


Figure 5-16. Novel amines produced in this chapter using (A) m-toluic acid, (B) p-toluic acid, and (C) 3,4-dimethylbenzoic acid as substrates for the engineered pathway in pQR1062

The novel amines presented here in Figure 5-16, if aminated at the hydroxyl group as shown, will be chiral. Chiral amines are particularly interesting to industry as they are difficult to produce using traditional synthetic chemistry methods in an enantiomerically pure state as discussed in the introduction, section 1.2.2. The structures of these products need to be further confirmed in order to determine any potential uses for these amines.

Chapter 6: Automating biological experiments

Chapter 6. **Automating biological experiments**

6.1 Background

6.1.1 *Streptomyces* and biological products for industry

An example of a species that is already used industrially and could be investigated as an alternative host organism are the *Streptomyces* species of gram positive bacteria. *Streptomyces* are part of the *Actinomyces* family of gram positive bacteria and have been exploited for their antimicrobial properties over the last 100 years. Though these bacteria are used in the manufacture of antibiotics, genetic tools to manipulate and engineer *Streptomyces* growth and product production have not been developed to the extent that tools for *E. coli* have been developed. In many cases the strains of *Streptomyces* that produced desired antibiotics have been developed in house by the pharmaceutical companies that produce them. This means that if tools have been developed they are under strict protection by the company that has developed them and are being kept as trade secrets. However, generally speaking, the development of these strains has been by using random mutagenesis and therefore it is difficult to identify exactly what has caused any positive or negative effects and subsequently then difficult to apply these changes to other species to give the same effect (Bekker et al. 2014). The lack of tools for genetic modification in *Streptomyces* species has meant that research with these bacteria is limited. *Actinomycetes* still have a wealth of unexplored biosynthesis pathways that may contain antimicrobial agents and interesting molecules. To increase our understanding of the biosynthesis pathways and to access these molecules better tools are needed to manipulate the bacteria easily. The aim of this study was to develop a cloning system in *Streptomyces* which could be used to clone the engineered pathway developed in chapter 5 into an example *Streptomyces* host organism and compare the conversion rate with the earlier conversion reported with *P. putida* and *E. coli* .

6.1.2 Cloning in *Streptomyces*

Streptomyces species have a small linear genome and are generally organised with a core in the centre containing genes for essential metabolic pathways and an outer region containing secondary biosynthetic pathways such as for antibiotic synthesis. These outer regions of the genome are not essential to survival and are often easily mutated in nature (Hopwood 2006). Conveniently for us, this also means that genes for biosynthetic pathways are often close to each other or organised in operons.

There are replicative plasmids available for cloning in *Streptomyces*, however the most popular method for cloning in gram positive bacteria is the introduction of DNA to integrate into the genome. There are a number of integration methods that each integrate into the genome at specific sites, thus requiring the recipient to have these sites in their genome. A popular method that was developed is the redirect method which uses flippase recognition target (FRT) sites using flippase, originating in *Saccharomyces cerevisiae* (Gust et al. 2004). Another is the Cre/*LoxP* method that allows integration into *loxP* sites using Cre recombinase (Herrmann et al. 2012). φC31 integration uses attachment sites (attP and attB) sites which are diverse across *Streptomyces* species (Combes et al. 2002). Genomes may also have multiple copies of the attachment/integration sites and therefore the number of insertion copies can be variable and unpredictable. With the aim to produce a robust and reproducible system this are therefore not ideal integration methods to use. There remains a need to for a reliable genome integration system that can replace or insert a single copy of the desired DNA sequence at a precise location within a genome. With this in mind a homologous recombination system was designed to use genome specific DNA sequences.

6.1.3 Engineering robustness and reproducibility in biology

Developing a cloning system for *Streptomyces* species was attempted in parallel to the development of an automated system for performing biological experiments. It is widely recognised that biology is highly variable; there are even examples earlier in this thesis in Chapter 3 and Chapter 5 where the cell growth of *P. putida* was highly variable even when growth was replicated under the exact same conditions. Causes of this variability can be due to experimental discrepancies or slight biological differences. To increase the reliability of an experiment it is repeated as many times as is necessary to reduce error. Conducting high numbers of replicates though only increases chances of error arising from variable factors such as different batches of growth media or variability in laboratory personnel technique or skill. High throughput methods are slowly making their way into biological laboratories, thus enabling scientists to increase the n number of studies. Still, with highly complicated biological experiments this requires excellent focus to reduce mistakes. Bringing engineering principles into biology is a concept that has gained traction to address this issue of variability. Standardising parts is one system that is being used more, particularly in synthetic biology fields; an example of this is Biobricks (Knight 2003). Another area is removing human and experimental error by introducing digital and automated systems for designing and conducting experiments. Using automated systems reduces human error as the computer will only follow inputs, so if a 96 well plate is being used the ingredients for each well will match exactly what has been recorded. Automated systems also reduce experimental error by ensuring consistent pipetting and measurements which can often vary slightly when performed by hand.

Automated systems also accurately record all details of the experiment. This ensures that the experiment is reproducible and the correct information is passed on or published for subsequent research. Aside from the accurate recording of information, automated systems also aid in reproducibility due to

their consistent performance. Due to this reduced variability in performance experimenters can also be confident that any differences in results are due to true differences between samples and not due to mistakes or experimental variables.

There are a number of automated systems for biological experiments available. In this study, the operating system, Antha (Synthace Ltd.) and the Pipetmax® (Gilson) was used to develop an end to end molecular biology experiment beginning from the design of primers to transformation of DNA into competent cells.

6.1.4 Automated process design using Antha

The objective of this study is to create an end to end biological experiment from amplification of genome fragments to cloning and validation. There are a number of automated liquid handling tools available, the most flexible and widely used platform is the Tecan system which allows multiple functions using a single liquid handling platform. Antha is an operating system which can be connected to and coordinate the protocols of multiple pieces of equipment so that an experiment can be planned, designed and executed from beginning to end using a single software platform. The Antha operating system is still under development so this study was designed to continue the development and test in a real world experiment the usability and the performance of Antha. Results of the study would be fed back to developers of Antha at Synthace Ltd.

6.2 Automation of cloning for a *Streptomyces* vector

A cloning experiment to construct a vector containing the homology arms derived from *S. lividans* genome, reporter gene *gusA* and neomycin resistance was designed as described in appendix C. The first step in this process was to PCR amplify the relevant homology arms from the *Streptomyces lividans* genome.

6.2.1 PCR from *Streptomyces lividans* genome

Primers for PCR of homology arms were designed manually. *Streptomyces lividans* was prepared for genome PCR by harvesting spores from agar plates grown for 5 days and freeze thawing in 50 ul dimethyl sulfoxide (DMSO) at – 80°C every 2 hours for 3 freeze thaw cycles. The PCR reaction mixture was then prepared using the Gilson Pipetmax™ controlled by the Antha software. Polymerase used was the Q5 polymerase (NEB). The PCR reaction mix was prepared as per manufacturer's instructions, 1 µL of freeze thawed *Streptomyces lividans* spores were used per PCR reaction. Heat cycling was performed using the T100™ Thermal Cycler (BioRad).

6.2.2 Running a gel

To confirm success of the genomic PCR an agarose gel was run to check the size of DNA fragments amplified. E-gel™ 96 Gels with Sybr™ Safe DNA gel stain, 2% (Invitrogen) were used. The set up of the pipetmax for loading and running the gel is shown in Figure 6-1 below.

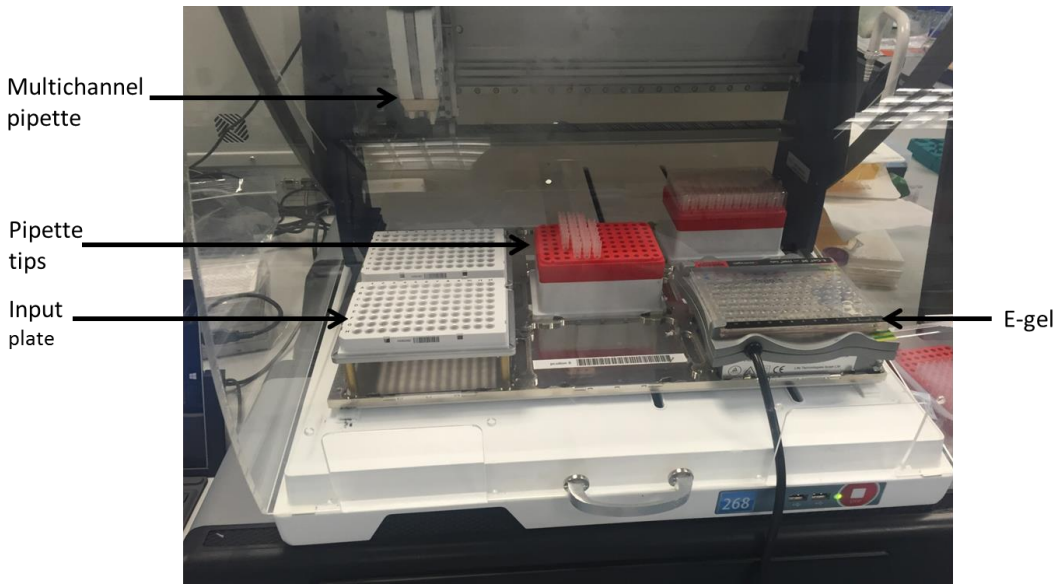


Figure 6-1. Gilson Pipetmax® set up to load and run PCR reactions for validation. The input plate contains completed PCR reactions ready to load onto the E-gel™ 96 Gels with Sybr™ Safe DNA gel stain, 2% (Invitrogen). The E-gel® must always be positioned as shown here due to the power cable exiting the Pipetmax®.

The gel was loaded using the Gilson Pipetmax® and Antha software and visualised using a Dark Reader blue transilluminator (Clare Chemicals). Regions which were amplified, annealing temperatures and corresponding agarose gel lanes are detailed in

One limitation with the Antha software as it is currently is the lack of control over the position of the output wells and therefore the order of loading the PCR reactions in the gel was not intuitive. In addition to this the 96 well E-gels® have staggered wells (as discussed later in section 6.5.2) making the gel more difficult to read.

Table 6-1. Position and length of amplified regions of the *S. lividans* genome and the corresponding lanes on the agarose gels in Figure 6-2 and Figure 6-3. For all amplifications, annealing temperatures used in PCR reactions were 70.0 °C, 69.3 °C, 68.0 °C, 66.1 °C, 63.8 °C, 62.0 °C, 60.7 °C and 60.0 °C. The lanes in agarose gels correspond to the temperatures from left to right for each fragment.

Amplified region of *S. lividans* genome (bp)	Homology arm and size	Agarose gel lanes
3,249,242-3,250,277	Left 1kb	Figure 6-3, lanes 24-32
3,247,378-3,250,277	Left 3kb	Figure 6-3, lanes 15-22
3,244,783-3,247,452	Left 2.5kb	Figure 6-3, lanes 2-9
3,247,480-3,250,277	Left 2.5kb	Figure 6-3, lanes 11-14 and 33-36
3,250,518-3,251,263	Right 1kb	Figure 6-2, lanes 24-32
3,250,518-3,253,099	Right 3kb	Figure 6-2, lanes 15-22
3,250,518-3,252,381	Right 2.5kb	Figure 6-2, lanes 2-9
3,252,316-3,254,433	Right 2.5kb	Figure 6-2, lanes 11-14 and 43-45

Figure 6-2. Gradient PCR results of right hand homology arms amplified from *Streptomyces lividans* genomic DNA run on an E-gel® 96 gels with Sybr® Safe, 2%. There were 8 PCR reactions per pair of PCR primers. The two sets of 2.5 kb homology arms make up the single 5 kb homology arm after vector assembly. There is also a 3 kb and 1 kb homology arm. All PCRs for the 2.5 kb fragments were successful. There were also bands visible for 3 kb fragments at every melting temperature. For 1 kb homology arms bands were visible at higher annealing temperatures of 68.0 °C and 69.3 °C. For all amplifications, annealing temperatures used in PCR reactions were 70.0 °C, 69.3 °C, 68.0 °C, 66.1 °C, 63.8 °C, 62.0 °C, 60.7 °C and 60.0 °C. The lanes in agarose gels correspond to the temperatures from left to right for each fragment.

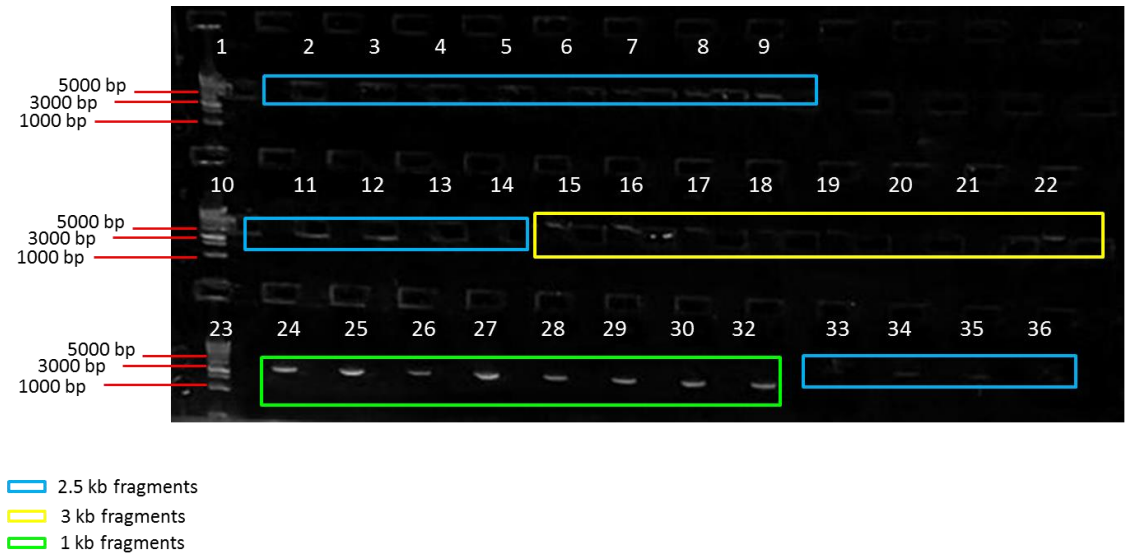


Figure 6-3. Gradient PCR results of left hand homology arms amplified from *Streptomyces lividans* genomic DNA run on an E-gel® 96 gels with Sybr® Safe, 2%. There were 8 PCR reactions per pair of PCR primers. The two sets of 2.5 kb homology arms would have made up the single 5 kb homology arm after vector assembly however no bands were observed for one of the 2.5 kb arms, in the middle row and only faint bands in the top and bottom rows indicating unsuccessful PCRs. There is also a 3 kb and 1 kb homology arm. PCRs to extract the 1 kb homology arm from the genome appear particularly successful as clear bands were visible. Faint bands were visible for the 3 kb fragments at all temperatures. For all amplifications, annealing temperatures used in PCR reactions were 70.0 °C, 69.3 °C, 68.0 °C, 66.1 °C, 63.8 °C, 62.0 °C, 60.7 °C and 60.0 °C. The lanes in agarose gels correspond to the temperatures from left to right for each fragment.

DNA markers were smeared on both agarose gels (Figure 6-2 and Figure 6-3). This could be due to the fast running of the gels at a high current which are run in 12 minutes. This results in poor resolution between DNA fragments as is seen particularly in the DNA marker of Figure 6-2. Amplification of both sets of 3 kb homology arms and for all 1 kb homology arms. One of the 2.5 kb fragments for the 5 kb left homology arm amplification from the genome was unsuccessful therefore going forward it was not possible to construct a vector with a 5 kb homology arm each side of the neomycin resistance gene that was to be inserted. The PCR of the right hand 5 kb homology arm was successful and still usable. Therefore the study was modified to also create vectors with unbalanced homology arm lengths. The remainder of each PCR

reaction not loaded onto the gel was purified using a Qiagen PCR cleanup kit to prepare the DNA for vector assembly.

6.2.3 Vector assembly

Once all DNA elements had been confirmed the vector could be assembled using the SapI restriction enzymes and T4 DNA ligase. The reaction was set up using the Gilson Pipetmax™ and Antha software as follows:

1 µL of each part (50-100 ng)
2 µL Cutsmart buffer
1 µL SapI
1 µL T4 DNA ligase (added after heat denaturation)
Deionized water up to 10 µL

The mixture, except for the T4 DNA ligase, once set up by the Gilson Pipetmax™ was incubated at 37 °C for 1 hour before transformation. The reaction was stopped by heat inactivation at 65 °C for 20 minutes. The T4 DNA ligase was then added and the mixture was incubated at room temperature for 1 hour. Transformation was performed manually with NEB 10beta chemically competent cells as per manufacturer's instructions. Cells were plated on selective LB agar containing 50 µg/mL apramycin and incubated at 37 °C overnight. Colonies were observed as shown in Table 6-2.

6.3 Validation and comparison with manual cloning methods

6.3.1 Comparison with manual cloning

The process was also completed manually using identical parameters for reactions. Table 6-2 shows the number of colonies observed after transformation of both the automated and manual experiments.

Table 6-2. Number of colonies observed after transformation of both automated and manual cloning experiments

Label	Right homology arm length (kb)	Left homology arm length (kb)	Number of colonies from automation study	Number of colonies by manual cloning
A1	3	3	25	8
B1	1	1	100	50
C1	3	1	100	7
D1	5	1	50	5
E1	1	3	5	3
F1	5	3	4	6

Transformations were performed identically and the volumes of cells plated were equal for all samples. There were more colonies observed on plates from the automated cloning process, except for F1, compared with manual cloning. The reasons for this depend on whether the cloning was successful or not and therefore validation of the colonies was needed before any conclusions could be made.

6.3.2 Validation by colony PCR

To validate whether the cloning had been successful colony PCR was performed to verify the sequence of the plasmid. A protocol was designed using Antha to pick colonies for colony PCR. To pick a colony using Antha and the Gilson Pipetmax® the location of the colony must be precisely specified. This was overcome by creating an agarose plate in the size of a standard rectangular 1536 well plate (Figure 6-4) and selecting what would be the well location of the desired colony. A range of colonies from each transformation from the automated process were streaked onto a section of this agar plate and labelled A1 to F1 for ease of tracking. To work out what well location each desired colony was positioned at the agar plate was laid on top of a 1536 well plate as shown in Figure 6-5. Well positions that

corresponded to the desired colony were noted and input into the Antha protocol for colony picking.

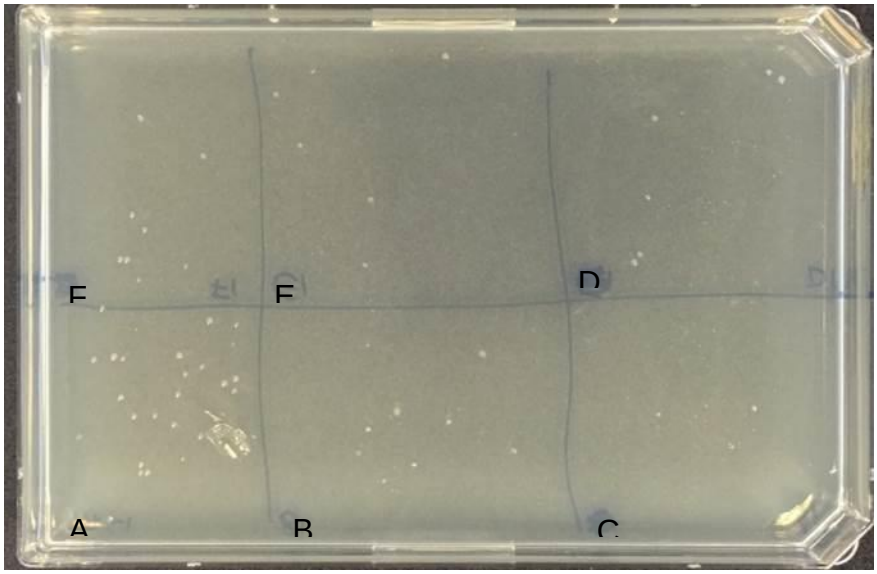


Figure 6-4. Colonies were restreaked onto an agar plate with the dimension of a standard 1536 well plate in preparation for picking by the Gilson Pipetmax®. Sections were labelled A1-F1.

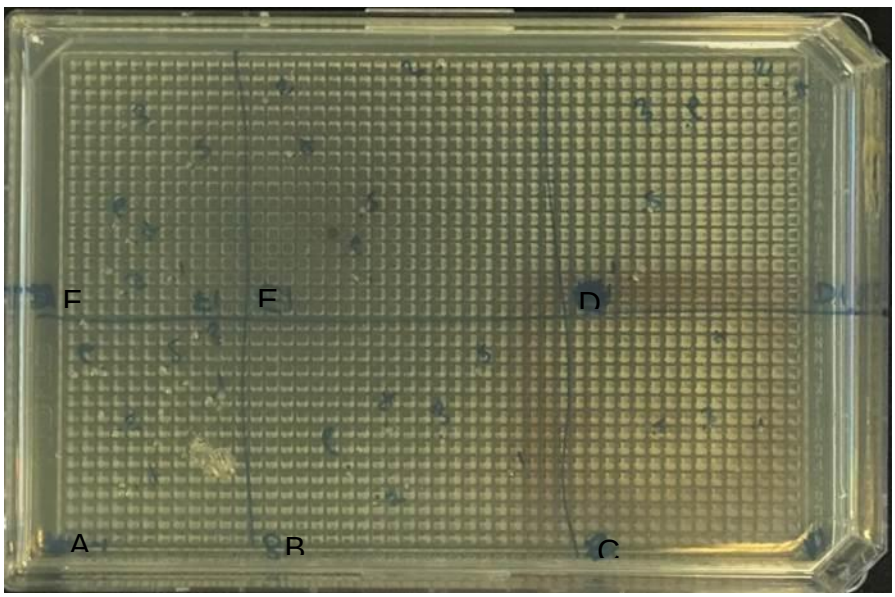


Figure 6-5. The agar plate was laid over a 1536 well plate so that the well number that corresponded to each desired colony to be picked could be noted.

Six colonies per transformation were picked using the Gilson Pipetmax® controlled by Antha OS and inoculated into 50ul water in preparation for colony PCR.

Primers for PCR were designed so that a region where two parts joined, one part being a homology arm, would be amplified using the Antha software. The method for designing primers was set up by defining parameters as follows:

- length of primer between 20 – 30 base pairs;
- GC content between 40 and 70% (higher GC allowance was selected due to the naturally high GC content of *Streptomyces* species genomes);
- the starting position in the genome to begin searching for suitable forward primers;
- the starting position in the reverse complement of the genome to begin searching for reverse primers;

Colonies were prepared for colony PCR by resuspension in 50 µL deionized water and heated to 50 °C for 10 minutes. 1 µL of the denatured colony was used for the PCR reaction which was set up as previously described. After thermal cycling the PCR reaction mixture was observed using manual gel electrophoresis on a 1% agarose gel. Manual gel electrophoresis was used on this occasion due to time limitations. Modifications to the Antha gel running protocol were being made in parallel to correct the gel loading onto the E-gels so that the order of samples was more logical and gel analysis was more user friendly.

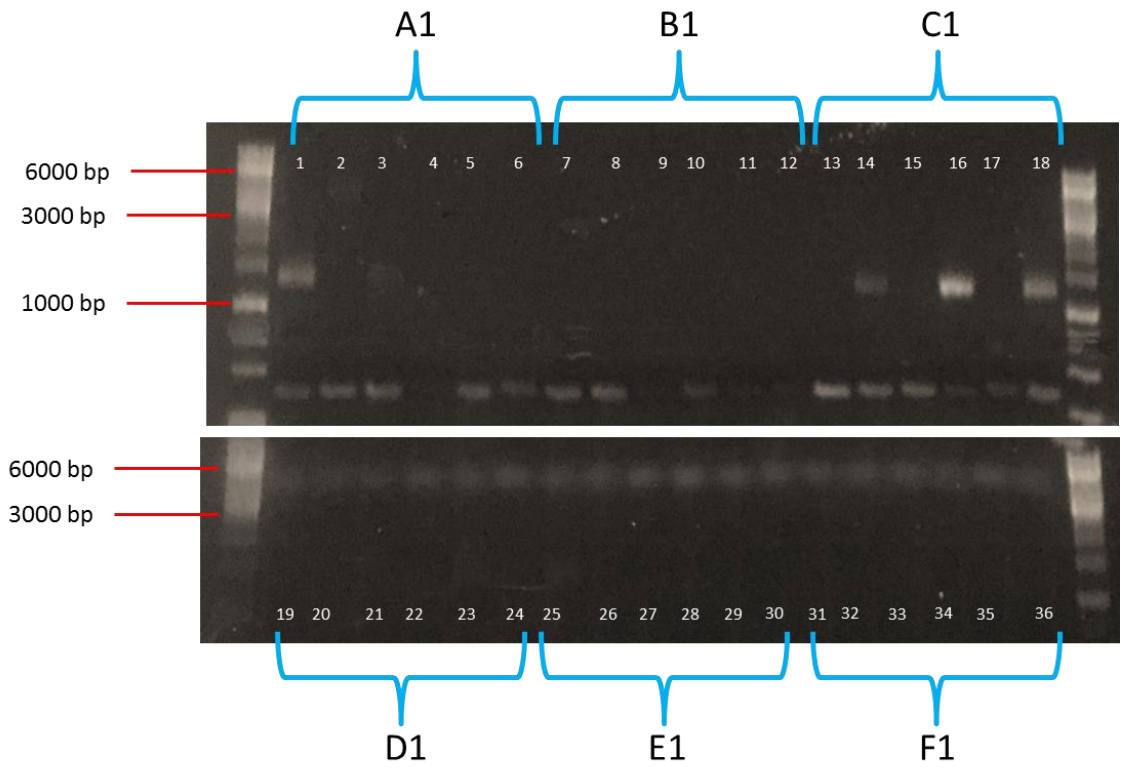


Figure 6-6. Agarose gel of colony PCR reactions to validate cloning.
Six colonies from each transformation were picked and inoculated into water using Antha OS and the Gilson Pipetmax®. PCR was performed using primers that would produce a fragment containing a joint between two parts of the vector assembly. From this gel it appears as though PCR reactions run in lanes 1, 14, 16 and 18 were successful. Bands observed in lanes 19-36 are faint and with poor resolution so it is unclear whether these represent positive PCR reactions. Bands at very low molecular weights in lanes 1-18 are assumed to be primers. Labels A1-F1 indicate which transformation the colony originated from.

From the gel in Figure 6-6 it would appear that colonies tested in lanes 1, 14, 16 and 18 may be positive clones. These lanes corresponded to colonies from sections A1 (lane 1) and C1 (lanes 14, 16 and 18). These colonies were sent for sequencing for sequence confirmation.

Colonies that showed positive results during colony PCR were selected to send for sequence verification. Unfortunately sequencing results confirmed that both the manual and automated cloning had been unsuccessful. The

plasmid sequences that were observed showed artifacts of cloning where plasmid contains fragments of homology arms and the vector backbone allowing for replication in *E. coli.* The full assembled vector was not observed.

6.4 Discussion

6.4.1 Cloning was unsuccessful using both automated and manual methods

Though assembly was not successful, this appeared to be a biological limitation and not a limitation of the automated process. The main aims of this study were to design and perform a biological experiment using the Antha software created by Synthace Ltd. The creation of a *S. lividans* vector was a secondary objective. The main objective, therefore, was achieved as the experiment was executed from beginning to end almost fully automated. Protocols to set up a PCR reaction, load and run a gel and design primers were successfully developed with Antha and executed using the Gilson Pipetmax™ when liquid handling was required. The assembly of the vector appears to be a biological issue. *Streptomyces* genomes have particularly high GC content and therefore cloning is often problematic. The homology arms chosen for this assembly were no exception to this, however the PCR from the genome was expected to be the limiting step but this was fairly successful. Possible reasons for the cloning artifacts observed could be over digestion with SapI or possible SapI recognition sites in the homology arms that were unknown due to a mutation in the *S. lividans* cells used compared with the genome sequence used from Genbank.

There were colonies observed from both automated and manual transformations despite the cloning being unsuccessful. As mentioned above, one reason for this was the appearance of cloning artifacts meaning

that vectors may have contained the apramycin resistance gene but the vector was not assembled correctly. There were much higher numbers of false positive colonies from the automated process and this highlights a potential limitation of the process. This increased number of false positives may arise from cross contamination as the pipette is moving to different positions above the open well plates. Therefore there is the potential for small droplets to end up in wells that they should not be in. This could be minimised by using foil plate seals which the pipette tip in theory should pierce with ease when dispensing liquid but would only leave a small opening to each well.

6.4.2 Challenges in usability with Antha

A number of challenges were overcome when using the Antha software, particularly relating to the Gilson Pipetmax. Difficulties arose when determining the exact height and positioning of the 96 well plates. During the PCR protocol 96 well plates were kept on Eppendorf PCR cooler plates, this altered the height of the plates and needed to be updated on the programming of the protocol. These PCR cooler plates were frozen at -20 °C for 24 hours before use. If the plates were not frozen exactly horizontally they would freeze unevenly resulting in a slight height difference at one end of the plate compared to the other. This meant that pipetting by the Pipetmax™ occasionally resulted in tips reaching the bottom of the plate and bending or breaking, occasionally getting stuck in the plate or the tips not being ejected in to the waste properly.

Challenges also arose with gel loading. The current for the gel electrophoresis is powered through the e-gel platform. It was possible to install this in the Pipetmax™, however it could only be positioned in the bottom right hand corner cell of the Pipetmax™ due to the position of the power cable attaching to the e-gel platform. This meant that this needed to be specified in the protocol. The e-gels used have a staggered well format

and therefore the default system for the computer was to load alternate rows first (i.e. odd numbered rows) followed by the even numbered rows (Figure 6-7). This meant after running the gel it became more complicated to identify samples. This was corrected so that gels were loaded in a more user friendly order, this issue was resolved but further limitations to the E-gel system in general are discussed below.

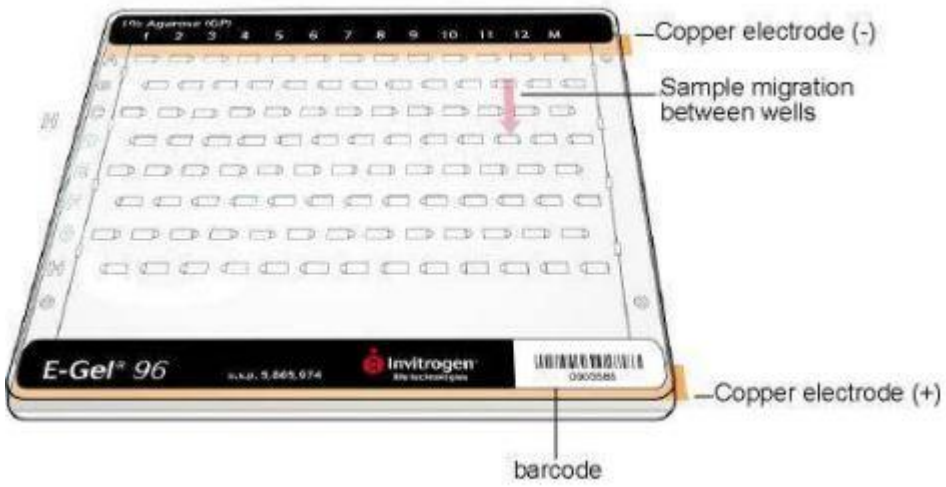


Figure 6-7. The E-gel 96. Taken from the Life technologies E-gel technical guide. The default system for loading the E-gel 96 was to load rows A, C, E… and then B, D, F…. This was corrected for more user friendly identification of samples so that the gel was loaded A, B, C as expected.

6.4.3 Limitations of the E-gels

The 96 well e-gels are not user friendly themselves as the way the lanes are staggered means that samples run between the wells in the rows below, making identification of samples difficult. In future the 48 well format would be much more user friendly. The limitation with this would be that with the Pipetmax™ it is only possible to fit the e-gel in one location due to the power cable so only one gel could be run per protocol. To overcome this, it may be possible to use a different liquid handler that could accommodate multiple power cables exiting the cabinet.

A limitation that is specific to the E-gels and cannot be overcome using the Antha software was the short running length meaning that small variations in DNA movement through the gel results in what appears to be large differences in DNA size. This is a resolution issue and an example of this can be observed in Figure 6-3. The 1 kb fragments run on this gel appear to line

up with the 3 kb marker on the ladder on the left but gradually run further down the gel to line up with the 1 kb marker as we move across the lanes to the right. The marker in a number of cases is not clearly resolved which may be overcome by using a marker with fewer molecular weight bands. To overcome the difference in running of the samples, it would be necessary to understand why this is happening across the gel to be able to reduce the effect.

Chapter 7: Future work

Chapter 7. **Future work**

7.1 Investigating more *P. putida* strains

In Chapter 3, many *P. putida* strains did not transform using the electroporation method used. In addition to this, there is a wider collection of strains available in the Ward lab, which were not tested due to time constraints and a lack of high throughput methods. For a more complete study, a wider range of *P. putida* strains should be transformed with pQR1050 and their conversion of benzoate to 2-HMSA measured.

7.2 Full identification of the novel molecule produced after transaminase conversion of 2-HMSA

The novel molecule produced after transaminase conversion of 2-HMSA was not fully confirmed. To confirm the structure of this molecule, NMR would be performed. The limitation in this study that was not overcome was the purity and the low concentration of the product. To gain sample that would be suitable for NMR analysis, the reaction should be conducted on a larger scale, for example using a 5 L bioreactor rather than in a shake flask experiment. This would require some additional method development as the methods used in chapter 3 are unlikely to transfer to a bioreactor based experiment. The amine product would also need to be isolated which should be performed using preparative HPLC.

7.3 Properties and potential uses of the amine product

Once purified, the amine can be further investigated. One use of the amine described in chapter 4 is as a precursor to picolinic acid. Methods to convert the amine to picolinic acid could be investigated, either using techniques as described in (Riegert et al. 1998) or by investigating further enzymes that may convert the amine further.

7.4 Investigating the engineered pathway in *Streptomyces*

Due to time constraints the cloning in *S.* lividans was not completed. This should be continued to produce a vector that can be easily cloned into *Streptomyces* species and express the TOL pathway truncate. The production of 2-HMSA could be compared with the production in *E. coli* and in *P. putida* strains. Furthermore the engineered pathway could also be cloned into *Streptomyces* to produce the novel molecule and investigate the industrial feasibility of cloning an engineered pathway into *Streptomyces* species.

Chapter 8: Conclusion

Chapter 8. **Conclusion**

8.1 Restatement of the aims

The aims of the research as set out in the introduction, chapter 1, were as follows:

1. As a proof of principle, to show the creation of a novel amine using an engineered metabolic pathway and a transaminase from the UCL toolbox of transaminases
2. To explore an engineered pathway in one or more industrially relevant alternative host organisms and compare the activity of the pathway to the activity in *E. coli.*
3. To design and conduct a biological experiment using the Antha software with relevant automation technology.

The remainder of this chapter will discuss the extent to which these aims were met and future work to leading on from these aims.

The aim was to show as a proof of principle that novel molecules can be produced by engineering metabolic pathways. A novel amine was produced by engineering a fragment of the TOL *meta*cleavage pathway and a transaminase from the UCL toolbox of transaminases. The amine was verified using mass spectrometry but it was not isolated and the absolute amount that was produced was not calculated. Therefore, thought the first aim was met in essence there is still work to be done to validate the process and confirm the conversion rates. This validation would be essential to make any judgements on the feasibility of this method of production for industry, which is the rationale behind this research project and the applicableness of this research to the real world. In addition to this structural work to confirm the conversion of the products predicted in chapter 5 starting with various

methylbenzoates as the starting material is something that also remains to be completed.

With regards to the second aim, which is the investigation of alternative host organisms, the objective was partially fulfilled. *P. putida* was investigated successfully and compared to *E. coli.* It was found that particular strains showed similar activity in the conversion of 2-HMSA by the truncate of the TOL *meta*cleavage pathway. There are a number of limitations to this investigation which were discussed in Chapter 3. The main limitation being that only one strain of *E. coli* was used as a comparison and may not be the ideal strain. Another limitation being that only LB media was used when better results may have been obtained using a minimal media optimised for each strain specifically. However with regards to the aim of this investigation, the main objective was met in the respect that it was shown that *P. putida* is a feasible alternative option for industrial research in bacteria due to similar enzyme expression and activity and similar growth times compared to the current standard that is *E. coli.* A shortcoming of this research project with respect to the second aim is the failure to express an engineered pathway in *Streptomyces lividans.* With more time, a different cloning strategy should be adopted to obtain positive clones for expression in *Streptmyces lividans.*

The third aim, regarding the automation of a biological experiment was also partially successful. A biological experiment was conducted, though the process was not fully automated due to limitations of the liquid handling machines and lack of other automated equipment available such as 4°C incubation for the transformation. In addition to this the experiment was unsuccessful, which appears to be due to the cloning strategy rather than the automation aspect. It would be beneficial for future work to include a positive control where the every step of the experiment is known to work well manually and therefore the automation process can be truly validated.

8.2 Final remarks

All in all, the main objectives of this project were met, though with more time the research could be made more robust. With the increase in automation, maybe this would be possible within a shorter timeframe in the future.

References

Abril, M. a et al., 1989. Regulator and enzyme specificities of the TOL plasmid-encoded upper pathway for degradation of aromatic hydrocarbons and expansion of the substrate range of the pathway. *Journal of bacteriology*, 171(12), pp.6782–6790.

Ager, D.J. et al., 2001. Novel biosynthetic routes to non-proteinogenic amino acids as chiral pharmaceutical intermediates. *Journal of Molecular Catalysis B: Enzymatic*, 11(4–6), pp.199–205.

Andrianantoandro, E. et al., 2006. Synthetic biology: new engineering rules for an emerging discipline. *Molecular systems biology*, 2, p.2006.0028.

Asano, Y., Yamamoto, Y. & Yamada, H., 2014. Catechol 2,3-Dioxygenase-catalyzed Synthesis of Picolinic Acids from Catechols. *Bioscience, Biotechnology and Biochemistry*, 58(11), pp.2054–2056.

Ávalos, M. et al., 2008. A Family of Hydrogels Based on Ureido-Linked Aminopolyol-Derived Amphiphiles and Bolaamphiphiles: Synthesis, Gelation under Thermal and Sonochemical Stimuli, and Mesomorphic Characterization. *Chemistry - A European Journal*, 14(18), pp.5656–5669.

Baeshen, M.N. et al., 2015. Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives. *Journal of Microbiology and Biotechnology*, 25(7), pp.953–962.

Bagdasarian, M. et al., 1981. Specific-purpose plasmid cloning vectors II. Broad host range, high copy number, RSFlOlO-derived gene cloning in Pseudomonas. *Gene*, 16, pp.237–247.

Bagdasarian, M.M. et al., 1983. Activity of the hybrid trp-lac (tac) promoter of Escherichia coli in Pseudomonas putida. Construction of broad-host-range, controlled-expression vectors. *Gene*, 26(2–3), pp.273–282.

Bagdasarian, M.M. et al., 1981. Specific-purpose plasmid cloning vectors II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*. *Gene*, 16(1–3), pp.237–247.

Baker, M., 2016. 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), pp.452–454.

Ballard, D.G.H. et al., 1983. A biotech route to polyphenylene. *Journal of the Chemical Society, Chemical Communications*, (17), p.954.

Baltz, R.H., 2007. Antimicrobials from Actinomycetes: Back to the Future. *Microbiol.*, 2(3), pp.125–131.

Barabási, A.-L. & Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2), pp.101–113.

Bayly, R.C., Dagley, S. & Gibson, D.T., 1966. The metabolism of cresols by species of Pseudomonas. *The Biochemical journal*, 101(2), pp.293–301.

Beites, T. & Mendes, M. V, 2015. Chassis optimization as a cornerstone for the application of synthetic biology based strategies in microbial

secondary metabolism. *Frontiers in microbiology*, 6, p.906.
Bekker, V. et al., 2014. Tools for metabolic engineering in Streptomyces. *Bioengineered*, 5(5), pp.293–9.
Benner, S.A. & Sismour, A.M., 2005. Synthetic biology. *Nature reviews. Genetics*, 6, pp.533–543.
Berłowska, J. et al., 2016. Simultaneous Saccharification and Fermentation of Sugar Beet Pulp for Efficient Bioethanol Production. *BioMed Research International*, 2016, pp.1–10.
Bilal, M. et al., 2018. Metabolic engineering and enzyme-mediated processing: A biotechnological venture towards biofuel production – A review. *Renewable and Sustainable Energy Reviews*, 82, pp.436–447.
de Boer, H.A., Comstock, L.J. & Vasser, M., 1983. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences*, 80(1), pp.21–25.
Bookser, B.C. & Zhu, S., 2001. Solid phase extraction purification of carboxylic acid products from 96-well format solution phase synthesis with DOWEX 1 x 8-400 formate anion exchange resin. *Journal of Combinatorial Chemistry*, 3(2), pp.205–215.
Breuer, M. et al., 2004. Industrial Methods for the Production of Optically Active Intermediates. *Angewandte Chemie International Edition*, 43(7), pp.788–824.
Brosius, J., Erfle, M. & Storella, J., 1985. Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. *The Journal of biological chemistry*, 260(6), pp.3539–41.
Cai, W. & Zhang, W., 2018. Engineering modular polyketide synthases for production of biofuels and industrial chemicals. *Current Opinion in Biotechnology*, 50, pp.32–38.
Caspi, R. et al., 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*, 44(D1), pp.D471–D480.
Chater, K.F., 2006. Streptomyces inside-out: a new perspective on the bacteria that provide us with antibiotics. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 361(1469), pp.761–8.
Choi, K.H., Kumar, A. & Schweizer, H.P., 2006. A 10-min method for preparation of highly electrocompetent Pseudomonas aeruginosa cells: Application for DNA fragment transfer between chromosomes and plasmid transformation. *Journal of Microbiological Methods*, 64(3), pp.391–397.
Clark, K. et al., 2016. GenBank. *Nucleic Acids Research*, 44(D1), pp.D67–D72.
ClinCalc, 2019. The top 300 drugs of 2019. Available at: https://clincalc.com/DrugStats/Top300Drugs.aspx.
Cobb, R.E., Wang, Y. & Zhao, H., 2015. High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. *ACS synthetic biology*, 4(6), pp.723–8.
Combes, P. et al., 2002. The Streptomyces Genome Contains Multiple

Pseudo-attB Sites for the C31-Encoded Site-Specific Recombination System. *Journal of Bacteriology*, 184(20), pp.5746–5752.

Costa, I.C.R., de Souza, R.O.M.A. & Bornscheuer, U.T., 2017. Asymmetric synthesis of serinol-monoesters catalyzed by amine transaminases. *Tetrahedron: Asymmetry*, 28(9), pp.1183–1187.

Cozzetto, D. & Jones, D.T., 2017. Computational Methods for Annotation Transfers from Sequence. In Humana Press, New York, NY, pp. 55–67.

D. Patil, M. et al., 2018. Recent Advances in ω-Transaminase-Mediated Biocatalysis for the Enantioselective Synthesis of Chiral Amines. *Catalysts*, 8(7), p.254.

Dammeyer, T., Timmis, K.N. & Tinnefeld, P., 2013. Broad host range vectors for expression of proteins with (Twin-) Strep-tag, His-tag and engineered, export optimized yellow fluorescent protein. *Microbial cell factories*, 12(1), p.49.

Denome, S.A. et al., 1993. Metabolism of dibenzothiophene and naphthalene in Pseudomonas strains: complete DNA sequence of an upper naphthalene catabolic pathway. *J. Bacteriol.*, 175(21), pp.6890–6901.

Deszcz, D. et al., 2015. Single active-site mutants are sufficient to enhance serine:pyruvate α-transaminase activity in an ω-transaminase. *The FEBS journal*, 282(13), pp.2512–26.

Durante-Rodríguez, G., de Lorenzo, V. & Martínez-García, E., 2014. The Standard European Vector Architecture (SEVA) plasmid toolkit. *Methods in molecular biology (Clifton, N.J.)*, 1149, pp.469–78.

EMBL, 2018. Protein expression, E. coli; Improving protein solubility. Available at: https://www.embl.de/pepcore/pepcore_services/protein_expression/ecoli/improving_protein_solubility/ [Accessed June 3, 2018].

Endy, D., 2005. Foundations for engineering biology. *Nature*, 438(7067), pp.449–53.

Farny, N.G., 2018. A Vision for Teaching the Values of Synthetic Biology. *Trends in Biotechnology*, 36(11), pp.1097–1100.

Feist, C.F. & Hegeman, G.D., 1969. Phenol and Benzoate Metabolism by Pseudomonas putida: Regulation of Tangential Pathways. *J. Bacteriol.*, 100(2), pp.869–877.

Ferber, D., 2004. Microbes Made to Order. *Science*, 303, pp.158–161.

Finn, R.D. et al., 2014. Pfam: the protein families database. *Nucleic Acids Research*, 42(D1), pp.D222–D230.

Friello, D.A. et al., 1976. XYL, a Nonconjugative Xylene-Degradative Plasmid in Pseudomonas Pxy. *Journal of Bacteriology*, 127(3), pp.1217–1224.

Fürste, J.P. et al., 1986. Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP expression vector. *Gene*, 48(1), pp.119–31.

Gawin, A., Valla, S. & Brautaset, T., 2017. The XylS/Pm regulator/promoter system and its use in fundamental studies of bacterial gene expression, recombinant protein production and metabolic engineering. *Microbial biotechnology*, 10(4), pp.702–718.

Ghislieri, D. & Turner, N.J., 2013. Biocatalytic Approaches to the Synthesis of

Enantiomerically Pure Chiral Amines. *Topics in Catalysis*, 57(5), pp.284–300.

Gómez, R. V. & Varela, O., 2009. Synthesis of Polyhydroxy [*n*]-Polyurethanes Derived from a Carbohydrate Precursor. *Macromolecules*, 42(21), pp.8112–8117.

Gomm, A. & O'Reilly, E., 2018. Transaminases for chiral amine synthesis. *Current Opinion in Chemical Biology*, 43, pp.106–112.

Grand View Research Inc, 2017. Amino Acids Market Size; Share 2012-2022 | Global Industry Report. *Grand View Research*, pp.1–273. Available at: https://www.millioninsights.com/industry-reports/amino-acids-market [Accessed September 21, 2017].

Grant, R.S., Coggan, S.E. & Smythe, G.A., 2009. The physiological action of picolinic Acid in the human brain. *International journal of tryptophan research : IJTR*, 2, pp.71–9.

Griffin, M. & Magor, A.M., 1986. Possible Uses of Micro-organisms in the Manufacture of Plastics and Synthetic Fibres. *Biotechnology and Genetic Engineering Reviews*, 4(1), pp.263–290.

Gunsalus, I. et al., 1975. Microbiology - 1974. In Washington D.C.: American Society for Microbiology, pp. 207–212.

Guo, F. & Berglund, P., 2017. Transaminase biocatalysis: optimization and application. *Green Chemistry*, 19(2), pp.333–360.

Gupta, V. et al., 2017. BioBlocks: Programming Protocols in Biology Made Easier. *ACS Synthetic Biology*, 6(7), pp.1230–1232.

Gust, B. et al., 2004. Lambda red-mediated genetic manipulation of antibiotic-producing Streptomyces. *Advances in applied microbiology*, 54, pp.107–28.

Harayama, S., Polissi, A. & Rekik, M., 1991. Divergent evolution of chloroplast-type ferredoxins. *FEBS Letters*, 285(1), pp.85–88.

Harder, P.A. & Kunz, D.A., 1986. Characterization of the OCT plasmid encoding alkane oxidation and mercury resistance in Pseudomonas putida. *Journal of Bacteriology*, 165(2), pp.650–653.

Herrmann, S. et al., 2012. Site-Specific Recombination Strategies for Engineering Actinomycete Genomes. *Applied and Environmental Microbiology*, 78(6), pp.1804–1812.

Hill, G.A. & Robinson, C.W., 1975. Substrate Inhibition Kinetics : Phenol Degradation by Pseudomonas putida. *Biotechnology and Bioengineering*, XVII, pp.1599–1615.

Hobom, B., 1980. Surgery of genes-at the doorstep of synthetic biology. *Medizinische Klinik*, 75(24), pp.14–21.

Höhne, M. et al., 2010. Rational assignment of key motifs for function guides in silico enzyme identification. *Nature chemical biology*, 6(11), pp.807–13.

Holloway, B.W., 1955. Genetic Recombination in Pseudomonas aeruginosa. *Microbiology*, 13(3), pp.572–581.

Hopwood, D.A., 2006. Soil To Genomics: The *Streptomyces* Chromosome. *Annual Review of Genetics*, 40(1), pp.1–23.

Horn, J.M., Harayama, S. & Timmis, K.N., 1991. DNA sequence

determination of the TOL plasmid (pWWO) xylGFJ genes of Pseudomonas putida: implications for the evolution of aromatic catabolism. *Molecular microbiology*, 5(10), pp.2459–74.

Humble, M.S. & Berglund, P., 2011. Biocatalytic Promiscuity. *European Journal of Organic Chemistry*, 2011(19), pp.3391–3401.

Hwang, B.-Y. et al., 2005. Revisit of aminotransferase in the genomic era and its application to biocatalysis. *Journal of Molecular Catalysis B: Enzymatic*, 37(1–6), pp.47–55.

Jackson, G., 1996. *Metabolic engineering of a toluene degradation pathway*. University College London.

Jeandet, P. et al., 2013. Engineering microbial cells for the biosynthesis of natural compounds of pharmaceutical significance. *BioMed research international*, 2013, p.780145.

Jenkins, R.O., Stephens, G.M. & Dalton, H., 1987. Evidence for more than one form of toluene cis -glycol dehydrogenase from Pseudomonas putida NCIB11767. *FEMS Microbiology Letters*, 44(2), pp.209–214.

Jimenez, J.I. et al., 2002. Genomic analysis of the aromatic catabolic pathways from Pseudomonas putida KT2440. *Environmental Microbiology*, 4(12), pp.824–841.

Kanehisa, M. & Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1), pp.27–30.

Kaulmann, U. et al., 2007. Substrate spectrum of ω-transaminase from Chromobacterium violaceum DSM30191 and its potential for biocatalysis. *Enzyme and Microbial Technology*, 41(5), pp.628–637.

Kazlauskas, R.J. & Bornscheuer, U.T., 2012. 7.22 Enzyme Catalytic Promiscuity: Expanding the Catalytic Action of Enzymes to New Reactions. *Comprehensive Chirality*, pp.465–480.

Kieser, T. et al., 2000. *Practical Streptomyces Genetics*,

Kita, H. et al., 2008. Replication of genetic information with self-encoded replicase in liposomes. *Chembiochem : a European journal of chemical biology*, 9(15), pp.2403–10.

Knight, T., 2003. Idempotent Vector Design for Standard Assembly of Biobricks.

Kunz, D. a. & Chapman, P.J., 1981. Catabolism of pseudocumene and 3-ethyltoluene by Pseudomonas putida (arvilla) mt-2: Evidence for new functions of the TOL (pWWO) plasmid. *Journal of Bacteriology*, 146(1), pp.179–191.

Kusumawardhani, H., Hosseini, R. & de Winde, J.H., 2018. Solvent Tolerance in Bacteria: Fulfilling the Promise of the Biotech Era? *Trends in biotechnology*, 36(10), pp.1025–1039.

Lambert, S. et al., 2014. Altered desferrioxamine-mediated iron utilization is a common trait of bald mutants of Streptomyces coelicolor. *Metallomics*, 6(8), p.1390.

Lee, S.H., Lee, S.Y. & Park, B.C., 2005. Cell Surface Display of Lipase in Pseudomonas putida KT2442 Using OprF as an Anchoring Motif and Its Biocatalytic Applications Cell Surface Display of Lipase in Pseudomonas putida KT2442 Using OprF as an Anchoring Motif and Its Biocatalytic

Applications. *Applied and environmental microbiology*, 71(12), pp.8581–8586.

Leuchtenberger, W., Huthmacher, K. & Drauz, K., 2005. Biotechnological production of amino acids and derivatives: current status and prospects. *Applied microbiology and biotechnology*, 69(1), pp.1–8.

Li, P. et al., 2015. An efficient blue-white screening based gene inactivation system for Streptomyces. *Applied Microbiology and Biotechnology*, 99(4), pp.1923–1933.

Lichman, B.R. et al., 2015. One-pot triangular chemoenzymatic cascades for the syntheses of chiral alkaloids from dopamine. *Green Chem.*, 17(2), pp.852–855.

de Lorenzo, V. et al., 1998. Mini-transposons in microbial ecology and environmental biotechnology. *FEMS Microbiology Ecology*, 27, pp.211–224.

De Lorenzo, V., 1994. Designing microbial systems for gene expression in the field. *Trends in biotechnology*, 12(9), pp.365–71.

Luisi, P.L., Ferri, F. & Stano, P., 2006. Approaches to semi-synthetic minimal cells: a review. *Die Naturwissenschaften*, 93(1), pp.1–13.

Luisi, P.L. & Stano, P., 2011. Synthetic biology: minimal cell mimicry. *Nature chemistry*, 3(10), pp.755–6.

Ma, S.K. et al., 2010. A green-by-design biocatalytic process for atorvastatin intermediate. *Green Chem.*, 12(1), pp.81–86.

Magrioti, V., Hadjipavlou-Litina, D. & Constantinou-Kokotou, V., 2003. Synthesis and In vivo anti-inflammatory activity of long-chain 2-amino-alcohols. *Bioorganic & Medicinal Chemistry Letters*, 13(3), pp.375–377.

Malik, M.S., Park, E.-S. & Shin, J.-S., 2012. ω-Transaminase-catalyzed kinetic resolution of chiral amines using l-threonine as an amino acceptor precursor. *Green Chemistry*, 14(8), p.2137.

Mansy, S.S. et al., 2008. Template-directed synthesis of a genetic polymer in a model protocell. *Nature*, 454(7200), pp.122–5.

Mars, A.E. et al., 2001. Bioconversion of limonene to increased concentrations of perillic acid by Pseudomonas putida GS1 in a fed-batch reactor. *Applied microbiology and biotechnology*, 56(1–2), pp.101–7.

Marten, S. & Naguschewski, M., 2011. Determination of 17 AQC derivatized Amino acids in baby food samples. *Knauer*, 9, pp.3–11.

Martínez-García, E. et al., 2014. Pseudomonas 2.0: genetic upgrading of P. putida KT2440 as an enhanced host for heterologous gene expression. *Microbial cell factories*, 13(1), p.159.

Martins Dos Santos, V. a P. et al., 2004. Insights into the genomic basis of niche specificity of Pseudomonas putida KT2440. *Environmental microbiology*, 6(12), pp.1264–86.

McFadden, B.A. & Howes, W. V, 1961. Pseudomonas indigofera. *Journal of bacteriology*, 81(6), pp.858–62.

Medema, M.H., Breitling, R. & Takano, E., 2011. Synthetic biology in Streptomyces bacteria. *Methods in enzymology*, 497, pp.485–502.

Mermod, N. et al., 1986. Vector for regulated expression of cloned genes in a

wide range of gram-negative bacteria. *Journal of bacteriology*, 167(2), pp.447–454.

Mermod, N., Harayama, S. & Timmis, K., 1986. New route to bacterial production of indigo. *Nature Biotechnology*, 4, pp.321–324.

Mi, J. et al., 2014. De novo production of the monoterpenoid geranic acid by metabolically engineered Pseudomonas putida. *Microbial cell factories*, 13(1), p.170.

Mora, J., Hottenstein, C. & DeSilva, B., 2017. Ligand Binding Assays in the Regulated Bioanalytical Laboratory. In Springer, Cham, pp. 177–228.

Muratore, K.E. et al., 2013. Molecular function prediction for a family exhibiting evolutionary tendencies toward substrate specificity swapping: Recurrence of tyrosine aminotransferase activity in the Iα subfamily. *Proteins: Structure, Function, and Bioinformatics*, 81(9), pp.1593–1609.

Murray, K. et al., 1972. The metabolism of benzoate and methylbenzoates via the meta-cleavage pathway by Pseudomonas arvilla mt-2. *European journal of biochemistry*, 310, pp.301–310.

Myronovskyi, M. et al., 2011. Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes. *Applied and environmental microbiology*, 77(15), pp.5370–83.

NEB, 2019. Quality At NEB. Available at: https://international.neb.com/about-neb/quality-at-neb [Accessed March 2, 2019].

Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V. a P., et al., 2002. Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. *Environmental microbiology*, 4(12), pp.799–808.

Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A.P., et al., 2002. Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. *Environmental microbiology*, 4(12), pp.799–808.

Nestl, B.M. et al., 2014. New generation of biocatalysts for organic synthesis. *Angewandte Chemie (International ed. in English)*, 53(12), pp.3070–95.

Ni, Y. et al., 2013. Proteomic analysis of Pseudomonas putida reveals an organic solvent tolerance-related gene mmsB. *PloS one*, 8(2), p.e55858.

Nikel, P.I. et al., 2016. From dirt to industrial applications: Pseudomonas putida as a Synthetic Biology chassis for hosting harsh biochemical reactions. *Current opinion in chemical biology*, 34, pp.20–29.

Nikel, P.I. et al., 2015. The Glycerol-Dependent Metabolic Persistence of Pseudomonas putida KT2440 Reflects the Regulatory Logic of the GlpR Repressor. *mBio*, 6(2), p.e00340-15-.

Nikel, P.I. & de Lorenzo, V., 2014. Robustness of Pseudomonas putida KT2440 as a host for ethanol biosynthesis. *New biotechnology*, 31(6), pp.562–71.

Noireaux, V. & Libchaber, A., 2004. A vesicle bioreactor as a step toward an artificial cell assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), pp.17669–74.

O' Sullivan, L., 2000. *Metabolic engineering of the toluene degradation pathway of Pseudomonas putida mt-2*. University College London.

Pérez-Pantoja, D. et al., 2014. The differential response of the Pben promoter of Pseudomonas putida mt-2 to BenR and XylS prevents metabolic conflicts in m-xylene biodegradation. *Environmental microbiology*.

Pfeifer, B. a et al., 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science (New York, N.Y.)*, 291(5509), pp.1790–2.

Poblete-Castro, I. et al., 2016. Host Organism: Pseudomonas putida. In *Industrial Biotechnology*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, pp. 299–326.

Poblete-Castro, I. et al., 2012. Industrial biotechnology of Pseudomonas putida and related species. *Applied microbiology and biotechnology*, 93(6), pp.2279–90.

Pollard, D.J. & Woodley, J.M., 2007. Biocatalysis for pharmaceutical intermediates: the future is now. *Trends in biotechnology*, 25(2), pp.66–73.

Poux, S. et al., 2014. Expert curation in UniProtKB: a case study on dealing with conflicting and erroneous data. *Database*, 2014(0), pp.bau016-bau016.

Purcell, O. et al., 2013. Towards a whole-cell modeling approach for synthetic biology. *Chaos (Woodbury, N.Y.)*, 23(2), p.025112.

Qu, Y. et al., 2012. Optimization of indigo production by a newly isolated Pseudomonas sp. QM. *Journal of basic microbiology*, 52(6), pp.687–94.

Ramos-Gonzalez, M.-I. et al., 2003. Genetic Engineering of a Highly Solvent-Tolerant Pseudomonas putida Strain for Biotransformation of Toluene to p- Hydroxybenzoate. *Applied and environmental microbiology*, 69(9), pp.5120–5127.

Rehm, B.H. a, 2010. Bacterial polymers: biosynthesis, modifications and applications. *Nature reviews. Microbiology*, 8(8), pp.578–92.

Reineke, W. & Knackmuss, H.-J., 1979. Construction of haloaromatics utilising bacteria. *Nature*, 277(5695), pp.385–386.

Rheinwald, J.G., Chakrabarty, A.M. & Gunsalus, I.C., 1973. A transmissible plasmid controlling camphor oxidation in Pseudomonas putida. *Proceedings of the National Academy of Sciences of the United States of America*, 70(3), pp.885–9.

Richter, N. et al., 2014. Synthesis of pharmaceutically relevant 17-α-amino steroids using an ω-transaminase. *Chemical communications (Cambridge, England)*, 50(46), pp.6098–100.

Richter, N. et al., 2014. ω-Transaminases as catalysts for the amination of sterically demanding cyclic compounds. *Unpublished data*.

Richter, N. et al., 2015. ω-Transaminases for the amination of functionalised cyclic ketones. *Org. Biomol. Chem.*, 13(33), pp.8843–8851.

Riegert, U. et al., 1998. Distal cleavage of 3-chlorocatechol by an extradiol dioxygenase to 3-chloro-2-hydroxymuconic semialdehyde. *Journal of bacteriology*, 180(11), pp.2849–53.

Ro, D.-K. et al., 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440, pp.940–943.

Roselle, C. et al., 2017. Evaluation of a digital dispenser for direct curve dilutions in a vaccine potency assay. *Journal of Immunological Methods*, 442, pp.20–28.

Rückert, C. et al., 2015. Complete genome sequence of Streptomyces lividans TK24. *Journal of biotechnology*, 199, pp.21–2.

Sadowski, M.I., Grant, C. & Fell, T.S., 2016. Harnessing QbD, Programming Languages, and Automation for Reproducible Biology. *Trends in Biotechnology*, 34(3), pp.214–227.

Sala-trepat, J.M. & Evans, W.C., 1971. The meta Cleavage of Catechol by Azotobacter Species. *European journal of biochemistry*, 20, pp.400–413.

Savile, C.K. et al., 2010. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science (New York, N.Y.)*, 329(5989), pp.305–9.

Schmid, A. et al., 2001. Industrial biocatalysis today and tomorrow. *Nature*, 409, pp.258–268.

Schnoes, A.M. et al., 2009. Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies A. Valencia, ed. *PLoS Computational Biology*, 5(12), p.e1000605.

Schrempf, H. & Dyson, P., 2011. Streptomycetes. *Microbial biotechnology*, 4(2), pp.138–40.

Sehl, T. et al., 2012. TTC-based screening assay for ω-transaminases: A rapid method to detect reduction of 2-hydroxy ketones. *Journal of Biotechnology*, 159(3), pp.188–194.

Senior, E., Bull, A.T. & Slater, J.H., 1976. Enzyme evolution in a microbial community growing on the herbicide Dalapon. *Nature*, 263(5577), pp.476–479.

Sheldon, R.A., 2016. Engineering a more sustainable world through catalysis and green chemistry. *Journal of the Royal Society, Interface / the Royal Society*, 13(116), p.20160087-.

Sheridan, R. et al., 1998. Rational engineering of the TOL meta-cleavage pathway. *Biotechnology and Bioengineering*, 58(2–3), pp.240–249.

Shewan, J.M., Hobbs, G. & Hodgkiss, W., 1960. A determinative scheme for the identification of certain genera of gram-negative bacteria, with special reference to the Pseudomonadaceae. *Journal of Applied Bacteriology*, 23(3), pp.379–390.

Shin, J.-S. & Kim, B.-G., 2009. Transaminase-catalyzed asymmetric synthesis of l-2-aminobutyric acid from achiral reactants. *Biotechnology Letters*, 31(10), pp.1595–1599.

Shin, J.S. et al., 2003. Purification, characterization, and molecular cloning of a novel amine:pyruvate transaminase from Vibrio fluvialis JS17. *Applied Microbiology and Biotechnology*, 61(5–6), pp.463–471.

Silva-Rocha, R. et al., 2013. The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic acids research*, 41(Database issue), pp.D666-75.

Simon, O. et al., 2015. Analysis of the molecular response of Pseudomonas putida KT2440 to the next-generation biofuel n-butanol. *Journal of*

proteomics, 122, pp.11–25.
Simon, R.C. et al., 2013. Recent Developments of Cascade Reactions Involving ω-Transaminases.
Speelmans, G., Bijlsma, A. & Eggink, G., 1998. Limonene bioconversion to high concentrations of a single and stable product, perillic acid, by a solvent-resistant Pseudomonas putida strain. *Applied Microbiology and Biotechnology*, 50(5), pp.538–544.
Stanier, R.Y., Palleroni, N.J. & Doudoroff, M., 1966. The Aerobic Pseudomonads a Taxonomic Study. *Journal of General Microbiology*, 43(2), pp.159–271.
Stewart, J.D., 1997. A Chemist's Perspective on the Use of Genetically Engineered Microbes as Reagents for Organic Synthesis. *Biotechnology and Genetic Engineering Reviews*, 14(1), pp.67–144.
Stewart, J.D., 2001. Dehydrogenases and transaminases in asymmetric synthesis. *Current opinion in chemical biology*, 5(2), pp.120–9.
Subrizi, F. et al., 2019. Aminopolyols from Carbohydrates: Amination of Sugars and Sugar-Derived Tetrahydrofurans with Transaminases. *Angewandte Chemie International Edition*.
Takano, H., Hagiwara, K. & Ueda, K., 2015. Fundamental role of cobalamin biosynthesis in the developmental growth of Streptomyces coelicolor A3 (2). *Applied microbiology and biotechnology*, 99(5), pp.2329–37.
Teruhiko Ishikawa, * et al., 2000. Application of the Silicon-Tether Strategy for Controlling the Regioselectivity and Diastereoselectivity of Intramolecular Nitrone Cycloadditions for Aminopolyol Synthesis.
Thoralf Gross et al., 2002. Synthesis of Primary Amines: First Homogeneously Catalyzed Reductive Amination with Ammonia.
Tierrafría, V.H. et al., 2011. Disruption of the siderophore-binding desE receptor gene in Streptomyces coelicolor A3(2) results in impaired growth in spite of multiple iron-siderophore transport systems. *Microbial biotechnology*, 4(2), pp.275–85.
Troeschel, S.C. et al., 2012. Novel broad host range shuttle vectors for expression in Escherichia coli, Bacillus subtilis and Pseudomonas putida. *Journal of Biotechnology*, 161(2), pp.71–79.
Tucker, J.L., 2006. Green Chemistry , a Pharmaceutical Perspective. *Organic process research & development*, 10, pp.315–319.
Turner, N.J., 2018. Editorial overview: Biocatalysis and biotransformation: The golden age of biocatalysis. *Current Opinion in Chemical Biology*, 43, pp.A1–A3.
Vanier, F.O., 1998. *An evaluation of the process engineering options for a multi step biotransformation.* University College London (University of London).
Villegas-Torres, M.F. et al., 2015. Multi-step biocatalytic strategies for chiral amino alcohol synthesis. *Enzyme and Microbial Technology*, 81, pp.23–30.
Villegas Torres, M., 2014. *Design and characterization of novel biocatalytic cascades for chiral amino alcohol synthesis*. UCL (University College London).

Ward, J. & Wohlgemuth, R., 2010. High-Yield Biocatalytic Amination Reactions in Organic Synthesis. *Current Organic Chemistry*, 14(17), pp.1914–1927.

Weber, T. et al., 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, 43(W1), pp.W237–W243.

Weeks, A.M. & Chang, M.C.Y., 2011. Constructing de novo biosynthetic pathways for chemical synthesis inside living cells. *Biochemistry*, 50(24), pp.5404–18.

Wildey, M.J. et al., 2017. High-Throughput Screening. *Annual Reports in Medicinal Chemistry*, 50, pp.149–195.

Williams, P.A. & Murray, K., 1974. Metabolism of benzoate and the methylbenzoates by Pseudomonas putida (arvilla) mt-2: evidence for the existence of a TOL plasmid. *Journal of bacteriology*, 120(1), pp.416–23.

Wohlgemuth, R., 2010. Biocatalysis--key to sustainable industrial chemistry. *Current opinion in biotechnology*, 21(6), pp.713–24.

Wong, C.L. & Dunn, N.W., 1974. Transmissible plasmid coding for the degradation of benzoate and m-toluate in Pseudomonas arvilla mt-2. *Genetical research*, 23(2), pp.227–32.

Woodley, J.M., 2008. New opportunities for biocatalysis: making pharmaceutical processes greener. *Trends in biotechnology*, 26(6), pp.321–7.

Woodley, J.M., Breuer, M. & Mink, D., 2013. A future perspective on the role of industrial biotechnology for chemicals production. *Chemical Engineering Research and Design*, 91(10), pp.2029–2036.

Worsey, M.J. & Williams, P.A., 1975. Metabolism of toluene and xylenes by Pseudomonas Metabolism of Toluene and Xylenes by Pseudomonas (putida (arvilla) mt-2 : Evidence for a new function of the TOL plasmid. *Journal of bacteriology*, 124(1), pp.7–13.

Wu, J. et al., 2011. Toward Improvement of Erythromycin A Production in an IndustrialSaccharopolyspora erythraea Strain via Facilitation of Genetic Manipulation with an ArtificialattB Site for Specific Recombination. *Applied and Environmental Microbiology*, 77, pp.7508–7516.

Yong, X. et al., 2014. Enhanced cadmium resistance and accumulation in Pseudomonas putida KT2440 expressing the phytochelatin synthase gene of Schizosaccharomyces pombe. *Letters in applied microbiology*, 58(3), pp.255–61.

Chapter 8: Appendices

Chapter 9. **Appendices**

9.1 Appendix A - Primers

9.1.1 Cloning for *P. putida* expression

The following primers were used in constructing the broad host range plasmid used in chapters 3, 4 and 5.

Table 9-1 Cloning pQR1050

Primer name	Primer sequence (5'-3')
TOLxyzIteFWD	GGAATCTCTAGAAAGGCCTACCCCTTAGGCTTTATGCAAC
TOLxyzIteREV	GAAAATAAGCTTGAATTCCTCGAGAAACTCGCCGAAGCGCGC

Table 9-2 Cloning pQR1062

Primer name	Primer sequence (5'-3')
ARTAmFWD	ATTAGTCGGTTCGGCTAATTTTGTTTAACTTTAAGAAGG
ArTAmREV	AACAGCTTGCCGCTGTCAGTGGTGGTGGTGGTG
pQR1050FWD	CAGCGGCAAGCTGTTCGACAATGTC
pQR1050REV	GCCGAACCGACTAATTCACCGCTAATG

9.1.2 Cloning for *Streptomyces* expression

The following primers were used in constructing the DNA fragments used for cloning in chapter 6.

Table 9-3 Constructing the vector backbone from pSET152

Primer name	Primer sequence (5'-3')
pSET152-For	GAGCTACTGGTACCGTGCCAGCTGCATTAATGAATCGGCCAACG
pSET152-For-lacZa	GAGCTACTGGTACCCTACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGC
pSET152-Rev	TGCAGGTCGACGGATCTTTTCCGCTGCATAACCCCATGGTCATCGAG
SapI-mut1-For	CGGTTTGCGTATTGGGCACTCTTCCGCTTCCTCGCTCA
sapI-mut1-Rev	TGAGCGAGGAAGCGGAAGAGTGCCCAATACGCAAACCG
SapI-mut2-	CAGGTGGCTCAAGGAGAAGAGTCTTCAGAAGGAAGGTC

For	C
SapI-mut2-Rev	GGACCTTCCTTCTGAAGACTCTTCTCCTTGAGCCACCTG
SapI-mut3-For	TCCGCGAAGTCGCTCGTCTTGATGGAGCGCATGG
SapI-mut3-Rev	CCATGCGCTCCATCAAGACGAGCGACTTCGCGGA
SapI-mut4-For	CTTTTCCTCAATCGCCCTTCGTTCGTCTGGA
SapI-mut4-Rev	TCCAGACGAACGAAGGGCGATTGAGGAAAAG

Table 9-4 Primers for the PCR of homology arms from *Streptomyces lividans*

Primer name	Primer sequence (5'-3')
LHA-P1-For-Lividans-des	GAGCTGCAGCTCTTCTGAAACGAGTCGGTCGCCTACGCGAAGGAACGTCACGCCTT
LHA-P1A-For-Lividans-des	GAGCTGCAGCTCTTCTGAAATCCAGTTCAAGATCGCCGACATGGAGATGAAGG
LHA-P2-Rev-Lividans-des	GAGCTGCAGCTCTTCTTCCTTGATTCGATCAAACGCGTGCCTGTGGGTGACCGCT
LHA-P2A-Rev-Lividans-des	GAGCTGCAGCTCTTCTTCCCGGAGCCCTCATTTTGACTTAGGTGAGCCTAACCTAAG
LHA-P3-For-Lividans-des	GAGCTGCAGCTCTTCTAGGCGAATCAACTTAGGTTAGGCTCACCTAAGTCAAAATGAGGGCT
LHA-P3A-For-Lividans-des	GAGCTGCAGCTCTTCTAGGCAGGCACGCGTTTGATCGAATCAACTTAGGTTAGGC
LHA-P4-Rev-Lividans-des	GAGCTGCAGCTCTTCTCAGTCATACGGCGAACTCCTGGAACGCGATCGACTGCT
LHA-P4A-Rev-Lividans-Des	GAGCTGCAGCTCTTCTCAGCGATCGACTGCTCGACCGGGTAGTACTCGC
RHA-P1-For-Lividans-des	GAGCTGCAGCTCTTCTATAGTTCCAGGAGTTCGCCGTATGAGCCGCTTGAGCA
RHA-P1A-For-Lividans-des	GAGCTGCAGCTCTTCTATACTCGGACGCGAGTACTACCCGGTCGAGCAGTC
RHA-P2-Rev-Lividans-des	GAGCTGCAGCTCTTCTGAGCGTGAAGATCGACAGGAGCTTCTTGTCGTCCGGCACGT
RHA-P2A-Rev-lividans-des	GAGCTGCAGCTCTTCTGAGGCGTAGAAGCTGTGCAGCAGCGGGACGTAGTAGG
RHA-P3-for-Lividans-des	GAGCTGCAGCTCTTCTCTCACGGACGTCTTCGACTGCTTCTTCCGCTTC
RHA-P3A-ForLividans-des	GAGCTGCAGCTCTTCTCTCCACAGCTTCTACGCCTACGACCTGGTGTACATGCCG
RHA-P4-Rev-lividans-des	GAGCTGCAGCTCTTCTAAACCCACGTTCCGTCCTTCTGGATGATTTCC

RHA-P4A-Rev-Lividans-des	GAGCTGCAGCTCTTCTAAAACCCAGTTCCCTGCGGAGCAGACTGACGTTC
LHA-1kb-For-liv-des	GAGCTGCAGCTCTTCTGAACTCAAGGAGCAGGGCCGGCTGTACTCCTTCTAC
RHA-1kb-Rev-liv-des	GAGCTGCAGCTCTTCTAAAGTGGCCCTCGGTCATGCCGGTCTCGAC
LHA-3kb-For-liv-des	GAGCTGCAGCTCTTCTGAAGGAGGAACAACTCCGCGAGCAGGGCTGAG
RHA-3kb-Rev-liv-des	GAGCTGCAGCTCTTCTAAACAGTACTGGGGCCTGGACCGCACCAG

9.2 Appendix B – Data analysis

9.2.1 Cloning results analysis

Cloning results were confirmed by either restriction digestion and DNA agarose gel or by DNA sequencing or both. Cloning in chapter 3 of the TOL xylXYZLTE genes into pMMB67EH to create pQR1050 was confirmed using DNA gel analysis and by sequencing analysis below.

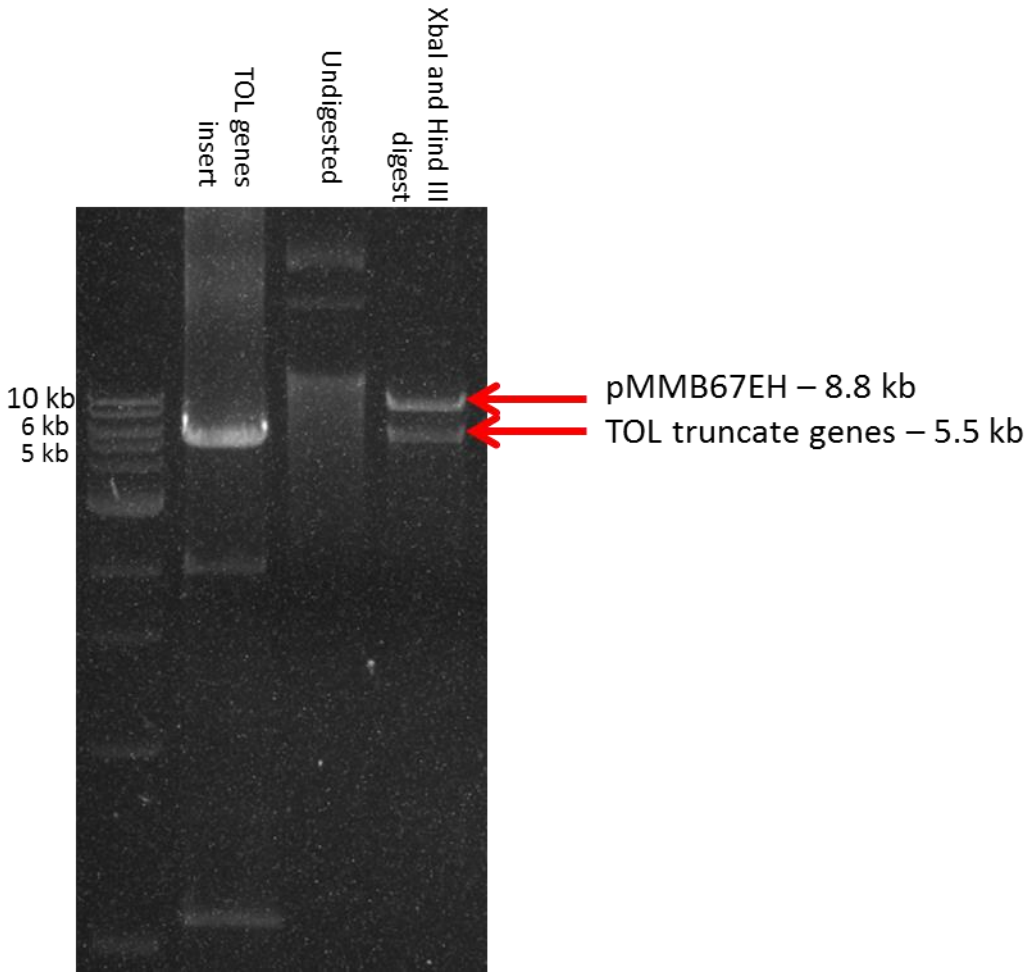


Figure 9-1. DNA agarose gel indicating the successful insertion of the TOL xylXYZLTE genes into pMMB67EH.
The first lane shows the PCR amplified TOL xylXYZLTE genes from pQR226. The second lane shows the undigested plasmid isolated from the transformation of the cloning product described in section 3.2.1. The third lane shows the cloning product after digestion with *Xba*I and *Hin*dIII restriction enzymes. This indicates that the insert was successfully cloned into the pMMB67EH plasmid.

The DNA sequence of the isolated plasmid was also analysed. Primers for sequencing were as described in Table 9-5.

Table 9-5. Primers used for DNA sequencing of pQR1050 and pQR1060

Description	DNA sequence 5' – 3'
pMMB67EH forward	CCGACATCATAACGGTTCTGGC
pMMB67EH reverse	GGCGTTTCACTTCTGAGTTCGGC
TOL xylE forward	GGGGAGATTACAACTACCC
TOL xylX reverse	GCCTTGTTTCCACTCCTAAAGCG

Using these primers DNA sequencing analysis confirmed the sequence of pQR1050 and pQR1060.

9.2.2 MBA/acetophenone HPLC analysis

Percent conversion of transaminases using the MBA/acetophenone assay was calculated by HPLC analysis. Concentration of acetophenone in the reaction mixture were calculated based on the standard curve below in

Figure **9-2**.

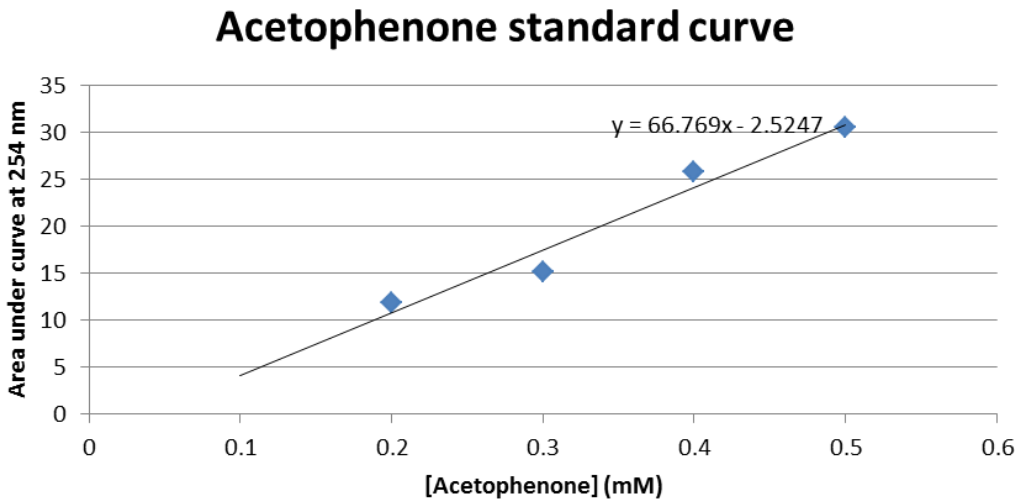


Figure 9-2. Acetophenone standard curve used to calculate concentration of acetophenone in the transaminase MBA assays.
The concentration of acetophenone was ultimately used to calculate percent conversion of 2-HMSA to amine.

9.2.3 LCMS

The following LCMS trace was used to elucidate the structure of the novel amine in section 4.3.3.

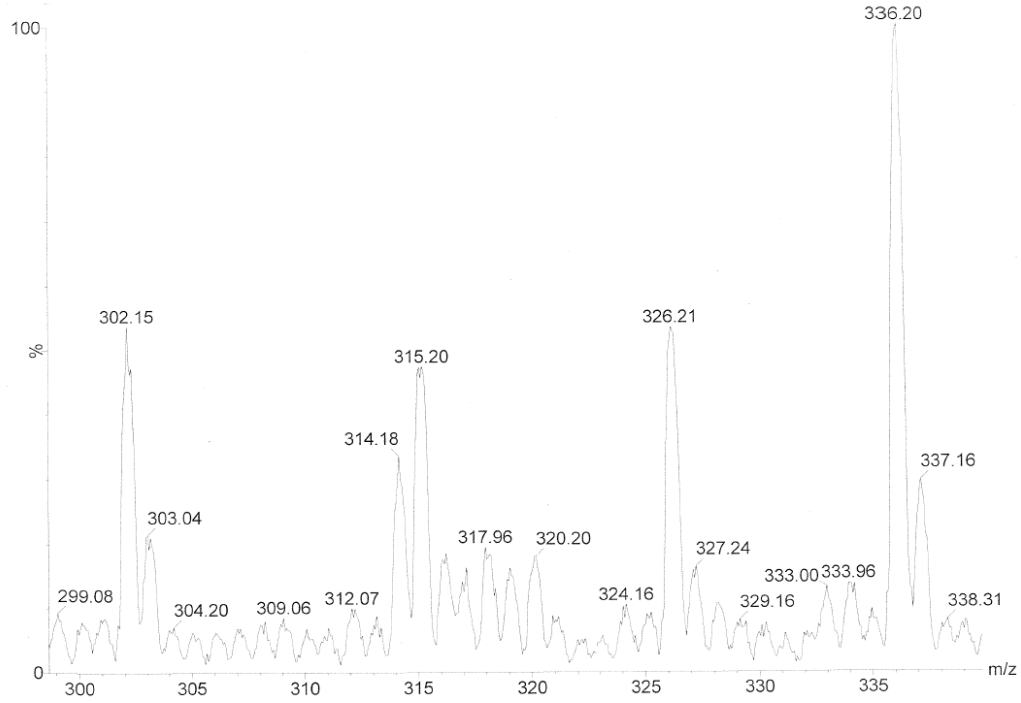


Figure 9-3. LCMS trace resulting from the predicted HPLC purified derivatised amine product.
Peaks at 314.8 and 336.20 were predicted to be the novel amine product without and with a sodium adduct respectively.

9.3 Appendix C – *Streptomyces* cloning experimental design

The following description of the cloning design for *Streptomyces* is applicable to chapter 6.

9.3.1 Experimental design

To begin an investigation into cloning in *Streptomyces* the species was selected based on abundance of previous knowledge and work. The particular strain of *Streptomyces lividans* without any naturally occurring plasmid was chosen so that the organism was simple to work with.

The aim of this study was to design a cloning tool that would be easily transferable between *Actinomyces* species; therefore the integration site of the plasmid must be easily changeable and be flexible enough to integrate in any desired genome sequence. To achieve this homologous recombination was investigated. Positions for integration were chosen based on region of the genome – not too close to the ends due to the frequent modification of the ends in *Streptomyces* species but also not in the centre due to the essential nature of these genes. Modifying the genome must not result in any inhibition of growth or DNA regions necessary for survival in this case, though the method ultimately should allow the modification of these regions if so desired by the investigator. During homologous recombination the desired DNA to be introduced into the bacterium is flanked by sequences matching the desired integration site in the genome. 1 kb flanking have previously been used with a Crispr/Cas9 system in *Streptomyces* species (Cobb et al. 2015); therefore this was the smallest size of recombination arm designed in order to maximise chances of success. However larger recombination sites not only results in a larger plasmid but also increases chances of disrupting essential DNA sequences in the genome.

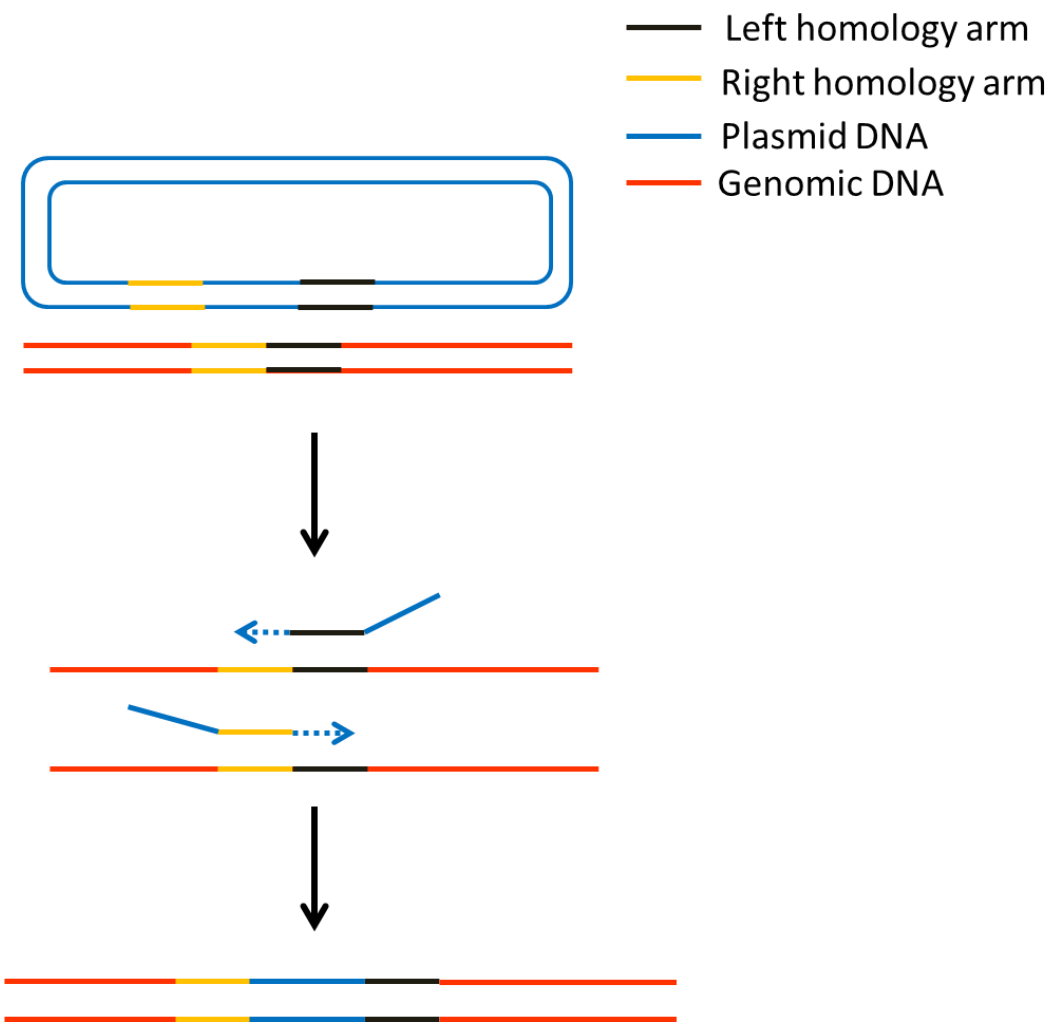


Figure 9-4. Using homology arms, it is possible to integrate DNA into genomic DNA by the process of homologous recombination.
The most desired result is a double crossover as shown here, where homology arms either side of the plasmid DNA of interest results in integration of only that target DNA and not the full plasmid. It is also possible to achieve a single crossover where the entire plasmid can be integrated into the genome.

9.3.2 Integration site selection

A range of homology arm sizes were investigated which added to the requirement for a region in the genome to design homology arms. The region needed to be in a locus that would not have an effect on growth or survival of the bacteria and also needed to be large enough to not have an effect on

flanking genome regions or other nearby metabolic pathways. The *Streptomyces lividans* genome is approximately 8.3 Mbp long (Rückert et al. 2015). Assuming the core genes for growth and core *metab*olism are within the central 3 Mbp of the genome and the outer 2 Mbp at each end of the linear genome are frequently mutated the ideal region to investigate is approximately between 2 and 3 Mbp in from each end of the genome. A maximum length of 5 kb per homology arm was investigated and therefore the ideal location needed to be 10 kb in length. We aimed to investigate two suitable regions to minimise failure due to any unpredictable effects of modifying the DNA.

To search for genes, operons and biosynthetic pathways within the *Streptomyces lividans* genome a variety of online tools were used including the GenBank database, the KEGG database, the BioCyc database and antiSMASH secondary *metab*olite mining and annotation (Kanehisa & Goto 2000; Clark et al. 2016; Caspi et al. 2016; Weber et al. 2015). After searching for large enough regions of the genome that would only effect one pathway and filtering this into pathways that would not effect growth, two regions of the *Streptomyces lividans* genome were selected for the design of homology arms. The regions selected contain genes relevant to cobalamin biosynthesis and desferrioxamine biosynthesis. The cobalamin synthesis pathway genes are located in an operon on the *Streptomyces coelicolor* genome; and although only the *cobN* gene is annotated in the *S. lividans* genome these species are similar and further analysis of nearby genes would lead us to assume that the cobalamin biosynthesis pathway is also arranged in an operon in this genome. This operon in *S. lividans* is located between base pairs 2 194 980 and 2 211 537. This location fits our previous requirements of a position not too close to the ends of the genome but also not in the centre of the genome. Cobalamin, also known as vitamin B12, is a cofactor containing cobalt and is only used in some prokaryotes. This would indicate that it is not essential for survival. There is literature describing the effects of

cobalamin deficiency in *S. coelicolor* which describes reduced growth in minimal media but not in rich media. This indicates a function for cobalamin in the growth and development of *Streptomyces* species but not in survival (Takano et al. 2015). Though there does appear to be an effect on growth, for our research purposes and a lack of abundance of other suitable sites this region was selected for the design of homology arms. The desferrioxamine biosynthesis pathway is located between base pairs 3 244 783 and 3 254 785 which is more towards the centre of the genome than initially planned but is far enough away from the cobalamin synthesis pathway that will be investigated. This is important because if the sites were too close then they may have the same negative effect on survival or growth and both would not be suitable for investigated cloning tools. Literature describes desferrioxamine as an iron siderophore that are found in abundance in *Streptomyces* species (Schrempf & Dyson 2011). Reported side effects arising from the inhibition of some desferrioxamine biosynthesis pathways in *S. coelicolor* are impaired growth or inhibition of spore formation (Tierrafría et al. 2011; Lambert et al. 2014). It is unknown whether either of these observations correspond to the equivalent desferrioxamine biosynthesis pathway as has been selected in the *S. lividans* genome. Therefore the effect of disrupting this pathway is unknown but may affect an aspect of cell growth. For the purpose of this study, this should not be a major hindrance. The main aim is to achieve homologous recombination in *Steptomyces lividans* and to determine success colonies need to be observed, further growth of these colonies is not necessary.

9.3.3 Vector design

There are two methods for transforming DNA into *Streptomyces;* these are conjugation and protoplasting (Kieser et al. 2000). Conjugation involves the use of *E. coli* ET12567/pUZ8002 containing the plasmid intended for transformation into *Streptomyces*. Protoplasting involves permeabilising the *Streptomyces* cells (to create protoplasts) and electroporation of these

protoplasts so that they will take up the new plasmid DNA. Conjugation traditionally has a good success rate and it used routinely in academic and industry research. Using the conjugation method requires a vector that contains the relevant DNA elements for conjugation as well as elements for transformation and replication in *E. coli*. A widely used vector for conjugation in *Streptomyces* research is the pSET152 vector. This vector contains all the necessary elements and also contains DNA encoding for the integrase protein φC31 and attP integration sites. For the purpose of this study, the integrase and integration sites in pSET152 would interfere with investigating integration by homologous recombination. These elements were therefore removed from the pSET152 vector using PCR to amplify the desired fragment of the backbone with appropriate restriction enzymes at each end and ligated to re-circularise the backbone. The pSET152 vector contains an apramycin resistance gene which does allow selection in both *E. coli* and *S. lividans.* With the φC31 integration the whole plasmid would be integrated into the genome, however using the homologous recombination method, the aim is to only introduce the fragment of DNA present between the two homology arms. The apramycin resistance gene is not located between the homology arms and therefore to select for positive clones in *S. lividans* after transformation neomycin resistance was chosen to be inserted between homology arms and thus be integrated into the *S. lividans* genome. With homologous recombination it is also possible for a single crossover to occur resulting in the entire plasmid being integrated into the genome. For this reason a reporter gene was also included in the plasmid outside of the homology arms, the GusA gene. GusA encodes for β-glucoronidase and provides a blue/white screen that has been widely used in *Steptomyces* and other bacteria (Myronovskyi et al. 2011; Li et al. 2015). Having this reporter outside the homology arms it is possible to quickly screen for double crossover mutants which will appear as white colonies and single crossover mutants will appear blue when plated onto selective agar containing X-gluc

(5-bromo-4-chloro-3-indolyl-beta-D-glucuronide). The resulting vector is shown in Figure 9-5 below.

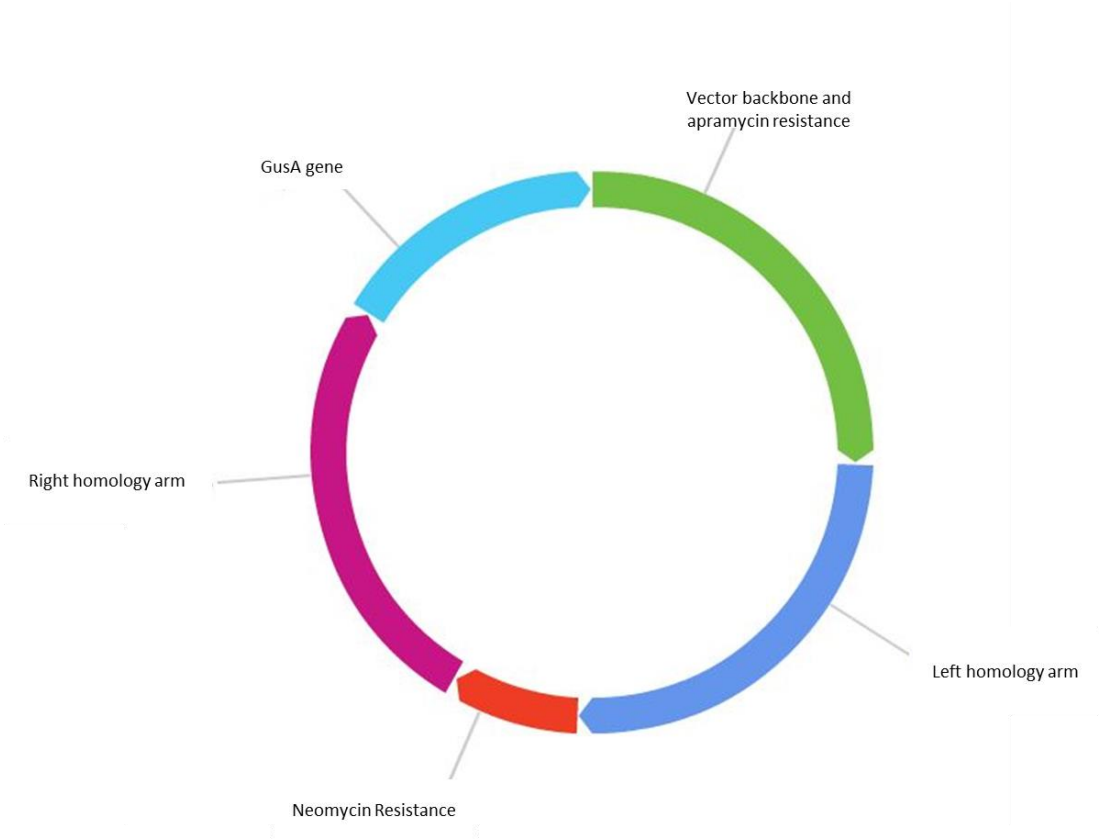


Figure 9-5. Vectors for conjugation into *Streptomyces* species in this study will follow this design.
The left and right homology arms will vary in size from 1 kb each to 5 kb each. The vector backbone contains genes necessary for conjugation and an apramycin resistance gene for selection in *E. coli* during cloning. The neomycin gene should be integrated into the *Streptomyces* genome after conjugation and allow selection of successful conjugants. The GusA gene is also a selection marker that will allow selection for *Streptomyces* colonies that have integrated the full plasmid into the genome as a result of an undesired single crossover.

One of the main objectives of this study is to create a vector that can be easily manipulated and adapted for cloning in a range of *Streptomyces* species and different locations in the genome. Therefore the homology arms must be easily interchangeable. One widely used system to facilitate this is the golden gate system which takes advantage of type IIS restriction

endonucleases which cut outside of their DNA recognition sequence thus leaving the restriction site intact. Using this system and creating a system where overhangs left by the restriction enzyme for each individual element in the vector (described as parts) are different means one restriction enzyme can be used for assembly of the entire vector (Figure 9-6). The same enzyme can also be used when single or few parts need to be altered, for example changing the homology arms.

Figure 9-6. The vector is designed so that the ends of each part have unique overlapping overhangs when digested with a type IIS endonuclease.
This ensures that the vector is assembled correctly and also allows for easily interchangeable parts. Base pairs indicated by brackets at the ends of each genetic part indicate the overhanging base pairs after digestion with SapI.

Designing the vector in this way with standard overhangs allowing for interchangeable parts is a step toward the more systematic design of

biological experiments. This type of experimental design allows for the move towards engineering biology by creating standard parts. The main challenge with this system is ensuring that the DNA elements do not contain the desired type IIS endonuclease recognition site. Therefore a number of the parts, including the vector backbone were mutated using site directed mutagenesis to remove SapI restriction sites.

9.4 Appendix D - Antha script

The following scripts were used in chapter 6 for automation.

9.4.1 Script for gradient PCR set up

The below script was used with the PipetMax™ to load 96 well plates ready for gradient PCR. This was used for PCR from the *Streptomyces* genome and for the colony PCR for construct validation.

```
protocol AutoPCR_Gradient_mmx
import (
"github.com/antha-lang/antha/antha/anthalib/wtype"
"github.com/antha-lang/antha/microArch/factory"
)
// Input parameters for this protocol (data)
Parameters (
// PCRprep parameters
Projectname string
Reactiontotemplate map[string]string // e.g. ["left homology arm"]:"templatename"
Reactiontoprimerpair map[string][2]string // e.g. ["left homology arm"]:"fwdprimer","revprimer"
RowGradientRatherthanColumn bool // if true, 12 replicates of each reaction will be set up, one set of reactions per row, else 8 reactions set up 1 set per column
)
// Data which is returned from this protocol, and data types
Data (
        Error error
        NumberOfReactions int
)
// Physical Inputs to this protocol with types
Inputs (
FwdPrimertype *wtype.LHComponent
RevPrimertype *wtype.LHComponent
Templatetype *wtype.LHComponent
Plate *wtype.LHPlate
)
// Physical outputs from this protocol with types
Outputs (
Reactions []*wtype.LHComponent
)
```

```
Requirements {
}
// Conditions to run on startup
Setup {
}
// The core process for this protocol, with the steps to be performed
// for every input
Steps {
	var Samplenumber int
	// if RowGradientRatherthanColumn == true,
	//12 replicates of each reaction will be set up,
	//one set of reactions per row,
	//else 8 reactions set up 1 set per column
	if RowGradientRatherthanColumn{
		Samplenumber = 12
	} else {
		Samplenumber = 8
	}
	var counter int
Reactions = make([]*wtype.LHComponent,0)
volumes := make([]wunit.Volume,0)
welllocations := make([]string,0)
/*
// add step to make mastermix first
mastermix := RunSteps(MakePCRmmx,
		Parameters{
		    WaterVolume: wunit.NewVolume(10,"ul"),
                                        ReactionVolume:
wunit.NewVolume(25,"ul"),
	BufferConcinX: 5,
   FwdPrimerName: Reactiontoprimerpair[reactionname][0],
   RevPrimerName: Reactiontoprimerpair[reactionname][1],
	 TemplateName: templatename,
   ReactionName: reactionname,
	FwdPrimerVol: wunit.NewVolume(1,"ul"),
	RevPrimerVol: wunit.NewVolume(1,"ul"),
	AdditiveVols: []wunit.Volume{wunit.NewVolume(5,"ul")},
	Templatevolume: wunit.NewVolume(1,"ul"),
	PolymeraseVolume: wunit.NewVolume(1,"ul"),
	DNTPVol:wunit.NewVolume(1,"ul"),
	Numberofcycles: 30 ,
	InitDenaturationtime: wunit.NewTime(30,"s"),
	Denaturationtime: wunit.NewTime(5,"s"),
	Annealingtime: wunit.NewTime(10,"s"),
	AnnealingTemp:  wunit.NewTemperature(72,"C"),  // Should  be
calculated from primer and template binding
```

```
        Extensiontime: wunit.NewTime(60,"s"), // should be calculated from
template length and polymerase rate
        Finalextensiontime: wunit.NewTime(180,"s"),
        Hotstart: false,
        AddPrimerstoMasterMix: false,
                WellPosition: wellposition,
            }, Inputs{
        FwdPrimer:FwdPrimertype,
        RevPrimer: RevPrimertype,
        DNTPS: factory.GetComponentByType("DNTPs") ,
        PCRPolymerase:factory.GetComponentByType("Q5Polymerase"),
        Buffer:factory.GetComponentByType("Q5buffer"),
        Water:factory.GetComponentByType("water"),
        Template: Templatetype,
        Additives:
[]*wtype.LHComponent{factory.GetComponentByType("GCenhancer")} ,
        OutPlate: Plate,
                    })
*/
for reactionname, templatename := range Reactiontotemplate {
        //wellposition := Plate.AllWellPositions(wtype.BYCOLUMN)[counter]
for j:=0;j< Samplenumber;j++{
//      for i:= 0;i < len(Reactions);i++{
        var wellcoords  = wtype.WellCoords{X:counter,Y:j}
        if RowGradientRatherthanColumn{
                wellcoords  = wtype.WellCoords{X:j,Y:counter}
        } else {
                wellcoords  = wtype.WellCoords{X:counter,Y:j}
        }
        wellposition := wellcoords.FormatA1()
result := RunSteps(PCR_vol_mmx,
            Parameters{
          //      WaterVolume: wunit.NewVolume(10,"ul"),
                                        MasterMixVolume:
wunit.NewVolume(17,"ul"),
   FwdPrimerName: Reactiontoprimerpair[reactionname][0],
   RevPrimerName: Reactiontoprimerpair[reactionname][1],
   TemplateName: templatename,
   ReactionName: reactionname,
        FwdPrimerVol: wunit.NewVolume(1,"ul"),
        RevPrimerVol: wunit.NewVolume(1,"ul"),
        Templatevolume: wunit.NewVolume(1,"ul"),
        PolymeraseVolume: wunit.NewVolume(1,"ul"),
        Numberofcycles: 30 ,
        InitDenaturationtime: wunit.NewTime(30,"s"),
        Denaturationtime: wunit.NewTime(5,"s"),
        Annealingtime: wunit.NewTime(10,"s"),
```

```
        AnnealingTemp: wunit.NewTemperature(72,"C"), // Should be calculated from primer and template binding
        Extensiontime: wunit.NewTime(60,"s"), // should be calculated from template length and polymerase rate
        Finalextensiontime: wunit.NewTime(180,"s"),
        PrimersalreadyAddedtoMasterMix: false,
                PolymeraseAlreadyaddedtoMastermix: true,
                WellPosition: wellposition,
            }, Inputs{
        FwdPrimer:FwdPrimertype,
        RevPrimer: RevPrimertype,
        PCRPolymerase:factory.GetComponentByType("Q5Polymerase"),
        Template: Templatetype,
        OutPlate: Plate,
        MasterMix: factory.GetComponentByType("Q5mastermix"),
            })

   Reactions = append(Reactions, result.Outputs.Reaction)
   volumes = append(volumes,result.Outputs.Reaction.Volume())
        welllocations = append(welllocations,wellposition)

   }
        counter++
        }
        NumberOfReactions = len(Reactions)
        Error = wtype.ExportPlateCSV(Projectname+".csv", Plate,Projectname+"outputPlate", welllocations, Reactions, volumes)
}
```

9.4.2 Script for loading and running E-gel®

The below script was used to load E-gels® from samples in a 96 well plate using the Gilson PipetMax™ and provide a file with the order of loading onto the gel for easy analysis.

```
// example protocol for loading a DNAgel
protocol DNA_gel_fromCSV
import (
        "fmt"
        "github.com/antha-lang/antha/antha/anthalib/wtype"
        "github.com/antha-lang/antha/antha/anthalib/mixer"
```

```
        inplate "github.com/antha-lang/antha/target/mixer"

)
// Input parameters for this protocol (data)
Parameters (

        ProjectName string
        Loadingdyeinsample bool
        InputCSVfile string
        Samplenumber int
        Watervol       Volume
        LadderVolume Volume
        Loadingdyevolume Volume
        DNAgelrunvolume Volume

        Mixingpolicy string //wtype.LiquidType

        //DNAladder Volume // or should this be a concentration?

        //DNAgelruntime time.Duration
        //DNAgelwellcapacity Volume
        //DNAgelnumberofwells int32
        //Organism                    Taxonomy                    //=
http://www.ncbi.nlm.nih.gov/nuccore/49175990?report=genbank
        //Organismgenome Genome
        //Target_DNA wtype.DNASequence
        //Target_DNAsize float64 //Length
        //Runvoltage float64
        //AgarosePercentage Percentage
) // polyerase kit sets key info such as buffer composition, which effects
primer melting temperature for example, along with thermocycle parameters

// Data which is returned from this protocol, and data types
Data (
  //    NumberofBands[] int
        //Bandsizes[] Length
        //Bandconc[]Concentration
        //Pass bool
        //PhotoofDNAgel Image
        Error error
)


// Physical Inputs to this protocol with types
Inputs (
   Water       *wtype.LHComponent
        Ladder *wtype.LHComponent
```

```
        Sampletotest *wtype.LHComponent//WaterSolution
        Loadingdye  *wtype.LHComponent//WaterSolution  //Chemspiderlink  //
not correct link but similar desirable
        DNAgel *wtype.LHPlate  // gel
        MixPlate *wtype.LHPlate// plate to mix samples if required

        //DNAladder *wtype.LHComponent//NucleicacidSolution
    //Water *wtype.LHComponent//WaterSolution

        //DNAgelbuffer *wtype.LHComponent//WaterSolution
        //DNAgelNucleicacidintercalator  *wtype.LHComponent//ToxicSolution
// e.g. ethidium bromide, sybrsafe
        //QC_sample *wtype.LHComponent//QC // this is a control
        //DNASizeladder *wtype.LHComponent//WaterSolution
        //Devices.Gelpowerpack Device
) // need to calculate which DNASizeladder is required based on target
sequence length and required resolution to distinguish from incorrect
assembly possibilities

// Physical outputs from this protocol with types
Outputs (
        Loadedsamples []*wtype.LHComponent//Gel
    //
)

// No special requirements on inputs
Requirements {
    // None
/* QC if negative result should still show band then include QC which will
result in band // in reality this may never happen... the primers should be
designed within antha too
control blank with no template_DNA */
}

// Condititions run on startup
// Including configuring an controls required, and the blocking level needed
// for them (in this case, per plate of samples processed)
Setup {
        /*control.config.per_DNAgel {
        load DNASizeladder(DNAgelrunvolume) // should run more than one
per gel in many cases
        QC          :=          mix          (Loadingdye(loadingdyevolume),
QC_sample(DNAgelrunvolume-loadingdyevolume))
        load QC(DNAgelrunvolume)
        }*/
}
```

```
// The core process for this protocol, with the steps to be performed
// for every input
Steps {

	// parse sample locations from file
	inputplate,err := inplate.ParseInputPlateFile(InputCSVfile)

	if err != nil {
		Errorf(err.Error())
	}

	// count number of colonies from inplate
	loadedsamples := make([]*wtype.LHComponent,0)
	wells := make([]string,0)
	volumes := make([]wunit.Volume,0)

	var DNAgelloadmix *wtype.LHComponent
	var loadedsample *wtype.LHComponent
	Water.Type = wtype.LTloadwater

	var counter int

	// work out sample volume

	// copy volume
	samplevolume := (wunit.CopyVolume(DNAgelrunvolume))

	// subtract volume of water
	samplevolume.Subtract(Watervol)
	/*
	// add ladder sample to first column
	loadedsample = MixInto(
	DNAgel,
	DNAgel.AllWellPositions(wtype.BYROW)[counter],
	mixer.Sample(Water,Watervol),
	mixer.Sample(Ladder, samplevolume),
	)

	loadedsamples = append(Loadedsamples,loadedsample)
	wells                                                    =
append(wells,DNAgel.AllWellPositions(wtype.BYROW)[counter])
	volumes = append(volumes,loadedsample.Volume())
	counter++
	*/
	for j:=0;j< Samplenumber;j++{
	for _, wellcontents := range inputplate.AllWellPositions(false){
```

```
	if inputplate.WellMap()[wellcontents].Empty() == false{

		// ready to add water to well
	waterSample := mixer.Sample(Water,Watervol)

	// get position, ensuring the list is by row rather than by column
	position := DNAgel.AllWellPositions(wtype.BYROW)[counter]

		//get well coordinates
	wellcoords := wtype.MakeWellCoordsA1(position)
	fmt.Println("wellcoords.X",wellcoords.X)


	// if first column add ladder sample
	if wellcoords.X == 0 {

	Ladder.Type,_ = wtype.LiquidTypeFromString(Mixingpolicy)

		laddersample := MixInto(
		DNAgel,
	DNAgel.AllWellPositions(wtype.BYROW)[counter],
	mixer.SampleForTotalVolume(Water,DNAgelrunvolume),
	mixer.Sample(Ladder, LadderVolume),
	)

	loadedsamples = append(loadedsamples,laddersample)
	wells = append(wells,position)
	volumes = append(volumes,laddersample.Volume())
	counter++

	}

	// refresh position in case ladder was added
	position = DNAgel.AllWellPositions(wtype.BYROW)[counter]


	// temporarily avoiding using last column; remove later
	// skips contents so fix properly!!
	/*
	if wellcoords.X == DNAgel.WlsX-1{
		counter++
		position = DNAgel.AllWellPositions(wtype.BYROW)[counter]
	}
		*/
	Sampletotest =  inputplate.WellMap()[wellcontents].WContents
```

```
	// load gel

	// add loading dye if necessary
		if Loadingdyeinsample == false {

		Loadingdye.Type,_ = wtype.LiquidTypeFromString("NeedToMix")

		DNAgelloadmixsolution := MixInto(
		MixPlate,
		"",
		mixer.Sample(Sampletotest,samplevolume),
		mixer.Sample(Loadingdye,Loadingdyevolume),
		)
		DNAgelloadmix = DNAgelloadmixsolution
		}else {

		DNAgelloadmix = Sampletotest

		}

	// Ensure  sample will be dispensed appropriately:


	// comment this line out to repeat load of same sample in all wells using first sample name
	//DNAgelloadmix.CName = Samplenames[i]//[i] //originalname + strconv.Itoa(i)

	// replacing following line with temporary hard code whilst developing protocol:
	DNAgelloadmix.Type,_ = wtype.LiquidTypeFromString(Mixingpolicy)

	//DNAgelloadmix.Type = "loadwater"

	loadedsample = MixInto(
	DNAgel,
	position,
	waterSample,
	)

	loadedsample = Mix(loadedsample,mixer.Sample(DNAgelloadmix, samplevolume))

	loadedsamples = append(loadedsamples,loadedsample)
	wells = append(wells,position)
	volumes = append(volumes,loadedsample.Volume())
```

```
        counter++
        }
        }

}
        Loadedsamples = loadedsamples

        // export to file
        //wtype.AutoExportPlateCSV(ProjectName+".csv",DNAgel)
        Error = wtype.ExportPlateCSV(ProjectName+"_gelouput"+".csv",
DNAgel, ProjectName+"gelouput", wells, Loadedsamples, volumes)
        // Then run the gel
```