

Toward Engineering the Substrate Specificity of a PHA Synthase (PhaC)

Alex Kane

A thesis submitted to the Victoria University of
Wellington in fulfilment for the degree of Master of
Science in Biotechnology

Victoria University of Wellington

2019

Abstract

Manufacturing of high-grade plastics from petroleum-based feedstocks is a high-cost, unsustainable process resulting in expensive products. My overall goal was to engineer the pathway of bacterial bio-polyester formation, in order to produce high-grade bio-plastics. More specifically, the aim was to introduce aromatic rings into the main-chain of the polyhydroxyalkanoate (PHA) polymer currently produced by specialist bacteria. This research aimed to create these bio-plastics from renewable resources, rather than relying on petroleum-based sources.

A key enzyme for this process is the polyhydroxyalkanoate synthase, PhaC. This enzyme is capable of polymerizing activated hydroxybutyrate-CoA monomers. I began with the establishment of a system that allowed the use of directed evolution. I constructed a minimal plasmid for the expression of PhaC and a second plasmid with the CoA ligase genes required for substrate activation. I generated error-prone PCR libraries of the *Cupriavidus necator phaC_a*, *Chromobacterium* sp. USM2 *phaC_b* and an ancestrally reconstructed *phaC_b-LCA* that contained differing spectra of mutations. A life-or-death selection was employed to select for PhaC variants able to polymerise aromatic substrates based upon the toxicity of the un-polymerized aromatic hydroxy-acid monomers. I determined the minimum inhibitory concentrations (MICs) for six of these monomers in *Escherichia coli* for downstream selection. Lastly, I adapted a Nile red screening method to test wild-type PHA accumulation of PhaC enzymes.

Selections for mutants capable of polymerizing aromatic monomers were implemented on the libraries generated from *phaC_a* and *phaC_b*. Whereas, the library generated from *phaC_b-LCA* was screened for variants with increased wild-type activity. Selections yielded no candidates for further testing. However, the screen isolated several variants with increased wild-type activity. These variants may serve as a new scaffold for further mutagenesis experiments to achieve the overall goal; to produce a high-grade bioplastic.

Acknowledgments

Firstly, I would like to thank my supervisor A/Prof Wayne Patrick for all of his guidance and help with my project. I have had many ups and downs throughout this project but Wayne has stayed positive throughout and knew when a ‘kick up the ass’ was needed. It is great to have someone so passionate and driven as Wayne to guide me through my first experience with scientific research. I really cannot put into words how grateful I am that you managed to put up with me for almost two years, that is no mean feat.

The whole of the E³/M³/MME labs have been instrumental in the design, implementation and support of my project and I am very grateful for their continued support and guidance. I was warmly welcomed by a group of strangers and I will leave with friends for life. From being blind-drunk to having the most intellectual conversations of my life, I could not have hoped for a better group of people to do this with.

I want to give a special mention to Tiffany Oulavallickal for over-seeing my project at the beginning and helping with anything whenever I was lost or unsure. She has taken time out of her busy days and given me more than I could have ever asked for. Mike Fairhurst has also been of great assistance. Thanks for all of your patience with my constant pestering and ‘dumb’ questions.

Thank you to my family for their continued support and all of my friends who have helped me along the way. A thank you to Otago and Victoria Universities for allowing me to complete my Masters and to Studylink for the financing.

There is nothing I would change about my time doing this project (that may be a lie), and I would again like to thank everyone involved. I could not have done this without you guys.

The research idea for making aromatic polyesters outlined in this thesis was developed by Scion. I acknowledge the financial support from Scion through subcontracts to the University of Otago and Victoria University of Wellington from the MBIE Endeavour Fund Bioinspired Aromatic Polyesters Smart idea programme C04X1604 (1st October 2016 - 30th September 2018). The provision of plasmid constructs and the helpful guidance by colleagues at Scion is appreciated.

Abstract	i
Acknowledgements	ii
Table of Contents	iii
List of Figures	v
List of Tables	vi
Chapter 1	1
Introduction	
1.1 Bio-based plastics	1
1.2 PHA synthase	4
1.2.1 Ancestral sequence reconstruction of PhaC	7
1.3 Directed Evolution	7
1.4 Genetic Selection for altered specificity of PhaC	9
1.5 Screen for highly active PhaC variants	10
1.6 Aims	11
Chapter 2	12
Methods and materials	
2.1 Regulatory approval	12
2.2 Materials	12
2.3 Methods	13
2.3.1 Media	13
2.3.2 Freezer stocks	13
2.4 DNA techniques	14
2.4.1 DNA extraction and purification	14
2.4.2 Primer design and storage	14
2.4.3 Primers used	15
2.4.4 Agarose gel electrophoresis	15
2.4.5 High-fidelity PCR	16
2.4.6 Error-prone PCR	16
2.4.7 Colony PCR	17
2.4.8 Whole-circle PCR	17
2.4.9 Sequencing	17
2.4.10 Restriction digests	18
2.4.11 Ligation reactions	18
2.4.12 Electro-competent cells and transformation	18
2.5 Determination of minimum inhibitory concentrations	19
2.5.1 MIC Set-up	19
2.5.2 LB-agar plate MIC	20
2.6 Selection strategy	20
2.6.1 Library plating	20
2.6.2 Agar-based selection of epPCR libraries	20
2.6.3 Winners and further selections	21
2.6.4 Single plasmid isolation	21

2.7	Screening for PHA production	22
2.7.1	Viable Nile red colony staining	22
2.7.2	Plate based Nile red fluorescence screen	22

Chapter 3 **24**

Directed evolution and selection of PhaC_a

3.1	Introduction	24
3.2	MIC analysis	24
3.2.1	MIC determination	24
3.3	Construction of ligase-only plasmid	26
3.4	Error-prone PCR library generation	27
3.4.1	Error-prone PCR	27
3.4.2	Error-prone PCR library analysis	29
3.5	Genetic selections for cells with an improved PhaC_a	32
3.5.1	Selection of variant PhaC _a enzymes	32
3.6	Chapter 3 discussion	34

Chapter 4 **38**

Directed Evolution of PhaC_b

4.1	Introduction	38
4.2	Error-prone PCR library generation	38
4.2.1	Error-prone PCR	38
4.2.2	Error-prone PCR library analysis	39
4.3	Genetic selections for cells with an improved PhaC_b	42
4.3.1	Basic wild-type MCI test	42
4.3.2	Selections of variant PhaC _b enzymes	44
4.3.3	Position of mutations on the structure of PhaC _b	46
4.4	Chapter 4 discussion	47

Chapter 5 **50**

Directed Evolution of PhaC_b-LCA

5.1	Introduction	50
5.2	Error-prone PCR library generation	50
5.2.1	Error-prone PCR	50
5.2.2	Error-prone PCR library analysis	51
5.3	Two-tier screen for PHA accumulation	54
5.3.1	Agar-based screen for PHA accumulation	54
5.3.2	Microplate-based screen for PHA accumulation	55
5.4	Chapter 5 discussion	59

Chapter 6 **62**

Key Findings and Future Directions

6.1	Minimal selection pressure for PhaC to polymerise the aromatic substrates	62
------------	--	-----------

6.2	Expressing PhaC may impose a fitness cost on the cell	63
6.3	PhaC_b-LCA gained increased wild-type activity	64
6.4	PhaC_b-LCA as a scaffold for further mutagenesis	65
6.5	Perspectives on the utility of error-prone PCR	66
6.6	Concluding remarks	67
<u>References</u>		<u>69</u>

List of Figures

Figure 1.1: A cross section of a bacterial cell with PHA inclusions	2
Figure 1.2: Comparison of PHA to Kevlar	3
Figure 1.3: Displaying native polymerisation pathway of PHA from sugar	4
Figure 1.4: Crystal structure of the catalytic domain of PhaC _a	6
Figure 1.5: Aromatic hydroxy acid substrate polymerisation	10
Figure 3.1: Initial MIC determination using a doubling dilution series	25
Figure 3.2: Whole-circle PCR to remove PhaC _a (1802 bp) from pGETS-MH.	27
Figure 3.3: PhaC _a error-prone PCR products	28
Figure 3.4: 1% agarose gel showing the colony PCR screen.	33
Figure 3.5: One step reaction showing the addition of a CoA molecule onto 3-hydroxybenzoic acid monomer	36
Figure 4.1: Error-prone PCR products of <i>phaC_b</i> .	39
Figure 4.2: Agarose gels showing the colony PCR screen of the <i>phaC_b</i> gene ‘winners’ after genetic selection.	44
Figure 4.3: Structure of the catalytic domain of PhaC _b with colour-coded mutations.	47
Figure 5.1: 1% agarose gel showing the epPCR products of <i>phaC_b-LCA</i> .	51
Figure 5.2: Viable colony staining of the PHA accumulating cells from the pooled <i>phaC_b-LCA</i> epPCR library.	55
Figure 5.3: Bar graph showing 92 <i>phaC_b-LCA</i> variants and respective fluorescent intensities.	56
Figure 5.4: Homology model of PhaC _b -LCA with mutations from Table 5.5 mapped onto the structure.	58

List of Tables

Table 2.1: Characteristics of <i>E. coli</i> used in this project	12
Table 2.2: Characteristics of plasmids used in this project	13
Table 2.3: Primers used in this project	15
Table 2.4: Table showing the general settings for thermal cycler PCR reactions	16
Table 2.5: Table showing the general settings for thermal cycler epPCR reactions.	17
Table 3.1: Table showing Six potential aromatic hydroxy acid substrates	25
Table 3.2: MICs for six aromatic hydroxy acid substrates	26
Table 3.3: Comparison of the mutational spectrums of <i>phaC_a</i> variants from the 4 ng and 20 ng starting templates	31
Table 3.4: Composition of the 4 ng <i>phaC_a</i> mutant library as estimated by PEDEL-AA	31
Table 3.5: Sequencing results of <i>phaC_a</i> variants from genetic selections	34
Table 4.1: Comparison of the mutational spectrum of the <i>phaC_b</i> variants from the 2 ng and 4 ng starting templates	41
Table 4.2: Composition of the 2 ng <i>phaC_b</i> epPCR library as estimated by PEDEL-AA	42
Table 4.3: Composition of the 4 ng <i>phaC_b</i> epPCR library as estimated by PEDEL-AA	42
Table 4.4: Sequencing results of the <i>phaC_b</i> variants from genetic selections.	46
Table 5.1: Comparison of the mutational spectrums of <i>phaC_b-LCA</i> variants from the 2 ng and 4 ng starting templates.	53
Table 5.2: Composition of the 2 ng <i>phaC_b-LCA</i> epPCR library as estimated by PEDEL-AA	54
Table 5.3: Composition of the 4 ng <i>phaC_b-LCA</i> epPCR library as estimated by PEDEL-AA	54
Table 5.4: Sequencing results of <i>phaC_b-LCA</i> variants from PHA accumulation screen.	57

Chapter 1

Introduction

1.1 Bio-based plastics

Over recent years, there has been a global increase of concentration to find alternative resources and production schemes, be it for power, plastic or waste products. This increase in concern arises from the rapid depletion of the world's energy and manufacturing resources. For this project, I concentrate on the production of high grade plastics, which is currently very expensive and risks being unsustainable in the long term. This is largely due to the dependence on oil and gas for petroleum based plastics. What makes this problem even more imminent is that this fossil fuel reserve is thought to be totally depleted by 2030 (Chedid *et al.*, 2007). Whether or not this is an accurate estimate, synthesis of both high-grade and general plastics from alternate resources is vital. The ability to biologically produce a high-grade plastic that is structurally similar to traditional petroleum-based plastics may be the future of the plastic industry.

Here, I propose two bio-based substitutes, a high-grade bioplastic and a natural polyester produced by bacteria. These have the advantage of cutting production costs with the potential added benefit of being more environmentally friendly. Lee and Na, (2013), detail the risks that climate change and other environmental factors pose for the sustainable manufacturing of plastics. This literature heightens the promise that PHAs may be able to possess material properties similar to the polymers that are presently used.

There is currently no plastic manufacturing industry in New Zealand and yet we import and process around 250,000 tons/year (Aoteaplas, 2017). New Zealand, however, has a large amount of unused carbon waste products from the logging industry, biodiesel refineries, and milk processing plants. The ability to produce a 'green' plastic derived from these waste materials opens a window to take low-value waste by-products and generate high-value end-products. In nature, many bacteria are found with enzymes capable of producing polyhydroxyalkanoates (PHAs) (**Figure 1.1**).

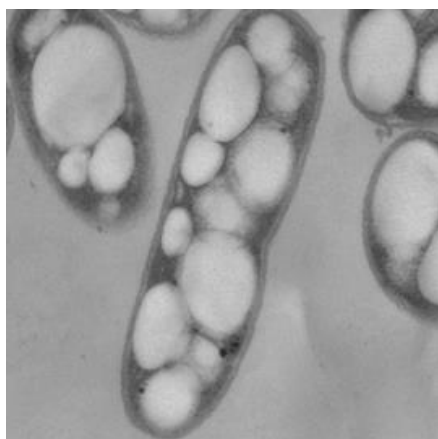


Figure 1.1: A cross-section of a bacterial cell with PHA inclusions. <https://tinyurl.com/yd9hpr4s>.

These are carbon and energy reserve polyesters produced by bacteria, which have attracted industrial attention for decades due to their thermoplastic and biodegradable properties (Anderson and Dawes, 1990). Efficient production of these PHAs derived from the aforementioned waste products could change this attention into action.

PHAs are diverse in both composition and size. More than 150 different hydroxyalkanoate (HA) monomers are known to be constituents of PHAs, many of which have desirable industrial properties (Steinbuechel and Valentin, 1995). According to Sagong *et al.* (2018), PHAs can be classified into three groups of short, medium and long chain length (SCL, MCL and LCL respectively). This is based on the number of carbon atoms in the monomer units: SCL-PHAs contain C3 to C5, MCL-PHAs contain between C6 and C14 and the LCL-PHAs comprise more than C14.

Choi and Lee (2004), showed that higher molecular weight PHAs have more mechanical strength and are able to be processed into strong films or fibres. Following on from this, Agus *et al.* (2006) displayed that acidic conditions and higher temperature are factors in the production of PHAs with increased molecular weight.

The ability to generate natural polyesters from waste was shown to be possible by Chanasit *et al.*, (2016). This research group managed to use a native *Pseudomonas* strain to efficiently produce PHAs using biodiesel waste as the sole carbon source. Another example is from Chung *et al.*, (2011), who created a PHA from 3-hydroxydodecanoate (3-HDD) monomers using mutant strains of *Pseudomonas* bacterium. The group knocked out key enzymes involved in β -oxidation and fatty acid synthesis to improve the production of the necessary

substrates for higher-grade PHA production. They managed primarily to produce a PHA comprised of 3-HDD that possessed increased thermo-tolerant properties. This demonstrates the possibility of creating a more diverse range of PHAs using metabolically engineered bacteria.

The initial goal of my project was to create a new bio-synthetic pathway for aromatic polyester biosynthesis. **Figure 1.2** displays three structures: A) low grade PHA, B) high-grade aromatic Kevlar and C) a potentially biologically produced aromatic polymer. Aromatic plastics are well-known and are a commodity in medical and industrial settings. An example displayed below is Kevlar, a strong and versatile polymeric plastic. The ability to generate a material biologically that contains aromatic rings would increase the performance and strength of the produced bioplastic compared to that of native PHAs. In saying that, the ability to increase the production of the PHAs by bacteria may provide the advancement needed to meet commercial demand of these biodegradable thermoplastics.

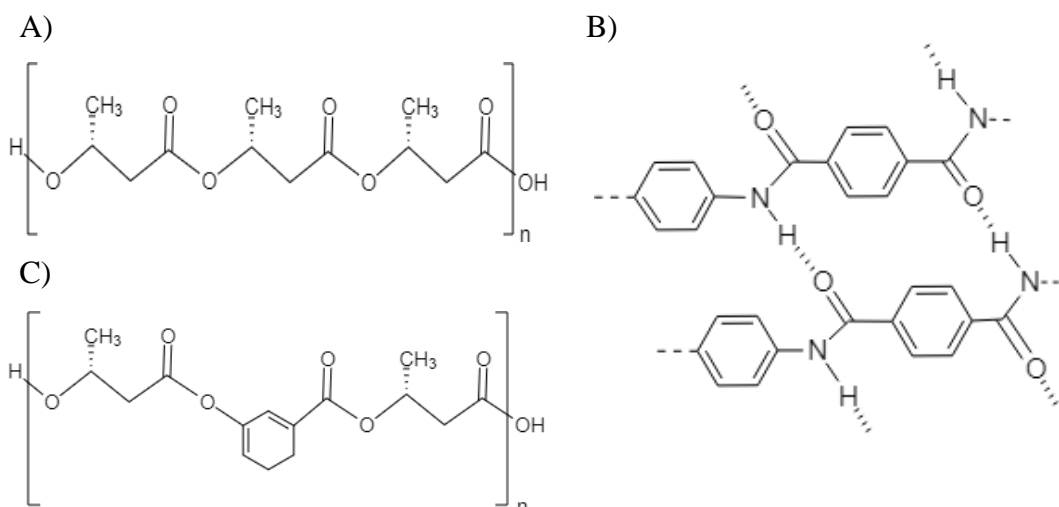


Figure 1.2: Comparison of PHA to Kevlar. A) shows the structure of the bacterially produced aliphatic bio-polymer, PHA. B) shows a section of the high-grade aromatic plastic known as Kevlar. C) shows the PHA backbone with aromatic rings incorporated into the main-chain of the polymer.

The native pathway of PHA synthesis is pictured in **Figure 1.3**. The critical step in creating a new biosynthetic pathway for aromatic polyesters is to engineer the substrate specificity of the synthase. This enzyme, polyhydroxyalkanoate synthase (PhaC) enzyme is found in more than 200 bacterial species known to be capable of synthesizing PHAs (Hiroe *et al.*, 2012). The production of bacterially produced

polyesters is largely dependent on the catalytic efficiency of the enzymes involved, specifically that of PhaC (Bhubalan *et al.*, 2011). Engineering this enzyme is critical for the possibility of this production scheme to overthrow traditional, yet unsustainable, petrochemicals.

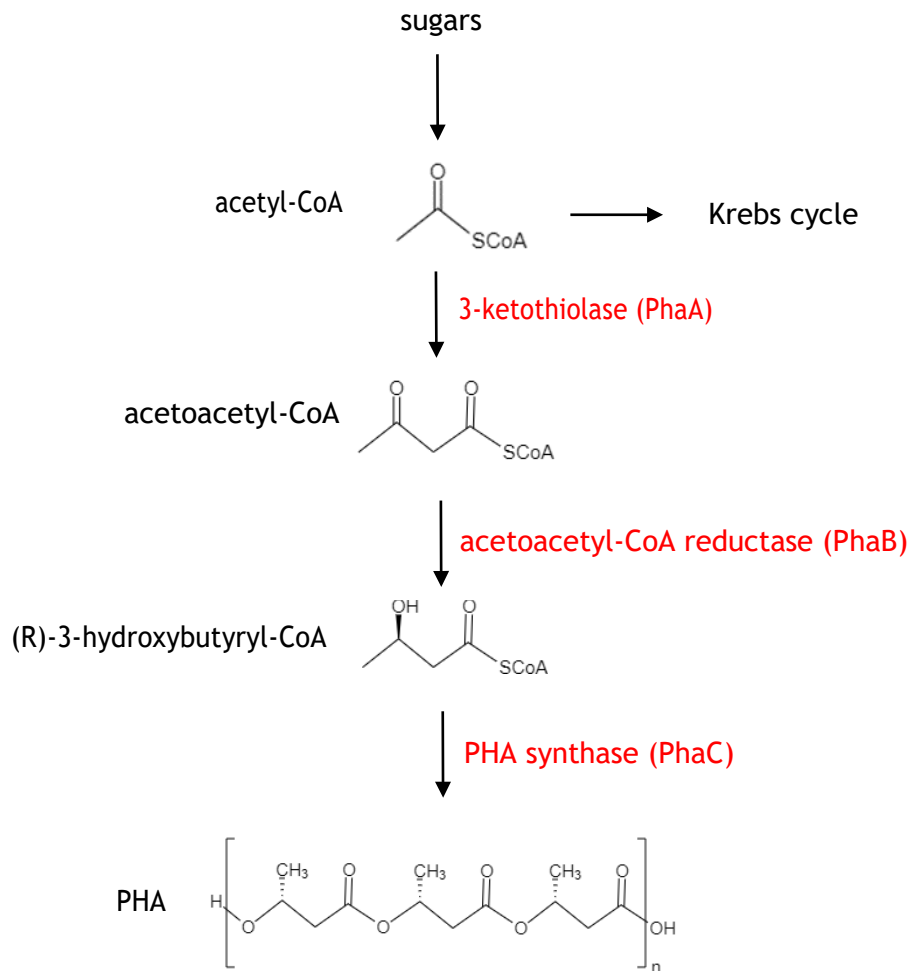


Figure 1.3: Displaying native polymerisation pathway of PHA from sugar. Enzymes required for these reactions are shown in red. The structure of the intermediates for each step of the pathway.

1.2 PHA Synthase

The PHA synthase (PhaC) enzyme polymerizes carbon compounds, principally 3-(R)-hydroxybutyryl-CoA, by hydrolyzing the CoA moiety and producing aliphatic bio-polyesters termed PHAs, pictured in **Figure 1.1**. Polymerisation is performed in bacteria as a means of carbon and energy storage under nutrient limited conditions (Steinbuechel and Valentin, 1995) and in the presence of excess carbon source

(Gumel *et al.* 2013). *C. necator* can use its PhaC to synthesise a large amount of PHA from simple carbon compounds such as sucrose and fructose (Byrom, 1992).

PhaC enzymes isolated from different bacterial species have a large variation in activity and the PHA produced (Meng *et al.* 2014). For example, the highly active PhaC from *Chromobacterium sp.* USM2 favours the production of short chain length PHAs and utilises 3-hydroxybutyrate up to 3-hydroxyhexanoate. PhaC enzymes isolated from most *Pseudomonas* strains of bacteria, however, often favour the creation of medium chain length PHAs and have the ability to utilise large substrates such as 3-hydroxydodecanoate. The substrates utilised by PhaC enzymes from different bacterial species depend largely on the PHA it produces.

To proceed in this project three PhaC enzymes were chosen to be engineered for altered specificity and/or activity. These were from *Cupriavidus necator* (PhaC_a), *Chromobacterium sp.* USM2 (PhaC_b) and an ancestrally reconstructed enzyme based off *Chromobacterium sp.* USM2 and its associated enzyme family (PhaC_b-LCA). *C. necator* PhaC was originally chosen due to the width of the substrate channel, ~4 Å (Wittenborn *et al.*, 2016). This means only a small change may be needed for this enzyme to adopt the proposed aromatic substrates into its channel. The second two enzymes were chosen for their superior polymerizing ability; *in vitro* assays showed an ~5-fold increase in PHA accumulation between PhaC_a and PhaC_b (Bhubalan *et al.*, 2011). Further crude *in vitro* assays by our collaborators at Waikato University showed that PhaC_b-LCA possessed a higher catalytic efficiency than PhaC_b.

There was little known about PhaC's overall mechanism of catalysis or any structural information until recently, when Wittenborn *et al.* (2016) published the first structure of the catalytic domain of the PhaC_a enzyme. This group mutated the catalytic cysteine at amino acid position 319 to an alanine for easier crystallization. More recently, Kim *et al.* (2017) solved the structure containing all three catalytic residues in the active site (**Figure 1.4**), showing the complete catalytic pocket. The structure suggests that PhaC must become dimerised in order to become active. PHAs are produced by PhaC in a processive manner using a Cys-His-Asp catalytic triad (Kim *et al.*, 2017). There is a second potential mechanism utilised by PhaC in which there is a non-processive ping-pong reaction switching between the two active sites at the dimer interface. This is currently less favoured due to the distance between the two active sites upon dimerisation.

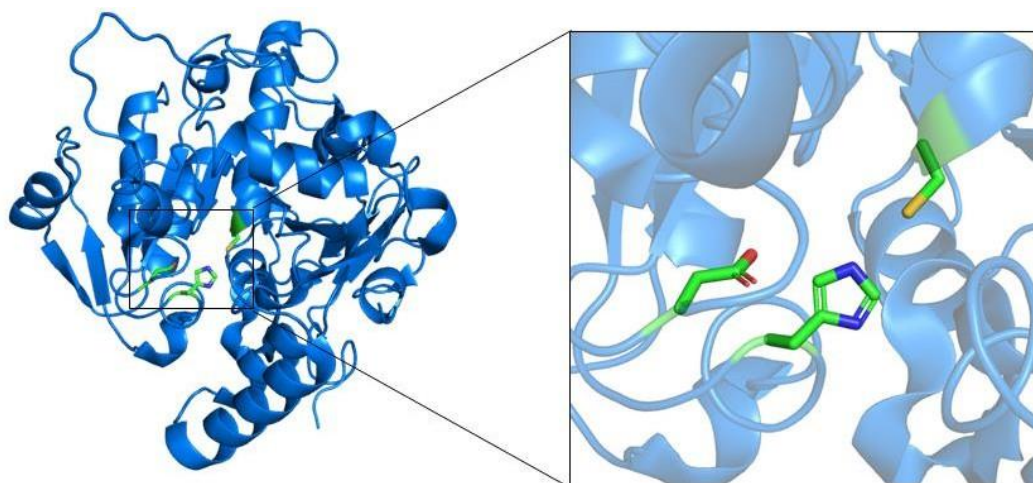


Figure 1.4: The crystallography-determined structure of the catalytic domain of PhaC_a (PDB ID: 5HZ2). Key active site residues are shown in green sticks. The inset shows the orientation of the residues that make up the catalytic triad (C319, D480 and H508).

PhaC_a and PhaC_b have been well characterised recently (Chek *et al.*, 2017; Wittenborn *et al.*, 2016). They share an amino acid homology of 46%, the relative position of the Cys-His-Asp catalytic triad is well conserved between the two enzymes and the subdomains resemble one another (Chek *et al.*, 2017). This group also proposed that the disparity in amino acid sequence is due to potential different modes of dimerisation. The differences in sequence homology instilled hope that the two enzymes provide separate pathways to the final goal of aromatic polymerisation.

Wittenborn *et al.* (2016), established both the entry and egress routes of the substrates and products involved in PHA production. Hydrogen bonding between conserved residues of PhaC synthases is thought to open the egress channel for product extrusion. Due to the proposed aromatic substrates being larger (140 – 180 g/mol compared to 100 g/mol), mutations in these residues may be required to confer a more suitable exit route for the polymerized chain. **Figure 1.4** displays the active site of PhaC.

1.2.1 Reconstruction of an Ancestral PhaC

Ancestral reconstruction (ASR) involves building a phylogenetic tree of modern genes, and using this tree to infer the sequences at ancient nodes (Thornton, 2004). The inferred sequences are often synthesised and recombinantly expressed in a laboratory setting to determine the ancient catalytic properties (Hobbs *et al.* 2015). Using this method, it is possible to reproduce enzymes from as early as the Precambrian era which Gaucher *et al.* (2003) showed to be more thermostable than extant forms. More recent common ancestors may also be devised from ASR as shown by Gumulya *et al.* (2018) where they produced a thermostable ancestral CYP3 enzyme. These ancestral enzymes may then prove to be decidedly better starting points for enzyme evolution of biocatalysts due to favourable attributes such as thermostability and increased activity.

In this project, I worked with an ancestral PhaC enzyme that had previously been reconstructed by Dr Emma Andrews (University of Waikato). She built a phylogenetic tree of enzymes with >65% amino acid sequence identity to PhaC_b. The sequence of the last common ancestor of these enzymes was inferred using maximum likelihood, as described previously (Hobbs *et al.* 2015). The ancient enzyme, PhaC_b-LCA, could be solubly expressed at high levels and was highly active *in vitro*. It is postulated that this enzyme will provide a more suitable starting point for directed evolution experiments due its high stability and activity.

1.3 Directed Evolution

The goal of this research was to use directed evolution to find PhaC variants with altered substrate specificities and/or higher activities. Directed evolution is a commonly used engineering strategy for improving biochemical functions of proteins through rounds of mutation and selection/screening (Bloom and Arnold, 2009). Romero and Arnold (2006), describe directed evolution as a biological optimisation process, as they believe these molecules are not at the end of their evolutionary trajectories. Using this technique better enzymes can be determined as it is just a matter of selecting and screening for the right properties. It has proved particularly effective for engineering biotechnologically relevant enzymes. For example, Savile *et al.* (2010) increased the production of the antidiabetic compound, sitagliptin, by combining *in silico* design and directed evolution for the engineering

of an enzyme to gain altered activity. The initial production scheme was an expensive and inefficient process. Through using directed evolution, a more economical and environmentally friendly alternative was created.

Error-prone PCR acts through random generation of mutations in copies of a single gene, producing a large and diverse library of mutants (Abou-Nader and Benedik, 2010). Large and diverse libraries of mutants are generated through this technique, some of which may have the desired catalytic ability. epPCR procedures are modified versions of standard PCR techniques that enhance the natural error rate of the polymerase, generating random mutations (Cirino *et al.* 2003). Sylvestre *et al.*, (2006) stated that the mutations are spread over the entire sequence when performing epPCR. This may seem counterintuitive when trying to alter the specificity of the active site, but often distal mutations from the active site can generate a similar or improved effect opposed to closer mutations (Morley and Kazlauskas, 2005).

There are three fundamental requirements for directed evolution experiments: 1) introducing a suitable amount of variation into the gene of interest, 2) obtaining a large library of variants and 3) designing an appropriate screen/selection process.

In regard to mutation rate, there is balance that needs to be achieved. A high rate produces distinct variants but yields a large number of truncated proteins. A mutation rate that is too low results in a library with less distinct full-length variants and a number of wild-type sequences. Patrick and Firth (2008), describe the optimal mutation rate as giving a maximally diverse library that contains a large number of full-length distinct variants. To obtain a large library, restriction cloning and transformation must be optimised. To achieve this, large scale restriction digests must be performed providing high quality vector DNA (Copp *et al.*, 2014). Ideally a small streamlined plasmid will be used for efficient transformation into the host cell. The development of a suitable screen or selection is often posited as the most difficult part of a directed evolution experiment. A selection is designed to isolate variants with a specific activity and is often 'life-or-death' making them very stringent in what you get out. A screen is more liberal and views the whole library for a specific activity and requires controls as a threshold for selecting variants with increased activities.

Arnold (1996) explains that epPCR can generate a virtually unlimited number of mutants and hence a high-throughput selection strategy is required and must be designed for this field of research. In regard to the first aim of my project, native PhaC cannot incorporate the proposed aromatic hydroxy acids into the main-chain of

the polymer it currently produces, as these substrates will be toxic to the bacteria harbouring this enzyme. A selection process must be employed that singles out variant PhaC enzymes capable of polymerising the toxic aromatics into a non-toxic polymer. For my second aim, a quantitative screen was developed to find variants that were capable of producing more PHA than their wild-type predecessor.

Wittenborn *et al.*, (2016), established both the entry and egress routes of the substrates and products involved in PHA production. Hydrogen bonding between conserved residues of PhaC synthases is thought to open the egress channel for product extrusion. Due to the proposed aromatic substrates being larger (140 – 180 g/mol compared to 100 g/mol), mutations in these residues may be required to confer a more suitable exit route for the polymerized chain. **Figure 1.1** displays the active site of PhaC.

1.4 Genetic Selection for Altered Specificity of PhaC

Genetic selections are a way of isolating variants in a library through linking an activity of interest to the survival of the organism (Packer and Liu, 2015). In this study, I made use of the fact that aromatic hydroxy acid monomers, such as 3-hydroxybenzoic acid, are toxic to *E. coli*. Evolved PhaC variants, capable of detoxifying these substrates by polymerising them, should enable *E. coli* to survive at increased concentrations of the substrates.

In order to implement the genetic selection, another important consideration is the activation of the substrate. In the native pathway for PHA biosynthesis, acetyl-CoA is converted into 3-hydroxybutyryl-CoA through the actions of the enzymes encoded by *phaA* and *phaB* (**Figure 1.3**). For PhaC to polymerise aromatic substrates, it is essential to co-express a CoA ligase enzyme that can first activate the substrate.

Three CoA ligases were chosen due to their activity toward the aromatic hydroxy acid substrates: 4-coumarate CoA ligase (4CL), Hydroxybenzoate CoA ligase (HB) and phenylacetate CoA ligase (PCL). A paper by Yuan *et al.*, (2014) describes the specificity of 4CL isolated from the plant *Lonicera japonica* (or Japanese honeysuckle). It is thought to have the ability to prime two of the potential substrates *p*-coumaric acid and hydroxycinnamic acid. HB was isolated from the bacteria *Thauera aromatic* and has specificity toward 3-hydroxybenzoic acid and 4-

hydroxybenzoic acid (Laempe *et al.*, 2001). PCL was isolated from *Penicillium chrysogenum* and was included to activate 3-hydroxy phenylacetic acid (Koetsier *et al.*, 2009).

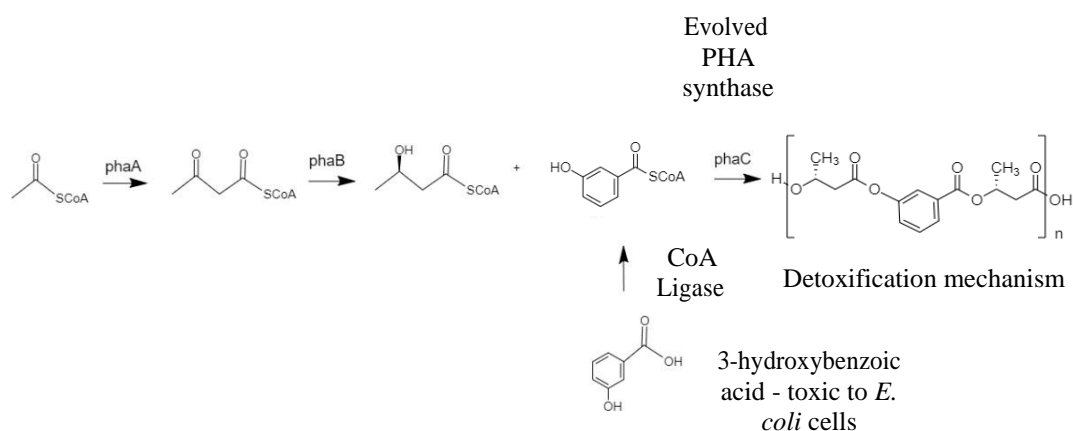


Figure 1.5: Aromatic hydroxy acid substrate polymerisation. Figure showing the proposed reaction mechanism for inclusion of aromatic substrates for polymerization by an evolved PhnC enzyme. Evolved PHA synthase are variants from epPCR libraries capable of aromatic polymerisation.

1.5 Screen for Highly Active PhnC Variants

Screening is a powerful tool that is often coupled with epPCR experiments to isolate enzymes with altered substrate specificity activities (Dougherty and Arnold, 2009). The first rule of thumb when designing a screen or selection is ‘you get what you screen for,’ (Schmidt-Dannert and Arnold, 1999), so it is paramount they are directly for the property of interest. In this project a functional screen was designed to isolate variants from epPCR libraries that were capable of producing PHAs at a higher efficiency than wild-type. The screen was based on the use of a fluorescent dye, Nile red.

Nile red (9-diethylamino-5H-benzo[*a*]phenoxazine-5-one), is the oxidised form of the basic oxazine Nile blue A. It is a lipophilic dye that stains intracellular lipid deposits in hydrophobic environments and was first documented to stain lipid droplets more than thirty years ago (Greenspan *et al.* 1985). More recently, the dye has been reported to have high sensitivity toward staining PHAs accumulating inside bacterial cells (Speikermann *et al.* 1999). This paper was influential in enabling the identification of PHA-producing bacteria without the use of expensive and time

intensive processes such as transmission and scanning electron microscopy and gas chromatography (TEM, SEM and GC).

Nile red is an intensively fluorescent stain in hydrophobic environments making it ideal for visualisation of hydrophobic PHA inclusions in bacteria. Nile red-stained PHA inclusions were shown by, Zuriani *et al*, (2013), to have excitation values between 520 and 550 nm, and emission wavelength between 590 and 630 nm. This paper provided a novel method for staining liquid cultures of PHA producing cells and the quantification through fluorescent analysis. I modified this technique to become a high-throughput approach for library screening.

1.6 Aims of this Thesis

The overall aim of this thesis was to alter the substrate specificities and/or the activities of three PhaC enzymes: PhaC_a, PhaC_b and PhaC_b-LCA using directed evolution.

The first specific aim was to use epPCR to create a library of variant PhaC_a genes and to isolate mutants capable of polymerising aromatic hydroxy acid substrates. This also required construction of a plasmid containing the necessary CoA ligases for priming of the aromatic hydroxy acid substrates for polymerisation. Minimum inhibitory concentrations needed to be determined for the subsequent selection process and lastly, the genetic selection protocol for isolating desired mutants was implemented. Progress toward these aims is described in **Chapter 3**.

The second specific aim was to create two further epPCR libraries of the highly active PhaC_b enzyme. The same genetic selection was implemented and selected mutants were sequenced. These experiments are the subject of **Chapter 4**.

The third specific aim was to create two final epPCR libraries using the *phaC_b-LCA* gene as the template. A screen was designed and used to identify mutants capable of accumulating PHA at a higher rate than the wild-type enzyme. This research is described in **Chapter 5**.

By completing these aims I hoped to gain functional PhaC variants capable of either the production of aromatic bio-polyesters or increased native PHA accumulation.

Chapter 2

Methods and Materials

2.1 Regulatory Approval

The development of all genetically modified organisms was approved under approval number GMD102337 by the Environmental Protection Authority of New Zealand. All work was performed under PC2 conditions and all procedures were carried out at room temperature unless stated otherwise.

2.2 Materials

Chemicals in this study were purchased from Merck (USA), unless stated otherwise. Molecular biology enzymes and associated buffers were purchased from New England Biolabs (NEB, USA), unless stated otherwise. All reagents used in this project were of analytical grade. Solutions, media and antibiotics were prepared with de-ionised water, hereafter described as Milli-Q, and sterilised by autoclaving (121 °C, 100 kPa for 20 min), or sterile filtering. All microbiology experiments were conducted using several strains of *Escherchia coli* and plasmids. These are displayed in **Table 2.1** and **Table 2.2**, respectively. The antibiotics used in this project were added to a final concentration of 100 µg/mL, apart from kanamycin which was 30 µg/mL.

Table 2.1: Characteristics of *E. coli* used in this project

Strain	Genotype	Reference
<i>E. coli</i> E cloni 10G	F- <i>mcrA</i> Δ(<i>mrr</i> - <i>hsdRMS</i> - <i>mcrBC</i>) <i>endA1 recA1</i> Φ80d <i>lacZ</i> Δ <i>M15</i> Δ <i>lacX74</i> <i>araD139</i> Δ(<i>ara, leu</i>)7697 <i>galU galK rpsL nupG λ-tonA</i>	Lucigen Inc.
<i>E. coli</i> BL21(DE3)-gold	<i>ompT gal dcm lon</i> <i>hsdSB</i> (rB- mB-) λ(DE3 [<i>lacI</i> <i>lacUV5-T7 gene 1 ind1</i> <i>sam7 nin5</i>])	Agilent Inc.

Table 2.2: Characteristics of plasmids used in this project

Plasmid Name	Resistance Marker	Obtained from
pGETS-MH	ampicillin	Scion, Rotorua, NZ.
pCDF-HisB-C	spectinomycin	Novagen Inc
pBluescript II KS+	ampicillin	Alting-Mees and Short, 1989
pGETS-Ligase	ampicillin	Scion, Rotorua, NZ. (Modified by Alex Kane)
pCDF-PhaC _a	spectinomycin	Genscript, USA
pGETS-CBA	ampicillin	Scion, Rotorua, NZ.
pET28a-RnI1	kanamycin	Llana Gerber, Massey University.
pET28a-PhaCb	kanamycin	Scion, Rotorua, NZ.

2.3 Methods

2.3.1 Media

Luria-Bertani (LB) medium was used for all experiments. Liquid medium contained pre-formulated LB powder dissolved at 25 g L⁻¹, (ForMedium, Hunstanton, UK). LB-agar plates were prepared adding agar (15 g L⁻¹, ForMedium), autoclaving, cooling the mixture to ~55 °C, adding antibiotic if required and pouring into sterile Petri dishes using aseptic technique.

2.3.2 Freezer Stocks

All bacterial cell lines used in this study were stored at -80 °C. Stocks generated in this study were created by combining overnight cultures 4:1 with sterile 80% v/v glycerol, and stored in 1.5 mL cryogenic vials (Thermo Scientific, USA). These stocks were used for inoculating liquid media or streaking plates.

2.4 DNA Techniques

2.4.1 DNA extraction and purification

Plasmid extraction was carried out utilizing either the NucleoSpin plasmid prep kit (Macherey-Nagel, Germany) or the Monarch plasmid mini-prep kit (NEB, USA), following the manufacturer's guidelines. More efficient elution was achieved by heating the elution buffer to 70 °C, adding it to the spin column, and incubating for five min at before elution. PCR amplicons were purified using either the PureLink PCR purification kit (Thermo Scientific, USA) or the Nucleospin Gel and PCR clean-up kit (Macherey-Nagel, Germany). Extraction of DNA from agarose gels was performed using the Nucleospin Gel and PCR clean-up kit (Macherey-Nagel, Germany), following the manufacturers guidelines.

2.4.2 Primer Design and Storage

All primers were designed using Snapgene 4.0 (GSL BioTech, USA). Primers used in this study were synthesised by Integrated DNA Technologies (IDT, USA) or Macrogen Inc (South Korea). For long-term storage, primers were re-suspended in TE buffer (10 mM Tris-Cl pH 8.0, 0.1 mM EDTA) to a final concentration of 100 µM and stored at -20 °C. For working stocks, aliquots were diluted in Milli-Q to a final concentration of 10 µM.

2.4.3 Primers Used

Table 2.3 Primers used in the project

Primer Name	5' → 3'
PhaC1.For	GCGTGTTGACTATTTTACCTCTGGCG
PhaC1mid.For	GCTGATTGAAAGCGGTGGTGAAAGC
PhaC1.Rev	CTGACCGCTTTCTTTCCGAGGCAG
Lambda_PR.For	Phos - ATGTACTAAGGATCCTGTTGTGAGAGAGAC
Lambda_PR.Rev	Phos - GCAACCATTATCACCGCCAGAGG
Pre-phaC_seq.Rev	CTTTTACCTTCTTGGGTGCTTGC
pGETS_CoA.For	Phos - CTCGAGTGCCAATCAAGGAGTG
pGETS_CoA.Rev	Phos - TTAGTACATGCAACCATTATCACCGC
PhaC_XhoI_Gibson.Rev	AGCCTAGGTAACTCGAGTGTTATGCTTTTGCTTTAAC
ChrPhaC_lig.For	TAGGATCCTGTTGTGAGAGAGACAATCAAATCATGCAGCA GTTTGTTAATAGCCTG
PhaC1 seq 1592R (WK)	CACCATTCCAAAACAGCAGA
BamHI_Anc_ChrPhaC.For	TAAGGATCCTGTTGTGAGAGAGACAATCAAATCATGAGCC AGGATACCCAGG
Anc_ChrPhaC.Rev	CTCCGAGCTCCTCGAGTTAAC
pGETS_PhaB.Rev	CATGCCGCCGGTCACATACGC
T7.Rev	GCTAGTTATTGCTCAGCGG
T7.For	TAATACGACTCACTATAGGG
pET_Upstream.For	ATGCGTCCGGCGTAGA

2.4.4 Agarose Gel Electrophoresis

For size quantification of DNA products, 1 – 10 μ L were run on 0.8 or 1 % agarose gels (0.8 or 1% w/v agarose in TAE buffer), dependent on the size of the DNA product. Gels were set and submersed in TAE buffer in an electrophoresis chamber (Bio-Rad, USA). DNA samples, along with 1x KAPA loading dye (Kapa Biosystems, USA), were added into wells and run at 110 V, 400 A for 25 min. DNA was stained with 0.01 % (v/v) ethidium bromide and visualized under UV light.

Kapa Universal Ladder (Kapa Biosystems, USA) was used as a reference for DNA size.

2.4.5 High-fidelity PCR

Phusion DNA polymerase (NEB, USA) was used for all high-fidelity PCRs. Reactions were set up in 1 x HF or GC buffer, with 0.5 μ M of both forward and reverse primers, 200 μ M dNTPs, 3% DMSO and 1 unit of Phusion polymerase in a total volume of 20 or 50 μ L. The amount of template DNA added ranged from 1 – 20 ng. PCR was performed using a C1000TM thermal cycler (BioRad, USA), following the settings in **Table 2.3**. Products from PCR reactions were visualized following the agarose gel electrophoresis protocol in section **2.4.4**.

Table 2.4: Table showing the general settings for thermal cycler PCR reactions.

Step	Temperature (°C)	Time (Min)
Initial Denaturation	98	2:00
Denaturation	98	0:20
Annealing	50 – 70	0:30
Extension - 25 cycles	72	0:15 – 2:00 (depending on size)*
Final Extension	72	5:00

*Extension time per kb is 15 – 30 sec.

2.4.6 Error-prone PCR

Error-prone mutant libraries were generated using the GeneMorph II Random Mutagenesis Kit (Agilent Technologies, USA), which employs Mutazyme II DNA polymerase. Standard epPCR thermal cycler settings were followed (**Table 2.4**), in the C1000TM thermal cycler (BioRad, USA) and the procedure was taken from “Hanson-Manful *et al.* (2013)”. Amplicon size and concentration was verified using agarose gel electrophoresis and the DNA was purified (sections **2.4.1** and **2.4.4**). Cleaned amplicons were cloned into designated backbone vector via restriction digestion, ligation and transformation into appropriate strain of *E. coli* (sections **2.4.10**, **2.4.11** and **2.4.12**).

Table 2.5: Table showing the general settings for thermal cycler epPCR reactions.

Step	Temperature (°C)	Time (Min)
Initial Denaturation	98	1:00
Denaturation	95	0:30
Annealing	55 – 65	0:30
Extension (15-60 sec/kb) - 30 cycles	72	1:00 – 2:00 (depending on size)
Final Extension	72	5:00

2.4.7 Colony PCR

Colony PCR utilized the KAPA Taq ReadyMix from KapaBiosystems. Colonies were picked from selective agar plates and placed into 50 µL of Milli-Q. These were then incubated at 95 °C for 10 min to lyse cells and allow access to DNA for amplification. 20 µL PCR reactions were made up as follows: 10 µL of KapaTaq ReadyMix, 1 µL of forward and reverse primers, 1 µL of lysed cell suspension and 7 µL of Milli-Q. Thermal cycler setting were the same as **Table 2.4**; although extension time of the Taq polymerase was 0:30 – 1:00 min/kb.

2.4.8 Whole-circle PCR

Whole-circle PCR was carried out in the C1000™ thermal cycler (BioRad, USA) and Phusion polymerase (NEB, USA). Whole-circle PCR required the use of 5' phosphorylated primers, for subsequent ligation steps to work. The reactions were set up with the PCR protocol (Section 2.4.5 and **Table 2.4**). Extension times were increased depending on the size of the plasmid being amplified. DpnI digestion was performed to degrade methylated DNA (i.e. the plasmid template) at 37 °C for 1 hour and 65 °C for 20 min to inactivate the enzyme. Ligation steps followed by transformation were then performed (**Sections 2.4.11 and 2.4.12**).

2.4.9 Sequencing

All Sanger sequencing in this project was performed by Macrogen Inc. (Seoul, South Korea). Sequencing results for this project were visualised using Snapgene 4.0 (GSL BioTech, USA).

2.4.10 Restriction Digests

Plasmid DNA and PCR products were digested using restriction endonucleases following the manufacturers guidelines. Reactions typically contained ~1 µg of DNA, 0.5 µL of appropriate enzyme (10 Units), and 1x CutSmart buffer made up to a volume of 20 µL. Reactions were incubated in a thermal cycler at 37 °C for ~1 hour and heat inactivated at 65 °C for 20 mins where applicable. Digested fragments were visualised by agarose gel electrophoresis and bands of interest were isolated by gel extraction (Section 2.4.4 and 2.4.1). DNA fragments were stored at 4 °C if not used immediately. Digestion with DpnI was also performed on PCR products generated from plasmid DNA. These reactions typically contained ~1 µg of DNA, 1 µL of DpnI (20 Units), and 1x CutSmart buffer made up to a volume of 20 µL. Reactions were incubated in a thermal cycler at 37 °C for ~1 hour and heat inactivated at 65 °C for 20 mins.

2.4.11 Ligation Reactions

DNA fragments with sticky ends were ligated using ~200 U of T4 DNA ligase, 1x ligation buffer and a 3:1 molar ratio of insert to vector made up to a volume of 10 – 20 µL. Ligations were incubated overnight at 16 °C and then heat inactivated at 65 °C for 10 min. DNA was stored at 4 °C if not used immediately.

2.4.12 Electro-competent Cells and Transformation

This protocol is modified from Sambrook & Russell (2001). Bacterial strains were grown on LB agar plates, supplemented with antibiotic if necessary, at 37 °C overnight. A single colony was picked from the plate to inoculate a 5 ml culture in LB broth, supplemented with antibiotic if necessary. This was incubated overnight at 37 °C in a shaking incubator set at 200 rpm.

Five-hundred millilitres of LB broth were inoculated with the 5 ml culture and the growth was monitored using a spectrophotometer at 600 nm until the optical density (OD) reached between 0.35 and 0.4.

The 500 ml cell culture was decanted into ten 50 ml tubes that were pre-cooled on ice. The cells were then chilled for 10 mins and pelleted using a Sorvall Lynx 6000 centrifuge (Thermo Scientific, USA) at 1800 x g for 10 mins at 4 °C. The

supernatant was discarded and pellets were pooled into six 50 ml tubes, re-suspended in 50 mL of chilled autoclaved Milli-Q and pelleted again. This step was repeated pooling the pellets into four 50 ml tubes. The supernatant was discarded and the pellets were pooled into one 50 ml tube and re-suspended in chilled 10 % (v/v) glycerol. The supernatant was discarded and the cells were re-suspended in ~1 ml of 10% (v/v) glycerol. Aliquots of 50 μ l were transferred into microcentrifuge tubes and immediately stored at -80 °C.

Aliquots of 50 μ l of electro-competent *E. coli* cells were thawed on ice. Once thawed, 1-100 ng of purified DNA, was pipetted into the cells. The DNA and cell mixture was pipetted into a pre-chilled 0.2 mm-gap electroporation cuvette (Bulldog-Bio, USA). Cells were then pulsed at 2500V, 25 μ F and 200 Ω , followed by the immediate addition of 500 μ L of room temperature SOC or LB media. Cells were transferred to sterile test tubes and recovered in the shaking incubator at 37 °C for 1 hr at 200 rpm before plating on LB-agar containing respective antibiotic.

2.5 Determination of Minimum Inhibitory Concentrations

2.5.1 MIC Set-up

Minimum inhibitory concentrations were found through the following method (Wiegand *et al.*, 2008). Firstly, 5 ml aliquots of LB broth were inoculated with the test strain with appropriate antibiotic and grown in a shaking incubator overnight at 37 °C and 200 rpm. A 70-mM stock solution was made of each aromatic hydroxy acid by dissolving it in LB broth. This concentrated stock was the starting point for a 2-fold serial dilution series, covering 12 concentrations from 70 Mm to 0.038 Mm. Once dilutions were made, 0.990 ml was transferred to each respective well in a 96 deep-well plate using a multichannel pipette. 5 μ L of the grown culture was added into each well. The plate was then sealed with a breathable membrane and incubated overnight at 37 °C in the plate shaker at 300 RPM. Once overnight incubation had occurred, the plate was spun down using the Eppendorf 5804 centrifuge at 1900 x g for 30 min. The Supernatant was discarded and pellets were examined and photographed. The MIC was defined as the point where no pellet was formed after centrifugation.

2.5.2 LB-agar plate MIC

Minimum inhibitory concentrations were also determined by an LB-agar plate method. LB agar was made up containing respective hydroxyl acid substrate at increasing concentrations plus necessary antibiotic. A cell suspension of *E. coli* 10G containing correct wild-type PhaC enzyme was spread onto the prepared LB-agar. Plates were incubated overnight at 37 °C and analysed the next day for the cut-off concentration that inhibited bacterial growth.

2.6 Selection Strategy

2.6.1 Library Plating

All libraries were spread onto agar containing either hydroxyl acid substrates or Nile red and glucose. *E. coli* containing epPCR libraries were plated to ensure that ~5 copies of each variant was present, *e.g.* for a library containing 100,000 variants, 500,000 cells from a culture were plated. This over-sampling was to ensure that, on average, every variant in the library was represented at least once on the plate. To calculate the number of cells from a culture to spread, the following method was followed: 1-5 ml of LB broth with appropriate antibiotic was inoculated using the correct freezer stock. Cells were grown at 37 °C in a shaking incubator at 200 RPM for 1-3 h until the optical density at 600 nm was ~ 0.1. The following equation was used to determine the amount of culture that contains the desired number of cells: $(2.5 \times 10^8) \times OD_{600} = \text{no. of cells per millilitre of culture}$. Required no. of cells/ no. of cells per millilitre of culture = volume of culture to plate.

2.6.2 Agar-Based Selection of epPCR Libraries

To perform agar-based selections, LB-agar was made containing respective hydroxyl acid substrates at the appropriate concentrations, as well as the antibiotic required for plasmid maintenance. LB-agar was supplemented with IPTG (100 µM) for induction where necessary. LB-agar was then poured into 150 mm Petri dishes, using aseptic technique. Plates were used the same day as the hydroxyl acid monomers can weaken the agar and at high enough concentrations, prevent it from solidifying. Libraries were plated, following library plating protocol **2.5.1**, on increasing

concentrations of the hydroxyl acid substrate plates. The concentration range was based off a pre-determined MIC for cells harbouring the wild type *phaC* genes (determined as described in **2.4.1**). Ten microlitres of a 1:1000 dilution was plated onto an LB-agar plate containing suitable antibiotic as a countable measure to calculate how many cells were actually plated. Plates were incubated overnight at 37 °C and analysed the following day.

2.6.3 Winners and Further Selections

Plates containing a large number of colonies were harvested by pipetting 2 – 4 ml of LB broth and using a glass spreader to scrape the cells from the surface of the plate. Freezer stocks (**2.2.3**) were made of these enriched cells. Further selections were then performed with the enriched stocks following (**2.4.2**) and plating them on higher concentrations of hydroxyl acid LB-agar. Further selections were performed until there were few colonies or no growth on the plates. Individual colonies were picked off high concentration plates where only a few colonies grew and screened by colony PCR (**2.1.7**) for the presence of the gene of interest. The selected mutations in the PCR products were identified by sequencing (**2.1.9**).

2.6.4 Single Plasmid Isolation

To obtain single plasmids from cells containing two plasmids the following procedure was performed. Cells containing two plasmids were inoculated in 5 ml of LB broth with antibiotic appropriate to only the plasmid to be isolated. These were grown overnight in a shaking incubator at 37 °C and 200 RPM. The plasmids were then purified from these cultures (**2.4.1**) and transformed into competent E coli 10G cells (**2.4.12**). Transformed cells were then plated on LB-agar containing antibiotic specific for the plasmid to be isolated. Plates were incubated overnight at 37 °C. The following day, multiple colonies from the plates were picked and spot-plated on two LB-agar petri dishes, one with antibiotic specific for the plasmid to be isolated and one with antibiotic specific to the second plasmid. These were then incubated overnight at 37 °C and analysed the next day. Colonies that grew solely on LB-agar containing antibiotic specific to the plasmid to be isolated were picked for colony PCR screen or cultured for downstream plasmid purification. Colonies that grew on both antibiotic plates were discarded as they must still contain both plasmids.

2.7 Screening for PHA Production

2.7.1 Viable Nile Red Colony Staining

Nile red solution was made up by adding 0.25 mg of Nile red into 1 ml of DMSO and stored at room temperature wrapped in foil. Nile red staining of PHA was achieved using the protocol from Spiekermann *et al.* (1999). Amendments were made by adding 0.22 μ m sterile filtered glucose to a final concentration of 20 g/L and appropriate antibiotic once the LB-agar had cooled to ~55 °C. PHA producing cells were plated or streaked and incubated at 37 °C overnight. Agar plates were exposed to blue light at 470 nm (SafeImager2.0 Blue Light Transilluminator, Thermo Scientific, USA) to visualise the accumulation of PHAs.

2.7.2 Plate Based Nile Red Fluorescence Screen

A two-stage screen was designed that combined agar plate-based screening and 96-well microplate-based fluorescence assays. In the first stage, LB-agar was supplemented with excess glucose (20 g/L) and Nile red (0.5 mg/L) and poured into 150 mm Petri dishes. Fifty thousand cells containing *phaC_b*-LCA variants were spread onto these plates and incubated at 37°C for 16-24 h. The plates were exposed to blue light at 470 nm enabling fluorescent colonies to be identified by eye. In the second stage, the protocol of Zuriani *et al.* (2013) was adapted. Each well of a deep-well, 96-well plate was filled with 1 ml of LB medium that had been supplemented with excess glucose (20 g/L) and respective antibiotic. A total of 92 highly-fluorescent colonies were picked from the library plates, and used to inoculate a 96-well plate for rescreening. The other four wells were used for: positive controls expressing wild type *PhaC_b*-LCA (two wells – A1 and D6); a cells-only negative control (no *PhaC*, well A2); and a control for cross-contamination (growth medium only, well D7). In parallel, an LB-agar Nunc Omnitray (Thermofisher, USA), an agar plate in the shape of a 96 well plate, was poured. Before incubation a 96-well pin replicator was used to inoculate the Omnitray from the deep-well 96 well plate containing the variants as a stock of individual variants. The deep-well plate was covered with AeraSeal film (Excel Scientific, USA) and incubated at 30 °C for ~12 h (400 RPM shaking).

After this incubation, 50 μ l aliquots were taken from each well and placed in a sterile Nunc® 96-well plate containing 250 μ l of LB broth. The OD₆₀₀ of each diluted culture was measured using an Enspire™ 1300 Multilabel Reader (Perkin Elmer, USA).

Next, cells were pelleted by centrifugation at 2,100 x g for 15 minutes. The supernatants were discarded and pellets were resuspended in 300 μ l of water supplemented with Nile red (3.1 μ g/ml). The plate was left to incubate at room temperature, in the dark, for 30 min. After this incubation, the cells were re-pelleted (2,100 x g, 15 min), the supernatants were discarded and the pellets resuspended in 300 μ l water. The Nile red fluorescence intensity in each well was measured with excitation at 535 nm and emission at 605 nm. Raw fluorescence readings were corrected for the OD₆₀₀ readings taken above.

Chapter 3

Directed Evolution of PhaC_a

3.1 Introduction

In this chapter, error-prone PCR (epPCR) and a genetic selection were used in an attempt to alter the substrate specificity of PhaC_a. As discussed in **Section 1.2**, this particular PhaC enzyme (from *Cupriavidus necator*) was chosen because of its large substrate entry channel (Wittenborn *et al.*, 2016) which would potentially allow access of bulky aromatic substrates.

The life-or-death selection was based on finding PhaC variants that could detoxify aromatic hydroxy acids by polymerising them. Before Libraries could be constructed and selected from, the first goal of this chapter was to determine minimum inhibitory concentrations (MICs) of the potential aromatic hydroxy acid substrates.

3.2 MIC Analysis

3.2.1 MIC Determination

The minimum inhibitory concentrations (MICs) of six aromatic hydroxy-acid substrates (**Table 3.1**) were tested using the cloning strain of *E. coli* known as E. cloni 10G (**Table 2.1**) harbouring the wild-type PhaC_a enzyme. The wild-type enzyme was not posited to have any effect on the MIC value as it cannot natively polymerise the toxic aromatic substrates into a non-toxic polymer. However, it was included to as a control for downstream selections so in both experiments PhaC_a was constitutively expressed. MICs were determined in a two-stage process. First, a wide range of concentrations were tested, along a doubling dilution series (**Figure 3.1**). This suggested that the MIC for each aromatic hydroxy-acid was between 4.4 and 17.5 mM. Next, each MIC was determined more precisely by testing each substrate at 1 mM increments. The results are presented in **Table 3.2**.

The MIC data suggested that *E. coli* is sensitive to these aromatic hydroxy-acid substrates at relatively low concentrations. The lowest MIC was found for salicylic acid at 5 mM. A practical benefit of a low MIC is that it allows a lower amount of substrate to be used in the downstream life-or-death selection process.

Table 3.1: Table showing six potential aromatic hydroxy acid substrates

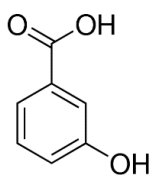
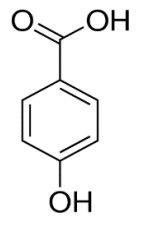
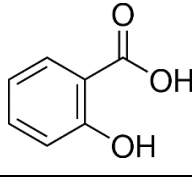
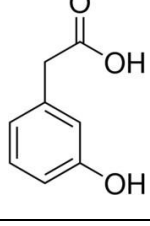
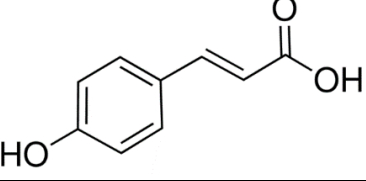
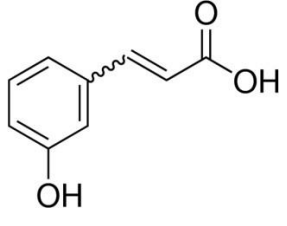
Name (Molecular Weight)	Structure
3-Hydroxybenzoic acid (Mol wt. 138.12)	
4-Hydroxybenzoic acid (Mol wt. 138.12)	
Salicylic acid (Mol wt. 138.12)	
3-Hydroxy Phenylacetic acid (Mol wt. 152.15)	
<i>p</i> -Coumaric acid (Mol wt. 164.16)	
<i>trans</i> -3-Hydroxycinnamic acid (Mol wt. 164.16)	



Figure 3.1: Initial MIC determination using a doubling dilution series. The example shown is 3-hydroxybenzoic acid with the concentration of the hydroxy acid substrate displayed in mM over the respective well. Cell pellets after overnight incubation are clearly visible in the wells with less toxic substrate, as the concentration increases the cell pellets decrease until no growth occurs.

In addition to MICs, **Table 3.2** also shows the pH of each hydroxy-acid at 17.5 mM in un-buffered solution. I hypothesised that the strength of the acid might play a role in its ability to inhibit the growth of *E. coli*. However, I found that the pH at 17.5 mM did not correlate with the MIC for each substrate, showing that the strength of the acid is not the only determinant in the inhibition of *E. coli* growth. Instead, it appears that the structures of the aromatic hydroxy-acids are responsible for their differing levels of toxicity.

Table 3.2: MICs for six aromatic hydroxy acid substrates

Substrate	MIC (mM)	pH @ 17.5 mM
3-Hydroxybenzoic Acid	8	4.3
4-Hydroxybenzoic Acid	9	4.6
Salicylic Acid	5	4.2
3-Hydroxy Phenylacetic Acid	10	4.4
<i>p</i> -Coumaric Acid	8	4.7
<i>trans</i> -3-Hydroxycinnamic Acid	7	4.5

3.3 Construction of Ligase-only Plasmid

The six substrates require non-native CoA ligases for priming (**Figure 1.5**). Accordingly, a plasmid was designed that encoded the three CoA ligases: 4-coumarate CoA ligase, hydroxybenzoate CoA ligase and phenylacetic CoA ligase. These together have specificity toward all six potential aromatic substrates, and was termed pGETS-Ligase (**Table 2.2**).

The pGETS-Ligase plasmid was constructed by whole-circle PCR. pGETS-MH, received from Scion (Rotorua, NZ), was used as the template for the whole-circle PCR. This plasmid contained *phaCa* along with the genes encoding the three ligases, under the simultaneous control of the constitutive lambda promoter. Phosphorylated primers, pGETS_CoA.for and pGETS_CoA.rev (**Table 2.3**), were designed to remove *phaCa* by amplifying the backbone of pGETS-MH either side of the gene. The PCR was designed so the ligase genes would still be under the constitutive control of the lambda promoter.

The whole-circle PCR yielded the desired product (**Figure 3.2**). The ligase-only plasmid is 8,630 bp. There is not a large difference in size between this and the original pGETS-MH plasmid (10,432 bp), in the agarose gel (**Figure 3.2**). This is due to the pGETS ligase-only PCR product being linear as opposed to pGETS-MH which is a circularised plasmid. In gel electrophoresis, it is known that linear DNA runs slower than circularized DNA (Shin and Day, 1995). Circularisation of the PCR product was performed by ligation and this was then transformed into *E. coli* 10G (**Section 2.4.8**). Three colonies were picked from the transformation plate and the plasmid DNA was prepared from them (**Section 2.4.1**). Then sequencing of the upstream and downstream of the ligation site confirmed the correct plasmid was constructed.

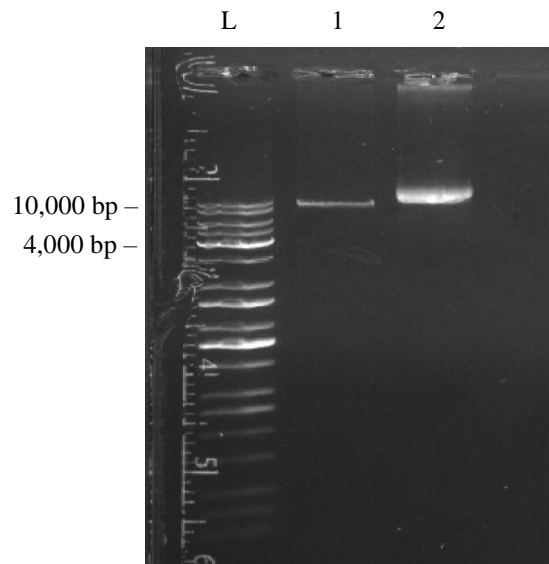


Figure 3.2: Whole-circle PCR to remove *phaC* (1,802 bp) from pGETS-MH. 0.8% agarose gel. L) Kapa universal ladder with size reference on the left, 1) Ligase-only pGETS-MH whole circle PCR product (8630 bp). 2) pGETS-MH plasmid (10,432 bp).

3.4 Error-Prone PCR Library Generation

3.4.1 Error-Prone PCR

When constructing epPCR libraries, it is important to optimise the number of mutations in the amplicon of interest - in this case, *phaCa*. The optimal mutation rate is difficult to deduce. A high mutation rate results in a large number of premature stop codons and deleterious mutations yet produces mostly distinct individual

variants. A low mutation rate results in less distinct individual variants and also less premature stop codons and deleterious variants.

The standard way to vary mutation rate in an epPCR is by altering the amount of template DNA in the reaction. The mutation frequency depends on the error-rate of the polymerase and the number of times that each template is duplicated in the reaction (Hanson-Manful and Patrick, 2013). A lower amount of template DNA will undergo a larger number of duplications during the PCR cycles, due to the copying being performed by an error-prone polymerase, it will result in a high mutation frequency. A higher template amount results in fewer duplications, and hence, fewer mutations in the amplicon.

The *phaCa* was amplified from pGETS-MH in epPCRs with three different amounts of template DNA (0.8 ng, 4 ng and 20 ng). The outcome of each epPCR was assessed by running aliquots of the products on a 1% agarose gel (**Figure 3.3**).

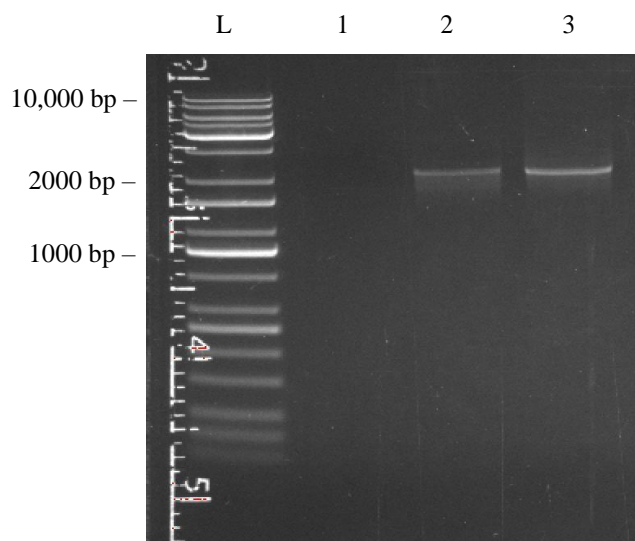


Figure 3.3: *PhaCa* error-prone PCR products. 1% agarose gel; L) KAPA universal ladder with size reference on the left. 1, 2 and 3) epPCR products from 0.8, 4 and 20 ng starting template concentrations respectively.

Unfortunately, the reaction with 0.8 ng template failed and this library was not pursued further. The reaction was not performed again as the consumables are costly and based off the mutation rate of the 4 ng template library (**Table 3.3**), the 0.8 ng template would have produced a very high mutation rate which would have resulted in a number of deleterious mutants.

3.4.2 Error-prone PCR Library Analysis

The epPCR products generated from 4 ng and 20 ng of template DNA were cloned using restriction enzymes BamHI and XhoI into the pCDF-PhaC vector backbone (**Table 2.2**). The ligated products were used to transform *E. coli* strain E. cloni 10G, which already harboured the pGETS-Ligase plasmid. The two plasmid system gave rise to a number of considerations that needed to be addressed. This included the phenomenon known as plasmid incompatibility which arises when two plasmids share common replication controls (Datta and Hedges, 1971). pCDF-phaC_a and pGETS-Ligase plasmids (**Table 2.2**) fortunately contain distinct origins of replication and antibiotic resistance markers, meaning these two plasmids are compatible and will be propagated in the same cell line during replication. Random colonies were then picked and the *phaC_a* genes were sequenced to determine the mutation rate and spectrum of mutations in each library. One library was chosen for further testing and selection based off the mutation rate and bias found in each library.

The library constructed from 4 ng of starting template was predicted to have a higher mutation rate. Sequencing of randomly-chosen clones generated 12,540 bp of useable sequence (corresponding to ~7.1 *phaC_a* genes). The sequencing reads were obtained from an initial 'test' ligation of a portion of the epPCR product. A total of 31 mutations were found over this sequence data resulting in an average of 4.4 mutations per *phaC* gene (**Table 3.3**).

The library generated from the 20 ng starting template is posited to have a low mutation rate due to the higher amount of initial template DNA. Sequencing of the chosen variants generated 7,930 bp of usable sequence (this covered the *phaC_a* gene ~4.5 times). In total, the number of mutations found was 21 which resulted in an average of 4.7 mutations per *phaC_a* gene (**Table 3.3**). Ideally, there would be a higher sequence coverage to ensure that the mutation rate and bias are statistically relevant, although screening by colony PCR and sequencing is a rate-limiting step.

Comparatively, the mutation rate per *phaC_a* gene is similar between the two starting template concentrations - contrary to the theory that a lower starting concentration will generate a higher mutation rate as more doublings occur. It is often hard to predict the mutation rate for clones and this may have arisen due to random chance, or technical error in the method.

There was a notable bias towards mutations at guanine bases (*i.e.* G → N, where N is any of the other three nucleotides) in both libraries. While guanine is the most common base in the *phaCa* gene, making up 30% of the total bases, these mutations are at a higher prevalence than would be assumed by statistical inference. This is indicative of some bias or random chance toward mutations at G bases.

The ratio of transitions/transversions in a perfectly unbiased library should be 0.5, due to there being four possible transition mutations (purine → purine or pyrimidine → pyrimidine) compared to eight possible transversions (purine → pyrimidine or pyrimidine → purine). The ratio reported here is 0.87 which shows only a small amount of bias toward transitions throughout the amplicons.

The most interesting result is the transition/transversion ratio of 1.63 found in the 20 ng library. The 4 ng starting template transition/transversion ratio was 0.87, which is closer to the ideal value of 0.5. A ratio of 1.63 is a clear indicator of bias toward transitions throughout the library and may result in a number of duplicated variants of those already present in the library.

The G:C content of *phaCa* is 50.1% which should, in theory, provide an even mutational spectrum. However, epPCR acts by generating random mutations throughout the amplicon of interest and bias is likely to occur in these experiments. A study undertaken by Rasila *et al.*, (2009) describe all random mutagenesis experiments to be biased toward transition mutations. Ideally, a non-biased method would generate a ratio of 0.5 as opposed to ~1 found experimentally by Rasila *et al.*, (2009). Specific to Mutazyme polymerase, Vanhercke *et al.*, (2005) confirms that there was bias with an equal distribution of transitions to transversions. A bias toward these mutations is likely to create bias in other mutational analysis as well.

Taking this into account, the decision was made to create a full-sized library of the 4 ng starting template library by ligating and transforming the total epPCR product for further analysis and selections.

The 4 ng library data in **Table 3.3** was used as the input to analyse the library with PEDEL-AA (Firth and Patrick, 2008). PEDEL-AA is software that uses the wild-type sequence and mutation data to estimate a variety of parameters about the library. These parameters are summarised in **Table 3.4**, with the total library size of 1×10^6 variants being estimated after large-scale ligation and transformations (**sections 2.4.11** and **2.4.12**). The program estimated that the library of 1×10^6 clones actually contained $\sim 7 \times 10^5$ full-length distinct variants. The other 3×10^5 variants either contained insertions, deletions or premature stop codons or were wild-type *phaCa*, or

were duplicated copies of variants already present in the library.

Table 3.3: Comparison of the mutational spectrums of *phaC_a* variants from the 4 ng and 20 ng starting templates.

Type(s) of mutations	Frequency (4 ng)	Proportion of total (4 ng)	Frequency (20 ng)	Proportion of total (20 ng)
Transitions				
A → G, T → C	5	16.1%	1	4.8%
G → A, C → T	8	25.8%	12	57.1%
Transversions				
A → T, T → A	2	6.5%	1	4.8%
A → C, T → G	0	-	1	4.8%
G → C, C → G	6	19.4%	3	14.2%
G → T, C → A	7	22.6%	3	14.2%
Summary of bias				
Transitions/ Transversions	0.87	N/A	1.63	N/A
AT → GC/GC → AT	0.28	N/A	0.13	N/A
A → N, T → N	7	22.6%	3	14.3%
G → N, C → N	24	77.4%	18	85.7%
Mutation rate				
Mutations per kb	2.5	N/A	2.7	N/A
Mutations per <i>phaC_a</i> gene*	4.4	N/A	4.6	N/A

* The cloned *phaC_a* insert was 1,803 bp.

Table 3.4: Composition of the 4 ng *phaC_a* mutant library as estimated by PEDEL-AA

Property	Estimate
Total library size	1.0 x 10 ⁶
Number of variants with no indels or stop codons	8.6 x 10 ⁵
Mean number of amino acid substitutions per variant	3.1
Un-mutated (wild-type) sequences (% of total library)	3.7%
Number of distinct full-length proteins in the library	7.0 x 10 ⁵

3.5 Genetic Selections for cells with an improved *PhaC_a*

3.5.1 Selections of variant *PhaC_a* enzymes

The agar-based genetic selections followed the correct trend, as concentration increased a lower number of colonies formed on the plates. Three initial plate selections were performed by plating the library onto 0.5x MIC, 1x MIC and 1.5x MIC for each respective aromatic hydroxy acid substrate. Growth was noted after 1 and 2 days' growth. For all substrate plates no growth was seen on 1.5x MIC after 1-day incubation. However, after 2 days' incubation, a low number of colonies had formed on 3-hydroxybenzoic acid, 4-hydroxybenzoic acid, *trans*-3-hydroxycinnamic acid and hydroxy phenylacetic acid. This was consistent for 1x MIC for the same substrates, although the growth rate and health of colonies was improved because of the lower concentration of toxic substrates.

The colony morphology on each plate differed largely. For the most part they were unhealthy, small and irregular colonies. Initially, it was thought to be contaminant or satellite colonies but upon screening, bands correlating to *phaC_a* were confirmed by colony PCR. The high concentration of toxic substrates stresses the cells and is the main reason for slow growth and unhealthy colonies.

A step-wise scheme was implemented so cut-off concentrations could be deduced - *i.e.* the concentration where the libraries can no longer grow. Once this was determined it is much easier to identify a narrow range of concentrations for further selections and whether the library variants are better than wild-type. This was performed as a means to avoid plating on the highest concentration and working my way down – costing time and consumables.

Selections were carried out at increasing substrate concentrations until the library was enriched for the fittest individual clones. The next step was to sequence the *phaC_a* variants that were being expressed in these clones. A complicating factor was that each clone contained both the pCDF-*phaC_a* plasmid and the pGETS-Ligase plasmid. The forward sequencing primer PhaC1.for (**Table 2.2**) annealed equally well to both plasmids, because they both contain the same lambda promoter sequence for controlling expression of their cloned genes. To get around this problem, I developed a method in which only the pCDF-*phaC_a* was isolated (**2.5.4**).

Using this method, I was able to isolate the pCDF-*phaC_a* plasmids that encoded *phaC_a* variants from maximal concentrations of each substrate. The *phaC_a* genes were amplified from these plasmids for subsequent sequencing (**Figure 3.4**).

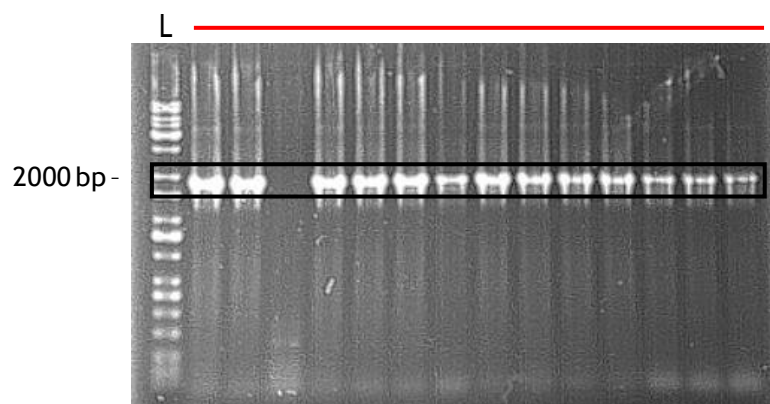


Figure 3.4: 1% agarose gel showing the colony PCR screen. Post-implementation of the single plasmid isolation of the genetic selection ‘winners’. Selected from 3-hydroxybenzoic acid, hydroxycinnamic acid, *p*-coumaric acid and 3-hydroxy phenylacetic acid. L) Kapa universal ladder, red line) indicates lanes containing colonies amplified for *phaC_a*, black box) PCR product (1869 bp).

In total, 30 *phaC_a* variants were sequenced. However, a number of the sequencing results showed multiple chromatogram peaks which could indicate multiple DNA sequences present. In some cases, there were multiple chromatogram peaks and overlapping sequences, the primers did not anneal at the correct point for the sequencing to occur, and large deletions or insertions of ~100 – 300 bp were noted. Two further sequences were wild type. Out of the 30 samples sequenced, only four were candidates to be genuinely improved variants. The details of these four mutants are summarised in **Table 3.5**.

Two of these variants contained the same T deletion in codon 577, resulting in an extended protein containing 11 more amino acids. The same mutation arose even when grown on separate substrates. It will be interesting to express and purify these variants for *in vitro* activity assays to determine if they remain catalytically active. Lastly, the variant S6 from hydroxycinnamic acid yielded short sequencing reads which waned at the 3’ end which meant the potential mutations in these regions were unable to be called.

Table 3.5: Sequencing results of *phaC_a* variants from selections

SEQ ID	Substrate	Mutations
S3	Hydroxy phenylacetic acid	Deletion of T in codon 577; frameshift extends ORF from 589 codons to 600 codons. Forward sequence inconclusive.
S6	Hydroxycinnamic acid	R101L; sequence deteriorates 3' end – may contain more mutations.
S10	Hydroxycinnamic acid	Deletion of T in codon 577; frameshift extends ORF from 589 codons to 600 codons.

3.5 Chapter 3 Discussion

This chapter has described the establishment of a novel genetic selection protocol, the generation of the first epPCR libraries, and implementation of the genetic selection.

The initial goal was to perform MIC testing of the hydroxyl acid substrates on *E. coli* harbouring wild-type PhaC_a enzyme, to find out what concentration is toxic to the cells. I did this in order to gain a threshold for the downstream genetic selections. It was posited that while there are concentrations of hydroxyl acid monomers which are toxic to cells containing wild-type PhaC (which are unable to polymerise these substrates), these concentrations will not be toxic to cells containing the variant PhaC_a enzymes (which will be able to polymerise and detoxify these normally toxic substrates). I established the MIC values for six hydroxyl acid substrates. The MIC values gathered were largely similar, ranging from 5 mM to 10 mM. The inhibition of the aromatic hydroxy acid substrates was not determined by the pH values determined in **Table 3.2**. As the structures overall are similar, I considered that the structures of the substrates may be a factor in their toxicity.

I then designed and constructed a plasmid with the pGETS backbone, encoding only the three ligases that have been shown by our collaborators to successfully ligate CoA onto the proposed hydroxy acid substrates. Achieving this was straightforward using whole-circle PCR and recircularization. In parallel, *phaC_a*

was cloned into a different plasmid, pCDF, by a commercial supplier (Genscript, USA). For the two-plasmid system to work, it was important to have a plasmid containing a compatible origin of replication and a different antibiotic resistance marker. pCDF met these criteria and was therefore chosen.

The primary focus of the chapter was to construct epPCR libraries of *phaCa* and apply the life-or-death selection protocol. I created two libraries using 4 ng and 20 ng of starting template DNA. These contained similar mutational spectra with the goal of carrying out selections on the least biased of the two libraries. The 4 ng library was chosen for further genetic selections and was up-scaled with an estimated 1,000,000 variants. Obtaining a large library was beneficial for subsequent high-throughput selections. A higher number of variants increases the chance for a variant with the desired attributes. Along with this, the PEDEL-AA analysis showed 7×10^5 full-length distinct variants. Drummond *et al.*, 2005 states that the effective library size is the number of unique variants that you can screen/select.

While straightforward in principle, two technical factors complicated the selection protocol. First, I discovered that high concentrations of hydroxy acid substrates (particularly salicylic acid and 4-hydroxybenzoic acid) prevented LB-agar plates from solidifying. This had not been a problem when the MICs were determined, because this was done in liquid growth medium. This was averted by using a higher percentage of agar (2% w/v compared to 1.5% in regular media) when plates were poured. Second, downstream PCR screening of selected variants was difficult because of the two different plasmid in each cell. While this was circumvented by retransformation and patch plating (**Section 3.4.1**), this process was inefficient and time-consuming.

Ultimately, the genetic selection only yielded three candidate sequences (**Table 3.6**). The majority of the sequenced candidates were obvious false positives, including wild-type *phaCa* and sequences that did not align to the pCDF-*phaCa* plasmid at all.

The largest proportion of sequenced candidates encoded truncated PhaC_a variants or empty vectors. Of the three most likely candidates, two shared the same single-base deletion near the 3' terminus of *phaCa*. The resulting frameshift altered the amino acid sequence of the last 12 residues and extended the length of the protein by 11 amino acids. The third candidate only contained one mutation (R101L) in the first 500 codons of the gene, but multiple attempts at resequencing the 3' end were

unsuccessful. When all of the sequencing data are considered, it appears most likely that every sequenced variant was a false positive.

The idea of behind the life-or-death selection was to isolate variants capable of detoxifying the normally toxic aromatic hydroxyl acid substrates by polymerising them. As Arnold (1994) famously stated ‘you get what you screen (or select) for’. At its heart, my selection was not for PhaC_a activity, but for cell growth in the presence of toxic substrates. There were multiple ways for the host cells to achieve this.

One possibility was that the toxic substrates were rendered non-toxic through the action of the CoA ligases, rather than PhaC_a. It is possible that ligation of a CoA molecule onto their carboxyl group, (**Figure 3.5**) may neutralise the toxicity of the substrates. The cells would be under no selective pressure to polymerise the substrates if they are already detoxified through the addition of CoA. However, if this was the case, the accumulation of these CoA derivatives must eventually become toxic for the cells, as there were substrate concentrations above which cells could not grow.

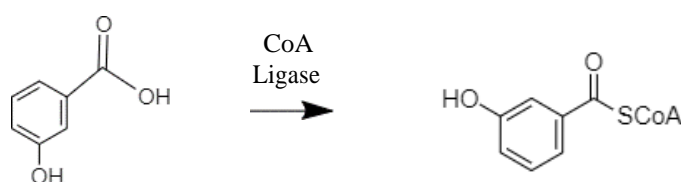


Figure 3.5: One-step reaction showing the addition of a CoA molecule onto 3-hydroxybenzoic acid monomer.

One problem with this explanation is that *E. coli* cells tend to have a low intracellular concentration of CoA molecules and wasting them on this would be detrimental to other critical CoA requiring pathways. Aerobically respiring *E. coli* grown on nutrient rich medium contain free CoA at ~2 mM (Bennet *et al.*, 2009). In the selection scheme, excess of 2 mM of substrates were used and once they are transported inside the *E. coli* cell, there will not be enough free CoA to detoxify all the substrate present. Leading on from this, another argument is that the CoA ligases were not capable of supporting high rates of substrate activation when they were expressed in *E. coli*. In future work, it would be useful to test this experimentally.

It is also possible that the results of the selection were because the over-expression of an unnecessary protein (PhaC_a) caused a reduction in cell fitness (Andrews and Hegeman, 1976; Dekel *et al.*, 2005); and therefore the fittest variants were those that did not express PhaC_a. The *phaC_a* gene was under the control of the

native lambda promoter meaning PhaC_a is constitutively expressed by a strong promoter. Elvin *et al.*, (1990) utilised this bacteriophage promoter as a means for overexpressing proteins in *E. coli* which portrays its high level of expression. Protein biosynthesis is an energy-intensive process and has been estimated to account for ~50% of the energy consumption in rapidly growing bacterial cells (Russel and Cook, 1995).

There is no direct requirement for PhaC_a to be expressed in *E. coli* cells unless it is detoxifying the hydroxyl acid substrates through polymerisation. As mentioned previously, the selection pressure for variant PhaC_a enzymes to do this may have been lost through the addition of CoA molecules onto the toxic substrates. The constitutive expression of PhaC_a - an unnecessary protein - may then have resulted in *E. coli* cells selectively trying to stop or reduce expression.

Overall, it appears that preventing the expression of *phaC_a* is beneficial to the host cells. Throughout the experiments described in this Chapter, including the initial cloning (before epPCR library construction), plasmids accumulated mutations that inactivated the promoter or led to non-functional PhaC_a variants. The life-or-death selection only isolated false positives. I hypothesised that two major contributing factors were the choice of *C. necator* PhaC and its constitutive expression. In the following Chapter, I tested these by re-designing the experimental process, using a PhaC enzyme isolated from *Chromobacterium* sp. USM2 and the t7 controllable promoter.

Chapter 4

Directed Evolution and Selection of PhaC_b

4.1 Introduction

This chapter describes epPCR and genetic selection performed on the enzyme PhaC_b (PhaC from *Chromobacterium sp.* USM2). This was chosen as a secondary route to find a variant PhaC capable of polymerising aromatic compounds. Bhubalan *et al.*, (2011) describes PhaC_b as having PHA accumulation five times higher than that of the previous enzyme PhaC_a, isolated from *Cupriavidus necator*. The high activity and the fact that the structure and mechanism of PhaC_b had been determined made this a good candidate for potential mutagenesis to gain a novel function.

The epPCR libraries and life-or-death selection for improved variants largely followed the same experimental pathway as described for PhaC_a in **Chapter 3**. A key difference was the use of an inducible expression system (pET28), to ensure that PhaC_b was not expressed until cells were in the presence of hydroxy acid substrates. It was hoped this would prevent the highly functional (but mildly toxic) PhaC_b variants dropping out of the libraries during cloning and storage (pre-selection).

4.2 Error-prone PCR Library Generation

4.2.1 Error-prone PCR

In the construction of the epPCR libraries here, I decided to continue with a 4 ng starting template amount due to the favourable mutation rate found in **Section 3.2.2**. It was also decided to make a second library with a starting gene template amount of 2 ng. This was done in an attempt to increase the mutation rate, as a lower amount of template DNA results in a higher number of duplications in the PCR and therefore, more mutations in the amplicon (Hanson-Manful and Patrick, 2013). It is often hard to predict the optimal mutation rate to gain a desired activity. Therefore, creating two libraries with distinct rates provides a higher chance of finding favourable mutants.

As shown in **Figure 4.1**, *phaC_b* was successfully amplified in both epPCRs. The yield of both epPCRs was similar between the two reactions and gave clean bands at the required size.

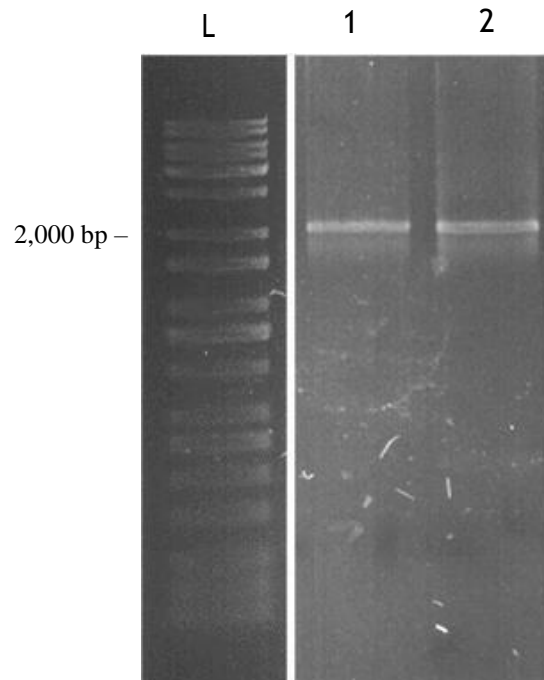


Figure 4.1: Error-prone PCR products of *phaCb*. 1% agarose gel; L) KAPA universal ladder with size reference on the left. 1 and 2) epPCR products of *phaCb* from 2 and 4 ng starting template amounts respectively. 1 μ l aliquots of each 50 μ l epPCR were run on the gel. The break in the gel between lanes L and 1, represents two irrelevant lanes that have been cropped out. Expected product size: 1988 bp.

4.2.2 Error-prone PCR Library Analysis

The epPCR products generated from 2 and 4 ng of template DNA were cloned into the pET28(a) vector backbone using restriction enzymes NdeI and XhoI. The vector was chosen as it contained the IPTG-inducible T7 promoter, adding a control for the expression of *phaCb*. The cloned products were used to transform a DE3 strain of *E. coli* known as BL21(DE3)-gold; a DE3 strain (with the T7 RNA polymerase gene in its chromosome) was needed in order to get expression from the T7 promoter.

The library generated from 2 ng of starting template was predicted to have a higher mutation rate. **Table 4.1** summarises the sequencing analysis of randomly-chosen clones from this library. In total, 18,744 bp of useable sequence data (correlating to ~11 *phaCb* genes) was analysed. This *phaCb* library showed a low amount of bias in the mutational spectrum with a 50/50 distribution of mutations from A/T \rightarrow N and G/C \rightarrow N. The transition/transversion ratio was 1.3 which is higher than the expected value of 0.5 (the value if no bias was present); however, experience with this project (**Table 3.3**) suggests that this value is more often above

0.5 than it is below. Three insertions were also found, suggesting that a portion of the library will contain frameshifts that render variants non-functional.

The library created from 4 ng of starting template was predicted to have a lower mutation rate than the 2 ng starting template library. Randomly picked colonies generated 16,101 bp of useable sequence data (covering the *phaC_b* gene ~9.5 times). As predicted, this library had a mutation rate of 5.1 mutations per *phaC_b* gene, lower than that of the 2 ng starting template library.

The frequency of the different possible mutations is relatively free of bias and suggests a moderately even distribution of mutations over the sequenced genes. There is a slight bias toward GC → AT nucleotide substitutions. Yet, by comparison with the 4 ng and 20 ng *phaC_a* libraries displayed in **Table 3.3**, it contains much less bias toward these mutations. Pleasingly, only one deletion was found over the analysed sequences of the 4 ng template.

The G:C content of the *phaC_b* gene is 47% so a relatively even distribution of the different types of mutations was expected, in the absence of polymerase-induced bias. This was reflected in these results and the remainder of the epPCR product was ligated and transformed to create the full-size library.

The scaled up libraries were both comprised of an estimated 250,000 clones. The mutation data in **Table 4.1** were used to analyse the composition of the libraries (**Table 4.2** and **4.3**) using PEDEL-AA (Firth and Patrick, 2008). The analysis of the 2 ng starting template estimated that there will be around six amino acid changes per variant in the library. Due to the high mutation rate and low starting template amount the number of un-mutated wild-type sequences is low and the number of distinct full length variants is relatively high.

Overall, sequence analysis of the 2 and 4 ng starting template libraries showed that they differed in their mutation rates and spectra, and the 4 ng starting template was also scaled up into a larger library.

The PEDEL-AA analysis showed that the two *phaC_b* libraries have approximately the same number of full length individual variants without stop codons or indels (1.7×10^5). That is, in both cases ~68 % of individual variants are unique, while the remainder are duplicated sequences, truncated sequences, or un-mutated wild-type. The 4 ng library has a higher percentage of wild-type sequences which was anticipated because the mutation rate is lower, meaning that the probability of a clone having zero mutations is higher.

Table 4.1: Comparison of the mutational spectrum of the *phaC_b* variants from the 2 ng and 4 ng starting templates.

Type(s) of mutations	Frequency (2 ng)	Proportion of total [%] (2 ng)	Frequency (4 ng)	Proportion of total [%] (4 ng)
Transitions				
A → G, T → C	21	22.1	8	16.7
G → A, C → T	31	32.6	16	33.3
Transversions				
A → T, T → A	18	18.9	8	16.7
A → C, T → G	7	7.4	3	6.2
G → C, C → G	9	9.5	3	6.2
G → T, C → A	6	6.3	9	18.8
Insertions and Deletions				
Insertions	3	3.2	0	-
Deletions	0	-	1	2.1
Summary of bias				
Transitions/ Transversions	1.3	N/A	1.04	N/A
AT → GC/GC → AT	0.76	N/A	0.44	N/A
A → N, T → N	46	50	19	40.4
G → N, C → N	46	50	28	59.6
Mutation rate				
Mutations per kb	4.8	N/A	3	N/A
Mutations per <i>phaC</i> gene*	8.6	N/A	5.1	N/A

* The cloned *phaC_b* insert was 1988 bp.

Table 4.2: Composition of the 2 ng *phaC_b* epPCR library as estimated by PEDEL-AA

Property	Estimate
Total library size	2.5 x 10 ⁵
Number of variants with no indels or stop codons	1.7 x 10 ⁵
Mean number of amino acid substitutions per variant	5.96
Un-mutated (wild-type) sequences (% of total library)	0.17%
Number of distinct full-length proteins in the library	1.7 x 10 ⁵

Table 4.3: Composition of the 4 ng *phaC_b* epPCR library as estimated by PEDEL-AA

Property	Estimate
Total library size	2.5 x 10 ⁵
Number of variants with no indels or stop codons	1.9 x 10 ⁵
Mean number of amino acid substitutions per variant	3.6
Un-mutated (wild-type) sequences (% of total library)	2.2%
Number of distinct full-length proteins in the library	1.7 x 10 ⁵

4.3 Genetic Selections for Cells with Improved PhaC_b

4.3.1 Basic Wild-Type MIC Test

Having switched host strain from *E. coli* E cloni 10G in **Chapter 3**, to *E. coli* BL21(DE3)-Gold, it was necessary to verify that the new strain was inhibited by the same concentrations of the hydroxy acid substrates. *E. coli* BL21(DE3)-Gold cells containing the pET28(a)-PhaC_b and pGETS-Ligase plasmids were spread on LB-agar that contained increasing concentrations of each hydroxy acid. In each case, the MIC was the same as that shown for E cloni 10G in **Table 3.2**.

With this data in hand, the life-or-death selection experiments could proceed. During selection I supplemented all plates with excess IPTG at 100 μM. This was a high enough concentration for high level expression of *phaC_b*. For each hydroxy acid plate a similar experimental approach to **Chapter 3** was implemented with three

concentrations of 0.8 x MIC, 1 x MIC and 1.2 x MIC. The results from **Chapter 3** provided the insight to utilise a tighter concentration range for initial selections. All substrates, apart from salicylic acid, had growth at 1.2 x MIC concentrations (ranging from 8.4 – 12 mM). 1.5 x MIC plates were also prepared and yet no colonies were present from either of the libraries.

Colonies were then picked from the highest concentration of substrate that had grown, Colony PCR was then used for confirmation that correctly sized *phaCb* inserts were present. Unfortunately, there was a large variation in the yield and purity of these reactions as seen in **Figure 4.2** and hence DNA purification was performed to gain uniform samples to send for sequencing. A reason for the messy PCRs may have been because of a low annealing temperature of the primers

Compared with the experiments in **Chapter 3**, an additional benefit of switching to the T7 promoter was that the upstream sequencing primer no longer annealed to the pGETS-Ligase plasmid. This allowed direct PCR-based screening of the clones from maximal substrate concentrations.

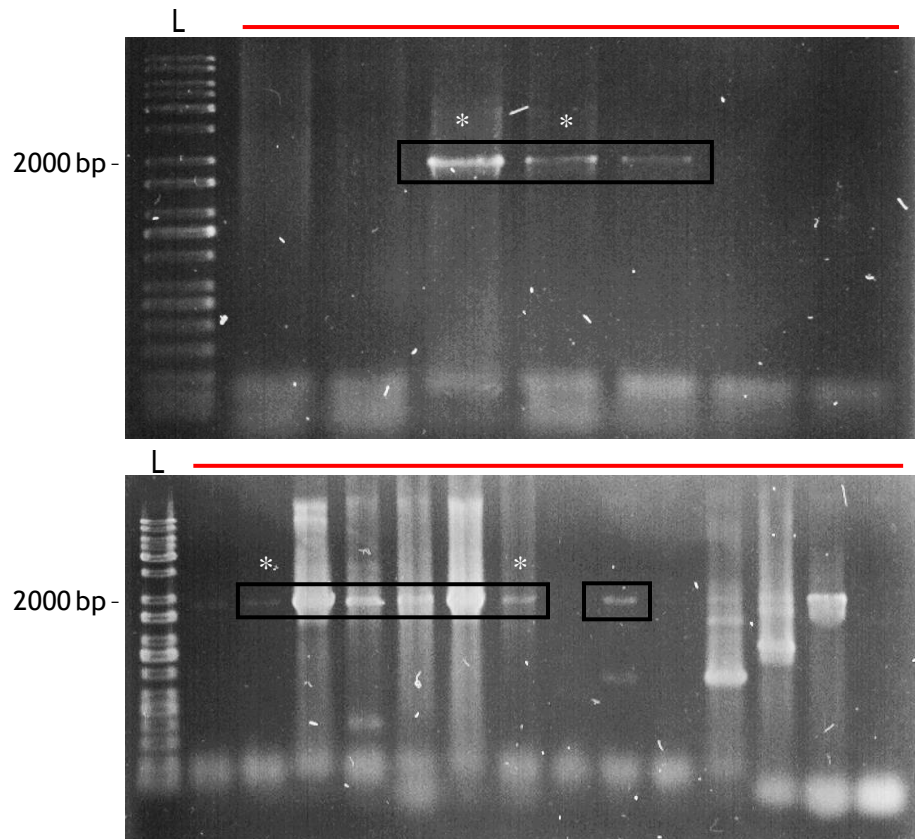


Figure 4.2: Agarose gels showing the colony PCR screen of the *phaC_b* gene ‘winners’ after genetic selections. L) Kapa universal ladder, size reference on the left, red line) indicates lanes containing colonies amplified for *phaC_b*, black boxes) display positive hits of colony PCR screen that were verified by sequencing. NB: the four PCR products that came back as positive hits for *phaC_b* are identified by a white asterisk above the band.

4.3.2 Selections of variant PhaC_b enzymes

The inserts in the 10 most promising plasmids (**Figure 4.2**) were sequenced. Surprisingly, and despite the positive PCR screen, six of these did not contain any *phaC_b* sequence. The four that contained a mutant *phaC_b* are denoted by a white asterisk on **Figure 4.2**. The PCR screen generated messy products containing multiple bands, although the band correlating to *phaC_b* is present in all ten samples sent for sequencing. However, the four that contained the gene were clean products and did not contain any non-specific bands so this may have played a role in whether *phaC_b* was present.

The six clones that did not contain PHA genes were also sequenced, three of which were found to contain the gene present in the original pET28 vector, *rml1*, which was used for library creation. The other three sequences had identities to

hypothetical *E. coli* proteins and an *E. coli* tetracycline resistance protein. This was done for interest sake and no further analysis was performed.

The four *phaC_b* containing sequences found from the genetic selections had a high number of mutations, short sequencing reads or premature stop codons. There is a significant chance that S1, S4 and S6 will be catalytically dead based on the result of selections from **Chapter 3** and the low success rate of isolating *phaC_b* during selections.

To deduce the effects of the mutations found in **Table 4.4**, a basic analysis was performed using an online computational tool known as PROVEAN (Choi and Chan, 2015). This program uses the protein sequence and specific mutations of the sequence as input. It then performs a BLAST search to find homologous sequences and uses this to predict the ‘**Protein Variation Effect Analysis**’. It provides scores for each individual mutation. The higher the score the more likely it is to be deleterious and vice versa (scores equal to or below -2.5 are deleterious and scores above -2.5 are neutral). This analysis was performed on S1, S4 and S6.

For S1, the mutation I279S is largely deleterious with a PROVEAN score of -5.6, effectively meaning overall this variant has a high chance of being non-functional.

S4 contained 3 deleterious mutations with scores ranging from -4.7 to -6.0.

Fortunately, Variant S6 contained two mutations which were deemed neutral and may be a candidate for further testing. However, the sequencing reads did not cover the full gene so there is a level of uncertainty. S7 contains a premature stop codon which results in an 85 residue non-functional protein. This was a basic *in silico* analysis and further *in vitro* activity assays will be required to confirm these findings.

Table 4.4: Sequencing results of *phaCb* variants from genetic selections

Seq ID	Substrate	Mutations
S1 – 2 ng lib.	hydroxy phenylacetic acid	V5F, H19L, Y92C, E149E (silent), G239S, S251S (silent), M256I, I279S, S454S (silent), E504D.
S4 – 2 ng lib.	3-hydroxybenzoic acid	D61N, S168N, E343A, L379Q, R490H, N501D.
S6 – 4 ng lib.	3-hydroxybenzoic acid	I198I (silent), T203I, A315V.
S7 – 4 ng lib.	<i>trans</i> -3-hydroxycinnamic acid	W25S, Y85STOP

*lib. = library.

4.3.3 Position of mutations on the structure of PhaCb

The structure of the catalytic domain of PhaCb was solved and characterised by others during this project. The catalytic domain is comprised of the α/β core subdomain (residues 175 – 318, and 439 – 562) and the CAP subdomain (residues 319 – 438) (Chek *et al.*, 2017). The α/β core contains the active site, while the CAP subdomain is important in dimerisation. The mutations in **Table 4.5** have been mapped onto the structure, where possible, to visualise the position and effect each mutation imposes on the overall functionality of the enzyme (**Figure 4.3**). Mutations that are not found in the solved structure have been omitted from further analysis as there is no structural data to support any claims made about these substitutions.

The distance from the active site of the deleterious mutations range from 13.5 to 31.0 Å (taken from the key catalytic residue, H477 to the most proximal atom on the side-chain of the un-mutated residue). R490H: 13.5 Å, L379Q: ~20 Å, I279S: 31.0 Å, E343A: 25.3 Å. This provokes the idea that distal residues can play an important role in function and that mutations in these residues can have deleterious effects, causing changes in conformational sampling to render inactive enzymes (Wang *et al.*, 2006).

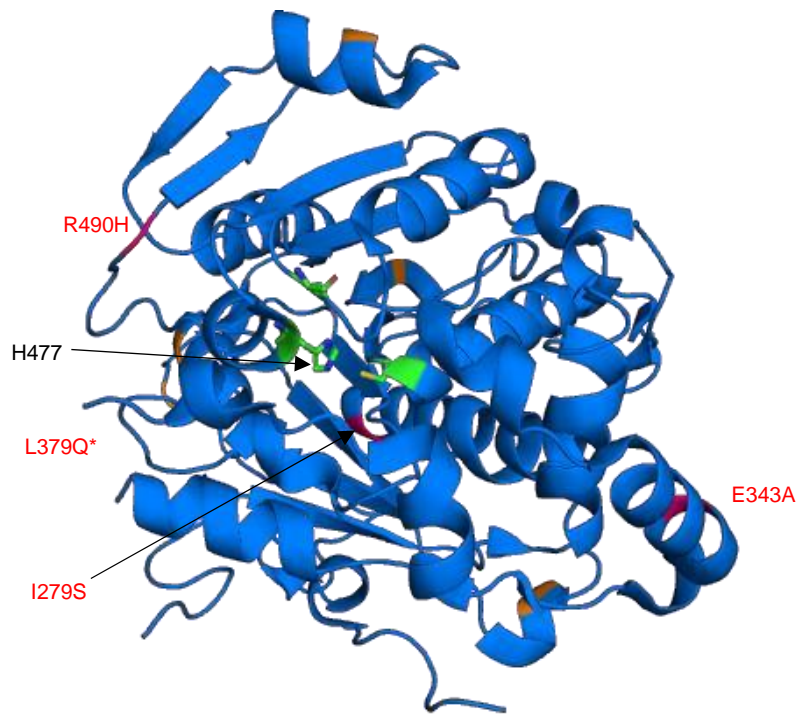


Figure 4.3: The crystallographically-determined structure of the catalytic domain of PhaCb (PDB ID: 5XAV). Key active site residues are shown in green sticks. (H477 is labelled as it was used for distance measurements). The inset shows the orientation of the residues that make up the catalytic triad (C291, D447 and H477). Mutations are denoted by; orange: neutral mutations and hot-pink: deleterious mutations (amino acid change described in red). * = section of the structure not solved, although a deleterious mutation was found in this region.

4.4 Chapter 4 Discussion

This chapter described the generation of two error-prone PCR libraries using the highly active *phaCb* gene as a template and implementation of the same life-or-death selection process that was described in **Chapter 3**.

The libraries generated from 2 and 4 ng starting template of *phaCb* contained much less bias and more suitable mutation rates compared to those generated in **Chapter 3**. The spread of the mutations was much more even, each possible nucleotide substitution had a higher representation and the bias toward transitions (established from the transition/transversion ratio) seemed in line with the experimental findings of Rasila *et al.*, (2009). These libraries also followed the established mutational rate trend - *i.e.* a lower starting template amount gives rise to a higher number of mutations per daughter sequence. The 2 ng library had a mutation rate of 8.6 mutations/*phaCb* while the 4 ng library contained 5.1 mutations/*phaCb*.

There was an overall trend in the enzyme engineering field stating that a lower mutation rate creates mutational diversity yet retains functionality in order to find improved clones (Arnold, 1998). However, more recent studies have suggested that higher mutation rate libraries contain a disproportionately high number of improved variants or variants with a novel function (Daugherty *et al.*, 2000; Drummond *et al.*, 2005). Based off this literature, the high number of mutations found in both the *phaC_b* libraries was satisfactory.

The introduction of an inducible expression system for *phaC_b*, ensured that PhaC_b was not expressed until cells were plated on media containing IPTG and the hydroxy acid substrates. It was hoped that this would avoid any highly functional variants from being rejected by the host cell and drop out of the library pre-selection. Based on the lack of success of in isolating *phaC_b* genes from the genetic selection, I do not think that controlling the expression of this gene helped overall. My argument from **Chapter 3** that *phaC* is expending unnecessary cellular energy and is being selected against by the host cells is still valid. However, the switch in plasmid did make downstream colony PCR and sequencing much easier.

During selections in this Chapter, the PCR screen performed showed 10 positive hits correlating to *phaC_b*. Yet, when I inoculated cultures, grew them overnight (without IPTG, so there was no expression of *phaC_b*) for DNA extraction and subsequently sequenced them, only 4 contained *phaC_b*. This may have been due to *E. coli* rejecting phaC_b and selecting against it during the selections where *phaC_b* was being expressed.

One downside of switching from pCDF (**Chapter 3**) to pET28 (**Chapter 4**) was the larger size of the latter. pCDF-phaC_a is 5079 bp, whereas pET28-phaC_b is 6997 bp. These changes inherently gave rise to complications with the purification and transformation of the pET28-PhaC_b plasmid. The DNA purification kits used were optimised for DNA sizes between 500 – 5000 bp, with the upper and lower limits yielding lower concentrations of DNA. There was also the potential that my electro competent cells were of poor quality increasing the effect of the transformation bottleneck - a phenomenon in the molecular biology field which explains the limit of DNA library members that can be transformed into a host cell to be selected or screened *in vivo* (Packer and Liu, 2015). The transformation bottle neck increases proportionally with the size of DNA fragments being transformed, so larger plasmids will have a lower transformation efficiency. Thus, it was not

surprising that the pET28 based *phaC_b* libraries were smaller, at ~250,000 variants each, than the pCDF-based *phaC_a* library in **Chapter 3**.

The three possibly functional variants found through selections contained a large number of mutations throughout the *phaC_b* gene. Upon a PROVEAN analysis and based on the results found in **Chapter 3** - in which the selection scheme isolated non-functional and truncated enzymes - I posited that two of these enzymes would not be functional. However, further *in vitro* and *in vivo* testing will confirm this theory. The idea that PhaC_b imposes a fitness cost on *E. coli* seems to be more plausible, given that the vast majority (27 out of 30) of clones emerging from the selections were clearly false positives (*e.g.* plasmids that lacked any *phaC_b* insert).

As a whole, the results presented in **Chapters 3 and 4** provide strong evidence that the selection scheme was not selecting for PhaC variants capable of polymerising CoA-primed aromatic substrates. Instead, I selected for variants with deleterious mutations, premature stop codons and also absent variants (*i.e.* variants with no *phaC* insert). The failure to find plausible candidates for further study drove the development of a screen (rather than selection). This is described in the next Chapter.

Chapter 5

Directed Evolution of PhaC_b-LCA

5.1 Introduction

The life-or-death selection used with PhaC_a (**Chapter 3**) and PhaC_b (**Chapter 4**) failed to yield variants with altered substrate specificities. This chapter describes a new approach. The goal was to engineer an ancestrally reconstructed PhaC (PhaC_b-LCA) for improved wild-type activity (*i.e.* the polymerisation of 3-*R*-hydroxybutyrate-CoA). The ancestral reconstruction was performed by collaborators at the University of Waikato (Drs Emma Andrews and Emma Summers, and Prof Vic Arcus), resulting in the successful production of a soluble and highly active ancient enzyme. Indeed, preliminary activity assays by the Waikato team found the activity of PhaC_b-LCA to be higher than that of PhaC_b, which in turn is 5-fold more active than PhaC_a (Bhubalan *et al.*, 2011).

Engineering this enzyme for even higher wild-type activity could be a possible step towards the introduction of aromatic substrates into the main-chain of the PHA it currently produces.

A novel high-throughput screen was designed using a previously described intracellular PHA stain known as Nile red (Spiekerman *et al.*, 1999). The multi-tiered screen comprised an initial agar plate-based screen and a quantifiable 96-well microplate screen. **Chapter 5** describes the generation of an epPCR library coupled with this screen to isolate variants with increased wild-type PHA synthase activity.

5.2 Error-prone PCR Library Generation

5.2.1 Error-prone PCR

Error-prone PCR libraries were constructed using 2 and 4 ng of template DNA, as described for PhaC_b in **Chapter 4**. Two libraries were made in order to gain two different mutational rates in the *phaC_b-LCA* amplicon. The epPCR products were generated using primers PhaC1.For and pGETS_PhaB.Rev (**Table 2.3**). These were then purified and the outcome of each epPCR was assessed by running aliquots of the DNA on a 1% agarose gel (**Figure 5.1**). The products displayed are from the second trial of epPCR, as the first trial resulted in the presence of non-specific bands

from the primers annealing incorrectly. An increase in the primer annealing temperature gave rise to cleaner products.

The epPCR products were cloned into the pGETS-CBA plasmid (**Table 2.2**) using the restriction enzymes BamHI and XhoI and the ligated products were used to transform *E. coli* strain E. cloni 10G. The pGETS-CBA plasmid has the genes of interest under the control of the constitutive lambda promoter: in this case, mutant *phaC_b-LCA*, *phaB* and *phaA*. These genes make up the full cassette required for native PHA polymerisation from sugars as in **Figure 1.3**. For this project glucose was utilised.

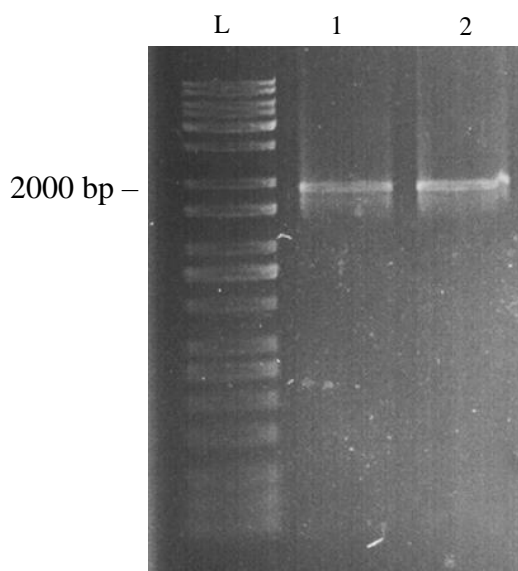


Figure 5.1: 1% agarose gel showing the epPCR products of *phaC_b-LCA*. L) Kapa universal ladder, size reference on the left. 1) 2 ng template epPCR product, 2) 4 ng template epPCR product. Expected product size: 1904 bp.

Chapter 5.2.2 Error-prone PCR Library Analysis

As described previously (**Sections 3.4.2** and **4.2.2**), randomly-chosen variants were sequenced from each of the two libraries to determine their mutational spectra. A total of 7,664 bp of sequence was analysed from the 2 ng starting template library (corresponding to ~4.5 *phaC_b-LCA* genes) and the results are summarised in **Table 5.1**.

In total, 53 mutations were found over the base pairs analysed. This results in an average of 12.2 mutations per *phaC_b-LCA* gene. The transition/transversion ratio is 1.2, suggesting a bias toward transitions throughout the epPCR amplicons. This

trend was also noted in the libraries constructed previously (**Tables 3.3** and **4.1**). Five indels were identified, corresponding to an average of ~1 per *phaC_b-LCA* sequence.

The mutational spectrum of the 4 ng starting template library was generated through the analysis of 7,587 bp of sequence correlating to ~4.5 *phaC_b-LCA* genes. The results are summarised in **Table 5.1**. A mutation rate of 11.0 mutations/*phaC_b-LCA* was observed for the 4 ng starting template library for *phaC_b-LCA*.

There is bias toward AT → N mutations compared to GC → N. This may be due to *phaC_b-LCA* having a lower G:C content of 45.6 % likened to that of *phaC_a* and *phaC_b* (50.1 and 47.0 % respectively). The transition/transversion ratio of 0.95 is close to the expected bias shown by Rasila *et al.*, (2009).

Out of the ~4.5 mutant *phaC_b-LCA* sequences analysed 4 indels were found. These create non-functional proteins due to the large frameshifts in the open reading frames (ORFs), impacting the number of full-length functional variants in the library. This is reflected in the results in **Table 5.3** displaying the PEDEL-AA estimations of the libraries composition. A total of 1.3×10^4 mutants were distinct full-length variants meaning the high number of mutations and rate of indels in the sequence resulted in 7×10^3 variants that contained insertions, deletions or premature stop codons.

I hypothesised that the new PHA synthase activity screen (described in more detail below) would make it straightforward to eliminate non-functional variants, such as those with an insertion or deletion, from downstream analysis. Therefore the ligation and transformation steps were scaled up, resulting in both libraries containing $\sim 2 \times 10^4$. The data in **Table 5.1** was used as an input to analyse both libraries using PEDEL-AA (Firth and Patrick, 2008). These results are summarised in **Table 5.2**. As expected, the number of full-length distinct variants is quite low for both libraries: at 1.1×10^4 and 1.3×10^4 for the 2 and 4 ng starting template libraries respectively. This correlates to just over 50% of each library containing full-length mutated genes, lower than usual due to the number of indels found in both libraries. At the same time, the activity screen is lower throughput than the life-or-death selection, so the small size of the library did not affect its usability.

The two libraries of 2 and 4 ng starting templates contained relatively similar mutational spectra and rates. However, they did not contain a large number of variants. Due to this, the libraries were pooled together to increase the likelihood of finding highly active variants during subsequent screening.

Table 5.1: Comparison of the mutational spectrums of *phaC_b-LCA* variants from the 2 ng and 4 ng starting templates.

Type(s) of mutations	Frequency (2 ng)	Proportion of total [%] (2 ng)	Frequency (4 ng)	Proportion of total [%] (4 ng)
Transitions				
A → G, T → C	15	28.3	13	27.7
G → A, C → T	11	20.8	8	16.9
Transversions				
A → T, T → A	16	30.2	13	27.7
A → C, T → G	1	1.9	3	6.4
G → C, C → G	1	1.9	3	6.4
G → T, C → A	4	7.5	3	6.4
Insertions and Deletions				
Insertions	1	1.9	3	6.4
Deletions	4	7.5	1	2.1
Summary of bias				
Transitions/ Transversions	1.2	N/A	0.95	N/A
AT → GC/GC → AT	1.1	N/A	1.5	N/A
A → N, T → N	32	66.7	28	66.6
G → N, C → N	16	33.3	14	33.3
Mutation rate				
Mutations per kb	6.9	N/A	6.2	N/A
Mutations per <i>phaC_b-LCA</i> gene*	12.2	N/A	11.0	N/A

* The cloned *phaC_b-LCA* insert was 1904 bp.

Table 5.2: Composition of the 2 ng *phaC_b-LCA* epPCR library as estimated by PEDEL-AA.

Property	Estimate
Total library size	2×10^4
Number of variants with no indels or stop codons	1.1×10^4
Mean number of amino acid substitutions per variant	8.6
Un-mutated (wild-type) sequences (% of library; Poisson est.)	0.011%
Number of distinct full-length proteins in the library (Poisson est.)	1.1×10^4

Table 5.3: Composition of the 4 ng *phaC_b-LCA* epPCR library as estimated by PEDEL-AA

Property	Estimate
Total library size	2×10^4
Number of variants with no indels or stop codons	1.3×10^4
Mean number of amino acid substitutions per variant	7.9
Un-mutated (wild-type) sequences (% of library; Poisson est.)	0.024%
Number of distinct full-length proteins in the library (Poisson est.)	1.3×10^4

5.3 Two-tier Screen for PHA Accumulation

5.3.1 Agar-Based Screen for PHA Accumulation

A two-tier screen was performed to analyse the combined *phaC_b-LCA* library for increased PHA accumulation. In the first round of screening, the library was grown on LB-agar plates supplemented with excess glucose (20 g/L) and Nile red. Nile red is an intracellular lipid stain and also stains PHA inclusions inside cells while maintaining the viability of the bacterial colonies (Spiekerman *et al.*, 1999). Colonies that synthesise PHA are fluorescent, and can be readily detected by the naked eye when the plate is viewed under blue-light trans illumination.

When the library was plated, a small proportion (~2%) of the colonies were fluorescent **Figure 5.2**. This suggested that there were active variants in the library, and that these were candidates for having PHA synthase activity greater than wild-

type. The largest and most fluorescent colonies were picked and subjected to a second, more quantitative, round of microplate-based screening.

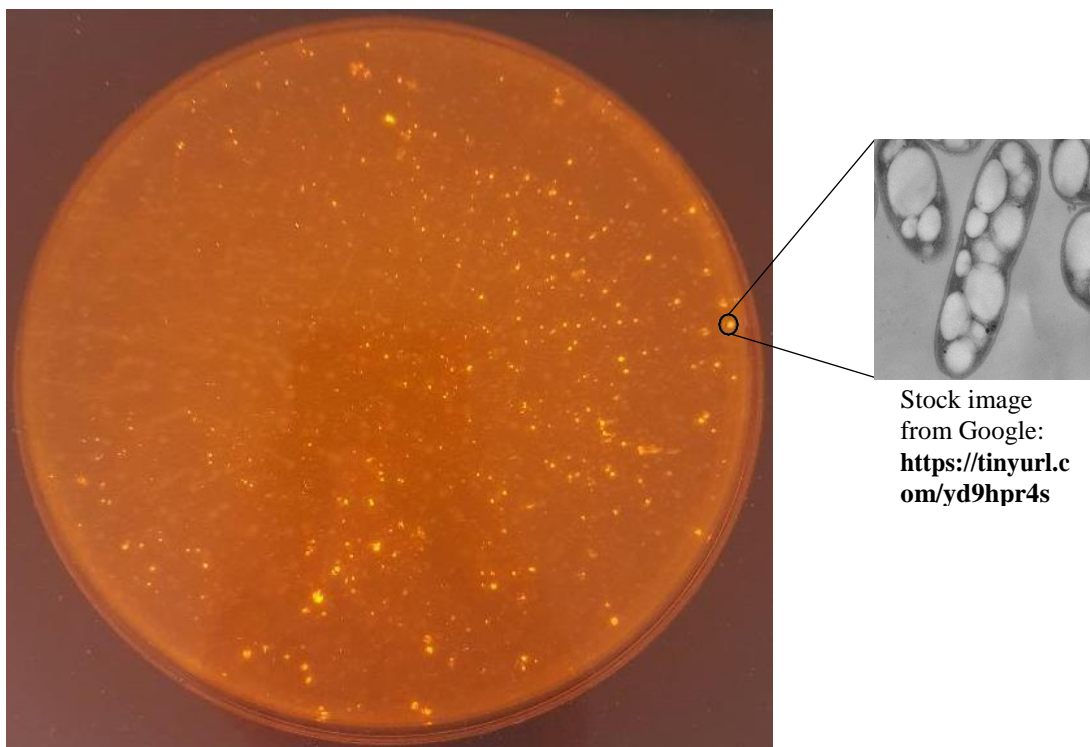


Figure 5.2: Viable colony staining of PHA accumulating cells from the pooled *phaC_b-LCA* epPCR library. Left: LB-agar containing excess glucose and Nile red, ~10,000 variants plated. The plate was photographed under blue-light transillumination ($\lambda = 470$ nm). Right: stock image of the appearance of PHA granules accumulating inside a bacterial cell.

5.3.1 Microplate-Based Screen for PHA Accumulation

In the second tier of the screen, the 92 most fluorescent colonies were picked into a 96-well microplate, along with two positive controls (expressing wild-type *PhaC_b-LCA*) and two negative controls (one pGETS-ligase empty vector, the other media-only). The first negative control contained cells that were unable to produce PHA, to observe the background staining of lipids present in the bacteria. The second negative control was media only, performed as a measure of residual staining after wash steps. The colonies were grown in liquid LB medium with excess glucose. After overnight growth, they were pelleted and resuspended in a buffer containing Nile red. Once each well had been stained, normalised fluorescence measurements of

the individual variants were made based off the excitation and emission wavelength of Nile red; 535 nm and 605 nm respectively (Zuriani *et al.*, 2013).

The results of the plate-based screen are shown in **Figure 5.3**. In technical duplicates [2], 28 variants displayed fluorescence measurements above wild-type. There were also around 50 fluorescent colonies picked from the plate that displayed little or no fluorescence when re-grown in liquid medium.

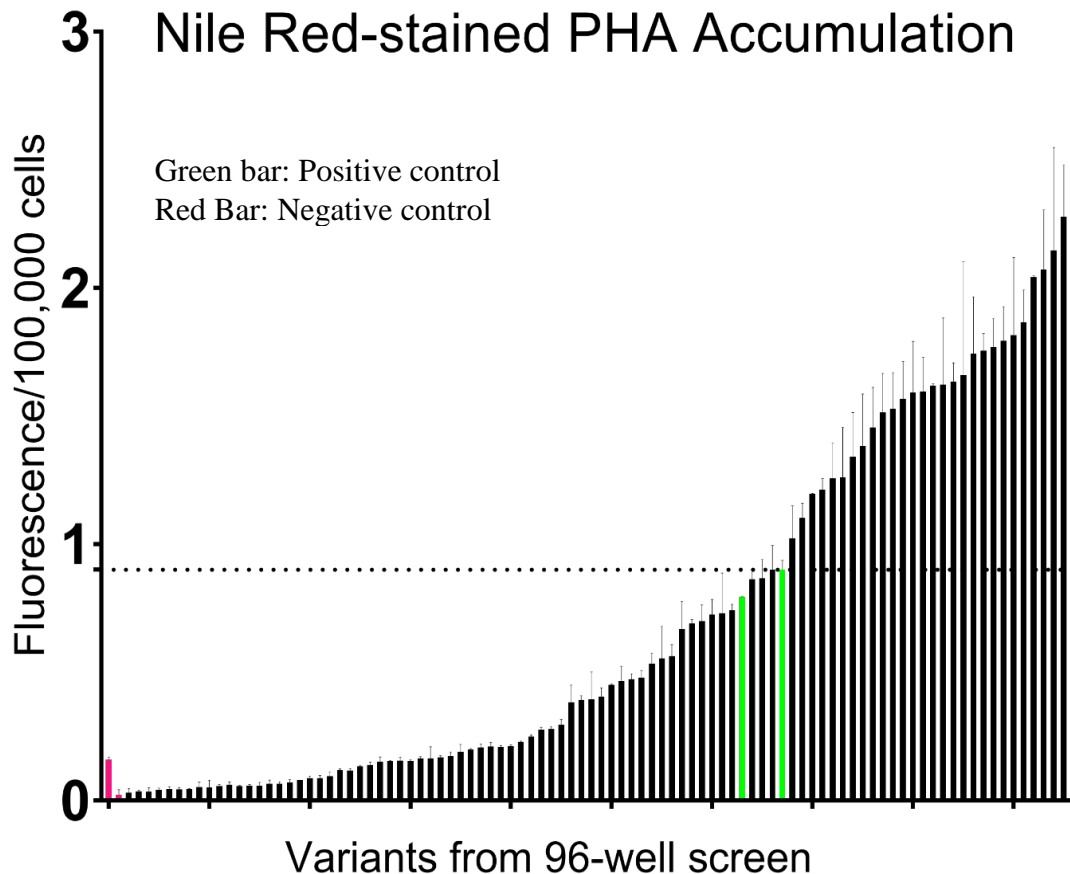


Figure 5.3: Bar graph showing 92 *phaC_b-LCA* variants and respective fluorescent intensities (n = 2, technical replicates). Positive control (green bars): cells containing wild type *phaC_b-LCA*. Negative control one (first red bar): Cells not producing PHA. Negative control two (second red bar): Media only. Error bars: SEM of technical replicates.

The 28 variants that outperformed wild-type were retested in an independent experiment, also done in technical duplicates. Ten of the 28 performed highly in both independent biological replicates consistently displaying fluorescence around two times that of the positive control. In one final test, these ten variants were spread onto LB-agar plates containing excess Nile red and glucose and grown overnight to ensure they still produced fluorescent colonies. Two of the variants exhibited no

fluorescence in this test. The eight variants which remained fluorescent were analysed by sequencing (**Table 5.4**).

Table 5.4: Sequencing results of *phaC_b-LCA* variants from PHA accumulation screen.

SEQ ID	Mean Fluorescence (\pm SEM)	Substrate	Mutations	Comments
A7	2.17 (0.47)	LB-agar + Glucose	M34K, V139A, A318V, K402I, D408E	5 aa substitutions
C6	2.14 (0.01)	LB-agar + Glucose	Q26H, Q62L, D124G , V159V (silent), G369D	4 aa substitutions
B4	1.97 (0.25)	LB-agar + Glucose	W73-->Stop, K176T, K480K (silent), G484G (silent)	Stop-codon, although cells are fluorescent so may be read-through. 2 aa substitutions
C5	1.92 (0.60)	LB-agar + Glucose	D18G, P43P (silent), E101E (silent), K141N, E213D, P347S, K402T, P404P (silent), V459L, R577R (silent)	6 aa substitutions
F8	1.89 (0.27)	LB-agar + Glucose	L113M, D124G , A318A (silent), R351S, K360K (silent), N387D, A463A (silent)	4 aa substitutions
H6	1.86 (0.13)	LB-agar + Glucose	I16F, S32G, I44V, Q62K, L72Q, A146S, K194E, N357Y, A361V	9 aa substitutions, 300 bp in the middle are not covered by sequencing reads
A11	1.85 (0.44)	LB-agar + Glucose	A52A (silent), F138L, M175L, I359T, N515Y	4 aa substitutions
A10	1.72 (0.52)	LB-agar + Glucose	V394V (silent), L437Q	1 aa substitution

*The mean fluorescence is scaled to the positive control of 1.0

*Conserved mutations between clones are highlighted red.

The variants contained between one and nine amino acid substitutions and there were a total of 35 substitutions. In 5 cases, charged amino acids were replaced by non-charged, and in 2 cases, non-charged were replaced by charged. These had the potential to be disruptive. To assess the mutations further, the Phyre2.0 server

(Kelley *et al.*, 2015) was used to build a homology model of the PhaC_b-LCA protein (**Figure 5.4**). PhaC_b-LCA has 67% amino acid identity with its contemporary counterpart PhaC_b. As a result, the model was built with 100% confidence over 378 of 590 amino acids, using the solved structure of PhaC_b as a template. The structure of PhaC_b covers residues 175 – 562 - *i.e.* the catalytic domain of the enzyme. The model of PhaC_b-LCA should therefore be the catalytic domain of the ancient enzyme.

When the mutations were mapped onto the modelled structure of PhaC_b-LCA, none were found to be near the active site (**Figure 5.4**). Instead, the majority were found outside the catalytic domain of the protein. It is known that other PhaC enzymes are active as dimers (Chek *et al.*, 2017; Sagong *et al.*, 2018). Based on homology with the known interface in PhaC_b, I can postulate that the potential dimerisation of this enzyme would not be affected by the mutations found.

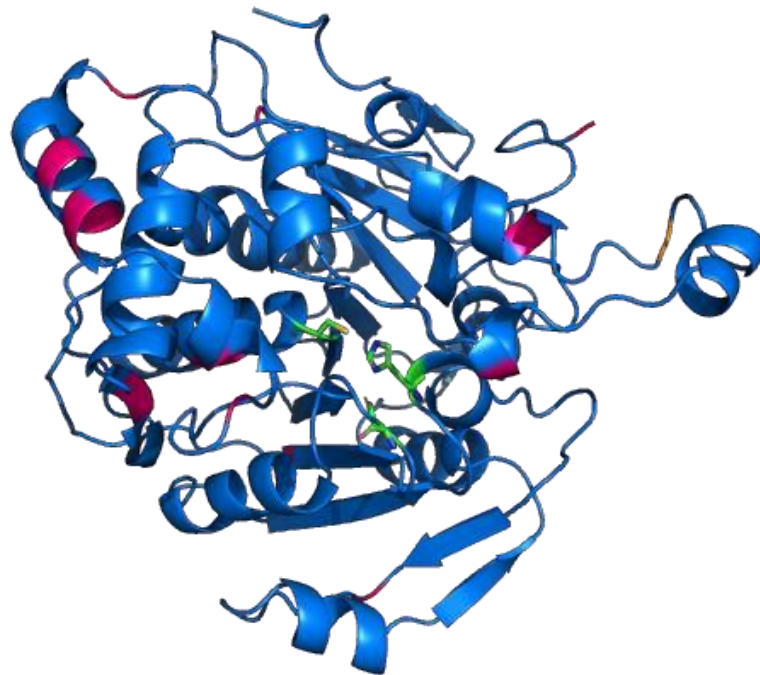


Figure 5.4: Homology model of PhaC_b-LCA with mutations from Table 5.5 mapped onto the structure. Hotpink) amino acid substitutions, Orange) residues mutated in two or more variants. Proposed active site is shown in green sticks. Note that no amino acid substitutions were found proximal to the active site.

5.4 Chapter 5 Discussion

The goal of the experiments described in this chapter was to increase the PHA accumulation of the ancestrally reconstructed PhaC_b-LCA enzyme. PhaC_b-LCA had a higher wild-type activity than both PhaC_a and PhaC_b, making it a suitable candidate

for further mutagenesis experiments. Increasing the catalysis of this enzyme would produce a superior PhaC than those found in nature. In order to achieve this goal, a plate based screen was used instead of the life-or-death selections used in earlier experiments

Two epPCR libraries of *phaCb-LCA* were created, each containing ~20,000 variants. After sequence analysis, the libraries were pooled and screened as one due to the similarities in mutational rates and biases. The similarities between these libraries may be attributed to several factors, one being the amount of template DNA used in the reactions. A book chapter by Chusacultanachai and Yuthavong, (2004) describes that in using a 10-fold difference in the amount of starting template DNA they noticed a 2-fold change in the mutation rate. Here, I used a 2-fold difference and noted only small changes in mutation rate.

The initial LB-agar based screening method was designed to allow the identification of PHA producing cells in the total library. Highly fluorescent colonies were picked and subjected to the further round of screening. The second tier of the screen was designed to quantify the amount of fluorescence of stained-PHA produced by each variant PhaC_b-LCA. The amount of fluorescence measured is assumed to be directly proportional to the amount of PHA produced by the cell (Zuriani *et al.*, 2013). This two-tiered approach allowed me to rapidly identify mutant PhaC_b-LCA enzymes with a high Nile red fluorescence. This method is inexpensive and takes three days from start to finish. This screening technique can be implemented on any known PhaC enzyme and may provide an initial determination of *in vivo* activity avoiding the expensive approaches currently used (scanning and transmission electron microscopy and gas chromatography – SEM, TEM and GC).

One potential pitfall was the large variation in Nile red fluorescence that was noted between replicates. This has been a trend in the literature regarding PHA production, with both the amount and composition of these biodegradable thermoplastics being variable when produced biologically (Chen *et al.*, 2004; Kelwick *et al.*, 2015). For example, Chen *et al.*, (2004) used engineered *E. coli* strains to maximise the production of PHA. Performing three independent experimental replicates this group noted a PHA accumulation (\pm the standard deviation of the means) in one of their strains of ~1.5 g/L (\pm 0.5). This is consistent with the large variation between replicates that I have found in this project.

In spite of these limitations, the two-tiered screen ultimately identified PhaC_b-LCA variants that reproducibly outperformed the wild-type (**Table 5.5**).

Variant B4 was of note as it contains a stop-codon 73 amino acids into the sequence. This should result in a small, truncated non-functional protein and yet the cells remained fluorescent when re-plated onto LB-agar with excess glucose and Nile red. One possibility is that there is translational read-through of the stop codon TAG. This has been reported before and termed a 'leaky' stop codon. (Ryden and Isaksson, 1984; Wei *et al.*, 2016).

It is known that TGA is the leakiest of the three stop codons (Wei *et al.*, 2016). However, there is still the possibility that TAG may have been read-through in this instance. Translation has been shown to be noisy and stochastic (Kaern *et al.*, 2005), meaning that stop codon read-through of the variant B4 may be more probable than what is inferred by statistics. The constitutive expression of *phaC_b-LCA* by the lambda promoter may play a role in the potential read-through of this stop codon.

Before commencing with the screening, one concern was the high mutation rate in the pooled library (~11 – 12 mutations per *phaC_b-LCA* gene). This was thought to be too high to expect any improved variants. Nevertheless, it did prove possible to identify mutants that conferred ~2-fold greater Nile red fluorescence in the screen. Sequencing revealed that a number of the variants that were most active (**Table 5.5**) did not contain a mutation rate as high as the values calculated in **Table 5.1**. They had an average of ~5.9 mutations at the DNA level, and ~4.4 amino acid substitutions, much less than the expected 11 – 12 mutations. However, high mutation rate libraries have been noted in the literature to be enriched for functional variants as they contain more distinct sequences. This was shown by Zacco and Gherardi, (1999), where they created epPCR libraries of the TEM-1 β -lactamase enzyme and found functional variants with up to 27.2 mutations/gene.

In a directed evolution experiment where variants with multiple mutations are identified, it is often hard to decipher which mutation or collection of mutations specifically confer the increase. It is also difficult to distinguish whether mutations affect the specific activity of the enzyme or whether they increase the stability or expression level of the protein. For example, a number of the mutations found were silent and did not result in amino acid substitutions (**Table 5.5**). It is well known that synonymous mutations can affect the process of translation dramatically (Gingold and Pilpel, 2011). The silent mutations found in highly active variants may be beneficial, if they result in increased expression levels.

In an effort to shed more light on the nature of the selected mutations, I built a homology model of PhaC_b-LCA (**Figure 5.4**). The structural information available is specific to the catalytic domains of PhaC, which are the most important for catalysis. However, mutations occurring in other regions of the protein may confer structural changes that have small effects on catalytic ability.

In general, the only trend that emerged was that none of the improved variants had mutations within 10 Å of the key catalytic residues. It seems likely that some or all the beneficial mutations are acting by stabilising the protein and/or enhancing its soluble intracellular expression.

It will be important to explore these possibilities further in future work. Expressing and purifying PhaC_b-LCA and the improved variants, and assaying their activities *in vitro*, will reveal whether expression levels and/or activities are altered by the mutations. It will also be useful to introduce the individual mutations into PhaC_b-LCA, to analyse their impact on enzyme function and determine which have the greatest effect. Given the large total number of mutations, it would also be interesting to randomly recombine them using DNA shuffling (Stemmer, 1994).

Ultimately, in terms of increased PHA production, a more stable enzyme with a high intracellular concentration is highly useful as a platform for further mutagenesis experiments. In this chapter, I managed to develop a novel screen for isolating variants with increased wild-type activity and by implementing this, find several improved variants for further testing. These variants may be a key step in the path towards the main aim of this research, to create a high-grade biopolymer using bacteria.

Chapter 6

Key Findings and Future Directions

6.1 Minimal selection pressure for PhaC to polymerise the aromatic substrates

In **Chapters 3** and **4**, I set out to evolve the enzymes PhaC_a and PhaC_b respectively, to alter their substrate specificity towards aromatic hydroxy acid substrates. epPCR experiments have a few key requirements for success, including gaining a suitable mutation rate in the gene of interest, obtaining a large and diverse library, and implementing a suitable screen/selection. I obtained both suitable mutation rates and large and relatively diverse libraries. However, a problem arose during the implementation of the genetic selection strategy.

The genetic selections gave rise to largely deleterious mutations, truncated mutants and cells lacking PhaC_a or PhaC_b. This may have been due to the aromatic hydroxy acid substrates being rendered non-toxic through the actions of the CoA ligase enzymes present in the cell. Ligating a CoA molecule to the substrates removes the carboxylic acid functional group (**Figure 3.6**), and this may have decreased the selective pressure for PhaC_a/PhaC_b to detoxify the substrates through polymerisation.

The genetic selections in **Chapters 3** and **4** did not yield mutants capable of polymerising the aromatic hydroxy acid substrates. These experiments demonstrated how difficult it can be to select for a specific enzymatic trait within the complex layers of regulation in a bacterial cell. Arnold's famous maxim is that 'you get what you select [or screen] for' (Schmidt-Dannert and Arnold, 1999). In contrast, screens – such as that developed in **Chapter 5** – can provide a more direct method for assaying the activity of interest.

It appears that the best way to achieve hydroxy acid polymerisation in future will be to adapt the high-throughput screen from **Chapter 5**. To achieve this a number of considerations will have to be taken into account. One of these is the ability of Nile red to stain the aromatic polymer that may be produced. This stain is the cornerstone of these experiments and a substitute will have to be found if Nile red is incapable of staining this new polymeric substance. A possible experiment to

determine this is to test the potential of Nile red to stain a similar aromatic substance – such as Kevlar. Alternatively, a commercial stain that has the ability to stain Kevlar could be utilised.

It might also prove beneficial to avoid depleting the intracellular pool, in order to identify variants with the highest possible yields of polymer. It would be worth investigating feeding experiments, in which either the activated hydroxy acid-CoA adduct or excess free CoA are added to the growth medium.

6.2 Expressing PhaC may impose a cost on the cell

Another potential issue with the life-or-death selection for PhaC_a variants is the constitutive expression of PhaC_a during all steps of the experimental process. In nature, PhaC enzymes are used to synthesise intracellular carbon stores, when carbon is in excess of the cell's requirements. In the experiments of **Chapter 3**, PhaC_a expression was unnecessary for the cell. Over-expressing an enzyme that does not perform a critical role is a waste of energy (Dekel *et al.*, 2005); the selection pressure may have effectively been to stop expressing PhaC_a.

The experiments in **Chapter 4** had a very similar design and outcome to those in **Chapter 3**: epPCR library creation and genetic selections which yielded non-functional PhaC_b variants. Given the lack of success in **Chapter 3**, expression of PhaC_b was made inducible by using pET28. The idea was to express PhaC_b only during the life-or-death selection step, when it was hoped that the PhaC enzyme was playing an essential role for the cell.

In spite of this altered experimental design, the genetic selection again yielded variants that were all predicted to be non-functional. This suggests that even transient expression of PhaC may be selected against, especially when the host cells are under stressful environmental conditions.

Future experiments should include a growth curve analysis comparing cells with high level inducible expression of *phaC* from pET28 and those with constitutive expression from pCDF-*phaC* to those with no expression of *phaC*. An expression test using the titratable pBAD vector system would also be informative.

Analysing growth curves may be able to distinguish whether or not the continuous expression of *phaC* is detrimental to cell health and growth. Analysing the area under the curve (AUC) for cells with high levels of PhaC and no PhaC will provide an insight to whether or not the enzyme imposes a fitness cost to the cells.

The pBAD expression system (Guzman *et al.*, 1995) is a titratable expression system which uses P_{BAD} promoter from the arabinose operon. This promoter's expression is turned on by arabinose (the inducer) and in the absence of arabinose expresses downstream proteins at very low concentrations. Were each of the *phaC* genes to be cloned downstream of the P_{BAD} promoter, it would be possible to vary the expression of PhaC by altering the concentration of arabinose. Growth curves could then be measured for cells harbouring these constructs to see whether there is a concentration dependence for PhaCs potential toxicity.

6.3 PhaC_b-LCA gained increased wild-type activity

In the experiments described in **Chapter 5**, I switched focus and investigated the possibility of increasing the native PHA synthase activity of the ancestrally reconstructed enzyme, PhaC_b-LCA. epPCR libraries were constructed and subjected to a novel two-tier screen to isolate variants with increased PHA accumulation.

Preliminary evidence from our collaborators at the University of Waikato suggested that the ancestral enzyme has higher PHA synthase activity than its contemporary counterpart, PhaC_b. I found it possible to increase activity further through epPCR and screening. A number of variants were found that reproducibly out-performed the wild-type enzyme in the screen. Interestingly, only two out of the 34 mutations identified in 8 improved variants were conserved across two variants. This highlights the diversity of mutations that can increase the activity of this ancient enzyme and the plasticity this enzyme has for potential further engineering. At this stage, it is difficult to distinguish which individual mutations are conferring the increases in activity, so further work must be done to determine this.

Firstly, it will be important to express and purify PhaC_b-LCA and the improved variants in order to compare their activities *in vitro*. This will provide us with an idea of whether the expression levels or activities are affected by the mutations present in each variant. It will also isolate the highest performing variant enzyme in comparison to the wild-type enzyme, this will be important for future work below.

Secondly, to find which of the mutations are important, it will be necessary to create enzyme variants that each contain one of the 34 identified mutations. Construction of these mutants was beyond the scope of this project.

An initial parameter could be looking at mutations found only in the catalytic domain of the enzyme. This would neglect the distal mutations that may play a role in the enzymes stability as opposed to directly altering the catalysis of the enzyme. A second idea is to map the individual mutations distance from the active site. In directed evolution experiments, mutations distal from the active site have been found to play a key role in gaining the wanted function (Cahn *et al.*, 2016). However, in general it is much more common for proximal mutations to active site to dramatically affect catalysis (Lee and Goodey, 2011). For this reason, it could be useful to reduce the number of mutants that have to be made and screened.

6.4 PhaC_b-LCA as a scaffold for further mutagenesis

Overall the results in **Chapters 3 – 5** suggest that there is no evolutionary path for the PhaC enzymes to accommodate aromatic substrates in a single round of mutagenesis. A step-wise approach may be the key to finding variants capable of aromatic polymerisation. Enhancing the wild-type activity of PhaC_b-LCA through the experiments in **Chapter 5** may have provided a better scaffold for further rounds of mutagenesis.

The first thing to do next would be to express/purify the winners found and test whether there is any activity toward the CoA-primed aromatic hydroxy acid substrates. If there are detectable levels of polymerisation of the aromatic substrates, then the variants displaying this can be used for further rounds of epPCR. Another potential experiment that could help identify the most suitable scaffold protein would be analysing what PHAs can be produced by the variant enzymes.

The improved variants may have some or all of: increased stability, increased catalytic efficiency or increased expression levels. These three things could exhibit a positive effect for further mutagenesis aimed toward gaining a variant enzyme with aromatic polymerisation capabilities.

I may have created variant enzymes harbouring over-compensatory mutations which lead to increased stability, catalysis or expression. An enzyme with increased stability is often termed ‘robust,’ this is beneficial as a template for mutagenesis experiments as destabilising mutations, that lead to the desired function, are more well tolerated (Bloom *et al.*, 2006). The same can be said for the other two traits, as

the effect of mutations which increase the catalytic efficiency may be eroded by further mutations leading towards aromatic polymerisation. This is why it is imperative to find which mutations confer which of these enhanced enzyme traits.

The structure of PhaC_b-LCA is currently unknown. While I built a high-confidence homology model (**Figure 5.4**), solving the structure will provide insight into the differences between this ancient enzyme and the extant enzymes from which its sequence was inferred. Another noteworthy structure to elucidate would be the highest-performing variant found from the screening protocol in **Chapter 5** and further *in vitro* assays. The comparison of the wild-type enzyme and the variant would then highlight any changes in the structure which contribute to the activity/stability of the mutant enzyme.

Short of polymerising aromatic substrates, any variants capable of polymerising a monomer with a higher number of carbon atoms may be valuable. Natively, the enzymes PhaC_a and PhaC_b create short chain length PHAs (Wittenborn *et al.*, 2016; Chek *et al.*, 2017), which are low-value and have limited uses. For example, when glucose is the carbon source in excess, it is metabolised to acetyl-CoA and fed into the PHA pathway shown in **Figure 1.3**. The resulting monomer, 3-hydroxybutyryl-CoA, undergoes polymerisation to polyhydroxybutyrate, which has limited uses to become a viable replacement for petroleum derived plastics (Li *et al.*, 2016). A PhaC variant that could polymerise larger aliphatic monomers would provide a route to a higher-grade elastomeric plastic produced by bacteria which could replace certain petroleum based plastics, a step toward sustainability. An example is

6.5 Perspectives on the utility of error-prone PCR

I also wish to add some comments on epPCR as a technique. It is used to create mutant libraries, some of which harbour variants with desirable qualities. epPCR has been used by many researchers to create novel variants capable of certain functions that were otherwise inconceivable if a rational approach was undertaken (van Vugt-Lussenburg *et al.*, 2007; Tracewell and Arnold, 2009; Maddock *et al.*, 2017; Liu *et al.*, 2019). At the same time, the results described in this thesis underscore the variability of the technique.

Initially, in **Chapter 3** I used 4 ng and 20 ng of template DNA. The literature predicts a higher mutation rate for 4 ng template compared with 20 ng (Hanson-

manful and Patrick, 2013). However, upon analysing the mutational spectra in each library, I observed a similar mutation rate between them, with 20 ng template actually leading to a slightly higher mutation rate (4.6 mutations per *phaCa* gene compared to 4.4).

The libraries described in **Chapter 4** had a uniform distribution of all potential mutations and harboured mutation rates comparable to the amount of starting template. The libraries constructed in this chapter were textbook, with little bias and the expected mutational rates.

In **Chapter 5**, the two template amounts of 2 and 4 ng yielded libraries with near identical mutation rates. I believe that the difference between 2 and 4 ng of starting template is negligible and that larger changes in the template amounts will be needed to yield distinct mutation rates and libraries. Another example is the comparison of the mutation rate of the 4 ng template libraries from **Chapters 3, 4 and 5** (2.5, 3 and 6.2 per kb of *phaC*, respectively) shows a large amount of variability in the epPCR. This demonstrates the highly stochastic nature of epPCR.

Screening or selecting is often time-consuming and requires a carefully planned protocol. It is important that if this process is to be carried out on two or more separate libraries then they must be distinct from one another, otherwise it is often more beneficial to pool together similar libraries as was done in **Chapter 5**.

My results from **Chapter 5** found enhanced variants containing a range of two to nine amino acid substitutions in the *phaCb-LCA* gene. This suggests that mutation rate can vary widely and still produced enhanced variants, highlighting that even libraries which may seem unsatisfactory upon initial analysis can yield improved variants.

6.6 Concluding Remarks

The main aim of this thesis was to introduce aromatic substrates into the main chain of the PHA currently produced by bacteria, increasing the value and versatility of this biopolymer. To achieve this, the key enzyme involved in the production of PHAs, PhaC, was subjected to directed evolution and a life-or-death selection. Ultimately, the desired activity of PhaC was not observed. However, progress toward this goal has been made.

In particular, this thesis describes the successful identification of PhaCb-LCA variants that outperformed the native enzyme by ~2-fold in a novel two-tier screen

for wild-type activity. Further characterisation of these variants *in vitro* will enable one to be chosen as an ideal scaffold for further directed evolution to achieve the main aim of this research, aromatic polymerisation. I believe this step-wise approach may be able to generate mutants capable of polymerising aromatic substrates into the main chain of the polymer it currently produces. If this can be achieved, it will lead to less reliance on petroleum-based plastics by taking low-quality waste products and generating high-grade bioplastics in bacteria.

References

- Abou-Nader, M., & Benedik, M. J. (2010). Rapid generation of random mutant libraries. *Bioengineered*. **1**, 337–340.
- Alting-Mees, M. A., & Short, J. M. (1989). pBluescript II: gene mapping vectors. *Nucleic Acids Res.* **17**, 9494.
- Anderson, A. J., & Dawes, E. A. (1990). Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiol. Rev.*, **54**, 450–472.
- Andrews, K. J., & Hegeman, G. D. (1976). Selective disadvantage of non-functional protein synthesis in *Escherichia coli*. *J. Mol. Evol.* **8**, 317 – 328.
- Aoteaplas.co.nz. (2017). Manufacturers of Plastic Bottles, Flexible Packaging Tubes and small Injection Moulded Products in New Zealand. [online] Available at: http://www.aoteaplas.co.nz/technical/nz_plastic_production_statistics [Accessed 27 Apr. 2017].
- Arnold, F. H. (1998). Enzyme engineering reaches the boiling point. *Proc. Natl Acad. Sci. USA.* **95**, 2035 – 2036.
- Bennet, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van, S. J., Dien, J. D., & Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **5**, 593 – 599.
- Bhubalan, K., Chuah, J-A., Shozui, F., Brigham, C. J., Taguchi, S., Sinskey, A. J., Rha, C., & Sudesh, K. (2011). Characterization of the highly active polyhydroxyalkanoate synthase of *Chromobacterium* sp. strain USM2. *Appl. Environ. Microbiol.* **77**, 2926 – 2933.
- Bloom, J. D., & Frances F. H. (2009). In the light of directed evolution: Pathways of adaptive protein evolution. *PNAS.* **106**, 9995 – 10000.
- Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. (2006). Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA.* **103**, 5869–5874.
- Byrom, D. (1992) Production of poly-3-hydroxybutyrate: poly-3-hydroxyvalerate copolymers, *FEMS Microbiol. Lett.* **103**, 247–250.
- Cahn, J. K. B., Baumschlager, A., Brinkmann-Chen, S., & Arnold, F. H. (2016) Mutations in adenine-binding pockets enhance catalytic properties of NAD(P)H-dependent enzymes. *Protein Eng. Des. Sel.* **29**, 31 – 38.
- Chanasit, W., Hodgson, B., Sudesh, K., & Umsakul, K. (2016). Efficient production of Polyhydroxyalkanoates (PHAs) from *Pseudomonas mendocina* PSU using a biodiesel liquid waste (BLW) as the sole carbon source. *Biosci. Biotech. Bioch.* **80**, 1440–1450.
- Chedid, R., Kobrosly, M., & Ghajar, R. (2007). A supply model for crude oil and natural and natural gas in the middle east. *Energy. Policy.* **35**, 2096 – 2109.
- Chen, J.Y., Liu, T., Zheng, Z., Chen, J.C., & Chen, G.Q. (2004). Polyhydroxyalkanoate synthases PhaC1 and PhaC2 from *Pseudomonas stutzeri* 1317 had different substrate specificities. *FEMS. Microbiol. Lett.* **234**, 231–237.
- Choi, J., & Lee, S. Y. (2004). High level production of supra molecular weight poly((R)-3-hydroxybutyrate) by metabolically engineered *Escherichia coli*. *Biotechnol. Bioprocess. Eng.* **9**, 196 –200.

- Choi, Y., & Chan, A. P. (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. **31**, 2745 – 2747.
- Chung, A. L., Jin, H. L., Huang, L. J., Ye, H. M., Chen, J. C., Wu, Q., & Chen, G. Q. (2011). Biosynthesis and characterization of poly(3-hydroxydodecanoate) by beta-oxidation inhibited mutant of *Pseudomonas entomophila* L48. *Biomacromolecules*. **12**, 3559-3566.
- Chusacultanachai, S., & Yuthavong, Y. (2004) Random Mutagenesis Strategies for Construction of Large and Diverse Clone Libraries of Mutated DNA Fragments. In: Melville S.E. *Parasite Genomics Protocols. Methods in Molecular Biology™*, vol 270. Humana Press.
- Cirino, P. C., Mayer, K. M., & Umeno, D. (2003). Generating mutant libraries using error-prone PCR. *Methods. Mol. Biol.* **231**, 3–9.
- Copp, J.N., Hanson-Manful, P., Ackerley, D.F. & Patrick, W.M. (2014). Error-prone PCR and effective generation of gene variant libraries for directed evolution. *Methods. Mol. Biol.* **1179**, 3-22.
- Datta, N., & Hedges, R. W. (1971). Compatibility groups among f_i^- R factors. *Nature*. **234**, 222 – 223.
- Daugherty, P. S., Chen, G., Iverson, B. L. & Georgiou, G. (2000). Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. *Proc. Natl Acad. Sci. USA*. **97**, 2029 – 2034.
- Dekel, E., Mangan, S., & Alon, U. (2005). Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Phys. Biol.* **2**, 81 – 88.
- Dougherty, M. J., & Arnold, F. H. 2009. Directed evolution: New parts and optimized function. *Curr. Opin. Biotechnol.* **20**, 486 – 491.
- Drummond, D. A., Iverson, B. L., Georgiou, G., & Arnold, F. H. (2005). Why high error-rate random mutagenesis libraries are enriched in functional and improved proteins. *J. Mol. Biol.* **350**, 806 – 816.
- Elvin, C. M., Thompson, P. R., Argall, M. E., Philip Hendr, N., Stamford, P. J., Lilley, P. E., & Dixon, N. E. (1990). Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene*. **87**, 123–126.
- Firth A. E., & Patrick W. M. (2008). GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Res.* **36**, 281–285.
- Gaucher, E. A., Thomson, J. M., Burgan, M. F., & Benner, S. A. (2003). Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature*. **425**, 285–288.
- Gingold, H., & Pilpel, Y. (2011). Determinants of translation efficiency and accuracy. *Mol. Sys. Biol.* **7**, 481 – 494.
- Greenspan, P., Mayer, E. P., & Fowler, S. D. (1985). Nile red: A selective fluorescent stain for intracellular lipid droplets. *J. Cell Biol.* **100**, 965-973
- Gumel, A. M., Annuar, M. S. M., & Chisti, Y. (2013). Recent Advances in the Production, Recovery and Applications of Polyhydroxyalkanoates. *J Polym Environ.* **21**: 580 – 605.

Gumulya, Y., Baek, J.-M., Wun, S.-J., Thomson, R. E. S., Harris, K. L., Hunter, D. J. B., Behendroff, J. B. Y. H., Kulig, J., Zheng, S., Wu, X., Wu, B., Stok, J. E., De Voss, J. J., Schenck, G., Jurva, U., Andersson, S., Isin, E. M., Boden, M., Guddat, L., & Gillam, E. M. J. (2018). Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat Catal*, **1**, 878–888.

Guzman, L.-M., Belin, D., Carson M. J., & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.* **177**, 4121 – 4130.

Hanson-Manful, P., & Patrick, W. M. (2013). Construction and analysis of randomized protein-encoding libraries using error-prone PCR. *Methods. Mol. Biol.* **996**, 251-267.

Hiroe, A., Tsuge, K., Nomura, C. T., Itaya, M., & Tsuge, T. (2012). Rearrangement of gene order in the phaCAB operon leads to effective production of ultrahigh-molecular-weight poly[(R)-3-hydroxybutyrate] in genetically engineered *Escherichia coli*. *Appl. Environ. Microbiol.* **78**, 3177–3184.

Hobbs, J. K., Prentice, E. J., Groussin, M. & Arcus, V. L. (2015). Reconstructed ancestral enzymes impose a fitness cost upon modern bacteria despite exhibiting favourable biochemical properties. *J. Mol. Evol.* **81**, 110 – 120.

Kaern, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451 – 464.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858.

Kelwick, R., Kopniczky, M., Bower, I., Chi, W., Chin, M. H. W., Fan, S., Pilcher, J., Strutt, J., Webb, J. A., Jensen, K., Stan, G-B., Kitney, R., & Freemont, P. (2015). A forward-design approach to increase the production of poly-3-hydroxybutyrate in genetically engineered *Escherichia coli*. *Plos One.* **10**, e0117202.

Kim, J., Kim, Y. J., Choi, S. Y., Lee S. Y., & Kim, K. J. (2017). Crystal Structure of *Ralstonia eutropha* Polyhydroxyalkanoate Synthase C-terminal Domain and Reaction Mechanisms. *Biotechnol. J.* **12**, 1600648.

Koetsier, M. J., Jekel, P. A., van den Berg, M. A., Bovenberg, R. A. L., & Janssen, D. B. (2009). Characterization of a phenylacetate-CoA ligase from *Penicillium chrysogenum*. *Biochem. J.* **417**, 467–476.

Laempe, D., Jahn, M., Breese, K., Schagger, H., & Fuchs, G. (2001). Anaerobic metabolism of 3-hydroxybenzoate by the denitrifying bacterium *Thauera aromatic*. *J. Biol.* **183**, 968 – 979.

Lee, J., & Goodey, N. M. (2011). Catalytic contributions from remote regions of enzyme structure. *Chem. Rev.* **111**, 7595 – 7624.

Lee, G. N., & Na, J. (2013) Future of microbial polyesters. *Microb. Cell. Fact.* **12**, 54.

Li, Z., Yang, J., & Loh, X. J. (2016). Polyhydroxyalkanoates: opening doors for a sustainable future. *NPG Asia Mater.* **8**, e265.

Liu, B., Peng, Q., Sheng, M., Hu, S., Qian, M., Fan, B., & He, J. (2019). Directed evolution of sulfonylesterase and characterization of a variant with improved activity. *J. Agric. Food Chem.* **67**, 836–843.

- Maddock, D. J., Gerth, M. L., & Patrick, W. M. (2017). An engineered glycerol dehydratase with improved activity for the conversion of meso-2,3-butanediol to butanone. *Biotechnol. J.* **12**, 1700480.
- Meng, D. C., Shen, R., Yao, H., Chen, J. C., Wu, Q., & Chen, G. Q. (2014). Engineering the diversity of polyesters. *Curr. Opin. Biotechnol.* **29**, 24 – 33.
- Morley, K. L., & Kazlauskas, R. J. (2005). Improving enzyme properties: when are closer mutations better? *Trends. Biotechnol.* **23**, 231–237.
- Packer, M. S., & Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nat. Rev. Gen.* **16**, 379 – 394.
- Patrick, W.M. & Firth, A.E. (2005). Strategies and computational tools for improving randomized protein libraries. *Biomol. Eng.* **22**, 105-112.
- Russel, J. B., & Cook, G. M. (1995). Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Micorbiol. Rev.* **59**, 48 – 62.
- Ryden, S. M., & Isaksson, L. A. (1984). A temperature-sensitive mutant of *Escherichia coli* that shows enhanced misreading of UAG/A and increased efficiency for some tRNA nonsense suppressors. *Mol Gen Genet.* **193**, 38–45
- Sagong, H. Y., Son, H. F., Choi, S. Y., & Kim, K. J. (2018). Structural Insights into Polyhydroxyalkanoates Biosynthesis. *Trends in Biochemical Sciences*, **43**, 790 – 805.
- Sambrook, J. F., & Russel D. W. (2001) *Molecular Cloning: A Laboratory Manual; Edition 3*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory.
- Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F.J., Brands, J., Devine, P. N., Huisman, G. W., & Hughes, G. J. (2010). Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*, **329**, 305–309.
- Schmidt-Dannert, C., & Arnold, F. H. (1999). Directed evolution of industrial enzymes. *Trends. Biotechnol.* **17**, 135 – 136.
- Shin, S., & Day, L. A. (1995). Separation and size determination of circular and linear single-stranded DNAs by alkaline agarose gel electrophoresis. *Anal. Biochem.* **226**, 202–206.
- Spiekermann, P., Rehm B. H. A., Kalscheuer, R., Baumeister, D., & Steinbuchel, A. (1999). A sensitive viable colony staining method using Nile red for direct screening of bacteria that accumulate PHA and other lipid storage compound. *Arch. Microbiol.* **171**, 73-80.
- Steinbuchel, A., & Valentin, H. E. (1995) Diversity of bacterial polyhydroxyalkanoic acids. *FEMS Microbiol. Lett.* **128**, 219 –228.
- Stemmer, W. P. C. (1994). DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA.* **91**, 10,747–10,751.
- Stubbe, J., & Tian, J. (2003) Polyhydroxyalkanoate (PHA) homeostasis: the role of the PHA synthase. *Nat. Prod. Rep.* **20**, 445–457.
- Sylvestre, J., Chautard, H., Cedrone, F., & Delcourt, M. (2006). Directed evolution of biocatalysts. *Org. Process Res. Dev.* **10**, 562–571.
- Thornton, J. W. (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366 – 375.

- Tracewell, C. A., & Arnold, F. H. (2009). Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr Opin Chem Biol.* **13**, 3 – 9.
- van Vugt-Lussenburg, B. M. A., Stjernschantz, E., Lastdrager, J., Oostenbrink, C., Vermeulen, P. E., & Commandeur, J. N. M. (2007). Identification of critical residues in novel drug metabolizing mutants of cytochrome P450 BM3 using random mutagenesis. *J Med Chem.* **50**, 455 – 461.
- Vanhercke, T., Ampe, C., Tirry, L., & Denolf, P. (2005). Reducing the mutational bias in random protein libraries. *Anal. Biochem.* **339**, 9 – 14.
- Wang, L., Goodey, N. M., Benkovic, S. J., & Kohen, A. (2006). Coordinated effects of distal mutation on environmentally coupled tunnelling in dihydrofolate reductase. *Proc. Natl Acad. Sci. USA.* **103**, 15753 – 15758.
- Wei, Y., Wang, J., & Xia, X. (2016). Coevolution between Stop Codon Usage and Release Factors in Bacterial Species. *Mol. Biol. evol.* **33**, 2357 – 2367.
- Wiegand, I., Hilpert, K., & Hancock, R. E. W. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Prot.* **3**, 163–175.
- Wittenborn, E. C., Jost, M., Wei, Y., Stubbe, J., & Drennan, C. L. (2016). Structure of the catalytic domain of the class I polyhydroxybutyrate synthase from *Cupriavidus necator*. *J. Biol. Chem.* **291**, 25264–25277.
- Yuan, Y., Yu, S., Yu, J., Zhan, Z., Li, M., Liu, G., Wang, X., & Huang, L. (2014). Predicting the function of 4-coumarate: CoA ligase (LJ4CL1) in *Lonicera japonica*. *Int. J. Mol. Sci.* **15**, 2386–2399.
- Zhang, X., Luo, R., Wang, Z., Deng, Y., & Chen G. Q. (2009). Application of (R)-3-hydroxyalkanoate methyl esters derived from microbial polyhydroxyalkanoates as novel biofuels. *Biomacromolecules.* **10**: 707 – 711.
- Zuriani, R., Vigneswari, S., Azizan M. N. M., Majid, M. I. A., & Amirul, A. A. (2013). A high throughput Nile red fluorescence method for rapid quantification of intracellular bacterial polyhydroxyalkanoates. *Biotechnol. Bioprocess. Eng.* **18**, 472 – 478.