# The development of a novel, CoA-dependent, organic acid modification system as a link to a biosynthetic pathway for the nylon 6 monomer, 6-aminocaproic acid

Thomas William Paul Hickman

UCL

Thesis submitted to UCL for the degree of Doctor of Philosophy

in

Biochemical Engineering

December 2019

## Declaration

I, Thomas William Paul Hickman confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm this has been indicated in the thesis.

## Abstract

The inherently unsustainable production chain of the nylon 6 monomer, 6-aminocaproic acid (6-ACA), has motivated a number of investigations into developing sustainable alternatives. Previous research at UCL developed a biosynthetic pathway from cyclohexanol to 6-ACA. However, the requirement for petroleum-derived cyclohexanol, limited the pathway's sustainability. Therefore, this investigation aimed to develop a biosynthetic pathway for renewable cyclohexanol production, from an intermediate of central metabolism. The pathway developed here, utilises eight enzymes for the production of cyclohexanol, through the CoA-dependent biotransformation of either shikimate or 3-phosphshikimate. Cyclohexanol production was not fully realised, however, exploring new biosynthetic routes for cyclohexanol production revealed several new enzymes and systems that were characterised and constructed.

Four enzymes (MycA1-4), identified in *Streptomyces rishiriensis*, were recombinantly expressed and assayed. The acyl-CoA reductase, MycA2, showed activity with cyclohex-1-enecarbonyl CoA and 3-hydroxycyclohex-1-enecarbonyl CoA. MycA4, a proposed hydratase/phosphatase, was inactive, attributed to either protein misfolding, or an alternative catalytic mechanism. Combined activity of ChCoADH (*Synthrophus aciditrophicus*), an acyl-CoA dehydrogenase and BadK (*Rhodopseudomonas pseudopalustris*), an enoyl-CoA hydratase, facilitated the conversion of cyclohexanecarbonyl CoA into 2-hydroxycyclohexanecarbonyl CoA. Activity analysis of four Hotdog-fold thioesterases, revealed RpaL (*Rhodopseudomonas palustris)* to have a promiscuous substrate acceptance, and to catalyse the required hydrolysis of 2-hydroxycyclohexanecarbonyl CoA to 2-hydroxycyclohexanocarboxylate.

A limited availability of CoA esters prevented the experimental analysis of many enzymatic steps. Resultantly, the CoA ligase, AliA (*Rhodopseudomonas pseudopalustris*), was screened and found to produce aromatic and alicyclic

CoA esters. Finally, a three-module system for the rational production of modified organic acids, via CoA ester intermediates, was tested. By coupling (1) broad AliA activity, (2) CoA modification (ChCoADH/ BadK) and (3) promiscuous RpaL activity, theoretically, a number of modified organic acids could be derived. Initial testing showed this system capable of producing cyclohex-1-enecarboxylate and 2-hydroxycyclohexanecarboxylate from cyclohexanecarboxylate. However, further exploitation of the CoA-utilising enzymes in endogenous metabolism, will undoubtedly expand this systems potential for rationally producing modified organic acids of commercial interest.

## Impact statement

The global use of petroleum-derived chemicals is notoriously finite and detrimental to the environment. Consequently, alternative technologies seeking to remove this petroleum dependence have the potential for profound impacts in both academic and industrial research, as well as possible international implications. Previous research at UCL developed a biosynthetic pathway from cyclohexanol to the nylon 6 monomer, 6-aminocaproic acid (6-ACA). However, its primary limitation was an inherent requirement of petroleum-derived cyclohexanol. The work presented here saught to tackle this limitation, by developing a sustainable petroleum-independent route to cyclohexanol. Despite cyclohexanol production not being fully realised, by characterising a number of key enzymatic steps, highlighting limitations and areas for improvement, as well as suggesting implementation strategies, this work will undoubtedly serve as a useful foundation for future academic investigations pursuing completion of this pathway.

Additionally, this work has highlighted a caveat to the use of CoA esters, which is a general lack of commercial availability. Therefore, the production of CoA esters directly through promiscuous CoA ligase activity, or through combined CoA ligase and CoA ester modification (both developed in this thesis), provides an alternative route for the production of currently unavailable CoA esters. This will have impacts within academic research, as it will facilitate the functional characterisation of CoA ester-utilising enzymes, aiding pathway design investigations such as this.

Furthermore, through the development of the cyclohexanol pathway, a CoA ester-utilising, organic acid modification system was proposed and developed. This system may be an extremely useful tool for the production of modified organic acids. Through rational design, modified organic acids produced in this manner may contain functionalities of direct industrial interest, or, represent useful synthons for more complex syntheses.

## Acknowledgements

I would firstly like to extend my gratitude and thanks to my PhD supervisors Professors John Ward and Helen Hailes for their support and guidance throughout the 4 years of this project. You are two of the most knowledgeable people I have ever met and I feel privileged to have been your PhD student.

Thank you to Damien Baud and Laure Benhamou for synthesising NAC and CoA esters and to Eve Carter for helping me submit, what seemed like an endless supply of HRMS samples.

Thank you to my friends in the department: Trish, Fair, Alex, Henry. You guys have kept me sane these last 4 years and I am really going to miss seeing you every day.

Finally, I would like to thank my parents Jayne and Steve, and my sister Rachel. The support and encouragement that you three have given me, through the ups and the downs, is what has kept me going through this PhD. I am endlessly grateful for everything you have done for me and so I would like to dedicate this thesis to you.

# Table of Contents

# List of Figures

## List of Tables

## List of abbreviations

| | |
|---|---|
| 6-ACA | 6-aminocaproic acid |
| 3-PS | 3-phosphoshikimate |
| CoA | Coenzyme A |
| CHC CoA | Cyclohexanecarbonyl CoA |
| 2-HCHC CoA | 2-hydroxycyclohexanecarbonyl CoA |
| 2-HCHC | 2-hydroxycyclohexanecarboxylic acid |
| TE | Thioesterase |
| DC | Decarboxylase |
| CHC | Cyclohexanecarboxylic acid |
| CH-2-eneC CoA | Cyclohex-2-enecarbonyl CoA |
| 3-HCHC CoA | 3-hydroxycyclohexanecarbonyl CoA |
| 3-HCH-1eneC CoA | 3-hydroxycylohex-1-enecarbonyl CoA |
| 3-HCHC | 3-hydroxycyclohexanecarboxylic acid |
| 3-HCH-1-eneC | 3-hydroxycylohex-1-enecarboxylic acid |
| CH-1-eneC CoA | Cyclohex-1-enecarbonyl CoA |
| CH-1-eneC | Cyclohex-1-enecarboxylic acid |
| 4-HCHC | 4-hydroxycylohexanecarboxylic acid |
| 4-HCH-1-eneC | 4-hydroxycylohex-1-enecarboxlic acid |
| 4-HCHC CoA | 4-hydroxycylohexanecarbonyl CoA |
| 4-HCH-1-eneC CoA | 4-hydroxycylohex-1-enecarbonyl CoA |
| CH-1, 3-eneC | Cyclohex-1,3-enecarboxylic acid |
| CH-1,3-eneC CoA | Cyclohex-1-,3-enecarbonyl CoA |
| CH-3-eneC | Cyclohex-3-enecarboxylic acid |
| CH-3-eneC CoA | Cyclohex-3-enecarbonyl CoA |
| 3,5-diCHC | 3,5-dihydroxycyclohexanecarboxylic acid |
| Bz | Benzoic acid |
| Bz CoA | Benzoyl CoA |
| 4-HCHC | 4-hydroxyclyclohexanecarboxylic acid |
| 4-oxoCHC | 4-oxocyclohexanecarboxylic acid |

| | |
|---|---|
| 2-oxoCHC | 2-oxocyclohexanecarboxylic acid |
| *t* 2-aCHC | *trans* 2-aminocyclohexanecarboxylic acid |
| DLBH CoA | DL-β-Hydroxybutyryl CoA |
| *n*HD CoA | *n*-heptadecanoyl CoA |
| *i*Val CoA | Isovaleryl CoA |
| *i*But CoA | Isobutyryl CoA |
| CHC NAC | Cyclohexanecarbonyl N-acetylcysteamine |
| HMG CoA | β-Hydroxy β-methylglutaryl CoA |
| DNA | Deoxyribonucleic acid |
| $Fc^+PF_6^-$ | Ferrocenium hexafluorophosphate |
| NAC | N-acetylcysteamine |
| DTNB | 5,5'-dithio-bis(2-nitrobenzoic acid) |
| TNB | 2-nitro-5-thiobenzoate |
| ATP | Adenosine triphosphate |
| HRMS (ES-) | High resolution mass spectrometry, negative electrospray |
| HRMS (ES+) | High resolution mass spectrometry, positive electrospray |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| HPLC | High-pressure liquid chromatrography |
| RBS | Ribosome binding site |
| ORF | Open reading frame |
| TB | Terrific broth |
| LB | Lysogeny broth |
| His-tag | Hexahistidine-tag |
| Rpm | Revolutions per minute |
| Tris | Tris(hydroxymethyl)aminomethane |

# 1. Chapter 1. Introduction

## 1.1. The advent of synthesizing novel organic compounds

The desirable active properties of organic compounds had long been known, with historical records of the therapeutic properties of willow bark, now known to contain salicin, dating back over 3,500 years to the Egyptians and Sumerians (Desborough and Keeling, 2017)**.** However, isolation from plant sources was the primary production route for these compounds, which entailed an inherent caveat on their use. In 1828 the German chemist Friedrich Wöhler serendipitously synthesised urea, proving for the first time that a simple organic compound could be derived from an inorganic source (Nicolaou, 2018). This was an extremely significant discovery as it opened the doors to an unexplored field of science and ultimately lead to the advent of the 'synthetic chemistry revolution'. With an ever improving understanding of molecular structures, chemists were able to start implementing rational design into their syntheses.

It was soon found that organic constituents of coal and oil could be valorized through chemical modification into useful compounds, mimicking those found in nature such as the dyes, indigo and alizarin (Welham, 1963), as well as the salicin derivate, salicylic acid. However, the medicinal capacity of salicylic acid was observed to be accompanied by unpleasant side effects such as nausea and tinnitus. In an effort to eradicate these side effects salicylic acid was modified into acetylsalicylic acid, commonly known as aspirin, which was found to have the same therapeutic efficacy without the unpleasant side effects (Sneader, 2000; Desborough and Keeling, 2017).

This notion of synthesizing novel organic compounds as more effective replacements to their naturally derived counterparts led to an exponential increase in the design and production of novel organic compounds. Through this process a vast new ‘space’ in organic chemistry was conceived that was no longer bound by the constraints imposed by natural products. This chemical design space allowed bespoke compounds to be synthesised with functionalities tailored towards specific industrial demands. One such novel compound was the synthetic polyamide, nylon, which was successfully developed as a synthetic alternative to natural polymers like silk (Keifer, 2000).

### 1.2. The synthetic polymer, nylon

Nylon is a generic name given to a family of synthetic polyamides known for their versatile applications as fibers and thermoplastics. Carothers developed the first member of this family, nylon 6,6 at DuPont’s research facility in 1935, which was the first fully synthetic fiber to be manufactured at an industrial scale in 1939 (Singh, 2011). His initial research focused on using dibasic acids and glycols to form polyesters. However, despite successful polymerisation, these polymers were found to have low melting points and a tendency to dissolve in most organic liquids, making them unsuitable industrial scale manufacture (Smith and Hounshell, 1985).

Carothers next used dibasic acids with diamines and found that these monomers could undergo a polycondensation reaction forming large polyamide chains, which were found to have much higher melting points than the former polyesters (Smith and Hounshell, 1985; Keifer, 2000). Most notably, adipic acid and hexamethylenediamine, through a polycondensation reaction, produced the first synthetic polyamide named ‘nylon 6, 6’ due to both the dibasic acid and diamine containing six carbon atoms. Specifically, equimolar amounts of adipic acid and hexamethylenediamine are firstly dissolved in methanol and mixed to form a salt. This salt can then undergo polymerisation when heated to 280 °C, forming nylon 6,6 (Chu, 2013) (Figure 1.1).

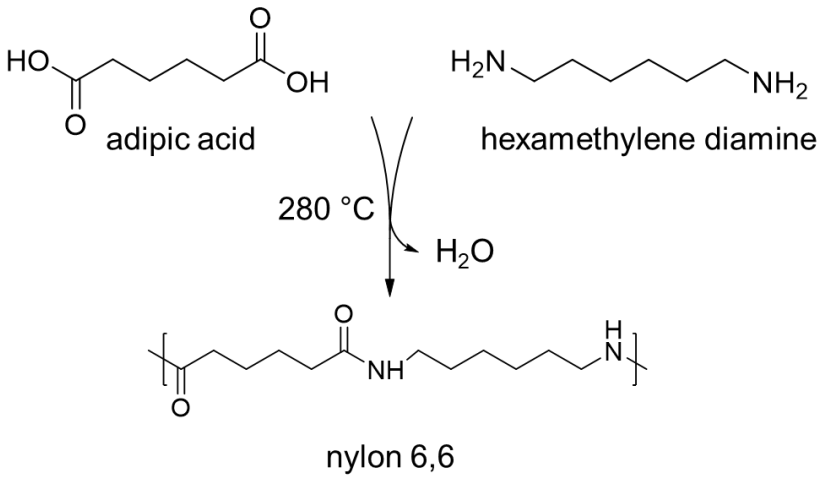



**Figure 1.1. The traditional chemical synthesis of nylon 6,6.**

Polymerisation of the synthetic polyamide nylon 6,6 occurs through heating adipic acid and hexamethylene diamine (nylon salt) to 280 °C forming water as a by-product.

This polyamide 'nylon 6,6', was found to have extremely desirable properties such as a high melting temperature (268.8°C), a strong resistance to abrasion and high tensile strength, and resultantly nylon 6,6 was targeted for industrial scale production. As a silk-like material when spun into fibers, it was initially targeted towards hosiery, where it was an immediate success and captured over 30% of the nationwide hosiery market within a year of it being introduced to the US market (Keifer, 2000). Additionally, the synthesized polymer could alternatively be molded into solid thermoplastics giving rise to a range of alternative applications. Furthermore, it was soon observed that variations in the carbon chain length of both the diacid and diamine monomers resulted in other viable alternative polyamides. These new polymers had dramatic variations in their structural properties, leading to nylon having a range of functionalities and thus many commercial applications including: shoe soles, automotive components, airbags, conveyor belts, toothbrush bristles, surgical sutures and many more (McKeen, 2014).

Following the development of nylon 6,6, DuPont filed a patent preventing competitors from synthesising this polyamide (Carothers, 1938). However, this

did not prevent alternative polyamides from being synthesized through alternative mechanisms; in 1938, Schlack, working at I.G Farbenindustrie, developed an alternative polymer that was produced from a single alicyclic amide monomer, Ɛ-caprolactam. Upon heating to 250 °C with 10% water, Ɛ-caprolactam undergoes ring cleavage and self-polymerisation resulting in a long polyamide chains with comparable properties to nylon 6,6 (Flynn, 2002; Barth, 2011; Chu, 2013) (Figure 1.2). This polyamide was named ‘nylon 6’ due to the six carbon atoms within its monomer, Ɛ-caprolactam.

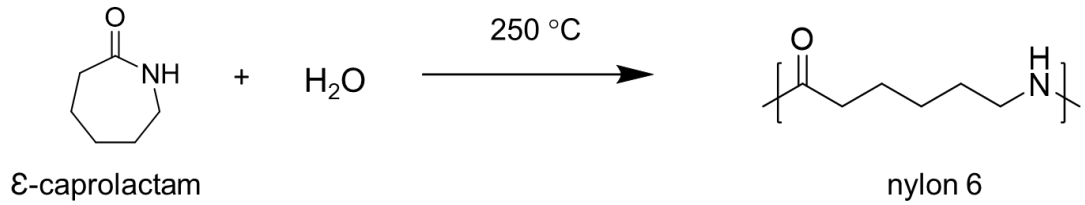



**Figure 1.2. The traditional chemical synthesis of nylon 6.**

Ɛ-Caprolactam is heated to 250 °C in the presence of 10% water resulting in ring cleavage to 6-aminocaproic acid (6-ACA), which polymerises into the polyamide, nylon 6.

Interestingly, despite the existence of alternative polyamides produced from various monomers, the combined production of nylon 6 and nylon 6, 6 account for approximately 90% of the current global nylon market. In 2016 global nylon production exceeded 6 million tons/year representing an estimated market size of USD 25.66 billion, which is predicted to rise to USD 41.13 billion by 2025 (Turk *et al.*, 2016; Grand view research, 2019). However, despite this vast production scale and the overall success of nylon as a novel synthetic polymer, there are a number of sustainability issues associated with its chemical production.

### 1.3. Limitations of chemically produced nylon

The requirement of petroleum-derived benzene as the main precursor to the substrates of both nylon 6 and nylon 6,6, reinforces the global dependency on a finite and environmentally detrimental carbon source (Rase, 2000; Zong *et al.*, 2017). The combustion of petroleum greatly contributes to the international issue of global warming, attributed to vast emissions of $CO_2$. In fact, an IPCC report showed that 78% of global greenhouse gas (GHG) emissions between 1970 to 2010, were attributed to the $CO_2$ derived from fossil fuel combustion (IPCC, 2014). Despite this, the global emission of petroleum-derived $CO_2$ has increased a further 5.6% between 2010 and 2016 (International Energy Agency, 2018). In addition, the finite nature of petroleum-derived fuels and chemicals means a dependency is fundamentally unsustainable in the long-term. Furthermore, the requirement of crude oil-derived benzene entails further caveats including: a limited availability (0.01% - 1% of crude oil), difficulties associated with its purification, as azeotropes cannot be removed through fractional distillation (Garcia Villaluenga and Tabe-Mohammadi, 2000) and an uncertain supply-demand chain, as benzene is often produced as a by-product of more in-demand products such as gasoline (Verma and Des Tombe, 2002). Moreover, the chemical synthesis of nylon 6,6 contains a number of undesirable steps. There are a range of alternative chemical routes to synthesising nylon 6,6, however, the conventional process requires the initial production of adipic acid. Adipic acid also has a number of chemical synthetic routes, all of which are derived from petrochemicals (Polen, Spelberg and Bott, 2013)**.** Its production begins with petroleum-derived benzene, which is reduced to cyclohexane. Cyclohexane then undergoes air oxidation, a process that requires both high temperatures (125-165°C) and high pressures (8 – 15 atm), producing KA oil, a mixture of cyclohexanol or cyclohexanone. In addition to requiring undesirable reaction conditions, this step is also inefficient with a 4-11% conversion. The KA oil is then used to produce adipic acid through its oxidative cleavage catalysed by the oxidant nitric acid (Figure 1.3), which is derived from natural gas (Ren *et al.*, 2009; Van Duuren *et al.*, 2011; Peñate *et al.*, 2012; Van De Vyver and Román-Leshkov, 2013).

However, this step is undesirable as not only is nitric acid a strong oxidizing agent, leading to the corrosion of reaction vessels and thus a safety hazard, it also emits nitrous oxide ($N_2O$), a putative greenhouse gas that contributes between 5-8% of global anthropogenic emissions, which was 5.4 million metric tons of $CO_2$ equivalent (MMT $CO_2$ Eq.) in 2014 (EPA, 2014; Hwang and Sagadevan, 2014).

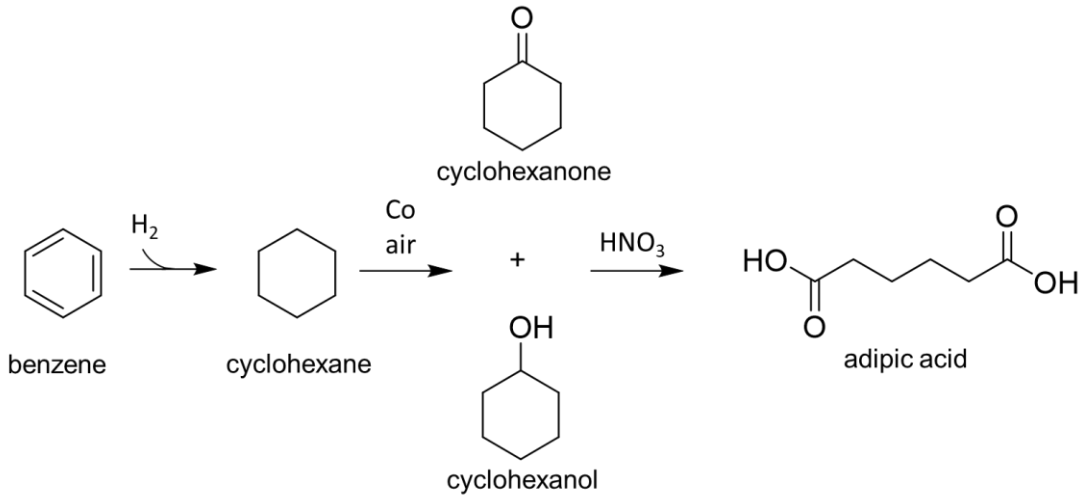



**Figure 1.3. The chemical production of adipic acid.**

Petroleum-derived benzene, is reduced to cyclohexane and then air oxidised to a KA oil mixture (cyclohexanone and cyclohexanol), which is finally oxidized to adipic acid.

It has been previously estimated that for every kilogram of adipic acid produced, approximately 300 g of $N_2O$ is released into the atmosphere (Hwang and Sagadevan, 2014). Despite this traditional production route, utilising nitric acid still accounts for approximately 95% of the global adipic acid production, which is 2.5 million tonnes/ year (Hwang and Sagadevan, 2014; Yuzawa *et al.*, 2018). Many investigations have looked into efficiently catalysing this process through different chemical mechanisms, however, it also often requires steps including high temperatures and pressures (Ueda *et al.*, 1997; Fang *et al.*, 2002; Deshayes *et al.*, 2008)**.** Furthermore, the primary chemical production route of hexamethylenediamine requires the use of petroleum-derived butadiene for the production of adiponitrile via

hydrocyanation, which is further hydrogenated in ammonia over a cobalt or iron catalyst to hexamethylenediamine (Dros *et al.*, 2015).

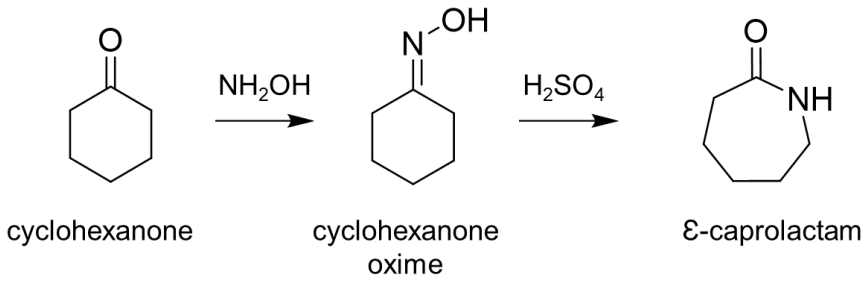



**Figure 1.4. The chemical production of Ɛ-caprolactam.**

The primary chemical production route of Ɛ-caprolactam from cyclohexanone via a cyclohexanone oxime intermediate.

A similar sustainability concern is inherent to the production of Ɛ-caprolactam, the main precursor of nylon 6, as there is a requirement of cyclohexanone, a component of petroleum-derived KA oil. Furthermore, the conversion of cyclohexanone into Ɛ-caprolactam requires the use of environmentally detrimental oxidising agents such as hydroxylamine sulfate (Carraher, 1978). Reacting these two compounds together forms a cyclohexanone oxime, which then further requires the use of highly dangerous oleum (fuming sulfuric acid) to produce Ɛ-caprolactam (Figure 1.4). In addition, the Ɛ-caprolactam production chain accounted for 3% of global emissions of the greenhouse gas nitrous oxide between 1990-2012 (Li *et al.*, 2014). It was recently estimated that the nitrous oxide emissions from Ɛ-caprolactam production in the United States in 2016 was 2 MMT $CO_2$ Eq. (EPA, 2018).

It seems clear that the environmental impacts associated with the continuous use of petroleum-derived substrates, combined with undesirable methods of chemical synthesis, limits the overall sustainability of producing nylon 6 and 6, 6. As a result, alternative production technologies are required that can directly tackle these issues, thereby confronting the global reliance on petroleum-derived chemicals. Various mandates have been set in attempts to facilitate the gradual transition to sustainable alternatives, for example the European

council targeted a 10% blend of bioethanol with petroleum by 2020 and various carbon taxes set internationally (Howard *et al.*, 2013). With authoritative bodies likely to further incentivise the development and industrial uptake of sustainable alternatives, combined with the finite and environmentally detrimental impacts associated with the use of petroleum derived chemicals, it is inevitable that alternative chemical synthetic routes will become the industrial standard (ICS, 2007). One such alternative approach may be through use of synthetic biology.

### 1.4. Synthetic biology as an alternative to chemical synthesis

Synthetic biology is a field of science that was officially recognized towards the end of the 20$^{th}$ century, when molecular biology began to actively incorporate the use of engineering principles to manipulate cellular organisms towards a desired function (Cameron, Bashor and Collins, 2014). The use of technologies such as recombinant gene cloning, DNA sequencing and DNA synthesis allowed scientists to isolate and heterologously express genes from organisms of interest as well as design and construct novels genes, leading to enzymes with unique functions. Moreover, through the use of engineering principles, characterised genes can be rationally used as building blocks and, in a bottom-up approach, used to produce novel genetic systems, with unique functionalities. Examples of these functionalities include; genetic toggle switches, oscillators and biosensors (Collins, Gardner and Cantor, 2000; Atkinson *et al.*, 2003; Nguyen *et al.*, 2010; Rogers *et al.*, 2015).

In addition to these functions, synthetic biology, combined with metabolic engineering techniques, can be used to construct genetic pathways designed to re-route metabolic flux towards the production of specific chemicals. In this way, genetically engineered organisms containing recombinant genetic pathways, combined with specific genomic alterations, can be considered cell factories for chemical production. Furthermore, advances in synthetic biology over the last two decades, including the development of high-throughput

automated processes, combined with *in silico* design and modeling, has provided a unique platform technology for the industrial up-scale of cell factories (Jullesson *et al*. 2015). In this manner, the metabolic production of chemicals may present an opportunity to tackle a number of limitations inherent to chemical production routes.

As mentioned previously, chemical production routes, such as those associated with nylon 6 and 6,6, often require petroleum-derived substrates, imparting a reliance on a finite and unsustainable source. However, constructing an organism with a metabolic pathway that can produce a desired chemical from an intermediate of central metabolism, would link the production of the desired chemical with the organism's growth. Therefore, instead of directly using a petroleum-derived substrate, a renewable carbon source (such as glycerol from biodiesel waste, or lignocellulosic biomass hydrolysates) could be used as the feedstock for a modified organism, thereby, indirectly functioning as a substrate for the production of the desired compound. Furthermore, implementation of enzymatically-catalysed metabolic pathways bypasses the requirement for harsh organic solvents associated with chemical synthesis processes. Moreover, the use of *in vivo* synthetic pathways means undesirable high pressures would no longer be required and reaction temperatures could be reduced to that of the optimal growth temperature of the chassis organism. In this way alternative biological synthetic routes may represent a more sustainable production method to chemicals.

As a result of the clear potential of engineering microbial cell factories for chemical production, a number of companies from a range of sectors, including food, fuel, chemical and pharmaceutical, have invested into research in this area (Jullesson *et al*. 2015). Examples of chemicals that are commercially produced through cell factories include D-lactic acid (Cargill), Isobutanol (Gevo), Succinic acid (Bioamber), Farnesene (Amyris) and Biodiesel (Solazyme) (Jullesson *et al*. 2015). Yet, despite these examples there remains to be a major caveat to the commercialization of biosynthetically

derived compounds, which is the associated low product titers and yields (Curran and Alper, 2012). Understanding the complexity of metabolic systems can be extremely challenging, and as a result, product titers can often be limited by competitive reactions, toxicity of intermediates, unwanted back-reactions and rate limiting steps. Moreover, traditional chemical production routes have infrastructure in place that has been purposely refined to maximise production efficiency and provide the system with reliably high yields. For industries to switch to alternative metabolic production routes, it would not only require the implementation of new infrastructure, and thus a high capital investment, it would also mean the transition from a pre-existing reliable production chain to a novel process with uncertain reliability. To overcome this inherent barrier to both industrial integration and direct commercialization of bio-based chemicals, increasing product titers and yields are of paramount importance for providing this technology with a competitive edge and making it more attractive to investors.

Genomatica have shown that an *E. coli* strain could be transformed with a biosynthetic genetic pathway (BDO pathway) and produce the non-natural commodity chemical, 1, 4-butanediol (1, 4-BDO). Initial production yields were approximately 18 g/l (Yim *et al.*, 2011). However, through iteratively re-designing components of the host metabolism, such as eliminating drains of redox and ATP, as well as developing a synthetic biology tool box for systematic and combinatorial fine-tuning of the biosynthetic pathway, metabolic flux could be forced towards 1,4–butanediol, reaching an industrially competitive titer of >140 g/L (Andreozzi *et al.*, 2016; Culler, 2016)**.**

Furthermore, biosynthetic pathways have also shown commercial success as an alternative route to the production of organic compounds currently derived from plant extracts, most notably with the antimalarial compound artemisinin. Traditional production methods are performed through the cultivation and isolation of artemisinin from the dried leaves and influorescences of the herb *Artemisia annua*. However, due to the long growth cycle of *A. annua* (12-18

months), combined with its sensitivity to environmental change, both the harvests and price of artemisinin are prone to fluctuation. In addition, the production yields are as low as 0.5% (Hale *et al.*, 2007), where from 1000 kg of dried leaves, 5 kg of Artemisinin is recovered. As a result, it was previously estimated that a total of 17,000 hectares would have to be dedicated to growing *A. annua* to meet the global demand of 100 million treatments each year (Hale *et al.*, 2007). Moreover, the downstream purification process predominantly requires petroleum ether for the extraction of artemisinin (Lapkin *et al.*, 2010; Zhang *et al.*, 2018), further imposing an environmental limitation to this production chain.

In order to overcome these limitations, the semi-synthetic artemisinin project, funded by the Bill and Melinda Gates foundation was founded. The primary goal of this project was to design a semi-synthetic pathway to artemisinin using a biosynthetic pathway to produce artemisinic acid, which could then be chemically converted into artemisinin. For this biosynthetic pathway a *Saccharomyces cerevisae* strain was developed, overexpressing nine genes from its endogenous mevalonate pathway as well as heterologously expressing five genes from *A. annua*. This engineered strain of *S. cerevisae* was observed to achieve industrially competitive titres of 25 g/L artemisinic acid (Paddon and Keasling, 2014). This compound could then be isolated and chemically converted to artemisinin. The current chemical conversion has been shown to have a 55% yield of metabolically-derived artemisininic acid to artmesinin and is used to manufacture 60 metric tons/ year (Kung *et al.*, 2018). This example clearly underscores the potential of using synthetic biology to design and construct novel metabolic systems that can provide alternative routes for the synthesis of specific chemicals. In doing so, limitations previously associated with traditional production routes are bypassed and a more environmentally friendly production chain is created. Furthermore, areas of land that may have previously been dedicated to the growth of *A. annua* may be freed for the growth of food crops.

Although synthetic biology is still in its infancy, it has already shown to be a useful technology, with positive economic, environmental and sustainability implications for industry. With ongoing research and literature published in this field the application of synthetic biology to create efficient metabolic systems to replace current chemical synthetic routes is likely to become more common. Furthermore, with the cost and speed of technologies such as DNA sequencing and synthesis constantly improving, cloning procedures that often require a lot of time and troubleshooting can be bypassed by direct DNA synthesis. As a result, this will increase both the speed and feasibility of constructing genetic systems consisting of multiple genes (Jullesson *et al*. 2015). In addition, the application of engineering techniques such as Design of Experiment (DoE), where a systematic approach is used to efficiently identify an optimal output, can further improve the speed and ability to identify optimal conditions required for the production of industrially competitive titers (Kumar, Bhalla and Rathore, 2014). With investigations such as these showing industrially competitive product titers, combined with global concerns such as a reliance of petrochemicals and global warming, it is inevitable that industry will turn to alternative technologies such as this for chemical synthesis. With this in mind, identifying chemical production chains that contain useful processes and address sustainability issues, are prime targets for the design and implementation of alternative biological production routes.

One such chemical is the nylon 6 monomer, 6-aminocaproic acid (6-ACA), which is the linear form of Ɛ-caprolactam. As previously mentioned, nylon 6 is currently produced at a vast industrial scale, however, production of its monomer substrate requires harsh chemicals, high temperatures and pressures, as well as the production of environmentally detrimental by-products (Carraher, 1978; Li *et al.*, 2014). In addition, the substrates used for chemical synthesis are sourced from crude oil thus contributing to the global reliance on such finite sources. Therefore, this production chain is an obvious target for the construction of an alternative biological production system.

## 1.5. Biological synthesis of nylon 6

The requirement for a renewable alternative route to the production of the nylon 6 monomer, Ɛ-caprolactam, has motivated a number of research groups into developing novel pathways. Many groups have focused on chemo-enzymatic routes where biological systems are used to produce specific metabolites, which are then chemically converted into the final product. For example, Frost *et al* developed and patented a novel route, whereby metabolically derived lysine was chemically converted into the precursor Ɛ-caprolactam (Frost, 2005). In addition, a novel route to 6-aminocaproic acid was proposed via a chemical step from lysine to 6-aminohex-2-enoic acid, followed by an enzymatic conversion to 6-aminocaproic acid (6-ACA) (Raemakers-Franken *et al.*, 2005; Turk *et al.*, 2016). These novel production routes are a move in the right direction, however, they do not fully curb the requirement of petroleum-derived chemicals.

More recently, research undertaken by Turk *et al* has developed a fully fermentative route to the production of 6-ACA from glucose. This pathway was designed to target α-ketoglutarate (AKG) from the citric acid cycle and through chain elongation, catalysed by NifV from *Azotobacter vinelandii*a and AksD, AksE and AksF from *Methanococcus aeolicus*, coupled with the activity of the decarboxylase, KdcA, from *Lactococcus lactis* and a transaminase from *Vibrio fluvialis* (Vfl), lead to the production of 6-ACA (Figure 1.5.) (Turk *et al.*, 2016). The results from this investigation showed this pathway to successfully synthesise 6-ACA from AKG. As such, this system represents a desirable route to 6-ACA, as glucose is acting as both the carbon source, required for growth, and the substrate for the production of 6-ACA. Furthermore, the use of glucose removes the need for petroleum-derived chemicals. However, as with many novel metabolic pathways, the shortfall of this system was its low final titre of 6-ACA (160 mg/L), imparting an initial barrier to upscaling.

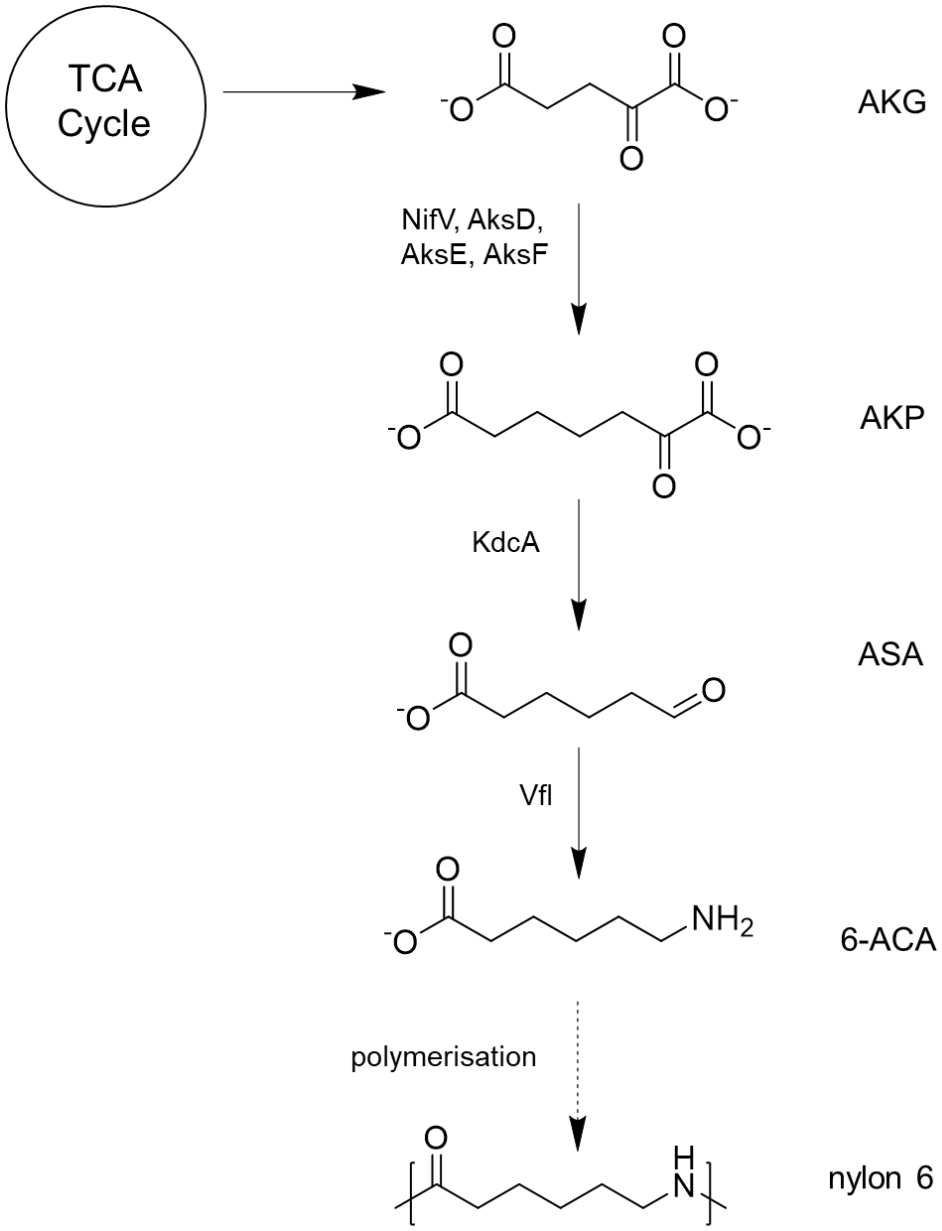



**Figure 1.5. The AKP biosynthetic pathway developed previously.**

The AKP biosynthetic pathway developed by Turk *et al* 2016. Here, α-ketoglutarate (AKG) from central metabolism is converted to α-ketopimelate (AKP), by the combined activity of NifV *Azotobacter vinelandii*), AksD, AksE and AksF (*Methanococcus aeolicus*). AKP is then decarboxylated to adipate semialdehyde (ASA) by KdcA activity. Finally the transaminase, Vfl, converted ASA into 6-aminocaproic acid (6-ACA). 6-ACA is the linear form of Ɛ-caprolactam and can form the polyamide, nylon 6, through polymerisation (dashed line). However this final polymerisation was not undertaken in the investigation.

Interestingly, this study showed that the total titre of related pathway intermediates exceeded 2 g/L, suggesting that through optimisation, this final titre might be significantly improved (Turk *et al.*, 2016). Of course, with previous optimisation studies showing significant increases in titres are achievable, such as 1,4–butanediol (1,4-BDO) at Genomatica, it is not unfeasible to suggest this to be a viable production route in the future. However, currently this pathway represents a strong proof of principle for the complete metabolic production of the nylon 6 monomer, 6- ACA.

## 1.6. Metabolic production of the nylon 6 monomer, 6-ACA, from cyclohexanol

Previous research at UCL has also been motivated to develop an alternative route for the production of the nylon 6 precursor. Specifically, doctoral research developed a biosynthetic route to 6-ACA from cyclohexanol. This pathway functions through the activity of four enzymes from *Acinetobacter calcoaceticus*: (1) ChnA, a cyclohexanol dehydrogenase, (2) ChnB, a cyclohexanone monooxygenase, (3) ChnC, a dehydrogenase and (4) ChnD, a dehydrogenase, coupled with the activity of a transaminase (TAm) from *Pseudomonas putida* (Figure 1.6.). This pathway has been fully constructed and expressed within *E. coli* and has so far shown the system to successfully convert cyclohexanol into 6-ACA, with a titer of 441.4 mg/L (Jackson, 2017). However, due to the requirement for cyclohexanol as a substrate, this system remains to be dependent on a non-renewable, petroleum-derived chemical, limiting its economic and environmental viability.

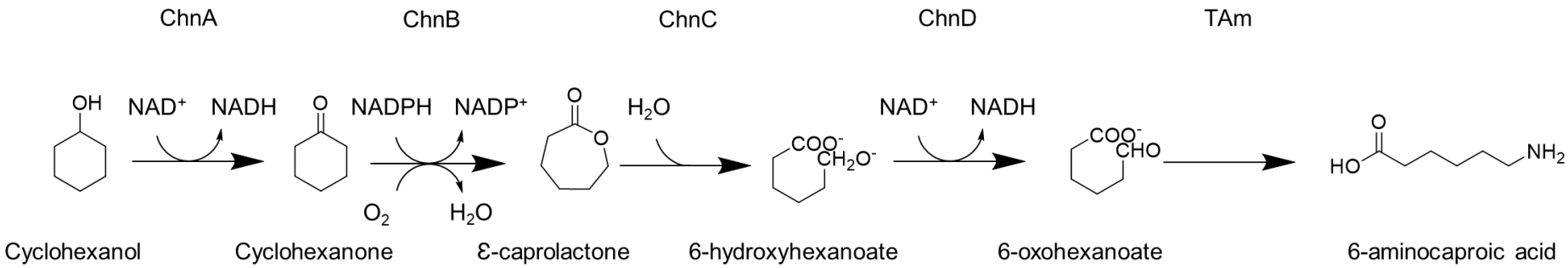



**Figure 1.6. The 6-ACA pathway developed at UCL.**

The biosynthetic pathway previously constructed at UCL converts cyclohexanol into the linear nylon 6 precursor, 6- aminocaproic acid (6-ACA). This biotransformation requires four enzymes from *Acinetobacter calcoaceticus*: (1) ChnA, a cyclohexanol dehydrogenase, (2) ChnB, a cyclohexanone monooxygenase, (3) ChnC, a dehydrogenase and (4) ChnD, a dehydrogenase and a transaminase (TAm) from Pseudomonas putida.

Therefore, the research presented in this thesis aims to further develop the described biosynthetic pathway, by removing its requirement for petroleum-derived cyclohexanol. This will be achieved through the provision of an additional biosynthetic pathway that can synthesise cyclohexanol from central metabolism. In doing so, cyclohexanol would no longer be required as an external substrate and instead the organism's carbon source would indirectly be the substrate of the pathway. Furthermore, if this system could be designed to utilize carbon derived from non-petroleum, non-food sources, such as hemicellulose-derived glucose or xylose, it would represent a renewable production process.

## 1.7. The biosynthetic cyclohexanol pathway

The biosynthetic pathway described within this study consists of eight enzymes that will convert either shikimate or 3-phosphoshikimate (3-PS), from the shikimate pathway of central metabolism, into cyclohexanol (Ghosh, Chisti and Banerjee, 2012). The construction of this pathway was split into four sections, each focused on synthesising a specific intermediate of the pathway: (A) Cyclohexanecarbonyl CoA (CHC CoA) production, (B) 2-hydroxycyclohexanecarbonyl CoA (2-HCHC CoA) production, (C) 2-hydroxycyclohexanecarboxylic acid (2-HCHC) production and (D) cyclohexanol production (Figure 1.7). The development and characterisation of sections A, B and C are described within chapters 3, 4 and 5, respectively.

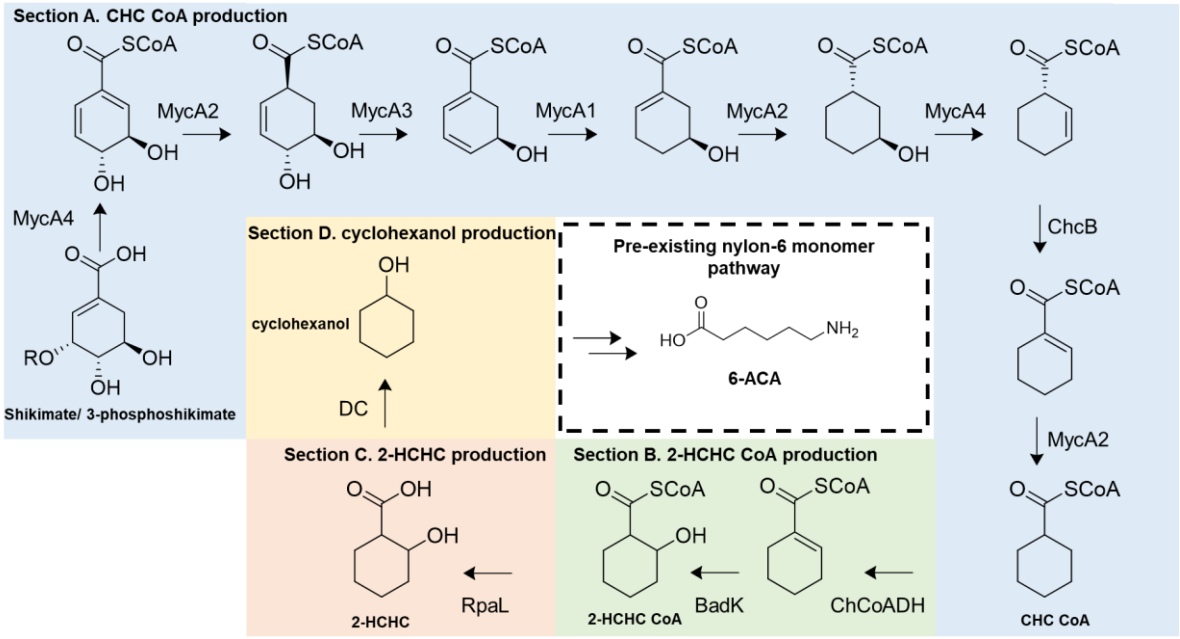



**Figure 1.7. The proposed biosynthetic cyclohexanol pathway.**

Development of the pathway was split into four sections: Section A (blue box) will link the production of CHC CoA from intermediates of central metabolism (shikimate or 3-phosphoshikimate) using four enzymes (MycA1-4) from the CHC CoA pathway endogenous to *S. rishiriensis.* Section B (green box) will use ChCoADH a dehydrogenase from *Syntrophus aciditrophicus* and BadK from *Rhodopseudomonas pseudopalustris*, for the production of 2-HCHC CoA from CHC CoA. Section C (orange box) will use the thioesterase (TE) activity of RpaL from *Rhodopseudomonas palustris*, for the production of 2-HCHC from 2-HCHC CoA. Finally, section D (yellow box) will use a decarboxylase (DC) activity to produce cyclohexanol from 2-HCHC. Cyclohexanol could then be used as a substrate for production of the nylon 6 monomer, 6-ACA, via the biosynthetic nylon 6 monomer pathway (dashed box).

### 1.7.1. Section A. CHC CoA production from central metabolism

The first section (section A) of the biosynthetic cyclohexanol pathway requires the activity of four enzymes that are endogenous to *Streptomyces rishiriensis*. These enzymes are encoded within a four gene operon (*mycA1, mycA2, mycA3* and *mycA4*) (CHC CoA operon), that is putatively required for the synthesis of the cyclohexanecarboxylic acid (CHC)-side chain of mycotrienin I and mycotrienin II (Sugita *et al.*, 1982). These four enzymes are: (1) MycA1, a 2,4-dienoyl-CoA reductase, (2) MycA2, a 1-cyclohexenylcarbonyl coenzyme A reductase, (3) MycA3, an acyl-CoA dehydrogenase and (4) MycA4, a bifunctional hydratase/phosphatase-CoA ligase (based on proposed function). The CHC CoA operon from *S. rishiriensis* was specifically identified for a number of reasons outlined below.

In a previous investigation, a homologous CHC CoA operon from *S. collinus* was observed to be essential for the synthesis of a (CHC)-derived side chain of the polyketide ansatrienin A (mycotrienin I) (Cropp, Wilson and Reynolds, 2000). Here, the CHC CoA operon was isolated from *Streptomyces collinus* and cloned into a pSE34 shuttle vector, which was expressed within two alternative species: *S. avermitilis* and *S. lividans*. The wild-type of these two species are unable to produce ansatrienin A, or any other CHC-containing polyketides (Cropp, Wilson and Reynolds, 2000). However, heterologous expression of the CHC CoA operon from *S. collinus* resulted in the production of doramectin, an antiparasitic CHC-containing avermectin, as well as high levels of ω-cyclohexyl fatty acids. Furthermore, it was shown that the deletion of c*hcA* (an *mycA2* homolog), resulted in an *S. collinus* strain unable to synthesis ansatrienin A (Wang *et al.*, 1996; Cropp, Wilson and Reynolds, 2000). The catalytic mechanism of the enzymes encoded within this operon has been proposed, through deuterium and $^{13}C$ incorporation studies, to function through an eight-step reduction pathway, starting with either shikimic acid or shikimate-3-phosphate and ending with CHC CoA (Figure 1.8) (Moore

*et al.*, 1993; Cropp, Wilson and Reynolds, 2000). The combined results of these investigations, confirmed the requirement of this operon for synthesising CHC CoA. Therefore, it was identified for incorporation within the biosynthetic cyclohexanol pathway.

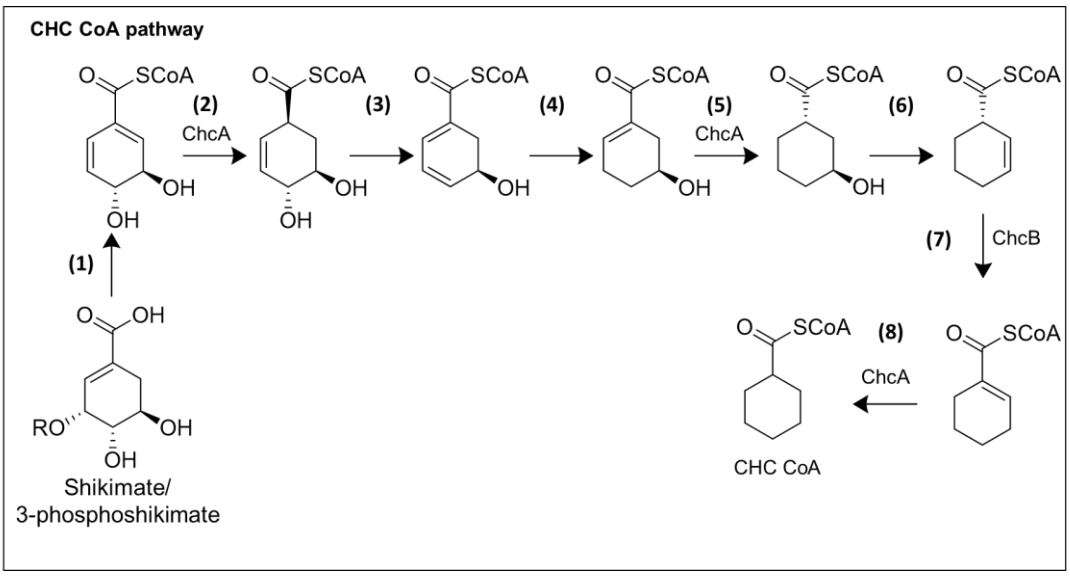



**Figure 1.8. The CHC CoA pathway.**

Shows the proposed 8-step route from shikimate or 3-phosphoshikimate, catalysed by enzymes encoded within the CHC CoA operon from *S. collinus*. The three steps catalysed by ChcA (2, 5, 6) and single step by ChcB (7) are the only characterised steps, all others are proposed from deuterium and $^{13}C$ incorporation studies.

Due to associated patents and an incomplete DNA sequence, obtaining the CHC CoA operon from *S. collinus* was not possible. Furthermore, at the beginning of this investigation there were no commercially available organisms with a sequenced genome and an annotated CHC CoA operon. Therefore, a number of polyketide-producing organisms were assessed for the identification of a commercially available organism that produced a polyketide containing a CHC side chain. This process lead to the identification of *S. rishiriensis,* which had been previously shown to produce the CHC-containing polyketides: mycotrienin I, mycotrienin II and trienomycin A (Sugita *et al.*, 1982), and so it was predicted that this organism would contain a CHC CoA operon that would catalyse the same biotransformation as the homologous

operon in *S. collinus*. Therefore, the genome of *S. rishiriensis* was to be assessed for the CHC CoA operon for incorporation within the cyclohexanol pathway.

With the exception of ChcA and ChcB, which have both been functionally characterized previously (Wang *et al.*, 1996; Chen *et al.*, 1999; Patton, Cropp and Reynolds, 2000), the proposed function of all enzymes within this cluster are attributed to their amino acid sequence identities. Homologs of this gene cluster have been found in a number other *Streptomycetes* that are also producers of certain polyketides containing a CHC side chain, including *S. flaveolus* and *S. seoulensis* (Qu, Lei and Liu, 2011; Song *et al.*, 2015). However, only the CHC CoA pathway from *S. collinus* has been functionally analysed. Therefore, the proposed biosynthetic cyclohexanol pathway was designed using the CHC CoA pathway from *S. collinus* as an initial reference point.

Analysis of the intermediates produced through the natural CHC CoA pathway was next required to determine the most suitable intermediate to be re-routed through to the production of cyclohexanol. The intermediate 3-hydroxycyclohexanecarbonyl CoA (3-HCHC CoA) was initially considered (product of step 5- Figure 1.8). Theoretically, 3-HCHC CoA could then by hydrolysed to 3-HCHC by thioesterase (TE) activity and finally decarboxylated to cyclohexanol by decarboxylase (DC) activity. However, MycA4 is proposed to catalyse the biotransformation of 3-HCHC CoA to cyclohex-2-enecarbonyl CoA (CH-2-eneC CoA), in addition to the initial step of the pathway. As a result, this second catalytic function would directly compete with TE activity and so presents an obstacle to targeting the 3-HCHC CoA intermediate toward cyclohexanol production. Therefore, an alternative route was considered that would utilise CHC CoA, derived from the CHC CoA pathway, which would then be enzymatically hydroxylated by the activity of two enzymes involved in the degradation of cyclohexane carboxylic acid (CHC) from *Rhodopseudomonas pseudopalustris* and *Syntrophus aciditrophicus*.

### 1.7.2. Section B. Biosynthesis of 2-HCHC CoA from CHC CoA

Following the production of CHC CoA by the sequential activity of MycA1-4, the following enzymatic steps will functionalise the CoA-activated CHC with a hydroxyl group. This bioconversion would be catalysed by two enzymatic steps. The first will use the previously characterized cyclohexanecarboxylic-CoA dehydrogenase (ChCoADH) from *Syntrophus aciditrophicus*, an enzyme required in the fermentation of benzoate, to derive cyclohex-1-enecarbonyl CoA (CH-1-eneC CoA) from CHC CoA (Kung *et al.*, 2013). The second enzyme, an enoyl-CoA hydratase (BadK) from *Rhodopseudomonas pseudopalustris,* required in the degradation of CHC, will convert CH-1-eneC CoA into 2-hydroxycycloxanecarboxylic acid (2-HCHC CoA) (Egland *et al.*, 1997).

When initially undertaking the work within this section, the uncharacterized acyl-CoA dehydrogenase (BadJ) from *R. pseudopalustris* was identified as the most suitable enzyme for producing CH-1-eneC CoA from CHC CoA. BadJ is encoded within the same operon as BadK and catalyses the preceding step within the CHC degradation pathway, which forms a branch into benzoate degradation (Figure 1.9). However, following activity analysis (see chapter 4) BadJ was found to be inactive when used in vitro, and so ChCoADH from *S. aciditrophicus* was identified as a suitable replacement. This second step, catalysed by BadK has not previously been experimentally shown, however, its function has been inferred from its amino acid sequence and so was to be assessed within this investigation. Therefore, the initial focus of this section of the investigation was to isolate the genes encoding BadJ (*badJ*) and BadK (*badK*) from the CHC operon in *R. psuedopalustris* and then individually assess their enzyme activity. Following characterisation, their activity would be coupled, facilitating the production of 2-HCHC CoA from CHC CoA.

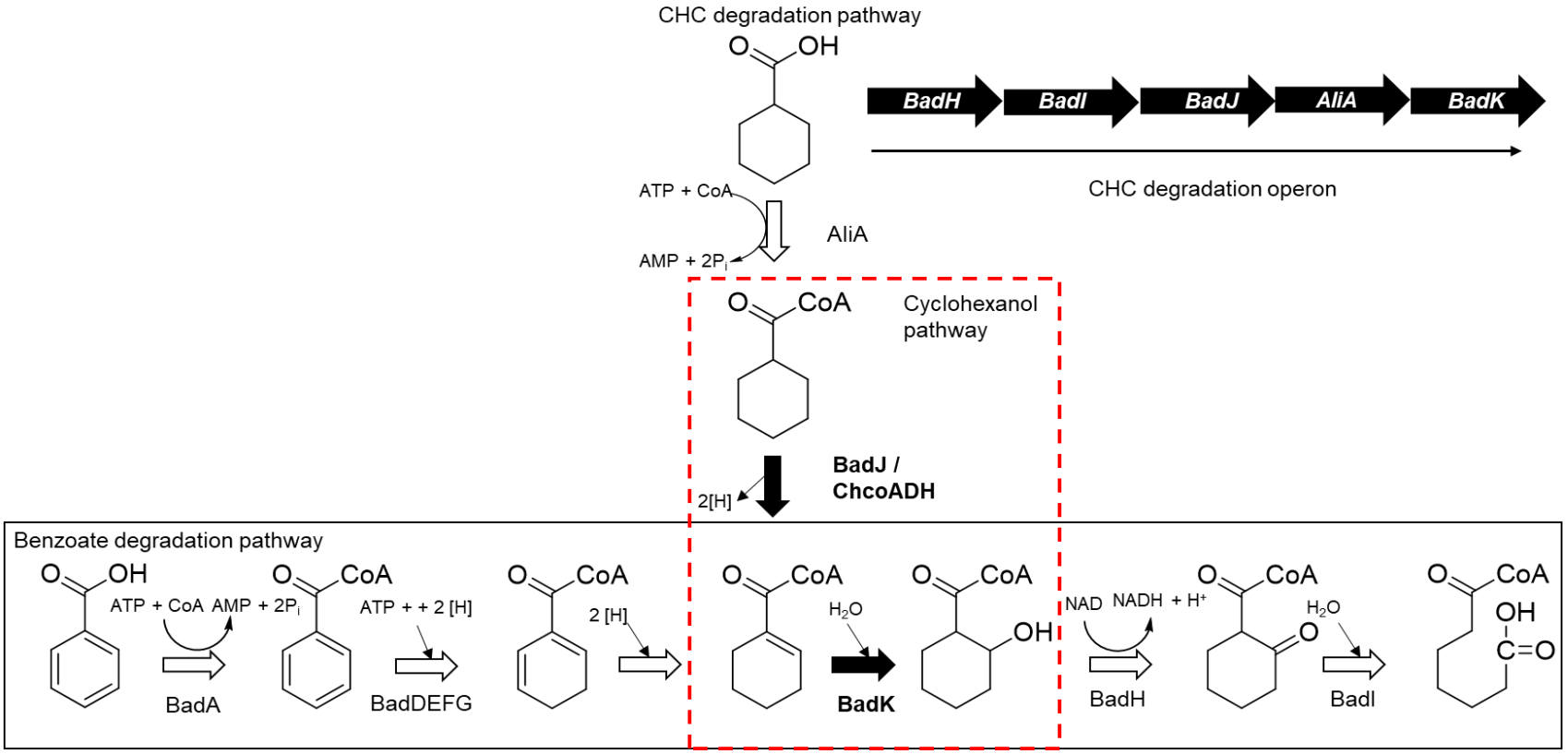



**Figure 1.9. The anaerobic degradation of benzoate and CHC.**

The anaerobic degradation pathway of benzoate (black box) and cyclohexane carboxylic acid (CHC) from *R. pseudopalustris.* The two enzymatic steps (filled black arrows) identified for incorporation within the biosynthetic cyclohexanol are highlighted (red box). The dehydrogenase ChCoADH from S. *aciditrophicus* is also shown as it catalyses the same biotransformation as BadJ. The five gene CHC operon encoding the genes of BadH, BadI, BadJ, AliA and BadK within *R. pseudopalustris* is shown.

### 1.7.3. Section C. 2-HCHC production from 2-HCHC CoA

Following the production of 2-HCHC CoA the next enzymatic step requires thioesterase (TE) activity to hydrolyse the thioester bond forming 2-hydroxycyclohexanecarboxylic acid (2-HCHC) and CoA (Figure 1.10). However, no previous TE had been experimentally shown to be able to catalyse this step, therefore, analysis of the literature was required to identify suitable candidate TEs to catalyse the biotransformation.

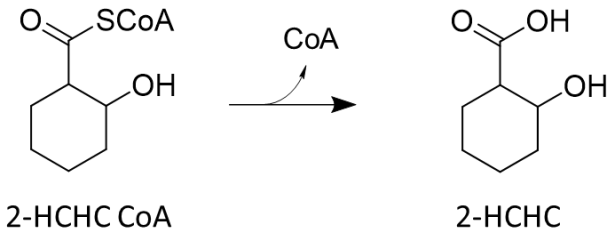



**Figure 1.10. The required TE activity with 2-HCHC CoA.**

The TE activity requires the hydrolysis of 2-hydroxycyclohexanecarbonyl CoA (2-HCHC CoA) into 2-hydroxycyclohexanecarboxylic acid (2-HCHC) and CoA.

TEs (EC 3.1.2.1-27) are a diverse and widespread class of enzyme that are categorized into superfamilies I and II, which are subcategorized into 25 families based on their primary and tertiary structures on the ThYme (Thioester-active enzYmes) database (Gonzalez *et al.*, 2012). Superfamily I consist of $\alpha/\beta$ hydrolase-fold TEs and superfamily II share a common 'hotdog' fold, coined due to a seven-stranded anti-parallel $\beta$-sheet, referred to as the 'bun', being wrapped around a five-turn $\alpha$-helix 'sausage' (Dillon and Bateman, 2004). Both classes of TEs function by catalysing the hydrolysis of thioesters, however, members of family I are commonly a component of a multi-subunit domain protein such as a polyketide synthases (PKS), whereas superfamily II are commonly independent enzymes (Lenfant *et al.*, 2013). Furthermore, superfamily II have been previously shown to target both CoA and Acyl Carrier Protein (ACP) thioesters with promiscuous activity.

The Hotdog fold domain was initially identified within FabA from *Escherichia coli* and subsequently in 4-hydroxybenzoyl-CoA (4-HB-CoA) TE from *Pseudomonas* sp. Strain CBS. Despite TEs within this superfamily having a conserved hotdog-fold a lot of variation has been found within both the tertiary and quaternary structures. The tertiary structure of these TEs has been shown to range from hotdog-fold monomers to fusion proteins containing two tandem hotdog fold domains, as well as hotdog fold domains linked to domains with different functions (Dillon and Bateman, 2004). Furthermore, the crystal structures of hotdog fold TEs resolved to date show a plethora of different quaternary structures, including hotdog-fusion monomers, homodimers (front to front and back to back), tetramers and hexamers (Pidugu *et al.*, 2009).

With this understanding it was clear that members of the Hotdog-fold superfamily would represent the most appropriate candidate enzymes due to be being found as independent enzymes and having been shown to have promiscuous activity towards CoA esters. Four TEs were identified as suitable candidates for isolation and activity analysis with 2-HCHC CoA, these included: 1- FcbC from *Arthrobacter sp*. Strain AU, 2- PA2801 from *Pseudomonas aeruginosa* PAO1, 3- RpaL from *Rhodopseudomonas palustris* HaA2 and 4- YbdB from *E. coli* K-12 MG1655. All four of these TEs were members of subfamilies within the Hotdog-fold superfamily, with FcbC, PA2801 and YbdB within the 4-HB-CoA subfamily and RpaL within the TesB-like subfamily.

However, an immediate obstacle for this section of the investigation was the lack of commercial availability of 2-HCHC CoA. As a result, activity analysis of these four TEs had the prerequisite of successful BadJ (or ChCoADH) and BadK activity to produce 2-HCHC CoA from commercially available CHC CoA. Furthermore, as all of the intermediates of the cyclohexanol pathway prior to 2-HCHC CoA are CoA esters, the ideal TE would have specific activity towards 2-HCHC CoA and no activity with any of the other preceding CoA ester intermediates. Therefore, each of the four TEs was also screened with CHC

CoA and CH-1-eneC CoA (derived from ChCoADH activity) to determine the specificity of each TE.

### 1.7.4. Section D. cyclohexanol production from 2-HCHC

The final step of the proposed cyclohexanol pathway is mediated by decarboxylase (DC) activity, converting 2-HCHC into cyclohexanol with the release of $CO_2$ (Figure 1.11). There are no examples within the literature of a DC that is able to catalyse this step, therefore, a number of candidate DCs will need to be identified and screened with 2-HCHC. However, one should consider that the β-hydroxyl group within 2-HCHC may limit the viability of identifying an enzyme that can catalyse this step, as decarboxylation steps often require a β-ketone (Clayden, Greeves and Warren, 2012). As previously shown, the second compound in the previously developed, 6-ACA pathway is cyclohexanone (Figure 1.6). Therefore, methods for enzymatically oxidising 2-HCHC to 2-oxocyclohexanecarboxylic acid (2-oxoCHC), or directly producing 2-oxoCHC from the cyclohexanol pathway, would provide an additional decarboxylation route into the 6-ACA pathway (Figure 1.11).

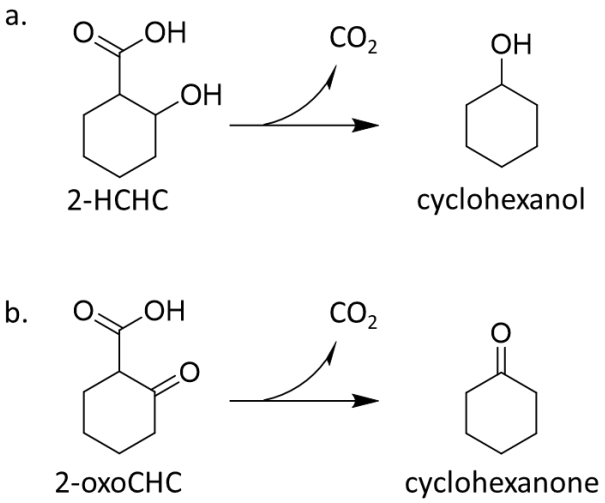



**Figure 1.11. The required decarboxylation step.**

Two decarboxylation steps were proposed, required for (a) the decarboxylation of 2-HCHC into cyclohexanol (the substrate of the 6-ACA pathway) and (b) the decarboxylation of 2-oxoCHC into cyclohexanone (the second compound in the 6-ACA pathway).

Unfortunately, due to a lack of time, this section of the cyclohexanol pathway could not experimentally explored within this investigation. However, a route for the production of the alternative DC substrate, 2-oxoCHC, is enzymatically explored within chapter 4.

### 1.7.5. Development of a novel organic acid production system

A recurrent caveat to characterising enzymatic steps within the biosynthetic cyclohexanol pathway, was the general lack of commercially available CoA esters. As a result, considerable work was focused on biologically synthesising CoA esters, to be used as substrates for enzymatic assays. Therefore, the final chapter of this thesis outlines the work undertaken for the production of a range of CoA esters and the resultant proposal of a biological system through which to synthesis modified organic acids.

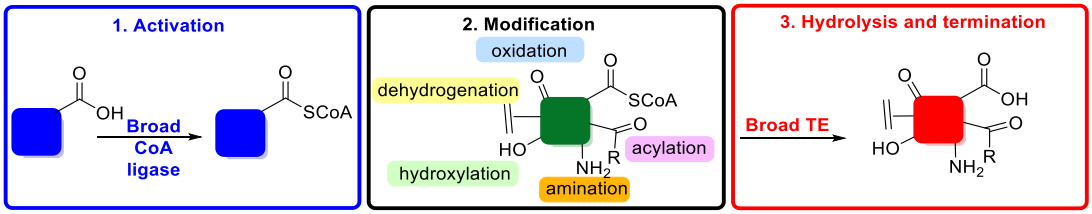



**Figure 1.12. The organic acid modification system.**

The 3 modules of the proposed biosynthetic organic acid modification system. Module 1 (activation) will utilise a broad spectrum CoA ligase to activate a range of organic acids into their corresponding CoA esters. Module 2 (modification) will add bespoke functional complexity to the CoA ester using enzymes from a developed toolbox of CoA-dependent enzymes. Module 3 (termination) will utilise a broad spectrum TE to hydrolyse a range of CoA esters into their corresponding organic acids.

Initial work focused on identifying a broad spectrum CoA ligase that could derive alicyclic CoA esters, such as the intermediates of the CHC CoA pathway from *S. rishiriensis*. As seen previously (Figure 1.9), AliA from *R. pseudopalustris* catalyses the first step in the CHC degradation pathway converting CHC into CHC CoA and is encoded within the same CHC operon

as BadJ and BadK. Therefore, AliA was isolated and its capacity for the enzymatic production of a range of CoA esters was assessed. However, once a number of CoA ester modifying enzymes had been isolated, from *S. rishiriensis* (MycA1-4), *R. pseudopalustris* (BadJ, BadK) and *S. aciditrophicus* (ChCoADH), as well as four TEs, a novel biosynthetic system for the synthesis of a myriad of organic acids, via CoA ester intermediates, soon became apparent. This proposed system would function in three defined modules (Figure 1.12).

The first module (activation) is a common step using a broad spectrum CoA ligase, facilitating the activation of a plethora of organic acids into their corresponding CoA esters. The second module (modification) is a specific step employing one or more CoA ester modification enzymes to specifically functionalise the CoA ester as desired. The final module (termination) is another common step that would utilize a broad spectrum TE to hydrolyse the modified CoA ester into its corresponding organic acid. This novel biosynthetic system has two desirable functionalities, the first is the synthesis of a range of CoA esters, particularly relevant for screening of CoA-dependent enzymes within the cyclohexanol pathway, and the second, for the synthesis of modified organic acids.

CoA is an ubiquitous and essential cofactor that is used by over 4% of the total enzymes within *E. coli* alone (Strauss, 2010). The formation of a CoA ester through a covalent thioester bond between an acyl group and the terminal thiol group of a CoA, facilitates an innumerable number of biotransformations. These biotransformations are found in both central metabolism, including: fatty acid oxidation, and the biosynthesis of glycolipids and sterol, as well as secondary metabolism, including: the biosynthesis of polyketides, non-ribosomal protein synthesis, flavonoids and lignin (Webb and Smith, 2011). The wealth of enzymes known to utilise CoA esters for a variety of different biochemical processes emphasizes the potential of the proposed biosynthetic for the production of a myriad of CoA esters, and following TE activity, modified

organic acids. Furthermore, as the substrate acceptance of CoA ester modifying enzymes is explored and expanded beyond their native substrates, possibly though rational mutagenesis, the modification potential of this system will expand further.

As explained at the beginning of this chapter, the advent of synthesising natural and novel organic acids was a major discovery, ultimately revolutionizing the field of chemistry and leading to the development of a number of industries including textiles and pharmaceuticals. This was due to the properties inherent to many organic acids. The conception of a biosynthetic system for the rational synthesis of modified organic acids, as described here, unites synthetic biology with this desire to design and synthesise novel organic acids. The work shown within this investigation (see chapter 6) provides an initial proof of principle for this system, however, as enzyme characterisation is expanded so may the viability and the scope of this organic acid modification system, as a biosynthetic route for the production of novel organic acids, that may directly have interesting functionalities or represent desirable synthons for further syntheses.

### 1.8. Hypothesis and aims

The hypothesis of this investigation was that a biosynthetic pathway can be developed to produce cyclohexanol from an intermediate of central metabolism. This hypothesis was tested by developing and functionally analysing the cyclohexanol pathway in defined sections (Figure 1.7), with each section forming a chapter within this thesis. Theoretically, cyclohexanol produced in this way could then be converted to the nylon 6 monomer, 6-ACA, using a previously developed pathway (Figure 1.6). The requirement of CoA ester-utilising enzymes within the cyclohexanol pathway presented an interesting route through which to modify organic acids. Therefore, a novel biosynthetic organic acid modification system was also developed and

assessed for proof of principal functionality. The specific aims of this project were as follows:

The first aim of this investigation was to sequence the *S. rishiriensis* genome and identify the CHC CoA operon that it was predicted to contain. Following identification and determination of the open reading frames (ORF) of each gene (*mycA1-4)*, they were to be individually cloned into pET28a expression vectors for recombinant expression and purification of the corresponding proteins. Then the activity of each enzyme was to be characterized, however, the proposed substrates of the CHC CoA pathway (shikimic acid and 3-phosphoshikimate) were the only commercially available compounds. Therefore, the characterisation of each enzyme would have to follow the sequential pathway order (i.e. the product of the first enzyme is the substrate of the next) (see chapter 3).

The second aim was to identify, construct and characterise an enzymatic route from CHC CoA to 2-HCHC CoA. Initially this was planned to be through sequential BadJ and BadK (*R. pseudopalustris*) activity. Therefore, the primary aim of this section of work was to clone these two genes, and recombinantly express and purify their corresponding enzymes. Following issues with characterising BadJ, the alternative acyl CoA dehydrogenase (ChCoADH) from *S. aciditrophicus,* which had previously been characterised, was identified for synthesis and activity analysis (see chapter 4).

The third aim was to identify and clone a number of thioesterases (TEs) to then be recombinantly expressed into their corresponding enzymes, with the aim of finding a TE that could hydrolyse 2-HCHC CoA into 2-HCHC. Four TEs were identified for this aim: FcbC, PA2801, RpaL and YbdB. This aim entailed the prerequisite of successful completion of the second aim for the production of commercially unavailable 2-HCHC CoA (see chapter 5).

The final aim was to develop a proof of principle for the proposed organic acid modification system. For the first (activation) module, the CoA ligase (AliA) from *R. pseudopalustris* was to be cloned, expressed and purified, for subsequent activity screening with a number of CoA esters. For the second (modification) module, CoA ester modifying enzymes cloned within the previous aims were to be screened with CoA esters derived from AliA activity, for the development of modified CoA esters. Finally, for the third (termination) module, the four TEs recombinantly expressed within the third aim were to be assessed for promiscuity. The most promiscuous TE was to then be purified and characterised further, and finally coupled with the activation and modification modules to provide a proof of principle for the proposed organic acid modification system (see chapter 6).

# 2. Chapter 2. Materials and methods

## 2.1. Bacterial strains

A total of six organisms were experimentally used within this investigation (Table 2.1). *Streptomyces rishiriensis* (DSM 40489) and *Rhodopseudomonas pseudopalustris* (DSM 123) were both ordered from the Leibniz institute (© DSMZ). Upon receiving samples, freeze-dried organisms were re-suspended in 25% glycerol, which was then used to inoculate recommended media. *R. pseudopalustris* (DSM 123) was grown on solid 'media 27' in the dark at room temperature for 72 hours. *Streptomyces rishiriensis* (DSM 40489) was grown in liquid 'media 65' for 48 hours at 28 °C. Following growth, these cultures were mixed with glycerol (20% final concentration) and stored as 1 ml aliquots at -80 °C. *Pseudomonas aeruginosa* PAO1 and *Escherichia coli* K21 MG1655 were not cultured in this investigation, however, their purified in-house genomes (extracted previously) were used as template DNA for PCR amplification of *pa2801* and *ybdB*, respectively. *E. coli* BL21 (DE3) and TOP10 cells were used in transformation procedures and subsequent recombinant protein expression and DNA replication (see section 2.2).

**Table 2.1. Table of organisms used.**

All organisms that were used within experimental procedures, the source from which they were obtained and their respective genotypes.

| Organism | Source | Genotype |
|---|---|---|
| *Escherichia coli BL21* (DE3) | Invitrogen ™ | F– *ompT hsdS*$_B$ ($r_B^-$, $m_B^-$) *gal dcm* (DE3) |
| *Escherichia coli* TOP10 | Invitrogen ™ | F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ– *rps*L(StrR) *end*A1 *nup*G |
| *Escherichia coli* K12 MG1655 | In-house collection | F- lambda- *ilv*G- *rfb*-50 *rph*-1 |
| *Streptomyces rishiriensis* (DSM 40489) | Leibniz institute (DSMZ) | wild-type |
| *Rhodopseudomonas pseudopalustris* (DSM 123) | Leibniz institute (DSMZ) | wild-type |
| *Pseudomonas aeruginosa* PAO1 | In-house collection | wild-type |

## 2.2. Transformation procedures

Recombinant plasmids produced through cloning methods were used to transform One shot ™ chemically competent *Escherichia coli* TOP10 cells (Invitrogen ™ cat#C404003), from which the commercial protocol was followed (available at: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/oneshottop10_chemcomp _man.pdf). For growth of *E.coli* TOP10 strains on solid media, LB agar containing appropriate antibiotic was uniformly inoculated, using a sterile spreader and incubated overnight at 37 °C in a static incubator (Galaxy R $CO_2$ incubator). Transformed cells grew as single colonies, which were picked, using a sterile inoculation loop, and used to inoculate 10 ml of Lysogeny Broth (LB) media containing appropriate antibiotics. Cultures were then incubated overnight at 37 °C in a shaking incubator, shaking at 250 rpm (New Brunswick Innova 4330 Incubator shaker). Following overnight growth, 500 µl of culture was added to 500 µl of 50% glycerol and then stored at -80 °C for long-term storage. *E. coli* TOP10 cells were used primarily for the long-term storage (-80 °C) and subsequent replication and isolation of recombinant plasmids. For recombinant protein expression, purified recombinant plasmids were used to transform chemically competent one shot ™ *E. coli* BL21 (DE3) cells (Invitrogen ™ cat# C600003) according to the commercial protocol (available at:https://assets.thermofisher.com/TFS-Assets/LSG/manuals/oneshotbl21_ man.pdf.) For the co-transformation of recombinant plasmids with the Takara chaperone plasmids (Takara Bio Inc.), both plasmids were simultaneously used to transform chemically competent one shot ™ *E. coli* BL21 (DE3) cells. Following the addition of 50 µg of each plasmid to the thawed cells the standard commercial protocol was used.

## 2.3. Chemicals, media and stock solutions

The standard working concentration of antibiotics used were: 50 μg/ml Kanamycin (KAN), 25 μg/ml Chloramphenicol (CAM) and 5 ng/ml Tetracycline (inducer). 1000x stocks of all antibiotics were prepared, filter sterilised (0.2 μm) and stored at -20 °C. Isopropyl-β-D-galactopyranoside (IPTG) was used at a working concentration of 0.1-1 mM and was prepared as a 1 M stock and stored at -20 °C. L-Arabinose was prepared as a 200 mg/ml stock solution and used at a working concentration of 0.5 mg/ml. Premade LB agar and LB broth (VWR) was used for growth of all *E. coli* cultures. *R. pseudopalustris* (DSM 123) was grown on solid 'media 27' made according to: https://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium27.pdf, with the addition of 1.5% agar. *S. rishiriensis* (DSM 40489) was grown in liquid 'media 65' made according to: https://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium65.pdf.

All CoA and NAC esters were prepared as 10 mM stock solutions in water and stored at -80 °C. 10 mM stock solutions of all organic acids, used for screening CoA ligase activity, were prepared in water and stored at 4 °C. ATP (NEB) was prepared as 100 mM stock solutions in water and stored at -80 °C. DTNB (Sigma-Aldrich) was prepared as a 10 mM stock solution in 100 mM TrisHCL pH8 and prepared fresh for each use. Ferrocenium hexafluorophosphate (Insight Biotechnology limited) was prepared fresh prior to each use as a 10 mM stock solution. All CoA ester substrates used were bought from Sigma-Aldrich with the exception of CHC CoA and the NAC ester, CHC NAC. These two substrates were chemically synthesised by Laure Benhamou and Damien Baud within the UCL Chemistry Department. CHC CoA later became commercially available and was purchased from CoALA bioscience (Austin, Texas). The source of organic acid substrate used for determining CoA ligase activity was: cyclohexanecarboxylic acid (CAS-98-89-5), benzoic acid (CAS- 65-85-0), 3,5-dihydroxycyclohexanecarboxylic acid (CAS-804428-26-0), 4-oxocyclohexanecarboxylic acid (CAS- 874-61-3), 3-phosphoshikimate

(CAS-63959-45-5), *trans*-2-aminocyclohexanecarboxylic acid (CAS-5691-19-0) and 2-oxocyclohexanecarboxylic acid (CAS-18709-01-8) were all purchased from Sigma Aldrich. *cis/trans*-3-hydroxycyclohexanecarboxylic acid (CAS-606488-94-2) and 4-hydroxycyclohexanecarboxylic acid (CAS-17419-81-7) were purchased from Insight biotech limited and cyclohex-3-enecarboxylic acid (CAS-4771-80-6) was purchased from Fischer scientific and shikimic acid was bought from Generon (CAS- 138-59-0).

## 2.4. Recombinant plasmids and plasmid extractions

For the extraction of recombinant plasmids glycerol stocks of transformed *E. coli* TOP10 cells were used to inoculate 10 ml of Lysogeny Broth (LB) media containing appropriate antibiotics. This culture was then incubated overnight at 37 °C in a shaking incubator (shaking at 250 rpm). Following overnight growth, 10 ml cultures were pelleted through centrifugation (Eppendorf centrifuge 5810R) (10,000x *g*, 10 minutes) and plasmids were purified according to the QIAprep spin Miniprep kit (Qiagen) protocol. A list of all recombinant plasmids is shown in Table 2.2.

**Table 2.2. A table of all constructed recombinant plasmids.**

Shown are all of the recombinant plasmids constructed in this investigation and the chapter in which they were produced.

| Chapter | gene/ operon | Organism | Vector | plasmid name |
|---|---|---|---|---|
| **Chapter 3** | CHC CoA operon (*mycA4321*) | *S. rishiriensis* | pET29a | pQR2786 |
| | *mycA1* | *S. rishiriensis* | pET28a | pQR2787 |
| | *mycA2* | *S. rishiriensis* | pET28a | pQR2788 |
| | *mycA3* | *S. rishiriensis* | pET28a | pQR2789 |
| | *mycA4* | *S. rishiriensis* | pET28a | pQR2790 |
| **Chapter 4** | *badJKH* | *R. pseudopalustris* | pET29a | pQR2791 |
| | *badJ* | *R. pseudopalustris* | pET28a | pQR2792 |
| | $badJ_{ATG}$ | *R. pseudopalustris* | pET28a | pQR2793 |
| | *badK* | *R. pseudopalustris* | pET28a | pQR2794 |
| | $badK_{ATG}$ | *R. pseudopalustris* | pET28a | pQR2795 |
| | *badH* | *R. pseudopalustris* | pET28a | pQR2796 |
| | $badH_{NheI}$ | *R. pseudopalustris* | pET28a | pQR2797 |
| | *chcoADH* | *S. aciditrophicus* | pET29a | pQR2798 |
| **Chapter 5** | *fcbC* | *Arthrobacter sp* . Strain AU | pD451-SR | pQR2799 |
| | *pa2801* | *P. aeruginosa* PAO1 | pET29a | pQR2800 |
| | *rpaL* | *R. palustris* | pET29a | pQR2801 |
| | *ybdB* | *E. coli* K-12 MG1655 | pET29a | pQR2802 |
| **Chapter 6** | *aliA* | *R. pseudopalustris* | pET28a | pQR2803 |
| | $rpaL_{His}$ | *R. palustris* | pET28a | pQR2804 |

## 2.5. DNA synthesis

The sequence manipulation suite online tool (available at: www.bioinformatics.org/sms2/rev_trans.html) was used for codon optimisation of *rpaL*, *fcbC* and *chcoADH* for expression in *E. coli*. The gene, *fcbC,* was synthesised by ATUM (California, United States) into pD451-SR, an IPTG inducible vector (pQR2799). The gene, *rpaL*, from *Rhodopseudomonas palustris* HaA2 was synthesised by Eurofins (Luxemburg) into pET29a and designed to contain an N-terminal *Nde*I restriction site and a C-terminal *Xho*I restriction site (pQR2801). These restriction sites were used to clone *rpaL* into the expression vector, pET28a (pQR2804). *chcoADH,* was synthesised and sub-cloned into pET29a by GenScript ® (New Jersey, United States), its open reading frame (ORF) encodes a C-terminal His-Tag (pQR2798).

## 2.6. Polymerase chain reaction (PCR)

**Table 2.3. The thermocycler conditions used for all PCRs.**

| Step | Temperature | Time |
|---|---|---|
| Initial denaturation | 98 °C | 2 minutes |
| 35 cycles | 98 °C | 10 seconds |
| | 50-72 °C | 30 seconds/ kb |
| Final extension | 72 °C | 2 minutes |
| Hold | 4 °C | |

Each PCR contained: 0.02 U/µl Q5 High-Fidelity DNA polymerase (NEB), 200 µM dNTPs, 1x Q5 Reaction buffer, 0.5 µM of each primer and template DNA (genomic- 1 ng- 1 µg, plasmid- 1 pg- 1 ng), which was brought to a final volume of 25 µl with Milli-Q water (MilliPore Q-POD). If the GC content of the construct was above 60% then 1x GC enhancer was used. 10 ng of plasmid DNA or 1 µg of genomic DNA was used as the template. Thermocycler conditions used for PCR amplification are shown in Table 2.3. When initially amplifying a DNA

fragment a temperature gradient was used for the annealing step ± 5 °C from the primer with the lowest annealing temperature. Primers were designed to have annealing temperatures within 2 °C of each other, all primers sequences used in this investigation are shown below in Table 2.4.

## 2.7. Circular polymerase extension cloning (CPEC)

The two genes, *pa2801* and *ybdB,* were amplified from the reference genomes: *P. aeruginosa* PAO1 (NC_002516.2) and *E. coli* K12 MG1655 (NC_000913.3), respectively. For the identification of *pa2801*, the characterised protein (accession: 3QY3_A) (Gonzal*ez et al*, 2012) was used to identify the nucleotide sequence (Gene ID: 879843) within the *P. aeruginosa* PAO1 reference genome. For identification of y*bdB* the previously characterised enzyme (accession: P0A8Y8) was used to search the reference genome (NC_000913.3) leading to the identification of the nucleotide sequence (Gene ID: 945215) used in this study. The genes *pa2801* and *ybdB* were cloned into pET29a using Circular Polymerase Extension Cloning (CPEC), following the methods stated previously (Quan and Tian, 2011). Here, primers were required for both amplification and introduction of overlapping regions between the vector and inserts (underlined in Table 2.4), allowing complementary base-pair binding and polymerase extension for a successful CPEC reaction.

**Table 2.4. A table of all primers sequences used.**

A list of all primers sequences and cloning method used for each gene. Underlined sequence represent overlapping sequences between vector and insert required for CPEC.

| Gene | Forward primer sequence | Reverse primer sequence | Cloning method |
|---|---|---|---|
| ***pa2801*** | 5'-CGAGGAGAGAATTCATGGCTGACAGACAATTGC-3' | 5'-GCTTTGACTAGTTCAGGCGATCGCGGCG-3' | CPEC |
| **pET29a *(pa2801)*** | 5'-CGATCGCCTGAACTAGTCAAAGCCCGAAAGGAAG-3' | 5'- GTCAGCCATGAATTCTCTCCTCGAATTTAGCAGCAGCGG-3' | CPEC |
| ***ybdB*** | 5'-GCTAAATTCGAGGAGAGAATTCATGATCTGGAAACGCCATTTAAC-3' | 5'-CTTTCGGGCTTTGACTAGTTCATCCCAAAACTGCCG-3' | CPEC |
| **pET29a *(ybdB)*** | 5'-CGGCAGTTTTGGGATGAACTAGTCAAAGCCCGAAAG-3' | 5'-CCAGATCATGAATTCTCTCCTCGAATTTAGCAGCAGCGG-3' | CPEC |
| ***badJ*** | 5'-GAGTCACTGTCATATGGTGAATCCTTATCTGAACGACG-3' | 5'-TCTACTCAAGGGATCCTCAGGCCGGCAC-3' | Restriction-ligation (*Nde*I + *Bam*HI) |
| ***badJ*** $_{ATG}$ | 5'-GAGCTCACTGTCATATGAATCCTTATCTGAACGACG-3' | 5'-TCTACTCAAGGGATCCTCAGGCCGGCAC-3' | Restriction-ligation (*Nde*I + *Bam*HI) |
| ***badK*** | 5'-TATTCTCAGTGGATCCAGGAGATTCAGAGTGCCAGAGAACG-3' | 5'-GTGTCAACTTGAAGCTTTCAGCGATGGACGAAGC-3' | Restriction-ligation (*Bam*HI + *Hin*dIII) |
| ***badK*** $_{ATG}$ | 5'-GAGCTCGACTGTCATATGCCAGAGAACGTG-3' | 5'-GTGTCAACTTGAAGCTTTCAGCGATGGACGAAGC-3' | Restriction-ligation (*Nde*I + *Hin*dIII) |
| ***badH*** | 5'-TTCACGAGTTAAGCTTAGGAGATGTAGTATGGCGCGTCTGC-3' | 5'-GAGGTACTTGCTCGAGTCAACCGTTCATCGTCAGC-3' | Restriction-ligation (*Hin*dIII + *Xho*I) |
| ***badH*** $_{NheI}$ | 5'-TTCACTAGTTGCTAGCAGGAGATGTAGTATGGCGCGTCTGC-3' | 5'-GAGGTACTTGCTCGAGTCAACCGTTCATCGTCAGC-3' | Restriction-ligation (*Nhe*I + *Xho*I) |
| ***mycA4*** | 5'-AGCGTGGACTCATATGCACCTTCTCGTCAAAGG-3' | 5'-ACTACTTCGACTCGAGTCAGCCGGCGTCCTC-3' | Restriction-ligation (*Nde*I + *Xho*I) |
| ***mycA3*** | 5'-AGCGTGGACTCATATGACGGACGAGCAG-3' | 5'-ACTACTTCGACTCGAGTCATCGGTTCGTTCCC'3- | Restriction-ligation (*Nde*I + *Xho*I) |
| ***mycA2*** | 5'-AGCGTGGACTCATATGAACAGCCCGCAC-3' | 5'-ACTACTTCGACTCGAGTCAGGCATCGCGC-3' | Restriction-ligation (*Nde*I + *Xho*I) |
| ***mycA1*** | 5'-AGCGTGGACTCATATGTTGGACCAGGTCTTCAG-3' | 5'-ACTACTTCGACTCGAGCTACCCGGTCCTCCTC-3' | Restriction-ligation (*Nde*I + *Xho*I) |
| ***mycA4321*** | 5'-TGGGCTAGGCTTGCATATGCACCTTCTCGTCAAAG-3' | 5'-AATACATAACTCTCGAGCTACTACCCGGTCCTCCTC-3' | Restriction-ligation (*Nde*I + *Xho*I) |
| ***aliA*** | 5'-AGCGTGGACTCATATGGAATTCGATGCCG-3' | 5'-ACTACTTCGACTCGAGCTAAAGGTCGTTCTGC-3' | Restriction-ligation (*Nde*I + *Xho*I) |

## 2.8. Standard restriction-ligation cloning procedure

For the production of all other recombinant plasmids, a standard procedure requiring (1) PCR, (2) restriction endonuclease digestion and (3) T4 ligase activity was used. Primers were used to add synthetic flanking restriction sites, as well as an additional 5' region containing 10 nucleotides to facilitate the restriction endonucleases binding and digesting the linear PCR product. Following amplification, PCR products were separated on agarose gels (1%) and identified PCR products were purified from the gel using the Monarch ® gel extraction kit (NEB). Purified PCR products were digested overnight at 37 °C with relevant restriction endonucleases (Table 2.4) and pET28a or pET29a were digested overnight with the same restriction endonucleases. Following incubation, 1 µl of alkaline phosphatase (CIP) (NEB) was added to the vector digestion and incubated at 37 °C for one hour to remove phosphate monoesters, preventing self-ligation of the vector. Digested vectors and PCR products (inserts) were run on 1% agarose gels, from which they were assessed and purified. Following purification, the vector (0.02 pmol) and insert (0.02 pmol) were then ligated overnight at room temperature at a 1:1 molar ratio using T4 ligase (NEB). Ligation reactions were then heat shocked at 65 °C and 5 µl of the reaction was used to transform One shot ™ chemically competent *E. coli* TOP10 cells (Invitrogen ™ cat#C404003) (see chapter 2.2). Colonies were then assessed for the recombinant plasmid through digestion using the same restriction endonucleases as used in the cloning procedure. Plasmids that contained an insert of the correct nucleotide length were then sequenced using the commercial T7 forward primer and T7 terminator reverse primer (Eurofins).

## 2.9. Recombinant protein expression

Following overnight growth, 1:50 volume of a starter culture was added to either terrific broth (TB) or LB liquid media in a conical flasks containing 50 µg/ml KAN. A spectrophotometer (Jenway 7315) was then used to monitor

culture growth at 600 nm**.** Once an optical density of 0.5-0.8 $OD_{600}$ (exponential growth) was achieved, cultures were induced with 0.1-1 mM Isopropyl-β-D-galactopyranoside (IPTG) and further incubated (time, temperature and IPTG concentration stated in results) for protein expression. Following induced growth cultures were pelleted and transferred into 50 ml falcon tubes and were either immediately used or stored at -20 °C. Co-expression of recombinant plasmids with the Takara chaperone plasmids required liquid growth media to be prepared with 25 μg/ml CAM in addition to KAN. Additionally, chaperone inducers were added to liquid media prior to inoculation (Table 2.5). Once cultures were observed to reach exponential phase they were induced with 0.1-1 mM IPTG and grown as stated in the results.

**Table 2.5. The Takara chaperone plasmids.**

A table of all the Takara chaperones plasmids, their encoded chaperone proteins and the respective inducer for each plasmid.

| Plasmid | chaperones | Inducer |
|---|---|---|
| pG-Tf2 | groES - 10 kDa<br>GroEL - 60 kDa<br>tig - 56 kDa | L-Arabinose<br>Tetracycline |
| pTF16 | tig - 56 kDa | L-Arabinose |
| pGro7 | groES - 10 kDa<br>GroEL - 60 kDa | L-Arabinose |
| pKJE7 | dnaK - 70 kDa<br>dnaJ - 40 kDa<br>grpE - 22 kDa | Tetracycline |
| pG-KJE8 | dnaK - 70 kDa<br>dnaJ - 40 kDa<br>grpE - 22 kDa<br>groES - 10 kDa<br>GroEL - 60 kDa | L-Arabinose |

## 2.10. Clarification of cell lysate using Bugbuster

For the lysis of *E.coli* BL21 (DE3) cells using Bugbuster protein extraction reagent (Novagen, Inc. 70584-3), cultures were pelleted through centrifugation

for 20 minutes at 10,000 rpm (Eppendorf centrifuge 5810R) and re-suspended in Bugbuster in a 1/25$^{th}$ volume of the original culture volume. After a 15 minute incubation at room temperature, suspensions were centrifuged for 20 minutes at 4 °C and 11,500 rpm (Eppendorf centrifuge 5424R), separating soluble protein from the insoluble components of the cell suspension.

### 2.11. Clarification of cell lysate using sonication

For the lysis of *E.coli* BL21 (DE3) cells through sonication, cultures were pelleted through centrifugation for 20 minutes at 10,000 rpm (Eppendorf centrifuge 5810R) and re-suspended in a 1/25$^{th}$ volume of buffer, either 50 mM Tris-HCL pH 7 or IMAC 1 buffer (see chapter 2.13) for purification. Re-suspended cells were then lysed using a Soniprep 150 (© 2018 MSE Centrifuges). Sonication conditions used were: 10 seconds on, 10 seconds off for 10 cycles. Re-suspended cells were kept on ice throughout sonication. Following sonication, lysed cells were clarified through centrifugation for 20 minutes at 4 °C and 11,500 rpm (Eppendorf centrifuge 5424R). The resultant supernatant was composed of the soluble protein and the pellet composed of the insoluble protein and cell debris. Soluble protein could then be used for enzyme assays, recombinant protein purification or SDS gel analysis.

### 2.12. Re-suspension of insoluble protein

Insoluble pellets produced through centrifugation of lysed cells were re-suspended in 8 M urea, incubated at room temperature for 30 minutes and then prepared in the same manner as all other samples for analysis using SDS gel electrophoresis (see chapter 2.17).

## 2.13. Preparation of a Sepharose nickel column

Before each purification a 10 ml sepharose (Merck) column was washed, and re-charged with nickel. The following sequential eight wash steps were used: (1) 20 ml water, (2) 20 ml EDTA (0.32 M), (3) 20 ml water, (4) 10 ml NaOH (1 M), (5) 20 ml water, (6) 20 ml $NiSO_4$, (7) 10 ml water, (8) 30 ml IMAC 1 buffer (10 mM imidazole, 50 mM $Na_2HPO_4$, 500 mM NaCl pH7.4). The column was then either stored in 20% ethanol at 4 °C or loaded with soluble protein.

## 2.14. His-tagged protein purification

For the purification of soluble recombinant protein, firstly, cell pellets were re-suspended in IMAC 1 buffer (10 mM imidazole, 50 mM $Na_2HPO_4$, 500 mM NaCl pH7.4) and lysed through sonication (see chapter 2.10)**.** Following lysis and clarification, the soluble protein fraction (supernatant) was transferred on to a nickel column. Once all soluble protein had entered the column the following wash steps were used: 40 ml IMAC 1 buffer (10 mM imidazole, 50 mM $Na_2HPO_4$, 500 mM NaCl pH7.4), 40 ml IMAC 2 buffer (50 mM imidazole, 50 mM $Na_2HPO_4$, 500 mM NaCl pH7.4), 20 ml IMAC 3 buffer (100 mM imidazole, 50 mM $Na_2HPO_4$, 500 mM NaCl pH7.4) and finally 15 ml IMAC 4 elution buffer (500 mM imidazole, 50 mM $Na_2HPO_4$, 500 mM NaCl pH7.4). As the IMAC 4 elution buffer eluted from the column the protein content was monitored by taking 5 µl of an eluting droplet and adding it to 200 µl of Bradford reagent (Bio-Rad). Upon detection of eluting protein, the eluate was collected. The purified protein was then precipitated by the addition of ammonium sulphate directly to the eluate (50% final concentration) and stored at 4 °C for long-term. For subsequent use of purified protein an aliquot (100 – 1000 µl) of precipitated stock was pelleted through centrifugation for 20 minutes at 4 °C and 15,000 rpm (Eppendorf centrifuge 5424R). The pelleted protein was then re-suspended in water and could then be quantified and used in assays.

## 2.15. Bradford assay

A Bradford assay was used to quantify the total soluble protein of clarified lysate containing each of the four TEs. Here, known concentrations of Bovine Serum Albumin (BSA) (25 µg/ml, 50 µg/ml, 75 µg/ml, 100 µg/ml, 125 µg/ml, 250 µg/ml and 500 µg/ml) were prepared in 1 ml. 800 µl of each concentration was incubated with 200 µl of Bradford reagent and absorbance values at 595 nm were obtained. These were used to produce a standard curve (Figure 2.1). A 1 ml dilution of the clarified lysate containing each of the TEs was then used in the same manner to obtain an absorbance ($A_{595}$) value within the range of the BSA calibration curve. This value could then be used to determine the protein concentration within each dilution by using the equation determined for the standard calibration curve.

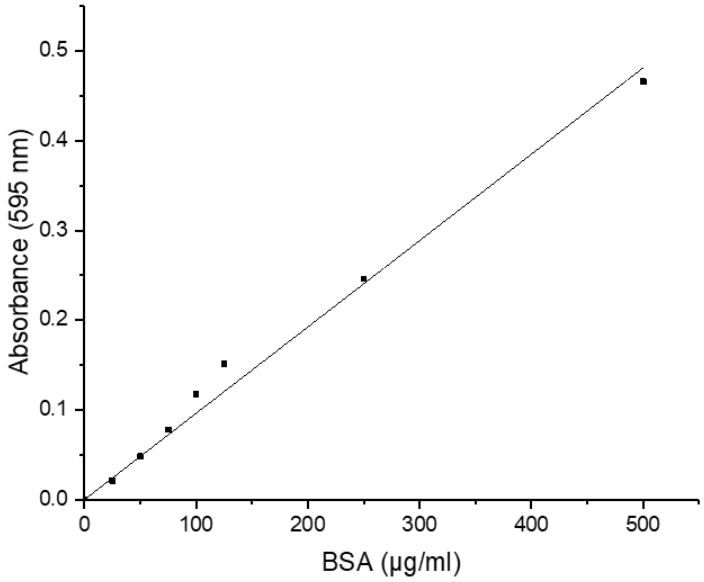



**Figure 2.1. A standard curve of BSA.**

A standard curve of BSA was produced using the absorbance values at 595 nm with a range of standard concentrations of BSA, following incubation with Bradford reagent. y= 0.001x, $R^2$= 0.9898

## 2.16. Quantification of purified protein using $A_{280}$ absorbance

For the determination of purified protein concentrations, the extinction coefficient and molecular weight was firstly calculated using the online analysis tool available at: https://web.expasy.org/protparam/. Then a nanodrop (Thermo scientific, Nanodrop 2000c) was used, to determine protein concentration using absorbance values at 280 nm ($A_{280}$).

## 2.17. SDS-PAGE gels

For molecular weight analysis of denatured protein monomers, from soluble and insoluble protein fractions, samples were analysed through sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Samples were heated to 98 °C for five minutes in 4x Laemmli sample buffer (Bio Rad, UK) containing 2-mercaptoethanol in a 20 µl total volume. Denatured samples were then loaded on to a pre-cast gel (Bio Rad) and run for 35 minutes at 200 volts**.** All SDS gels were then stained with Instant Blue (Expedeon) for one hour and then washed overnight with water**.** For size determination the protein colour standard (NEB) was loaded and run with all samples.

## 2.18. Agarose gels

For DNA analysis agarose gels were made using 1% agarose (Sigma aldrich) dissolved in 1xTBE (Invitrogen) boiled using a microwave. Following the addition of SyBR safe (Invitrogen), agarose gels were left for 30 minutes to set. 6x loading dye (NEB) was added to samples and they were loaded on to the gel. The 2-log ladder (NEB) was run alongside samples for determining the size of linear DNA fragments.

## 2.19. MycA4 assay

One of the putative functions of MycA4 activity was to activate an organic acid with CoA, therefore an end-point assay was monitored in a similar way to AliA using DTNB to monitor CoA concentrations through absorbance at 405 nm. Assays contained: 50 mM TrisHCL pH8, 1.5 mM $MgCl_2$, 2 mM ATP, 0.4 mM CoA, 5 µg purified MycA4, 1 mM shikimate or 3-phosphoshikimate, made up to 500 µl with MiliQ water. This assay was also repeated with the addition of 10 µg MycA2 and 1 mM NADPH. All assays were run for 30 minutes at 25 °C, in triplicate, in a 96-well plate (star lab). For absorbance readings 50 µl aliquots were quenched with TFA (0.2% final concentration) and then 50 µl of quenched reaction was added to 150 µl of 10 mM DTNB for determination of absorbance values. Alternative assay conditions were also assessed in an attempt to find active conditions. These included: addition of 1 mM PEP, addition of 1.5 mM KCl, altering the pH to 7.5, using 50 mM HEPES buffer, increasing MycA4 to 50 µg and addition of 0.5 mM $FMNH_2$/ 0.5 mM $FADH_2$.

## 2.20. MycA2 reductase assay

For determining MycA2 reductase activity with CH-1-eneC CoA, ChCoADH was used to produce CH-1-eneC CoA from commercial CHC CoA (see chapter 2.21). For analysis of activity with 3-HCH-1-eneC CoA, sequential activity of AliA and ChCoADH was required for the production of 3-HCH-1-eneC CoA from 3-HCHC. Prior to assessment of MycA2 activity, the purified enzymes required for CH-1-eneC CoA and 3-HCH-1-eneC CoA production were removed using a 10 kDa Vivaspin 500 (GE Healthcare). 5 µg MycA2 was then added to 100 µl of CH-1-eneC or 3-HCH-1-eneC CoA, 0.1-0.6 mM NADPH and the reaction mixture was brought to a final volume of 150 µl with MiliQ water. Activity was determined though the oxidation of NADPH to NADP, monitored at 340 nm using a ClARIOstar plate-reader (BMG labtech). For the conversion of absorbance values into concentrations of NADPH a calibration curve was produced, using the determined 340 nm absorbance values of NADPH at a range of concentrations (Figure 2.2). All assays were run in

triplicate for 30 minutes at 25 °C. HPLC analysis was also used for determining activity. Here, immediately after (t = 0) and 30 minutes (t = 30) after the addition of MycA2 a 75 µl aliquot of the assay was quenched with 75 µl 0.4% TFA (0.2% final TFA) and run on the HPLC using the program in 2.30.

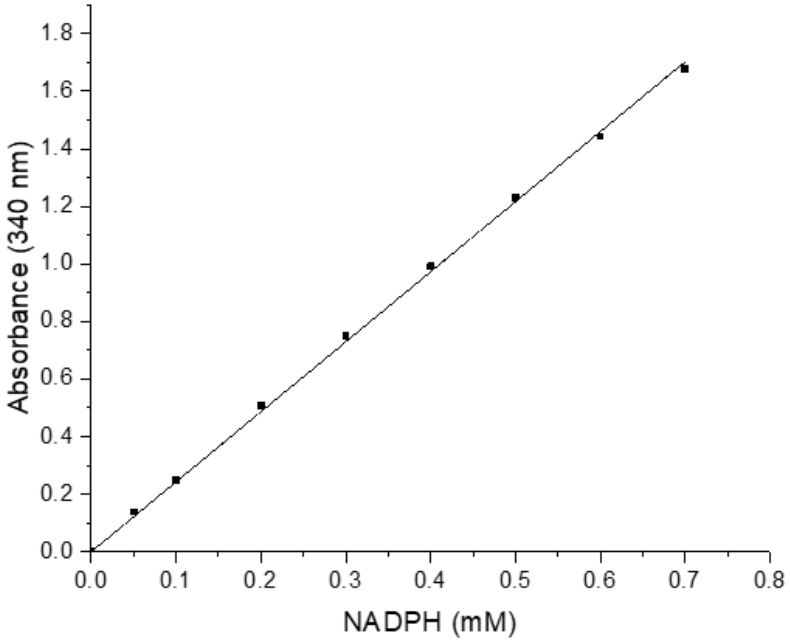



**Figure 2.2. An NADPH standard curve.**

A standard curve produced using the absorbance values at 340 nm with a range of concentrations of NADPH. The oxidised product, $NADP^+$, does not absorb light at 340 nm. Therefore, the oxidation of NADPH to $NADP^+$, monitored at 340 nm, was used to determine reductase activity of MycA2. $y = 2.4323x$, $R^2 = 0.9989$

## 2.21. BadJ and ChCoADH dehydrogenase assay

Dehydrogenase activity (ChCoADH or BadJ) was measured by following the CHC CoA dependent reduction of the artificial electron acceptor ferrocenium hexafluorophosphate ($Fc^+PF_6^-$) to ferrocene (300 nm) for 30 minutes at 25 °C using a CLARIOstar plate-reader (BMG labtech). The reaction mixture (250 µl) contained 10 µg purified enzyme, 50 mM TrisHCL pH 8, 0.5 mM CHC CoA, 0.2-1 mM $Fc^+PF_6^-$ and 1.5 mM $MgCl_2$. Concentration of $Fc^+PF_6^-$ was

determined using a standard curve of absorbance values of known concentrations of $Fc^{+}PF_6^{-}$ (Figure 2.3).

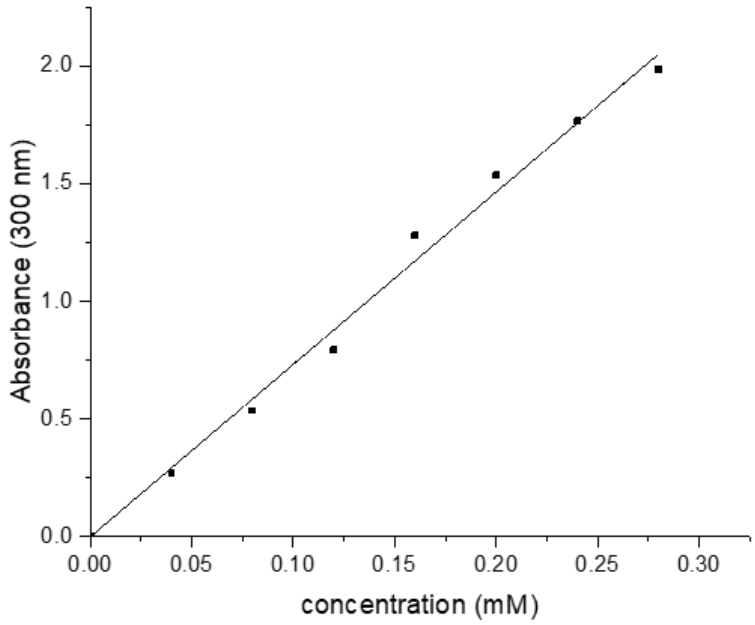



**Figure 2.3. An Fc+PF6- standard curve.**

A standard curve produced using the absorbance values at 300 nm with a range of concentrations of the artificial electron acceptor, ferrocenium hexafluorophosphate ($Fc^{+}PF_6^{-}$). The reduced product of $Fc^{+}PF_6^{-}$, ferrocene, has no absorbance at 300 nm. Therefore, the reduction of $Fc^{+}PF_6^{-}$ was monitored to indirectly determine the oxidation activity of ChCoADH. $y = 7.3239x$, $R^2 = 0.9916$.

Ferrocenium salts are one-electron oxidising agents and so for every mole of oxidised CoA ester, two moles of $Fc^{+}PF_6^{-}$ are reduced (Connelly and Geiger, 1996). Consumption of CHC CoA was also used to determine ChCoADH activity, assessed through HPLC analysis, indicated by a drop in peak area over the 30 minute time course. As there is no commercial standard of CH-1-eneC CoA, HPLC was also used to identify its production. Here, a novel peak that formed following dehydrogenase activity was purified using HPLC fraction collection (chapter 2.30) and confirmed to correspond to be CH-1-eneC CoA through High-resolution mass spectrometry (HRMS).

## 2.22. BadK hydratase assay

For the assessment of CH-1-eneC CoA-dependent hydratase activity, CH-1-eneC CoA was firstly synthesised through dehydrogenase activity with CHC CoA (see section 2.21). 20 µg of purified BadK was directly added to 1 ml of reaction mixture containing CH-1-eneC CoA and incubated for 30 minutes at 25 °C. A 75 µl aliquot was taken immediately after (t = 0) and 30 minutes (t= 30) after the addition of BadK and quenched with 75 µl 0.4% TFA (0.2% final concentration). Consumption of CH-1-eneC CoA was then used to determine BadK activity, assessed through HPLC analysis, indicated by a drop in peak area over the 30 minute time course. As there is no commercial standard of 2-HCHC CoA, HPLC was also used to identify its production. Here, a novel peak that formed following BadK activity was purified using HPLC fraction collection (section 2.30) and confirmed to correspond to be 2-HCHC CoA through HRMS.

## 2.23. BadH dehydrogenase assay

For the assessment of BadH dehydrogenase activity, 2-HCHC CoA firstly had to be synthesised by sequential ChCoADH and BadK activity (see section 2.23), however, 50 mM TrisHCL pH 9 was used in these steps as a previous investigation of BadH from *R. palustris* found a 54% decrease in activity at pH 8 and only a 14% decrease at pH 9, compared to activity at the optimal pH 9.5 (Pelletier and Harwood, 2000). Following 2-HCHC CoA synthesis, all enzymes were removed using a 10 kDa Vivaspin ® 500 centrifugal concentrator (Sartorius). BadH dehydrogenase activity was then monitored in a reaction mixture (total 275 µl) containing 200 µl 2-HCHC CoA reaction mixture, 1 mM $NAD^+$, 1 µg purified BadH and 10 mM hydrazine hydrate (or without hydrazine hydrate). Activity was determined spectrophotometrically by monitoring the reduction of $NAD^+$ to NADH at 340 nm. A standard curve using the 340 nm absorbance values for NADH concentrations (0.01–0.2 mM) was used for conversion of $A_{340}$ values into NADH concentrations (Figure 2.4).

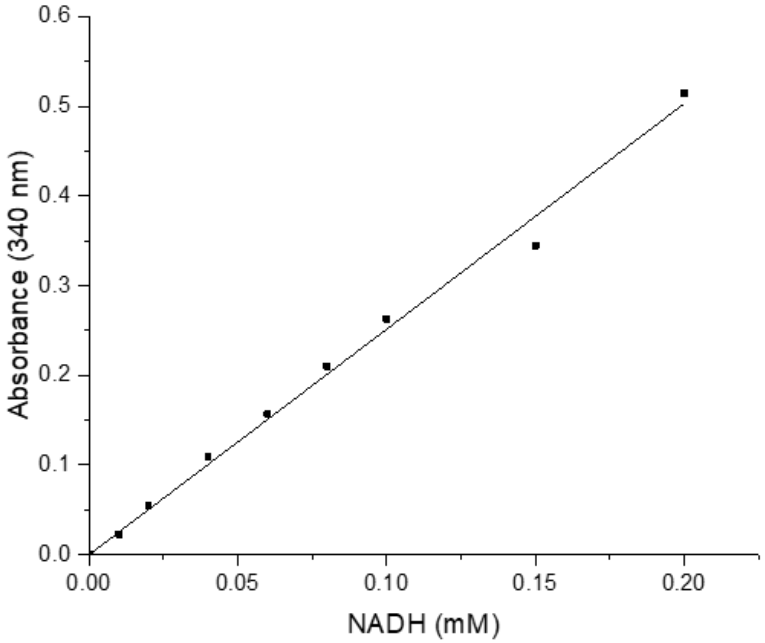



**Figure 2.4. An NADH standard curve.**

A standard curve produced using absorbance values determined at 340 nm with a range of concentrations of NADH. The oxidised product, $NAD^+$, does not absorb light at 340 nm. Therefore, the reduction of $NAD^+$ to NADH, monitored at 340 nm, was used to determine oxidase activity of BadH. $y = 2.5149x$, $R^2 = 0.9919$.

## 2.24. Screening TE activity with CHC CoA, CH-1-eneC CoA and 2-HCHC CoA

Prior to TE analysis a reaction mixture containing CHC CoA, CH-1-eneC CoA and 2-HCHC CoA was produced using sequential ChCoADH and BadK with CHC CoA (see section 2.21-2.22). 0.5 mM $Fc^+PF_6^-$ was used to prevent complete consumption of the 0.5 mM CHC CoA to CH-1-eneC CoA (the reduction of 1 mole of CHC CoA requires 2 moles of $Fc^+PF_6^-$). Following the production of the substrate CoA esters, a 10 kDa Vivaspin was used to remove ChCoADH and BadK, prior to analysis of TE activity. The standard assay (200 µl total volume) contained 200 µg of clarified lysate (made to 20 µl with water) containing recombinantly expressed FcbC, PA2801, RpaL, YbdB, or a

negative control (cell expressing unmodified pET28a) and 180 µl of premade substrate containing reaction mixture (containing the CoA ester substrates). A 75 µl aliquot was taken immediately after the addition of clarified lysate (t= 0) and quenched with 75 µl TFA (0.2% final). Following incubation of the assay at 25 °C for 30 minutes a second 75 µl aliquot was quenched with 75 µl TFA (0.2% final). HPLC analysis was then used to determine activity through a drop in the peak area of CHC CoA, CH-1-eneC CoA and 2-HCHC CoA.

### 2.25. Screening TE activity with commercial CoA esters

TE activity was determined by using the 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB) colorimetric assay (Ellman, 1959). In the presence of free thiol groups DTNB is cleaved and the 2-nitro-5-thiobenzoate (TNB) released can be detected at 405 nm. For one mole of thiol, one mole of TNB is produced, allowing CoA concentration to be directly inferred from TNB concentration. Assays for all TEs were run at 25 °C and contained: 20 µg clarified lysate, 100 mM Triethanolanime (TEA) buffer (pH 8), 0.2 mM CoA ester, 0.4 mM DTNB and ultrapure Mili-Q water (MilliPore Q-POD) to a total volume of 250 µl. Reactions were started by the addition of the CoA ester and run for 30 minutes. Control reactions using lysate from cultures containing unmodified pET29a, expressed in the same conditions as the recombinant TEs, was used to account for any background hydrolysis of CoA esters. CoA concentrations determined in control reactions (containing no enzyme) were subtracted from TE reactions for determination of TE-hydrolysed CoA. All assays were performed in triplicate in a 96-well plate and monitored using a TECAN Safire$^{2}$ $^{TM}$ (Invitrogen $^{TM}$) microplate reader. To calculate the concentration of CoA produced, a standard curve was produced using known concentrations of CoA. Each standard concentration was measured at 405 nm and contained 0.4 mM DTNB, 100 mM of TEA and water (250 µl final volume) (Figure 2.5).

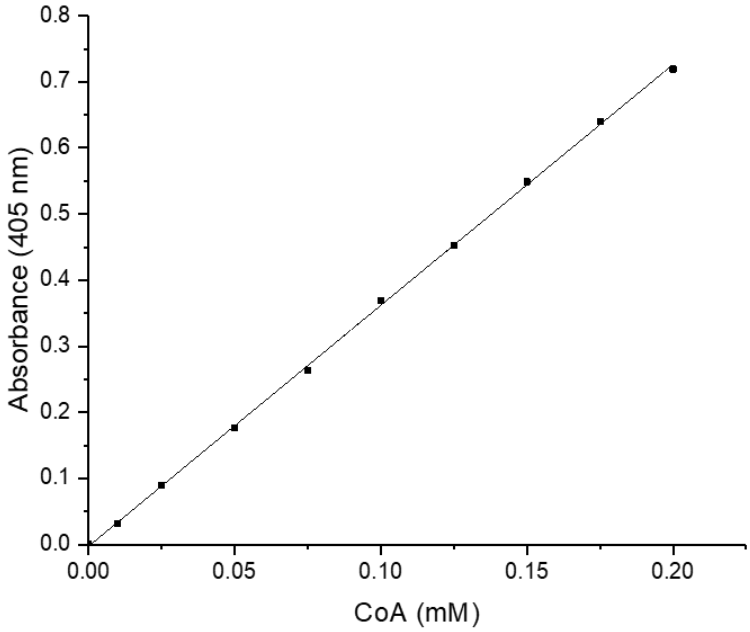



**Figure 2.5. A DTNB standard curve for assessing TE activity.**

A standard curve produced using absorbance values determined at 405 nm with a range of CoA concentrations with 0.4 mM DTNB. In the presence of a thiol functional group (CoA) DTNB is cleaved, yielding TNB, a coloured species that can be measured at 405 nm. Therefore, CoA concentration and thus TE activity, could be determined indirectly by the production of TNB. $y = 3.6277x$, $R^2 = 0.9995$

## 2.26. Determining the CoA ligase activity of AliA.

For the determination of CoA ligase activity an endpoint assay was used. Each assay was run at 25 °C and contained 50 mM TrisHCL pH8, 1.5 mM $MgCl_2$, 0.4 mM CoA, 2 mM ATP, 5 µg purified AliA, 1 mM organic acid (substrate) and made to 150 µl with MiliQ water**.** Immediately after the addition of substrate a 50 µl aliquot of the assay was quenched with 50 µl 0.4% TFA (0.2% final TFA). After 10 minutes incubating at 25 °C, a second 50 µl aliquot of the assay was quenched with TFA (0.2% final concentration). Then the two quenched aliquots were either analysed through HPLC (see section 2.30) or assessed using the DTNB assay. To determine the concentration of CoA using the DTNB assay, 50 µl of each quenched time point was added to 150 µl of 10

mM DTNB (7.5 mM final). A CLARIOstar plate-reader (BMG labtech) was used to determine the absorbance at 405 nm of each sample. All assays were performed in triplicate in a 96-well plate. Absorbance values were then converted into a CoA concentration using a standard curve of known CoA concentrations (Figure 2.6).

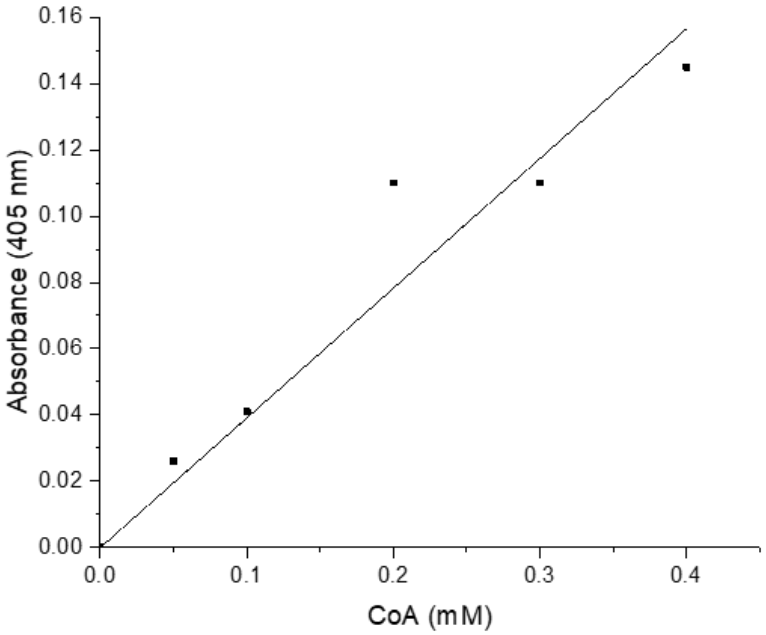



**Figure 2.6. A DTNB standard curve for assessing CoA ligase activity.**

A standard curve produced using absorbance values determined at 405 nm with a range of CoA concentrations with 7.5 mM DTNB. In the presence of a thiol functional group (CoA) DTNB is cleaved, yielding TNB, a coloured species that can be measured at 405 nm. Therefore, CoA concentration and thus CoA ligase activity, could be determined by a decrease in TNB concentration. $y = 0.3914x$, $R^2=0.9248$

## 2.27. Determination of steady state kinetic constants of RpaL

The DTNB colorimetric assay was used to determine the production of CoA from the CoA thioesters. The same conditions as previously mentioned were used (see section 2.25) with the exception of 2 μg (0.73 μM) of purified RpaL rather than clarified lysate being used. Steady state kinetic measurements for RpaL were determined by measuring the initial rates ($v_0$) of CoA ester

hydrolysis as a function of time. Following the addition of a CoA ester the first time points showing an increase in CoA concentration were used to determine the initial rate ($v_0$). All reactions were run for 30 minutes, however, the initial rates were determined from time points obtained within the first 5 minutes. Varying concentrations of each of the six CoA esters (*i*Val CoA, Bz CoA, *i*But CoA, HMG CoA, DLBH CoA and *n*HD CoA) were assayed, in triplicate, to create substrate [S] saturation curves. These curves were then fitted to the Michaelis-Menten equation $v_0 = v_{max}[S]/ ([S] + k_m)$, using Origin 2017 software, allowing determination of maximal velocity ($v_{max}$) and Michaelis-Menten constant ($k_m$). Turnover ($k_{cat}$) could then be obtained using the equation $k_{cat} = v_{max}/ [E]$.

### 2.28. CoA ester and organic acid modification assays

Organic acid modification assays were set up in three stages. In step one (activation), organic acids were CoA activated according to chapter 2.26, with the addition of 1 mM $Fc^+PF_6^-$ and a total volume of 1 ml**.** In the second step (modification), 20 µg of each purified modification enzyme (ChCoADH, BadK) was added and incubated at 25 °C for 30 minutes. In reactions requiring the HPLC analysis of resultant CoA esters (Figures: 6.7, 6.8, 6.10, 6.12, 6.14, 6.16 and 6.18), following ChCoaDH/BadK activity, a 50 µl aliquot of the assay was quenched with TFA (0.2% final concentration) for HPLC analysis. For the aforementioned figures the third step was not preformed (Figure 6.7. etc.). Finally, in the third step (termination), 20 µg of purified RpaL was added to the reaction and incubated for 30 minutes at 25 °C. The resultant modified organic acids were determined through HRMS.

## 2.29. Size exclusion chromatography

Size exclusion chromatography (SEC) was used to determine the biological configuration of purified RpaL. An HPLC 1660 was used with an Agilent Zorbax 250 column. Phosphate buffered solution (PBS) degassed with helium was used as the mobile phase and run at a flow rate of 0.4 ml/min and a run time of 10 minutes. Purified RpaL was buffer exchanged after purification into PBS. An injection volume of 25 µl was used.

## 2.30. HPLC analysis and fraction collection

For the analysis of CoA esters a 25 minute method was created on an HPLC (Thermo scientific Dionex Ultimate 3000) using an ACE Acclaim$^{TM}$ 120 C18 5µM (120Å 4.6 X 150 mm) reverse phase column. The mobile phases used were: A- 50 mM sodium acetate pH 5 and B- acetonitrile. The HPLC sequence used for analysis of all CoA esters was as follows: 0-1 minute 2% mobile phase B, 1-10 minutes 2-11% B, 10-15 minutes 11-20% B, 15-16 minutes 20-30% B, 16-21 minutes 30% B, 21-22 minutes 30-2% B, 22-25 minutes 2% B. The flow rate was set to 1 ml/min throughout the run. Samples were filtered using a Whatman$^{TM}$ 0.2 µm filter unit (GE Healthcare) prior to being loaded on to the HPLC. Retention time and collection times for each compound analysed with HPLC are shown (Table 2.6).

**Table 2.6. The retention times and collection periods of all compounds analysed and collected through HPLC.**

| Compound | Retention time (minutes) | Collection period (minutes) |
|---|---|---|
| CoA | 7.5 | - |
| CHC CoA | 18.5 | - |
| CH-1-eneC CoA | 18.2 | 17.6 - 18.4 |
| 2-HCHC CoA | 14.7 | 14.2 - 15.1 |
| Benzoyl CoA | 17.1 | - |
| CH-3-eneC CoA | 18.1 | 18 - 18.15 |
| CH-1,3-eneC CoA (c1) | 17.1 | 16.975 - 17.250 |
| CH-1,3-eneC CoA (c2) | 17.5 | 17.375 - 17.6 |
| 3-HCHC CoA | 13.3, 13.4 | 13.1 - 13.6 |
| 3-HCH-1-eneC CoA | 12.9 | 12.7 - 13.1 |
| 4-HCHC CoA | 13.3 | 13.15 - 13.6 |
| 4-HCH-1-eneC CoA | 12.4 | 12.25 - 12.6 |
| 4-oxoCHC CoA | 12.5 | - |
| ATP | 1.8 | - |
| AMP | 3.5 | - |

## 2.31. CoA and CHC CoA HPLC calibration curves

Calibration curves for CoA and CHC CoA were determined using the peak area determined for a range of standard concentrations of each compound (Figure 2.7 & 2.8).

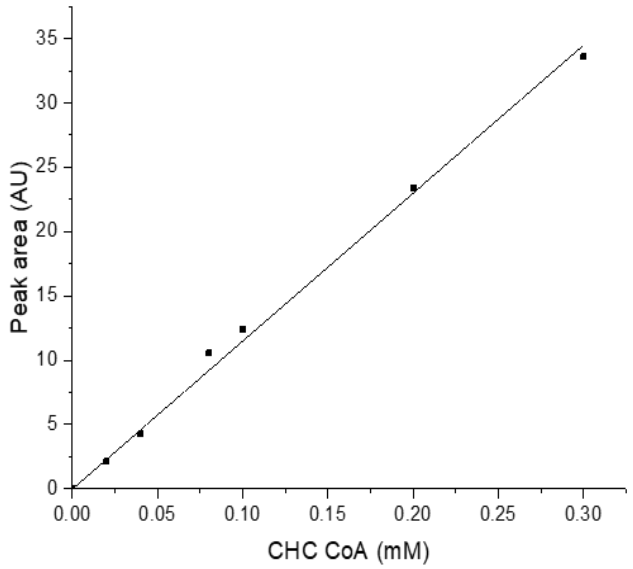



**Figure 2.7 . A CHC CoA standard curve- HPLC.**

A calibration curve determined using HPLC analysis to determine the peak area of a range of concentrations of commercial CHC CoA. y = 112.17x + 0.5609 $R^2$ = 0.9958

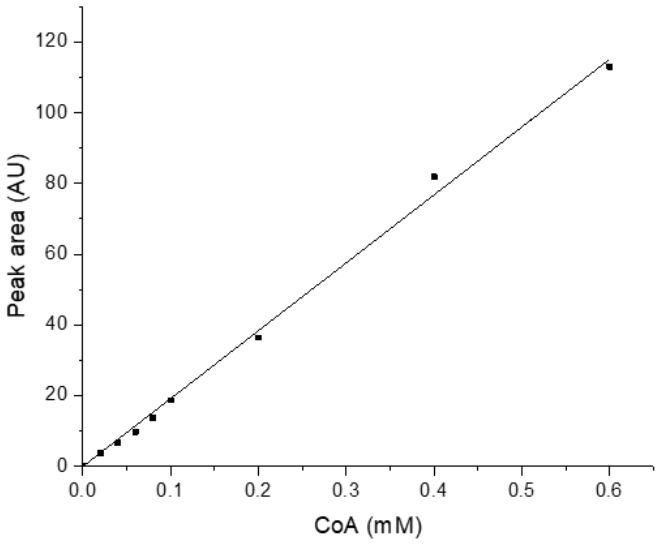



**Figure 2.8. A CoA standard curve- HPLC.**

A calibration curve determined using HPLC analysis to determine the peak area of a range of concentrations of commercial CoA. y = 194.63x - 1.0272 $R^2$ = 0.9968

## 2.32. Densitometry

For determining protein as a percentage of total protein on an SDS gel, densitometry was used with ImageJ software available at: https://imagej.nih.gov/ij/download.html

## 2.33. Statistical analysis

For analysis of significant variation between two groups a t-test was performed using GraphPad Prism 8 software.

## 2.34. Software for graphs

For presentation of data (except tables) origin ® 2017 and GraphPad Prism 8 software was used.

## 2.35. HRMS preparation and analysis

Samples for HRMS analysis were purified desalted using a ZipTip ® (Merck), following the user guide available at: http://www.merckmillipore.com/GB/en/product/ZipTip-Pipette-Tips,M M_NF-C5737#documentation. Prepared samples were then submitted to the UCL Chemistry department for HRMS analysis.

## 2.36. Bioinformatics

The genome of *S. rishiriensis* was purified using a GeneJET genomic DNA Purification Kit (Thermo Scientific) and sequenced by Eurofins Scientific (formerly GATC). For analysis and mining of the purified genome, all contiguous DNA sequences received from GATC were compiled into a custom BLAST database using Geneious prime software. For all other bioinformatic work (i.e. gene design, primer design) Snapgene software was used.

# 3. Chapter 3. Characterisation of the CHC CoA operon from *Streptomyces rishiriensis*

## 3.1. The CHC CoA operon from *Streptomyces rishiriensis*

The work described in chapter 3 focuses on identifying and constructing an enzymatic link between central metabolism and the previously developed, pathway to 6-ACA at UCL (Jackson, 2017). For this route, enzymes encoded by four genes (*mycA1*, *mycA2*, *mycA3* and *mycA4)* from the CHC CoA operon within *S. rishiriensis* were identified. In its native biological context, this operon is putatively required for the conversion of either shikimate or 3-phosphoshikimate into CHC CoA, which is then attached to polyketides such as: trienomycin A, mycotrienin II and mycotrienin I (Cropp, Wilson and Reynolds, 2000). Alternatively, the synthetic pathway proposed in this thesis will redirect CHC CoA for further enzymatic modification by recombinant enzymes described in chapters 4 and 5, ultimately producing cyclohexanol (Figure 3.1).

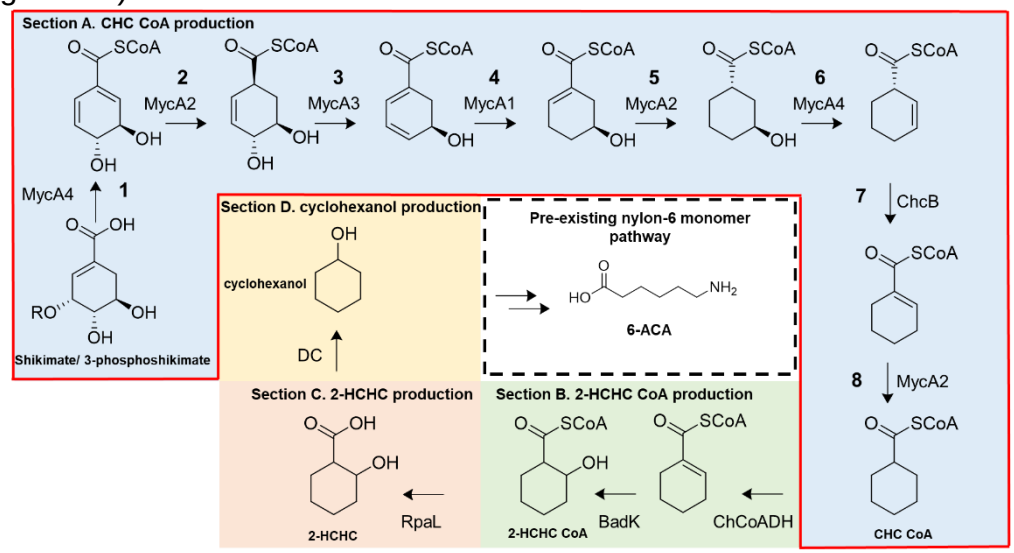



**Figure 3.1. The biochemical CHC CoA pathway.**

The section of the cyclohexanol pathway to be assessed in chapter 3 (outlined in red). In the proposed mechanism, shikimate or 3-phosphoshikimate is converted into CHC CoA (labelled) by MycA1-4. MycA1 is a proposed dienoyl-CoA reductase (step 4), MycA2 is a 1-cyclohexenylcarbonyl CoA reductase (steps 2, 5 & 8), MycA3 is a proposed acyl-CoA dehydrogenase (step 3), and MycA4 is proposed to function as a CoA ligase and phosphatase/hydratase (step 1 & 6).

At the time of undertaking this work, the *S. rishiriensis* genome had not been sequenced and so the CHC CoA operon was only proposed to be present. This likely presence was based on the understanding that a CHC CoA operon is essential for the production of the CHC side chains in polyketides and cyclohexyl fatty acids (Cropp, Wilson and Reynolds, 2000), and that *S. rishiriensis* was a known producer of CHC-containing polyketides, such as trienomycin A (Sugita *et al.*, 1982). Unfortunately, understanding of the enzymatic mechanism in the CHC CoA pathway is limited. A homologous CHC CoA operon from *S. collinus* had previously been cloned into a pAC12 vector and constitutively expressed within *Streptomyces avermitilis*, for assessment of the associated polyketide products. However, only one of the four enzymes encoded within this CHC CoA operon, ChcA (MycA2 homolog), has been recombinantly expressed in *E. coli* and had its activity assessed *(*Wang *et al*, 1996). The activity of the remaining three enzymes is yet to be assessed.

The initial aim of the work presented here was to confirm the presence of a CHC CoA operon within the genome of *S. rishiriensis*. Following its identification, the entire CHC CoA operon was cloned and recombinantly expressed within *E. coli*, to determine the solubility of all four enzymes when expressed in *E. coli*. As all four of these enzymes are required for the biosynthetic cyclohexanol pathway, their soluble recombinant expression will be essential to facilitate viable activity. Following determination of the conditions required for soluble expression of the CHC CoA operon genes, each gene was to be separately cloned, expressed and the resultant enzyme purified for activity analysis. However, a caveat to analysing enzyme activity was that none of the CoA ester intermediates of the CHC CoA pathway were commercially available. Shikimate and 3-phosphoshikimate, the proposed substrates of the first enzymatic step (MycA4), were the only commercially available substrates. This inherent limitation motivated later research (see chapter 6) to develop a biological method through which to produce various CoA ester intermediates in this pathway, facilitating activity analysis of the recombinantly expressed CHC CoA pathway enzymes described here.

## 3.2. Identification of CHC CoA operon in *Streptomyces rishiriensis*

*S. rishiriensis* is an organism that has been previously described to produce polyketides such as mycotrienin II, which contains a CHC moiety (Figure 3.2), and so it was proposed that its genome is likely to contain a CHC CoA operon.

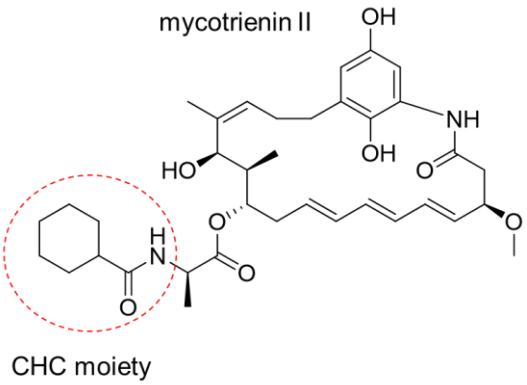



**Figure 3.2. The chemical structure of mycrotrienin II.**

The CHC side chain is highlighted which is produced by the CHC CoA pathway (red dashed circle).

Therefore, the S. rishiriensis (DSM 40489) genome was first isolated (Figure 3.3), sequenced and assessed for the presence of a putative CHC CoA operon.

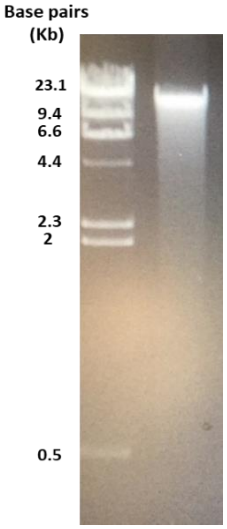



**Figure 3.3. An agarose gel of the purified *S. rishiriensis* genome.**

The sequencing output gave 2,970 contigs ranging from 120 bp to 382,592 bp. Contigs that were >1000 nucleotides (240 contigs) were then compiled into a BLAST database, which was used to search for the genes within the CHC CoA cluster. Genes from CHC CoA clusters previously identified in two other organisms, *S. collinus* and *S. flaveolus* were used for this process. The results of this BLAST analysis lead to the identification of four adjacent unidirectional open reading frames (ORFs) on a single contig (contig 39) (Figure 3.4). These four genes formed a 7193 bp operon, which was proposed to be the CHC CoA operon. For the identification of ribosome binding sites the 3' region of the 16s rRNA sequence of *S. rishiriensis* was used. The available *S. coelicolor* 16s rRNA sequence (Gene ID: 1096815) was used to identify a homolog in *S. rishiriensis* (see appendix 8.38). The 3' region of the 16s rRNA from *S. rishiriensis* was found to be 3'-UCUUUCCUCCA-5', which was then used to identify ribosome binding sites with complementary sequences. Putative ribosome binding sites were found upstream of the identified start codon of each gene (Table 3.1). The three genes: *mycA4*, *mycA3* and *mycA2* were found to have ATG (Met) start codons, conversely, *mycA1* had a GTG start codon. The RBS and start codon of *mycA2* was found within the ORF of the upstream gene, *mycA3*. An intragenic start codon was not found between *mycA4* and *mycA3*, which was found to have a 50 nucleotide non-coding interspace between the stop codon of *mycA4* and the start codon of *mycA3* (see appendix 8.1 for operon nucleotide sequence and appendix 8.2-9 for the nucleotide and polypeptides sequences of *mycA1-4*).

**Table 3.1. The 5' nucleotide sequence of the CHC CoA operon genes in *S. rishiriensis.***

Putative ribosome binding sites have been underlined and start codons are shown in bold font.

| gene | 5' nucleotide sequence |
|---|---|
| *mycA4* | 5'-AGGGAGCAGTAGGACCC**ATG** |
| *mycA3* | 5'-AGAGGTGGCAC**ATG** |
| *mycA2* | 5'-GAGGGAACGAACCG**ATG** |
| *mycA1* | 5' -GAGGGACGGCAG**GTG** |

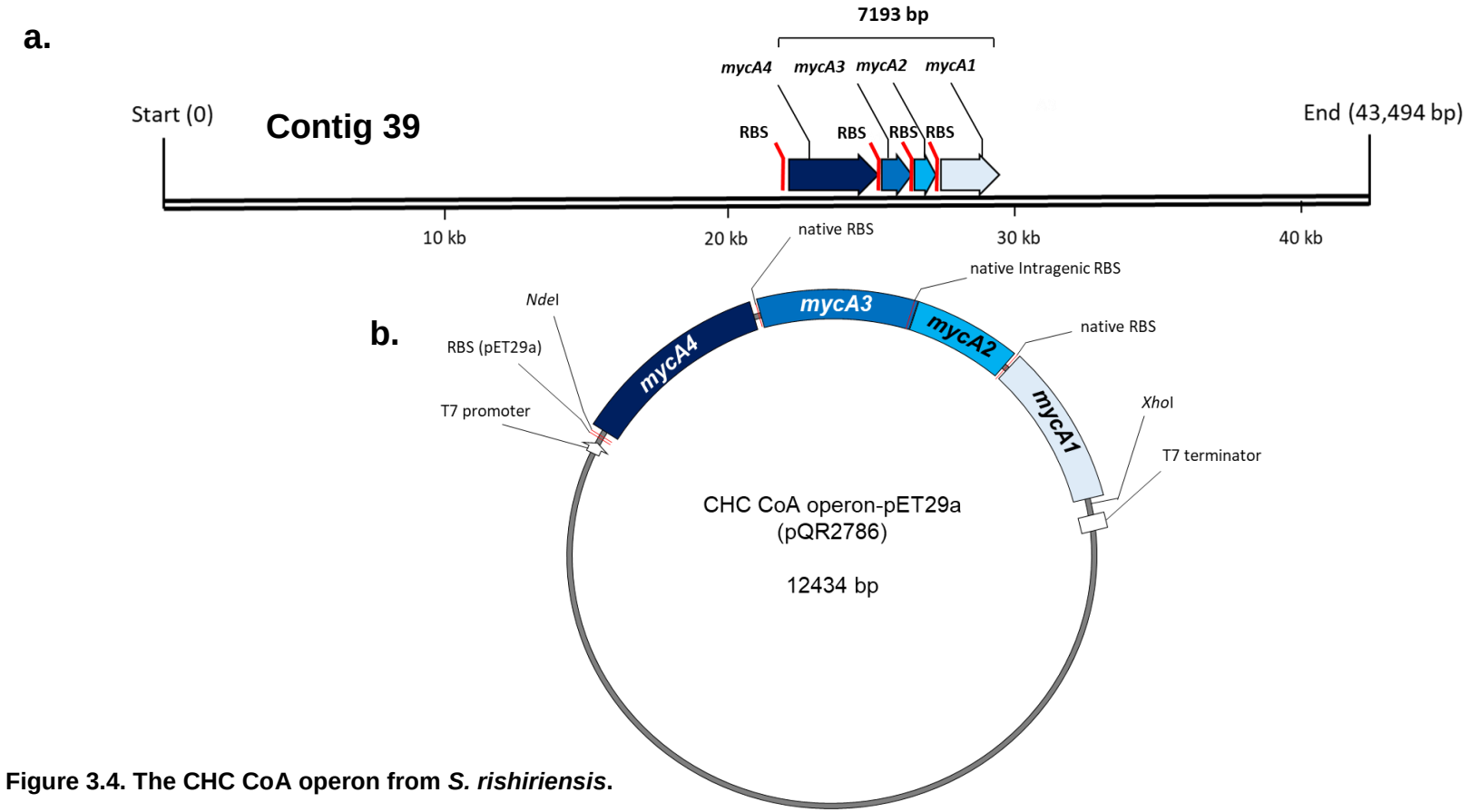



**Figure 3.4. The CHC CoA operon from *S. rishiriensis*.**

(a) The four genes (*mycA1, mycA2, mycA3*, and *mycA4*) encoded by the CHC CoA operon (contig 39) within *S. rishiriensis*. The four genes were found to be coded in same direction and each contained an upstream RBS. (b) The recombinant CHC CoA operon-pET29a plasmid (pQR2786), containing all four of the operon genes and their native RBS, except for *myca4,* which was translated from the pET29a RBS.

The penultimate step is an isomerisation previously shown to be catalysed by an enoyl CoA isomerase (ChcB) in S. collinus (Patton, Cropp and Reynolds, 2000). The gene encoding this enzyme, *chcB*, was found at another locus, away from the CHC CoA operon. A BLAST search using *chcB* sequence from S. collinus with the S. rishiriensis contigs revealed a 792 bp homologue with 91.5 % similarity (see appendix 8.37). This homolog was located on a separate contig (contig 11) to the CHC CoA operon, aligning with the spatial separation observed in S. collinus.

### 3.3. Cloning and expression of the CHC CoA operon

Using the identified sequence for the CHC CoA operon in *S. rishiriensis*, primers were designed to isolate the operon from the genome as a single DNA fragment. Additionally, these primers were designed to add flanking endonuclease restriction sites, required for cloning the cluster into pET29a. These restriction sites were an N-terminal *Nde*I and a C-terminal *Xho*I. Following PCR amplification (Figure 3.5a) the fragment was digested with *Nde*I and *Xho*I and ligated into a pET29a vector. After ligation, the recombinant plasmid was used to transform *E. coli* TOP10 cells. The plasmid of a successful colony was then purified and analysed by digestion with *Nde*I and *Xho*I. The results showed two linear DNA fragments: 5.2 kb and 7.2 kb (Figure 3.5b), corresponding to the linear sizes of pET29a and the CHC CoA operon, respectively, indicative of successful ligation. This result was further validated through DNA sequencing, confirming the CHC CoA operon had correctly ligated. This recombinant plasmid, pQR2786, was then used to transform *E. coli* BL21 (DE3) for heterologous expression of the operon.

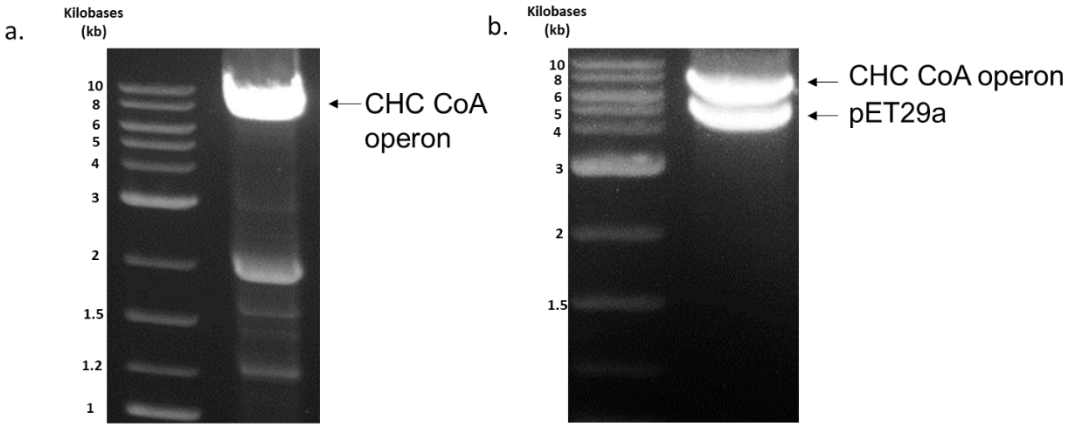



**Figure 3.5. Cloning the CHC CoA operon into pET29a (pQR2786).**

An agarose gel showing the cloning of *S. rishiriensis* cluster into pET29a. (a) The 7.2 kb PCR product of the CHC CoA operon (*S. rishiriensis*), amplified with flanking *Nde*I and *Xho*I restriction sites (black arrow). (b) An analytical *Nde*I and *Xho*I digest of a successful recombinant plasmid (pQR2786) showing linear pET29a (~5.2 kb) and CHC CoA operon (~7.2 kb) (black arrows).

The CHC CoA operon genes (*mycA4, mycA3, mycA2 and mycA1*) were then assessed for expression from pQR2786 within *E. coli* BL21 (DE3). The clarified lysate was analysed for soluble expression of MycA1, MycA2, MycA3, and MycA4. The results showed putative expression of MycA4, MycA3 and MycA1, however, no clear protein bands were found with a molecular weight corresponding MycA2 (Figure 3.6- lane 2). Furthermore, a protein of similar molecular weight to MycA3 was found in the lysate of *E. coli* BL21 (DE3) making it unclear whether MycA3 had expressed (Figure 3.6). To achieve soluble expression of all four enzymes encoded by the CHC CoA operon the Takara chaperone plasmid set was used (Table 2.5.).

Each of the five chaperone plasmids (pG-Tf2, pTf16, pGro7, pKJE7, pG-KJE8) were used with the recombinant pQR2786 plasmid to co-transform *E. coli* BL21 (DE3) cells. The soluble protein content of each of these five strains was then analysed following IPTG induced expression. The strain containing pQR2786 + pG-Tf2 showed putative expression of MycA3 and a low level of MycA4 (Figure 3.6- lane 3). Co-expression of chaperones from the pTf16

plasmid produced soluble MycA4 (6% of total soluble protein) and putative expression of MycA3, however, no expression of soluble MycA1 or MycA2 was observed (Figure 3.6- lane 4 and Table 3.3). Co-expression with chaperones encoding by pG-KJE8 showed soluble expression of MycA4 (4% of total soluble protein) and putative expression of soluble MycA3, yet no clear soluble expression of MycA1 and MycA2 (Figure 3.6- lane 7 and Table 3.3)

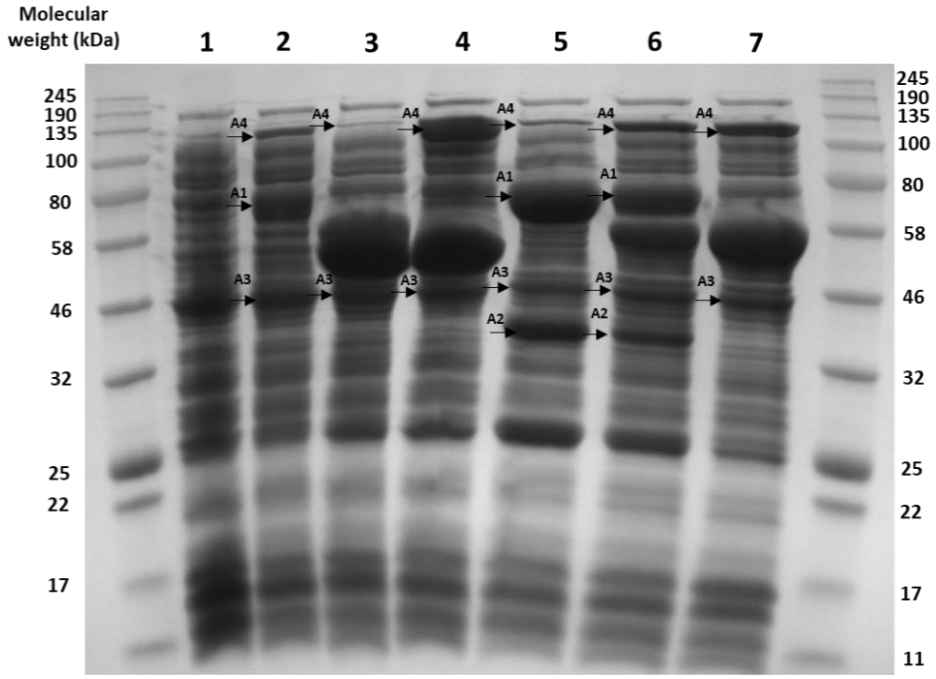



**Figure 3.6. Analysis of soluble MycA1-4 (pQR2786) expression.**

An SDS gel showing the soluble protein content, following 72 hours of 0.1 mM IPTG induced expression at 16 °C, of *E.coli* BL21 (DE3) cells transformed with: (1) pET29a (negative control), (2) pQR2786, (3) pQR2786 + pG-Tf2, (4) pQR2786 + pTf16, (5) pQR2786 + pGro7, (6) pQR2786 + pKJE7, (7) pQR2786 + pG-KJE8. The expected molecular weight of each enzyme was: MycA4 (A4) – 111.3 kDa, MycA3 (A3) - 45.5 kDa, MycA2 (A2) - 32.0 kDa and MycA1 (A1) - 75.2 kDa. 200 µg of soluble protein was prepared and loaded on the SDS gel following the method described previously (chapter 2.17). Cultures were grown in 300 ml of TB shaking at 170 rpm. All cultures were lysed using sonication (see chapter 2.11).

Conversely, protein bands putatively corresponding to all four CHC CoA operon enzymes (MycA1-4) were identified when co-expressed with the chaperones encoded within pGro7 and pKJE7 (Figure 3.6- lane 5 and 6,

respectively). Co-expression with pKJE7 resulted in a ~110 kDa protein band that was not found in the negative control, did not correspond to any of the chaperone proteins, and so was concluded to be MycA4 (2.8% of total soluble protein) (Table 3.3). A second protein band between ~70-75 kDa was observed that was proposed to be a mixture of dnaK (70 kDa) and MycA1 (75.2 kDa). It should be mentioned briefly, in single gene expression experiments shown later in this investigation, MycA2 (32 kDa) was found to migrate a comparable distance through an SDS gel as a 43 kDa protein marker (Figure 3.16). Therefore, a protein band observed at ~40 kDa, was thought to be a combination of dnaJ (40 kDa) and MycA2 (32 kDa). Finally, a ~45 kDa band was thought to putatively represent MycA3 (45.5 kDa), however, as mentioned previously, a protein of similar molecular weight is found in the negative control.

Co-expression with pGro7 produced of a large soluble protein band spanning ~60-75 kDa, which was thought to be a mixture of MycA1 (73 kDa) and groEL (60 kDa). A faint protein band with the molecular weight of ~110 kDa was observed that did not correspond to any inherent chaperones and so was proposed to represent MycA4. A third protein at ~40 kDa was observed that was not found within the negative control or found to correspond to either of the pGro7 chaperones and so was proposed to represent MycA2, and finally a protein band at ~45 kDa putatively corresponding to MycA3 (45.5 kDa).

It should be made clear that there are limitations associated with this data. As a result of the similarity in molecular weight of the chaperone proteins, and protein endogenous to *E. coli*, with MycA1, MycA2 and MycA3, conclusive identification of these enzymes could not be achieved. However, these results do show that, when expressed alone, only putative expression of MycA4, MycA3 and MycA1 is found. Therefore, for viable expression of the entire operon, co-expression of chaperone proteins will be required. Specifically, strains co-expressing pQR2786 with either pKJE7 or pGro7 have been shown to putatively contain MycA1-4 and so are good candidates for confirming the

expression of MycA1-4 through activity analysis. However, as none of these enzymes had been previously characterised, it was decided that each enzyme should be individually expressed, purified and characterised. Then, once functional conditions have been determined for each enzyme, activity can be assessed in strains expressing the CHC CoA operon as a whole to confirm the expression of each enzyme.

### 3.4. Individual cloning of *mycA1-4*

Once the CHC cluster had been cloned into pET29a, it was used as a template for cloning each of the four genes individually. The reason for using this template rather than the genomic DNA was to prevent unwanted non-specific primer binding. PCR was used to amplify each of the four genes (*mycA1-4*), which were designed to add an N-terminal *Nde*I restriction site and a C-terminal *Xho*I restriction site (Figure 3.7).

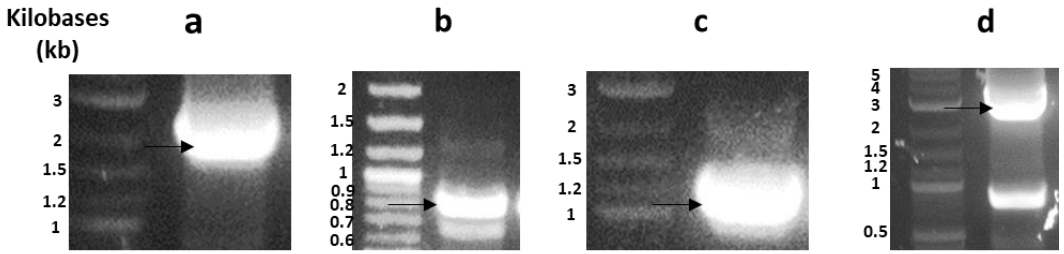



**Figure 3.7. Individual PCR amplification of *mycA1-4.***

An agarose gel showing the PCR amplification of (a) *mycA1*- 2127 bp, (b) *mycA2*- 834 bp, (c) m*ycA3*- 1170 bp and (d) *mycA4*- 3015 bp. Successful PCR amplification was found for each gene, highlighted by a black arrows.

Following successful PCR amplification, each gene was ligated into pET28a, forming four recombinant plasmids (*mycA1*- pQR2787, *mycA2*- pQR2788, *mycA3*- pQR2789, and *mycA4*- pQR2790) that were separately transformed into *E. coli* TOP10. The plasmids from two colonies of each transformation were isolated and digested with *Nde*I and *Xho*I to confirm the ligation of each gene. The results showed all four genes to have been successfully ligated into

pET28a (Figure 3.8). This was further confirmed by sequencing the extracted plasmids using the standard T7 promoter and T7 terminator primers from Eurofins. The pET28a vector was used to add an N-terminal His-Tag allowing each of the enzymes to be purified, which would facilitate enzymatic screening. Once all four of the CHC CoA operon genes had been individually cloned into pET28a vectors, these recombinant plasmids were used to transform *E. coli* BL21 (DE3) cells for the analysis of recombinant protein expression.

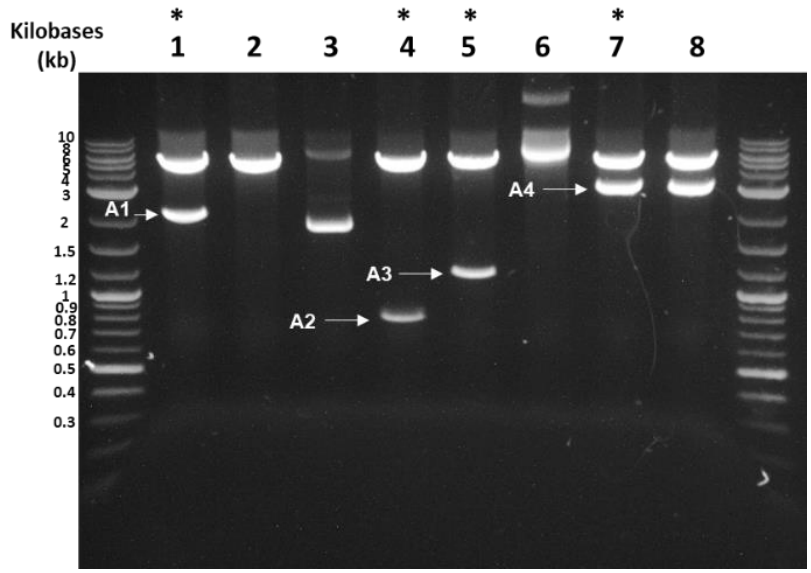



**Figure 3.8. The results of cloning *mycA1-4* into pET28a.**

An agarose gel showing the results of an *Nde*I-*Xho*I analytical digest of (1-2) m*ycA1*, (3-4) m*ycA2*, (4-6) *mycA3*, (7-8) m*ycA4*. Linear DNA fragments corresponding to each gene are indicated (white arrows). The results show successful ligation of *mycA1* in lane 1 (A1), *mycA2* in lane 4 (A2), *mycA3* in lane 5 (A3) and m*ycA4* in both lane 7 and 8 (A4). Asterisks are used to indicate which plasmids were used to transform *E. coli* BL21 (DE3) cells for protein expression.

## 3.5. Expression and purification of MycA4

MycA4 was the first enzyme to be expressed, as it was the only enzyme within the CHC CoA operon that had commercially available substrates. *E. coli* BL21 (DE3) cells that were transformed with pQR2790, containing *mycA4*, were

expressed at three different temperatures (37 °C, 25 °C and 16 °C) and at each temperature induced with two concentrations of IPTG (0.1 mM and 1 mM). The results showed no clear soluble expression of MycA4 for any of the conditions used (Figure 3.9). However, analysis of the insoluble pellet showed there to be a large protein band ~110 kDa in size when expressed at 37 °C induced with 0.1 mM and 1 mM IPTG, 25 °C induced with 0.1 mM and 1mM IPTG and at 16 °C when induced with 1 mM IPTG (Figure 3.10).

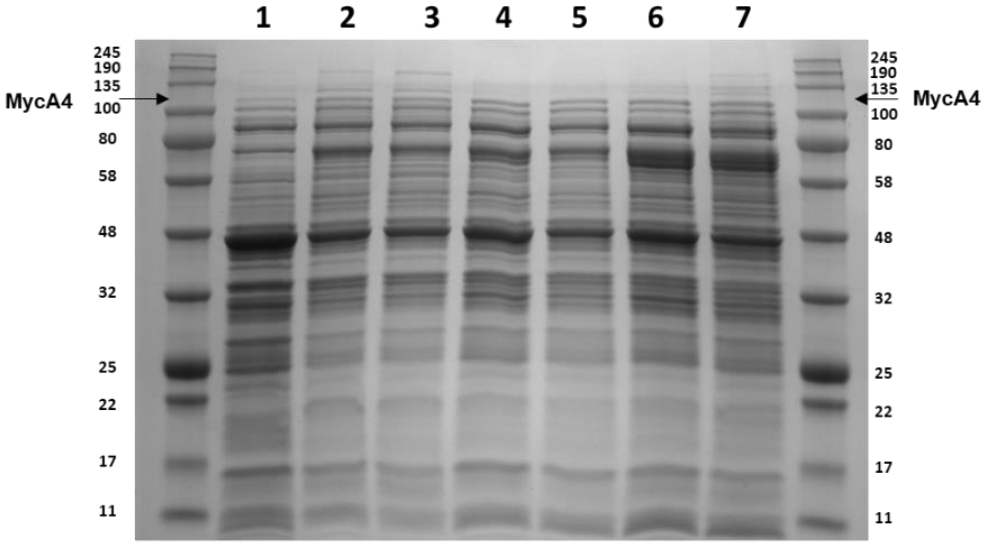



**Figure 3.9. Analysis of soluble MycA4 (pQR2790) expression.**

An SDS gel of the soluble protein fraction of *E.coli* BL21 (DE3) cells transformed with pQR2790 expressed in a range of conditions: (1) non-induced control, grown for 6 hours at 37 °C (2) 0.1 mM IPTG, 37 °C post-induction for 6 hours, (3) 1 mM IPTG, 37 °C post-induction for 6 hours (4) 0.1 mM IPTG, 25 °C post-induction for 24 hours, (5) 1 mM IPTG, 25 °C post-induction, for 24 hours, (6) 0.1 mM IPTG, 16 °C post-induction for 72 hours, (7) 1 mM IPTG, 16 °C post-induction for 72 hours. No clear soluble overexpression of MycA4 was observed at any expression condition (black arrows). Cultures were grown in 50 ml of TB shaking at 250 rpm (37 °C and 25 °C) or 170 rpm (16°C) and were lysed using Bugbuster (chapter 2.10). 100 ug of each clarified lysate was prepared and loaded on the SDS (chapter 2.17).

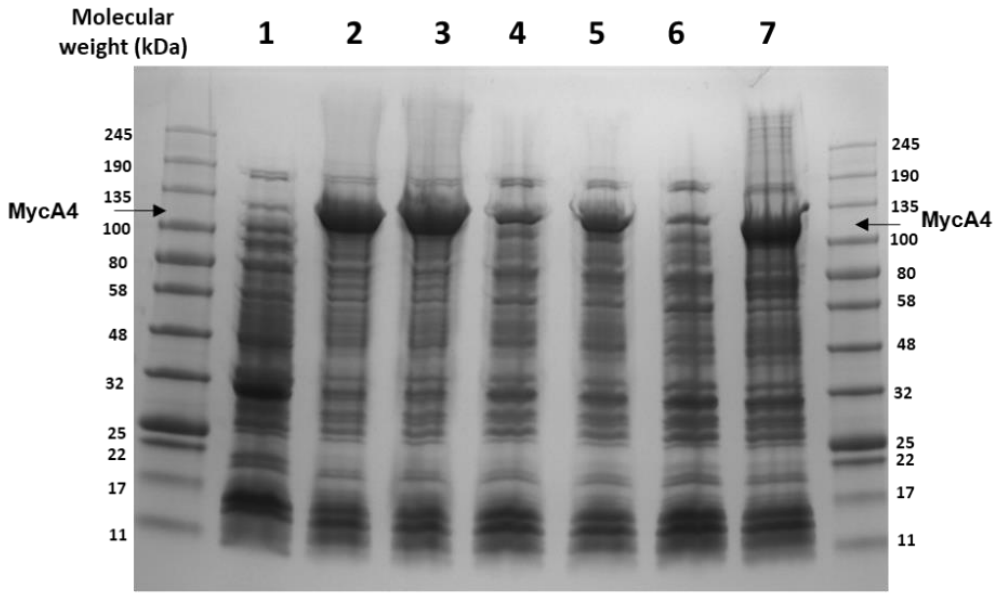



**Figure 3.10. Analysis of insoluble MycA4 (pQR2790) expression.**

An SDS gel of the insoluble protein fraction of *E.coli* BL21 (DE3) cells transformed with pQR2790 expressed in a range of conditions. (1) non-induced control, grown for 6 hours at 37 °C (2) 0.1 mM IPTG, 37 °C post-induction, (3) 1 mM IPTG, 37 °C post-induction, (4) 0.1 mM IPTG, 25 °C post-induction, (5) 1 mM IPTG, 25 °C post-induction, (6) 0.1 mM IPTG, 16 °C post-induction, (7) 1 mM IPTG, 16 °C post-induction. Protein putatively corresponding to MycA4 was found in lanes 2,3,4,5 and 7 (black arrows). Cultures were grown in 50 ml of TB shaking at 250 rpm (37 °C and 25 °C) or 170 rpm (16°C) and were lysed using Bugbuster (chapter 2.10), insoluble protein was prepared according to methods stated previously (chapter 2.12). Length of post-induction growth was, for 37 °C, 6 hours; 25 °C, 24 hours and 16 °C, 72 hours.

As a result of the observed insoluble expression, pQR2790 and all five of the chaperone plasmids were used to co-transform *E. coli* BL21 (DE3) to facilitate soluble protein expression. These five co-transformed strains were then assessed for soluble protein expression. The results showed there to be no clear soluble expression of MycA4 in any of the co-transformed strains (Figure 3.11).

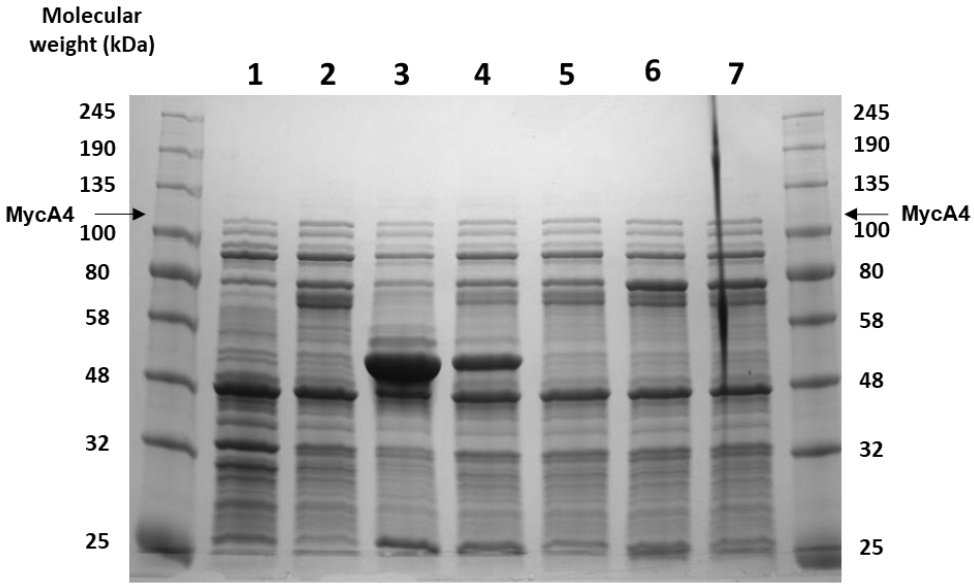



**Figure 3.11. Analysis of soluble MycA4 (pQR2790) expression with chaperones.**

An SDS gel showing the soluble protein content after 72 hours of 16 °C 1 mM IPTG-induced expression in *E.coli* BL21 (DE3) cells transformed or co-transformed with: (1) pET28a (negative control), (2) pQR2790, (3) pQR2790 + pG-Tf2, (4) pQR2790 + pTf16, (5) pQR2790 + pGro7, (6) pQR2790 + pKJE7, (7) pQR2790 + pG-KJE8. The results showed no clear soluble expression of MycA4 in any of the expressed strains analysed. Black arrows were used to indicate expected size of MycA4. Cultures were lysed using Bugbuster (chapter 2.10) and 100 µg of each clarified lysate was prepared and loaded on to the SDS gel according to methods previously described (chapter 2.17). Cultures were grown in 50 ml of TB media shaking at 170 rpm.

However, the insoluble pellet of *E. coli* BL21 (DE3) cells transformed with: pQR2790, pQR2790 + pG-Tf2, pQR2790 + pTf16, pQR2790 + pGro7, pQR2790 + pKJE7 and pQR2790 + pG-KJE8 were all found to contain protein corresponding to MycA4 (Figure 3.12). As co-expression of the CHC CoA operon (pQR2786) with pTf16, pKJE7 and pG-KJE8 had increased the size of the protein band putatively corresponding to MycA4, it was surprising that co-expression of MycA4 with the chaperones plasmids did not lead to any soluble expression. Therefore, it was thought that another factor may be resulting in the insolubility observed, so the effect of the cellular lysis method on protein solubility was investigated.

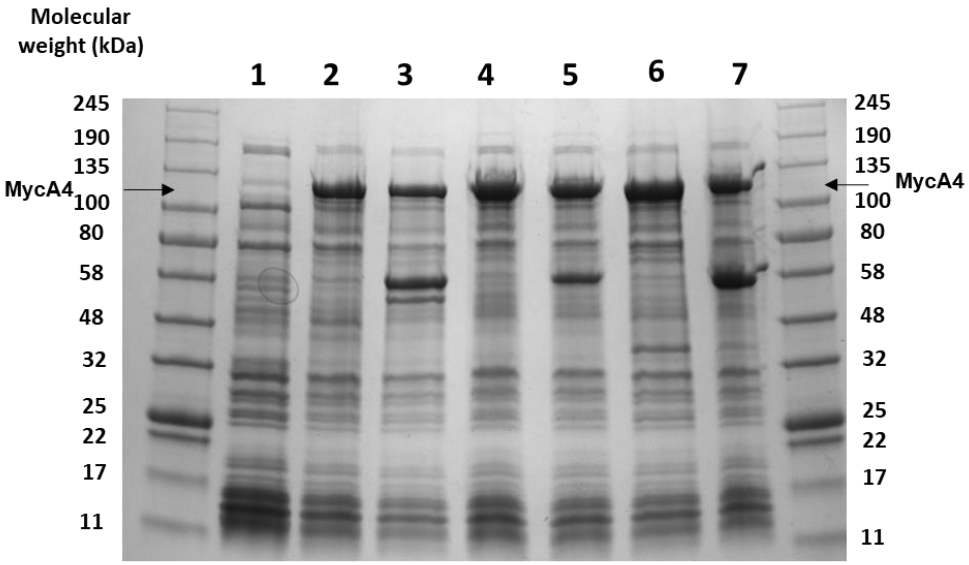



**Figure 3.12. Analysis of insoluble MycA4 (pQR2790) expression with chaperones.**

An SDS gel showing the insoluble protein content after 72 hours of growth at 16 °C, with 1 mM IPTG-induced expression in *E.coli* BL21 (DE3) cells, transformed or co-transformed with: (1) unmodified pET28a (negative control), (2) pQR2790, (3) pQR2790+ pG-Tf2, (4) pQR2790+ pTf16, (5) pQR2790+ pGro7, (6) pQR2790+ pKJE7, (7) pQR2790+ pG-KJE8. The results show a clear protein band between the 100 kDa and 135 kDa markers that is not found in the negative, most likely corresponding to MycA4. All cultures were lysed using Bugbuster (chapter 2.10), insoluble protein was prepared according to methods stated previously (chapter 2.12). Cultures were grown in 50 ml of TB media shaking at 170 rpm.

pQR2790 was co-expressed again with all of the chaperone proteins, however, this time rather than chemical lysis (Bugbuster), sonication was used to disrupt the cells. The results showed protein corresponding to MycA4 in all of the strains expressing the recombinant MycA4 plasmid, including a faint protein band in the strain containing only pQR2790 (Figure 3.13). Densitometry determined the strain co-expressing pQR2790 + pG-Tf2 to produce the highest amount of soluble MycA4 (3.6% of total protein), which was 50% more than when pQR2790 was expressed alone (Table 3.2). This result clearly showed that chemical lysis (Bugbuster) was a contributing factor leading to the protein insolubility observed when expressing the *mycA4* gene.

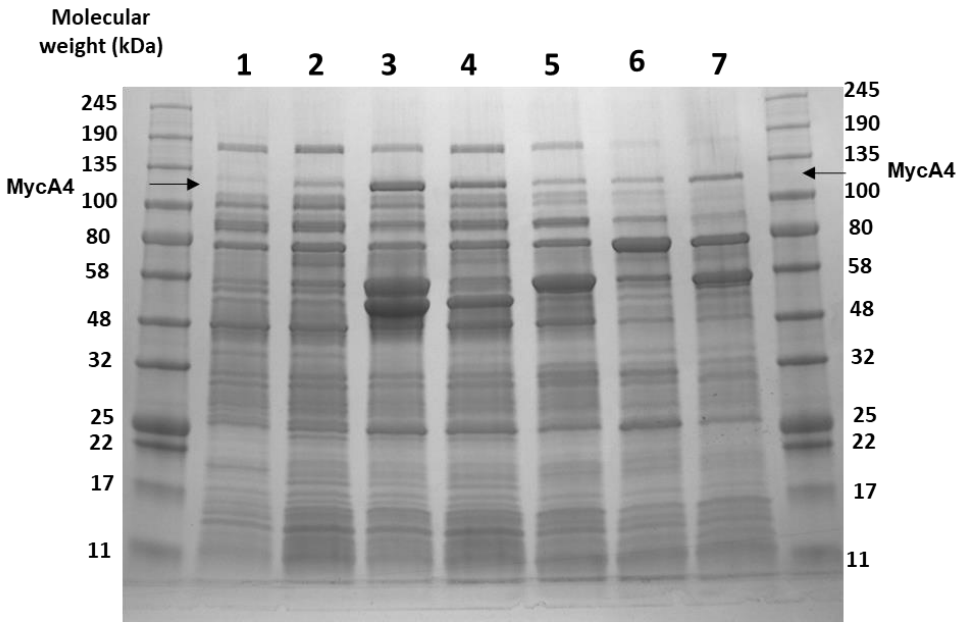



**Figure 3.13. Analysis of soluble MycA4 (pQR2790) expression with chaperones and lysis by sonication.**

An SDS gel showing the soluble protein content after 72 hours of growth at 16 °C, with 1 mM IPTG-induced expression in *E.coli* BL21 (DE3) cells, transformed or co-transformed with (1) pET28a (negative control), (2) pQR2790, (3) pQR2790+ pG-Tf2, (4) pQR2790+ pTf16, (5) pQR2790+ pGro7, (6) pQR2790+ pKJE7, (7) pQR2790+ pG-KJE8. The results show a clear protein band between the 100 kDa and 135 kDa marker that is not found in the negative, corresponding to MycA4. The largest protein band is observed in the strain co-transformed pG-Tf2. Black arrows indicate the expected size of MycA4. Cultures were lysed using sonication (chapter 2.11) and 100 µg of each clarified lysate was loaded on to the SDS gel following methods previously stated (chapter 2.17). Cultures were grown in 50 ml of TB media shaking at 170 rpm.

**Table 3.2. The composition of soluble MycA4.**

MycA4 is shown as a percentage of total protein within *E. coli* BL21 (DE3) strains expressing pQR2790, with and without five chaperone plasmids. Densitometry data was determined using the gel shown previously (Figure 3.13).

| Strain | MycA4 % of total protein |
|---|---|
| pQR2790 | 2.4 |
| pQR2790 + pG-Tf2 | 3.6 |
| pQR2790 + pTf16 | 3.1 |
| pQR2790 + pGro7 | 2.1 |
| pQR2790 + pKJE7 | 2.3 |
| pQR2790 + pG-KJE8 | 2.9 |

Following expression of MycA4 in the soluble fraction, MycA4 was purified using a nickel column (see chapter 2.14 for method). The soluble protein content of *E.coli* BL21 (DE3) cultures transformed with pQR2790 and co-transformed with pQR2790 + pG-Tf2, was loaded on to a nickel column for purification. The results showed recombinantly expressed MycA4 had bound to the nickel column and eluted within the 500 mM imidazole elution in both expression cultures (Figure 3.14). The protein concentration of eluate from cultures transformed with pQR2790 and co-transformed with pQR2790 + pG-Tf2 was 1.3 mg/ml and 5.26 mg/ml, respectively.

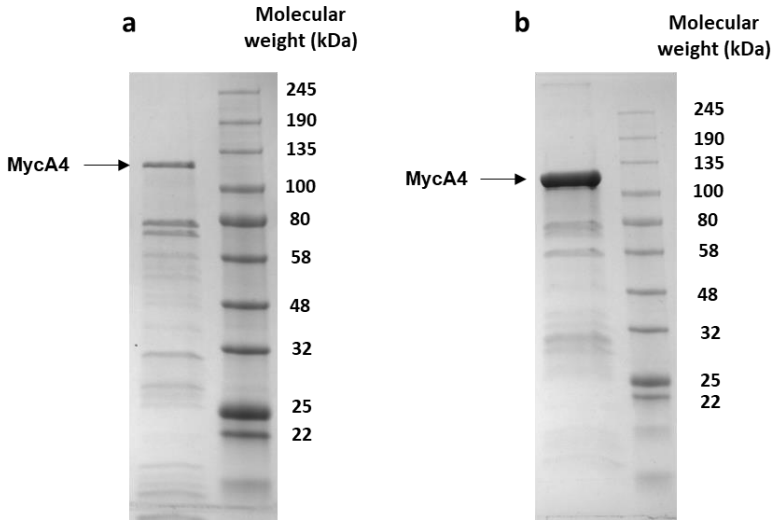



**Figure 3.14. His-tag purification of MycA4.**

An SDS gel showing the soluble protein content that eluted from a nickel column following a 500 mM imidazole wash (chapter 2.14). (a) *E. coli* BL21 (DE3) transformed with pQR2790 and (b) *E. coli* BL21 (DE3) containing pQR2790 + pG-Tf2. Protein bands corresponding to MycA4 are indicated with black arrows. 5 µl of each eluate was loaded on to the SDS gel according to methods stated previously (chapter 2.17). Soluble protein was quantified in each eluate (chapter 2.16) and was determined (a) 1.3 mg/ml and (b) 5.26 mg/ml. Cultures were grown in 300 ml TB media, for 72 hours at 16 °C after induction with 1 mM IPTG, shaking at 170 rpm.

### 3.6. Expression and purification of MycA3

The recombinant MycA3-pET28a plasmid, pQR2789, was then used to transform *E.coli* BL21 (DE3) for the analysis of recombinant protein expression. Analysis of an SDS gel showed the clarified lysate to have overexpressed a soluble ~43 kDa protein marker, corresponding to the expected size of MycA3 (45.5 kDa) (Figure 3.15)**.** The clarified lysate was loaded on to a nickel column and the protein corresponding to MycA3 bound to the column and was then found to elute within the 500 mM imidazole elution wash.

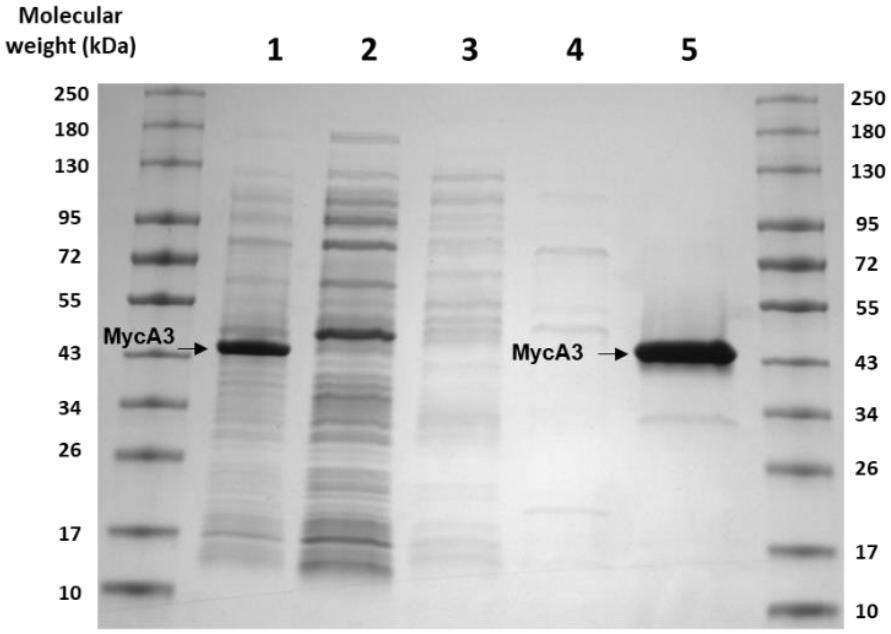



**Figure 3.15. The soluble expression and purification of MycA3 (pQR2789).**

An SDS gel showing the soluble protein content of *E. coli* BL21 (DE3) cells expressing pQR2789 and the eluted fractions following a nickel column purification. (1) Clarified lysate prior to purification (2) 10 mM imidazole wash (3) 50 mM imidazole wash (4) 100 mM imidazole wash (5) 500 mM imidazole elution. The results show the overexpression of a soluble protein at ~43 kDa in lane 1 (black arrow) corresponding to MycA3 (expected 45.5 kDa). MycA3 is bound to the nickel column and so was not observed in lane 2, 3 or 4. MycA3 is then eluted within the 500 mM imidazole wash eluate. The expression culture was lysed using sonication (see chapter 2.11) and then purified (see chapter 2.14). 5 µl of each collected eluate and clarified lysate was prepared and loaded on to the SDS gel (see chapter 2.17). The culture was induced with 1 mM IPTG, and then grown for 72 hours at 16 °C in 300 ml TB media, shaking at 170 rpm (see chapter 2.9).

### 3.7. Expression and purification of MycA2

The recombinant MycA2-pET28a plasmid, pQR2788, was used to transform *E.coli* BL21 (DE3) for recombinant protein expression. The results showed an overexpressed soluble protein between the 34 kDa and 43 kDa protein markers, which had not been seen in any of the previous negative control experiments, expressed under these conditions (16 °C post-induction, 1 mM IPTG, 72 hours). This protein was slightly larger than the expected size of MycA2 (32 kDa). However, the soluble lysate was loaded onto a nickel column and the overexpressed protein was found to bind and then correspondingly elute within the 500 mM imidazole wash, suggesting this protein to be MycA2 (Figure 3.16).

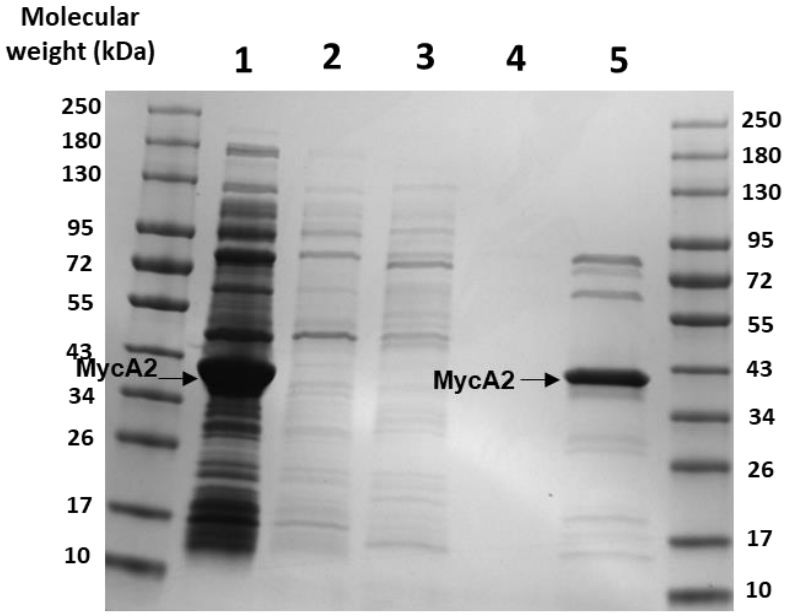



**Figure 3.16. The soluble expression and purification of MycA2 (pQR2788).**

An SDS gel showing the soluble protein content of (1) *E. coli* BL21 (DE3) cells expressing MycA2 (32 kDa) from pQR2788 and the eluate fractions following a nickel column purification. (2) 10 mM imidazole wash (3) 50 mM imidazole wash (4) 100 mM imidazole wash (5) 500 mM imidazole wash. The results show the overexpression of a soluble protein between 34-43 kDa, in lane 1 (black arrow), putatively MycA2, which bound to the nickel column. This protein is then found to elute within the 500 mM imidazole fraction (black arrow). 5 µl of each collected eluate and clarified lysate was prepared and loaded on to the SDS gel (see chapter 2.17). The culture was induced with 1 mM IPTG, and then grown for 72 hours at 16 °C in 300 ml TB media, shaking at 170 rpm (see chapter 2.9).

### 3.8. Expression and purification of MycA1

The recombinant MycA1 plasmid, pQR2787, was used to transform *E. coli* BL21 (DE3) cells for IPTG-induced expression. Multiple expression conditions were assessed including growth in TB and LB media and induction with 0.1 mM and 1 mM IPTG. Analysis of the whole cell lysate following 24 hours at 25 °C showed a protein in all expression conditions between 72 kDa and 95 kDa, which was not found in the negative control (Figure 3.17). The expected molecular weight of MycA1 was 75.2 kDa and so this protein was proposed to be MycA1.

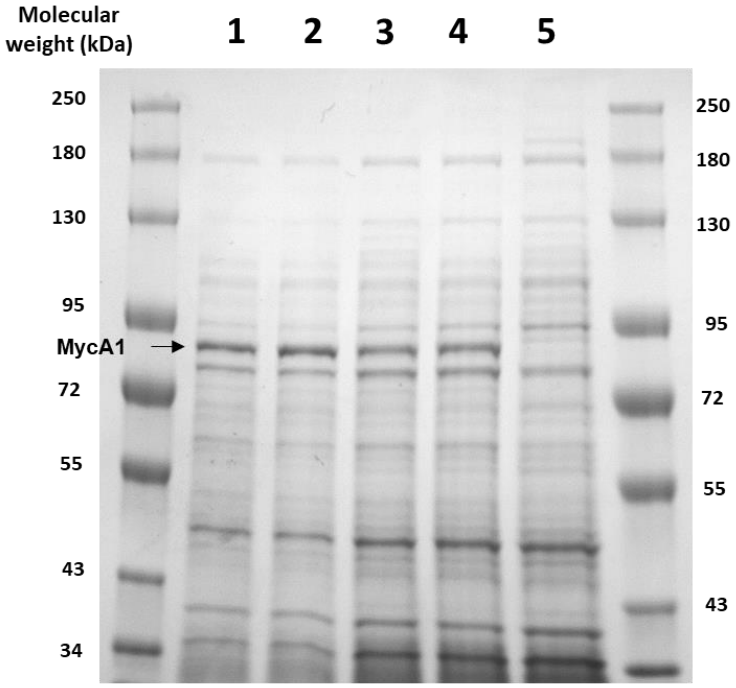



**Figure 3.17. Analysis of MycA1 (pQR2787) expression from whole cell lysate.**

An SDS gel showing the whole cell lysate from *E. coli* BL21 (DE3) cells expressing pQR2787 in a number of conditions: (1) LB media induced with 0.1 mM IPTG, (2) LB media induced with 1 mM IPTG, (3) TB media induced with 0.1 mM IPTG and (4) TB media induced with 1 mM IPTG. (5) The whole cell lysate of *E.coli* BL21 (DE3) cells transformed with an unmodified pET28a plasmid was used as a negative control. Putative MycA1 (75.2 kDa) is indicated with a black arrow. A 1 ml aliquot of each culture was normalised to an $OD_{600}$ of 2.5 and then 5 µl of each culture was prepared and loaded on to the SDS gel as previously described for clarified lysate (chapter 2.17). All cultures were grown in 50 ml media at 37 °C prior to IPTG induction, which was changed to 25 °C immediately after induction. In all conditions cultures were shaking at 250 rpm.

The clarified lysate obtained from all four cultures was then pooled together and loaded on to a nickel purification column. MycA1 was designed to contain an N-terminal His-tag and so could be purified if it was soluble. The results showed no purification of MycA1 suggesting it to be an insoluble protein (Figure 3.18).

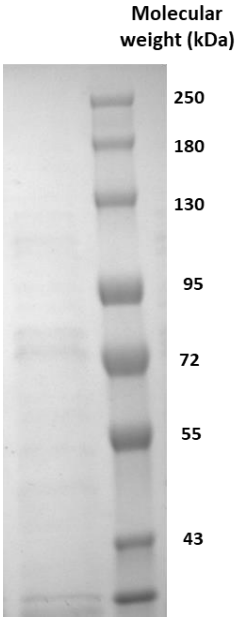



**Figure 3.18. Analysis of the His-Tag purification of MycA1 (pQR2787).**

An SDS gel showing 5 µl of the eluate (500 mM imidazole) from a nickel column that had been loaded with the pooled clarified lysates of the four expression cultures assessed previously (Figure 3.17). No clear soluble MycA1 is observed to have been purified. Samples were prepared and loaded on the SDS gel as previously described (chapter 2.17).

As the substrate from MycA1 was not commercially available and it had shown putative soluble expression when expressed as part of the CHC CoA operon, further time was not spent on attempting to confer soluble expression of MycA1. Instead the focus was turned to determining the activity of purified MycA4.

### 3.9. Activity analysis of MycA4

MycA4 is proposed to catalyse the first committed step of the CHC CoA operon by converting either shikimate or 3-phosphoshikimate (3-PS) to 3,4-dihydroxycyclohex-1,5-enecarbonyl CoA (3,4-diHCH-1,5eneC CoA) via a two-step mechanism (Figure 3.19). The two steps of this mechanism are proposed to be: (1) CoA activation and (2) either dehydration or dephosphorylation (depending on if shikimate or 3-phosphoshikimate is the substrate, respectively).

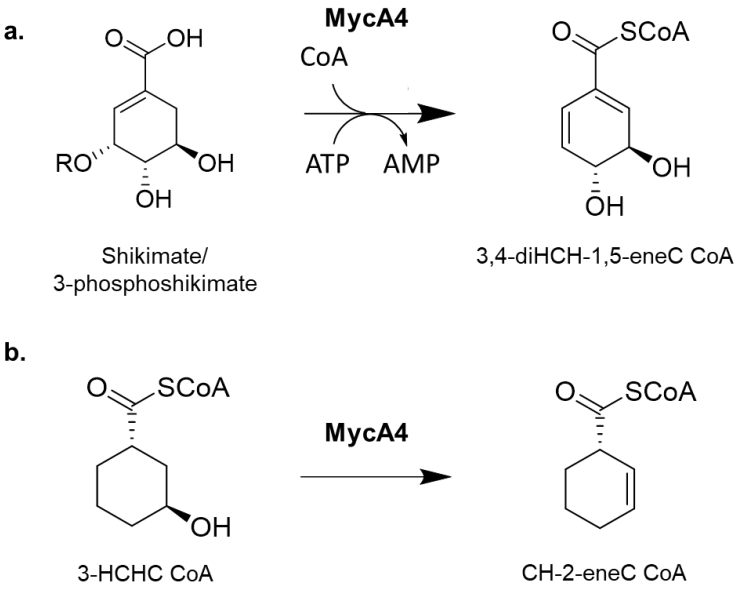



**Figure 3.19. The two proposed functions of MycA4 within the CHC CoA pathway.**

In the first function (a) shikimate or 3-phosphoshikimate is activated into a CoA ester and then dehydrated (shikimate) or dephosphorylated (shikimate-3-phosphate), forming 3,4-dihydroxycyclohex-1,5-enecarbonyl CoA (3,4-diHCH-1,5-eneC CoA). In the second proposed function (b) 3-hydroxycyclohexanecarbonyl CoA (3-HCHC CoA) is dehydrated into cyclohex-2-enecarbonyl CoA (CH-2-eneC CoA).

Initial assays were set up in parallel containing either shikimate or 3-phosphoshikimate in order to determine the native substrate of this enzyme. As there was no positive control available, a range of conditions had to be assessed with a range of substrates and cofactors, in order to determine the correct condition for catalysis. Due to this enzyme having a CoA ligase

domain, CoA should be used which will be added to the organic acid, allowing the assay to be monitored through the consumption of CoA. The results showed no significant consumption of CoA after a 30 minute assay with either shikimate or shikimate-3-phosphate as the substrate (Figure 3.20).

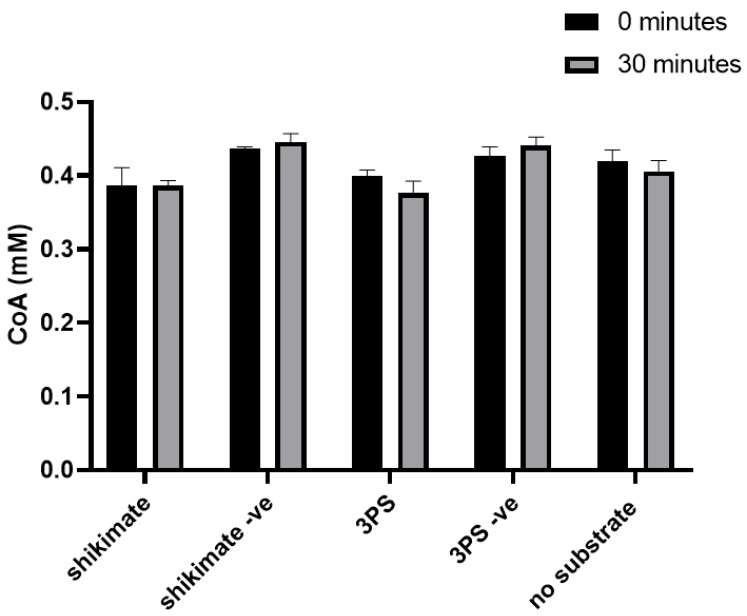



**Figure 3.20. Analysis of MycA4 activity.**

An end-point assay monitoring MycA4 activity through the consumption of CoA. Shikimate and 3-phosphoshikimate (3-PS) were both assessed as substrates, negative controls contained no MycA4 and 'no substrate' contains all assay components except the organic acid (shikimate or shikimate-3-phosphate). All reactions started with 0.4 mM of CoA and carried out as previously described (chapter 2.19).

It was thought that MycA4 may be experiencing product inhibition, therefore, the same assay was repeated containing MycA2, the next proposed enzyme in the CHC CoA pathway. However, the results showed no significant consumption despite the addition of MycA2 (Figure 3.21). In an attempt to identify the conditions in which MycA4 was active, a number of additional conditions were assessed including: using alternative buffers (HEPES), pyruvate as a cofactor (reverse EPSP synthase), different metal ions (KCl), using MycA4 within a clarified lysate, addition of phosphoenolpyruvate and addition of $FMNH_2$ and $FADH_2$, however, none of these conditions elicited the consumption of CoA. In addition, preliminary experiments were conducted using 3-HCHC CoA (synthesised in chapter 6.2) as a substrate for the second

function of MycA4. Unfortunately, no activity was observed and due to insufficient time a range of conditions could not be assessed with this second function.

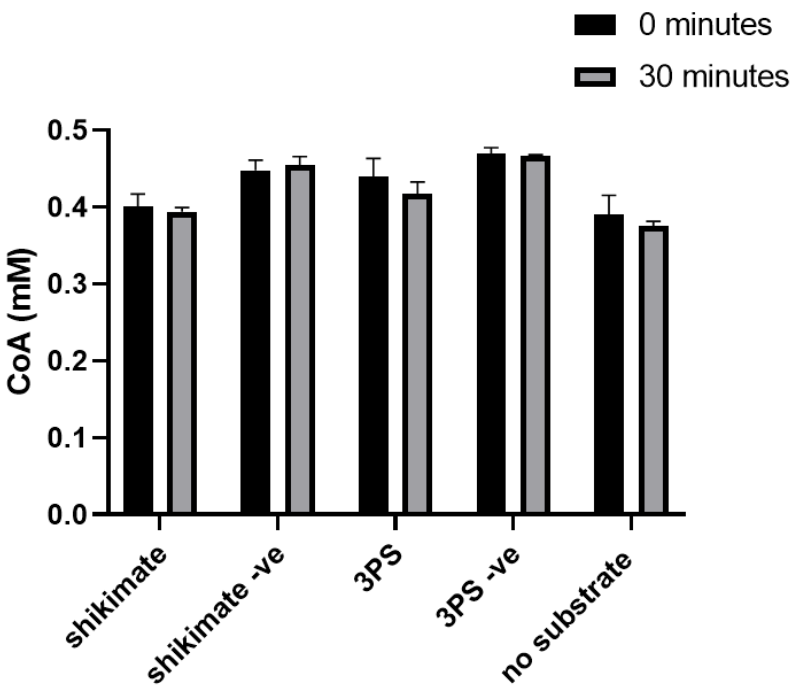



**Figure 3.21. Analysis of MycA4 activity with addition of MycA2.**

An end-point assay monitoring MycA4 activity with the addition of MycA2**,** through the consumption of CoA. Shikimate and 3-phosphoshikimate were both assessed as substrates, negative controls (shikimate –ve and 3PS –ve) contained no MycA4 and 'no substrate' contains all assay components except the organic acid (shikimate and 3-phosphoshikimate). All reactions started with 0.4 mM of CoA and carried out as previously described (chapter 2.19).

### 3.10. MycA2 activity

MycA2 is proposed to catalyse the α-β double bond reduction of three intermediates of the CHC CoA pathway: (1) cyclohex-1-enecarbonyl CoA, (2) 3-hydroxycyclohex-1-enecarbonyl CoA and (3) 3, 4 -dihydroxycyclohexa-1, 5-dienecarbonyl CoA (Figure 3.22).

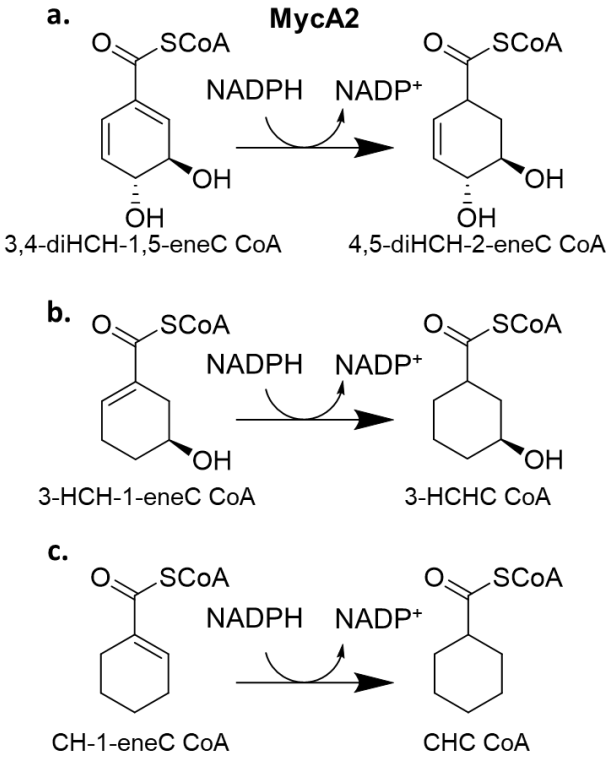



**Figure 3.22. The proposed functions of MycA2 within the CHC CoA pathway.**

The three proposed functions of MycA2 include: (a) Reduction of 3,4-dihydroxycyclohexa-1,5-dienecarbonyl CoA (3,4-diCH-1,5-eneC CoA) to 4,5-dihydroxycyclohex-2-enecarbonyl CoA (4,5-diCH-2-eneC CoA), (b) Reduction of 3-hydroxycyclohex-1-enecarbonyl CoA (3-HCH-1-eneC CoA) to 3-hydroxycyclohexanecarbonyl CoA (3-HCHC CoA) and (c) reduction of cyclohex-1-enecarbonyl CoA (CH-1-eneC CoA) to cyclohexanecarbonyl CoA (CHC CoA).

The homolog from *S. collinus*, ChcA, has been shown to be active with each of these three CoA esters. None of these CoA esters are commercially available, however, following the work in chapter 6, 1-cyclohexenecarbonyl CoA and 3-hydroxycyclohex-1-enecarbonyl CoA could be enzymatically

synthesised using a CoA ligase (AliA) and an acyl CoA dehydrogenase (ChCoADH). Therefore, following the removal of the CoA ligase and acyl-dehydrogenase, these two enzymatic products were used as substrates for screening MycA2 activity.

### 3.10.1. Cyclohex-1-enecarbonyl CoA reductase activity

As there is no chemical standard available for CH-1-eneC CoA, the exact concentration used could not be confirmed, however, if it was assumed that 100% of substrate (CHC CoA) had been converted into CH-1-eneC CoA, the concentration would be 0.4 mM. The reaction was monitored by the drop in concentration of NADPH at 340 nm. The results showed a significant drop in NADPH concentration compared to the negative control following the addition of MycA2, and the reaction was complete within six minutes (Figure 3.23).

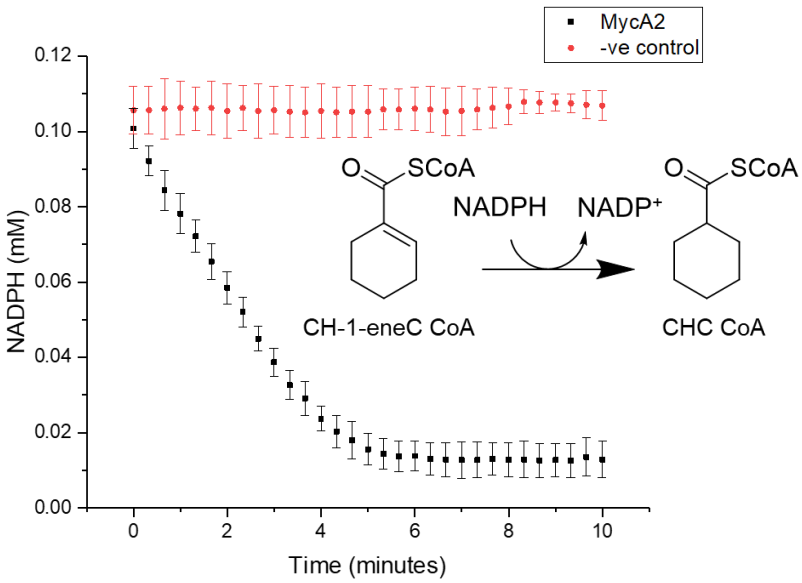



**Figure 3.23. MycA2 reductase activity with CH-1-eneC CoA.**

The reductive activity of MycA2 using CH-1-eneC CoA as a substrate. Activity was monitored through the consumption of NADPH as it is oxidised to NADP. The negative control shown contained no MycA2. Both assays were run in triplicate with standard error shown. Absorbance was $340_{nm}$ was used to monitor NADPH concentration. The assay was performed as described previously (chapter 2.20) with 0.1 mM NADPH.

A previous investigation had shown that an enoyl-CoA reductase from *S. collinus* (not ChcA) was inhibited by >0.2 mM of NADPH (Wallace *et al*, 1995). Although the tolerance of NAPDH towards ChcA had not been established, determining whether MycA2 was inhibited by similar concentrations of NADPH was thought to be an important consideration. In addition, when all of the CHC CoA operon-encoded enzymes have been characterised, understanding the maximal concentrations of redox cofactors, such as NADPH, will be crucial for optimal functionality. MycA2 was assayed against a fixed concentration of CH-1-eneC CoA and varying concentrations of NADPH (0.05 – 0.7 mM) to observe whether there was cofactor inhibition. The results showed an overall increase in the initial rate of reactions as the concentration increased from 0.05 mM to 0.7 mM (Figure 3.24), suggesting MycA2 to not exhibit the same NADPH inhibition found with the enoyl-CoA reductase from *S. collinus* (Wallace *et al*, 1995).

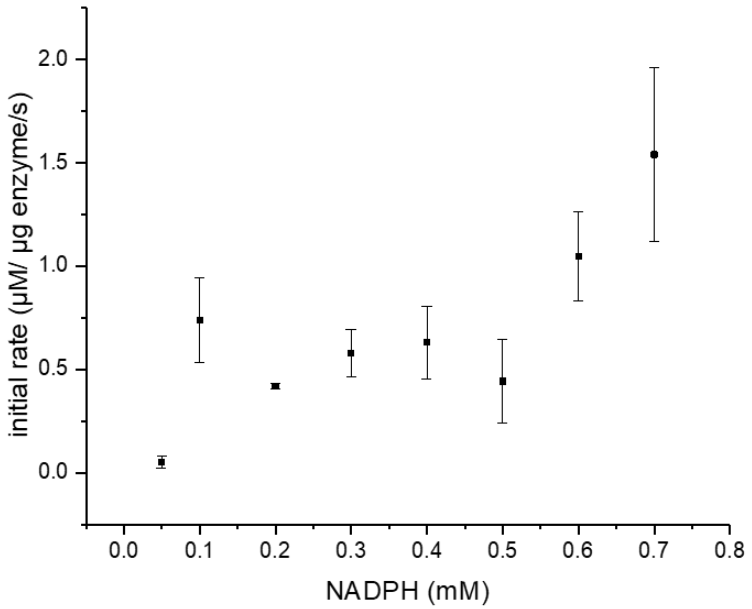



**Figure 3.24. Analysis of MycA2 activity with varying NADPH concentrations.**

Analysis of the effect of increasing NADPH concentration on the initial rate determined for MycA2 with CH-1-eneC CoA as a substrate. The concentration of CH-1-eneC CoA remained constant throughout all concentrations of NADPH. The assay was performed as described previously (chapter 2.20).

Two anomalies were observed: the first was that a concentration of 0.1 mM NADPH (0.74 $\mu M^{-1}\mu g^{-1}s^{-1}$) led to a higher initial rate than that found with 0.2-0.5 mM NADPH, and the second was addition of 0.5 mM of NADPH ( 0.44 $\mu M^{-1}\mu g^{-1}s^{-1}$) resulted in a lower initial rate than that of 0.3 mM and 0.4 mM NADPH. However, if NADPH activity inhibition occurred, then a certain concentration of NADPH would result in activity loss. As the highest initial rate, 1.54 $\mu M^{-1}\mu g^{-1}s^{-1}$, was with the highest concentration of NADPH (0.7 mM), it seems clear that NADPH is not inhibiting MycA2 activity.

To confirm this drop in NADPH present was due to the reduction of CH-1-ene-C CoA into CHC CoA, the reaction was also monitored using HPLC. The HPLC analysis confirmed that the peak area corresponding to CH-1-eneC CoA dropped by 44%, which was followed by the production of 0.17 mM CHC CoA after a 10 minute assay (Figure 3.25). Verification of the identity of the peak area corresponding to CH-1-eneC CoA was provided through HRMS analysis (chapter 4, Figure 4. 16).

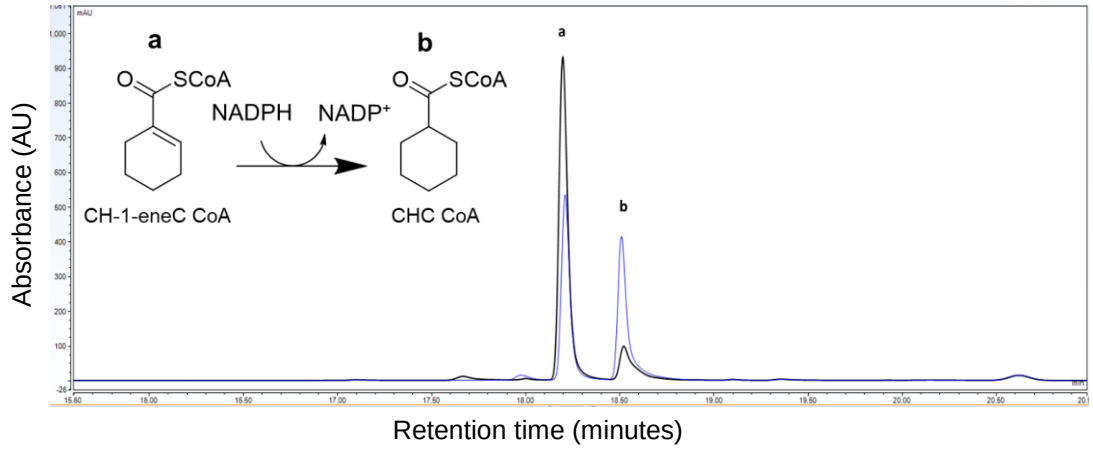



**Figure 3.25. HPLC analysis of MycA2 activity with CH-1-eneC CoA.**

HPLC analysis of MycA2 reductase activity with (a) CH-1-eneC CoA, forming (b) CHC CoA. The initial time point was taken immediately after the addition of MycA2 and the second time point was 10 minutes after MycA2 addition (blue line). The reaction was prepared according to chapter 2.20 with the addition of 0.2 mM NADPH.

### 3.10.2. 3-hydroxycyclohex-1-enecarbonyl CoA reductase activity

MycA2 was next screened for reductase activity with 3-hydroxycyclohex-1-enecarbonyl CoA (3-HCH-1-eneC CoA). Like with CH-1-eneC CoA, there is no commercial standard of this substrate, so the exact concentration, following its catalysis (see chapter 6), could not be determined. Analysis of the concentration of NADPH following the addition of MycA2 showed it to significantly decrease compared to the negative control, with 0.065 mM of NADPH being consumed after 10 minutes (Figure 3.26).

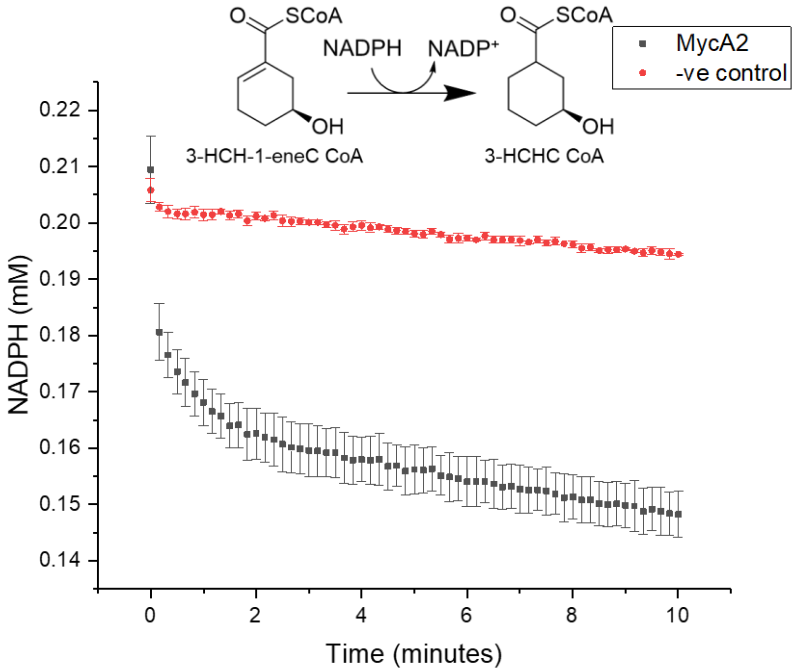



**Figure 3.26. Analysis of MycA2 activity with 3-HCH-1-eneC CoA.**

Monitoring the reductive activity of MycA2 through the consumption of NADPH using 3-HCH-1-eneC CoA as its substrate (black dots), over a 10 minute time period. The negative control (red dots) contains no purified MycA2. Both assays were run in triplicate with standard error shown. Absorbance at $340_{nm}$ was used to monitor NADPH concentration. The assay was performed as described previously (chapter 2.20), with the addition of 0.2 mM NADPH.

Furthermore, to determine that the observed consumption of NADPH corresponded to the reduction of 3-HC-1-eneC CoA to give 3-hydroxycyclohexanecarbonyl CoA (3-HCHC CoA), HPLC analysis was used

to monitor the change in concentration of both of these CoA esters. HPLC analysis showed the peak area corresponding to 3-HCH-1-eneC CoA dropped by 87% after 10 minutes, which corresponded to an increase in the peak area corresponding to 3H CHC CoA (Figure 3.27).

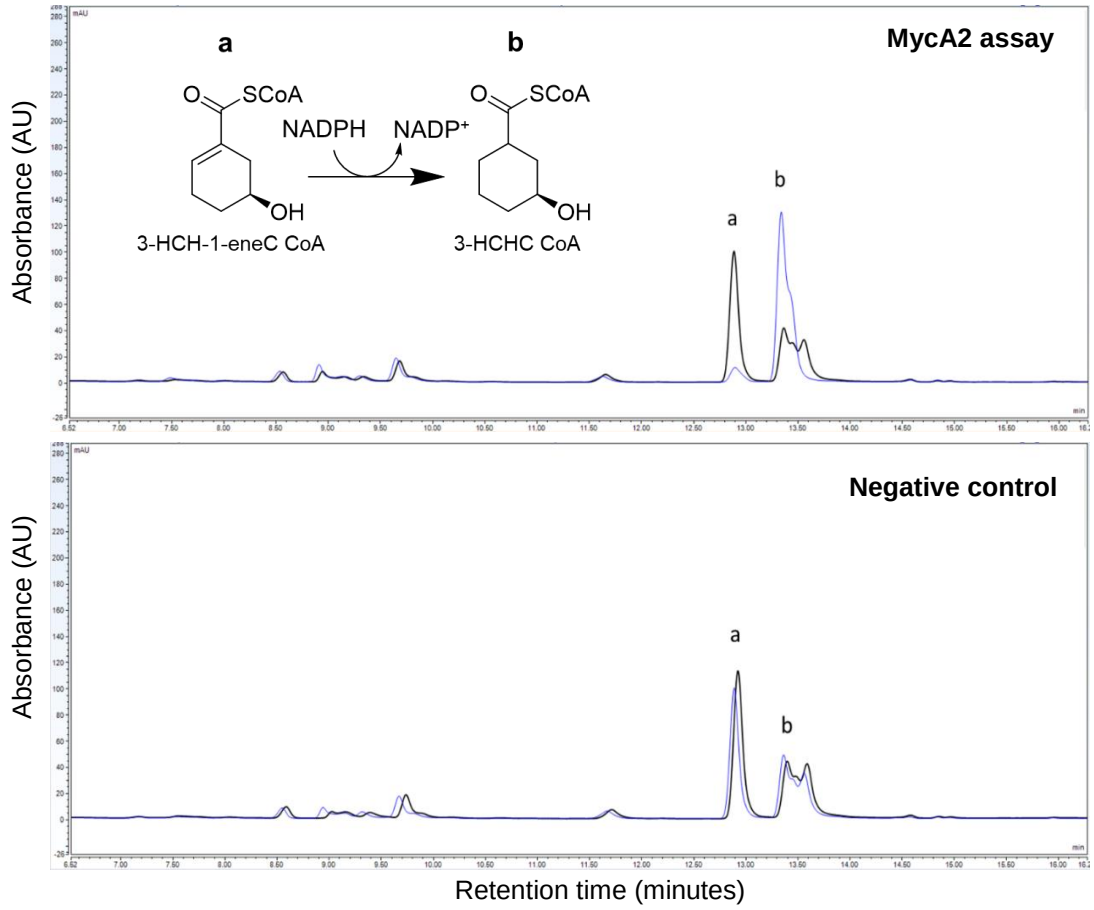



**Figure 3.27. HPLC analysis of MycA2 activity with 3-HCH-1-eneC CoA.**

HPLC analysis MycA2 reductase activity via the reduction of (a) 3-HCH-1-eneC CoA into (b) 3-HCHC CoA. The top panel shows the results with purified MycA2 immediately after the addition of MycA2 (black line) and 10 minutes after MycA2 addition (blue line). The bottom panel shows the negative control containing no purified MycA2 immediately after (black line) and 10 minutes after (blue line) the addition of water (instead of MycA2). The assay was performed as described previously (chapter 2.20), with the addition of 0.2 mM NADPH.

However, the exact 3-HCHC CoA stereoisomer that is produced is unknown. The results shown here clearly show that MycA2 from *S. rishiriensis* is able to catalyse the conversion of 3-hydroxycyclohex-1-enecarbonyl CoA to 3H CHC CoA. Verification of the identity of the HPLC peak area corresponding to 3-

HCHC CoA and 3-HCH-1-eneC CoA is described in chapter 6 (Figure 6.13 and Figure 6.15, respectively).

## 3.11. Discussion

### 3.11.1. MycA4 insolubility with chemical lysis (Bugbuster)

Initial experiments that were focused on the expression of *mycA4* into its corresponding polypeptide, found that chemical lysis using Bugbuster (invitrogen) probably caused protein insolubility. Bugbuster had been routinely used in prior experiments such as when lysing *E. coli* BL21 (DE3) cultures expressing recombinant thioesterases (TE) (chapter 5), and so Bugbuster was not immediately considered as a factor effecting protein solubility. Furthermore, when assessing a number of different conditions for optimal recombinant protein expression, smaller cultures sizes were used (50 ml or less). As a result, cell pellets were re-suspended in small buffer volumes imposing difficulties on lysis through sonication, due to submerging the probe within cell suspensions and the associated foaming. Despite varying the expression conditions and co-expressing *mycA4* (pQR2790) with various chaperone plasmids this was not sufficient to overcome the protein insolubility invoked by the addition of Bugbuster.

In a previous investigation it was observed that a much lower yield of soluble protein was produced following lysis with Bugbuster compared to sonication (Listwan *et al*, 2010). Furthermore, comparison of the soluble yields attained from a range of recombinant proteins lysed with Bugbuster, found the lowest soluble yields were associated with larger proteins (459 amino acids (aa) – 14% soluble yield) (Listwan *et al.*, 2010). This correlates with the observations in this study where smaller proteins such as the recombinant TEs (134-291 aa) expressed in chapter 5 remained soluble in Bugbuster, whereas MycA4 (1024 aa) was completely insoluble. Typical chemical lysis methods use (1)

detergent addition, (2) solvent addition, (3) acidification, (4) chaotrope addition and (5) metal chelation (Pieracci *et al.*, 2018). The exact components of Bugbuster are not known, so the specific cause of insolubility cannot be determined, however, it is clear that in addition to cell disruption a component of Bugbuster causes recombinant protein insolubility.

### 3.11.2. Improved recombinant MycA4 solubility using chaperone proteins

Despite the observation that using sonication for cell lysis rather than Bugbuster resulted in the soluble expression of MycA4, certain chaperone proteins were found to be required for improved soluble protein production. The strain transformed with pQR2790 + pG-Tf2 was identified as the highest producer of soluble MycA4 (3.6% of total soluble protein) and so was used for the purification step. The pG-Tf2 chaperone plasmid encodes the chaperone proteins: GroEL, GroES and Trigger Factor (TF). GroEL and GroES function together with GroEL forming a large barrel shaped protein complex with GroES functioning as a dome-shaped lid cofactor. Improperly folded substrate proteins, such as MycA4, contain hydrophobic regions on their surface which facilitates binding to the GroES-GroEL complex through hydrophobic interactions. The GroES-GroEL complex is then able to facilitate refolding of the substrate protein by the burial of non-polar amino acid residues into the interior of the substrate protein (Lu *et al*, 2019). This process results in the release of a substrate protein in a native or native-like conformation (Martin, 1998). Conversely, TF is the only known ribosome-associated chaperone from bacteria. It functions through the binding of nascent polypeptide chains as they are elongated from the ribosome, and together with the ribosomal surface, forms a ‘cage’ that aids native folding and the ‘burial’ of hydrophobic regions (Hoffmann, Bukau and Kramer, 2010; Singhal *et al.*, 2015). It is clear that the combined activity of GroES-EL and TF can aid misfolded MycA4, when individually expressed, leading to the formation of a soluble native, or near native, protein.

When expressed as part of the CHC CoA operon (pQR2786), co-expression with the pG-Tf2 chaperone plasmid did not increase the cellular composition of soluble MycA4 (0.8% of total protein), compared to when expressed alone (2.1% total protein). Furthermore, co-expression with pTF16, which only contains TF resulted in the highest cellular composition of soluble MycA4 (6.1% of total protein). This would suggest that there is not always a synergistic effect of using multiple chaperones proteins and in fact additional chaperones in some contexts, such as expression of an operon, limit soluble protein production. Furthermore, the highest composition of MycA4 within a clarified lysate was when pQR2786 was co-expressed with the pTf16 chaperone plasmid. The simultaneous expression of all four enzymes encoded within the CHC CoA operon may lead to lower cellular concentrations of mRNA containing *mycA4*, which when combined with the co-expression of the chaperone proteins encoded with the pTf16 plasmid, my confer more preferable conditions for soluble MycA4 expression.

**Table 3.3. The soluble MycA4 expressed from pQR2786.**

The composition of soluble MycA4 as a percentage of total protein within *E. coli* BL21 (DE3) strains expressing the CHC CoA operon (pQR2786) with and without five chaperone plasmids. Densitometry was used to calculate protein composition use gel shown in shown previously (Figure 3.6).

| Strain | MycA4 % of total protein |
|---|---|
| pQR2786 | 2.1 |
| pQR2786+ pG-Tf2 | 0.8 |
| pQR2786+ pTf16 | 6.1 |
| pQR2786+ pGro7 | 1.2 |
| pQR2786+ pKJE7 | 2.8 |
| pQR2786+ pG-KJE8 | 4.1 |

### 3.11.3. MycA4 inactivity

MycA4 is proposed to catalyse the first committed step of the CHC CoA pathway, which is the conversion of either shikimate or 3-phosphoshikimate (3-PS) into 3, 4 hydroxycyclohex-1,5-enecarbonyl CoA. This transformation is proposed to commence via a two-step mechanism involving CoA activation and subsequent dehydration or dephosphorylation (depending on if shikimate or 3-phosphoshikimate is the substrate, respectively). Analysis of the amino acid sequence showed there to be two distinct domains, a CoA ligase domain and a 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase domain. A CoA ligase functions in a two-step mechanism via an acyl-adenylate (acyl-AMP) intermediate, producing a CoA ester from CoA and an organic acid (Morgan-Kiss and Cronan, 2004)**.** This mechanism correlates to the CoA activation mechanism proposed with MycA4**.** However, an EPSP synthase converts 3-PS into 5-enolpyruvylshikimate-3-phosphate (EPSP) via an enolpyruvyl transfer from phosphoenolpyruvate (PEP) to the 5-hydroxyl group on 3-PS (Dos Santos *et al.*, 2017). It was not clear how the EPSP synthase mechanism correlated to the dehydration/ dephosphorylation step proposed for MycA4 (Figure 3.28).

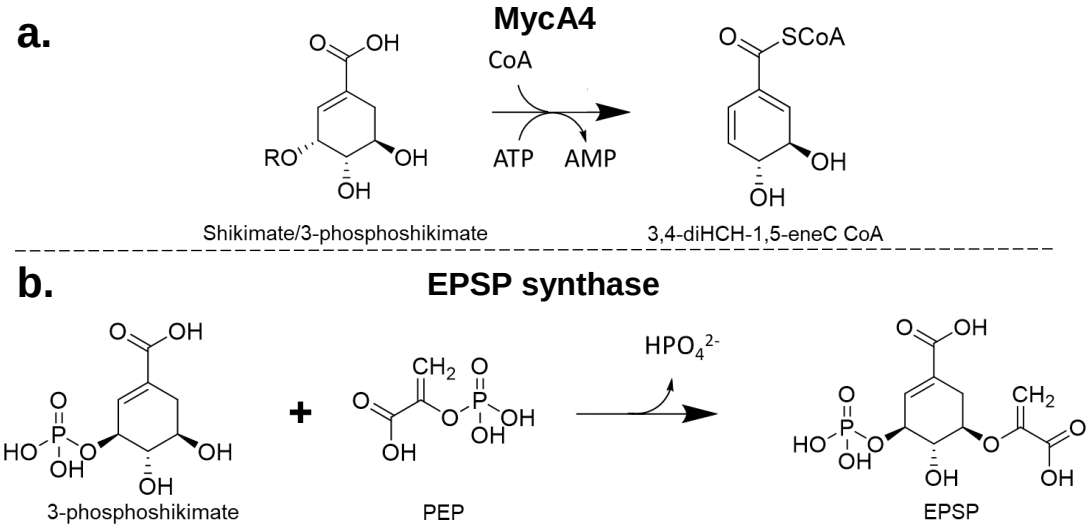



**Figure 3.28. The proposed functions of MycA4 and an EPSP synthase.**

The biochemical mechanism for (a) proposed MycA4 activity and (b) known EPSP synthase activity, which converts 3-phosphoshikimate and phosphoenolpyruvate (PEP) into 5-enolpyruvylshikimate-3-phosphate (EPSP). It is unclear how the EPSP synthase domain within MycA4 is used with its activity.

Analysis of the EPSP synthase domain of MycA4 showed it to contain a number of highly conserved residues shown to be involved in EPSP synthase catalysis such as: K22, D386, R27, D237, R121, Q166, R347, K343, E344 and H387 (Berti and Chindemi, 2009). Interestingly, the highly conserved arginine-R100 (*E.coli* K12- P0A6D3), which was shown to confer a 99% drop in activity in an R100K mutant and no detectable activity in an R100A mutation (Shuttleworth *et al.*, 1999), was not found to be conserved within MycA4. This might suggest that despite containing much of the enzymatic machinery required for EPSP synthase activity, it may function via a different mechanism. In a previous investigation it was found that a G96S mutation of the conserved sequence LGNA$\underline{G}^{96}$ in an EPSP synthase from petunia conferred hydrolase activity. Here, the mutant was unable to bind to PEP or function as an EPSP synthase and instead dehydrated EPSP to 3-phosphoshikimate (Padgette *et al.*, 1991). The corresponding motif in MycA4 is AGSS$\underline{G}^{93}$, which does not contain an equivalent G93S mutation, does contain two preceding serine residues that may confer a similar altered function. This AGSSG motif is not observed in any reviewed EPSP synthase (1,365 entries) within the uniprot database. However, it is found within the EPSP synthase domains of MycA4 homologs. With the observation that a single G96S mutation can abolish EPSP synthase activity and confer hydrolase activity, combined with its proposed mechanism, the assigned EPSP synthase domain of MycA4 was considered to function as a hydrolase/ phosphatase to catalyse to two proposed steps (Figure 3.19)

When purified MycA4 was screened with shikimate and 3-phosphoshikimate no detectable activity was observed despite a range of conditions being assessed. The only example in the literature of an MycA4 homolog showing recombinant activity was when the entire CHC CoA operon from *S. collinus* was cloned into an *E.coli-streptomyces* shuttle vector (pSE34) and constitutively expressed in *S. lividans* 1326 (Cropp, Wilson and Reynolds, 2000)**.** Here, the specific MycA4 catalysed activity was not assessed and so the substrates and active conditions were not determined. One possible explanation for the lack of activity found with MycA4 may be its

overexpression. Despite achieving soluble expression of MycA4 with the use of chaperones, it may not necessarily be folded correctly, and instead be in an inactive form. If this were indeed the case, then even if correct assay conditions were used, no product formation would be observed. With this understanding, rather than assessing a plethora of additional assay conditions, future investigations should endeavour to recombinantly express the individual m*ycA4* gene within a *Streptomyces* host, such as *S. lividans* 1326. Clarified lysate containing MycA4, could then be screened, using the same assay conditions used in this investigation, to assess its activity. As a homolog has been viably recombinantly expressed and shown to be active within this organism, an inactive enzyme could be ruled out as a causes of inactivity. Furthermore, if activity were to be found, through the consumption of CoA, then the inactivity observed within the current investigation could be attributed to improper folding of the recombinant protein.

### 3.11.4. Soluble overexpression of CHC CoA operon using chaperone proteins

When the four CHC CoA operon genes, were recombinantly expressed from pQR2786, in *E. coli* BL21 (DE3), there was found to be putative expression of only MycA4, MycA3 and MycA1. Co-transformation with five chaperone plasmids resulted in two strains that putatively were able to express all four of the CHC CoA enzymes 1- pQR2786 + pGro7 and 2- pQR2786 + pKJE7 (Figure 3.6). Therefore, these two strains represent good candidates for future work focused on developing the CHC CoA pathway section of the cyclohexanol pathway. However, one cannot be completely certain that each enzyme has been recombinantly expressed solely based on the SDS gel shown. Assay data for each enzyme expressed within these strains will be required to confirm its expression. As this investigation has characterised the reductase, MycA2, future work should screen MycA2 activity within the clarified lysate of the two mentioned candidate strains to confirm its expression. Unfortunately, due to the lack of commercial availability of the CoA

ester substrates for MycA1 and MycA3, assaying them will not be viable in the short-term. One possible alternative method to assess the expression of each enzyme may be to synthesise or re-clone the entire CHC CoA operon but add a His-Tag within the ORF of each gene. The modified CHC CoA operon could then be co-expressed with each of the chaperone plasmids and then purified using a nickel column. This would conclusively determine whether each of the four enzymes has been expressed, by-passing the requirement of assaying each enzyme.

### 3.11.5. MycA2 activity

The results shown here have determined MycA2 to be an active acyl CoA reductase that, like ChcA from *S. collinus,* is able to use CH-1-eneC CoA and 3-HCH-1-eneC CoA as substrates (Figure 3.29). Its third substrate 3, 4-dihydroxycyclohex-1-eneC CoA was not commercially available and so could not be screened for activity. However, further development and characterisation of the biosynthetic system as described in chapter 6, will

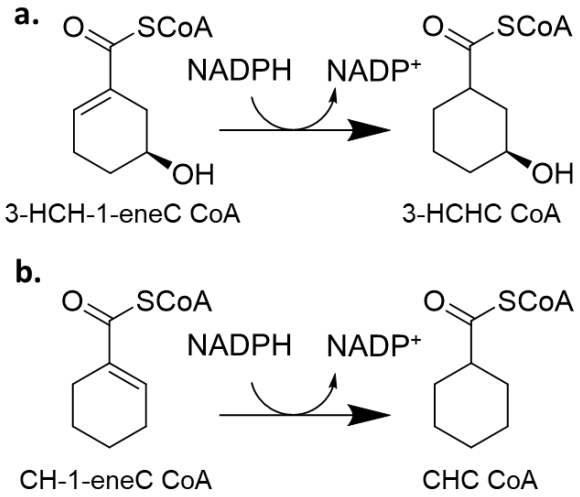



**Figure 3.29. The two MycA2 functions determined within this investigation.**

The two MycA2 functions determined in this chapter were: (a) The reduction of 3-hydroxycyclohex-1-enecarbonyl CoA (3-HCH-1-eneC CoA) to 3-hydroxycyclohexanecarbonyl CoA (3-HCHC CoA) and (b) the reduction of cyclohex-1-enecarbonyl CoA (CH-1-eneC CoA) to cyclohexanecarbonyl CoA (CHC CoA).

hopefully provide a route to the production of 3, 4-dihydroxycyclohex-1-eneC CoA for MycA2 activity assessment. The work shown here also showed that MycA2 is not inhibited by concentrations of NADPH up to 0.7 mM, which is important to know for when MycA2 is ultimately used within the cyclohexanol pathway and redox cofactors need to be optimally balanced.

As mentioned at the beginning of this chapter, the penultimate step within the CHC CoA pathway is an isomerisation of CH-2-eneC CoA to CH-1-eneC CoA. This step has been shown to be catalysed by ChcB from *S. collinus*, however, the gene encoding this enzyme is not located near the CHC CoA operon. A ChcB homolog was also found within the genome of *S. rishiriensis* sequenced in the investigation. Interestingly, in a previous investigation focused on mining the transcriptome of the CHC CoA operon-containing *S. flaveolus,* it was proposed this penultimate step was catalysed by an MycA2 homolog (Qu, Lei and Liu, 2011). With this in mind future work should screen MycA2 with CH-2-eneC CoA to conclusively determine whether ChcB activity is required within the CHC CoA pathway. If activity is not found with MycA2 then the ChcB homolog from *S. rishiriensis* will also need to be cloned and combined with the CHC CoA operon to facilitate production of CHC CoA from central metabolism.

### 3.11.6. Concluding remarks

In the work presented within this chapter, a CHC CoA operon has been identified within the newly sequenced genome of *S.rishiriensis*, which was predicted to be present by the associated production of CHC-containing polyketides by this organism. This operon was recombinantly expressed within *E. coli* BL21 (DE3) and, facilitated by co-expression of chaperone proteins, soluble expression of all four encoded enzymes was putatively found. Furthermore, each gene was individually cloned into pET28a and expressed within *E. coli* BL21 (DE3), with three of the four enzymes being soluble and expressed and purified (MycA2, MycA3 & MycA4). Despite a range of

attempted assay conditions no activity was found with MycA4, which may be due to the recombinant overexpression leading to an inactive enzyme. MycA2 was shown to be an active acyl-CoA reductase, with a comparable substrate preference to the previously described homolog from *S. collinus*, ChcA. A major limitation to screening the enzymes within this pathway was the lack of commercially available CoA ester substrates. However, this was curbed by the biosynthetic system developed in chapter 6 facilitating the production of MycA2 substrates for screening. Therefore, this work provides a strong foundation for future investigations on characterising the CHC CoA pathway and construction of this section of the biosynthetic cyclohexanol pathway.

The next chapter describes studies into the enzymatic mechanism through which CHC CoA is converted into 2-hydroxycyclohexanecarbonyl CoA (2-HCHC CoA). Ultimately, the CHC CoA required for this conversion will be derived from central metabolism (shikimate or 3-phosphoshikimate) via the CHC CoA pathway enzymes investigated in this chapter. Therefore, rather than attachment to a polyketide, CHC CoA will be rerouted towards the production of 2-HCHC CoA, which is two proposed enzymatic steps for the desired product of the biosynthetic cyclohexanol pathway, cyclohexanol.

# 4. Chapter 4. Linking CHC CoA with the existing nylon 6 synthetic pathway

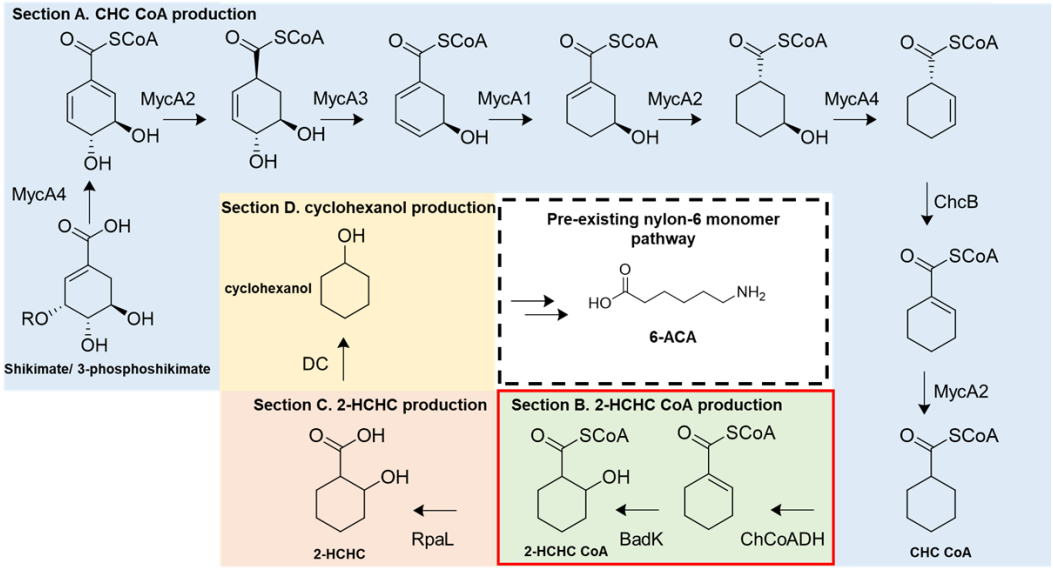



**Figure 4.1. The section of the cyclohexanol pathway assessed within chapter 4.**

The biocatalyic steps to be assessed within chapter 4 (outlined in red) are the oxidation of cyclohexanecarbonyl CoA (CHC CoA) to cyclohex-1-enecarbonyl CoA (CH-1-eneC CoA) and the hydration of CH-1-eneC CoA to 2-hydroxycyclohexanecarbonyl CoA (2-HCHC CoA).

Following the production of CHC CoA from central metabolism by the CHC CoA pathway from *S. rishiriensis*, Chapter 4 describes the studies performed to functionalise the CoA activated CHC with either a hydroxyl (2-HCHC CoA) (Figure 1) or ketone group (2-oxoCHC CoA) at C-2, using components of the anaerobic CHC degradation pathway from *Rhodopseudomonas pseudopalustris*. Three enzymes were required for these biotransformations: an acyl-CoA dehydrogenase (BadJ), a cyclohex-1-ene-1-carbonyl CoA hydratase (BadK) and a 2-HCHC CoA dehydrogenase (BadH). These enzymes are encoded within a five-gene operon (*aliA*, *badJ*, b*adK*, *badI* and *badH*) that together form the anaerobic CHC degradation pathway, which converges into the anaerobic benzoate degradation pathway (chapter 1-Figure 1.9). Once synthesised, 2-HCHC CoA and 2-oxoCHC CoA would represent two possible substrates for

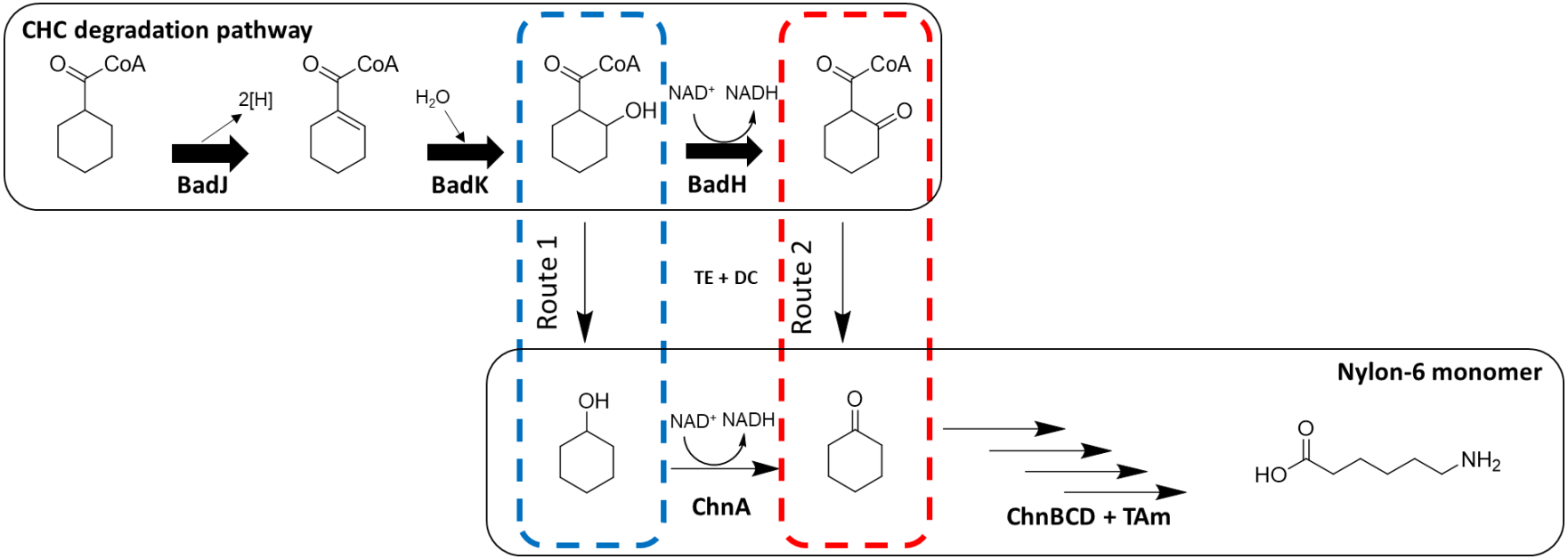



**Figure 4.2. The proposed links between CHC degradation and the 6-ACA pathway.**

Shown is the section of the CHC degradation pathway within *R. psuedopalustris* that is catalysed by BadJKH (top black box), the pre-existing synthetic nylon 6 pathway at UCL (bottom black box) and the two possible routes to link these pathways. Both route 1 and 2 share a common requirement of thioesterase (TE) and decarboxylase (DC) activity, however, they differ in their entry compound and exit compound into the nylon 6 pathway. Route 1 (blue dashed box) uses BadJ and BadK activity to produce 2-HCHC CoA (entry compound), which following TE and DC activity will derive cyclohexanol (exit compound), the substrate of the existing nylon 6 pathway. Route 2 (red dashed box) uses BadJ, BadK and BadH activity to produce 2-oxoCHC CoA (entry compound), which following TE ad DC activity will derive cyclohexanone (exit compound), the second compound in the existing nylon 6 pathway.

homologous routes, to the pre-existing nylon 6 pathway (Figure 4.2). This homologous route would employ sequential TE and DC (decarboxylase) activity to produce either cyclohexanol or cyclohexanone depending on which substrate was used.

When originally designing the synthetic route between CHC CoA and the nylon 6 monomer, the primary focus was producing cyclohexanol- the first compound in the nylon 6 pathway. This could theoretically be achieved through the combined activity of BadJ and BadK, hydroxylating CHC CoA to 2-HCHC CoA, which then, through TE and DC activity, would form cyclohexanol (route 1). However, after considering the CHC degradation pathway and its intermediates, a second route to the nylon 6 pathway became clear. The second route, like with route one, utilises BadJ and BadK activity, however, here their activity could be coupled with BadH, producing 2-oxoCHC CoA. Then, following TE and DC activity cyclohexanone would be produced, the second compound in the nylon 6 pathway (route 2). A previous publication found BadH from *R. palustris* to be inhibited by its product, furthermore, BadJ and BadK have not been previously characterised. Therefore, the rationale for producing both 2-HCHC CoA and 2-oxoCHC CoA was to determine which compound was the most viable to produce and then screen the TE activity (see chapter 5).

## 4.1. Cloning *badJ*, *badK* and *badH* into a synthetic operon

Standard restriction-ligation cloning was the chosen cloning strategy for constructing two synthetic operons containing either, *badJ* and *badK* (*badJK*) (route 1), or, *badJ, badK* and *badH* (*badJKH*) (route 2). PCR was used to amplify each gene directly from the genome of *R. pseudopalustris*. Here, primers were designed to introduce flanking restriction sites: *badJ* (*Nde*I and *Bam*HI), *badK* (*Bam*HI and *Hin*dIII) and *badH* (*Hin*dIII and *Xho*I), and a synthetic ribosome binding site (RBS) six base pairs upstream of *badK* and *badH*. The use of these restriction sites allowed these three genes to be cloned into the two synthetic operons, *badJKH* and *badJK* (Figure 4.4), as well as individually cloned into pET28a for purification and activity analysis. For the construction of *badJKH*, the PCR product of each gene (Figure 4.3) was digested with its corresponding synthetic restriction sites, as well as pET29a which was digested using *Nde*I and *Xho*I. All of these digested fragments

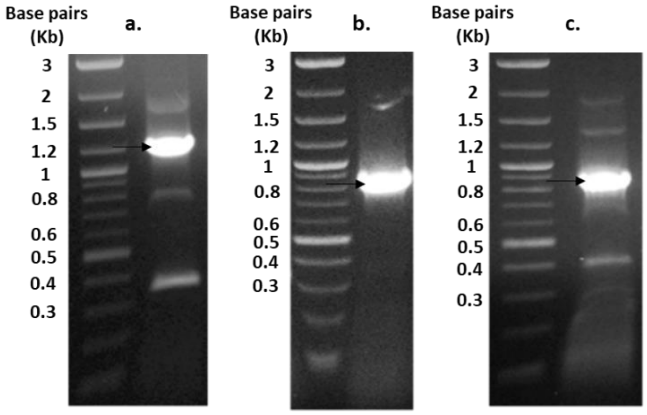



**Figure 4.3. PCR amplification of *badJ*, *badK*, and *badH*.**

An agarose gel showing amplified PCR products of (a) *badJ*- 1184 bp (b) *badK* – 821 bp and (c) *badH* – 812 bp (black arrows). Primers and conditions used in each PCR are described previously (chapter 2.6). 20 µl of a PCR was loaded onto a gel for each gene (chapter 2.18).

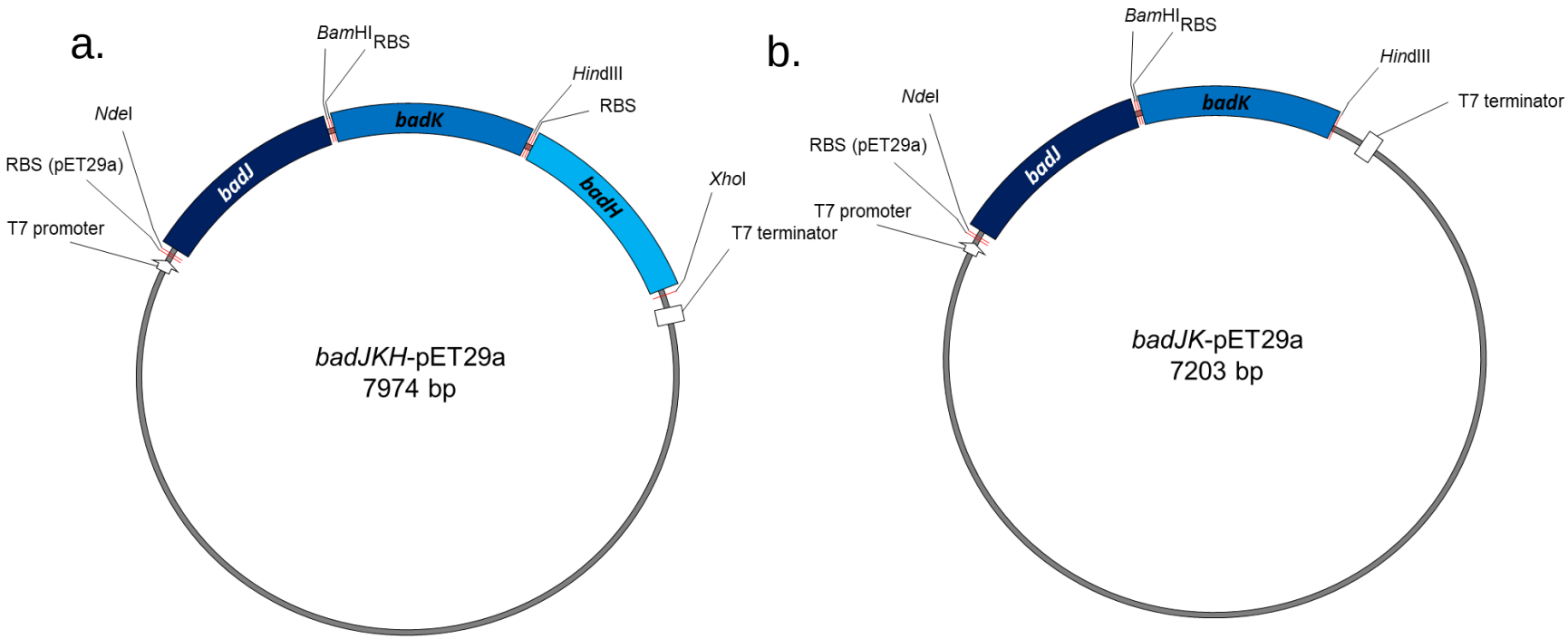



**Figure 4.4. A plasmid map of *badJKH* (pQR2791) and *badJK*.**

The synthetic (a) *badJKH* (pQR2791) and (b*) badJK* operons in pET29a, showing key genetic features such as: flanking restriction sites of each gene- *badJ* (*Nde*I and *Bam*HI), *badK* (*Bam*HI and *Hin*dIII) and *badH* (*Hin*dIII and *Xho*I), the T7 promoter, the ribosome binding sites (RBS) and the T7 terminator.

were then ligated in a single reaction, resulting in the synthetic *badJKH* operon (for nucleotide sequence see appendix 8.10). *E. coli* BL21 (DE3) cells were transformed with the *badJKH* plasmid (pQR2791) for assessment of IPTG-induced expression of each gene in the synthetic plasmid. Transformed BL21 (DE3) cells grown at 25 °C following IPTG induction were found to express soluble BadJ, the first gene in the cluster, however no soluble or insoluble expression of BadK or BadH was found (Figure. 4.5).

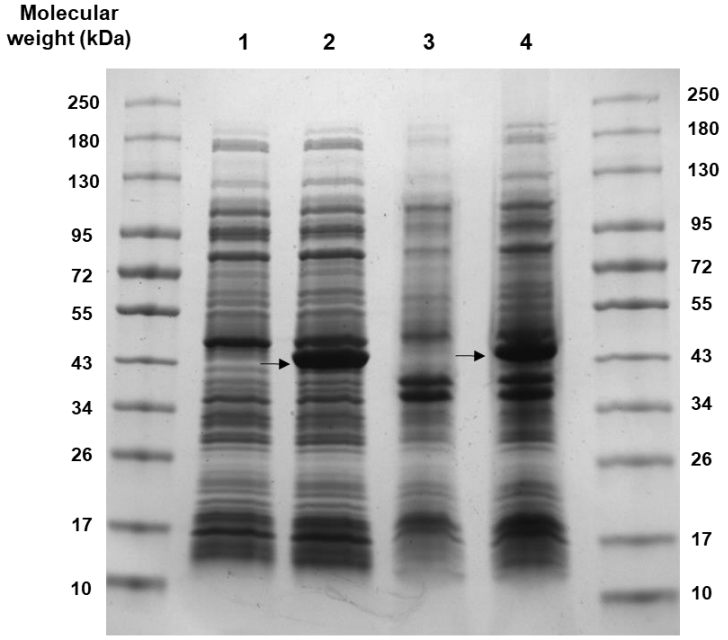



**Figure 4.5. Analysis of soluble and insoluble BadJ, BadK and BadH expression from pQR2791.**

An SDS gel showing soluble and insoluble protein obtained from *E. coli* BL21 cells transformed with either pQR2791, encoding *badJKH*, or, an unmodified pET29a plasmid (negative control). (1) Soluble lysate from the negative control, (2) soluble lysate from pQR2791, (3) insoluble protein from the negative control, (4) insoluble protein from pQR2791. The results show (2) a soluble protein at the approximate molecular weight of BadJ (41.4 kDa), that is not found in the negative control (black arrow). (4) The protein band putatively corresponding to BadJ is also found within the pQR2791 insoluble fraction (black arrow). However, no soluble or insoluble protein bands corresponding to BadK (27.9 kDa) or BadH (26.1 kDa) can be found in either the soluble or insoluble fractions of pQR2791. Following expression, Bugbuster was used for cell lysis (see chapter 2.10) and 150 µg of clarified lysate was prepared and loaded into each well in a total volume of 20 µl (see chapter 2.17). Cultures were induced with 0.1 mM IPTG and expressed at 25 °C for 24 hours in 50 ml TB media, shaking at 250 rpm.

As BadK and BadH were not found within either the soluble or insoluble protein fractions it was thought that the problem may be an issue with the gene design, such as the synthetic ribosome binding sites or the use of GTG start codons, preventing translation altogether. Therefore, rather than constructing *badJK* each gene was individually cloned into pET28a to elucidate the gene design and culture conditions that would facilitate the soluble expression of their corresponding polypeptides.

### 4.2. Cloning *badJ*, *badK* and *badH* individually into pET28a

The three genes: *badJ*, *badK* and *badH* were each digested from the *badJKH* operon, using their flanking restriction sites and individually sub-cloned into pET28a (*badJ*- pQR2792, *badK*- pQR2794 and *badH*- pQR2796). These three recombinant pET28a plasmids were then used to transform *E. coli* BL21 (DE3) cells for protein expression (Figure 4.6). Expression results showed soluble expression of BadH after 24 hours of IPTG-induced expression at 25 °C, however, both BadJ and BadK were found to be insoluble (Figure 4.6). As BadJ had been expressed partially within the soluble protein fraction from the *badJKH* operon (pQR2791), two factors were proposed to be leading to protein insolubility. Firstly, *badJ* contained both an ATG start codon from the *Nde*I restriction followed immediately by its native GTG start codon and secondly, due to the incorporation of 20 N-terminal codons encoding a His-tag and thrombin site (see appendix 8.11). GTG was also the start codon for *badK* and due to the choice of N-terminal restriction site, there were 38 codons between the start codon of *badK* and the pET28a start codon upstream of the N-terminal His-tag (see appendix 8.15). Primers were designed to replace the start codon in *badJ* and *badK* from a GTG to an ATG. Additionally, the primers used with *badK* were also used to switch the N-terminal restriction site in *badK* from a *Bam*HI to an *Nde*I, removing 18 codons from between the N-terminal His-tag and the gene start codon that had been added due to the position of the *Bam*HI restriction site in pET28a. Using their respective restriction sites, these new PCR products were then cloned into pET28a and used to transform *E.coli* BL21 (DE3) cells for protein expression. The modified genes and

resultant plasmids were named *badJ*$_{ATG}$ (pQR2793) (see appendix 8.13) and *badK*$_{ATG}$ (pQR2795) (see appendix 8.17), respectively.

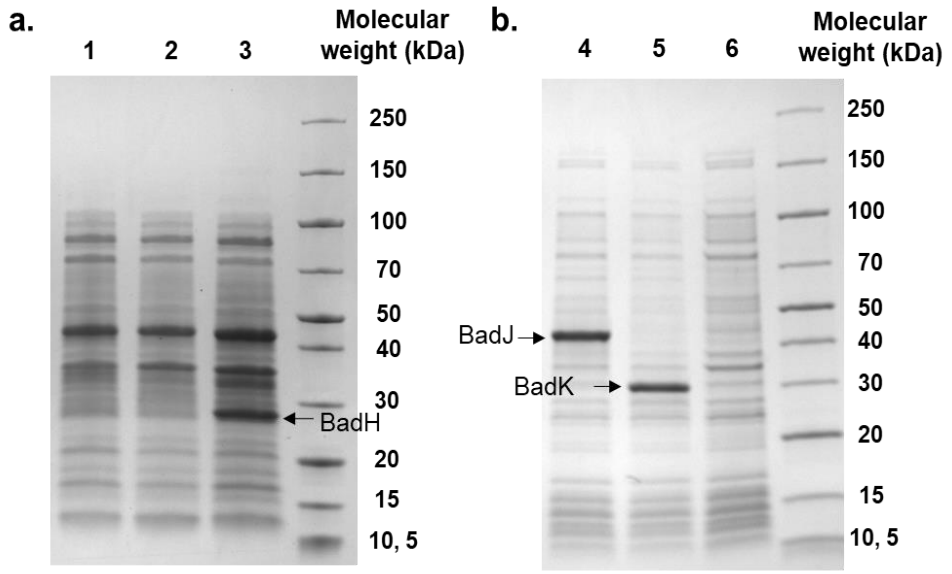



**Figure 4.6. Analysis of soluble and insoluble BadJ (pQR2792), BadK (pQR2794) and BadH (pQR2796) expression.**

An SDS gel showing the protein content within the (a) clarified lysate and (b) insoluble pellet of *E. coli* BL21 cells transformed with: (1&4) pQR2792, encoding *badJ*, (2&5) pQR2794, encoding *badK* and (3&6) pQR2796, encoding *badH*. (1) No clear soluble expression of BadJ (43.5 kDa) was observed, however, a native cellular protein of a similar molecular weight was found in all soluble protein samples. (4) An insoluble protein putatively corresponding to BadJ was observed that was not present within any other insoluble pellet (black arrow). (2) BadK (32 kDa) showed no clear soluble expression, however, (5) protein corresponding to BadK was observed within the insoluble pellet (black arrow). (3) A soluble protein corresponding to BadH (26.5 kDa) was found (black arrow). (6) No protein corresponding to BadH was observed within the insoluble pellet. Following expression, each culture was adjusted to an $OD_{600}$ of 1.5 in 1ml. Bugbuster was then used for cell lysis (see chapter 2.10) and 5 µl of each clarified lysate was loaded on to an SDS gel (see chapter 2.17). Insoluble pellets were re-suspended in an equal volume of 8 M urea as of Bugbuster used to re-suspend the 1 ml cell pellets (see chapter 2.12). 5 µl of each insoluble protein fraction was then loaded on to the SDS gel (see chapter 2.17). Cultures were induced with 1 mM IPTG and grown at 25 °C for 24 hours, shaking at 250 rpm in 50 ml TB media (see chapter 2.9).

An SDS gel analysis of *E. coli* BL21 (DE3) cells expressing *badJ*$_{ATG}$ from pQR2793 showed there to be no clear soluble expression of BadJ. The insoluble pellet contained protein at the expected molecular weight of BadJ, suggesting it to be insoluble (Figure 4.7).

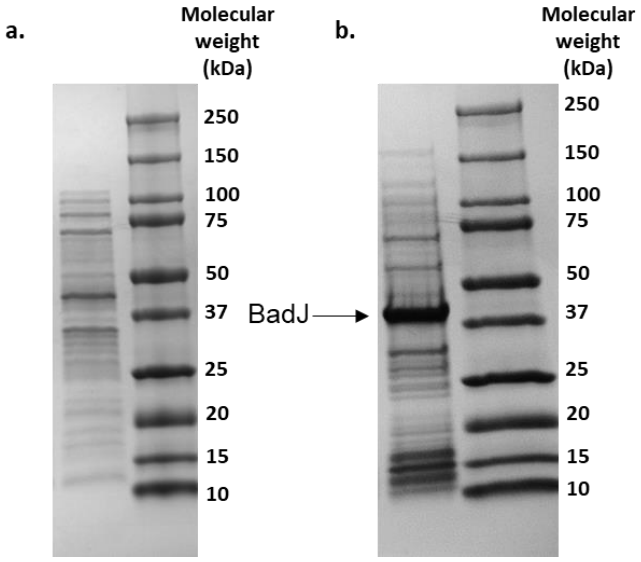



**Figure 4.7. Analysis of soluble and insoluble BadJ (pQR2793) expression.**

An SDS gel showing the protein content within the (a) soluble and (b) insoluble fractions of *E. coli* BL21 (DE3) cells expressing pQR2793, encoding *badJ*$_{ATG}$. (a) No clear overexpression of soluble BadJ (43.5 kDa) could be determined. (b) The insoluble protein fraction contained a large ~40-45 kDa protein, likely to be BadJ (black arrow). Following expression, the culture was adjusted to an OD600 of 0.5 in1ml. Bugbuster was used for cell lysis (see chapter 2.10) and 10 µl of clarified lysate was loaded on to the SDS gel (see chapter 2.17). The insoluble pellet was re-suspended in an equal volume of 8 M urea as of Bugbuster used to re-suspend the initial 1 ml cell pellet (see chapter 2.12). 10 µl of the insoluble protein fraction was then loaded on to the SDS gel. Cultures were induced with 0.1 mM IPTG and then grown for 24 hours at 25 °C at 250 rpm in 50 ml TB media (see chapter 2.9)

A similar insoluble protein band had not been found in a previous negative control, expressed in the same conditions (Figure 4.5, lane 3), further suggesting this protein band to be insoluble BadJ. A small ~47 kDa soluble protein band was observed, however, this was thought to be a protein endogenous to *E. coli*, as a similar protein band was previously found in a negative control expressed in the same conditions (Figure 4.5, lane 1). To

determine whether trace amounts of soluble BadJ, undetectable through SDS gel analysis, were being produced, the clarified lysate of a large culture (300 ml) expressing the *badJ*$_{ATG}$ gene from pQR2793, was run through a nickel column and assessed for purified BadJ. The results showed a large band of purified protein corresponding the molecular weight of BadJ, suggesting it had been successfully purified (Figure 4.8).

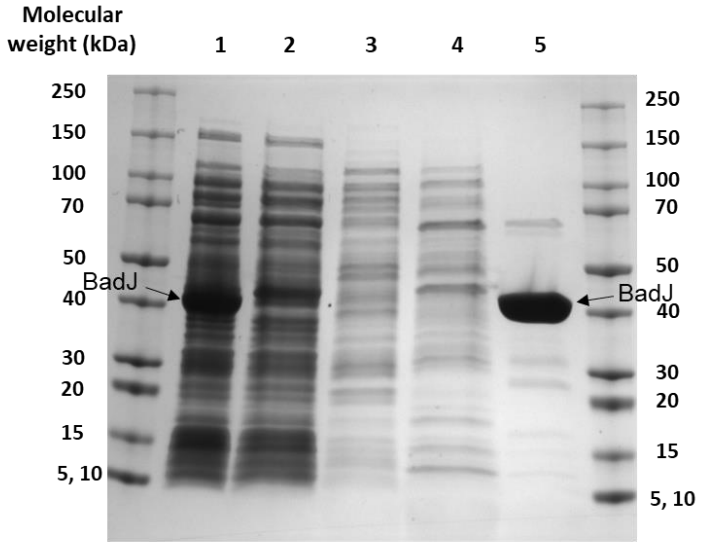



**Figure 4.8. The His-tag purification of BadJ (pQR2793).**

An SDS gel showing the soluble protein content of *E. coli* BL21 (DE3) cells following expression of *badJ*$_{ATG}$ from pQR2793, before and after elution from a nickel column. (1) Soluble protein (clarified lysate) prior to loading onto nickel column. (2) Soluble protein content eluted following a 10 mM imidazole wash. (3) Soluble protein within the eluate following a 50 mM imidazole wash. (4) Soluble protein eluted following a 100 mM imidazole wash. (5) Soluble protein eluted following a 500 mM imidazole elution wash. The results showed a purified protein corresponding to BadJ (43.5 kDa) within the 500 mM imidazole eluate. A large protein band putatively corresponding to BadJ was also found within the clarified lysate (lane 1). The expression culture was lysed using sonication (see chapter 2.11) and then purified (see chapter 2.14). 5 µl of each collected eluate and clarified lysate was prepared and loaded on to the SDS gel (see chapter 2.17). The culture was induced with 0.1 mM IPTG, and then grown for 24 hours at 25 °C in 300 ml TB media, shaking at 250 rpm (see chapter 2.9).

Furthermore, the clarified lysate was found to contain a large protein band, most likely containing both BadJ and the ~45 kDa protein endogenous to *E. coli*, which was not found when previously expressed in the same conditions at a smaller scale (50 ml). This endogenous protein was found within the first

wash fraction (10 mM imidazole) suggesting the rest of this large protein band within the clarified lysate had bound to the nickel column and was BadJ. This work was performed prior to the expression experiments in the previous chapter, which showed Bugbuster to cause insoluble expression of MycA4. Due to the volume of this culture, cells were lysed using sonication, therefore, these results combined with those of the previous chapter would suggest that Bugbuster had resulted in the observed insolubility found with BadJ. Analysis of *badK*$_{ATG}$ expression following the removal of 18 N-terminal codons and switching the GTG start codon for an ATG, showed the production of soluble BadK after 24 hours of IPTG-induced (0.1mM) expression from pQR2795 (Figure 4.9).

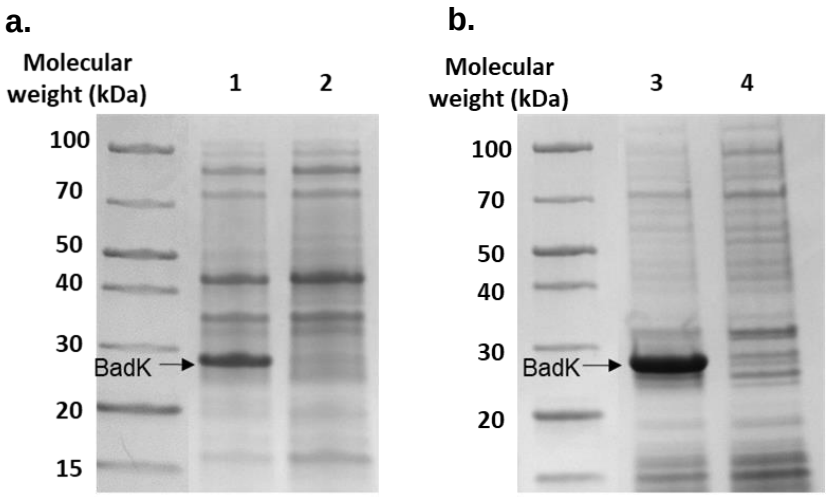



**Figure 4.9. Analysis of soluble and insoluble BadK (pQR2795) expression.**

An SDS gel showing the (a) soluble and (b) insoluble protein content of *E. coli* BL21 (DE3) cells expressing *badK*$_{ATG}$ from pQR2795, and unmodified pET28a (negative control). (1) The soluble protein of BL21 (DE3) cells expressing *badK*$_{ATG}$, a protein band corresponding to BadK was found (black arrow), which is not found in the negative control. (2) The soluble protein content of the negative control. (3) The insoluble protein content of cells expressing *badK*$_{ATG}$, a protein band corresponding to BadK can be seen (black arrow). (4) The insoluble protein content of the negative control. Following expression, the culture was adjusted to an $OD_{600}$ of 0.5 in 1ml. Bugbuster was used for cell lysis (see chapter 2.10) and 10 µl of clarified lysate was loaded on to the SDS gel (see chapter 2.17). The insoluble pellet was re-suspended in an equal volume of 8 M urea as of Bugbuster used to re-suspend the initial 1 ml cell pellet (see chapter 2.12). 10 µl of the insoluble protein fraction wa1 mM IPTG and then grown for 24 hours at 25 °C at 250 rpm in 50 ml TB media (see chapter 2.9).

Furthermore, analysis of the protein within the insoluble cell pellet also showed BadK to be present. As with *badK*, these samples were lysed using Bugbuster, suggesting that here, decreasing the number of unnecessary N-terminal codons between the N-HisTag and the gene start codon, can aid recombinant protein solubility when lysing cells with Bugbuster. BadK was then purified using the clarified lysate of a 300 ml culture lysed by sonication. The results showed a large protein band corresponding to BadK within the clarified lysate, which bound to the nickel column and was purified in the 500 mM imidazole wash (Figure 4.10).

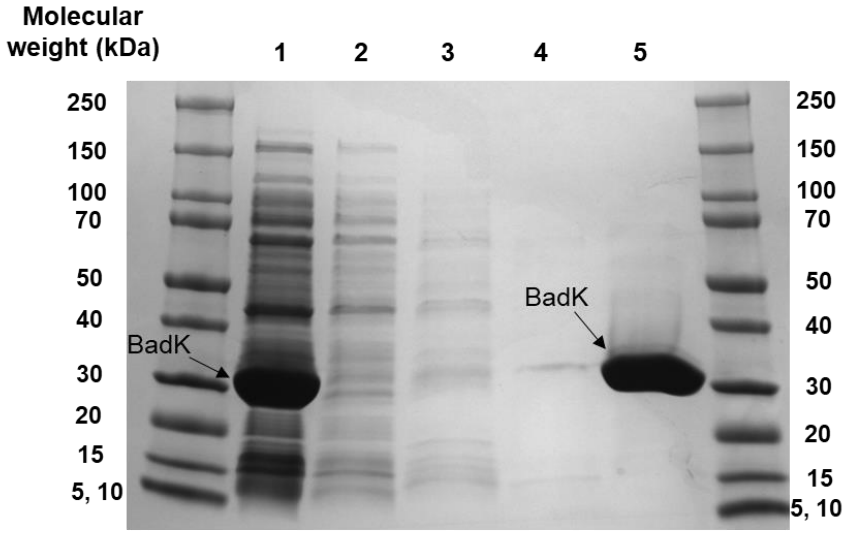



**Figure 4.10. The His-tag purification of BadK (pQR2795).**

An SDS gel showing the soluble protein content of *E. coli* BL21 (DE3) cells following expression of $badK_{ATG}$ from pQR2795, before and after elution from a nickel column. (1) Soluble protein (clarified lysate) prior to loading onto nickel column, showing a large protein band corresponding to BadK (30.1 kDa). (2) Soluble protein eluted following a 10 mM imidazole wash. (3) Soluble protein within the eluate following a 50 mM imidazole wash. (4) Soluble protein eluted following a 100 mM imidazole wash. (5) Soluble protein eluted following a 500 mM, showing a purified protein corresponding to BadK (30.1 kDa). The expression culture was lysed using sonication (see chapter 2.11) and then purified (see chapter 2.14). 5 µl of each collected eluate and clarified lysate was prepared and loaded on to the SDS gel (see chapter 2.17). The culture was induced with 0.1 mM IPTG, and then grown for 24 hours at 25 °C in 300 ml TB media, shaking at 250 rpm (see chapter 2.9).

When purifying BadH, it was found that rather than binding to the nickel column and eluting in the 500 mM imidazole fraction, as expected, it was immediately eluting in the 10 mM imidazole fraction (Figure 4.11). After reanalysing the DNA sequence a stop codon (TAG) was found at the junction between the *Hin*dIII site and the synthetic RBS (see appendix 8.19).

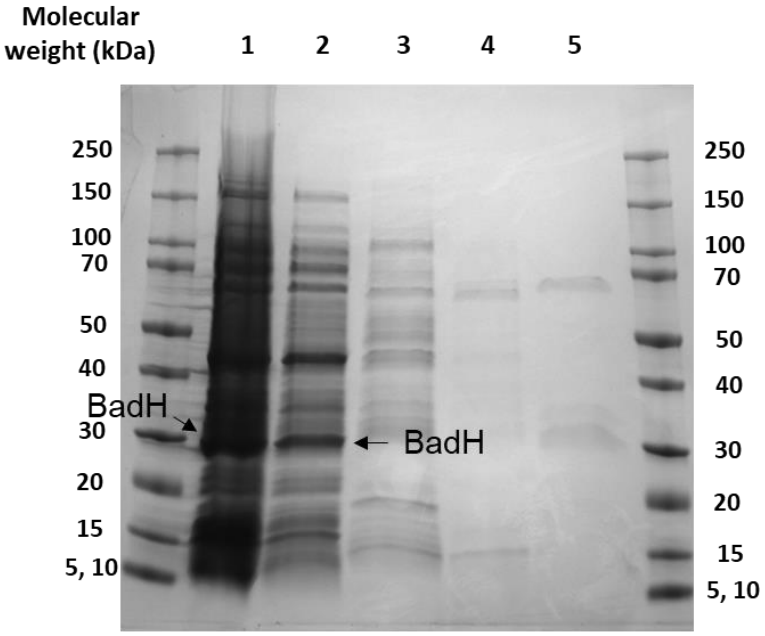



**Figure 4.11. Analysis of the His-tag purification of BadH (pQR2796).**

An SDS gel showing soluble protein of *E. coli* BL21 (DE3) transformed with pQR2796 expressing *badH,* before and after nickel column purification. (1) Clarified lysate prior to loading onto a nickel column, showing a protein band corresponding to BadH (26.1 kDa). (2) Soluble protein eluted following a 10 mM imidazole wash, a protein band putatively corresponding to BadH was observed. (3) Soluble protein within the eluate following a 50 mM imidazole wash. (4) Soluble protein eluted following a 100 mM imidazole wash. (5) Soluble protein eluted following a 500 mM, no protein corresponding to BadH was purified. The expression culture was lysed using sonication (see chapter 2.11) and then purified (see chapter 2.14). 5 µl of each collected eluate and clarified lysate was prepared and loaded on to the SDS gel (see chapter 2.17). The culture was induced with 0.1 mM IPTG, and then grown for 24 hours at 25 °C in 300 ml TB media, shaking at 250 rpm (see chapter 2.9).

As a result, *badH* may have been expressed from the synthetic RBS (AGGAGA) added through PCR, which meant there would be no N-terminal His-Tag and lead to the immediate elution of BadH in the first 10 mM imidazole

wash. This lack of additional N-terminal codons (encoding the His-tag) prior to the gene coding region might also explain why BadH was the only protein found to initially be soluble (Figure 4.6). The *badH* forward primer was redesigned, switching the *Hin*dIII restriction site to *Nhe*I, removing the undesired stop codon, allowing an N-terminal His-tag to be incorporated within the coding region of *badH* (see appendix 8.21). The modified gene was named $badH_{NheI}$ and once cloned into pET28a the resultant plasmid was named pQR2797. *E. coli* BL21 (DE3) cells were then transformed with the pQR2797 containing $badH_{NheI}$ for protein expression. Following expression, the corresponding polypeptide was found within the clarified lysate. The clarified lysate was then loaded onto a nickel column and BadH was found to elute within the 500 mM imidazole elution fraction (Figure 4.12).

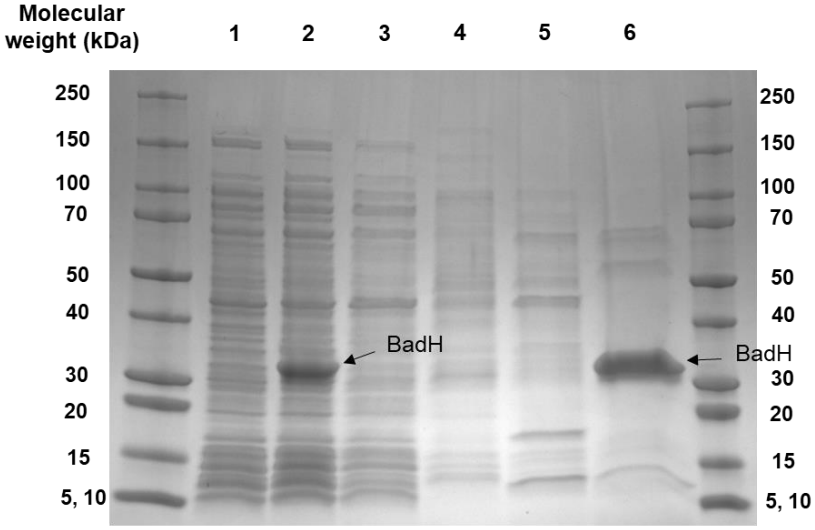



**Figure 4.12. The His-tag purification of BadH (pQR2797).**

An SDS gel showing soluble protein of *E. coli* BL21 (DE3) transformed with pQR2797 expressing $badH_{NheI}$, before and after nickel column purification. (1) Negative control-clarified lysate of cells expressing pET28a. (2) Clarified lysate prior to loading onto a nickel column, showing a protein band corresponding to BadH (29.1 kDa) (black arrow). (3) Soluble protein eluted following a 10 mM imidazole wash. (4) Soluble protein within the eluate following a 50 mM imidazole wash. (5) Soluble protein within the eluate following a100 mM imidazole wash. (6) Soluble protein eluted following a 500 mM imidazole wash, a soluble protein corresponding to BadH (29.1 kDa) was purified (black arrow). The expression culture was lysed using sonication (see chapter 2.11) and then purified (see chapter 2.14). 15 µl of each collected eluate and 2.5 µl clarified lysate were prepared and loaded on to the SDS gel (see chapter 2.17). The culture was induced with 0.1 mM IPTG, and then grown for 24 hours at 25 °C in 300 ml TB media, shaking at 250 rpm.

### 4.3. BadJ dehydrogenase activity

Following the expression and purification of BadJ it was assessed for the oxidation of CHC CoA to CH-1-eneC CoA. For this assay ferrocenium hexafluorophosphate ($Fc^+PF_6^-$) was used as an artificial electron acceptor for monitoring the assay. No significant activity was found with purified BadJ compared to the negative control (Figure 4.13). It was noted that despite being a putative FAD-dependent enzyme, the purified protein did not have a yellow colouration due to bound FAD. Therefore, prior to purification the clarified lysate containing BadJ was incubated on ice with 1 mg/ml of FAD. However, despite incubation with FAD no significant oxidase activity was found compared to the negative control containing no CHC CoA. Other factors were altered including; re-cloning BadJ with a C-terminal His-Tag, using lysate rather than purified protein, using FMN and NAD, however none of these conditions were found to confer dehydrogenase activity.

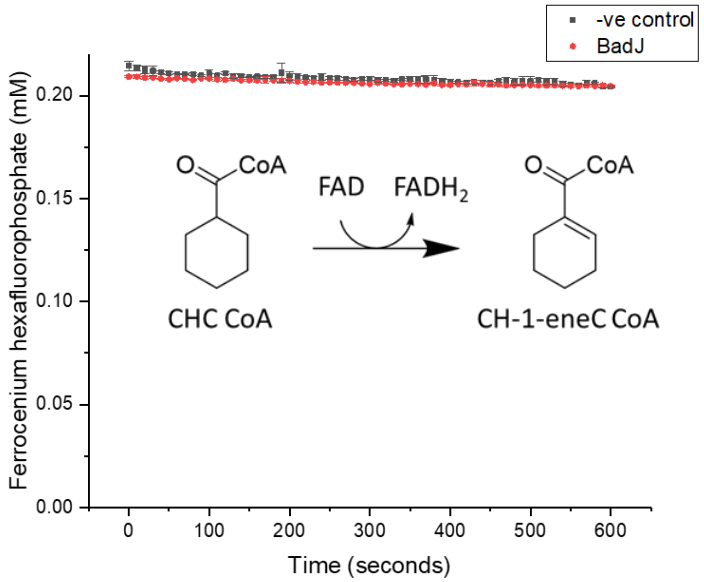



**Figure 4.13. Analysis of BadJ activity with CHC CoA.**

Assay results of purified BadJ (N-HisTag) using cyclohexanecarbonyl CoA (CHC CoA) as a substrate, compared to a negative control containing the same components but with no CHC CoA added. The absorbance at 300 nm was monitored as ferrocenium hexafluorophosphate ($Fc^+PF_6^-$) in its oxidised form absorbs strongly, however, once reduced it no longer absorbs at 300 nm. No significant drop in ferrocenium hexafluorophosphate concentration was found with BadJ compared to the negative control, indicating no dehydrogenase activity. The assay contained 0.5 mM CHC CoA, 0.2 mM $Fc^+PF_6^-$ and was performed as previously described (chapter 2.21).

## 4.4. ChCoADH, an alternative dehydrogenase from *Synthrophus aciditrophicus*

Due to the lack of activity observed with BadJ an alternative dehydrogenase was required to catalyse the oxidation of CHC CoA to CH-1-eneC CoA. A basic local alignment search tool (BLAST) was used to find a characterised dehydrogenase with homology to BadJ, however, this only revealed homology-inferred enzymes with no published characterisation studies. Therefore, a literature search was undertaken to find a dehydrogenase by its published activity irrespective of its homology to BadJ. This lead to the identification of the cyclohexanecarbonyl-CoA dehydrogenase (ChCoADH) from *Synthrophus aciditrophicus*. This dehydrogenase had been previously shown to oxidise CHC CoA to CH-1-eneC CoA (Kung *et al.*, 2013).

Prior to its synthesis, the ORF of the coding sequence of ChCoADH was codon optimised for expression in *E.coli* and a C-terminal His-Tag was added to the open reading frame (ORF). The altered nucleotide sequence was then synthesised (Genscript) and sub-cloned into the pET29a vector (see appendix 8.23). The recombinant plasmid, pQR2798, was then used to transform *E. coli* BL21 (DE3) cells and ChCoADH was shown to be expressed within the soluble fraction of cell lysate. Following soluble expression ChCoADH was His-tag purified using nickel column chromatography (Figure 4.14).

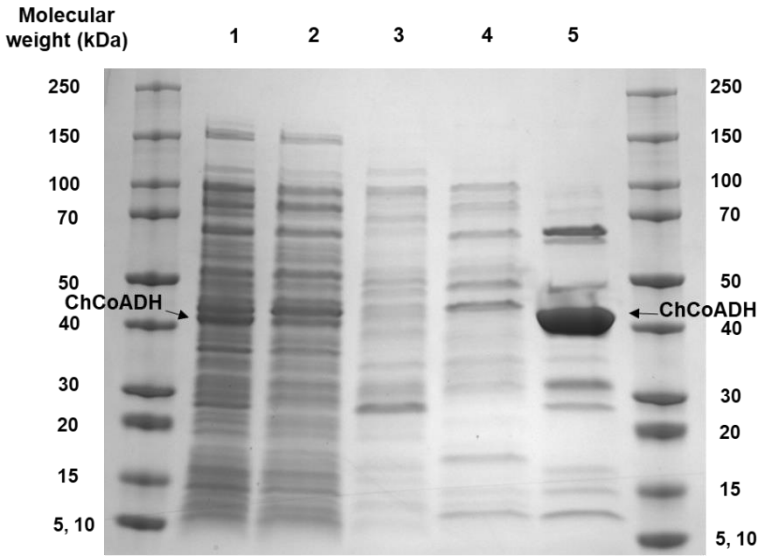



**Figure 4.14. The His-tag purification of ChCoADH (pQR2798).**

An SDS gel showing the soluble protein of *E.coli* BL21 (DE3) cells transformed with pQR2798, expressing ChCoADH, before and after purification with a nickel column. (1) Clarified lysate containing a small ~42 kDa protein band putatively corresponding to ChCoADH (black arrow). (2) Protein collected in eluate following 10 mM imidazole wash. (3) Protein collected in eluate following a 50 mM imidazole wash. (4) Protein collected in eluate following a 100 mM imidazole wash. (5) Protein collected in the eluate following a 500 mM imidazole elution wash, a large ~42 kDa protein band corresponding to ChCoADH is observed (expected size 42.9 kDa) (black arrow). The expression culture was lysed using sonication (see chapter 2.11) and then purified (see chapter 2.14). 15 µl of each collected eluate and 2.5 µl clarified lysate were prepared and loaded on to the SDS gel (see chapter 2.17). The culture was induced with 0.1 mM IPTG, and then grown for 24 hours at 25 °C in 300 ml TB media, shaking at 250 rpm (see chapter 2.9).

Purified CHCoADH was then used in the oxidative conversion of CHC CoA to CH-1-eneC CoA using the continuous $Fc^{+}PF_6^{-}$ assay. The assay results showed a significant drop in the concentration of $Fc^{+}PF_6^{-}$ compared to the negative control, indicative of the oxidation of CHC CoA to CH-1-eneC CoA (Figure 4.15).

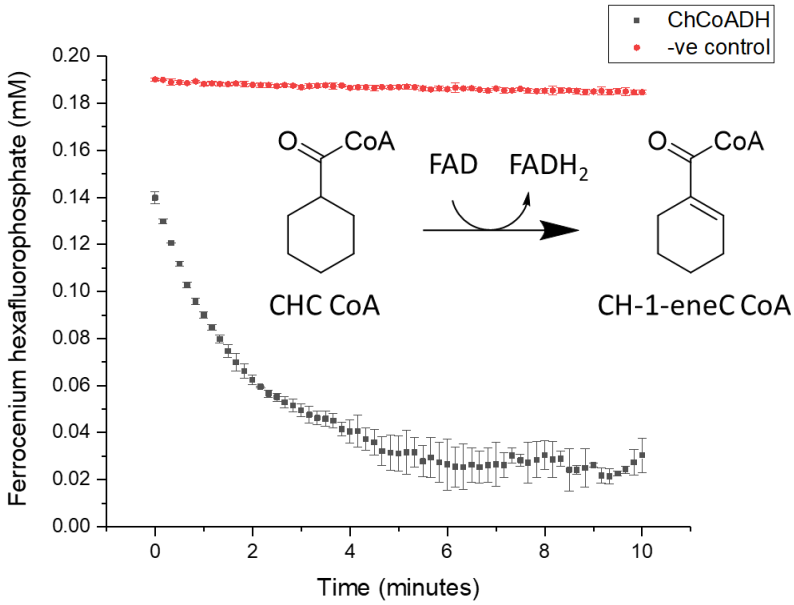



**Figure 4.15. Analysis of ChCoADH activity with CHC CoA.**

Assay results of purified ChCoADH oxidising CHC CoA to CH-1-eneC CoA and a negative control containing no CHC CoA. The assay monitored the reduction of the artificial electron acceptor, ferrocenium hexafluorophosphate into ferrocene, indicated by a drop in the absorbance at 300 nm. The assay contained 0.5 mM CHC CoA, 0.2 mM $Fc^{+}PF_6^{-}$ and was performed as previously described (chapter 2.21).

In order to confirm the production of CH-1-eneC CoA, HPLC analysis was used. The results found 90% (0.45 mM) of CHC CoA had been consumed 15 minutes after initiating the assay and a new peak with a retention time of 18.2 was observed (Figure 4.16).

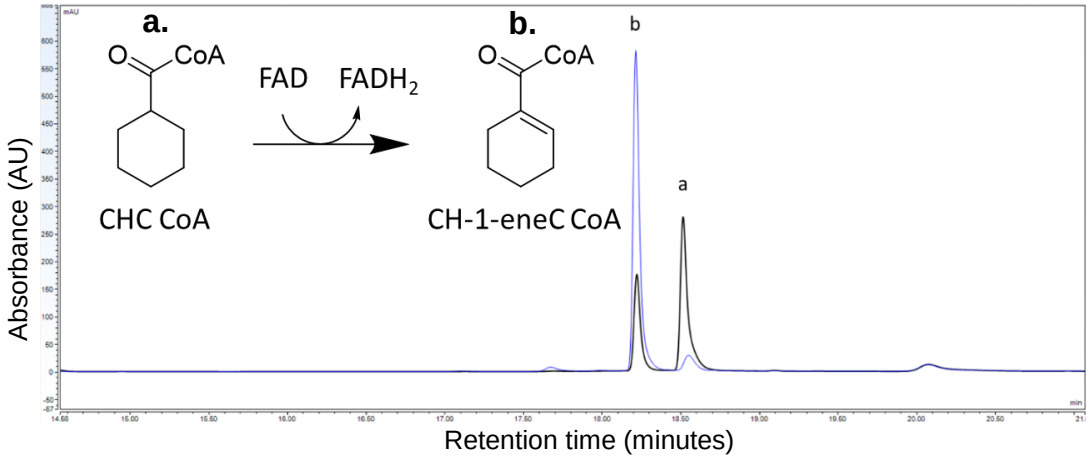



**Figure 4.16. HPLC analysis of ChCoADH activity with CHC CoA**

An HPLC trace immediately after addition of ChCoADH (black line) and 15 minutes after (blue line). (a) CHC CoA, A, (b) CH-1-eneC CoA. The results show CHC CoA being oxidised to CH-1-eneC CoA immediately after the addition of ChCoADH, with almost completely conversion after 15 minutes at 25 degrees. Assays contained 0.5 mM CHC CoA, 1mM Fc+PF6- and were performed as previously described (chapter 2.21). The HPLC method used is also described previously (chapter 2.30).

Due to there being no commercial standard of CH-1-eneC CoA it could not be concluded by the HPLC alone that this new peak was CH-1-eneC CoA. Therefore, this peak was collected from the HPLC and analysed by HRMS (ES-). Analysis of this collected peak showed it to have an *m/z* of 874.1685 (theoretical mass $[M-H]^-$ = 874.1654) , confirming its identity to be CH-1-eneC CoA and thus the successful oxidation of CHC CoA to CH-1-eneC CoA by ChCoADH had been achieved (Figure 4.17).

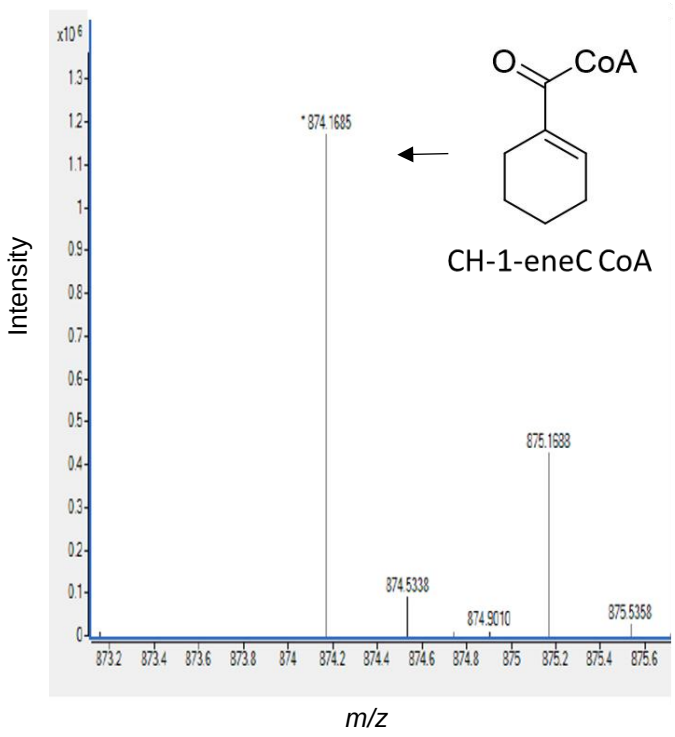



**Figure 4.17. HRMS (ES-) analysis of putative CH-1-eneC CoA.**

HRMS (ES-) results confirmed the new compound found through HPLC analysis was CH-1-eneC CoA (error <5 ppm). The HRMS was run in the negative mode to give an $[M-H]^-$ of 874.1685, representing the CH-1-eneC CoA deprotonated anion (theoretical mass $[M-H]^-$ 874.1654). The compound was purified and prepared for analysis as previously described (chapter 2.30 & 2.35).

## 4.5. BadK activity

Following confirmation of CH-1-eneC CoA production by ChCoADH, the activity of BadK could then be determined using CH-1-eneC CoA generated in situ. BadK activity was determined using HPLC to monitor a decrease in the peak area corresponding to CH-1-ene-C CoA at 18.2 minutes. Furthermore, as there is no commercial standard of 2-HCHC CoA, HPLC was also used to identify its production by searching for the formation of a new peak following BadK activity, which was not found in the negative control. The results showed a 30% decrease in the peak area corresponding to CH-1-eneC CoA, as well as a new peak forming with a retention time of 14.7 minutes (Figure 4.18). Furthermore, the CH-1-eneC CoA peak area was found to be consumed by 40% after one hour and 48% after two hours. As there is no commercial standard of CH-1-eneC CoA, the exact amount of substrate consumed could not be determined.

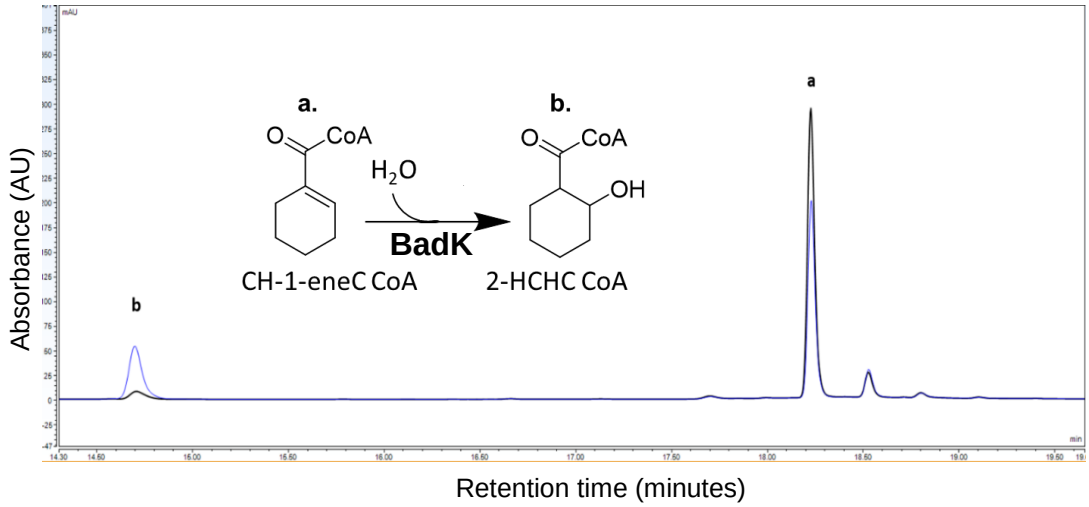



**Figure 4.18. HPLC analysis of BadK with CH-1-eneC CoA.**

An HPLC trace immediately after addition of BadK (black line) and 30 minutes after (blue line). (a) CH-1-eneC CoA, (b) 2-HCHC CoA. The results showed that after 30 minutes the peak corresponding to CH-1-eneC CoA had been consumed by 30%, accompanied by the formation of a new peak corresponding to 2-HCHC CoA. The assay was performed as previously described (chapter 2.22). The HPLC method is also previously described (chapter 2.30).

To confirm that the new peak forming following BadK activity was 2-HCHC CoA, the peak was collected from the HPLC and analysed using HRMS as before. The HRMS (ES-) found the *m/z* to be 892.1763 (theoretical $[M-H]^-$ 892.1760), determining this compound to be 2-HCHC CoA (Figure 4.19).

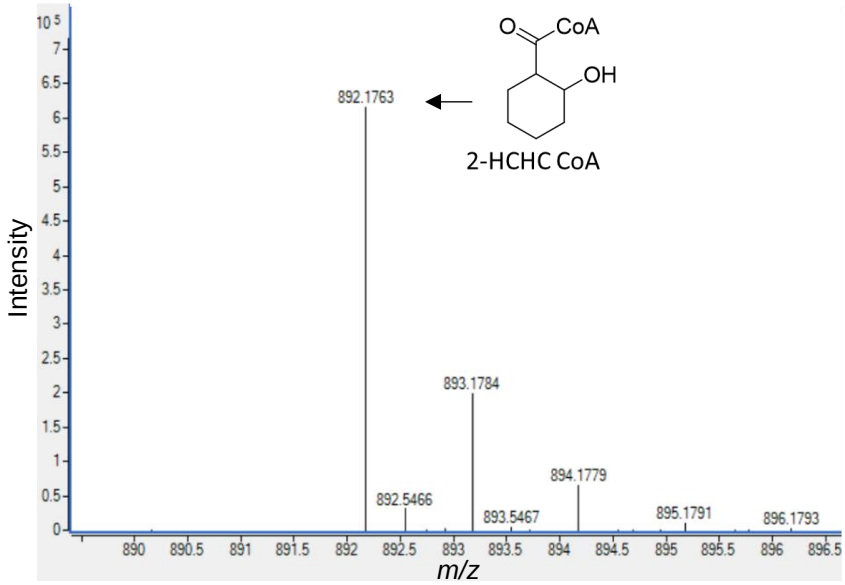



**Figure 4.19. HRMS (ES-) analysis of putative 2-HCHC CoA.**

HRMS (ES-) results confirmed the compound found following BadK activity to be 2-HCHC CoA (error <5 ppm). The HRMS was run in the negative mode to give an $[M-H]^-$ of 892.1763 representing the 2-HCHC CoA anion (theoretical mass $[M-H]^-$ = 892.1760). The compound was purified and prepared for analysis as previously described (chapter 2.30 & 2.35).

## 4.6. BadH activity

The NAD-dependent dehydrogenase, BadH, was next assessed for the oxidation of 2-HCHC CoA into 2-oxoCHC CoA (Figure 4.20). Due to there being no commercial standard of 2-HCHC CoA the activity of BadH was coupled with ChCoADH and BadK. Dehydrogenase activity was determined by monitoring the reduction of NAD to NADH at 340 nm.

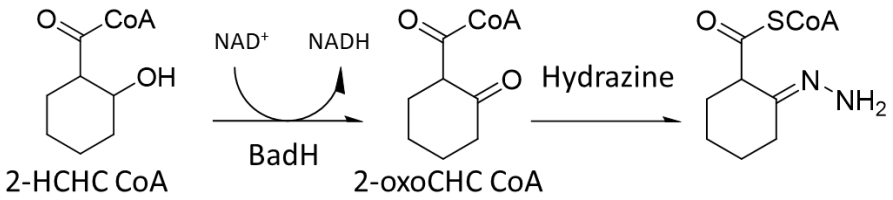



**Figure 4.20. The NAD+ dependent dehydrogenase activity of BadH.**

BadH functions by oxidising 2-hydroxycyclohexanecarbonyl CoA (2-HCHC CoA) into 2-oxocyclohexanecarbonyl CoA (2-oxoCHC CoA). Product inhibition was found with 2-oxoCHC CoA, therefore hydrazine was added to convert the ketone functional group into a hydrazone, preventing inhibition.

The results showed the formation of NADH within the first 5 minutes, however once reaching an NADH concentration of 0.03 mM the reaction was inhibited and no further production of NADH was observed throughout the 1 hour incubation (Figure 4.21- black line).

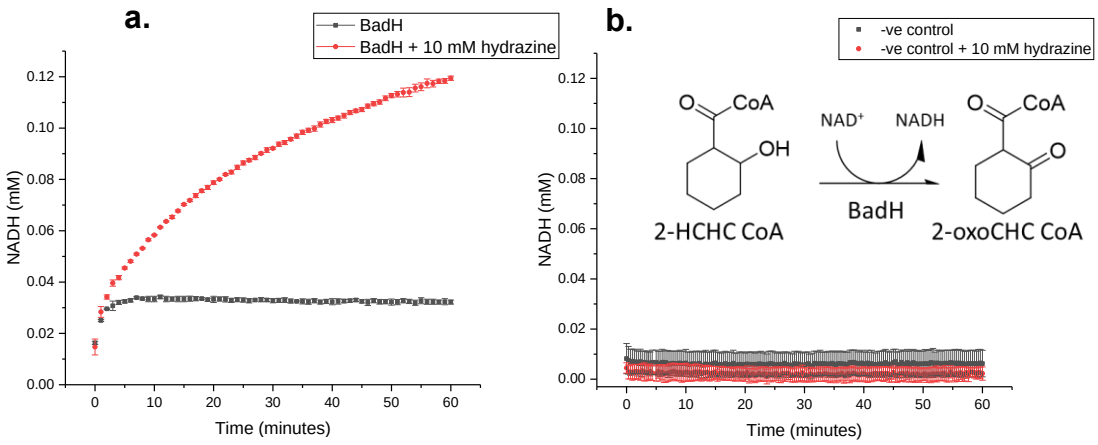



**Figure 4.21. Analysis of BadH activity with 2-HCHC CoA.**

The results of (a) purified BadH activity with and without the addition of hydrazine, using 2-HCHC CoA derived from BadK as the substrate and (b) a negative control with and without hydrazine, containing no purified BadH. Assays were performed in triplicate and performed as previously described (chapter 2.23).

Similarly, a previous publication investigating BadH from *R. palustris* showed it to also experience inhibition by its product, 2-oxoCHC CoA. However, this could be overcome with the addition either 2-oxocyclohexanecarboxyl-CoA hydrolase (BadI) to convert 2-oxoCHC CoA to pimelyl CoA, or hydrazine, irreversibly forming a hydrazone from the ketone functional group (Pelletier and Harwood, 2000). Therefore, 10 mM hydrazine hydrate was added to the assay to investigate whether this would prevent substrate inhibition. The results showed NADH formation beyond 0.03 mM and throughout the one hour incubation, with 0.11 mM of NADH being formed after one hour. This clearly suggests that with the addition of hydrazine product inhibition can be overcome to irreversibly converting the product, 2-oxoCHC CoA, into a hydrazone (Figure 4.20).

## 4.7. Discussion

### 4.7.1 Poor BadK and BadH expression from synthetic *badJKH* operon

Expression of the synthetic *badJKH* operon (pQR2791) resulted in only the first gene (*badJ*) directly downstream of the promoter being expressed into its corresponding polypeptide. Conversely, both BadK and BadH were absent from the soluble and insoluble lysate fractions. All three genes within this operon are transcribed by a single promoter into a single, polycistronic mRNA, as with many prokaryotic operons (Godbey and Godbey, 2014). As BadJ was expressed it would suggest that transcription is occuring, however, the RNA polymerase may be terminating prematurely, leading to an mRNA moelcule that does not include the coding reagions of *badK* or *badH*. It has been widely observed that RNA polymerase termination can occur prematurely, as a regulatory mechansim of gene expression (Bervoets and Charlier, 2019). Premature transcription termination can be mediated by G-C rich stem loop stuctures, causing the RNA polymerase (RNAP) to pause, then by a following string of Uracil nucleotides or by contact with Rho protein the RNAP dissociates from the DNA (Saba *et al.*, 2019). The GC content of *badJ* (69 %),

*badK* (64 %) and *badH* (67 %) is significantly higher than that of the *E. coli* genome (~50 %) (Mann and Chen, 2010), which may result in a higher abundance of intragenic G-C stem loops forming and thus increased premature transcription termination. Premature transcription termination may provide an explanation for limited expression of BadK and BadH, however, it does not explain why there is a complete absence of BadK and BadH in both the soluble and insoluble lysate fractions. This was thought to be caused by a combination of factors.

The native Shine-Dalgarno (SD) sequence of each gene was replaced with a synthetic SD seqeunce (AGGAGA) for optimal expression within *E.coli*. However, *badJ* used the RBS from pET29a that also has a pyrimidine (C and T) rich region preceeding the SD sequence (see appendix 8.10). This pyrimidine rich region has been shown to be an important factor for the initial binding of the ribosomal protein S1 to mRNA (Laursen *et al.*, 2005). It has been previously shown that removal or mutation of the pyrimidine nucleotides preceeding the SD sequence can lead to a significant drop in protein expression (Zhang and Deutscher, 1992). Therefore, the combined effect of prematurely termined mRNA molecules with the lack of a pyrimidine rich region upstream of the RBS may provide an explanation for why there was no observed soluble or insoluble expression of BadK or BadH. As this chapter determined route 1 to be the most viable route (Figure 4.2), construction of *badJK* plasmid will be the required. Therefore, new primers will need to be designed for *badK* to incorporarte the pyrimidine rich region directly upstream of the current ‘AGGAGA’ Shine Dalgarno sequence, facilitating ribosomal binding and thus translation.

### 4.7.2 Insoluble individual expression of BadJ and BadK

In the initial expression experiments of *badJ*, *badK* and *badH*, Bugbuster was routinely used for the lysis of small samples (1 ml aliquots of cultures) and it was not known that this could have a negative effect on the solubility of

corresponding recombinant proteins. Sonication was only used when an entire expression culture was to be lysed (for purification), due to there being a larger volume of resuspended cells, which the sonication probe could be submerged in. When BadJ was expressed from the synthetic *badJKH* operon (pQR2791) it was found within the soluble lysate fraction. However, when sub-cloned into pET28a (pQR2792 and pQR2792) and expressed under the same conditions it was found to be insoluble. In both cultures a 1 ml aliquot was pelleted and lysed with Bugbuster.

In the *badJKH* operon (pQR2791) the *badJ* start codon was directly downstream of the Shine-Dalgarno sequence with no additional codons between the RBS and gene (see appendix 8.10). However, when *badJ* was sub-cloned into pET28a (pQR2792) , there were 20 codons (encoding the His-tag and thrombin site) between the *badJ* start codon and the RBS (see appendix 8.11). It was thought that the prescence of both the native GTG start codon and synthetic *Nde*I ATG start codon, combined with the 20 additional N-terminal codons may be causing this insolublility. Consequently, $badJ_{ATG}$ (pQR2793) was produced, where the native GTG start codon was replaced with an ATG start codon (see appendix 8.13). Initially, there was only found to be insoluble BadJ expression, however, it was thought that running all of the soluble protein produced from a large (300 ml) culture expressing $badJ_{ATG}$ might allow any small amount of soluble BadJ, that could not be detected on an SDS gel, to be purified. BadJ was identified within the soluble lysate fraction and was purified through a Nickel column. However, the soluble lysate of the 300 ml culture clearly contained a higher proportion of BadJ (Figure 4.8) compared to the soluble lysate of the aliquot from the 50 ml culture (Figure 4.7), despite being grown and induced in the same conditions. It was initially thought this was due to the increased culture size. However, following the results found in chapter 3, where Bugbuster was shown to make MycA4 insoluble, it is thought that this may have been the cause of the insolubility found when attempting to express BadJ from *badJ* (pQR2792) and $badJ_{ATG}$ (pQR2793). Furthermore, the soluble expression of BadJ from the *badJKH* operon (lysed using Bugbuster), which does not contain an N-terminal His-tag,

suggests the addition of an N-terminal His-tag combined with the use of Bugbuster for cell lysis, leads to insoluble protein expression.

### 4.7.3. Soluble expression of BadK found by removing additional 5' codons

Initially *badK* was cloned using the *Bam*HI and *Hin*dIII restriction sites within the multiple cloning site (MCS) of pET28a (pQR2794), however, the resultant polypeptide was insoluble. The positioning of the *Bam*HI restriction site in the MCS meant there were 38 codons, encoding the His-tag, thrombin site and T7 tag sequence, between the first ATG start codon in pET28a and the GTG start codon of *badK* (see appendix 8.15). It was proposed that a contributing factor to the insolubility observed may be the long chain of N-terminal amino acids leading to improper folding of the extending polypeptide chain during translation. This theory was confirmed by soluble expression of BadK from $badK_{ATG}$ (pQR2795), where its *Bam*HI restriction site was switched with *Nde*I, removing 18 codons, leaving just the His-tag and thrombin site between the RBS and gene start codon (see appendix 8.17). Furthermore, BadK was still found in the insoluble protein fraction as well, implying that Bugbuster may still be causing insolubility. As with BadJ, when $badK_{ATG}$ (pQR2795) was expressed in a 300 ml culture and lysed by sonication, the clarified lysate showed a significantly larger composition of BadK, further confirming the negative effect lysing expression cultures with Bugbuster has on protein solubility.

With the replaced *Nde*I restriction site, $badK_{ATG}$ had the same number of codons between the first pET28a start codon and the gene start codon as $badJ_{ATG}$. However, when expressed and lysed with Bugbuster, BadK was partially soluble but BadJ was not. In a previous investigation comparing the amount of soluble protein produced using sonication with Bugbuster, a much lower yield was found with Bugbuster with larger proteins (Listwan *et al.*, 2010). The quaternary structures and sizes of BadJ and BadK are unknown,

however, a difference in molecular weight (BadJ- 43.6 kDa, BadK- 30.1 kDa) may explain why BadK was partially soluble and BadJ was not.

### 4.7.4. BadJ inactivity

The reason for the inactivity observed with BadJ was inconclusive, however analysis of the literature may give some insights. There are no previous investigations that have characterised BadJ, however, a number of studies have stated its proposed function, often referencing other studies that also have not characterised BadJ (Harwood *et al.*, 1998; Pelletier and Harwood, 2000; Hirakawa *et al.*, 2015). These investigations state that unpublished work by M.D, Emig heterologously expressed and purified flavin-containing BadJ (AliB*)* from *R. palustris* and showed it to oxidise CHC CoA into CH-1-eneC CoA. However, no publication has followed this unpublished work, so the exact conditions used to derive activity with BadJ were unknown. In these investigations BadJ was stated to be a flavin-containing enzyme, which was verified through BLAST analysis of the polypeptide chain which showed its conserved domains to match that of an FAD-dependent acyl-CoA dehydrogenase. Interestingly, in 1995 a study by Jan Kuver, which investigated the metabolism of CHC in *R. palustris,* found a dehydrogenase within the crude cell extract that could oxidise CHC CoA to CH-1-eneC CoA. In these assays either NAD or NADP was added to the cell extract and activity was monitored by the reduction of the artificial electron acceptor, ferrocenium ion (Kuver, Xu and Gibson, 1995). Furthermore, the amino acid sequence of dehydrogenase within the cell extract was not determined, so the activity found cannot be confidently linked to BadJ. Despite this investigation, based on the conserved domains it was assumed BadJ was a flavoprotein.

Flavoproteins have been found to both covalently and non-covalently bind to flavin, with approximately 10% of enzymes binding covalently (Starbird *et al.*, 2017). This covalent binding has been shown to be required for both increasing the thermodynamic driving force of the enzyme, as well as

preventing the dissociation of flavin in enzymes that are found in flavin-poor microenvironments (Starbird *et al.*, 2015). Early studies looking at cofactor binding proposed that flavinylation was an autocatalytic mechanism, however, more recent investigations have highlighted the importance of assembly proteins for the covalent binding of the flavin (McNeil and Fineran, 2013; McNeil *et al.*, 2014). In the current investigation it was observed that the purified BadJ was not yellow. Furthermore, despite incubation with both FAD and FMN pre and post purification, as well as the use of clarified lysate containing BadJ, no activity was found. This might suggest assembly factors are required for the formation of the BadJ apoenzyme, which are not present in *E. coli*. Therefore, future investigations on BadJ may require the identification and isolation of accessory proteins to facilitate the binding of FAD to BadJ.

### 4.7.5. Route 1 (2-HCHC CoA) will link CHC degradation pathway with nylon 6 pathway

Both BadK and BadH were shown to be active and able to produce 2-HCHC CoA and 2-oxoCHC, respectively. However, BadH was inhibited by its reaction product, 2-oxoCHC CoA, and so hydrazine was required to irreversibly convert the product ketone into a hydrazone allowing the reaction to run to completion. However, the formation of a hydrazone CoA ester is not useful for developing a route to the nylon 6 amino acid precursor, as there is no clear route through to the intermediate cyclohexanone. In a previous investigation, the product inhibition observed with BadH from *R. palustris* was overcome in two ways: 1- with hydrazine, as carried out in this study, 2- by introducing the next enzyme in the CHC degradation pathway, BadI to convert 2-oxoCHC CoA into pimelyl CoA. The next biochemical step in the proposed pathway will identity a TE that can hydrolyse the thioester bond in 2-HCHC CoA or 2-oxoCHC CoA. Theoretically, if a TE could be identified that had activity specifically with 2-oxoCHC CoA and none of the preceding intermediates, it could be used to flux the BadH reaction to completion. However, as 2-oxoCHC CoA is not

commercially available there is no clear method through which to screen it with a numbers of TEs. With the caveats presented here, the route to cyclohexanone via 2-oxoCHC CoA (route 2) (Figure 4.2) is currently not considered a viable route to the pre-existing nylon pathway. Therefore, the metabolic route via 2-HCHC CoA to cyclohexanol (route 1), is presented here as the most viable route to link central metabolism with the pre-existing nylon 6 pathway and so 2-HCHC CoA is the focus of TE activity screening in the following chapter.

## 4.8. Conclusion

The work presented in this chapter has shown the successful production of 2-HCHC CoA by the sequential activity of a heterologously expressed acyl CoA dehydrogenase (ChCoADH) from *S*. *Acidovorax* and BadK, a hydratase from *R. pseudopalustris*. BadJ the acyl-CoA dehydrogenase member of the CHC degradation pathway from *R. pseudopalustris* was found to be inactive when heterologously expressed in *E. coli*, however, this is possibly due to the absence of accessory proteins required for the covalent binding of FAD to the flavoprotein. BadH, a 2-HCHC dehydrogenase was shown to experience product inhibition towards 2-oxoCHC CoA, analogous to the previously characterised homolog from *R. palustris*, which could be overcome by the addition of hydrazine. However, due to this product inhibition, combined with a lack of commercial availability of 2-oxoCHC CoA, this route is not considered a viable route to the pre-existing nylon 6 pathway. Therefore, the focus of the next chapter is the route to cyclohexanol via 2-HCHC CoA, route 1. Here, a number of TEs will be screened to find one that can actively hydrolyse 2-HCHC CoA to 2-HCHC. Ideally this TE will not be active on any of the preceding intermediates encouraging the metabolic flux to be pulled through to 2-HCHC without the formation of any unwanted intermediates.

# 5. Chapter 5. Identification of a TE to hydrolyse 2-HCHC CoA into 2-HCHC

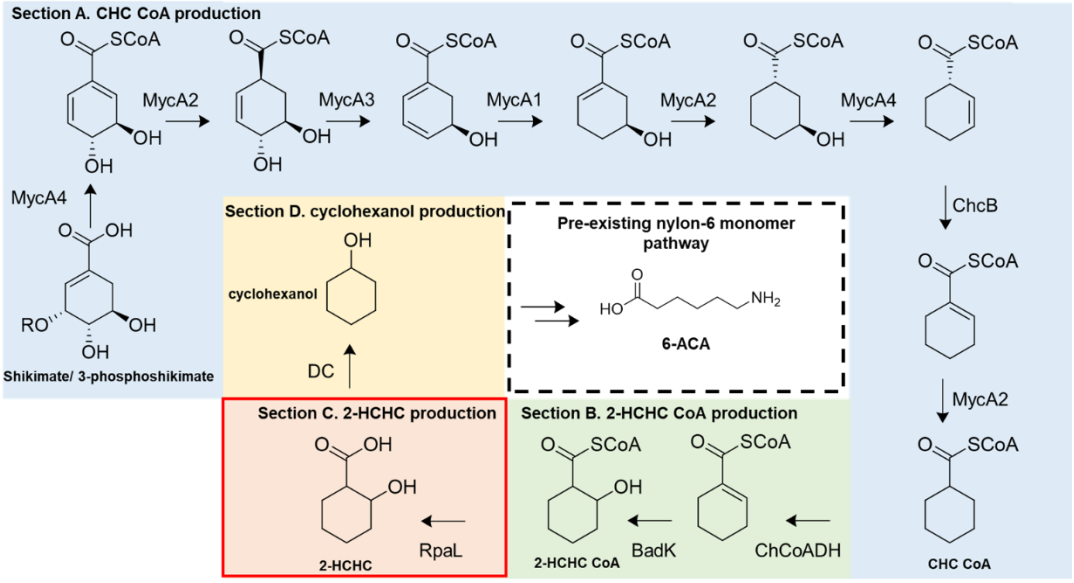



**Figure 5.1. The required TE activity within the cyclohexanol pathway.**

The section of the cyclohexanol pathway to be assessed within chapter 5 (outlined in red). The required TE activity will hydrolyse 2-hydroxycyclohexanecarbonyl CoA (2-HCHC CoA) to 2-hydroycyclohexanecarboxylic acid (2-HCHC) with the release of CoA.

Following the production of 2-hydroxycyclohexanecarbonly CoA (2-HCHC CoA) through sequential ChCoADH and BadK activity, a thioesterase (TE) was next required to hydrolyse the CoA ester, into CoA and 2-hydroxycyclohexanecarboxylic acid (2-HCHC) (Figure 5.1).

This organic acid would then be a single biochemical step (decarboxylation) away from cyclohexanol, a substrate for the previously developed nylon 6 monomer pathway. There were no identified examples of a TE that was able to catalyse this hydrolytic process in the literature and it is essential to achieve high throughput in the pathway. Therefore, four TEs were identified for assessment of their capacity for catalysing this step. Three of these TEs were members of the 4-HB-CoA subfamily: (1) FcbC (*Arthrobacter sp*. Strain AU, accession: Q04416), which has been previously shown to have activity towards short and medium chain aliphatic CoA esters, aromatic CoA esters

with various hydroxyl groups, as well as dihydroxylated and chlorinated aromatics (Song *et al.*, 2012), (2) PA2801 (*Pseudomonas aeruginosa* PAO1, accession: AAG06189.1), which has been shown to have activity towards a range of aliphatic CoA esters ranging from acetyl CoA to palmitoyl CoA (Gonzalez *et al.*, 2012) and (3) YbdB (*E. coli* K-12 MG1655, accession: P0A8Y8), which functions as a 2,3-dihydroxybenzoyl-*holo*EntB proof-reader in the biosynthesis of the siderophore enterobactin and has been shown to have inherent promiscuity, with a bias towards aryl-CoAs (Latham *et al.*, 2014; Wu *et al.*, 2014). The fourth TE was a member of the TesB-like subfamily (4) RpaL (accession: ABD05081) from *Rhodopseudomonas palustris* HaA2, which has not been previously characterised. These four TE were then screened for the desired activity of hydrolysing 2-HCHC CoA into 2-HCHC (Figure 5.1), for implementation within the cyclohexanol pathway.

## 5.1. Cloning and expression of four Thioesterases

For the isolation of the four TEs, both cloning and DNA synthesis was used. *fcbC* from *Arthrobacter sp*. Strain AU was synthesised by DNA2.0 into pD451-SR (pQR2799), an IPTG-inducible expression vector. DNA synthesis and sub-cloning into pET29a (*Nde*I and *Xho*I) by Eurofins (Luxembourg) was used for the production of *rpaL* (pQR2801) from *Rhodopseudomonas palustris* HaA2. Prior to synthesis, *rpaL* and *fcbC* were codon optimised for expression in *E. coli*. Alternatively, *pa2801* and *ybdB* were isolated through PCR from the reference genomes: *Pseudomonas aeruginosa* PAO1 (NC_002516.2) and

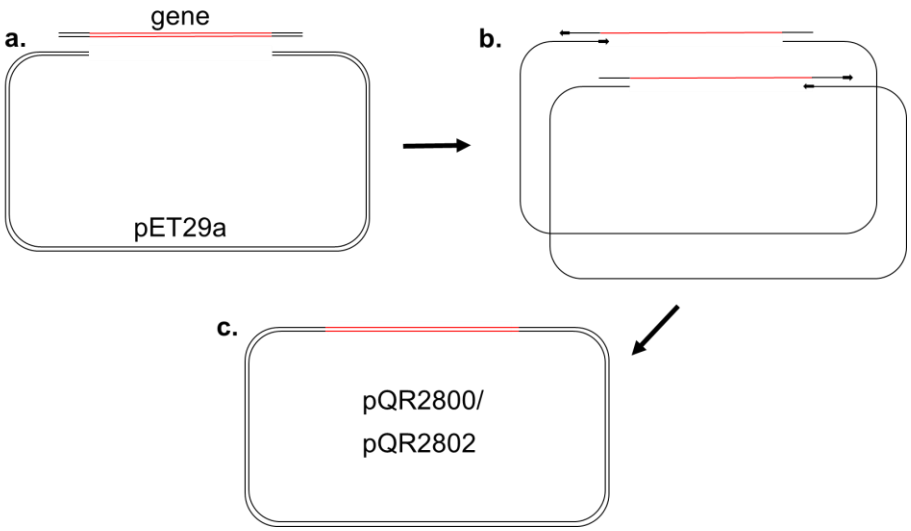



**Figure 5.2. The mechanism of Circular Polymerase Extension Cloning (CPEC).**

(a) The insert to be cloned (*pa2801* and *ybdB*) and the vector (pET29a) were amplified using PCR, the primers used introduced overlapping nucleotide sequences between the insert and vector. (b) The insert and vector use each other as templates for a polymerase to complete the full circle, stopping at its own 5' end. (c) The derived recombinant plasmid (pQR2800 and pQR2802) can then be directly used to transform *E. coli* TOP10 cells.

*E.coli* K12 MG1655 (NC_000913.3), respectively. These amplification products were then separately cloned into pET29a for expression using Circular Polymerase Extension Cloning (CPEC) (Figure 5.2). Primers were designed to introduce overlapping nucleotide sequences between *pa2801* and pET29a (Figure 5.3) and separately between *ybdB* and pET29a (Figure 5.4.), allowing complementary base-pair binding and polymerase extension in the CPEC reactions. This led to the formation of two recombinant plasmids containing *pa2801* (pQR2800) and *ybdB* (pQR2802).

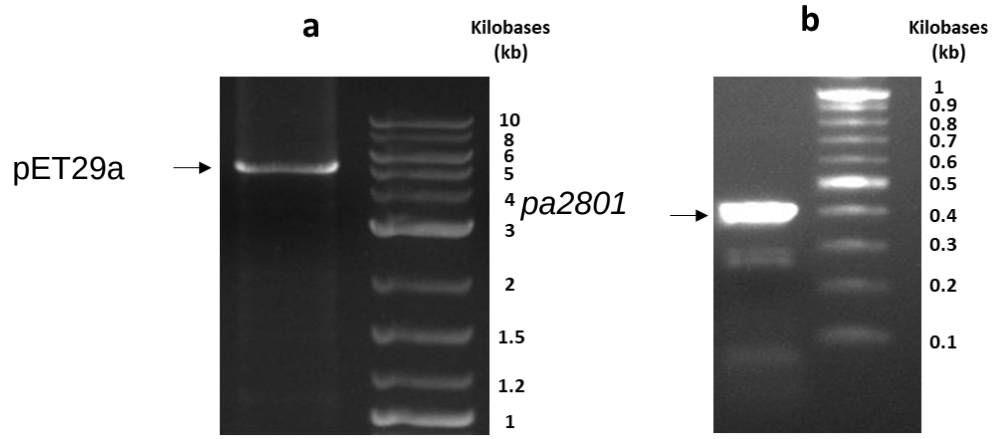



**Figure 5.3. PCR amplification of *pa2801* and pET29a for CPEC.**

An agarose gel showing (a) PCR amplified pET29a containing overlapping nucleotides with *pa2801* (b) PCR amplified *pa2801* (431 bp) containing flanking overlapping nucleotides with pET29a required for CPEC cloning. These two PCR products were used in a CPEC reaction for the formation of the recombinant plasmid, pQR2800.

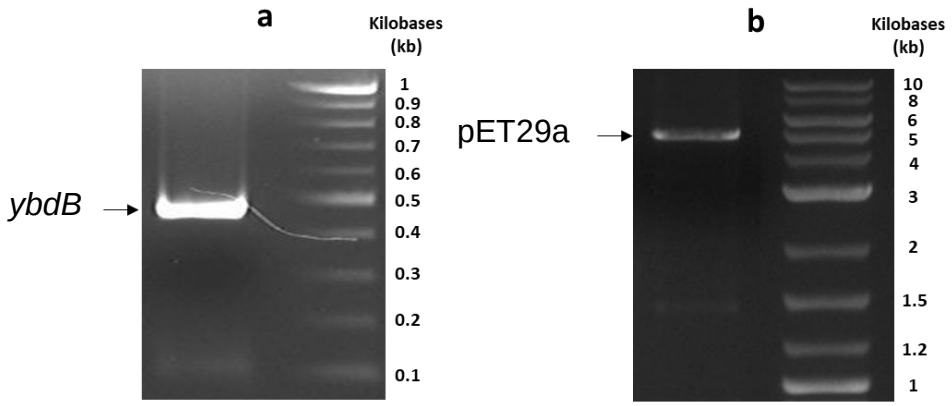



**Figure 5.4. PCR amplification of *ybdB* and pET29a for CPEC.**

An agarose gel showing (a) PCR amplified *ybdB* (455 bp) containing flanking overlapping nucleotides with pET29a required for CPEC cloning and (b) PCR amplified pET29a containing overlapping nucleotides with *ybdB*.

For the identification of *pa2801*, the characterised protein (accession: 3QY3_A) (Gonzalez *et al.*, 2012) was used to identify the nucleotide sequence (Gene ID: 879843) within the *P. aeruginosa* PAO1 reference genome. For identification of *ybdB* the previously characterised enzyme (accession:

P0A8Y8) was used to search the reference genome (NC_000913.3) leading to the identification of the nucleotide sequence (Gene ID: 945215) used here. Following CPEC cloning and DNA synthesis, all four recombinant TE plasmids were used to transform *E. coli* TOP10 cells, for plasmid replication and long-term storage. Purified TE plasmids were then used to transform *E. coli* BL21 (DE3) cells for expression of each TE. Following expression, the soluble protein content of each transformed *E.coli* BL21 (DE3) strain was assessed for the presence of each TE. The SDS gel results showed all four TEs to be expressed within the soluble fraction of lysed *E. coli* BL21 (DE3) cells (Figure 5.5). The size of each TE monomer was FcbC- 17.4 kDa, PA2801- 14.9 kDa, RpaL- 33.3 kDa and YbdB- 14.9 kDa.

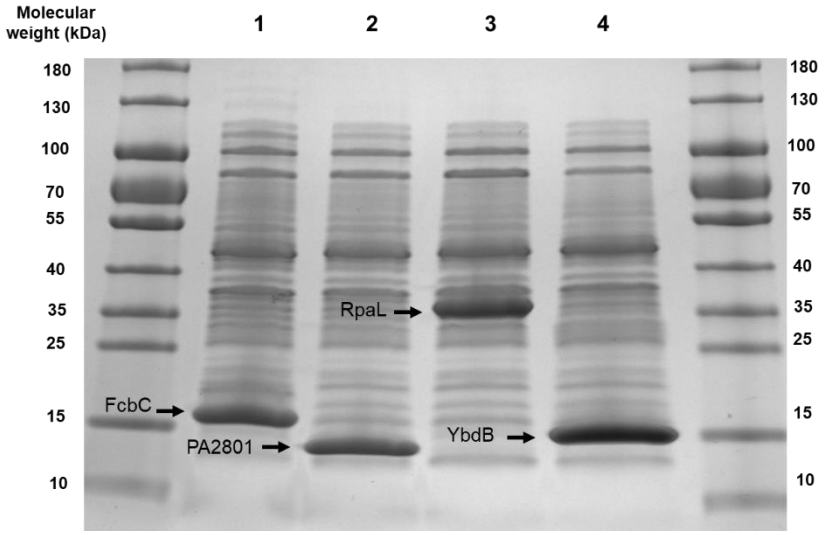



**Figure 5.5. Analysis of the soluble expression of four recombinant TEs.**

An SDS gel containing the clarified lysate from BL21 (DE3) cells expressing (1) FcbC- 17.4 kDa (pQR2799), (2) PA2801- 14.9 kDa (pQR2800), (3) RpaL- 33.3 kDa (pQR2801) and (4) YbdB- 14.9 kDa (pQR2802) (highlighted by black arrows). For each sample, 50 µg of total protein was loaded to each well, using a bradford assay for protein quantification (see chapter 2.15). The 50 ml cultures were induced with 1 mM IPTG and expressed for 5 hours at 37 °C in LB media (see chapter 2.9). All cultures were lysed using Bugbuster (see chapter 2.10). Samples were run alongside 5 µl of PageRuler$^{TM}$ prestained protein ladder for size determination.

## 5.2. Screening four TEs with 2-HCHC CoA, CH-1-eneC CoA and CHC CoA

Once each of the four TEs had been successfully cloned and expressed in *E. coli* BL21 (DE3), clarified lysate containing the soluble TEs was analysed for the desired activity to hydrolyse 2-hydroxycyclohexancarbonyl CoA (2-HCHC CoA) into 2-hydroxycyclohexanecarboxylic acid (2-HCHC) and CoA. In addition, the ideal TE for this step should only be able to hydrolyse 2-HCHC CoA and not CH-1-C CoA or CHC CoA. Therefore, all three of these CoA esters were used as substrates for analysing TE activity. However, both 2-HCHC CoA and CH-1-eneC CoA are not commercially available compounds, consequently these compounds were synthesised using the combined activity of ChCoaDH and BadK (see chapter 4 and chapter 2.24). Furthermore, as the TEs were screened as part of clarified lysate, a negative control using the clarified lysate of cells transformed with an unmodified pET29a plasmid was used to elucidate any CoA ester hydrolysis that may be catalysed by enzymes endogenous to *E. coli* BL21 (DE3). Due to the high cost of CHC CoA and lack of commercial availability of CH-1-eneC CoA and 2-HCHC CoA, replicates of these assays were not performed and so a standard error could not be calculated. As a result the data obtained was limited with regards to its accuracy. However, as the purpose of these experiments was to identify, rather than quantify, a specific TE activity, the accuracy of single assays was considered acceptable for this purpose. A standard instrumental error could not be found for the HPLC used in these experiments (see chapter 2.30), therefore, by assessing the relative errors determined in previous investigations (Sutherland, Nicolau and Kuti, 2011; Lourenco, 2012; Yilmaz, Asci and Erdem, 2014), a <5% change in peak area was considered to be within the error of the HPLC and so would not be attributed to enzyme activity or further investigated.

The hydrolysis results for the negative control showed a 32 µM (15%) drop in CHC CoA, a <1% reduction in CH-1-eneC CoA, a <1% reduction in 2-HCHC CoA and the formation of 33 µM CoA (Figure 5.6. & Table 5.1.).

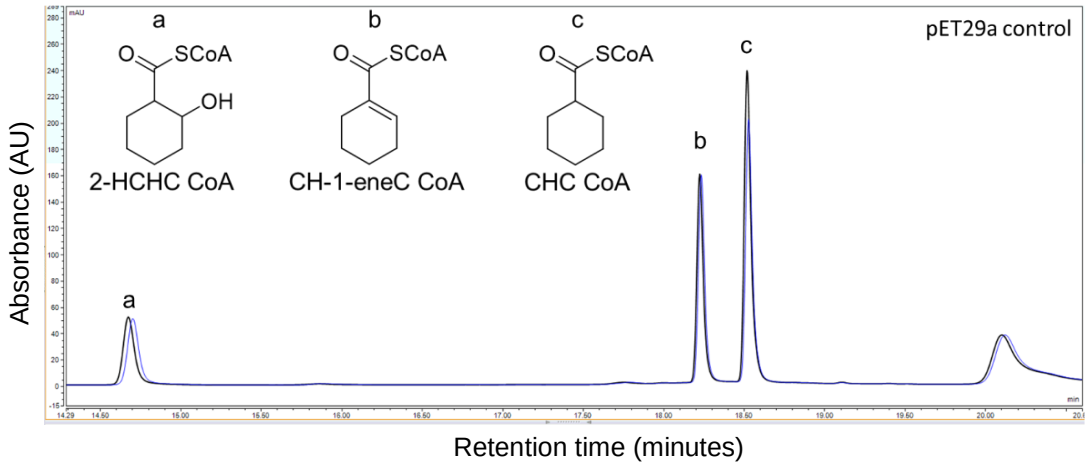
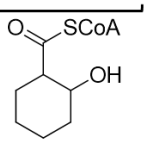
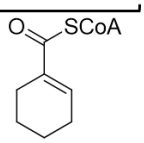
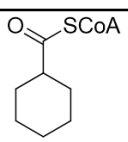



**Figure 5.6. HPLC analysis of TE activity with the negative control.**

HPLC analysis of the hydrolysis of (a) 2-HCHC CoA, (b) CH-1-eneC CoA and (c) CHC CoA, by the clarified lysate of *E. coli* BL21 (DE3) cells transformed with pET29a. Black line- t=0, blue line- t= 30. Assay components and conditions are outlined previously (chapter 2.24) and HPLC conditions were used as described previously (chapter 2.30).

**Table 5.1. The consumption of 2-HCHC CoA, CH-1-eneC CoA and CHC CoA and the production of CoA following TE activity.**

The structure of each CoA ester also is shown under its relevant column.

| TE | 2-HCHC CoA consumption (%) | CH-1-eneC CoA consumption (%) | CHC CoA consumption (%) | CHC CoA consumption (µM) | CoA production (µM) |
|---|---|---|---|---|---|
| pET29a control | <1 | <1 | 15 | 32 | 33 |
| FcbC | 3 | 50 | 10 | 21 | 49 |
| PA2801 | 2 | 20 | 69 | 144 | 108 |
| RpaL | 46 | 88 | 100 | 214 | 234 |
| YbdB | 1 | 49 | 19 | 43 | 55 |
| |  2-HCHC CoA |  CH-1-eneC CoA |  CHC CoA | | |

The clarified lysate of cells expressing FcbC from pQR2799 was next assessed for the consumption of CHC CoA, CH-1-eneC CoA and 2-HCHC CoA. The results showed a 21 µM (10%) consumption of CHC CoA, a 50% consumption of CH-1-eneC CoA, a 3% consumption of 2-HCHC CoA and the formation of 49 µM of CoA (Figure 5.7 & Table 5.1).

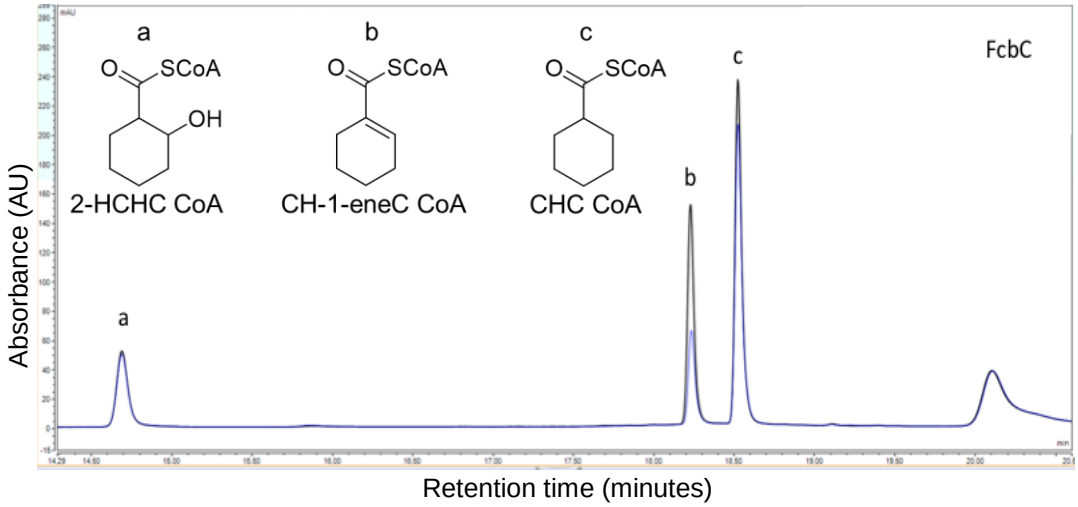



**Figure 5.7. HPLC analysis of TE activity with lysate containing FcbC.**

HPLC analysis of the TE hydrolysis of (a) 2-HCHC CoA (b) CH-1-eneC CoA (c) CHC CoA, by the clarified lysate of *E. coli* BL21 (DE3) cells transformed with pQR2799, expressing FcbC. Black line- t=0, blue line- t= 30. Assay components and conditions are outlined previously (chapter 2.24) and HPLC conditions were used as described previously (chapter 2.30).

The results of screening PA2801 activity within the clarified lysate with CHC CoA, CH-1-eneC CoA and 2-HCHC CoA showed a 144 µM (69%) consumption of CHC CoA, a 20% consumption of CH-1-eneC CoA, a 2% consumption of 2-HCHC CoA and the formation of 108 µM CoA (Figure 5.8 & Table 5.1).

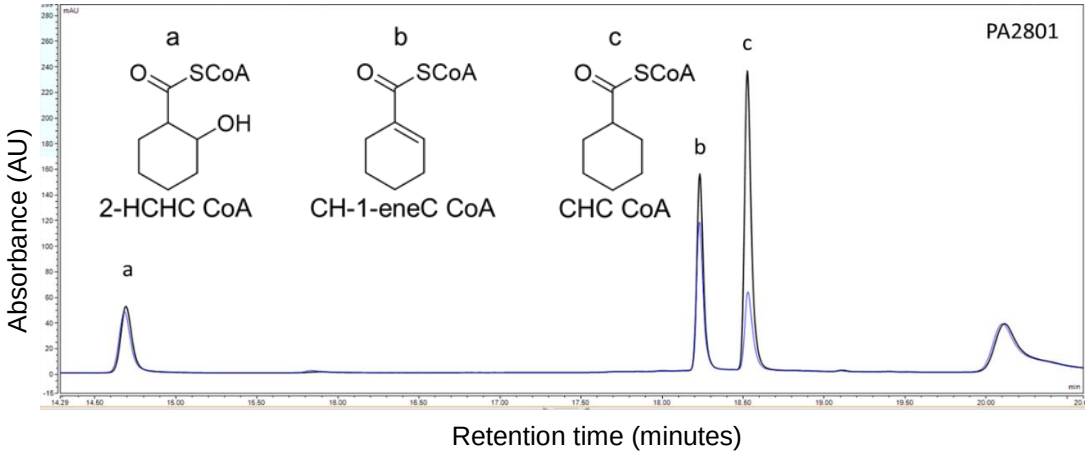



**Figure 5.8. HPLC analysis of TE activity with lysate containing PA2801.**

HPLC analysis of the TE hydrolysis of (a) 2-HCHC CoA (b) CH-1-eneC CoA (c) CHC CoA, by the clarified lysate of *E. coli* BL21 (DE3) cells transformed with pQR2800, expressing PA2801. Black line- t=0, blue line- t= 30. Assay components and conditions are outlined previously (chapter 2.24) and HPLC conditions were used as described previously (chapter 2.30).

Clarified lysate containing recombinantly expressed RpaL was next screened with CHC CoA, CH-1-eneC CoA and 2-HCHC CoA. The results showed CHC CoA to have been completely consumed (214 µM), an 88% consumption of CH-1-eneC CoA, a 46% consumption of 2-HCHC CoA and the formation of 234 µM of CoA (Figure 5.9 & Table 5.1).

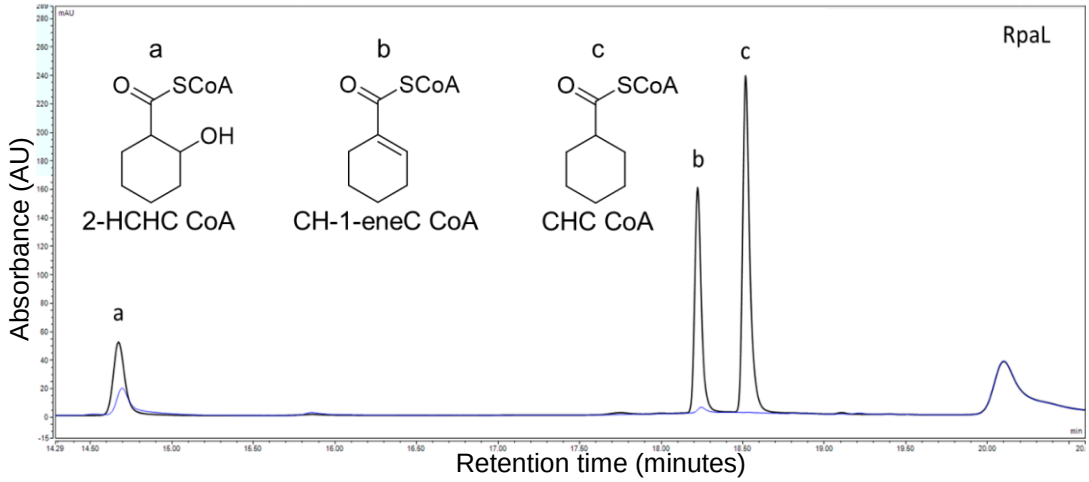



**Figure 5.9. HPLC analysis of TE activity with lysate containing RpaL.**

HPLC analysis of the TE hydrolysis of (a) 2-HCHC CoA (b) CH-1-eneC CoA (c) CHC CoA, by the clarified lysate of *E. coli* BL21 (DE3) cells transformed with pQR2801, expressing RpaL. Black line- t=0, blue line- t= 30. Assay components and conditions are outlined previously (chapter 2.24) and HPLC conditions were used as described previously (chapter 2.30).

Finally, YbdB within the clarified lysate was screened with CHC CoA, CH-1-eneC CoA and 2-HCHC CoA. The results showed a decrease of 43 µM for CHC CoA (19%), a 49% reduction in the peak area of CH-1-eneC CoA, a 1% reduction in the peak area of 2-HCHC CoA and the formation of 55 µM CoA (Figure 5.10 & Table 5.1).

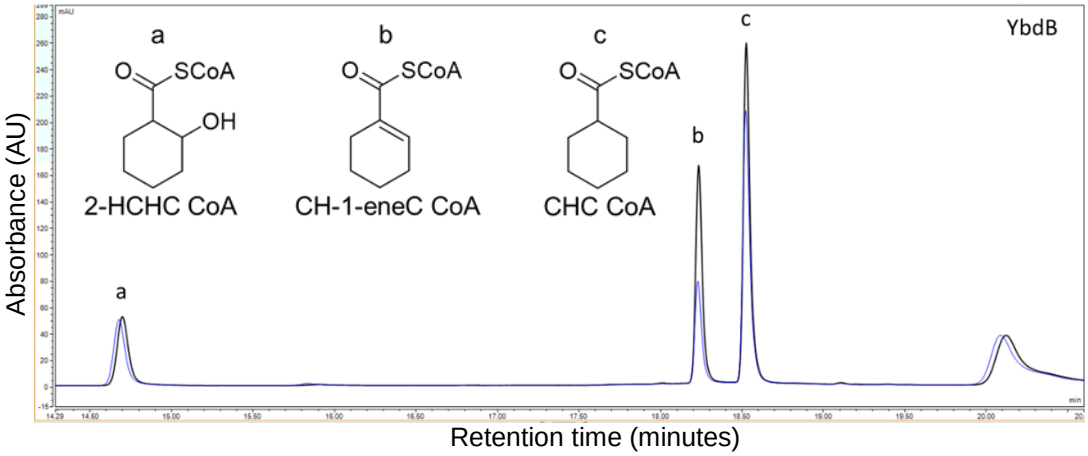



**Figure 5.10. HPLC analysis of TE activity with lysate containing YbdB**

HPLC analysis of TE hydrolysis of (a) 2-HCHC CoA (b) CH-1-eneC CoA (c) CHC CoA, by the clarified lysate of *E. coli* BL21 (DE3) cells transformed with pQR2802 expressing YbdB. Black line- t=0, blue line- t= 30. Assay components and conditions are outlined previously (chapter 2.24) and HPLC conditions were used as described previously (chapter 2.30).

## 5.3. Confirmation of 2-HCHC production

As a 46% drop in the peak area corresponding to 2-HCHC CoA was observed following the addition of clarified lysate containing RpaL to the reaction mixture, it was thought that this was as a result of 2-HCHC CoA being hydrolysed into 2-HCHC and CoA, however, this possibility needed to be confirmed. Therefore, following the RpaL reaction, the assay was analysed through HRMS to verify the production of 2-HCHC. The HRMS (ES-) results determined an *m/z* peak at 143.0721 (theoretical m/z of $C_7H_{12}O_3$ $[M\text{-}H]^-$ 143.0714) confirming the successful production of 2-HCHC via the RpaL-mediated hydrolysis of 2-HCHC CoA (Figure 5.11).

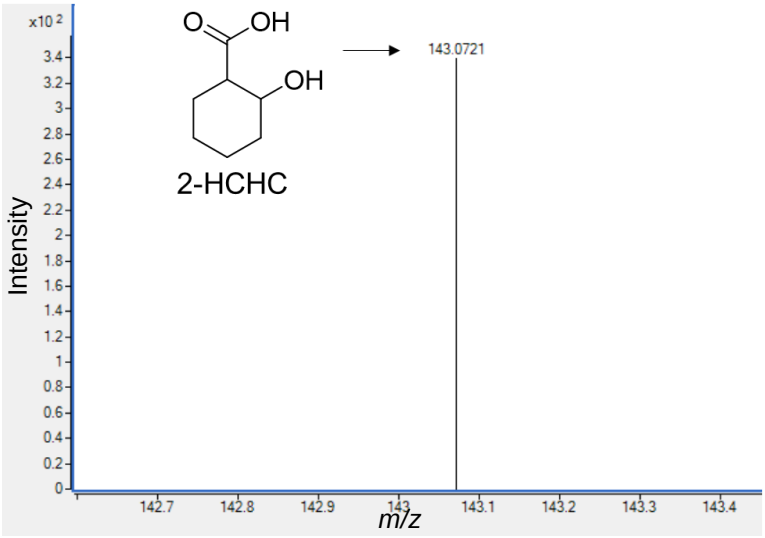



**Figure 5.11. HRMS (ES-) analysis of 2-HCHC production.**

HRMS (ES-) analysis to verify the production of 2-HCHC, following a 30 minute incubation of 2-HCHC CoA with RpaL (in clarified lysate). Results confirmed the production of 2-HCHC (error <5 ppm) shown as its anion $[M\text{-}H]^-$ with an *m/z* of143.0721. Theoretical m/z of $C_7H_{12}O_3$ $[M\text{-}H]^-$ 143.0714. The compound was purified and prepared for analysis following method stated previously (chapter 2.35).

## 5.4. Discussion and conclusions

### 5.4.1. An overview of results

The results described above assessed the capacity of four TEs (FcbC, PA2801, RpaL and YbdB) to hydrolyse CHC CoA, CH-1-eneC CoA and 2-HCHC CoA, with the objective of identifying one that is able to hydrolyse 2-HCHC CoA into 2-HCHC.

Firstly, hydrolysis of 32 µM of CHC CoA was found to occur in the negative control with 33 µM of CoA formed. Analysis of the annotated genome of *E. coli* BL21 (DE3) (Genbank: CP001509.3) shows seven endogenous TEs, including the YbdB homologue. None of these have been previously screened with CHC CoA, CH-1eneC CoA or 2-HCHC CoA, however, a number of them have been shown to have activity with a range of short, medium and long-chain saturated fatty acyl CoA esters (Bonner and Bloch, 1972; Li *et al.*, 2000; Eklöf *et al.*, 2009; McMahon and Prathera, 2014). Therefore, this observed breadth of activity towards saturated fatty acyl CoA esters might extend to CHC CoA, possibly providing an explanation for the 15% drop in CHC CoA observed in the negative control. Additionally, this reduction in CHC CoA could have been a result of non-enzymatic water hydrolysis. The small decreases in the peak area of CH-1-eneC CoA (<1%) and 2-HCHC CoA (<1%) observed, were <5%, therefore, were considered to be experimental error rather than enzymatic hydrolysis. Nonetheless, the consumption of each CoA ester observed within the negative control was used as a baseline to determine whether CoA ester hydrolysis, observed in lysate fractions containing a recombinantly expressed TE, was specifically catalysed by each recombinant TE.

Clarified lysate containing recombinantly expressed FcbC was found to consume 21 µM (10%) of CHC CoA, 11 µM less than the negative control, and produce 49 µM of CoA, 16 µM more than the negative control. Therefore, the

observed consumption of CHC CoA was not attributed to specific FcbC activity. The amount of clarified lysate used within each assay had been normalised by total protein content, as a result clarified lysate overexpressing FcbC would have a lower composition of endogenous enzyme that could hydrolyse CHC CoA, possibly resulting in the lower consumption of CHC CoA observed. The 49.8% consumption of CHC-1-eneC CoA was much higher than the <1% decrease found in the negative control, indicative of recombinant TE activity. Furthermore, the 3% decrease in the 2-HCHC peak area, although greater decrease than in the negative control (<1%), was <5% and so was considered to be experimental error rather than enzymatic hydrolysis. Therefore, the 28 μM difference between CHC CoA consumption and CoA formation was attributed to FcbC-mediated hydrolysis of CH-1-eneC CoA. FcbC has been previously shown to have a preference (highest specificity constant) for hydroxylated aromatic CoA esters, such as 4-hydroxybenzoyl CoA (4-HB CoA) ($5.4 \times 10^6$ $M^{-1}s^{-1}$), 3-hydroxybenzoyl-CoA (3-HB CoA) ($7.2 \times 10^5$ $M^{-1}s^{-1}$) and benzoyl CoA ($7.6 \times 10^3$ $M^{-1}s^{-1}$) (Zhuang *et al.*, 2003). Therefore, as CH-1-eneC CoA contains an α-β-double bond, it may be a preferential substrate, due to the conjugated thioester as is the case with benzoyl CoA, than either CHC CoA or 2-HCHC CoA. This might explain why selective activity is found with CH-1-eneC CoA.

A 144 μM (69%) consumption of CHC CoA was observed with clarified lysate overexpressing PA2801. This was 112 μM more than that found within the negative control, clearly attributing this consumption to be a result of PA2801 activity. Furthermore, the 20% decrease in peak area corresponding to CH-1-eneC CoA was considerably higher than that observed in the negative control, suggesting PA2801 to also have activity towards CH-1-eneC CoA. Finally, the 2% drop in peak area corresponding to 2-HCHC CoA, was greater than that observed in the negative control, however, it was <5% and so was considered to be a result of error rather than enzymatic hydrolysis. Interestingly, despite the combined consumption of CHC CoA and CH-1-eneC CoA, only 108 μM of CoA was observed. As this was lower than the consumption of CHC CoA alone, it was assumed that this was a result of endogenous enzyme activity

within the clarified lysate. The CoA activation of carboxylic acids within central metabolism is essential for viable growth and survival (Shimizu, 2013). Therefore, the consumption of CoA by CoA synthetases and other CoA-utilising enzymes endogenous to *E.coli* is a possible explanation for the lower than equivalent (to CHC CoA consumption) amount of CoA observed. PA2801 has been previously observed to have preferential activity towards saturated fatty acyl CoA-esters such as: octanoyl CoA, glutaryl CoA, lauroyl CoA and hexanoyl CoA (Gonzalez *et al.*, 2012). CHC CoA is a saturated alicyclic CoA ester and so fits the substrate preference of PA2801 and further shows that PA2801 can accept cyclic as well as linear aliphatic CoA esters.

Clarified lysate containing recombinantly expressed YbdB was found to consume 43 μM (19%) CHC CoA, only 4% more than the 32 μM (15%) reduction in the negative control. Therefore, without triplicate data allowing significant variation to be determined, it cannot be concluded that YbdB is able to catalyse the hydrolysis of CHC CoA. Additionally, only a 1% reduction in 2-HCHC CoA was observed, which was attributed to error rather than enzymatic hydrolysis. Finally, a 49% decrease in the peak area of CH-1-eneC CoA was found, which clearly determined YbdB to actively hydrolyse this CoA ester. In a previous investigation focused on elucidating the substrate spectrum of YbdB, it was revealed that it had a preference towards aromatic CoA esters, with the highest specificity constant found with benzoyl CoA ($7.1 \times 10^5\ M^{-1}s^{-1}$), which was found to decrease with hydroxylation: 4-HB-CoA ($5.8 \times 10^5\ M^{-1}s^{-1}$), 3-HB-CoA (not activity determined) (Latham *et al.*, 2014). As CH-1-eneC CoA contains an α-β-double bond, it may be a preferential substrate for YbdB, like with FcbC, due to the conjugated thioester as is the case with benzoyl CoA.

Clarified lysate containing overexpressed RpaL was found to result in the consumption of all three CoA esters. Complete consumption of CHC CoA was observed in addition to an 88% decrease in CH-1-eneC CoA and a 46% decrease in 2-HCHC CoA. RpaL has not previously been characterised,

however, it has shown here to have high hydrolytic activity towards hydroxylated, unsaturated and alicyclic CoA esters.

In the context of implementing a TE within the proposed cyclohexanol pathway, all four TEs were found to decrease the peak area corresponding to 2-HCHC CoA. However, the consumption of 2-HCHC CoA observed for FcbC, PA2801 and YbdB, was <5% and comparable to the negative control, therefore, it was attributed to error rather than enzymatic hydrolysis. Conversely, RpaL activity led to a considerably larger decrease of 2-HCHC CoA (46%) then the other TEs, suggesting it to be the only active TE with 2-HCHC CoA and so the best candidate for implementation into the cyclohexanol pathway. Despite the results showing a greater decrease of 2-HCHC CoA with clarified lysate containing RpaL, compared to the negative control, it could not be fully concluded that 2-HCHC CoA was being successfully hydrolysed into 2-HCHC. Therefore, RpaL activity was further analysed for the production of 2H-CHC, using HRMS analysis, which confirmed the production of 2-HCHC (Figure 5.11). From this, the decrease in 2-HCHC CoA could conclusively be attributed to the hydrolytic activity of RpaL forming 2-HCHC. Furthermore, this result provided successful confirmation of the biotransformation of CHC CoA to 2-HCHC production, which is a decarboxylation step away from cyclohexanol, a substrate of the pre-existing nylon 6 monomer pathway.

### 5.4.2. Considerations when integrating RpaL into the cyclohexanol biosynthetic pathway

Despite confirmation of the desired hydrolysis of 2-HCHC CoA to 2-HCHC, a limitation to the use of RpaL within the proposed synthetic metabolic system is the observation that it also hydrolyses both CHC CoA and CH-1-eneC CoA. By having this activity, RpaL would be competing with both ChCoADH and BadK for the production of CH-1-eneC CoA and 2-HCHC CoA, respectively. Consequently, this would limit the amount of 2-HCHC that could be produced

due to the formation of side products, such as CHC and CH-1-eneC. Furthermore, this activity may not be limited to just CHC CoA and CH-1-eneC CoA, RpaL may also be able to use the CoA-activated intermediates of CHC CoA pathway, which would further limit the production of 2-HCHC. Unfortunately, due to the lack of commercial availability of all the CoA ester intermediates of the CHC CoA (see chapter 3) pathway, RpaL activity with these compounds cannot be determined. However, for the viability of an efficient biosynthetic system for producing cyclohexanol, the use of RpaL will require consideration to preclude side-product formation. One possible mechanism would be to split the cyclohexanol pathway into two separate modules. The first could contain the CHC CoA operon, ChCoaDH and BadK, producing 2-HCHC CoA, and the second module could contain RpaL and a decarboxylase (Figure 5.12). Furthermore, through the implementation of enzymes to regenerate required cofactors (ATP, NADPH and FAD), this first module would lead to an accumulation of 2-HCHC CoA. Here, a glucose dehydrogenase (GDH) could be used to regenerate NADPH with the addition of glucose (Weckbecker and Hummel, 2005), and combined polyphosphate-AMP phosphotransferase (PAP) and polyphosphate kinase (PPK) could regenerate ATP with the addition of polyphosphate (Kameda *et al.*, 2001). Ferrocenium hexafluorophosphate is currently used to regenerate FAD, however, an alternative approach may be to use a ferredoxin and a ferredoxin—NADP+ reductase (EC 1.18.1.2). This combined activity would be able to utilise the NADP+ and $FADH_2$ formed to regenerate NADPH and FAD (Carrillo and Ceccarelli, 2003). Activation of the second module, through addition of purified RpaL and DC in an *in vitro* system, or through induced plasmid-based expression in a metabolic system, would then facilitate directed TE activity towards 2-HCHC CoA. The RpaL-derived 2-HCHC could then be decarboxylated to the final product, cyclohexanol. Furthermore, splitting the cyclohexanol pathway into two separate modules would be compatible with the pre-existing nylon 6 monomer biosynthetic pathway. This is due to none of this intermediates of the pre-existing pathway being CoA-activated, removing the risk of unwanted TE activity on these intermediates.

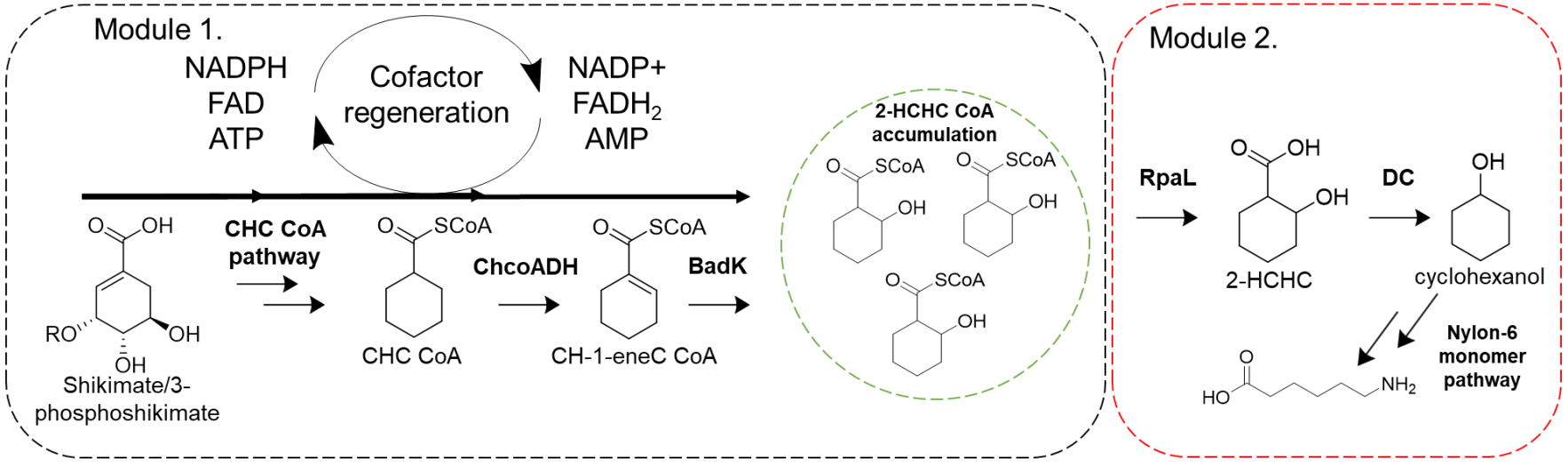



**Figure 5.12. A proposed method for implementing RpaL activity within the cyclohexanol pathway.**

Here, the enzyme components of the cyclohexanol pathway could be split into two modules. Module 1 would contain the CHC CoA pathway enzymes (MycA1-4 and ChcB) and the CHC degradation pathway enzymes (ChCoADH and BadK) for the conversion of shikimate or 3-phosphoshikimate to 2-HCHC CoA. Furthermore, the inclusion of enzymes to regenerate cofactors (NADPH, FAD and ATP) would theoretically facilitate an accumulation of 2-HCHC CoA. Following the accumulation of 2-HCHC CoA, Module 2, containing RpaL and a DC, would be activated facilitating RpaL activity to be directed towards 2-HCHC CoA, forming 2-HCHC. The 2-HCHC would then be converted to cyclohexanol via DC activity which could then enter the existing nylon 6 monomer pathway.

### 5.4.3. Alternative uses of a promiscuous TE

In the context of the cyclohexanol pathway, the promiscuous activity observed with RpaL clearly presents a caveat to its implementation. However, when considering promiscuous TE activity in alternative contexts, this could be seen as an extremely desirable attribute. Hypothetically, if a mechanism could be developed to take advantage of the wealth of known CoA ester modifying enzymes and pathways, this would present a route to the synthesis of a plethora of novel CoA esters. A TE with a broad substrate acceptance could then be implemented to hydrolyse these CoA esters into novel organic acids, which may possess desirable functionalities. However, the high price and general lack of commercial availability of CoA esters imposes a significant obstacle to the characterisation of CoA ester modifying enzymes, and would further impose a financial caveat to the synthesis of novel organic acids through this route. Therefore, an enzymatic method for the synthesis of a range of CoA esters would not only provide a desirable route for producing substrates for screening CoA ester modifying enzymes , it would also alleviate the financial limitation of producing novel organic acids in this manner.

With this understanding, the next chapter firstly describes the isolation and characterisation of a CoA ligase (AliA) for the enzymatic synthesis of a range of CoA esters. CoA ester modification enzymes, cloned within previous chapters of this thesis, were then coupled with CoA ligase activity for the synthesis of novel CoA esters. Next, the substrate promiscuity of the four TEs isolated within this chapter (FcbC, PA2801, YbdB and RpaL) was assessed, the most promiscuous of which (RpaL) was purified and characterised further. Finally, RpaL was coupled with CoA ligase and CoA ester modification activity as a proof of principle for the production of novel organic acids via the proposed CoA activation, modification and TE hydrolysis route.

# 6. Chapter 6. Development of an organic acid modification system

Throughout the work described in chapter 3-5, the lack of commercially available CoA esters, and high price of CoA esters that were available, was a frequently encountered problem. For example, at the beginning of this PhD, none of the 11 CoA ester intermediates of the proposed biosynthetic cyclohexanol pathway were commercially available. This immediately imposed a significant limitation for characterising the enzymes within the CHC CoA pathway (chapter 3), the CHC degradation pathway (chapter 4) and for screening TE activity (chapter 5). The chemical synthesis of CoA esters was one possible solution to this problem, however, this introduces additional issues such as the insolubility of CoA in organic solvents, the interference of functional groups with chemical reactants and the fundamental requirement of a chemistry laboratory, which is not always available to synthetic biologists. Furthermore, despite a number of chemical synthesis routes for the acylation of CoA existing, these are often specific descriptions for a single CoA ester, with no standard method existing for the production of a range of CoA esters (Peter *et al.*, 2016).

Therefore, development of a versatile method, or system, through which to synthesise CoA esters, would be an extremely desirable tool. Construction of such a system may be feasible by exploiting biological components for the production of a versatile biosynthetic system. Specifically, by employing the activity of a CoA ligase, a common method for the production of a range of CoA esters may be proposed. Hypothetically, if a promiscuous CoA ligase could be obtained, or synthetically derived, then a plethora of organic acids could be activated into their corresponding CoA esters. In this context, a limitation to this system would be the availability/ cost of the organic acid required for activation. However, if CoA ligase activity were to be coupled with enzymatic modification of the CoA-activated organic acid, then this would derive a system capable of developing CoA esters with minimal constraints

imposed by the initial substrate. This would be an extremely useful tool for the synthesis of CoA esters required for screening CoA-dependent enzymes. However, beyond the context of enzymatic assays, the CoA esters themselves may not be compounds of commercial interest.

As previously described in chapter 1, the rational synthesis of novel organic compounds is a globally significant area of science, due to associated functionalities. Therefore, a common biological route for the hydrolysis of these modified CoA esters into their corresponding organic acids, such as a promiscuous thioesterase (TE), would confer this biosynthetic system with the capacity to rationally produce novel organic acids. Furthermore, with the understanding that ~5% of all enzymes listed on the BRENDA database utilise CoA in some manner, there is no shortage of candidate enzymes for the system envisaged (Peter *et al.*, 2016).

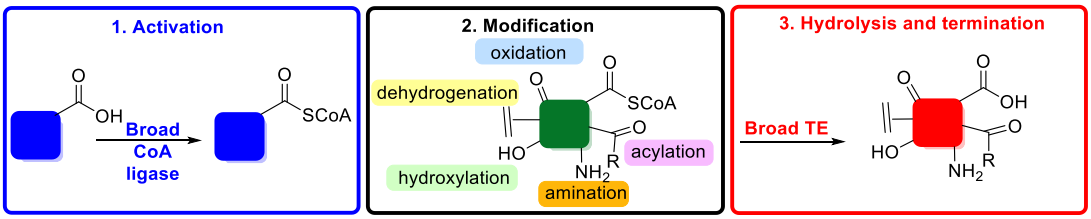



**Figure 6.1. The 3 modules of the proposed organic acid modification system.**

Module 1 (activation) will utilise a promiscuous CoA ligase to activate a range of organic acids into their corresponding CoA esters. Module 2 (modification) will add bespoke functionality to the CoA ester using enzymes from a developed toolbox of CoA-dependent enzymes. Module 3 (termination) will utilise a broad spectrum TE to hydrolyse a range of CoA esters into their corresponding organic acids.

Specifically, the CoA ligase, AliA, from *R. pseudopalustris* was isolated, initially for the production of commercially unavailable CHC CoA, but then later, for determination of its substrate promiscuity. Following the isolation and recombinant expression of: (1) a CoA ligase, (2) seven CoA ester (chapter 3 and 4) modifying enzymes and (3) four TEs (chapter 5), a three module system for the production of organic acids became apparent. The first module (activation) of this system would utilise the activity of a broad acting CoA

ligase, which would produce CoA esters from a range of organic acid substrates. In the second module (modification), a number of CoA-dependent enzymes and biological modules can be used, providing highly specific functional complexity to the CoA ester. Finally, the third component (termination) of this system would employ a broad spectrum thioesterase (TE) to hydrolyse the thioester bond, releasing the novel organic acid (Figure 6.1). The importance of the CoA ligase and TE having broad activity would mean they could be used with a plethora of substrates in a single common system. Moreover, the development of a 'toolbox' of CoA ester modifying enzymes, would facilitate a 'plug and play' approach to these enzymes in order to achieve the desired complex organic acid.

With this understanding, this chapter described studies to determine the substrate promiscuity of both AliA and the four TEs cloned within chapter 5. The strategy was then then that the TE that was found to have the most promiscuous activity would be purified and used within the termination step of this system. Finally, AliA and TE activity could be combined with the recombinant CoA ester modifying enzymes isolated within the previous chapters of this study (MycA2, ChCoADH, BadK) as a proof of concept for the proposed biosynthetic organic acid modification system (Figure 6.2)

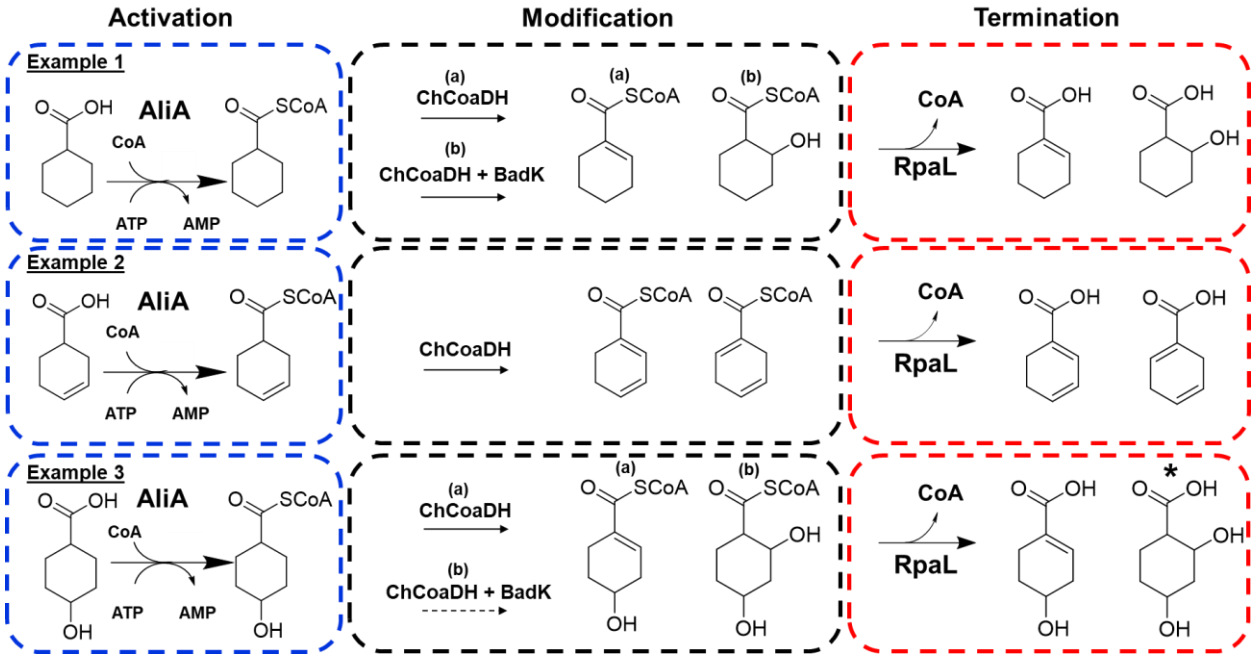



**Figure 6.2. Functional examples of the organic acid modification system.**

Three examples are shown that were used as a proof of concept for the proposed organic acid modification system. In the first 'activation' module (blue boxes) the CoA ligase, AliA, was used to active organic acids into corresponding CoA esters. The second module 'modification' (black boxes) used enzymes isolated within chapter 4, for the specific modification of CoA esters. The final module 'termination' (red box) used the promiscuous TE, RpaL, to hydrolse the modified CoA esters into their corresponding organic acids. Dashed arrow indicate proposed steps that were not experimentally assessed. Asterisks represent compounds that are proposed but were not experimentally produced.

## 6.1. Cloning, expression and purification of AliA

Primers were used for the isolation of a*liA* from the genome of *R. pseudopalustris,* which were also designed with an N-terminal *Nde*I and a C-terminal *Xho*I restriction site. The resultant recombinant plasmid, pQR2803 (see appendix 8.33 for nucleotide sequence), was used to transform *E.coli* BL21 (DE3) cells, from which AliA was recombinantly expressed and found to be soluble. AliA was then purified from the clarified lysate using the incorporated N-terminal His-tag to facilitate binding and purification from a nickel column. The soluble recombinant protein (Figure 6.3) was successfully purified and migrated through an SDS gel slightly further than the 58 kDa protein marker, marginally further than expected, based on its predicted molecular weight (61.8 kDa).

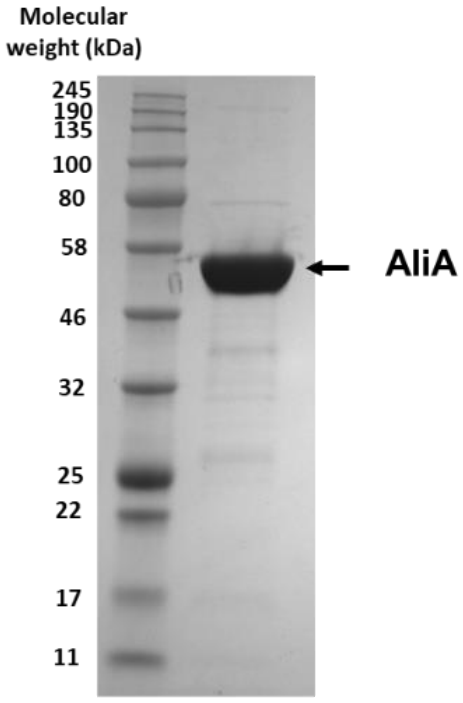



**Figure 6.3. The His-tag purification of AliA.**

An SDS gel showing recombinantly expressed and purified AliA (predicted size 61.8 kDa) following elution from a nickel column with 500 mM imidazole. AliA is indicated with a black arrow. 5 µl of a 500 mM imidazole eluate was prepared and loaded onto an SDS gel as previously described (chapter 2.17). AliA was expressed from *E.coli* BL21 (DE3) transformed with pQR2803, grown in 300 ml of TB media for 24 hours at 25 °C shaking at 250 rpm.

## 6.2. AliA activity and substrate promiscuity

Following purification, CoA ligase activity was assessed with a range of organic acids, including: cyclohexane carboxylic acid (CHC), 3,5-dihydroxycyclohexane carboxylic acid (3,5-diCHC), cyclohex-3-enecarboxylic acid (CH-3-eneC), benzoic acid (Bz), *cis/trans* 3-hydroxycyclohexanecarboxylic acid (c/t 3-HCHC), 4-hydroxyclyclohexanecarboxylic acid (4-HCHC), 4-oxocyclohexanecarboxylic acid (4-oxoCHC), shikimate, 2-oxocyclohexanecarboxylic acid (2-oxoCHC), 3-phosphoshikimate (3-PS) and *trans* 2-aminocyclohexanecarboxylic acid (*t*2-aCHC) (Figure 6.4).

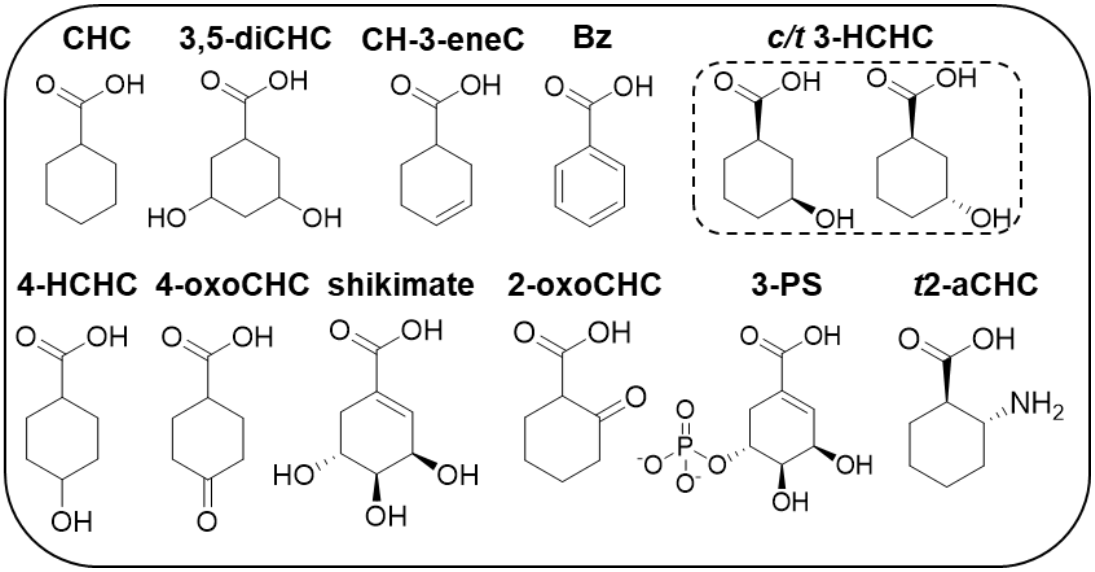



**Figure 6.4. All substrates used for screening AliA activity.**

The chemical structures of the substrates used for screening AliA activity. cyclohexanecarboxylic acid (CHC), 3,5-dihydroxycyclohexanecarboxylic acid (3,5-diCHC), cyclohex-3-enecarboxylic acid (CH-3-eneC), benzoic acid (Bz), *cis/trans* 3-hydroxycyclohexanecarboxylic acid (*c/t* 3-HCHC) (dashed box indicating racemate), 4-hydroxyclyclohexanecarboxylic acid (4-HCHC), 4-oxocyclohexanecarboxylic acid (4-oxoCHC), shikimate, 2-oxocyclohexanecarboxylic acid (2-oxoCHC), 3-phosphoshikimate (3-PS) and *trans* 2-aminocyclohexanecarboxylic acid (*t* 2-aCHC).

The results showed AliA to have significant activity (p value < 0.001) with six of these organic acids: CHC, CH-3-eneC, Bz, c/t 3-HCHC, 4-HCHC and 4-oxoCHC. However, no significant activity was found with the more complex CoA esters, containing multiple hydroxyl groups, phosphate groups and functional groups at the 2 position (Figure 6.5). CHC was used as a positive control as this is the natural substrate for AliA within the CHC degradation pathway. Of the 0.4 mM CoA added to each assay, 0.036 mM remained after the addition of CHC, 0.078 mM with CH-3-eneC, 0.11 mM with benzoate, 0.103 mM with 3-HCHC, 0.104 mM with 4-HCHC and 0.106 mM with 4-oxoCHC. Additionally, control assays were performed containing no AliA to determine whether a drop in CoA could be caused by the addition of the organic acid substrate, however, no significant drop (p value > 0.01) in CoA concentration was observed with any of the substrates (Figure 6.6).

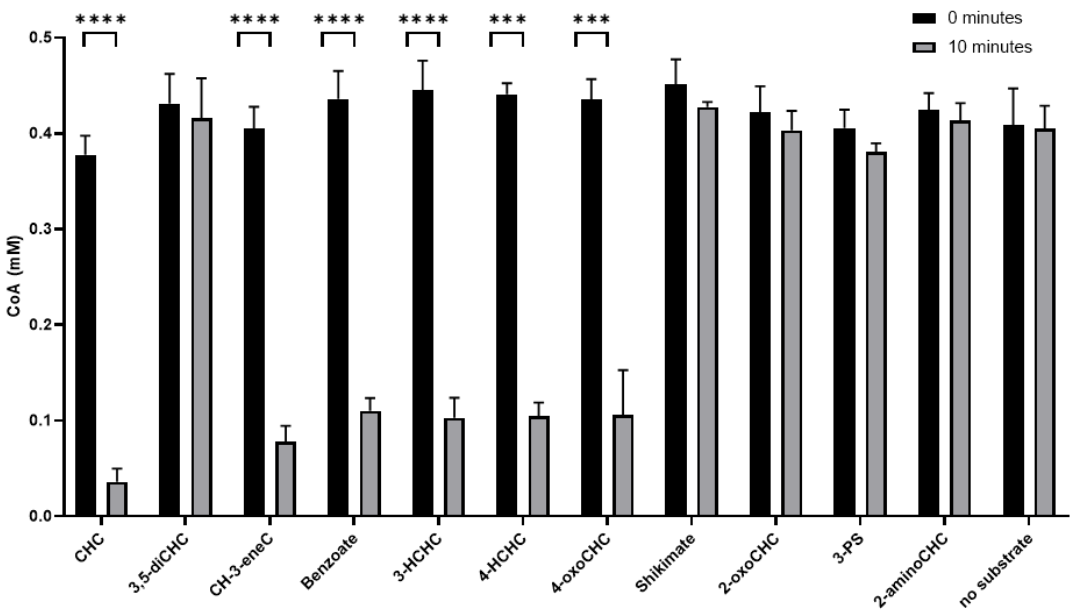



**Figure 6.5. Analysis of AliA activity with a range of substrates.**

An end-point assay determining AliA activity with a range of substrate. Activity was determined by a decrease in the concentration of CoA. Statistical significance (T-test) between the CoA concentration 0 minutes and 10 minutes after addition of substrate is signified with an asterisk. **** - p value < 0.0001, *** - p value < 0.001. Assays contained 0.4 mM CoA and were performed in triplicate and carried out according to the method described previously (chapter 2.26).

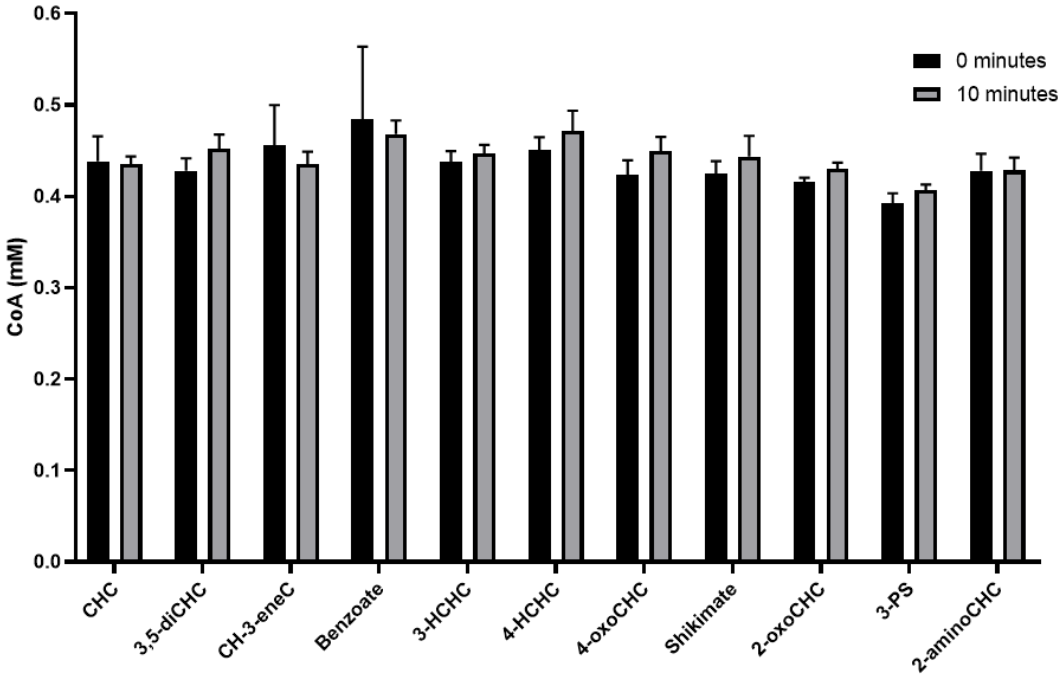



**Figure 6.6. Analysis of AliA activity- negative control.**

An end-point assay of control reactions containing no addition of AliA to determine whether any significant decrease of CoA is observed with the addition of a range of substrates. Statistical significance between the CoA concentration 0 minutes and 10 minutes after addition of substrate was determined using a T-test, however no significance (p value > 0.01) was found with any of the substrates. Assays contained 0.4 mM CoA and were performed in triplicate and carried out according to the method described previously (chapter 2.26).

## 6.3. Producing novel CoA esters by coupling AliA activity with CoA ester modification enzymes

### 6.3.1. Production of CH-1-eneC CoA and 2-HCHC CoA

CHC CoA was not commercially available prior to 2017 and although it has become available since then, it is still extremely expensive with prices varying between £300-850 per 10 mg (CoALA biosciences, Bertin Bioreagent). Therefore, producing CHC CoA through AliA activity was an extremely desirable alternative. Following the confirmation of AliA activity with CHC it was coupled with ChCoADH and BadK for the production of CH-1-eneC CoA and 2-HCHC CoA. The HPLC results confirmed the consumption of CoA following the addition of AliA, combined with the formation of CHC CoA (Figure 6.7).

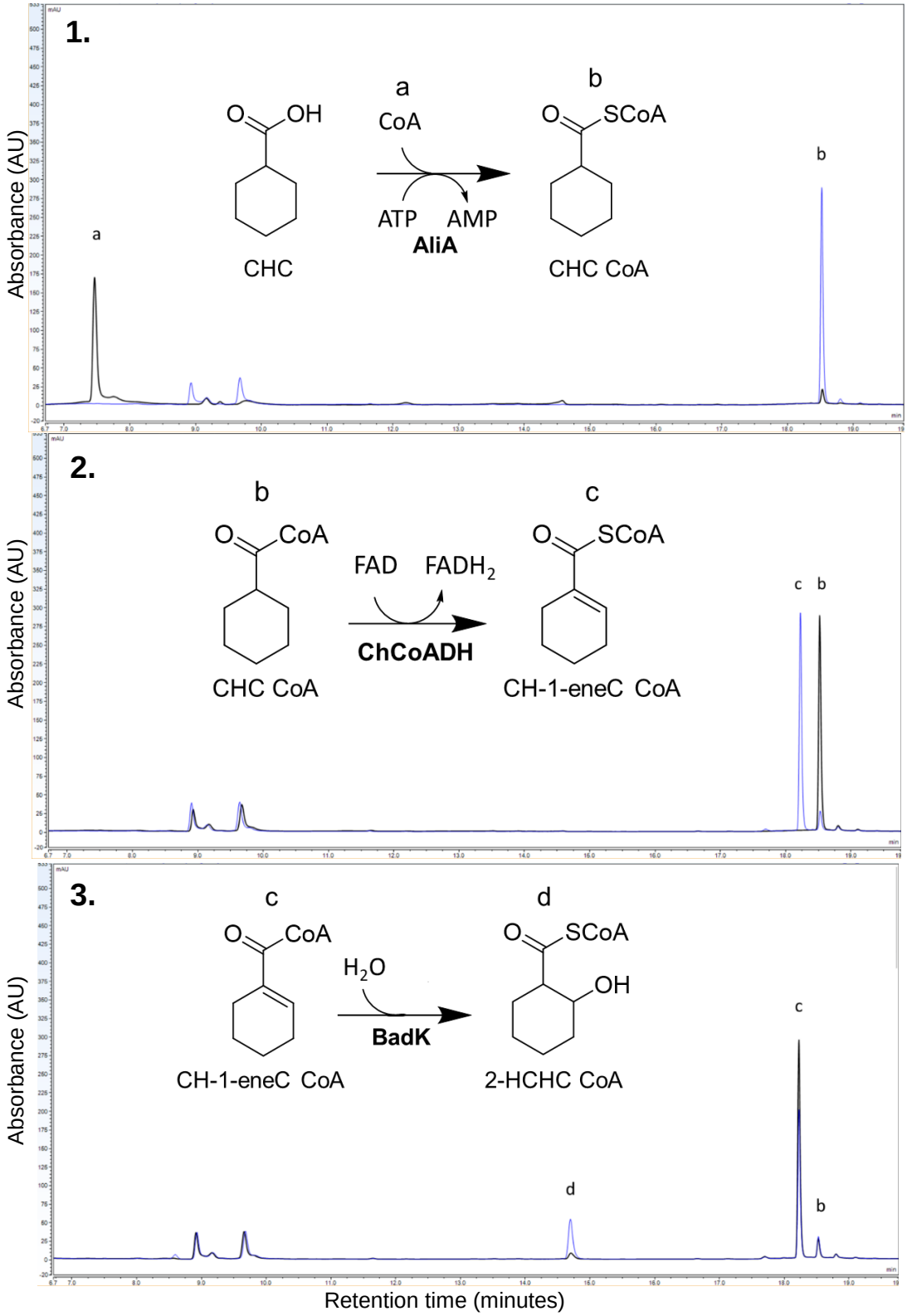



**Figure 6.7. Production of CHC CoA, CH-1-eneC CoA and 2-HCHC CoA from CHC.**

HPLC data of sequential (1) AliA activity, (2) ChCoADH activity and (3) BadK activity with CHC. Three CoA esters were produced: (b) CHC CoA, (c) CH-1-eneC CoA and (d) 2-HCHC CoA. Reactions were performed as previously described (chapter 2.28). Black line- t=0, blue line- after enzyme incubation.

Then with the addition of ChCoADH, CHC CoA was shown to be oxidised to CH-1-eneC CoA and finally, following the addition of BadK, 2-HCHC CoA was shown to be produced (Figure 6.7)**.** These results showed that AliA can be used as an alternative method for the production of CoA esters that can be further modified by CoA ester utilising enzymes. However, CHC was the natural substrate of AliA, as was CHC CoA for ChCoADH and CH-1-eneC CoA for BadK. Therefore, to determine the flexibility of this activation and modification system, activity was assessed with alternative organic acids.

### 6.3.2. Production of 4-HCH-1-eneC CoA from 4-HCHC

Next, 4-HCHC was used as the substrate for CoA activation via AliA activity, to determine whether the corresponding CoA ester could also be enzymatically modified. CoA activation of 4-HCHC was monitored using HPLC analysis. Following AliA activity, CoA consumption was found to be coupled with the formation of a novel peak, putatively corresponding to 4-HCHC CoA (Figure 6.8).

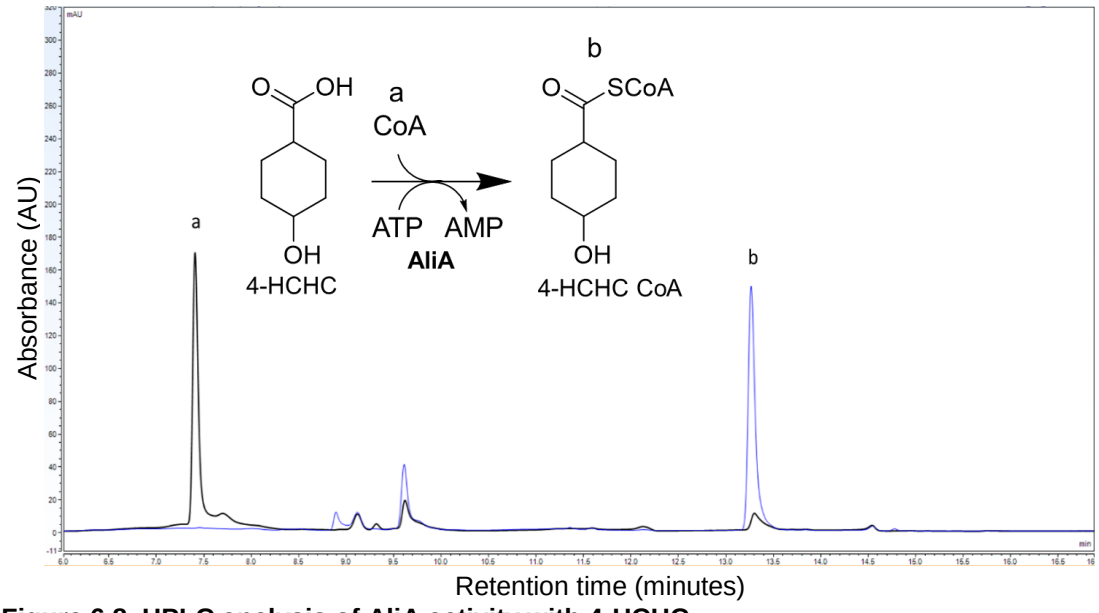



**Figure 6.8. HPLC analysis of AliA activity with 4-HCHC.**

HPLC data showing CoA ligase activity with 4-HCHC as the substrate. Black line- immediately after addition of AliA, blue line- reaction after incubation with AliA. a- CoA and b- 4-HCHC CoA. The reaction was performed as previously described (chapter 2.26).

To confirm that this new compound peak corresponded to 4-HCHC CoA it was collected from the HPLC (see chapter 2.30) and analysed using HRMS to determine its *m/z*. The HRMS (ES+) results confirmed this compound to be 4-HCHC CoA showing an *m/z* of 894.1896 (theoretical [M+H] - 894.1906) (Figure 6.9).

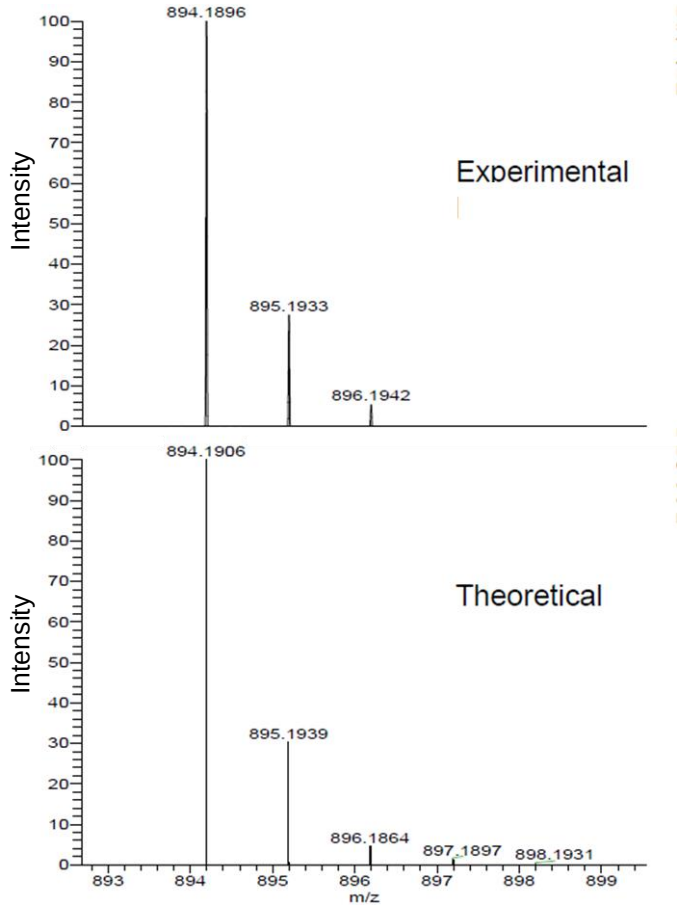



**Figure 6.9. HRMS (ES+) analysis of 4-HCHC CoA.**

HRMS (ES+) data of putative 4-HCHC CoA HPLC peak collected from the HPLC and detected as a cation. Theoretical [M+H] ($C_{28}H_{46}N_7O_{18}P_3S_1$ +H) 894.1906, observed *m/z* 894.1896. An observed error of <5 ppm confirmed this compound to be 4-HCHC CoA. The compound was purified and prepared for analysis following methods stated previously (chapter 2.30 & 2.35).

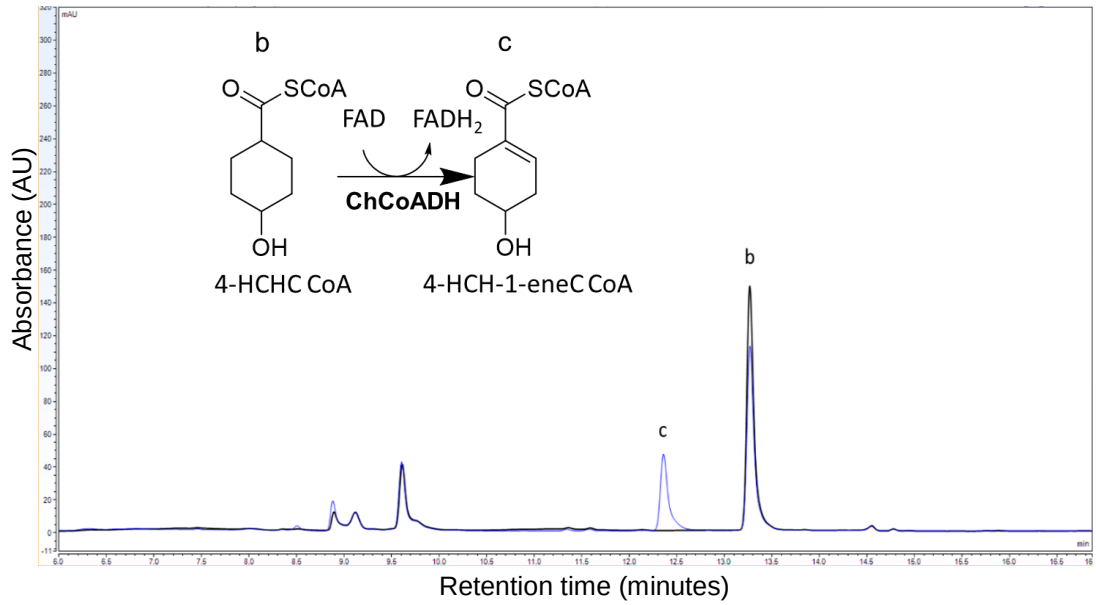



**Figure 6.10. HPLC analysis of ChCoADH activity with 4-HCHC CoA.**

HPLC analysis of ChCoADH activity using 4-HCHC CoA as the substrate directly after (black line) and 30 minutes after (blue line) addition of ChCoADH. The assay performed following the previously described method (chapter 2.28).

Following the confirmation of 4-HCHC CoA, ChCoADH was screened with this compound to determine whether it could produce the corresponding oxidised CoA ester, 4-HCH-1-eneC CoA. The results, following the addition of ChCoADH, showed a 22% decrease in the 4-HCHC CoA peak area, combined with the formation of a novel peak, putatively 4-HCH-1-eneC CoA (Figure 6.10). To confirm whether this new peak was 4-HCH-1-eneC CoA, it was collected following HPLC separation and analysed using HRMS as before. HRMS (ES+) confirmed the production of 4-HCH-1-eneC CoA determining its *m/z* to be 892.1728 (theoretical [M+H] 892.1749) (Figure 6.11).

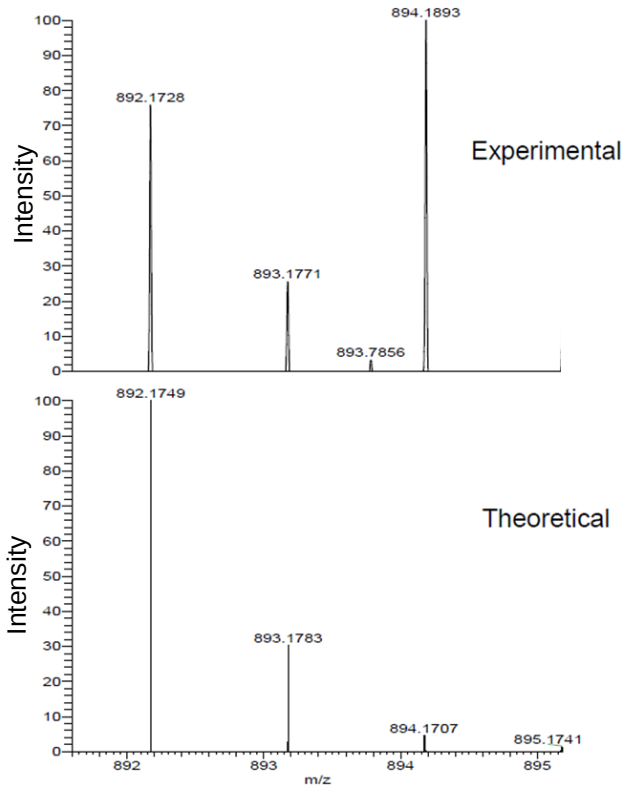



**Figure 6.11. HRMS (ES+) analysis of 4-HCH-1-eneC CoA.**

HRMS (ES+) data of putative 4-HCH-1-ene-C CoA peak collected from the HPLC and detected as a cation. Theoretical [M+H] ($C_{28}H_{44}N_7O_{18}P_3S_1$ +H) 892.1749, observed *m/z* 892.1728. An observed error of <5 ppm confirmed this compound to be 4-HCH-1-eneC CoA. The compound was purified and prepared for analysis following methods stated previously (chapter 2.30 & 2.35).

### 6.3.3 Production of 3-HCH-1-eneC CoA for screening MycA2 activity

As *cis/trans* 3-HCHC was observed to be a viable substrate for AliA activity, the derived CoA ester 3-HCHC CoA was evaluated as a possible substrate with ChCoADH for production of 3-HCH-1-eneC CoA. Both 3-HCHC CoA and 3-HCH-1-eneC CoA are not commercially available and represent the proposed product and substrate, respectively, for MycA2 activity (described in chapter 3). Therefore, a production method for both of these CoA esters was extremely desirable in the context of both screening MycA2 activity and increasing the substrate panel of ChCoADH within the organic acid

modification system. HPLC analysis of AliA activity with *cis/trans* 3-HCHC showed the complete consumption of CoA followed by the formation of two unknown overlapping peaks, likely representing each of the *cis/trans* stereoisomers of 3-HCHC CoA (Figure 6.12).

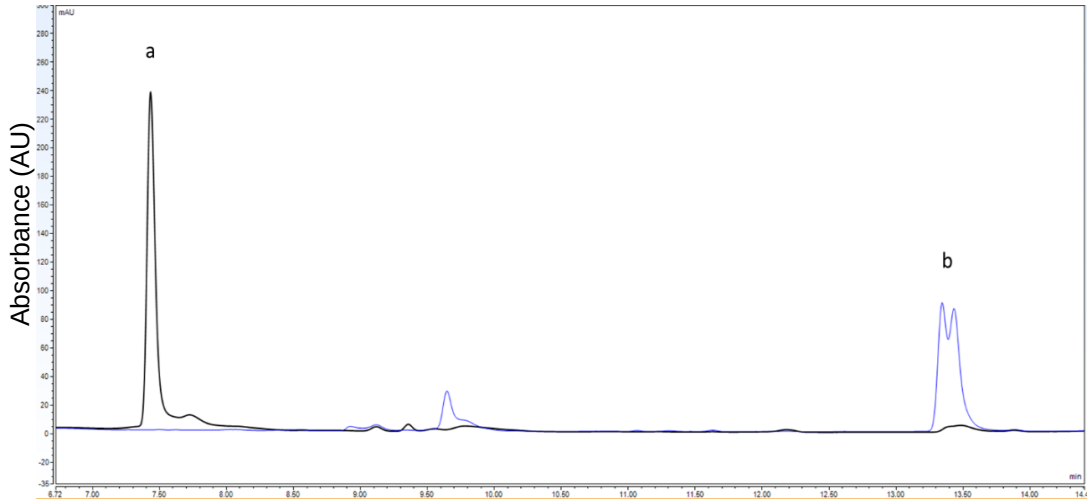



**Figure 6.12. HPLC analysis of AliA activity with cis/trans3-HCHC.**

HPLC analysis of CoA ligase activity using *cis/trans* 3-HCHC as the substrate directly after (black line) and 10 minutes after (blue line) addition of AliA. A double peak for cis/trans 3-HCHC CoA is found, likely representing each of the stereoisomers of *cis/trans* 3-HCHC CoA. The assay was performed followed the methods stated previously (chapter 2.26).

To confirm that these new overlapping peaks corresponded to *cis/*trans 3-HCHC CoA, they were collected in a single fraction from the HPLC and analysed through HRMS (ES+). This compound was found to have an *m/z* of 894.1891 corresponding to the [M+H] of 3-HCHC CoA ($C_{28}H_{46}N_7O_{18}P_3S_1$ theoretical [M+H] 894.1906), confirming the production of 3-HCHC CoA (Figure 6.13).

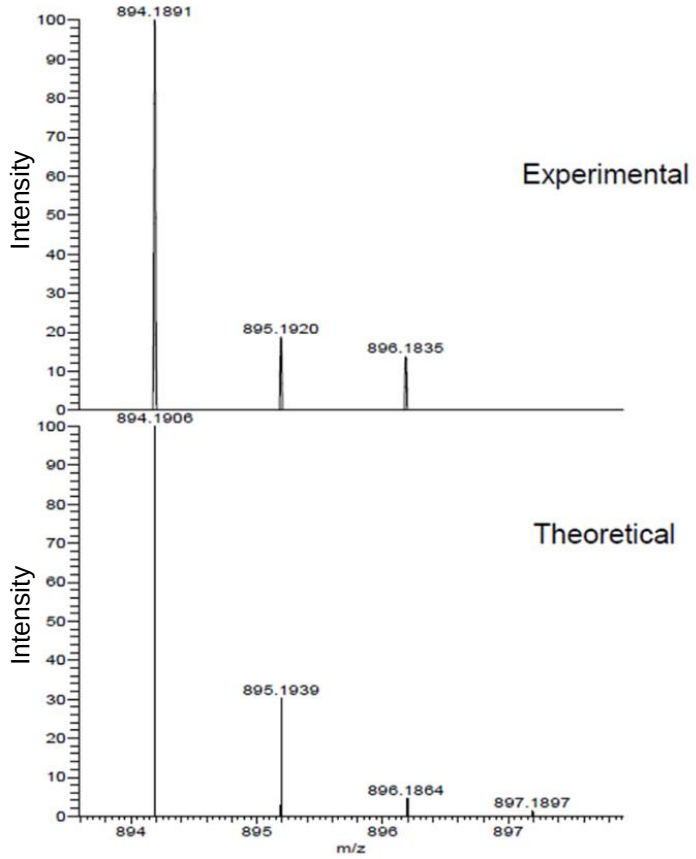



**Figure 6.13. HRMS (ES+) analysis of 3-HCHC CoA.**

HRMS (ES+) analysis of an unknown compound, purified through HPLC, putatively corresponding to 3-HCHC CoA. Theoretical [M+H] 894.1906, experimental *m/z* 894.1891. An observed error of <5 ppm confirmed this compound to be 3-HCHC CoA. The compound was purified and prepared for analysis following methods stated previously (chapter 2.30 & 2.35).

Following the confirmation of 3-HCHC CoA production it was used to screen ChCoADH activity. HPLC analysis showed a decrease in the peak area corresponding to *cis/trans* 3-HCHC CoA followed by the formation of a new unknown peak putatively corresponding to 3-HCH-1-eneC CoA (Figure 6.14).

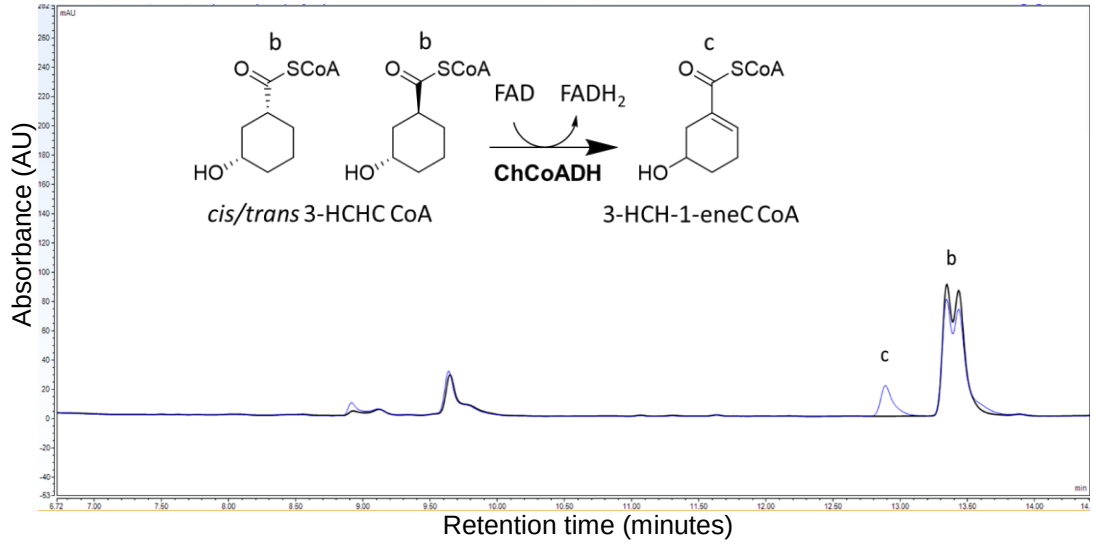



**Figure 6.14. HPLC analysis of ChCoADH activity with cis/trans 3-HCHC CoA.**

HPLC analysis of dehydrogenase activity using *cis/trans* 3-HCHC CoA as the substrate directly after (black line) and 30 minutes after (blue line) addition of ChCoADH. (b) *cis/trans* 3-HCHC CoA (c) *cis/trans* 3-HCH-1-eneC CoA. The assay was performed according to the methods stated previously (chapter 2.28).

This new peak was purified through HPLC fraction collection and analysed using HRMS (ES+). The theoretical [M+H] of 3-HCH-1-eneC CoA was 892.1749 and the observed *m/z* of the purified compound was 892.1737, confirming this compound to be 3-HCH-1-eneC CoA (Figure 6.15). With the confirmation of 3-HCH-1-eneC CoA it could then be used as a substrate for assessing MycA2 activity (chapter 3- Figure 3.26 & 3.27).

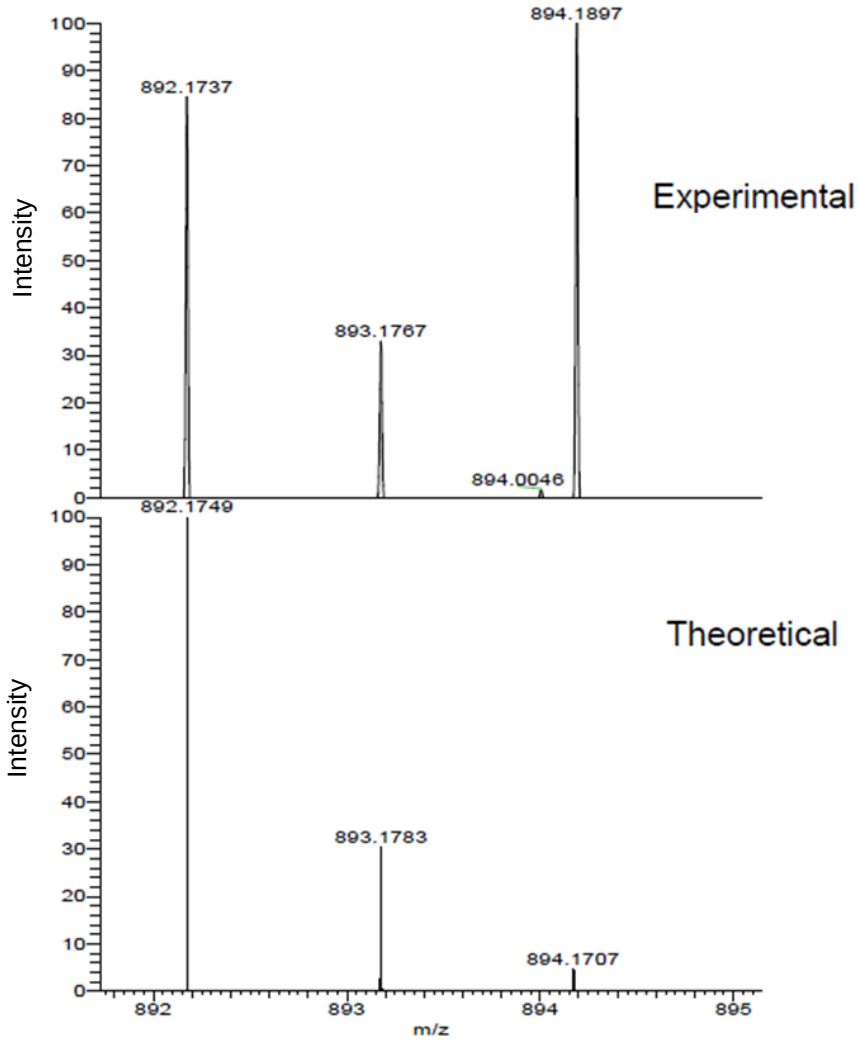



**Figure 6.15. HRMS (ES+) analysis of 3-HCH-1-eneC CoA.**

HRMS (ES+) analysis of an unknown compound collected through HPLC, putatively corresponding to 3-HCH-1-eneC CoA. Theoretical [M+H] 892.1749, experimental *m/z* 892.1737. An observed error of <5 ppm confirmed this compound to be 3-HCH-1-eneC CoA. The compound was purified and prepared for analysis following the methods stated previously (chapter 2.30 & 2.35).

### 6.3.4. Production of CH-1, 3-eneC CoA from CH-3-eneC

As AliA had been found to be active towards CH-3-eneC this activity was assessed by HPLC, so that ChCoADH activity could be monitored with the newly derived CoA ester, CH-3-eneC CoA.

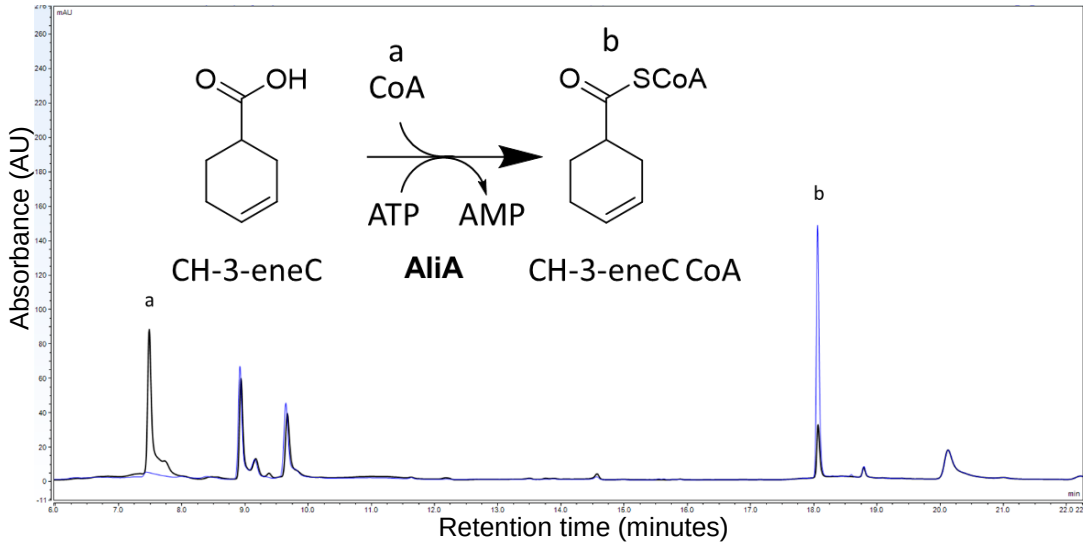



**Figure 6.16. HPLC analysis of AliA activity with CH-3-eneC as the substrate.**

Following the incubation with AliA (a) CoA is completely consumed and a new peak is found putatively corresponding to (b) CH-3-eneC CoA. Black line- immediately after AliA addition, blue line- 10 minutes after addition of AliA. The assay was performed according to the method stated previously (chapter 2.26).

HPLC analysis of the AliA activity showed the complete consumption of CoA, with the corresponding formation of a novel peak, putatively corresponding to CH-3-eneC CoA. This peak was collected and analysed by HRMS (ES+), which showed the compound to have an *m/z of* 876.5148 (theoretical [M+H] 876.18) (Figure 6.17). An observed error of 334 ppm was outside of the <5 ppm acceptable range for compound assignment. Therefore, this compound could not be conclusively identified as CH-3-eneC CoA. However, as AliA activity with CH-3-eneC had already been confirmed in triplicate (Figure 6.5), the CoA consumption and subsequent single HPLC peak formation was thought likely to be CH-3-eneC CoA. Therefore, the >5 ppm inaccuracy found is thought to be due to incorrect calibration of the HRMS and so analysis

should be repeated, however, due to insufficient time this could not be achieved.

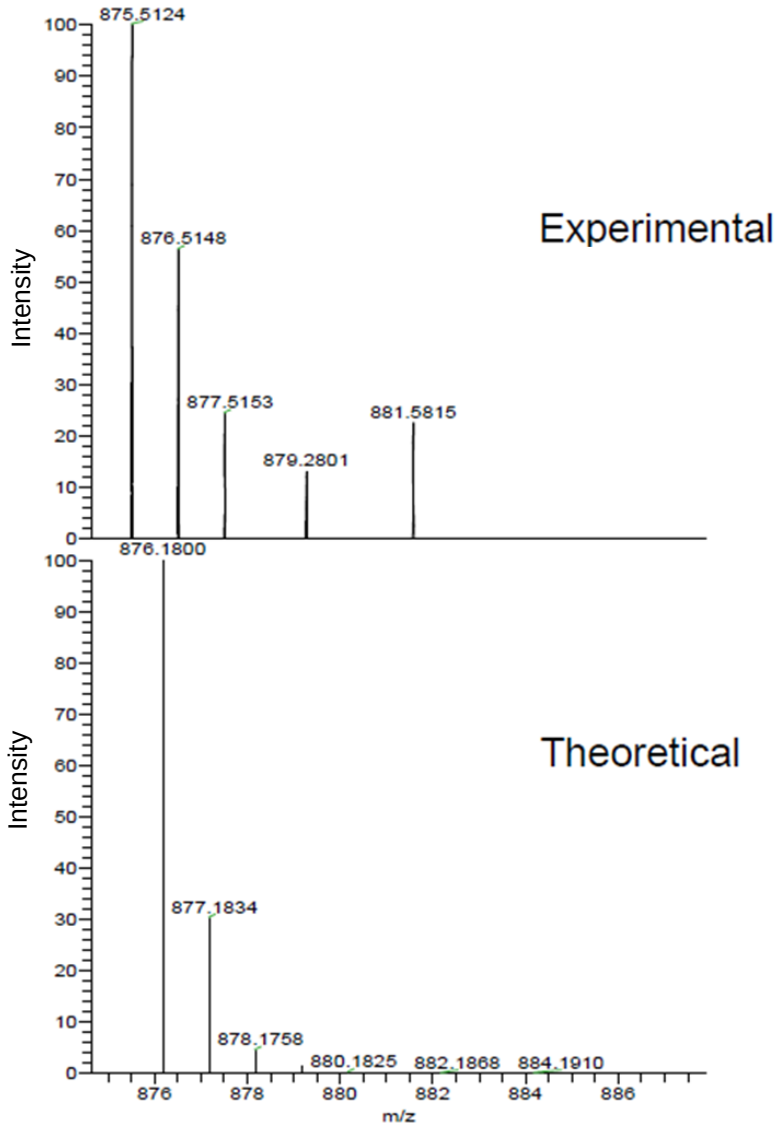



**Figure 6.17. HRMS (ES+) analysis of putative CH-3-eneC CoA.**

HRMS (ES+) analysis of an unknown compound collected through HPLC, putatively corresponding to CH-3-eneC CoA. Theoretical [M+H] 876.18, experimental *m/z of* 876.5148. An error of >5 ppm was determined and so this compound cannot be conclusively identified as CH-3-ene C CoA. The compound was purified and prepared for analysis following the methods stated previously (chapter 2.30 & 2.35).

Following the putative production of CH-3-eneC CoA it was next assessed whether it could be used as a substrate for ChCoADH dehydrogenase activity, for the production of CH-1, 3-eneC CoA. HPLC analysis showed a reduction

in the peak area putatively corresponding to CH-1-eneC CoA following a 30 minute incubation with ChCoADH, combined with the formation of two additional peaks, possibly corresponding to the two stereoisomers of CH-1, 3-eneC CoA (Figure 6.18).

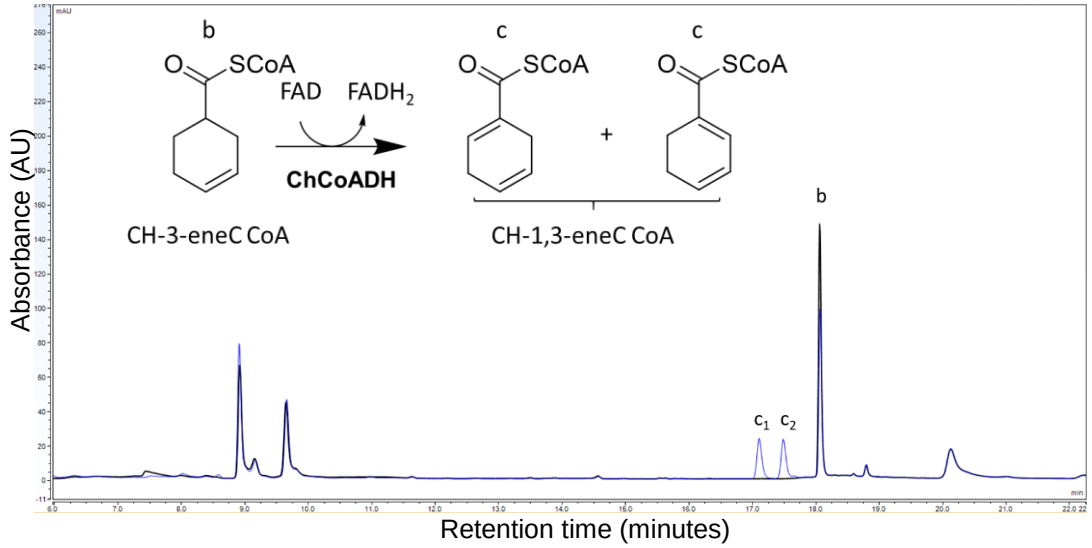



**Figure 6.18. HPLC analysis of ChCoaDH activity with putative CH-3-eneC CoA.**

HPLC analysis of ChCoADH dehydrogenase activity with the compound that is putatively CH-3-eneC CoA. Following the incubation with ChCoADH (b) CH-3-eneC CoA is consumed and (c) two new peaks are found, putatively corresponding to the two stereoisomers of CH-1,3-eneC CoA. Black line- directly after ChCoADH addition, blue lines- 30 minutes after addition of ChCoADH. CH-1,3-eneC CoA is labelled as $c_1$ and $c_2$ to signifying that each peak putatively represents two stereoisomers.

To confirm that these two new compounds were the two stereoisomers of CH-1, 3-eneC CoA, both peaks were collected and analysed using HRMS (ES+). The results determined the *m/z* of both compounds to be: $c_1$- 874.2950 and $c_2$- 874.2952 (theoretical *m/z* [M+H] 874.1643) (Figure 6.19). An observed error of 131 ppm (>5 ppm) for both compounds means, as with CH-3-eneC CoA, these two compounds cannot be confirmed as CH-1,3-eneC CoA isomers. However, as a corresponding reduction in CH-3-eneC CoA peak area was observed in this assay, it is likely that these two peaks correspond to the

CH-1,3-eneC CoA isomers. Like with CH-3-eneC CoA, an HRMS calibration issue is thought to be likely cause for the >5 ppm error observed and so this analysis should be repeated, however, due to insufficient time this could not be achieved.

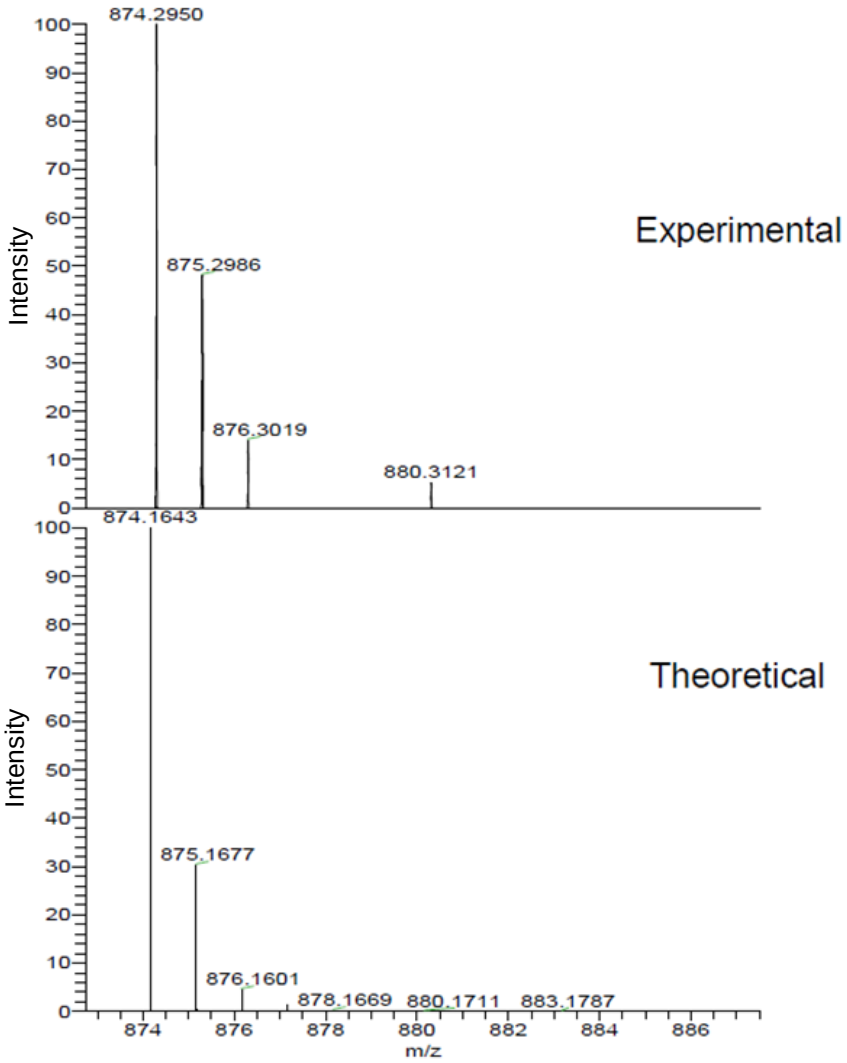



**Figure 6.19. HRMS (ES+) analysis of putative CH-1,3-eneC CoA.**

HRMS (ES+) analysis of an unknown compound collected through HPLC, putatively corresponding to CH-1, 3-eneC CoA ($c_1$). Theoretical [M+H] 874.1643, experimental *m/z* of 874.2950. Due to an observed error of >5 ppm this compound could not be conclusively identified and will need to be reanalysed. The compound was purified and prepared for analysis following the methods stated previously (chapter 2.30 & 2.35).

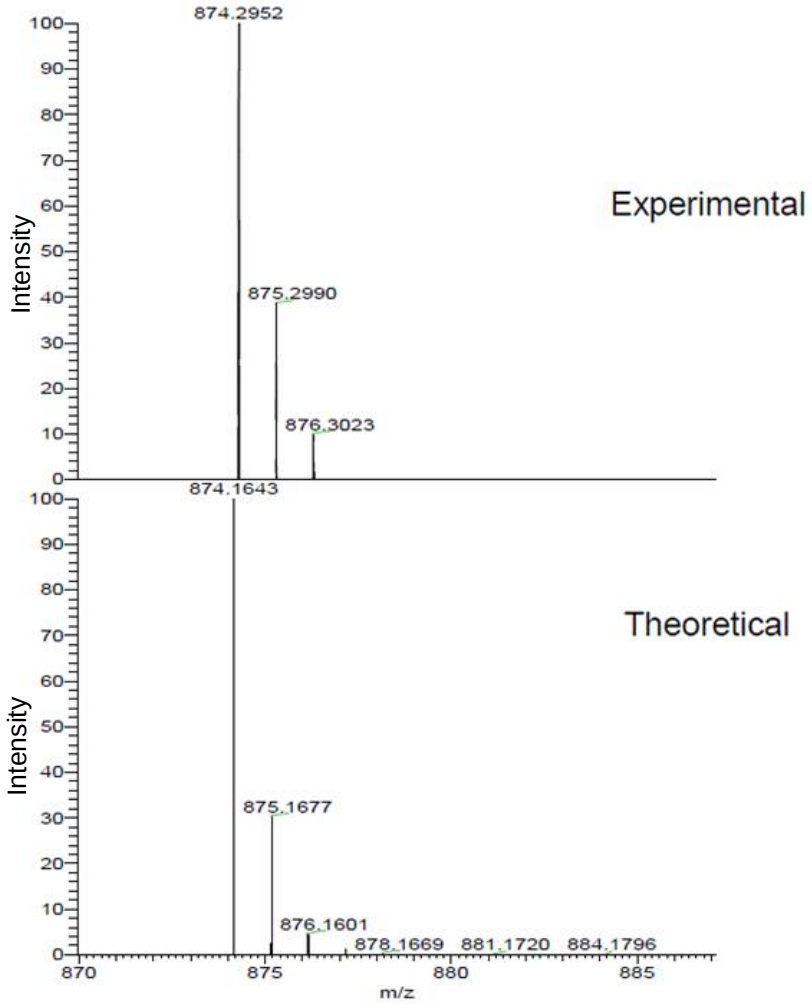



**Figure 6.20. HRMS (ES+) analysis of putative CH-1,3-eneC CoA.**

HRMS (ES+) analysis of an unknown compound collected through HPLC, putatively corresponding to CH-1, 3-eneC CoA ($c_2$). Theoretical [M+H] 874.1643, experimental *m/z of* 874.2952. Due to an observed error of >5 ppm this compound could not be conclusively identified and will need to be reanalysed. The compound was purified and prepared for analysis following the methods stated previously (chapter 2.30 & 2.35).

## 6.4. Identifying a promiscuous TE for implementation with the organic acid modification system

### 6.4.1. Activity screen of four TEs

The four TEs that were the focus of the previous chapter: FcbC, PA2801, RpaL and YbdB, were next screened with a range of commercially available CoA esters for the determination of a promiscuous TE for implementation with an organic acid production system (Figure 6.21).

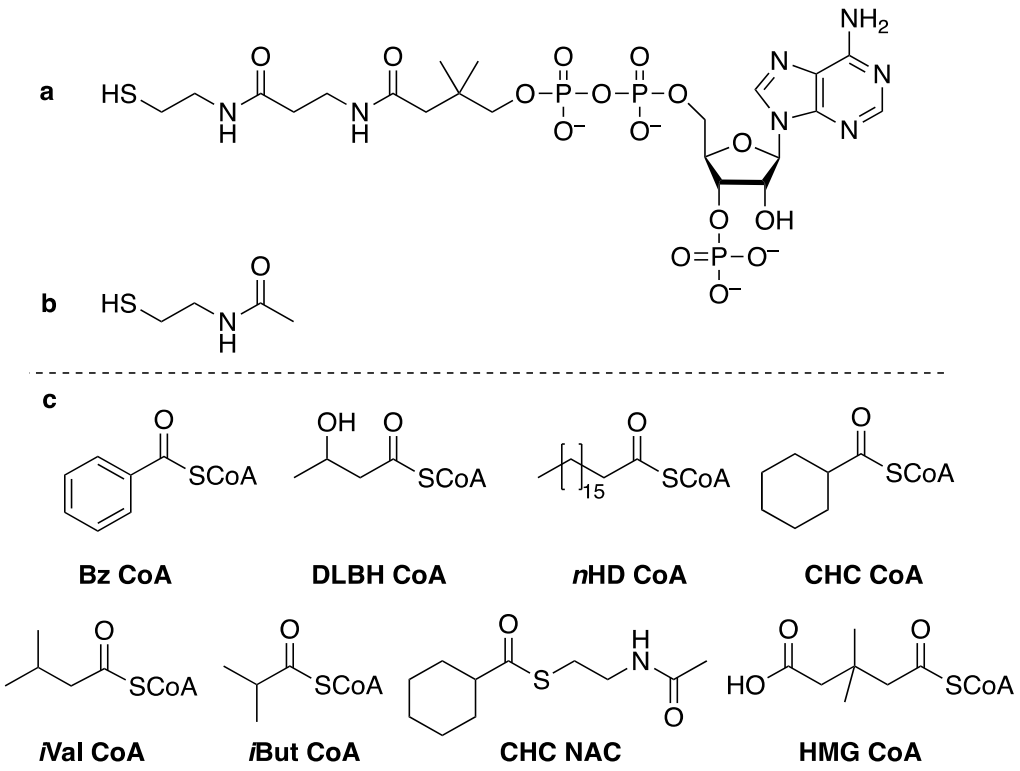



**Figure 6.21. The structure of CoA, NAC and all CoA ester used for screening TE activity.**

The chemical structure of (a) Coenzyme A (b) *N*-acetylcysteamine and (c) all CoA esters used as substrates for TE screening: Benzoyl CoA (Bz CoA), DL-β-Hydroxybutyryl CoA (DLBH CoA), *n*-heptadecanoyl CoA (*n*HD CoA), cyclohexanecarbonyl CoA (CHC CoA), isovaleryl CoA (*i*Val CoA), isobutyryl CoA (*i*But CoA), cyclohexanecarbonyl N-acetylcysteamine (CHC NAC) and β-Hydroxy β-methylglutaryl CoA (HMG CoA).

FcbC was found to have activity towards benzoyl CoA (Bz CoA) and the aliphatic substrate DL-β-Hydroxybutyryl CoA (DLBH CoA) (Figure 6.22). This result aligned with previous investigations which had looked into the activity of FcbC, and showed it to have activity towards a range of aromatic CoA esters, yet comparably little activity towards aliphatic CoA esters (Zhuang *et al.*, 2003).

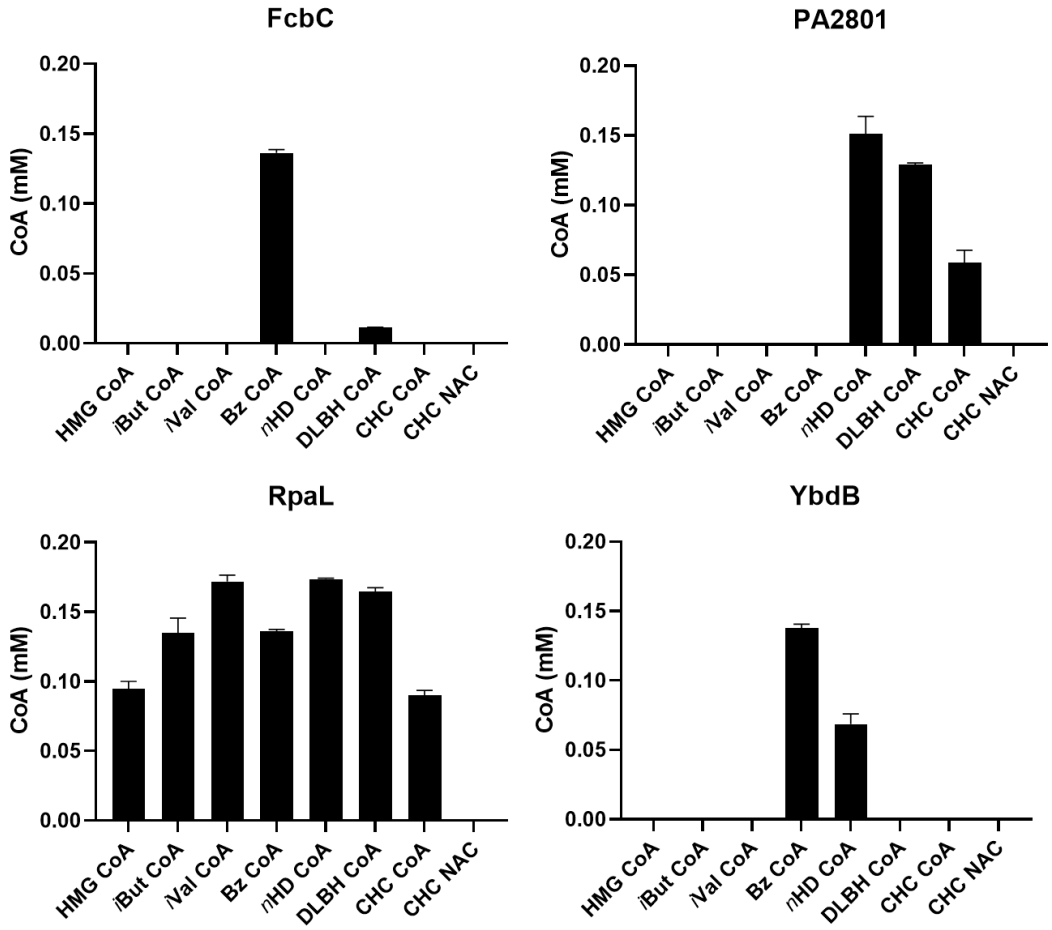



**Figure 6.22. Screening the activity of four TEs.**

The activity of each of the four recombinant TEs, as part of a clarified lysate, with a range of CoA esters and the synthetic CHC CoA mimic- CHC NAC. The TEs were assayed using the DTNB assay, which in the presence of a free thiol group (CoA) cleaves, releasing TNB, which is monitored at 405nm. All assays contained 0.2 mM CoA/NAC ester and were carried out according to the previously described method (chapter 2.25).

PA2801 was found to have activity towards the aliphatic CoA esters: DLBH CoA, *n*-heptadecanoyl CoA (*n*HD CoA) and the cyclic CoA ester, CHC CoA, however, no activity was observed with Bz CoA or either of the branched chain CoA esters, isovaleryl CoA (*i*Val CoA) and isobutyryl CoA (*i*But CoA). Interestingly, despite activity being found with CHC CoA, no activity was found with CHC NAC. RpaL was found to have activity towards all the CoA esters that were screened, including both aliphatic and aromatic CoA esters. RpaL, like all four TEs examined here, was unable to use CHC-NAC* despite having activity towards CHC CoA, possibly suggesting that hotdog fold TEs require the entire CoA to facilitate hydrolysis. YbdB was found to only have activity towards Bz CoA and *n*HD CoA, which aligned with a previous investigation showing it to have a preference towards aromatic and long chain fatty acyl CoA esters, with no activity towards short branched chain CoA esters such as β-methylcrotonyl-CoA (Latham *et al.*, 2014).

### 6.4.2. RpaL purification and size exclusion

As a result of the broad substrate specificity found within the clarified lysate of cells expressing *rpaL*, it was sub-cloned into pET28a (pQR2804), attaching an N-terminal His-Tag for Nickel column purification. Lysate containing expressed RpaL was loaded onto a Nickel column and RpaL was purified with 500 mM imidazole (Figure 6.23).

*Synthesised by Damien Baud and Laure Benhamou in the UCL Chemistry Department

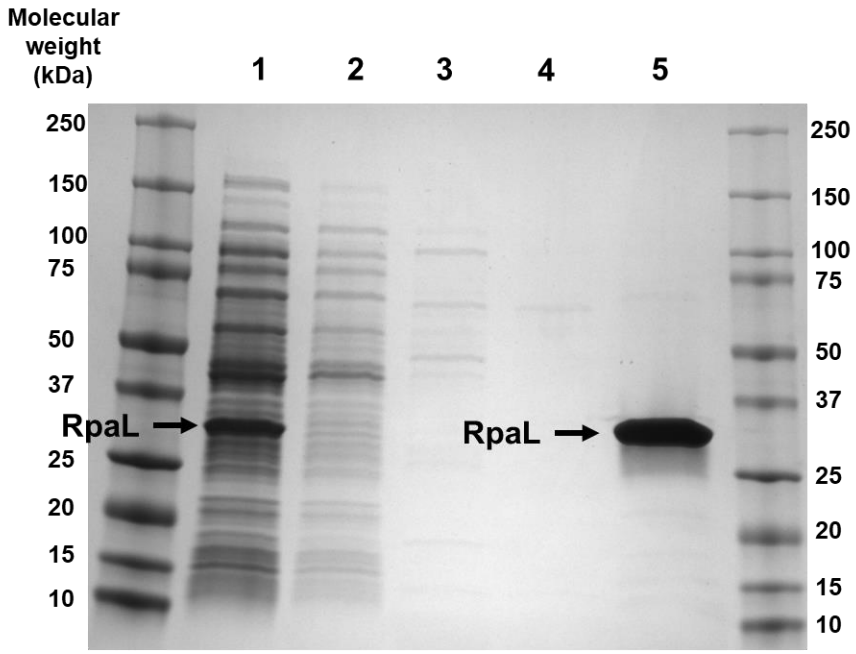



**Figure 6.23. The His-tag purification of RpaL (pQR2804).**

An SDS gel showing the results of the His-Tag purification of RpaL from soluble lysate using a nickel column. Lane 1 shows RpaL expressed within the soluble lysate. Lanes 2, 3 and 4 show protein eluted within the wash steps, containing 10 mM, 50 mM and 100 mM imidazole, respectively. Lane 5 shows purified RpaL eluted with 500 mM imidazole. 5 µl of the clarified lysate and from each elution fraction was prepared and loaded onto the gel as previously described (chapter 2.17). RpaL was expressed from *E.coli* BL21 (DE3) transformed with pQR2804, grown in 300 ml of TB media for 24 hours at 25 °C shaking at 250 rpm.

Size exclusion chromatography (SEC) was used to measure the size of purified RpaL against four standard proteins. RpaL was found to have a retention time of 6.09 minutes, indicating its size to be between 35-45 kDa, which would suggest its native biological conformation is a monomer (Figure 6.24). This differs with other members of the TesB-like subfamily where for example TesB, a medium chain acyl-CoA TE II from *E. coli* (1C8U), forms a dimer of double hot dog domains (Pidugu *et al.*, 2009).

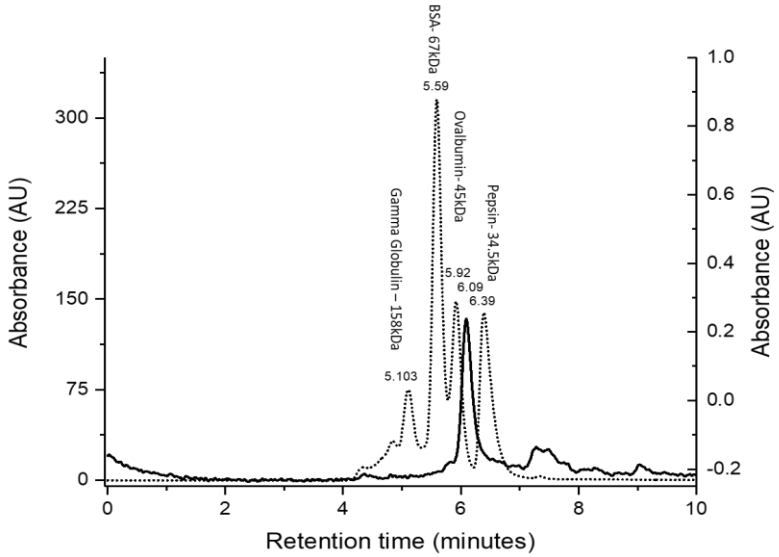



**Figure 6.24. SEC analysis of RpaL.**

Size exlcusion chromatography (SEC) results of RpaL and four standards. The dotted line indicates standards: 1-Gamma globulin (158 kDa) RT- 5.103, 2- BSA (67 kDa) RT- 5.59, 3- Ovalbumin RT- 5.92 (45 kDa) and 4- Pepsin RT- 6.39 (34.5 kDa). The solid line indicates RpaL, which had a retention time of 6.09 minutes. The experiment was performed according to the previously described method (chapter 2.29).

### 6.4.3. Steady state Kinetics of RpaL

The highest $k_{cat}/k_m$ ($1.6 \times 10^4$ $M^{-1}s^{-1}$) was found with DLBH CoA, with a similarly high specificity constant with *i*But CoA ($1.4 \times 10^4$ $M^{-1}s^{-1}$), suggesting a preference for short branched chain CoA esters (Table 6.1). However, a similarly high catalytic efficiency ($1.2 \times 10^4$ $M^{-1}s^{-1}$) was observed with Bz CoA, implying this high catalytic activity can occur on a range of structurally unrelated substrates. Slightly lower catalytic efficiencies were observed with *i*Val CoA and HMG CoA, suggesting that increasing the complexity of the CoA ester has a negative effect on enzyme activity. However, the lowest calculated $k_{cat}/k_m$, which was found with HMG CoA ($4.2 \times 10^3$ $M^{-1}s^{-1}$), was still found to be ~26% of the highest substrate, DLBH CoA. When determining the steady state kinetics of RpaL with *n*HD CoA, substrate inhibition was observed, with $V_0$ values increasing up to a substrate concentration of 0.3 mM, which was followed by a sharp $V_0$ decline. When attempting to fit the calculated $V_0$ values

with the modified Michaelis-Menten equation, V0 = $V_{max}$ [S]/ ($k_m$ + [S](1 + [S]/ $k_i$), a much steeper decrease in $V_0$ was observed than expected (Figure 6.25).

**Table 6.1. Steady state kinetic parameters with standard error for purified RpaL with CoA esters.**

| Substrate | $k_m$ (mM) | $v_{max}$ (μMs$^{-1}$) | $k_{cat}$ (s$^{-1}$) | $k_{cat}/k_m$(M$^{-1}$s$^{-1}$) |
|---|---|---|---|---|
| Isobutyryl CoA (iBut CoA) | 0.36 ± 0.16 | 1.22 ± 0.25 | 5.1 ± 1 | 1.4 x $10^4$ |
| Isovaleryl CoA (iVal CoA) | 0.11 ± 0.02 | 0.12 ± 0.01 | 0.51 ± 0.03 | 4.6 x $10^3$ |
| DL-β-Hydroxybutyryl CoA (DLBH CoA) | 0.24 ± 0.03 | 0.86 ± 0.04 | 3.74 ± 0.19 | 1.6 x $10^4$ |
| β-Hydroxy β-methylglutaryl CoA (HMG CoA) | 0.2 ± 0.05 | 0.2 ± 0.02 | 0.84 ± 0.1 | 4.2 x $10^3$ |
| Benzoyl CoA (Bz CoA) | 0.18 ± 0.02 | 0.5 ± 0.02 | 2.08 ± 0.07 | 1.2 x $10^4$ |

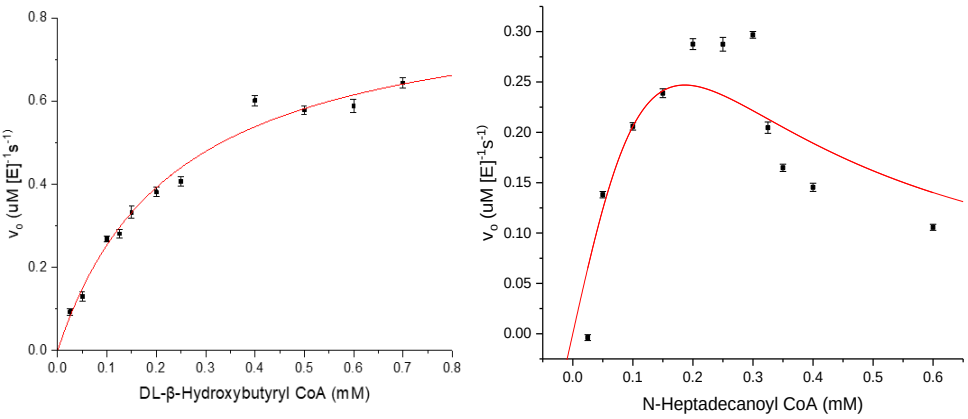



**Figure 6.25. Michaelis-Menten graphs of DL-β Hydroxybutyryl CoA and n-heptadecanoyl.**

Michaelis-Menten graphs of (a) DL-β Hydroxybutyryl CoA, which was the substrate with the highest specificity constant (1.6 x $10^4$ M$^{-1}$s$^{-1}$) and (b) *n*-heptadecanoyl, which showed exaggerated substrate inhibition compared to that predicted from the modified Michaelis-Menten equation.

## 6.5. A proof of principle for an organic acid modification system

### 6.5.1. Synthesis of CH-1-eneC from CHC

RpaL activity was then used with the CoA ester that had been generated via AliA activity followed by a modification step, as a proof of principle for the concept of this organic acid production system. As RpaL had previously been shown to consume CH-1-eneC CoA (see chapter 5), this activity was assessed with CH-1-eneC CoA that had been produced from CHC, by sequential AliA and ChCoADH activity. HPLC analysis of RpaL activity found an 85% reduction in the peak area corresponding to CH-1-eneC CoA and the formation of ~57 µM of CoA (Figure 6.26). Furthermore, HRMS (ES+) analysis of the reaction following RpaL activity confirmed the production of CH-1-eneC (Figure 6.27).

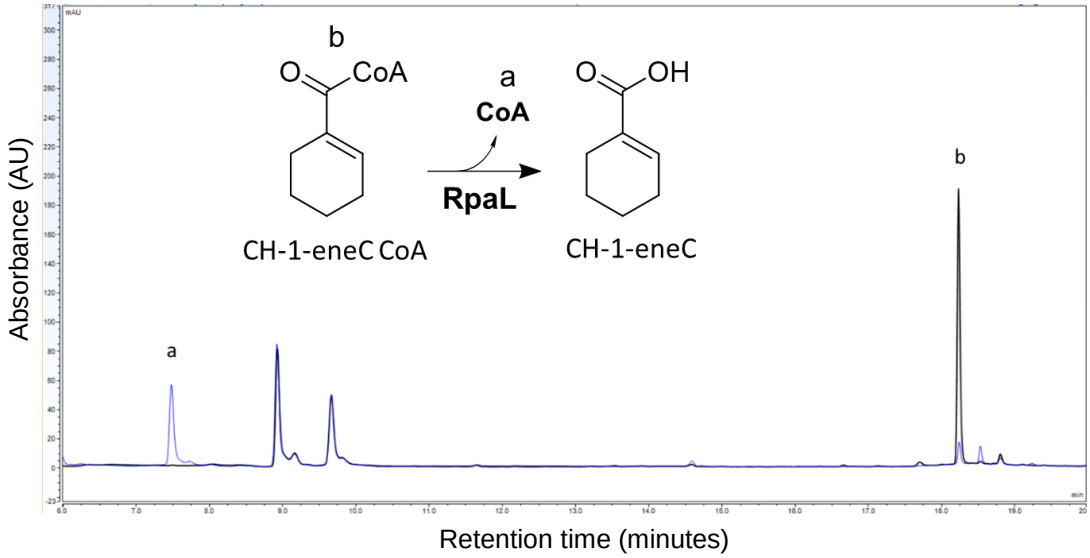



**Figure 6.26. HPLC analysis of TE activity with CH-1-eneC CoA derived from CHC.**

HPLC analysis of TE activity using CH-1-eneC CoA as the substrate immediately after (black line) and 30 minutes after (blue line) the addition of RpaL. (a) CoA, (b) CH-1-eneC CoA. CH-1-eneC CoA production from CHC, and RpaL activity were performed according to the method stated previously (chapter 2.28).

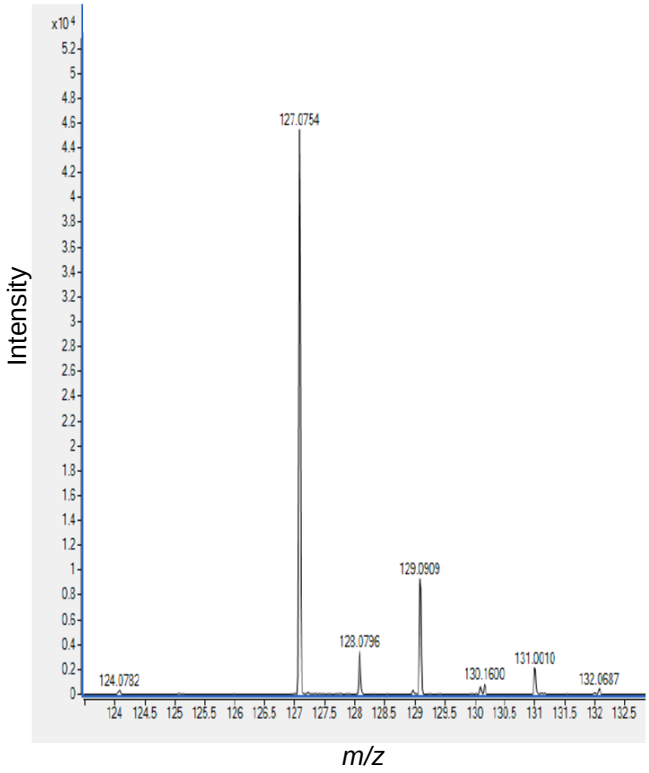



**Figure 6.27. HRMS (ES+) analysis of CH-1-eneC derived from CHC.**

HRMS (ES+) analysis of CH-1-eneC produced through sequential AliA, ChCoADH and RpaL activity. Theoretical [M+H] 127.0754, observed *m/z* 127.0754. An observed error of <5 ppm confirmed this compound to be CH-1-eneC. The compound was purified and prepared for analysis following method stated previously (chapter 2.35).

### 6.5.2. Synthesis of 2-HCHC from CHC

Next, as RpaL had been shown to also be able to produce 2-HCHC from 2-HCHC CoA derived from commercial CHC CoA, it was shown that 2-HCHC could be derived through sequential AliA, ChCoADH, BadK and RpaL activity using commercial CHC as the starting material. HPLC analysis determined a 75% reduction in the 2-HCHC CoA peak area following addition of purified RpaL (Figure 6.28). HRMS (ES-) analysis of the reaction mixture following RpaL activity confirmed the production of 2-HCHC (Figure 6.29).

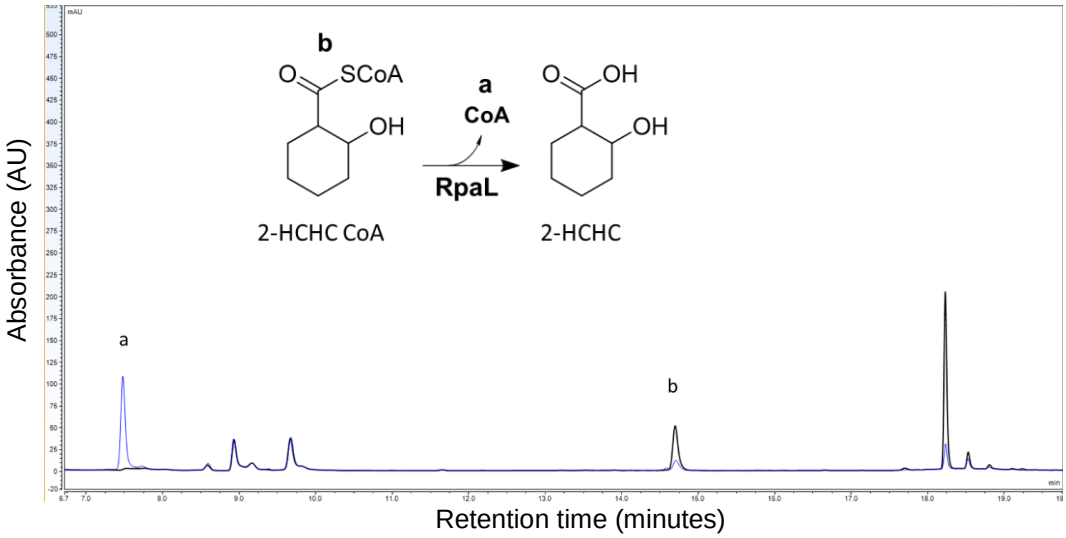



**Figure 6.28. HPLC analysis of TE activity with 2-HCHC CoA derived from CHC.**

HPLC analysis of RpaL thioesterase activity using 2-HCHC CoA, derived through AliA, ChCoADH and BadK activity, immediately after (black line) and 30 minutes after (blue line) the addition of RpaL. (a) CoA, (b) 2-HCHC CoA (c) CH-1-eneC CoA. 2-HCHC CoA production from CHC, and RpaL activity were performed according to the method stated previously (chapter 2.28).

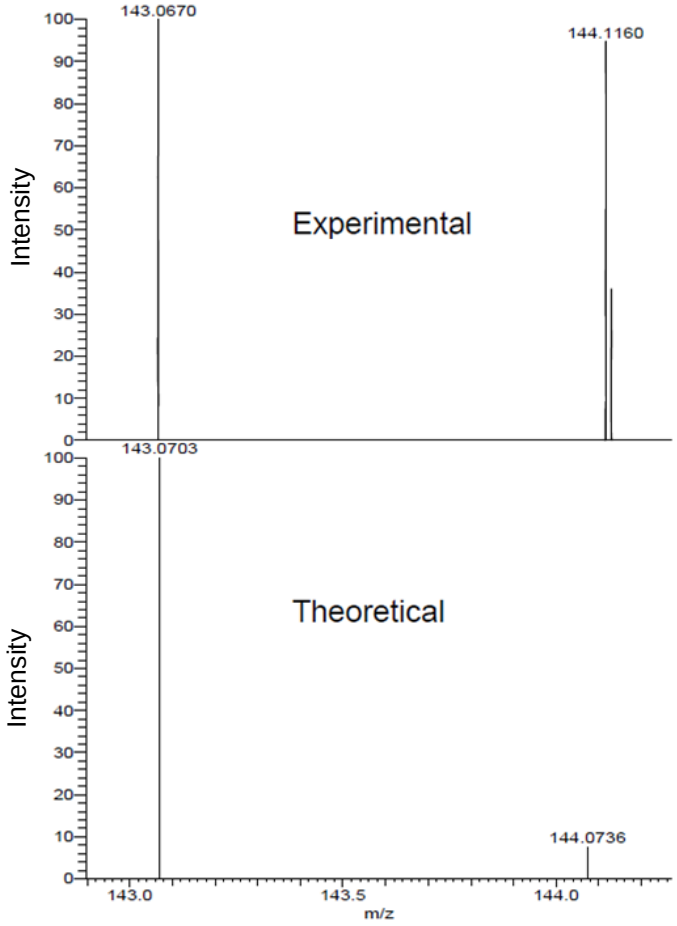



**Figure 6.29. HRMS (ES-) analysis of 2-HCHC derived from CHC.**

HRMS (ES-) analysis of 2-HCHC produced through sequential AliA, ChCoADH, BadK and RpaL activity. Theoretical $[M-H]^-$ 143.0703, observed *m/z* 143.0670. An observed error of <5 ppm confirmed this compound to be 2-HCHC. The compound was purified and prepared for analysis following method stated previously (chapter 2.35).

### 6.5.3. Synthesis of 4-HCH-1-eneC from 4-HCHC

RpaL activity was also assessed with the modified CoA ester 4-HCH-1-eneC CoA. Initial experiments showed that following incubation with RpaL a 49**%** decrease in the 4-HCH-1-eneC CoA peak area as well as an 84**%** decrease in 4-HCHC CoA peak area occurred. Unfortunately, due to issues associated with concentrating 4-HCH-1-eneC to a high enough concentration for HRMS analysis, the production of 4-HCH-1-eneC could not be confirmed. However, the results would putatively suggest that RpaL is able to hydrolyse 4-HCH-1-eneC CoA as well as 4-HCHC CoA into their corresponding organic acids (Figure 6.30).

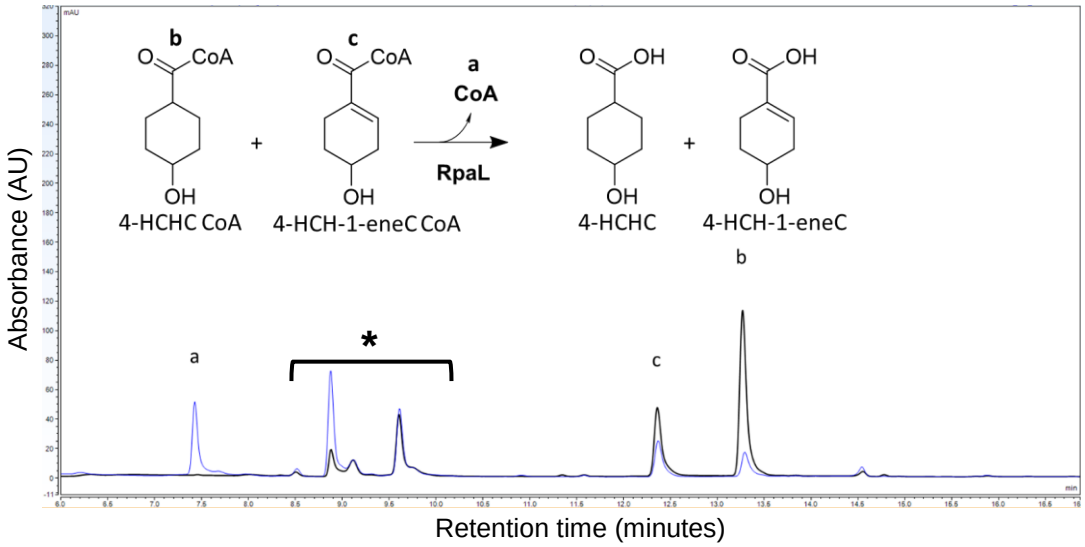



**Figure 6.30. HPLC analysis of TE activity with 4-HCH-1-eneC CoA.**

HPLC analysis of TE activity using 4-HCH-1-eneC CoA as the substrate immediately after (black line) and 30 minutes after (blue line) the addition of RpaL. (a) CoA (b) 4-HCHC CoA (c) 4-HCH-1-eneC CoA. The preceding production of 4-HCHC CoA and 4-HCH-1-eneC CoA from 4-HCHC, and subsequent RpaL activity (shown here) were performed according to the method stated previously (chapter 2.28). Peaks highlighted with an asterisk are considered unknown contaminant peaks as they have been found in a number of different samples analysed by the same HPLC program and equipment.

### 6.5.4. Synthesis of CH-1, 3-eneC from CH-3-eneC

Finally, CH-3-eneC CoA and CH-1, 3-eneC CoA derived from CH-3-eneC were used as substrates for RpaL activity to determine whether these CoA esters could be hydrolysed into their corresponding organic acids. HPLC analysis showed an 86% consumption of CH-3-eneC CoA, in addition to an 89% ($c_1$) and 90% ($c_2$) consumption in CH-1, 3-eneC CoA stereoisomers (Figure 6.31). Unfortunately, due to insufficient time, HRMS analysis was not used to confirm the production of these organic acids (CH-3-eneC and CH-1, 3-eneC). However, the observed consumption these CoA esters, following incubation with RpaL, suggests that the corresponding organic acids may have been produced.

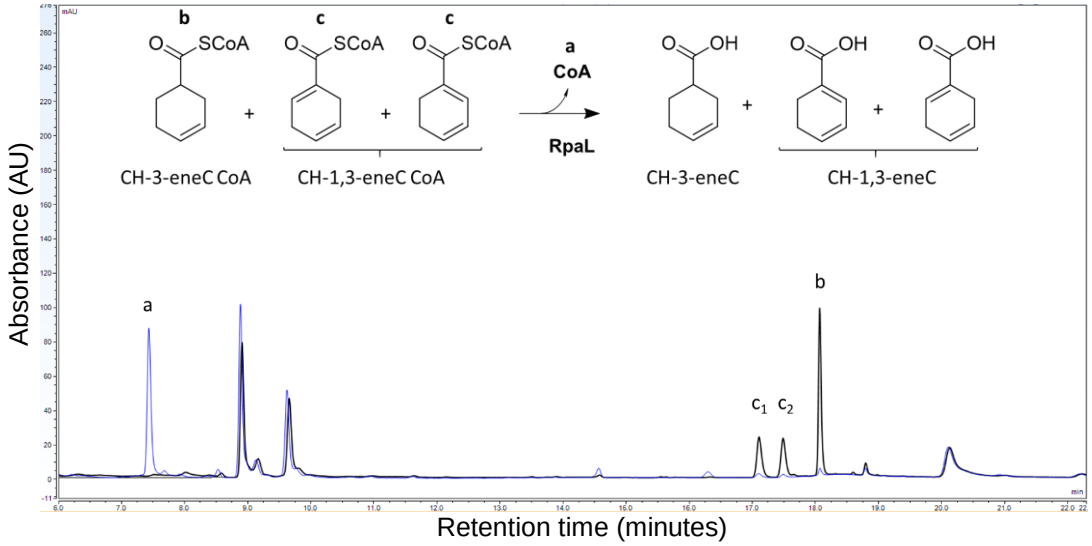



**Figure 6.31. HPLC analysis of TE activity with the putative compounds CH-3-eneC CoA and CH-1,3-eneC CoA.**

HPLC analysis of TE activity with the compounds putatively corresponding to (b) CH-3-eneC CoA and ($c_1$ and $c_2$) CH-3,1-eneC CoA, immediately after (black line) and 30 minutes after (blue line) the addition of RpaL. The CH-1, 3-eneC CoA peaks are labelled $c_1$ and $c_2$ to signify that each peak putatively represents a stereoisomers. The preceding production of CH-3-eneC CoA and CH-1,3-eneC CoA from CH-3-eneC, and subsequent RpaL activity (shown here) were performed according to the method stated previously (chapter 2.28).

## 6.6. Discussion

### 6.6.1. AliA activity preference

Screening AliA for CoA ligase activity with a range of alicyclic and aromatic organic acids, revealed a degree of promiscuity beyond its natural substrate CHC. Activity was found with derivatives of CHC including 3-HCHC, 4-HCHC, 4-oxoCHC and CH-3-eneC CoA. Interestingly, activity was also found with benzoate, suggesting AliA may have multiple functions within the metabolic degradation of benzoate, beyond just the CoA activation of CHC (Egland *et al.*, 1997). No activity was found with the more complex organic acids, such as shikimate, 3-phosphoshikimate and 3,5-dihydroxycyclohexanecarboxylic acid. Furthermore, no activity was found with 2-oxoCHC despite activity being found with 4-oxoCHC. This might suggest functional groups such as ketones, amines and possibly hydroxyls, at the 2' (ortho) position, inhibit activity due to steric reasons and their close proximity to the carboxylate. Moreover, the data may imply a slight degree of flexibility in functional groups at the 3 and 4' positions as activity was found with hydroxyl groups at both of these positions and a ketone at the 4' position. However, multiple hydroxyl groups, as with 3, 5-diCHC, shikimate and 3-phosphoshikimate showed no activity, clearly imply a preference for simple alicyclic organic acids. No linear aliphatic organic acids were screened with AliA, however, this would be an interesting avenue for future work, to further understand the extent of its substrate acceptance.

In the context of the biosynthetic organic acid modification system proposed in this chapter, the ideal candidate for the CoA 'activation' module, would be promiscuous allowing a plethora of organic acids to be activated into their corresponding CoA esters. With this in mind future work should consider screening additional CoA ligases or synthetically, through mutagenesis, expanding the substrate acceptance of AliA. In a recent investigation it was shown that rational mutation of amino acid residues within the carboxylate binding pocket of an acetyl CoA, conferred a preference switch from high

specificity towards acetate, to high specificity towards longer linear and branched chain organic acids (Sofeo *et al.*, 2019). A similar approach in future work could be applied to AliA for the development of mutants with the capacity to accept more complex organic acid substrates.

From a commercial and viability perspective, the practicality of using complex organic acids substrates in the biosynthetic organic acid modification system, invokes undesired caveats such as high costs and a possible lack of commercial availability of substrates. For example, the cost of CHC is £ 3.12/g (Merck, Germany), whereas 2-HCHC is £940/g (Merck, Germany) and 3-HCHC is £113/g (TCI UK Ltd. CAS: 606488-94-2). With this understanding the most viable organic acid modification system might use simple, inexpensive substrates for CoA activation. The desired complexity could then be achieved through the activity of CoA ester modification enzymes, rather than relying on pre-existing, chemically derived functionality. However, this will require continued research into developing and characterising a library of CoA ester modification enzymes for implementation.

### 6.6.2. Production of modified CoA esters

As mentioned throughout this investigation, the lack of commercially available CoA esters was a limitation to the characterisation of the enzymatic components of the biosynthetic cyclohexanol pathway. However, the combined activity of AliA, ChCoaDH and BadK lead to the synthesis of previously unavailable CoA esters, for use as both HPLC standards, as well as substrates for enzymatic assays. Specifically MycA2 activity (chapter 3) could not be screened due to both CH-1-eneC CoA and 3-HCH-1-eneC CoA not being commercially available. In this chapter it was shown that through the activity of three enzymes (AliA, ChCoADH, BadK) 12 different CoA esters were derived, showing the utility of this system. Given more time for further

substrate screening, and isolation of additional CoA ester modifying enzymes, this number could be significantly increased. The use of CoA esters is ubiquitous within the central and secondary metabolism of biological systems, with a plethora of enzymes acting on CoA esters in a range of interesting mechanisms. However, a constraint to their recombinant use has always been the limited commercial availability of CoA esters. Therefore, using a common biosynthetic system such as this to synthesis CoA esters could provide a straightforward method through which to derive CoA esters for the screening and characterisation of interesting CoA-dependent enzymes.

### 6.6.3. Aberrant TE activity with NAC thioesters

Due to the high cost and limited commercial availability of CoA esters, previous investigations have used *N*-acetylcysteamine (NAC) thioesters for screening TE activity. These thioesters contain the first part of the 4'-phosphopantetheine arm (Figure 6.21) of the CoA and so act as synthetic mimics of a CoA. Several investigations have shown NAC esters to be viable mimics of their CoA ester counterparts, specifically with TEs from the $\alpha/\beta$ hydrolase-fold superfamily (Korman *et al.*, 2010). However, there has been limited work assessing whether TEs from the hotdog fold superfamily can also use NAC esters as viable mimics. To gain insight, this study screened each TE with CHC-NAC as well as CHC CoA. Interestingly, the two TEs which were able to hydrolyse CHC CoA (PA2801 and RpaL) were unable to hydrolyse CHC-NAC, highlighting a stringent requirement for the entire CoA component, despite a promiscuous acceptance of CoA-activated organic acids.

Investigations into the structure and catalytic mechanism of hotdog fold TEs have shown there to be a conserved nucleotide binding site on the enzyme surface that binds to the 3'-phosphate (P) and 5'-pyrophosphate (PP) of the nucleotide moiety of the CoA (Song *et al.*, 2012; Wu *et al.*, 2014). In FcbC the

CoA 3'-P forms ions pairs with the side chains of Arg102 and Arg150 and the 5'-PP hydrogen bonds with Ser120 and Thr121. It was shown that the mutants FcbC T121A, R150A and R102A all resulted in significantly decreased $k_{cat}/K_m$, highlighting the importance of these interactions (Song *et al.*, 2012). As the NAC esters only mimic the first portion of the 4'-phosphopantetheine arm, none of these interactions between the surface of the protein and CoA nucleotide can occur. As such, this may explain why both PA2801 and RpaL could use CHC CoA as a substrate but not CHC NAC. Furthermore, with only a few exceptions, members of the α/β-hydrolase fold family predominantly target substrates bound to acyl carrier proteins (ACP) or peptidyl carrier proteins (PCP) (Cantu, Chen and Reilly, 2010). Here, substrates are covalently bound to a 4'-phosphopantetheine (PPT) arm, which lacks the nucleotide moiety of a CoA. This may suggest that interactions between the TE surface and substrate are not critical for hydrolysis, possibly explaining why NAC thioesters can be used as viable substrate mimics with α/β-hydrolase fold TEs and not Hotdog fold TEs.

### 6.6.4. The steady state kinetics determined for RpaL

RpaL was shown to be able to hydrolyse a variety of CoA esters including aromatics, short and long chain aliphatics and alicyclics. This ability to hydrolyse such a diverse range of CoA esters, with a high specificity constant, makes it an extremely desirable candidate for the final 'Termination' step of the proposed organic acid modification system. By implementing such a promiscuous TE within the termination step, this would mean that a plethora of CoA esters could be produced in the preceding 'activation' and 'modification' steps, which could all be hydrolysed by a single common enzymatic step, without the need for identifying a separate TE for each CoA ester.

Substrate saturation curves fitted with the Michaelis-Menten equation determined the highest specificity constant to be towards DLBH CoA, with decreasing specificity constants observed with substrates containing increased functional complexity. Furthermore, RpaL was able to hydrolyse the long-chain fatty acyl- N-heptadecanoyl CoA, at concentrations of 0.3mM, above which a dramatic decline in $V_0$ was observed that could not be fitted to the modified Michaelis-Menten equation. This may be a result of more complex substrate inhibition than can be modelled using the standard substrate model. As substrate inhibition was not observed with any of the other CoA esters it is also possibly due to the long acyl chain. In previous investigations that have screened hotdog fold TEs with fatty acyl CoA esters, substrate inhibition has also been observed with longer chain acyl-CoAs such as palmitoyl CoA and oleoyl CoA (Wei, Kang and Cohen, 2009). Here, it was suggested that micelle formation may result in apparent substrate inhibition, as it was shown that the $V_0$ of both substrates increased up to the critical micellar concentration (CMC), after which point the $V_0$ declined. It is possible that a similar effect has occurred in the current investigation and provides an explanation as to why a modified Michaelis-Menten equation does not fit the observed $V_0$ data.

### 6.6.5. The biosynthetic organic acid modification system and considerations for its use.

The work presented here focused on constructing and developing a proof of concept for a biosynthetic organic acid modification system. This system functions through the broad activity of a CoA ligase, activating a range of organic acids into corresponding CoA esters. Implementation of CoA ester modification enzymes would allow for bespoke, highly specific functionalisation of the CoA ester. Then, through the activity of the broad acting TE, RpaL, any modified CoA ester can be hydrolysed into the final modified organic acid. Importantly, promiscuity with the activity of AliA and RpaL facilitates a variety of organic acids to both enter (CoA activation) and

exit (CoA ester hydrolysis) the biosynthetic system, leading to the production of a number of modified organic acids

As a proof of concept for this system, the combined activity of AliA, ChCoADH, BadK and RpaL enabled the production CH-1-eneC and 2-HCHC (from CHC), and putative production of 4-HCH-1-eneC (from 4-HCHC) and both stereoisomers of CH-1, 3-eneC (from CH-3-eneC). It is clear there is an exciting potential for the application of this novel biosynthetic system, for the production of modified organic acids. However, it is acknowledged that this work has only provided relatively limited characterisation of modification possible within this system, primarily due to the small number of isolated enzymes for CoA ester modification.

As mentioned previously there are a vast number of CoA ester utilising enzymes, functioning throughout primary and secondary metabolism, from many organisms (Peter *et al.*, 2016). The diversity in these CoA ester utilising enzymes includes, but is not limited to: aminotransferases (Perret *et al.*, 2011), oxygenases (Zaar *et al.*, 2004; Rather *et al.*, 2010; Teufel *et al.*, 2010), reductases (e.g. MycA2- this study), dehydrogenases (e.g. ChCoADH- used in this study), isomerases (Patton, Cropp and Reynolds, 2000), hydroxylases (Shimizu, 2014) and a number of acyl transferases (Strauss, 2010). Furthermore, ongoing scientific research in this area will inevitably lead to the discovery and further characterisation of known and novel CoA ester-utilising enzymes. This will provide further opportunities for using these enzymes for the modification and production of novel CoA esters, and following RpaL activity, novel organic acids.

In the first chapter of this thesis salicylic acid was shown as amongst the first organic acids to be discovered to have therapeutic effects and to be targeted for chemical synthesis. Therefore, using the organic acid modification system described here combined with available literature, a route has been proposed

for its production. This route would utilise the characterised CoA activation of benzoic acid to benzoyl CoA, then following isolation and recombinant expression, screen the *ortho*-acyl CoA hydroxylase (RgC2'H, AtF6'H1, Ib1 and Ib2), previously shown to *ortho*-hydroxylate the aromatic ring of *p*-coumaroyl CoA and Feruloyl CoA (Shimizu, 2014). If *ortho*-hydroxylation of benzoyl CoA were found, forming salicyloyl CoA, then RpaL activity may then be able to hydrolyse it into salicylic acid (Figure 6.32). This is only a proposed method, as only the first step has been experimentally characterised, yet, it highlights potential of using this organic acid modification system.

Despite the potential of this biosynthetic system, there are a number of factors that will need to be considered regarding the method of use. Firstly CoA and ATP are essential substrates for AliA activity, in addition to the organic acid. Therefore, regeneration of these compounds would be essential for the viability of this system. ATP is converted into AMP and pyrophosphate ($PP_i$) following AliA activity, so a method for ATP regeneration from AMP would be required. This may be achieved through the combined activity of polyphosphate-AMP phosphotransferase (PAP), converting AMP into ADP and polyphosphate kinase (PPK), converting ADP to ATP (Kameda *et al.*, 2001; Shiba *et al.*, 2005). Conveniently, the activity of RpaL within the termination module will result in the regeneration of CoA, however, despite this convenience, due to the promiscuous activity of RpaL, its implementation method will need consideration. As mentioned in the previous chapter (chapter 5), when considering RpaL activity in the context of the biosynthetic cyclohexanol pathway, if RpaL were to be added at the same time as AliA and the specific CoA ester modification enzymes, than its activity would compete with all of the preceding enzymes. Therefore, a mechanism through which to effectively utilise the activity of RpaL is required.

One possible method, relevant to an *in vitro* system, would utilise the His-tag removal or RpaL from the system following its activity. Specifically, all enzymes used within this system would need to be designed to contain a

thrombin site between the gene ORF and the His-tag, this could be achieved using the pET28a vector. Following purification of each enzyme, thrombin could then be used to remove the synthetic His-tag due to the cleavage site in between the 6-His and protein provided by the vector. This would only need to be done once with each purified enzyme, prior to storage. Alternatively, RpaL would not have its His-tag removed. The components of the biosynthetic organic acid modification system could be added together, combined with required cofactor regeneration systems (such as PAP and PPK for ATP regeneration), resulting in the accumulation of the final modified CoA ester (step 1). Then, RpaL could be added to the system, resulting in the hydrolysis of the specific CoA ester into its corresponding modified organic acid (step 2) (Figure 6.32). Following complete conversion of the CoA ester into organic acid and free CoA, RpaL could then be specifically removed by washing the reaction through a nickel column. This implementation mechanism would significantly increase the viability of the biosynthetic organic acid modification system as CoA would not need to be constantly added to the system. With this understanding, the work presented here, combined with the proposed implementation method may provide the foundation for future research to further explore and characterise this biosynthetic organic acid modification system.

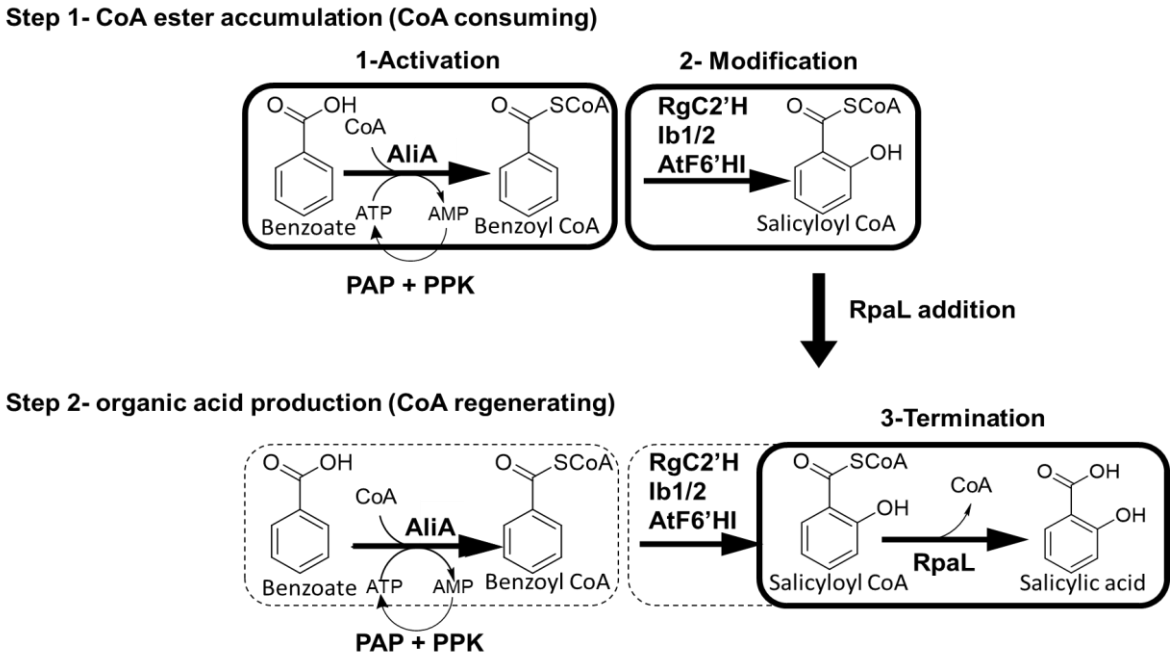



**Figure 6.32. An implementation method for the organic acid modification system**

Step 1- (activation module) AliA and all CoA ester modification enzymes (modification module) would be used for the production of the modified CoA ester (salicyloyl CoA). PAP and PPK would be used to regenerate ATP from AMP, facilitating final CoA ester accumulation. Step 2 – Following accumulation of modified CoA ester RpaL would be added for hydrolysis of the CoA ester into the modified CoA ester (salicylic acid) and the regeneration of CoA. Example shown uses characterised AliA-derived Benzoate activation to Benzoyl CoA (activation), for CoA modification a proposed step has been shown using either of the acyl-CoA *ortho*-hydroxylases: RgC2'H, IB1/2 or AtF6'HI, for the production of salicyloyl CoA, however this is an unknown biotransformation. Finally, RpaL would hydrolyse salicyloyl CoA into salicylic acid in the final 'termination' module.

# 7. Chapter 7. Discussion and future work

## 7.1. An overview of work achieved

The work undertaken within this investigation, had the primary aim of designing and constructing a petroleum-independent, biosynthetic link from central metabolism to cyclohexanol. In doing so, the potential sustainability of a previously developed biosynthetic pathway, converting cyclohexanol to 6-ACA, would be improved. A novel pathway was proposed, that used either shikimate or 3-phosphshikimate for the production of cyclohexanol, through 12 enzymatic biotransformations (Figure 7.1). Due to a number of caveats associated with the proposed pathway, the production of cyclohexanol was not fully realised. However, in the process of attempting to bypass these caveats, a novel biosynthetic system for the production of modified organic acids, via CoA ester intermediates, was proposed and tested.

One of the primary challenges of the cyclohexanol pathway was that none of the CoA ester pathway intermediates were commercially available. However, as shikimate and 3-phosphoshikimate were commercially available, it was thought this issue could be bypassed by characterising each enzyme sequentially, in the order of its proposed function. In this way, characterising the activity of the first enzyme within the cyclohexanol pathway, MycA4, was critical for the subsequent characterisation of all downstream enzymes. Yet, information on the biotransformation catalysed by MycA4 was extremely limited and no homolog had been previously characterised. Using the function of MycA4 that had been previously proposed, but not experimentally proven, it was assessed for hydratase/ phosphatase activity (Cropp, Wilson and Reynolds, 2000). However, despite assessing a range of experimental conditions no activity could be determined (chapter 3, Figure 3.20 & 21). As a result, this imparted a further limitation on characterising the activity of the downstream enzymes.

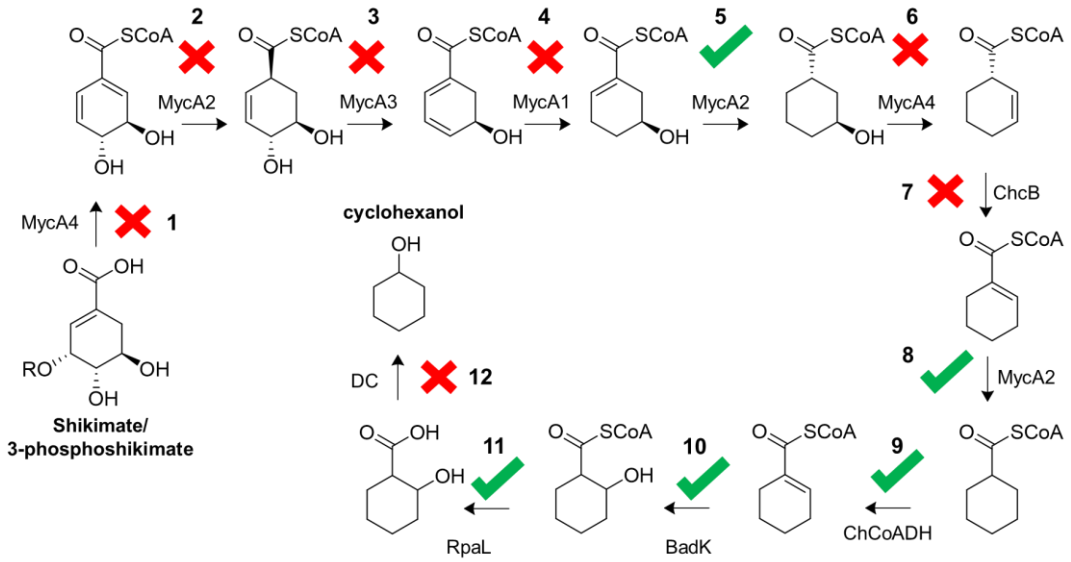



**Figure 7.1. The cyclohexanol pathway and the enzymatic steps successfully characterised in this investigation.**

Five of the total twelve steps (5, 8, 9, 10, and 11) were experimentally determined (green ticks) and seven remain undetermined (red cross). Analysis of step 5 was achieved using the combined activity of AliA and ChCoADH, for the production of 3-HCH-1-eneC CoA. MycA4 activity was assessed but no activity was found (step 1 and 6). CoA ester substrates required for step 2, 3, 4, 6 and 7 were not commercially available and so will require CoA ester production through the developed CoA modification system.

During this investigation CHC CoA was initially chemically synthesised and later became commercially available, which facilitated the characterisation of the enzymes downstream of the CHC CoA intermediate. As a result, the successful biotransformation of CHC CoA to 2-HCHC was determined by sequential activity of ChCoADH, BadK and RpaL (Figure 7.1, step 9-11) (see chapter 5, Figure 5.11). However, the limited amount of CHC CoA that could be chemically synthesised, combined with high cost and delivery time of purchasing commercial CHC CoA, lead to the decision to isolate the gene encoding the CoA ligase, AliA, from *R. pseudopalustris*. This CoA ligase had not been previously characterised, however, based on investigations on a homolog from *R. palustris*, it was proposed to catalyse the first step of the CHC degradation pathway; the CoA activation of CHC into CHC CoA (chapter 1, Figure 1.9.) (Pelletier and Harwood, 2000). The use of this recombinant enzyme for the production of CHC CoA was found to be a valuable alternative

production route, cutting out the requirement of chemical synthesis and expensive commercial procurement. CHC CoA produced through AliA activity was successfully used by ChCoADH and BadK in the production of 2-HCHC CoA (chapter 6, Figure 6.7).

As AliA had successfully produced CHC CoA for subsequent ChCoADH and BadK activity, it was assessed whether other CoA ester intermediates of the cyclohexanol pathway could be derived in this manner. The MycA2 substrate, CH-1-eneC CoA, had already been successfully derived through ChCoADH activity (chapter 6, Figure 6.7). Therefore, CH-1-eneC CoA produced in this way was used to screen MycA2 activity, confirming the reduction of CH-1-eneC CoA to CHC CoA (chapter 3, Figure 3.23). A degree of promiscuity in the substrate acceptance of AliA facilitated the production of a range of CoA esters, including another intermediate of the CHC CoA pathway, 3-HCHC CoA (chapter 6, Figure 6.5). This compound was the sixth compound of the CHC CoA pathway, and the second proposed substrate of MycA4. Production of this CoA ester allowed further preliminary screening of MycA4, however, no hydratase activity was found (chapter 3). Despite the inactivity observed with MycA4, it was assessed whether ChCoADH activity could be used to oxidise 3-HCHC CoA into the substrate of MycA2, 3-HCH-1-eneC CoA. This step was found to be successful and resultantly facilitated the assessment of MycA2 reductase activity with 3-HCH-1-eneC CoA, which was reduced to 3-HCHC CoA (chapter 3, Figure 3.26 & 3.27).

AliA activity combined with ChCoADH dehydrogenase activity had been found to facilitate the production of a range of CoA esters that were not commercially available and were required for enzymatic screening. However, it was further proposed that this function could form the basis of a biological system for the production of modified organic acids. The scientific literature shows a plethora of CoA utilising enzymes capable of functionalising CoA esters with a myriad of different functional groups (Peter *et al.*, 2016). Therefore, the coupled use of AliA, selected CoA ester modifying enzymes and finally, a promiscuous TE,

could theoretically, derive a myriad of rationally modified organic acids. For this final TE hydrolysis step (termination), four thioesterases were assessed with a range of commercially available CoA esters, which revealed the inherent substrate promiscuity of RpaL (chapter 6, Figure 6.22). The principle of this system was assessed and the combined activity of AliA, ChCoADH, BadK and RpaL was found to produce the modified organic acids: 2-HCHC and CH-1-eneC from CHC (chapter 6, Figure 6.29 & 6.27, respectively). Furthermore, this system was shown to putatively produce 4-HCH-1-eneC, and two stereoisomers of CH-1, 3-eneC, however, these will require further HRMS analytics to confirm their production. These proof of principle experiments have given a brief insight into the potential applications of this system for the production of both CoA esters and modified organic acids. As more CoA ester-utilising enzymes are obtained and characterised, the product portfolio of CoA esters will further expand. This will undoubtedly benefit the characterisation of CoA-utilising enzymes, such as those within the cyclohexanol pathway, as it will circumvent the availability issues associated with CoA esters. Furthermore, with an expanding possibility of CoA esters, the implementation of a broad-spectrum TE, such as RpaL, would enable the hydrolysis of these CoA esters into their corresponding organic acids, which through their rational design, may be of commercial interest.

Overall, this investigation has developed a biosynthetic pathway for the production of cyclohexanol from an intermediate of central metabolism. Despite only three of the twelve intermediates being commercially available, 5 of the enzymatic biotransformation were characterised. This characterisation was aided by the combined activity of AliA and ChCoADH, which facilitated the production of unavailable CoA ester intermediates. Finally, a three module, organic acid modification system was proposed and developed, using the combined activity of AliA, CoA ester modifying enzymes (such as ChCoADH) and the promiscuous TE, RpaL. Although limited insight has been given here, future work may build upon this system for both the rational production of CoA esters, required for enzymatic screening, and modified organic acids with potential commercial interest.

## 7.2. Considerations for future investigations into the cyclohexanol pathway

Future investigations focused on developing the cyclohexanol pathway should utilise components of the organic acid modification system developed here, for the production of the commercially unavailable CoA esters intermediates. For example, cyclohex-2-enecarboxylate (CH-2-eneC) is commercially available and so should be screened with AliA for the production of CH-2-eneC CoA. If successful this would allow both MycA2 and ChcB (once obtained) to be screened with this CoA ester to determine whether the activity of both enzymes is required for CHC CoA production (Figure 7.2).

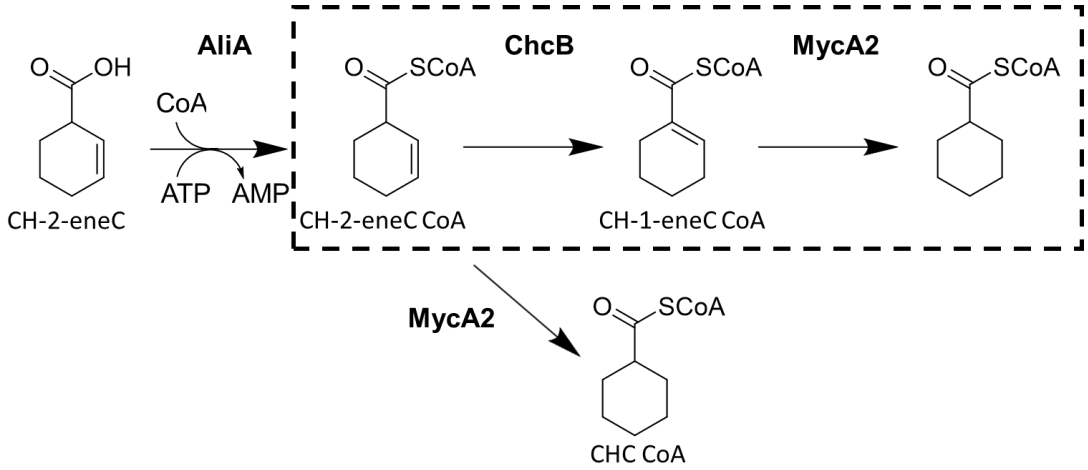



**Figure 7.2. CH-2-eneC CoA production, via AliA activity, for screening ChcB/ MycA2 activity.**

A proposed production route of CH-2-eneC CoA, using AliA activity with CH-2-eneC. Following the production of CH-2-eneC CoA, it could be used to screen MycA2 activity as well as ChcB. If MycA2 was found to be active with CH-2-eneC CoA, producing CHC CoA, then ChcB would not be required in the cyclohexanol pathway. The dashed box indicates the current proposed route to CHC CoA within the cyclohexanol pathway.

Due to corresponding organic acids not being commercially available, direct CoA ligase activity would not be a viable production route for the CoA ester intermediates of the second, third and fourth steps in the cyclohexanol pathway (Figure 7.1). However, a route for the production of the fourth CoA ester intermediate, 5-hydroxycyclohex-1, 3-enecarbonyl CoA (5-HCH-1, 3-

eneC CoA), would be to use the combined activity of AliA and ChCoADH with 3-HCHC, for the production of 3-HCH-1-eneC CoA (chapter 6, Figure 6.12). This is the proposed product of MycA1 activity, which may be a viable substrate for assessment of MycA1 activity in the reverse direction, forming 5-HCH-1, 3-eneC CoA (Figure 7.3). Alternatively, the commercially available organic acid, 5-hydroxycyclohex-3-enecarboxylate (5-HCH-3-eneC), could be screened with AliA activity, for the production of its corresponding CoA ester, 5-hydroxycyclohex-3-enecarbonyl CoA (5-HCH-3-eneC CoA). This CoA ester could then be screened with ChCoADH activity, for the production of 5-HCH-1, 3-eneC CoA, the substrate of MycA1 (Figure 7.3). If successful, this CoA ester could be used to screen the reductase activity of MycA1.

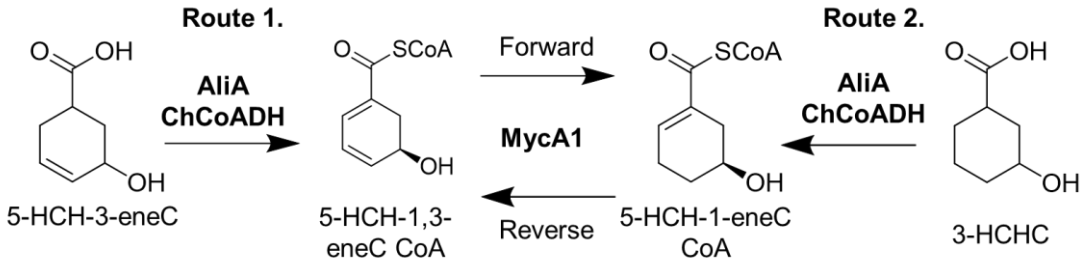



**Figure 7.3. Two proposed methods for screening MycA1 activity.**

Route 1 uses the combined activity of AliA and ChCoADH for the production of 5-HCH-1, 3-eneC CoA, the proposed substrate of MycA1, from 5-HCH-3-eneC. Route 2 uses combined activity of AliA and ChCoADH for the production of 5-HCH-1-eneC CoA, the proposed product of MycA1. Depending on which route is used MycA1 activity can then be screened in its forward/reduction (route 1) or reverse/oxidation (route 2) direction.

Furthermore, 5-H-CH-1, 3-eneC CoA could be used to assess the reverse reaction of MycA3 (Figure 7.4). If successful, the product, 4, 5-dihydroxycyclohex-2-enecarbonyl CoA (4, 5-diHCH-2-eneC CoA), could be used to assess the reverse activity of MycA2 (Figure 7.4) for the production of the second CoA ester, 3,4-dihydroxycyclohex-1,5-enecarbonyl CoA (3,4-diHCH-1,5-eneC CoA), in the cyclohexanol pathway. As MycA2 reductase activity has already been identified with CH-1-eneC CoA and 3-HCH-1-eneC

CoA (Figure 3.23 and 3.26, respectively), its reverse activity should firstly be assessed, to confirm the reversibility of its activity. Once confirmed, the reverse activity of MycA2 could be viably assessed with 4, 5-diHCH-2-eneC CoA, produced through the reverse MycA3 activity (Figure 7.4). All these examples are assuming the lack of MycA4 activity, however, if through the proposed methods in chapter 3 (see chapter 3.5.2), MycA4 activity could be achieved, the product of its activity could be used for screening the activity of downstream enzymes.

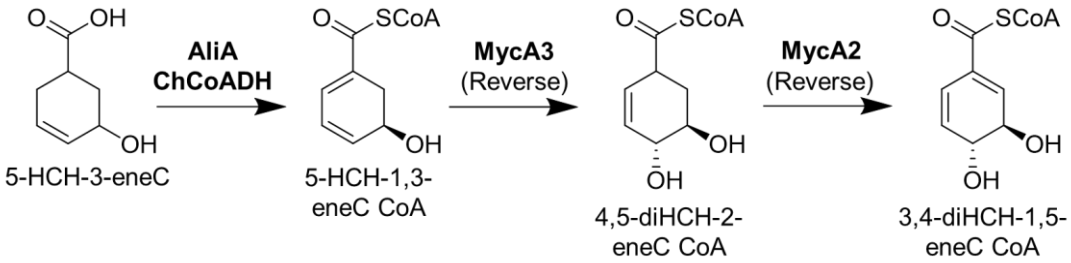



**Figure 7.4. A proposed production route for 5-hydroxycyclohex-1, 3-enecarbonyl CoA (5-HCH-1, 3-eneC CoA), the product of MycA3 activity.**

Production of this CoA ester, through sequential AliA and ChCoADH activity with commercially available 5-hydroxycyclohex-3-enecarboxylate (5-HCH-3-eneC), would allow the reverse activity of MycA3 to be assessed. Furthermore, if this activity successfully produced 4,5-dihydroxycyclohex-2-enecarbonyl CoA (4, 5-diHCH-3-eneC CoA), it could be used to screen the reverse reaction of MycA2, for the production of 3,4-dihydroxycyclohex-1, 5-enecarbonyl CoA (3,4-diHCH-1,5-eneC CoA).

Finally, as previously mentioned (chapter 1.7.4) DC activity was not assessed within this work. Unfortunately, there are no examples in the literature of a DC that can use either 2-HCHC or 2-oxoCHC as a substrate. Therefore, future work will need to identify a number of DCs that have been shown to have analogous activity to the desired DC step (Figure 1.11) and then screen their activity with 2-HCHC and 2-oxoCHC. Ideally this would lead to the identification of a DC that can use 2-HCHC for the production of cyclohexanol. Alternatively, if DC activity was found with 2-oxoCHC and not 2-HCHC, future work could make use of ‘route 2’ of the cyclohexanol pathway (see chapter 4, Figure 4.2), for the production of 2-oxoCHC CoA, which following TE activity would produce 2-oxoCHC. However, use of route 2 would require additional

work to address the product inhibition experienced by BadH when converting 2-HCHC CoA into 2-oxoCHC CoA. Additionally, 2-oxoCHC CoA would need to be screened with TE activity for the production of 2-oxoCHC. Conveniently, these two requirements may be complementary to one another. If a TE could be found to have highly specific activity with 2-oxoCHC CoA, this could limit the product inhibition of 2-oxoCHC CoA on BadH. In fact, enzymatic removal of 2-oxoCHC CoA has previously shown to reinstate BadH dehydrogenase activity with 2-HCHC CoA (Pelletier and Harwood, 2000). Moreover, identification of highly specific TE activity towards 2-oxoCHC CoA would also limit the undesired hydrolysis of preceding CoA ester intermediates within the cyclohexanol pathway.

### 7.3. Considerations for redesigning the cyclohexanol pathway

Following the exploration and development of the proposed cyclohexanol pathway, a number of design aspects represent areas for consideration and potential improvement. Firstly, the intermediate, 3-HCHC CoA, was initially considered as a possible target for TE activity (see chapter 1.7.1). However, as MycA4 was also proposed to use this intermediate, but could not be removed due to its requirement in the first step, this route was not experimentally explored. Instead, the cyclohexanol pathway was designed to use the entire CHC CoA pathway for the production of CHC CoA, which was then hydroxylated through sequential ChCoADH and BadK activity. The derived CoA ester, 2-HCHC CoA, then represented the alternative substrate for TE activity. This pathway required a total of 12 biotransformations, catalysed by nine enzymes. Alternatively, the pathway via 3-HCHC CoA TE hydrolysis, would only require 7 biotransformations catalysed by 6 recombinant enzymes (Figure 7.5). In a number of pathway engineering investigations there has been the observation of intermediate build ups, preventing optimal flux towards the product (Brockman and Prather, 2015; Lechner, Brunk and Keasling, 2016; Turk *et al.*, 2016). Furthermore, the overexpression of a large number of enzymes can impose a metabolic strain on the host organism (Pasini *et al.*, 2016). Therefore, identifying the shortest

metabolic route to a product would in theory, limit these issues. With this in mind, despite the unwanted second activity of MycA4, by containing fewer catalytic steps and requiring the expression of three less enzymes, the pathway via TE hydrolysis of 3-HCHC CoA, may in fact be a viable alternative route. Future work should assess this route by screening the activity of all four TEs isolated in this work with 3-HCHC CoA, derived through AliA activity. Additionally, it is critical that MycA4 activity is assessed, as its designated biotransformations are only proposed and have not been experimentally shown. In a previous investigation, which was focused on the CHC CoA operon from *S. collinus*, ChcB, an enzyme not encoded within the CHC CoA operon was required for the catalysis of an isomerisation step within the CHC CoA pathway (Patton, Cropp and Reynolds, 2000). Therefore, it is possible that the second proposed function of MycA4 is actually catalysed by an enzyme other than MycA1-4. If this is the case, then the alternative route, via TE hydrolysis of 3H-CHC CoA, would represent a shorter, and possibly more viable route to cyclohexanol than the route developed in this investigation (Figure 7.5).

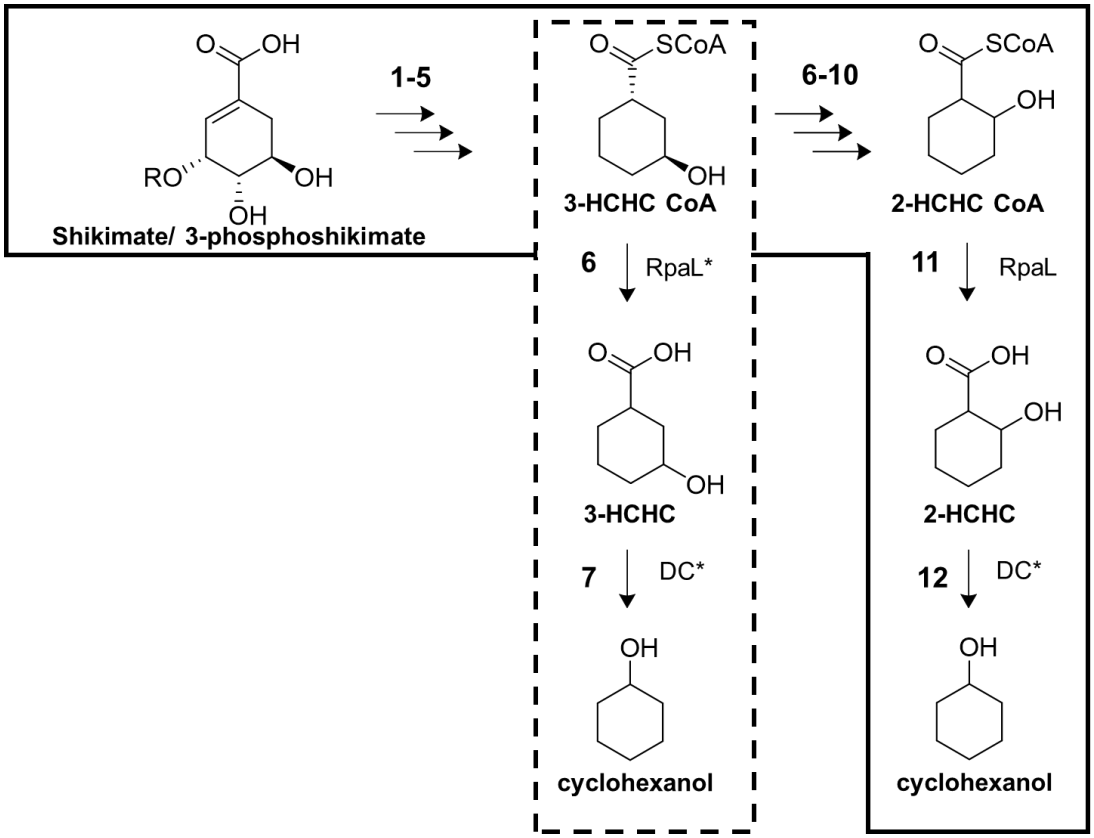



**Figure 7.5. Two possible routes to cyclohexanol from shikimate/ 3-phosphoshikimate.**

The route developed within this investigation required twelve biotransformations and employed TE hydrolysis with 2-HCHC CoA (outlined in black bold line). The alternative route (dashed black line) would employ TE activity with the CoA ester intermediate, 3-HCHC CoA, which following decarboxylase (DC) activity, would form cyclohexanol in seven total biotransformations. Astericks represent steps that have not been experimentally shown.

Furthermore, in the cyclohexanol pathway developed here, CH-1-eneC CoA is proposed to be produced by ChcB. Ideally, CH-1-eneC CoA could be directly hydrated by BadK, forming 2-HCHC CoA. However, MycA2 activity competes with BadK for the reduction of CH-1-eneC CoA to CHC CoA. Like with MycA4, MycA2 catalyses essential preceding steps within the pathway and so cannot be removed. Therefore, ChCoADH activity was required to oxidise CHC CoA back to CH-1-eneC CoA, to prevent the build-up of the dead-end intermediate, CHC CoA. Although ChCoADH activity, theoretically, restricts the accumulation of CHC CoA, its introduction means the requirement of two

additional biotransformation steps (sequential CH-1-eneC CoA reduction and oxidation) and the overexpression of another recombinant enzyme. In the fully constructed pathway, these additional steps may impose further stress on the host metabolism and a potential limitation on final product synthesis, due to possible intermediate build-up.

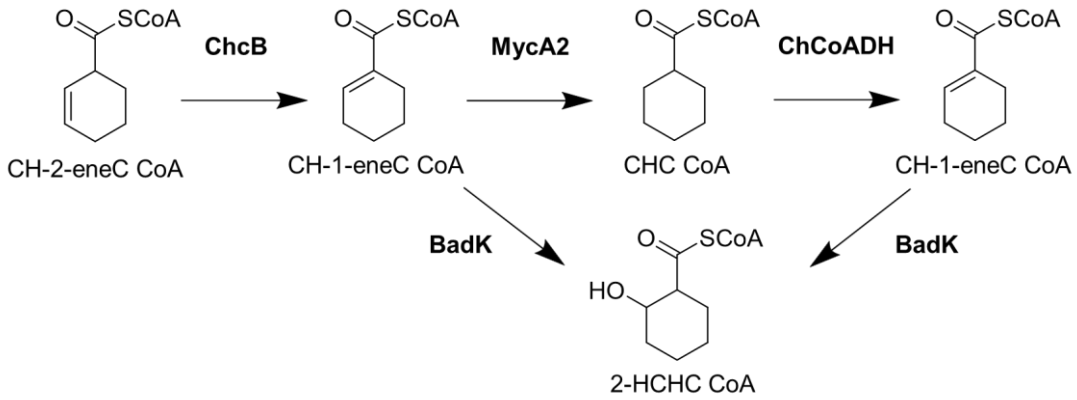



**Figure 7.6. The competing activity of BadK and MycA2 in the cyclohexanol pathway for CH-1-eneC CoA.**

The most direct route to cyclohexanol would hydrate, through BadK activity, ChcB-derived CH-1-eneC CoA, to 2-HCHC CoA. However, due to unwanted MycA2 activity, CH-1-eneC CoA will also be reduced to CHC CoA. Due to this competing activity, ChCoADH oxidation activity was also required to reform CH-1-eneC CoA to be used by BadK, preventing a build-up of the dead-end intermediate, CHC CoA.

With these points of consideration, future work should initially focus on determining the active conditions for MycA4, so that its activity can be determined with both of its proposed substrates. Based on these screening results the viability of the shorter cyclohexanol pathway, via 3-HCHC CoA TE hydrolysis, can more conclusively be determined. Moreover, successful determination of MycA4 activity would facilitate characterisation of the final upstream components of the cyclohexanol pathway, and bring the system closer to a scenario where the complete biotransformation of shikimate, or 3-phosphoshikimate, into cyclohexanol can be assessed as a proof of concept.

## 7.4. Chassis considerations for the final system

Due to well understood methods for growth and manipulation, *E.coli* BL21 (DE3) was used for all recombinant expression in this investigation. However, it may not represent the best chassis for the production of cyclohexanol. Therefore, once the cyclohexanol pathway has been full characterised and constructed, the optimal host for this system will need to be identified. The cyclohexanol pathway will utilise either shikimate, or 3-phosphoshikimate, from the shikimate pathway of the organism. Therefore, an organism that can produce high amounts of the substrate would be an extremely desirable feature.

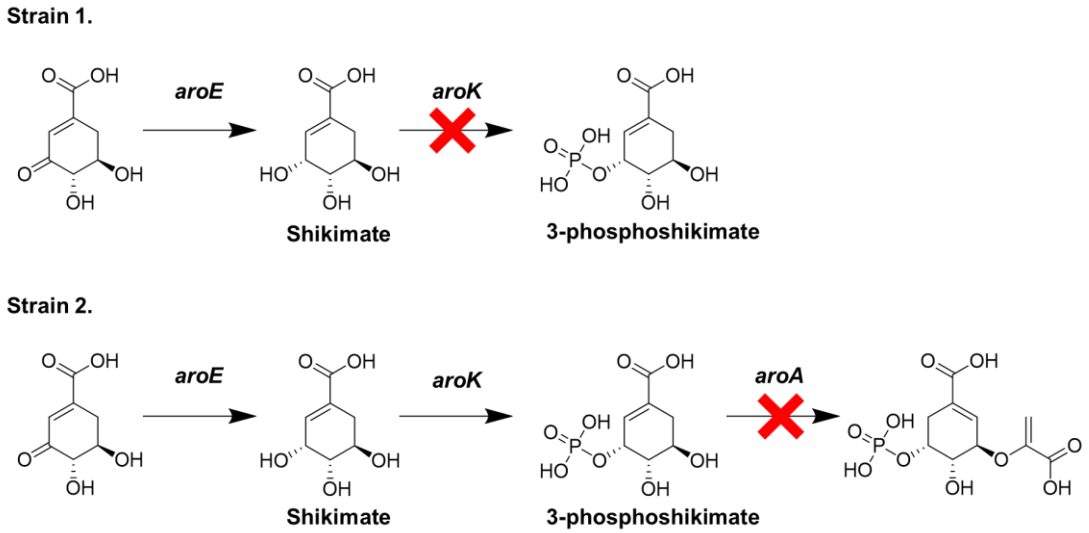



**Figure 7.7. A section of the shikimate pathway within two strains of C. glutamicum.**

Strain 1, has been previously developed and lead to the overproduction of shikimate, the deletion of the shikimate kinase gene, *aroK* (red cross), prevents the conversion of shikimate to 3-phosphoshikimate (Kogure *et al.*, 2016). Strain 2, is a proposed strain and would contain a functional copy of *aroK,* which would facilitate the conversion of shikimate into 3-phosphoshikimate. The 3-phosphoshikimate-1-carboxyvinyltransferase gene, *aroA*, would be deleted in this strain (red cross), enabling overproduction of 3-phosphoshikimate.

Notably, a previous investigation showed by overexpressing ten genes and inactivating another five genes, a strain of *Corynebacterium glutamicum* could be developed that was able to produce 141 g/L of shikimate (Kogure *et al.*,

2016). This would be an ideal candidate for the expression of the cyclohexanol pathway, as there would be a high amount of substrate for biotransformation into cyclohexanol. Furthermore, this strain of *C. glutamicum* was shown to derive similarly high titres of shikimate from lignocellulosic feedstock (Kogure *et al.*, 2016), demonstrating its potential for sustainable production of shikimate derived products. However, as MycA4 activity remains undetermined, it is unknown whether the cyclohexanol pathway will utilise shikimate or 3-phosphoshikimate. If the system requires 3-phosphoshikimate, then the described *C. glutamicum* strain will not be a feasible host. This is due to one of its deletion being shikimate kinase (*aroK*), required for the phosphorylation of shikimate into 3-phosphoshikimate (Kogure *et al.*, 2016; Averesch and Krömer, 2018). Therefore, for the accumulation of similarly high titres of 3-phoshoshikimate, the existing strain of *C. glutamicum* would need to be modified in two ways. Firstly, an active version of aroK would need to be reinstated within its genome, allowing 3-phosphoshikimte production from shikimate. Then to prevent consumption of 3-phosphoshikimate by the shikimate pathway, the downstream enzyme, 3-phosphoshikimatecarboxyvinyltransferase (aroA), would need to be deleted from the genome (Figure 7.7). Once the substrate for the cyclohexanol pathway has been deduced, a vector containing all of the pathway genes could be used to transform the chosen strain of *C. glutamicum*, for recombinant expression and assessment of cyclohexanol production. The high availability of substrate produced by this organism would hopefully facilitate g/L scale production of cyclohexanol. If this were found, then a strain could then be developed to contain both the cyclohexanol pathway and the previously developed 6-ACA producing pathway. This could be achieved by either co-transformation of both plasmids, or, genome integration of the pathways genes. Ultimately, this would result in the development of a *C. glutamicum* strain with the capacity to grow on lignocellulosic feedstock and produce the nylon 6 monomer, 6-ACA, via the CoA-utilising organic acid modification system developed here.

## 7.5. General discussion and conclusion

The work presented here has sought to develop an alternative production method for the nylon 6 monomer, 6-aminocaproic acid, using a synthetic biology framework, thereby alleviating sustainability issues inherent to its chemical production. However, the use of synthetic biology introduces its own sustainability issues, which were encountered within this investigation, and so will need to be addressed in future studies. At the outset of this investigation a hypothesis was set which was that a biosynthetic pathway can be developed that can produce cyclohexanol from an intermediate of central metabolism. However, a lack of enzymatic characterisation and commercially available substrates prevented the complete realisation of this pathway.

The limited characterisation of recombinant enzymes, as seen in this investigation, is a current bottleneck for the viable development of biosynthetic pathways. For this limitation to be overcome both academic and industrial groups need to continually characterise enzyme activities, adding to the existing scientific literature and databases. Although these types of publications may not have the highest impact factor, they provide invaluable data for the scientific community. Increases in publications looking into enzyme activities are therefore critical for improving the viability of developing pathways for chemical production. Currently, despite the existence of a number of databases (BRENDA, UniProt, KEGG and NCBI), it can be extremely difficult to identify required enzyme activities and key literature, with one often having to laboriously examine individual studies to find desired enzymes. As an increasing number of enzymes are characterised, databases should be constructed that will increase the accessibility of identifying desired enzymatic activities for implementation within biosynthetic pathways. For example, a database could be developed where one could input their desired chemical substrate and product, and the database could identify a number of biosynthetic pathways, using published enzyme activities, through which to catalyse this biotransformation. Although, this type of database may not be

near development at present, it is exactly this type of ‘big data’ that needs to be capitalised on to expand the utility and viability of synthetic biology as an alternative chemical production technology. Furthermore, as previously mentioned, limited product yields are a common caveat to the biological production of chemicals (Yim et al., 2011; Paddon and Keasling, 2014; Culler, 2016; Turk et al., 2016; Jackson, 2017), which is often caused by the metabolic production of unwanted by-products and rate-limiting enzymatic steps within the constructed pathway. As future investigations continue to elucidate the detailed metabolic network of organisms used for chemical production, more defined and rational approaches will enable the optimisation of biological pathways. Thus the combined characterisation of enzymes and metabolic system will ultimately facilitate the production of high product yields within biological systems.

Currently, the majority of investigations that have looked into the metabolic production of nylon monomers have been performed at the lab bench scale (Turk et al., 2016; Jackson, 2017). Yet, for the global production of nylon monomers through this method, vast production scale ups will be required to meet the consumer demand. At present fermentative scaling up is a largely undefined process containing a number of ‘unknowns’ including: the optimal method and machinery required to scale up fermentations, the effect of scaling up on product yields, the running and maintenance costs of large fermentations, the optimal media and carbon source for growth, the downstream processing of the product and the disposal method of biomass from genetically modified organisms (GMO). The current infancy of synthetic biology, therefore, provides an initial barrier to its industrial implementation as a platform technology for sustainable chemical production. However, this is a barrier that will be systematically resolved by academic and industrial research leading to an improved understanding of systems and processes. As the scientific understanding of the upscaling process improves, so too will the technology readiness and viability of using synthetic biology for the production of bulk and fine chemicals at an industrial scale. Furthermore, developments and standardisation of the upscaling process will attract investors to alternative

synthetic biology technologies, as they will be far less of an economic risk compared to standard chemical production processes.

A final consideration for the future of synthetic biology is its perception within society. The development of GMOs for the production of nylon monomers, and indeed other chemicals, will undoubtedly raise ethical concerns. For example, a possible societal issues may be faced in the form of a public stigma towards GMO-derived chemicals. Therefore, production chains will need to be developed such that GMOs are securely contained and correctly disposed of with no risk of exposure to the outside environment. Furthermore, as the field of synthetic biology expands, effective scientific communication will be required to continually address potential societal concerns. Through this means, the advantages of using synthetic biology as a sustainable alternative technology for chemical production can be effectively relayed to the public, promoting its consumer uptake and thus its status as a viable platform technology.

To conclude, while synthetic biology has the potential to be an exciting and sustainable alternative technology to that of petroleum-dependent production processes, it too currently has its own sustainability issues. However, these are predominantly a result of the field of synthetic biology being in its early stages. As research in this field is continually expanded, a collaborative emphasis must be made to effectively pool academic and industrial research and resources in the form of public databases and standardised processes. Through this route, these limitations can be tackled and the field of synthetic biology can meet its potential, becoming a competitive global technology. Therefore, although the outlined hypothesis could not be met in this investigation, it is clear that the continual sharing of data and information throughout the scientific community, will undoubtedly culminate in a field wielding the tools to meet this hypothesis and facilitate the sustainable production of nylon 6.

# 8. Chapter 8. Appendix

The appendix outlines all of the nucleotide and polypeptide sequences used within this investigation, as well as highlights key sequence features. All sequences are shown from 5’ to 3’.

### 8.1. The nucleotide sequence of the CHC CoA operon (pQR2786)

AGGAGATATACAT**ATG**CACCTTCTCGTCAAAGGAATTCAGCACCGTGTC
ACCGGTGAGATCCTCGTCCCGAACTCGAAATACCACGCGCACCGTGCG
CTGATCCTCGCCTCGCTCGCCGAAGGCGTGAGCCGCGTCCACGGACT
GTCCGACGCGCGGCACGTCGAGTACACCGTGCGACTGCTGCGCGACC
TCGGTGTGCGGATCGCCCGGGACGGTGACACCTTCGTGGTGCATGGA
CTCGGCGGGCGGTACAGGCCGCGCCGGGAGGCAGTGTCCGCCGGCA
GCTCCGGCACGACGCTGTACTTCATGATCGGTCTGGCCTCGCTGTCGG
ACCGCGCCGTGTCGGTGACCGGGCAGAAGTACTTCAGGCGACGCCCG
GTCGGCCCGCTGCTGCGCGCCCTCGAACAGCTGGGCGTCCAACTGGA
GTCGGCCGACGACTGCCCGCCCGTCCACGTACAGGGACGCAGGCCCA
CGGGCGGCCACGTCACCATCGCGGGAACCCTGTCCCAGTGGGTCTCG
GGCCTGATCCTGCTGGCCCCGTTCGCGACGCGGCACACCACCATCGA
GGTGGAGGGCGAACTCAACGAACGCCCCTACCTGGAACTGACCGTGG
CGATGATGCGGCAGTTCGGTCTCACCGTGACCGTCTCCGAGGACTGGC
GGCGCTTCGACATCGAGCCCGGCCAGCAGGCCCGCAGCGTCGAGCTG
ACGCTGCCGCCGGACATCGGCTCCGCCGCGTTCGGCATCGCGACCGC
GGCACTGCACCCGAGCGATGTGCTGCTGCGCGGGATCACCACGCTGG
ACGGCGGCCCGGCCGATCACCCCGAGTTCCACTTCCTCGACGTCGCA
CGCTCCATGGGCGTGCCGATGGAGACCGACGAGAGGGCCGGCGGTCT
GCGCATCCGGCAGGACACGCCGCGGCTCACCGCGGTCGACGTCGACT
GCCGCGACATCCCCGACATGCTGCCGATCCTGGCGACTCTCGCCACGT
TCGCGCGCGGCGAGTCGGTGTTCCGCAACATCGCGCACACCCGGCTC
AAGGAGTCCGACCGGGCCGCCGCCATGCTCCAGCTCAACTCCATGGG
CGGCCGCCTGGCACTGTCGGACGACGCGCTGCGGGTACGGGGCGTG
GACGGCCTGAGAGGGGCCAAGCTGTCCTCGTTCAACGACCACCGCATC
CTGATGTCGCTGGCGGTGGCGTCGTCGCGGGCGCGGGGGCACGCCA
CGCTGACCTATCCGAACGCGTACCGCATCTCGTACCCCACGTTCCTTG
ACGCGATGAACACGCTCACCGTGCCCATGTCGGTGGAGGACGGCTCG
GCGGGCCCCTCCGGCGGCGGTGCGACCGCCGGTGTCCCCGAGGCGA
CCGAGGCCGTCGGCCCCGAGGCCGCCTCGCGCGTCACTCTGCCTCGG
TGGCTGCGCCGAAGGGCCGAGGCCCGGCCGCACGACACCGCCGTGG
TGGACGTGCGCTCCGACCGCGACGAGATGATCACCTGGAGCGAGCTG
GCCGAGCGGGTGGACCGTGCGGCGGCCCTGTTGCTGCGGCTCGGGG
TACGGCCGGGCGAGAACGTGGCGTACCAGCTGCCGAACCGTGTCGAG
TTCGTGGTGCTGTCGCTGGCGGCCCTGCGCATCGGGGCGGTGTGCTG

CCCGGTCATCCCGTTCTTCCGCAAGCGCGAGCTCGGTTTCGTCCTGCG
GCGCTCCAAGGCGCGGGTCCTGGTCGTGGCGGACCGGCACCGCTCGC
GCCTGCCCGCCGAGGAGGTGCTCGACCTGGTCGCGGAGGACGGCCC
GGACCTGAATCTGGAGCACGTGGTGGTCCTGGCCGCCGCAGGCGGCT
CCGCCCGGCTTCCCGAGGCCCCCACGGGTGGGACGGCCGTGCACGA
CTGGGGGCAGGCGCTGTCGGAGACGACGGTCGACCGCGCGGTGCTC
GACGCCCTCGCGCCGACACCGGAGATGACGGCGCAGCTGCTCTTCAC
CTCGGGCACCACCAGCGAACCGAAGGGCGTCACCCAGCCCACGCGGA
ACCTGGTCAGGGCCGCCGCGATGGAGATCGGGCACCTCGGGCTCGGC
AAGCAGGACGCGATCTGGGTGCCGTCGCCGCTCGCCCACCAGACGGG
CTTCCTGTACGGCATGGTGCTCGCCCTGGTACTGGGCGTGCCGCAGAT
CCTGCAGCCCGAGTGGGACGCCAAGCGGGCACTGCAGTCGCTCAACG
AGCACCGGGCGACGTTCGTGCAGGCCGCGACGCCGTTCCTGTCCGAC
CTGGTGAAGGCCGTGGAGGAGAGCGGCGAGACGCCGCAGCACCTGC
GGATCTTCGTGGCGACGGGAGCCATGGTGCCGCGGGCGCTGGCGGA
GCGGGCCGGGCGGGTGCTGCGCACCTACGTGTGCGGGGCCTTCGGC
ACCACGGAGACCTGCCTGGGCGCGTTGTCGTCACCCGGCGACGAGCC
CCGGCAGCGCTGGGGCTCCGACGGCCGCGCGCTGGACGGCATCGAG
CTGCGCGTCACCGACGACGAAGGGCTGGTGCTGCCCGCGGGCAAGGA
GGGCAACTTCGAGCTGCGTTCCCCCACGGTCTTCGACGGTTACCTCGA
CCGCCCCGACCTGACCTCCCAGGCGTTCACCGAGGACGGCTGGTACC
GCACGGGCGATCTGGCGACGATCGACACCGACGGCTTCCTGCGGATC
ACCGGGCGGGTCAAGGACGTGATCAACCGCGGCGGCGAGAAGATACC
GGTCGCGGAGATCGAGCAACTGCTGTTCGGGCACCCCGCGGTGGAGG
ACGTCGCGGTGGTGGCCATGCCGGACGAGCGTCTCGGCGAGCGCGCC
TGCGCCTTCGTGGTCCCGGCGGCCGGTACGGAGCTGACCTTCGAGGA
GATGCGGCGCCACCTCGACGGCCACGAGGTGGCCAAGCAGTACTGGC
CGGAGCGTCTGGAGCGGATCGACGCCCTGCCGCGCAACCCCATCGGC
AAGGTGACGAAGTGGGAACTGCGGGCGCTGGCCCATGGTTTGCGGCC
CCACGACGAGGACGCCGGCTGACGCACGACGGACAGCACACCGCAAC
GACGCAACAAGAACAGAGGTGGCAC**ATG**ACGGACGAGCAGCAGTTCC
ACGAGCTGCGCACCCAGGTCGAACAGTGGGTGGAGGGACCCGGCGAG
CGCTGGGCCGAACGCATCGAGGAGACCGGGCAGGTGCCCGAGGAGCT
GTGGGCCGAGCTGAACGGCCTGGGGTTCCTGCGGATAGCCGCTCCCG
AGGAGTACGGCGGCCAGGGCCTCGACTTCGTGCGCTGGATGGAGCTG
ATGGAGATCTTCTCCCGCTCGCACGGATCCGTGCGCATGATCGTGCAC
GTGGTCAACGGCATCTGGCGAGCGATGGACGGCCACGCCGACGACGA
GCAGCGCAAGCGCTTCGTCGTTCCCTCCGTCACCGGTGAGATCAAGAT
CGCCTTCACCCTCACCGAACCGGGCAACGGCACCGGCGCGGACATCA
CCACGTCGGCGGTGCGCGAGGGCGACACGTACTACCTCTCAGGACGC
AAGCACCTGATCACCTTCGGGGTGCGCTGCGACTACTACCTGCTGGCC
GCCCGGGTCGAGGGCAGCACCGGGCACGAGGGCACGATGGCGCTGC
TGGTGCCGCGTGACGCCCCGGGCGTCACGGTCGAGGACACCTCCGAC
ACCATGGGCGTCCGGGGTACCGACCATGCCTCGCTGGTCTTCGACCG
CACCCCCGTGCCGGTCGACCACCGGCTCGGTGCGGAGGGCCAGGGC
CTGGAGGTGTTCCTCGGCGGCTTCCTCACCCCGTCCCGGATCTCCGTG

GCGATGAGCTGCGTGGGGCTCGCGCAGCGTGCCCAGCAGCTGGCCGT
CGAATACGCCCGCAACCGTGTGACCTTCGGCAAGACGCTGACCCAGCG
CCAGGTCATCCAGTTCATGCTCGCCGAGAACGCGGCGGACATCGAGG
CGGCACGGCAGCTGGTCCTGCACGCGGCGCGCCGCTTCGAGGACGGA
GCGGACGACGCCTCCATGCAGTCGTCGATGGCGAAGATGCACGCCGT
GACGATGCTGACGAACGTCACCGACAAGGCGCTCCAGGTGCACGGCG
GTCTCGGTTACTGGAAGTCGCAGAAGATCGAGCGCGTCTACCGCGACG
CACGCGCCCAGCGCTTCGAGGAAGGCACCAACGAGGTGCAGAAGGCC
GTCGTCTTCCGTGAGTTGCTCCAGCGCACGACGTCCCGGGACGGAGA
GGGAACGAACCG**ATG**AACAGCCCGCACCGTCAGCCGACGCCCGAGCG
CGAGCAGGTCTCGCTGATCACCGGCGCCTCGCGCGGCATCGGCCGCG
CCCTGGCCCTCACCCTCGCCGAACGCGGTGGCACCGTGGTCGTGAAC
TACAAGAAGAACGCCGACCTGGCCCAGAAGACCGTGGCCGAGGTCGA
GGAGGCCGGCGGCAGGGGCTTCGCGGTCCAGGCGGACGTCGAGACC
ACCGAGGGCGTCACGACGCTGTTCGACGAGGTGGCGCAGCGCTGCGG
ACGGCTCGACCACTTCGTCAACAACGCCGCGGCCAGCGCGTTCAAGAG
CATCCTCGACCTCGGCCCGCACCACCTGGACCGCTCGTACGCGATGAA
CCTGCGGCCGTTCGTCCTGGGGGCGCAGCAGGCCGTGAAGCTCATGG
ACAACGGCGGGCGGATCGTCGCGCTGTCCTCCTACGGTTCCGTCCGC
GCCTACCCCACCTACGCGATGCTCGGCGGCATGAAGGCCGCCATAGA
GTCGTGGGTGCGGTACATGGCGGTGGAGTTCGCGCCCTACGGCATCA
ACGTCAACGGGGTCAACGGCGGCCTGATCGACTCCGATTCGCTGGAGT
TCTTCTACAACGTGGAAGGCATGCCGCCCATGCAGGGCGTCCTCGACC
GCATTCCCGCGCGCCGTCCCGGCACCGTGCAGGAGATGGCCGACACC
ATCGCCTTCCTGCTCGGCGAGGGGGCGGGCTACATCACCGGGCAGAC
CCTCGTGGTCGACGGCGGGCTCAGCGTCGTCGCGCCGCCGTTCTTCG
CGGACGCGGGCGAGGCGCTCGCGCTGCCGCCCCGGCCCACGCGCGA
TGCCTGAACGCGACGACCGGGAAGAGGGACGGCAG**GTG**TTGGACCAG
GTCTTCAGGCCCTGCGTGATCGGTGGCCTGACAGTGCCGCACCGCATC
GTCATGGGTGCCATGCACCTGAACCTGGAGACCCTGGACGACGGTGG
CGCCGCCCTGGCCGCGTTCTACACGGAACGGGTGCGGGGCGGCGCA
GGCCTGATCGTCACCGGCGGCTCCGCGGTGAACCCGGCGGGTTCCGG
CGGCCCCGGATACGGCGTGCTCGACGACGACAAGCACCGGTCGGCGC
TGCGCCTGGCCGTCCGCGCGGTGCACGACGCCGGGGGGCTGATCGC
CCTTCAGCTCTTCCACGCCGGGCGGTACGCGCTGCCCGGCTCCCCCG
GGGCGGACGGCGCCGCGCCCCTCGCGCCGTCCGCCGTGTACAGCCG
GTTCTCCCGCACCGTTCCGCGGGCGATGACGCAAGAGCAGATCGCCG
AGACCCTCGCCGCCTTCGCCCGGGGGGGCGCGAAGGGCGTACGAGCTG
GGCTTCGACGCCGTCGAGGTGATGGGTTCCGAGGGCTATCTGATCAAT
CAGTTCACCGCGCCCCTCACCAACCGGCGCGACGACGCCTGGGGCGG
TGACGCCGGCCGGCGCGGGCGGTTCCCCGTCGAGGTGCTGCGGGCC
GTACGAGCCGCGGTGGGACCGGACTTCCCGGTGCTCTTCCGGATGTC
GGGCGCCGACCTCGTGGACGACGGCACCCCGCGGGAGGAGACGGCC
GCGCTCGCTGTGGAGCTCGCCCGAGCCGGCGCCGACGCGCTCGCGG
TGGGCGTCGGCTGGCACGAGTCCCCGGTGCCGACCGTGCAGGCCCAG
GTGCCGGCGGGCACCTGGGCCGTGTACGCCCAGCGGGTCAAGCGGG

CGCTGCGCGCGGCGGGTCACGGCACGCTGCCCGTCATCGCCTCCAAC
CGCTTCAACCGGCTCGCGCAGGCCGAGCAGGTCCTGGCCGGGGGTGA
CGTCGATCTGGTGGCCATGGCCCGCCCGTTCCTCGCGGACCCGGCCA
TCGTCGACAAGTCCCGTACCGGGGCGAGCGCGTCGGTGGACCTCTGC
ATCGCCTGCAACGAGGCGTGCATCGACCGGTCCTTCGGCACCGAGCG
GGTGTCCTGTCTGGTCAATCCGCGGGCGGGCCACGAGCGTGAGTTCC
CGGCCCGGCCGTCTCGCGCGCCGGGGCAGGTACGCCGTGATCGG
CGCGGGTCTCGCGGGACTGGAGGCGGCCCGCACCCTGGCGAGACTG
GGGCACAGGGTCGAGGTCTACGAGGCGGCCGACGAGCCCGGCGGCC
AGTTCCGGCTGGCCGCCCGGGTGCCGGGCAAGGCGGACTTCGCCGC
GACCGTACGACAGCGGGAGCAGGAACTCGGCCGTCTGAACGTGCCCG
TGCACATGGGGCGCCGCATCGGTGCCCGGGACGTGCCCGTCCTGCGC
GCCATGGACGGAGTGGTCCTCGCCACCGGGGTCCGCCCCCACCGGCC
GGACCTTCCCGGCGTCGATCTGCCGCATGTCCTCGACTACGTCCGGGC
GTTCGCCGACCCCGGCGCCCTCGGTCCGCGTGTCGCCGTCATCGGCG
GCGGCGGAATCGCGGTCGACCTGGCGCACCTCCTCACCAGTGGCTCC
TCCCCCGTCCCGACCCCGGAGGAGTTCCTCGCCTCCCGCGCCCTGAC
CGCACCGCCCGAAGCCGTCCCGGCGCCACGGACCACGGCCGCCCCG
GCCGGTGGCGCCCGTCAGCTCACCGTGATGCGGCGCGCCGGCAGGAT
CGGCGCCGGGATCGGCCCGTCCACCCGGTGGGTGGTCATGGAAGAAC
TGCGCGCTCGCGGAGTACGCCTGCTCACCGACGTCACGTACGAGGAG
ATCACCCCGGAGGGTGTGGTCGTCGTCGAGGCAGGAGGAGAACGGCG
GTTGGTCGCCGCCGACCACGTCGTCCTCGCCACCGGCCAGGAAAGCC
GGCGGGAGGTGGCCGCCGTGCTGCGCCGGGCCGACGTGCCGTTCGA
GACCGCCGGCGGCGCCGCCGACGCCCGGGCGCTCAACGCCGTGCGG
GCCACCGCGCAAGGACTGCGCGCCGCGCACCGCATCGTGGAACGGGC
CGAGTCGAGGAGGACCGGGTAGTAGCTCGAG

**Figure 8.1. The nucleotide sequence of the CHC CoA operon cloned into pET29a (pQR2786).**

The pET29a ribosome binding site (RBS) and proposed ribosome binding sites are highlighted in red. The RBS of *mycA2* is within the operon reading frame of *mycA3.* The restriction sites (*Nde*I and *Xho*I) are highlighted in turquoise. Open reading frames are highlighted in grey in the order of 1- *mycA4*, 2- *mycA3*, 3- *mycA2*, 4- *mycA1*. Start codons are in bold font and stop codons are underlined. The stop codon of *mycA3* and start codon of *mycA2* overlap.

## 8.2. The nucleotide sequence of *mycA1* (pQR2787)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCAT**ATG**TTGGACCAGGTCTTCAGGCC
CTGCGTGATCGGTGGCCTGACAGTGCCGCACCGCATCGTCATGGGTG
CCATGCACCTGAACCTGGAGACCCTGGACGACGGTGGCGCCGCCCTG
GCCGCGTTCTACACGGAACGGGTGCGGGGCGGCGCAGGCCTGATCGT
CACCGGCGGCTCCGCGGTGAACCCGGCGGGTTCCGGCGGCCCCGGA
TACGGCGTGCTCGACGACGACAAGCACCGGTCGGCGCTGCGCCTGGC
CGTCCGCGCGGTGCACGACGCCGGGGGGCTGATCGCCCTTCAGCTCT
TCCACGCCGGGCGGTACGCGCTGCCCGGCTCCCCCGGGGCGGACGG
CGCCGCGCCCCTCGCGCCGTCCGCCGTGTACAGCCGGTTCTCCCGCA
CCGTTCCGCGGGCGATGACGCAAGAGCAGATCGCCGAGACCCTCGCC
GCCTTCGCCCGGGGGGCGCGAAGGGCGTACGAGCTGGGCTTCGACG
CCGTCGAGGTGATGGGTTCCGAGGGCTATCTGATCAATCAGTTCACCG
CGCCCCTCACCAACCGGCGCGACGACGCCTGGGGCGGTGACGCCGG
CCGGCGCGGGCGGTTCCCCGTCGAGGTGCTGCGGGCCGTACGAGCC
GCGGTGGGACCGGACTTCCCGGTGCTCTTCCGGATGTCGGGCGCCGA
CCTCGTGGACGACGGCACCCCGCGGGAGGAGACGGCCGCGCTCGCT
GTGGAGCTCGCCCGAGCCGGCGCCGACGCGCTCGCGGTGGGCGTCG
GCTGGCACGAGTCCCCGGTGCCGACCGTGCAGGCCCAGGTGCCGGC
GGGCACCTGGGCCGTGTACGCCCAGCGGGTCAAGCGGGCGCTGCGC
GCGGCGGGTCACGGCACGCTGCCCGTCATCGCCTCCAACCGCTTCAA
CCGGCTCGCGCAGGCCGAGCAGGTCCTGGCCGGGGGTGACGTCGATC
TGGTGGCCATGGCCCGCCCGTTCCTCGCGGACCCGGCCATCGTCGAC
AAGTCCCGTACCGGGGCGAGCGCGTCGGTGGACCTCTGCATCGCCTG
CAACGAGGCGTGCATCGACCGGTCCTTCGGCACCGAGCGGGTGTCCT
GTCTGGTCAATCCGCGGGCGGGCCACGAGCGTGAGTTCCCGGCCCGG
CCGTCTCGCGCGCCGGGGCAGGTACGCCGTGATCGGCGCGGGTCT
CGCGGGACTGGAGGCGGCCCGCACCCTGGCGAGACTGGGGCACAGG
GTCGAGGTCTACGAGGCGGCCGACGAGCCCGGCGGCCAGTTCCGGCT
GGCCGCCCGGGTGCCGGGCAAGGCGGACTTCGCCGCGACCGTACGA
CAGCGGGAGCAGGAACTCGGCCGTCTGAACGTGCCCGTGCACATGGG
GCGCCGCATCGGTGCCCGGGACGTGCCCGTCCTGCGCGCCATGGACG
GAGTGGTCCTCGCCACCGGGGTCCGCCCCCACCGGCCGGACCTTCCC
GGCGTCGATCTGCCGCATGTCCTCGACTACGTCCGGGCGTTCGCCGAC
CCCGGCGCCCTCGGTCCGCGTGTCGCCGTCATCGGCGGCGGCGGAAT
CGCGGTCGACCTGGCGCACCTCCTCACCAGTGGCTCCTCCCCCGTCC
CGACCCCGGAGGAGTTCCTCGCCTCCCGCGCCCTGACCGCACCGCCC
GAAGCCGTCCCGGCGCCACGGACCACGGCCGCCCCGGCCGGTGGCG
CCCGTCAGCTCACCGTGATGCGGCGCGCCGGCAGGATCGGCGCCGG
GATCGGCCCGTCCACCCGGTGGGTGGTCATGGAAGAACTGCGCGCTC
GCGGAGTACGCCTGCTCACCGACGTCACGTACGAGGAGATCACCCCG
GAGGGTGTGGTCGTCGTCGAGGCAGGAGGAGAACGGCGGTTGGTCGC
CGCCGACCACGTCGTCCTCGCCACCGGCCAGGAAAGCCGGCGGGAGG
TGGCCGCCGTGCTGCGCCGGGCCGACGTGCCGTTCGAGACCGCCGG

CGGCGCCGCCGACGCCCGGGCGCTCAACGCCGTGCGGGCCACCGCG
CAAGGACTGCGCGCCGCGCACCGCATCGTGGAACGGGCCGAGTCGAG
GAGGACCGGGTAGCTCGAG

**Figure 8.2. The nucleotide sequence of *mycA1* cloned into pET28a (pQR2787).**

The ribosome binding site (RBS) is highlighted in red, restriction sites (*Nde*I and *Xho*I) are highlighted in turquoise. The open reading frame is highlighted in grey. Start codons (ATG) are in bold font and the stop codon is underlined. The N-terminal His-tag is highlighted in teal.

## 8.3. The polypeptide sequence expressed from *mycA1* (pQR2787)

**M**GSSHHHHHHSSGLVPRGSHMLDQVFRPCVIGGLTVPHRIVMGAMHLNL
ETLDDGGAALAAFYTERVRGGAGLIVTGGSAVNPAGSGGPGYGVLDDDKH
RSALRLAVRAVHDAGGLIALQLFHAGRYALPGSPGADGAAPLAPSAVYSRF
SRTVPRAMTQEQIAETLAAFARGARRAYELGFDAVEVMGSEGYLINQFTAP
LTNRRDDAWGGDAGRRGRFPVEVLRAVRAAVGPDFPVLFRMSGADLVDD
GTPREETAALAVELARAGADALAVGVGWHESPVPTVQAQVPAGTWAVYA
QRVKRALRAAGHGTLPVIASNRFNRLAQAEQVLAGGDVDLVAMARPFLAD
PAIVDKSRTGASASVDLCIACNEACIDRSFGTERVSCLVNPRAGHEREFPA
RPSRARRGRYAVIGAGLAGLEAARTLARLGHRVEVYEAADEPGGQFRLAA
RVPGKADFAATVRQREQELGRLNVPVHMGRRIGARDVPVLRAMDGVVLA
TGVRPHRPDLPGVDLPHVLDYVRAFADPGALGPRVAVIGGGGIAVDLAHLL
TSGSSPVPTPEEFLASRALTAPPEAVPAPRTTAAPAGGARQLTVMRRAGRI
GAGIGPSTRWVVMEELRARGVRLLTDVTYEEITPEGVVVVEAGGERRLVA
ADHVVLATGQESRREVAAVLRRADVPFETAGGAADARALNAVRATAQGLR
AAHRIVERAESRRTG-

**Figure 8.3. The polypeptide sequence of MycA1 expressed from mycA1 (pQR2787).**

The start Methionine (M) is in bold font. The N-terminal His-tag is highlighted in teal.

## 8.4. The nucleotide sequence of *mycA2* (pQR2788)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCAT**ATG**AACAGCCCGCACCGTCAGCC
GACGCCCGAGCGCGAGCAGGTCTCGCTGATCACCGGCGCCTCGCGCG
GCATCGGCCGCGCCCTGGCCCTCACCCTCGCCGAACGCGGTGGCACC
GTGGTCGTGAACTACAAGAAGAACGCCGACCTGGCCCAGAAGACCGTG
GCCGAGGTCGAGGAGGCCGGCGGCAGGGGCTTCGCGGTCCAGGCGG
ACGTCGAGACCACCGAGGGCGTCACGACGCTGTTCGACGAGGTGGCG
CAGCGCTGCGGACGGCTCGACCACTTCGTCAACAACGCCGCGGCCAG
CGCGTTCAAGAGCATCCTCGACCTCGGCCCGCACCACCTGGACCGCTC
GTACGCGATGAACCTGCGGCCGTTCGTCCTGGGGGCGCAGCAGGCCG
TGAAGCTCATGGACAACGGCGGGCGGATCGTCGCGCTGTCCTCCTACG
GTTCCGTCCGCGCCTACCCCACCTACGCGATGCTCGGCGGCATGAAG
GCCGCCATAGAGTCGTGGGTGCGGTACATGGCGGTGGAGTTCGCGCC
CTACGGCATCAACGTCAACGGGGTCAACGGCGGCCTGATCGACTCCGA
TTCGCTGGAGTTCTTCTACAACGTGGAAGGCATGCCGCCCATGCAGGG
CGTCCTCGACCGCATTCCCGCGCGCCGTCCCGGCACCGTGCAGGAGA
TGGCCGACACCATCGCCTTCCTGCTCGGCGAGGGGGCGGGCTACATC
ACCGGGCAGACCCTCGTGGTCGACGGCGGGCTCAGCGTCGTCGCGCC
GCCGTTCTTCGCGGACGCGGGCGAGGCGCTCGCGCTGCCGCCCCGG
CCCACGCGCGATGCCTGACTCGAG

**Figure 8.4. The nucleotide sequence of *mycA2* cloned into pET28a (pQR2788).**

The ribosome binding site (RBS) is highlighted in red, restriction sites (*Nde*I and *Xho*I) are highlighted in turquoise. The open reading frame is highlighted in grey. Start codons (ATG) are in bold font and the stop codon is underlined. The N-terminal His-tag is highlighted in teal.

## 8.5. The polypeptide sequence expressed from *mycA2* (pQR2788)

**M**GSSHHHHHHSSGLVPRGSHMNSPHRQPTPEREQVSLITGASRGIGRALA
LTLAERGGTVVVNYKKNADLAQKTVAEVEEAGGRGFAVQADVETTEGVTT
LFDEVAQRCGRLDHFVNNAAASAFKSILDLGPHHLDRSYAMNLRPFVLGA
QQAVKLMDNGGRIVALSSYGSVRAYPTYAMLGGMKAAIESWVRYMAVEFA
PYGINVNGVNGGLIDSDSLEFFYNVEGMPPMQGVLDRIPARRPGTVQEMA
DTIAFLLGEGAGYITGQTLVVDGGLSVVAPPFFADAGEALALPPRPTRDA-

**Figure 8.5. The polypeptide sequence of MycA2 expressed from *mycA2* (pQR2788).**

The start Methionine (M) is in bold font. The N-terminal His-tag is highlighted in teal.

## 8.6. The nucleotide sequence of *mycA3* (pQR2789)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCAT**ATG**ACGGACGAGCAGCAGTTCCA
CGAGCTGCGCACCCAGGTCGAACAGTGGGTGGAGGGACCCGGCGAGC
GCTGGGCCGAACGCATCGAGGAGACCGGGCAGGTGCCCGAGGAGCT
GTGGGCCGAGCTGAACGGCCTGGGGTTCCTGCGGATAGCCGCTCCCG
AGGAGTACGGCGGCCAGGGCCTCGACTTCGTGCGCTGGATGGAGCTG
ATGGAGATCTTCTCCCGCTCGCACGGATCCGTGCGCATGATCGTGCAC
GTGGTCAACGGCATCTGGCGAGCGATGGACGGCCACGCCGACGACGA
GCAGCGCAAGCGCTTCGTCGTTCCCTCCGTCACCGGTGAGATCAAGAT
CGCCTTCACCCTCACCGAACCGGGCAACGGCACCGGCGCGGACATCA
CCACGTCGGCGGTGCGCGAGGGCGACACGTACTACCTCTCAGGACGC
AAGCACCTGATCACCTTCGGGGTGCGCTGCGACTACTACCTGCTGGCC
GCCCGGGTCGAGGGCAGCACCGGGCACGAGGGCACGATGGCGCTGC
TGGTGCCGCGTGACGCCCCGGGCGTCACGGTCGAGGACACCTCCGAC
ACCATGGGCGTCCGGGGTACCGACCATGCCTCGCTGGTCTTCGACCG
CACCCCCGTGCCGGTCGACCACCGGCTCGGTGCGGAGGGCCAGGGC
CTGGAGGTGTTCCTCGGCGGCTTCCTCACCCCGTCCCGGATCTCCGTG
GCGATGAGCTGCGTGGGGCTCGCGCAGCGTGCCCAGCAGCTGGCCGT
CGAATACGCCCGCAACCGTGTGACCTTCGGCAAGACGCTGACCCAGCG
CCAGGTCATCCAGTTCATGCTCGCCGAGAACGCGGCGGACATCGAGG
CGGCACGGCAGCTGGTCCTGCACGCGGCGCGCCGCTTCGAGGACGGA
GCGGACGACGCCTCCATGCAGTCGTCGATGGCGAAGATGCACGCCGT
GACGATGCTGACGAACGTCACCGACAAGGCGCTCCAGGTGCACGGCG
GTCTCGGTTACTGGAAGTCGCAGAAGATCGAGCGCGTCTACCGCGACG
CACGCGCCCAGCGCTTCGAGGAAGGCACCAACGAGGTGCAGAAGGCC
GTCGTCTTCCGTGAGTTGCTCCAGCGCACGACGTCCCGGGACGGAGA
GGGAACGAACCGATGACTCGAG

**Figure 8.6. The nucleotide sequence of *mycA3* cloned into pET28a (pQR2789).**

The ribosome binding site (RBS) is highlighted in red, restriction sites (*Nde*I and *Xho*I) are highlighted in turquoise. The open reading frame is highlighted in grey. Start codons (ATG) are in bold font and the stop codon is underlined. The N-terminal His-tag is highlighted in teal.

## 8.7. The polypeptide sequence expressed from *mycA3* (pQR2789)

**M**GSSHHHHHHSSGLVPRGSHMTDEQQFHELRTQVEQWVEGPGERWAE
RIEETGQVPEELWAELNGLGFLRIAAPEEYGGQGLDFVRWMELMEIFSRSH
GSVRMIVHVVNGIWRAMDGHADDEQRKRFVVPSVTGEIKIAFTLTEPGNGT
GADITTSAVREGDTYYLSGRKHLITFGVRCDYYLLAARVEGSTGHEGTMAL
LVPRDAPGVTVEDTSDTMGVRGTDHASLVFDRTPVPVDHRLGAEGQGLE
VFLGGFLTPSRISVAMSCVGLAQRAQQLAVEYARNRVTFGKTLTQRQVIQF

MLAENAADIEAARQLVLHAARRFEDGADDASMQSSMAKMHAVTMLTNVTD
KALQVHGGLGYWKSQKIERVYRDARAQRFEEGTNEVQKAVVFRELLQRTT
SRDGEGTNR-

**Figure 8.7. The polypeptide sequence of MycA3 expressed from *mycA3* (pQR2789).**

The start Methionine (M) is in bold font. The N-terminal His-tag is highlighted in teal.

## 8.8. The nucleotide sequence of *mycA4* (pQR2790)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCAT**ATG**CACCTTCTCGTCAAAGGAATT
CAGCACCGTGTCACCGGTGAGATCCTCGTCCCGAACTCGAAATACCAC
GCGCACCGTGCGCTGATCCTCGCCTCGCTCGCCGAAGGCGTGAGCCG
CGTCCACGGACTGTCCGACGCGCGGCACGTCGAGTACACCGTGCGAC
TGCTGCGCGACCTCGGTGTGCGGATCGCCCGGGACGGTGACACCTTC
GTGGTGCATGGACTCGGCGGGCGGTACAGGCCGCGCCGGGAGGCAG
TGTCCGCCGGCAGCTCCGGCACGACGCTGTACTTCATGATCGGTCTGG
CCTCGCTGTCGGACCGCGCCGTGTCGGTGACCGGGCAGAAGTACTTC
AGGCGACGCCCGGTCGGCCCGCTGCTGCGCGCCCTCGAACAGCTGGG
CGTCCAACTGGAGTCGGCCGACGACTGCCCGCCCGTCCACGTACAGG
GACGCAGGCCCACGGGCGGCCACGTCACCATCGCGGGAACCCTGTCC
CAGTGGGTCTCGGGCCTGATCCTGCTGGCCCCGTTCGCGACGCGGCA
CACCACCATCGAGGTGGAGGGCGAACTCAACGAACGCCCCTACCTGGA
ACTGACCGTGGCGATGATGCGGCAGTTCGGTCTCACCGTGACCGTCTC
CGAGGACTGGCGGCGCTTCGACATCGAGCCCGGCCAGCAGGCCCGCA
GCGTCGAGCTGACGCTGCCGCCGGACATCGGCTCCGCCGCGTTCGGC
ATCGCGACCGCGGCACTGCACCCGAGCGATGTGCTGCTGCGCGGGAT
CACCACGCTGGACGGCGGCCCGGCCGATCACCCCGAGTTCCACTTCC
TCGACGTCGCACGCTCCATGGGCGTGCCGATGGAGACCGACGAGAGG
GCCGGCGGTCTGCGCATCCGGCAGGACACGCCGCGGCTCACCGCGGT
CGACGTCGACTGCCGCGACATCCCCGACATGCTGCCGATCCTGGCGA
CTCTCGCCACGTTCGCGCGCGGCGAGTCGGTGTTCCGCAACATCGCG
CACACCCGGCTCAAGGAGTCCGACCGGGCCGCCGCCATGCTCCAGCT
CAACTCCATGGGCGGCCGCCTGGCACTGTCGGACGACGCGCTGCGGG
TACGGGGCGTGGACGGCCTGAGAGGGGCCAAGCTGTCCTCGTTCAAC
GACCACCGCATCCTGATGTCGCTGGCGGTGGCGTCGTCGCGGGCGCG
GGGGCACGCCACGCTGACCTATCCGAACGCGTACCGCATCTCGTACCC
CACGTTCCTTGACGCGATGAACACGCTCACCGTGCCCATGTCGGTGGA
GGACGGCTCGGCGGGCCCCTCCGGCGGCGGTGCGACCGCCGGTGTC
CCCGAGGCGACCGAGGCCGTCGGCCCCGAGGCCGCCTCGCGCGTCA
CTCTGCCTCGGTGGCTGCGCCGAAGGGCCGAGGCCCGGCCGCACGAC
ACCGCCGTGGTGGACGTGCGCTCCGACCGCGACGAGATGATCACCTG
GAGCGAGCTGGCCGAGCGGGTGGACCGTGCGGCGGCCCTGTTGCTG
CGGCTCGGGGTACGGCCGGGCGAGAACGTGGCGTACCAGCTGCCGAA

CCGTGTCGAGTTCGTGGTGCTGTCGCTGGCGGCCCTGCGCATCGGGG
CGGTGTGCTGCCCGGTCATCCCGTTCTTCCGCAAGCGCGAGCTCGGTT
TCGTCCTGCGGCGCTCCAAGGCGCGGGTCCTGGTCGTGGCGGACCGG
CACCGCTCGCGCCTGCCCGCCGAGGAGGTGCTCGACCTGGTCGCGGA
GGACGGCCCGGACCTGAATCTGGAGCACGTGGTGGTCCTGGCCGCCG
CAGGCGGCTCCGCCCGGCTTCCCGAGGCCCCCACGGGTGGGACGGC
CGTGCACGACTGGGGGCAGGCGCTGTCGGAGACGACGGTCGACCGC
GCGGTGCTCGACGCCCTCGCGCCGACACCGGAGATGACGGCGCAGCT
GCTCTTCACCTCGGGCACCACCAGCGAACCGAAGGGCGTCACCCAGC
CCACGCGGAACCTGGTCAGGGCCGCCGCGATGGAGATCGGGCACCTC
GGGCTCGGCAAGCAGGACGCGATCTGGGTGCCGTCGCCGCTCGCCCA
CCAGACGGGCTTCCTGTACGGCATGGTGCTCGCCCTGGTACTGGGCGT
GCCGCAGATCCTGCAGCCCGAGTGGGACGCCAAGCGGGCACTGCAGT
CGCTCAACGAGCACCGGGCGACGTTCGTGCAGGCCGCGACGCCGTTC
CTGTCCGACCTGGTGAAGGCCGTGGAGGAGAGCGGCGAGACGCCGCA
GCACCTGCGGATCTTCGTGGCGACGGGAGCCATGGTGCCGCGGGCGC
TGGCGGAGCGGGCCGGGCGGGTGCTGCGCACCTACGTGTGCGGGGC
CTTCGGCACCACGGAGACCTGCCTGGGCGCGTTGTCGTCACCCGGCG
ACGAGCCCCGGCAGCGCTGGGGCTCCGACGGCCGCGCGCTGGACGG
CATCGAGCTGCGCGTCACCGACGACGAAGGGCTGGTGCTGCCCGCGG
GCAAGGAGGGCAACTTCGAGCTGCGTTCCCCCACGGTCTTCGACGGTT
ACCTCGACCGCCCCGACCTGACCTCCCAGGCGTTCACCGAGGACGGC
TGGTACCGCACGGGCGATCTGGCGACGATCGACACCGACGGCTTCCT
GCGGATCACCGGGCGGGTCAAGGACGTGATCAACCGCGGCGGCGAGA
AGATACCGGTCGCGGAGATCGAGCAACTGCTGTTCGGGCACCCCGCG
GTGGAGGACGTCGCGGTGGTGGCCATGCCGGACGAGCGTCTCGGCGA
GCGCGCCTGCGCCTTCGTGGTCCCGGCGGCCGGTACGGAGCTGACCT
TCGAGGAGATGCGGCGCCACCTCGACGGCCACGAGGTGGCCAAGCAG
TACTGGCCGGAGCGTCTGGAGCGGATCGACGCCCTGCCGCGCAACCC
CATCGGCAAGGTGACGAAGTGGGAACTGCGGGCGCTGGCCCATGGTT
TGCGGCCCCACGACGAGGACGCCGGCTGACTCGAG

**Figure 8.8. The nucleotide sequence of *mycA4* cloned into pET28a (pQR2790).**

The ribosome binding site (RBS) is highlighted in red, restriction sites (*Nde*I and *Xho*I) are highlighted in turquoise. The open reading frame is highlighted in grey. Start codons (ATG) are in bold font and the stop codon is underlined. The N-terminal His-tag is highlighted in teal.

## 8.9. The polypeptide sequence expressed from *mycA4* (pQR2790)

**M**GSSHHHHHHSSGLVPRGSHMHLLVKGIQHRVTGEILVPNSKYHAHRALIL
ASLAEGVSRVHGLSDARHVEYTVRLLRDLGVRIARDGDTFVVHGLGGRYR

PRREAVSAGSSGTTLYFMIGLASLSDRAVSVTGQKYFRRRPVGPLLRALEQ
LGVQLESADDCPPVHVQGRRPTGGHVTIAGTLSQWVSGLILLAPFATRHTT
IEVEGELNERPYLELTVAMMRQFGLTVTVSEDWRRFDIEPGQQARSVELTL
PPDIGSAAFGIATAALHPSDVLLRGITTLDGGPADHPEFHFLDVARSMGVP
METDERAGGLRIRQDTPRLTAVDVDCRDIPDMLPILATLATFARGESVFRNI
AHTRLKESDRAAAMLQLNSMGGRLALSDDALRVRGVDGLRGAKLSSFND
HRILMSLAVASSRARGHATLTYPNAYRISYPTFLDAMNTLTVPMSVEDGSA
GPSGGGATAGVPEATEAVGPEAASRVTLPRWLRRRAEARPHDTAVVDVR
SDRDEMITWSELAERVDRAAALLLRLGVRPGENVAYQLPNRVEFVVLSLAA
LRIGAVCCPVIPFFRKRELGFVLRRSKARVLVVADRHRSRLPAEEVLDLVAE
DGPDLNLEHVVVLAAAGGSARLPEAPTGGTAVHDWGQALSETTVDRAVLD
ALAPTPEMTAQLLFTSGTTSEPKGVTQPTRNLVRAAAMEIGHLGLGKQDAI
WVPSPLAHQTGFLYGMVLALVLGVPQILQPEWDAKRALQSLNEHRATFVQ
AATPFLSDLVKAVEESGETPQHLRIFVATGAMVPRALAERAGRVLRTYVCG
AFGTTETCLGALSSPGDEPRQRWGSDGRALDGIELRVTDDEGLVLPAGKE
GNFELRSPTVFDGYLDRPDLTSQAFTEDGWYRTGDLATIDTDGFLRITGRV
KDVINRGGEKIPVAEIEQLLFGHPAVEDVAVVAMPDERLGERACAFVVPAA
GTELTFEEMRRHLDGHEVAKQYWPERLERIDALPRNPIGKVTKWELRALA
HGLRPHDEDAG-

**Figure 8.9. The polypeptide sequence of MycA4 expressed from *mycA4* (pQR2790).**

The start Methionine (M) is in bold font. The N-terminal His-tag is highlighted in teal.

**8.10.The nucleotide sequence of the synthetic *badJKH* operon (pQR2791):**

TTTGTTTAACTTTAAGAAGGAGATATACATATG**GTG**AATCCTTATCTGAA
CGACGATCTGGTCGCGCTCGGCGATCAGGTGCGGCGATTTGCAGCCG
ATCGCATCGCACCCGGATTCCAGGAGCGCGACCGGACGCGCGTGCTC
GACCGCGGGCTGATGCGGGCGATGGGCGAGATGGGCTTCATCGCGCC
GGAATTGCCGGAGCAGTTCGGCGGGCTGGGCCTGGGCTGCCTTGCCG
CCGGCGTCATCCACGAGGAGGTCGCGCGCGCCGATCTCAGCATCTCC
TACATCAATCTGCTGGCGTCGCTGAACGGACAGATCCTGTCGCAATTCG
GCCGGCCGGAGGTGGTGAAGCCGTGGCTTGCGCGTCTGACCGCGGG
CGATGCGCTGGTGGCGATCGCGCTGACGGAAACGCGCGGCGGGTCG
GACGCCGCCAATCTGCGGCTGCGCATCGCGCGCGACGGCGACGACTA
CATCGTCAACGGCGAGAAGACCTCGATCTCCGCCGCCGACCAGGCCG
ACGCAGCCGTGGTGTTCGGCCGCACCGGAACGGTGGAGGCCGGCGC
GCACGGGGTGACGGCGCTGCTGATTCCGATGGACCTGCCGGGGATCA
GCCGCAGCCGCTTCGACTGCCACGGCCAGCGCGCGATCGGCCGCGG
CTCGCTGTTCTTCGAGAACGTCCGAGTCCCCGTTTCGCATCGCCTCGG
CGACGAGAACAAGGGCTTCGTCCAGGTGATGCAGGGCTTCGACTTCTC
CCGCGCGCTGATCGGACTGCAGGTGCTCGCGGTGGCGCGGGTCGCG
CTGGAGGAAACCTGGGCGCATGTGGCCGAGCGGCAGGCCTTCGGCAA
GCCGCTGTCGGCGTTCCAGGGCGTCGCGCATCCGCTCGCCGATCTCG
ACACAAAGGTCGAGGCCGCGCGCCTGCTGTGCCTGCAGGCGTTGTGG
CTGAAGGACAAGGGCGCGCCGCACAGCGCCGAAGCGGCGATGTGCAA
ATGGTGGGCGCCCAAGCTCGCTTACGACGTCGTGCATCAATGCCTGCT
GATGCACGGCCATGGCGGCTACGACCGCGGCGTGATGGAGCAGCGTT
TGCGCGATGTGCTCGGCTTCCAGATCGGCGACGGCACCGCCCAGATC
ATGAAGACCATCATTGCCCGCAGCCGGGCCGGCCGCGCCGCCGTGCC
GGCCTGAGGATCCAGGAGATTCAGA**GTG**CCAGAGAACGTGATCATCGT
CGAAAGGGAAGGGCGGGTCGGGATCGTCACGCTCAATCGGCCTGAAG
TCCTCAACGCGCTCAACGATGAGCTGATGGATGCGCTCGGCGCCGCG
CTGCTCGACTTCGACGCAGACGACGGCATCGGCGCCATCGTGATTGCC
GGTACGGCGCGGGCGTTTGCGGCAGGCGCCGACATCGCGGGCATGG
CGGAATGGAGCTACAGCGACGTCTACAGTTCGAACTTCATCACCCGGA
ACTGGGAGACCATCAAACGGGTCCGCAAGCCGGTGCTGGCGTCGGTC
GCGGGTCTCGCATTCGGGGGCGGATGCGAATTGGCGTTGGCGTGCGA
CATCATCATTGCCGCCCGCAACGCGAAGTTTGCGCTTCCGGAGATCAA
GCTCGGCCTGCTGCCGGGCGCAGGCGGCACGCAGCGCCTGCCGCGC
GCCATTGGCAAGGCCAAGGCGATGGACATGTGCCTGTCCGCACGCCC
GCTCGACGCTGAAGAGGCCGATCGCTATGGCCTCGTCTCACGCGTGGT
CGACGATGACAGGCTCCGCGAGGAAACCATGAAACTCGCCACGACGAT
CGCTTCGTTCTCGGCTCCGGCATTGATGGCGCTGAAGGAAAGTCTCAA
TCGCGCTTTTGAAATTCCGCTTGCGGAGGGCATCCTGTTCGAACGGCG
CGAACTGCACGCCCGCTTCGCGACCGCCGATGCCCGCGAGGGCATTC
GGGCCTTTCTGGAAAAGCGAAAGCCGAGCTTCGTCCATCGCTGAAAGC

TTAGGAGATGTAGT**ATG**GCGCGTCTGCAGAACAAGACAGTCGTGATCA
CGGGCGGCGGCGGCGGCATCGGCGGGGCGACGTGCCGCAAGTTTGC
GCAAGAAGGCGCCAAGGTCGCTGTGCTGGACCTCAATCTCGAAGCCGC
CGAGAAAGTCGCCGCCGAGATCCGCGCGGCCGGTGGCGTCGCCGAAG
CGTTCAAGTGCGACATCGCCGACCGGGCCAGCGTCGATGCCGCCGTG
GCGGCGACACAAACCAAGCTCGGTCCGATCGACGTGCTGGTGAACAAT
GCCGGCTGGGACATCTTCAAGCCGTTCACCAAGTCCGATCCCGGCGAA
TGGGAGAAGCTGATCGCGATCAATTTGACGGGCGCGCTGCATATGCAT
CACGCCATCCTGCCGGGCATGGTCGAGCGTCGCTACGGACGCATTGTC
AATGTCGCCTCCGACGCCGCCCGTGTCGGCTCGTCGGGCGAGGCGGT
GTATGCGGCGTGCAAGGGCGGTCTCGTCGCGTTCTCGAAGACGATCG
CGCGCGAGCACTCCCGTCACGGCATCACCGTCAACGTCGTCTGTCCCG
GCCCGACCGATACGGCGCTGCTGGCCGGCGTCACCAGCGGCGCGCC
CAATCCGGAAAAGCTGGTCGAAGCCTTCACCAAGGCGATCCCGCTCGG
ACGCCTCGGCCAGCCGGACGATCTCGCTGGCGCGATCGTCTTCTTCGG
CAGCGACGATGCGTCGTTCGTCACCGGCCAAGTGCTCAGCGTCTCCGG
CGGGCTGACGATGAACGGTTGACTCGAG

**Figure 8.10. The nucleotide sequence of the synthetic *badJKH* operon cloned into pET29a (pQR2791).**

Ribosome binding sites (RBS) are highlighted red, restriction sites are highlighted in blue. The three open reading frames are highlighted in grey and are in the order: 1- *badJ,* 2- *badK,* 3- *badH*. Start codons are in bold font and stop codons are underlined. The pyrimidine rich region upstream of the pET29a RBS has been highlighted in yellow.

## 8.11. The nucleotide sequence of *badJ* (pQR2792)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCATATG**GTG**AATCCTTATCTGAACGAC
GATCTGGTCGCGCTCGGCGATCAGGTGCGGCGATTTGCAGCCGATCG
CATCGCACCCGGATTCCAGGAGCGCGACCGGACGCGCGTGCTCGACC
GCGGGCTGATGCGGGCGATGGGCGAGATGGGCTTCATCGCGCCGGAA
TTGCCGGAGCAGTTCGGCGGGCTGGGCCTGGGCTGCCTTGCCGCCGG
CGTCATCCACGAGGAGGTCGCGCGCGCCGATCTCAGCATCTCCTACAT
CAATCTGCTGGCGTCGCTGAACGGACAGATCCTGTCGCAATTCGGCCG
GCCGGAGGTGGTGAAGCCGTGGCTTGCGCGTCTGACCGCGGGCGATG
CGCTGGTGGCGATCGCGCTGACGGAAACGCGCGGCGGGTCGGACGC
CGCCAATCTGCGGCTGCGCATCGCGCGCGACGGCGACGACTACATCG
TCAACGGCGAGAAGACCTCGATCTCCGCCGCCGACCAGGCCGACGCA
GCCGTGGTGTTCGGCCGCACCGGAACGGTGGAGGCCGGCGCGCACG
GGGTGACGGCGCTGCTGATTCCGATGGACCTGCCGGGGATCAGCCGC

AGCCGCTTCGACTGCCACGGCCAGCGCGCGATCGGCCGCGGCTCGCT
GTTCTTCGAGAACGTCCGAGTCCCCGTTTCGCATCGCCTCGGCGACGA
GAACAAGGGCTTCGTCCAGGTGATGCAGGGCTTCGACTTCTCCCGCGC
GCTGATCGGACTGCAGGTGCTCGCGGTGGCGCGGGTCGCGCTGGAGG
AAACCTGGGCGCATGTGGCCGAGCGGCAGGCCTTCGGCAAGCCGCTG
TCGGCGTTCCAGGGCGTCGCGCATCCGCTCGCCGATCTCGACACAAAG
GTCGAGGCCGCGCGCCTGCTGTGCCTGCAGGCGTTGTGGCTGAAGGA
CAAGGGCGCGCCGCACAGCGCCGAAGCGGCGATGTGCAAATGGTGGG
CGCCCAAGCTCGCTTACGACGTCGTGCATCAATGCCTGCTGATGCACG
GCCATGGCGGCTACGACCGCGGCGTGATGGAGCAGCGTTTGCGCGAT
GTGCTCGGCTTCCAGATCGGCGACGGCACCGCCCAGATCATGAAGACC
ATCATTGCCCGCAGCCGGGCCGGCCGCGCCGCCGTGCCGGCCTGAG
GATCC

**Figure 8.11. The nucleotide sequence of *badJ* cloned into pET28a (pQR2792).**

The ribosome binding site (RBS) is highlighted in red, restriction sites (*Nde*I and BamHI) are highlighted in turquoise. The open reading frame is highlighted in grey. Start codons (ATG and GTG) are in bold font and the stop codon is underlined. The N-terminal His-tag is highlighted in teal.

## 8.12. The polypeptide sequence expressed from *badJ* (pQR2792)

**M**GSSHHHHHHSSGLVPRGSHM**V**NPYLNDDLVALGDQVRRFAADRIAPGF
QERDRTRVLDRGLMRAMGEMGFIAPELPEQFGGLGLGCLAAGVIHEEVAR
ADLSISYINLLASLNGQILSQFGRPEVVKPWLARLTAGDALVAIALTETRGGS
DAANLRLRIARDGDDYIVNGEKTSISAADQADAAVVFGRTGTVEAGAHGVT
ALLIPMDLPGISRSRFDCHGQRAIGRGSLFFENVRVPVSHRLGDENKGFVQ
VMQGFDFSRALIGLQVLAVARVALEETWAHVAERQAFGKPLSAFQGVAHP
LADLDTKVEAARLLCLQALWLKDKGAPHSAEAAMCKWWAPKLAYDVVHQ
CLLMHGHGGYDRGVMEQRLRDVLGFQIGDGTAQIMKTIIARSRAGRAAVP
A-

**Figure 8.12. The polypeptide sequence of BadJ expressed from *badJ* (pQR2791).**

The start Methionine (M) is in bold font. A Valine (V) corresponding to the GTG start codon of *badJ* has also been highlighted in bold font. The N-terminal His-tag is highlighted in teal.

## 8.13. The nucleotide sequence of *badJ*$_{ATG}$ (pQR2793)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCAT**ATG**AATCCTTATCTGAACGACGAT
CTGGTCGCGCTCGGCGATCAGGTGCGGCGATTTGCAGCCGATCGCAT
CGCACCCGGATTCCAGGAGCGCGACCGGACGCGCGTGCTCGACCGCG
GGCTGATGCGGGCGATGGGCGAGATGGGCTTCATCGCGCCGGAATTG
CCGGAGCAGTTCGGCGGGCTGGGCCTGGGCTGCCTTGCCGCCGGCGT
CATCCACGAGGAGGTCGCGCGCGCCGATCTCAGCATCTCCTACATCAA
TCTGCTGGCGTCGCTGAACGGACAGATCCTGTCGCAATTCGGCCGGCC
GGAGGTGGTGAAGCCGTGGCTTGCGCGTCTGACCGCGGGCGATGCGC
TGGTGGCGATCGCGCTGACGGAAACGCGCGGCGGGTCGGACGCCGC
CAATCTGCGGCTGCGCATCGCGCGCGACGGCGACGACTACATCGTCAA
CGGCGAGAAGACCTCGATCTCCGCCGCCGACCAGGCCGACGCAGCCG
TGGTGTTCGGCCGCACCGGAACGGTGGAGGCCGGCGCGCACGGGGT
GACGGCGCTGCTGATTCCGATGGACCTGCCGGGGATCAGCCGCAGCC
GCTTCGACTGCCACGGCCAGCGCGCGATCGGCCGCGGCTCGCTGTTC
TTCGAGAACGTCCGAGTCCCCGTTTCGCATCGCCTCGGCGACGAGAAC
AAGGGCTTCGTCCAGGTGATGCAGGGCTTCGACTTCTCCCGCGCGCTG
ATCGGACTGCAGGTGCTCGCGGTGGCGCGGGTCGCGCTGGAGGAAAC
CTGGGCGCATGTGGCCGAGCGGCAGGCCTTCGGCAAGCCGCTGTCGG
CGTTCCAGGGCGTCGCGCATCCGCTCGCCGATCTCGACACAAAGGTCG
AGGCCGCGCGCCTGCTGTGCCTGCAGGCGTTGTGGCTGAAGGACAAG
GGCGCGCCGCACAGCGCCGAAGCGGCGATGTGCAAATGGTGGGCGC
CCAAGCTCGCTTACGACGTCGTGCATCAATGCCTGCTGATGCACGGCC
ATGGCGGCTACGACCGCGGCGTGATGGAGCAGCGTTTGCGCGATGTG
CTCGGCTTCCAGATCGGCGACGGCACCGCCCAGATCATGAAGACCATC
ATTGCCCGCAGCCGGGCCGGCCGCGCCGCCGTGCCGGCCTGACTCGA
G

**Figure 8.13. The nucleotide sequence of *badJ*$_{ATG}$ cloned into pET28a (pQR2793).**

The ribosome binding site (RBS) is highlighted in red, restriction sites (*Nde*I and *Xho*I) are highlighted in turquoise. The open reading frame is highlighted in grey. Start codons (ATG) are in bold font and the stop codon is underlined. The N-terminal His-tag is highlighted in teal.

## 8.14. The polypeptide sequence expressed from *badJ*$_{ATG}$ (pQR2793)

**M**GSSHHHHHHSSGLVPRGSH**M**NPYLNDDLVALGDQVRRFAADRIAPGFQ
ERDRTRVLDRGLMRAMGEMGFIAPELPEQFGGLGLGCLAAGVIHEEVARA
DLSISYINLLASLNGQILSQFGRPEVVKPWLARLTAGDALVAIALTETRGGSD
AANLRLRIARDGDDYIVNGEKTSISAADQADAAVVFGRTGTVEAGAHGVTA
LLIPMDLPGISRSRFDCHGQRAIGRGSLFFENVRVPVSHRLGDENKGFVQV

MQGFDFSRALIGLQVLAVARVALEETWAHVAERQAFGKPLSAFQGVAHPL
ADLDTKVEAARLLCLQALWLKDKGAPHSAEAAMCKWWAPKLAYDVVHQC
LLMHGHGGYDRGVMEQRLRDVLGFQIGDGTAQIMKTIIARSRAGRAAVPA-

**Figure 8.14. The polypeptide sequence of BadJ expressed from *badJ*$_{ATG}$ (pQR2793).**

The start Methionine (M) is in bold font. The N-terminal His-tag is highlighted in teal.

## 8.15. The nucleotide sequence of *badK* (pQR2794)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATGACTGGTGGACA
GCAAATGGGTCGCGGATCCAGGAGATTCAGA**GTG**CCAGAGAACGTGAT
CATCGTCGAAAGGGAAGGGCGGGTCGGGATCGTCACGCTCAATCGGC
CTGAAGTCCTCAACGCGCTCAACGATGAGCTGATGGATGCGCTCGGCG
CCGCGCTGCTCGACTTCGACGCAGACGACGGCATCGGCGCCATCGTG
ATTGCCGGTACGGCGCGGGCGTTTGCGGCAGGCGCCGACATCGCGGG
CATGGCGGAATGGAGCTACAGCGACGTCTACAGTTCGAACTTCATCAC
CCGGAACTGGGAGACCATCAAACGGGTCCGCAAGCCGGTGCTGGCGT
CGGTCGCGGGTCTCGCATTCGGGGGCGGATGCGAATTGGCGTTGGCG
TGCGACATCATCATTGCCGCCCGCAACGCGAAGTTTGCGCTTCCGGAG
ATCAAGCTCGGCCTGCTGCCGGGCGCAGGCGGCACGCAGCGCCTGCC
GCGCGCCATTGGCAAGGCCAAGGCGATGGACATGTGCCTGTCCGCAC
GCCCGCTCGACGCTGAAGAGGCCGATCGCTATGGCCTCGTCTCACGC
GTGGTCGACGATGACAGGCTCCGCGAGGAAACCATGAAACTCGCCACG
ACGATCGCTTCGTTCTCGGCTCCGGCATTGATGGCGCTGAAGGAAAGT
CTCAATCGCGCTTTTGAAATTCCGCTTGCGGAGGGCATCCTGTTCGAAC
GGCGCGAACTGCACGCCCGCTTCGCGACCGCCGATGCCCGCGAGGGC
ATTCGGGCCTTTCTGGAAAAGCGAAAGCCGAGCTTCGTCCATCGCTGA
AAGCTT

**Figure 8.15. The nucleotide sequence of *badK* cloned into pET28a (pQR2794).**

The ribosome binding sites (RBS) are highlighted in red, restriction sites (*Bam*HI and *Hind*III) are highlighted in turquoise. The open reading frame is highlighted in grey. Start codons (ATG and GTG) are in bold font and the stop codon is underlined. The N-terminal His-tag is highlighted in teal.

## 8.16. The polypeptide sequence expressed from *badK* (pQR2794)

**M**GSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSRRFR**V**PENVIIVERE
GRVGIVTLNRPEVLNALNDELMDALGAALLDFDADDGIGAIVIAGTARAFAA
GADIAGMAEWSYSDVYSSNFITRNWETIKRVRKPVLASVAGLAFGGGCELA
LACDIIIAARNAKFALPEIKLGLLPGAGGTQRLPRAIGKAKAMDMCLSARPLD
AEEADRYGLVSRVVDDDRLREETMKLATTIASFSAPALMALKESLNRAFEIP
LAEGILFERRELHARFATADAREGIRAFLEKRKPSFVHR-

**Figure 8.16. The polypeptide sequence of BadK expressed from *badK* (pQR2794).**

The start Methionine (M) is in bold font. A Valine (V) corresponding to the GTG start codon of *badK* has also been highlighted in bold font. The N-terminal His-tag is highlighted in teal.

## 8.17. The nucleotide sequence of *badK*$_{ATG}$ (pQR2795)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCAT**ATG**CCAGAGAACGTGATCATCGT
CGAAAGGGAAGGGCGGGTCGGGATCGTCACGCTCAATCGGCCTGAAG
TCCTCAACGCGCTCAACGATGAGCTGATGGATGCGCTCGGCGCCGCG
CTGCTCGACTTCGACGCAGACGACGGCATCGGCGCCATCGTGATTGCC
GGTACGGCGCGGGCGTTTGCGGCAGGCGCCGACATCGCGGGCATGG
CGGAATGGAGCTACAGCGACGTCTACAGTTCGAACTTCATCACCCGGA
ACTGGGAGACCATCAAACGGGTCCGCAAGCCGGTGCTGGCGTCGGTC
GCGGGTCTCGCATTCGGGGGCGGATGCGAATTGGCGTTGGCGTGCGA
CATCATCATTGCCGCCCGCAACGCGAAGTTTGCGCTTCCGGAGATCAA
GCTCGGCCTGCTGCCGGGCGCAGGCGGCACGCAGCGCCTGCCGCGC
GCCATTGGCAAGGCCAAGGCGATGGACATGTGCCTGTCCGCACGCCC
GCTCGACGCTGAAGAGGCCGATCGCTATGGCCTCGTCTCACGCGTGGT
CGACGATGACAGGCTCCGCGAGGAAACCATGAAACTCGCCACGACGAT
CGCTTCGTTCTCGGCTCCGGCATTGATGGCGCTGAAGGAAAGTCTCAA
TCGCGCTTTTGAAATTCCGCTTGCGGAGGGCATCCTGTTCGAACGGCG
CGAACTGCACGCCCGCTTCGCGACCGCCGATGCCCGCGAGGGCATTC
GGGCCTTTCTGGAAAAGCGAAAGCCGAGCTTCGTCCATCGCTGAAAGC
TT

**Figure 8.17. The nucleotide sequence of *badK*$_{ATG}$ cloned into pET28a (pQR2795).**

The ribosome binding site (RBS) is highlighted in red, restriction sites (*Nde*I and *Hind*III) are highlighted in turquoise. The open reading frame is highlighted in grey. Start codons (ATG) are in bold font and the stop codon is underlined. The N-terminal His-tag is highlighted in teal.

## 8.18. The polypeptide sequence of BadK expressed from *badK*$_{ATG}$ (pQR2795)

**M**GSSHHHHHHSSGLVPRGSH**M**PENVIIVEREGRVGIVTLNRPEVLNALNDE
LMDALGAALLDFDADDGIGAIVIAGTARAFAAGADIAGMAEWSYSDVYSSN
FITRNWETIKRVRKPVLASVAGLAFGGGCELALACDIIIAARNAKFALPEIKLG
LLPGAGGTQRLPRAIGKAKAMDMCLSARPLDAEEADRYGLVSRVVDDDRL
REETMKLATTIASFSAPALMALKESLNRAFEIPLAEGILFERRELHARFATAD
AREGIRAFLEKRKPSFVHR-

**Figure 8.18. The polypeptide sequence of BadK expressed from *badK*$_{ATG}$ (pQR2795).**

The start Methionine (M) is in bold font. The N-terminal His-tag is highlighted in teal.

## 8.19. The nucleotide sequence of *badH* (pQR2796)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCATGACTGGTGGACA
GCAAATGGGTCGCGGATCCGAATTCGAGCTCCGTCGACAAGCTTAGGA
GATGTAGT**ATG**GCGCGTCTGCAGAACAAGACAGTCGTGATCACGGGCG
GCGGCGGCGGCATCGGCGGGGCGACGTGCCGCAAGTTTGCGCAAGAA
GGCGCCAAGGTCGCTGTGCTGGACCTCAATCTCGAAGCCGCCGAGAAA
GTCGCCGCCGAGATCCGCGCGGCCGGTGGCGTCGCCGAAGCGTTCAA
GTGCGACATCGCCGACCGGGCCAGCGTCGATGCCGCCGTGGCGGCGA
CACAAACCAAGCTCGGTCCGATCGACGTGCTGGTGAACAATGCCGGCT
GGGACATCTTCAAGCCGTTCACCAAGTCCGATCCCGGCGAATGGGAGA
AGCTGATCGCGATCAATTTGACGGGCGCGCTGCATATGCATCACGCCA
TCCTGCCGGGCATGGTCGAGCGTCGCTACGGACGCATTGTCAATGTCG
CCTCCGACGCCGCCCGTGTCGGCTCGTCGGGCGAGGCGGTGTATGCG
GCGTGCAAGGGCGGTCTCGTCGCGTTCTCGAAGACGATCGCGCGCGA
GCACTCCCGTCACGGCATCACCGTCAACGTCGTCTGTCCCGGCCCGAC
CGATACGGCGCTGCTGGCCGGCGTCACCAGCGGCGCGCCCAATCCGG
AAAAGCTGGTCGAAGCCTTCACCAAGGCGATCCCGCTCGGACGCCTCG
GCCAGCCGGACGATCTCGCTGGCGCGATCGTCTTCTTCGGCAGCGAC
GATGCGTCGTTCGTCACCGGCCAAGTGCTCAGCGTCTCCGGCGGGCT
GACGATGAACGGTTGACTCGAG

**Figure 8.19. The nucleotide sequence of *badH* cloned into pET28a (pQR2796).**

The ribosome binding sites (RBS) are highlighted in red, restriction sites (*Hind*III and *Xho*I) are highlighted in turquoise. The open reading frame is highlighted in grey. Start codons (ATG) are in bold font and the stop codons are underlined. The N-terminal His-tag is highlighted in teal.

## 8.20. The polypeptide sequence expressed from *badH* (pQR2796)

**M**ARLQNKTVVITGGGGGIGGATCRKFAQEGAKVAVLDLNLEAAEKVAAEIR
AAGGVAEAFKCDIADRASVDAAVAATQTKLGPIDVLVNNAGWDIFKPFTKS
DPGEWEKLIAINLTGALHMHHAILPGMVERRYGRIVNVASDAARVGSSGEA
VYAACKGGLVAFSKTIAREHSRHGITVNVVCPGPTDTALLAGVTSGAPNPE
KLVEAFTKAIPLGRLGQPDDLAGAIVFFGSDDASFVTGQVLSVSGGLTMNG-

**Figure 8.20. The polypeptide sequence of BadH expressed from *badH* (pQR2796).**

The start Methionine (M) is in bold font. Due to an unwanted stop codon introduced into the nucleotide sequence no N-terminal His-tag was attached.

## 8.21. The nucleotide sequence of *badH*$_{NheI}$ (pQR2797)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCATATGGCTAGCAGGAGATGTAGT**AT**
**G**GCGCGTCTGCAGAACAAGACAGTCGTGATCACGGGCGGCGGCGGCG
GCATCGGCGGGGCGACGTGCCGCAAGTTTGCGCAAGAAGGCGCCAAG
GTCGCTGTGCTGGACCTCAATCTCGAAGCCGCCGAGAAAGTCGCCGCC
GAGATCCGCGCGGCCGGTGGCGTCGCCGAAGCGTTCAAGTGCGACAT
CGCCGACCGGGCCAGCGTCGATGCCGCCGTGGCGGCGACACAAACCA
AGCTCGGTCCGATCGACGTGCTGGTGAACAATGCCGGCTGGGACATCT
TCAAGCCGTTCACCAAGTCCGATCCCGGCGAATGGGAGAAGCTGATCG
CGATCAATTTGACGGGCGCGCTGCATATGCATCACGCCATCCTGCCGG
GCATGGTCGAGCGTCGCTACGGACGCATTGTCAATGTCGCCTCCGACG
CCGCCCGTGTCGGCTCGTCGGGCGAGGCGGTGTATGCGGCGTGCAAG
GGCGGTCTCGTCGCGTTCTCGAAGACGATCGCGCGCGAGCACTCCCG
TCACGGCATCACCGTCAACGTCGTCTGTCCCGGCCCGACCGATACGGC
GCTGCTGGCCGGCGTCACCAGCGGCGCGCCCAATCCGGAAAAGCTGG
TCGAAGCCTTCACCAAGGCGATCCCGCTCGGACGCCTCGGCCAGCCG
GACGATCTCGCTGGCGCGATCGTCTTCTTCGGCAGCGACGATGCGTCG
TTCGTCACCGGCCAAGTGCTCAGCGTCTCCGGCGGGCTGACGATGAAC
GGTTGACTCGAG

**Figure 8.21. The nucleotide sequence of *badH*$_{NheI}$ cloned into pET28a (pQR2797).**

The ribosome binding sites (RBS) are highlighted in red, restriction sites (*Nhe*I and *Xho*I) are highlighted in turquoise. The open reading frame is highlighted in grey. Start codons (ATG) are in bold font and the stop codon is underlined. The N-terminal His-tag is highlighted in teal.

## 8.22. The polypeptide sequence expressed from *badH*$_{NheI}$ (pQR2797)

**M**GSSHHHHHHSSGLVPRGSHMASRRCS**M**ARLQNKTVVITGGGGGIGGAT
CRKFAQEGAKVAVLDLNLEAAEKVAAEIRAAGGVAEAFKCDIADRASVDAA
VAATQTKLGPIDVLVNNAGWDIFKPFTKSDPGEWEKLIAINLTGALHMHHAI
LPGMVERRYGRIVNVASDAARVGSSGEAVYAACKGGLVAFSKTIAREHSR
HGITVNVVCPGPTDTALLAGVTSGAPNPEKLVEAFTKAIPLGRLGQPDDLA
GAIVFFGSDDASFVTGQVLSVSGGLTMNG-

**Figure 8.22. The polypeptide sequence of BadH expressed from *badH*$_{NheI}$ (pQR2797).**

The start Methionine (M) is in bold font. The N-terminal His-tag is highlighted in teal.

## 8.23. The nucleotide sequence of *chcoADH* (pQR2798)

AGGAGATATACAT**ATG**TATATTAACACCGAAACCGAAGATCTGAAAACC
GCGATTGATGCGATTCGCAAAGCGGTGAAAGATCGCATTGCGCCGCTG
GCGGCGGAAGTGGATGATAGCGGCGTGATTAAACCGGAAATTTATGAT
CTGCTGTGGGATCTGGGCCTGATGACCGTGACCTATCCGCCGGAATAT
GGCGGCAGCGAAACCAACCCGGGCACCCTGCTGTGCATTGGCTGCGA
AGAAATTGCGAAAGCGTGCGCGAGCACCGCGCTGCTGCTGATTATTCA
GGCGGTGGGCAGCTTTCCGCTGATGCATGGCGGCCGCAAAGAACTGC
TGGATCGCATTGCGCCGCGCATTGTGAACAACCGCGAACTGGCGGGCT
ATCTGGTGAGCGAACCGGGCGCGGGCAGCGATGTGAAAGCGATTCGC
ACCAAAGCGGTGAAAGATGGCAACGATTGGGTGATTAACGGCACCAAA
TGCTGGGCGACCAACGGCCCGATTGCGAGCTTTTATAGCTGCCTGTGC
CGCACCAAAGATGATAAAGGCGTGCAGGGCTATAGCTTTTTTCTGGTGG
AACGCAACACCCCGGGCCTGAGCGTGGGCAAAATTGAACATAAAATGG
GCATGCGCGGCAGCCAGACCAGCGAAGTGATTCTGGAAGATGTGCGC
GTGCCGGCGGAAAACCTGCTGGGCGAACTGAACAACGGCTTTAAACTG
GCGATGAAAGATTTTGATATGAGCCGCCCGGCGATTGCGGCGCAGGCG
CTGGGCATTAGCGAAGGCGCGTTTGCGCAGATGGAAACCTATAGCCGC
GAACGCTATACCTTTGGCAAACCGCTGTGCGAACATGGCATGATTACCC
AGATTATTGCGGATAGCGCGGCGCTGATTGAAGCGGGCCGCGGCCTG
ATTTATCAGGCGGCGGATCTGTATGATAAAGGCAAAAAAAACACCAAAC
TGGCGAGCATGGCGAAATTTTTTATGGGCGATGCGGCGGTGAAAATTA
CCACCGATGCGATTCAGGTGTTTGGCGGCTATGGCTATACCCATGATTA

TCCGGTGGAACGCATGTTTCGCGATGCGAAACTGACCCAGATTTTTGAA
GGCGCGAACCAGATTCAGCGCATTGTGGTGGCGCGCGAAATTCGCGAT
GAACAGAGCAAACATCATCATCATCATCATTAACTCGAG

**Figure 8.23. The nucleotide sequence of *ChcoADH* cloned into pET29a (pQR2798).**

The ribosome binding site (RBS) is highlighted in red, restriction sites (*Nde*I and *Xho*I) are highlighted in turquoise. The open reading frame is highlighted in grey. The start codon (ATG) is in bold font and the stop codon is underlined. The C-terminal His-tag is highlighted in teal.

## 8.24. The polypeptide sequence expressed from *chcoADH* (pQR2798)

**M**YINTETEDLKTAIDAIRKAVKDRIAPLAAEVDDSGVIKPEIYDLLWDLGLMT
VTYPPEYGGSETNPGTLLCIGCEEIAKACASTALLLIIQAVGSFPLMHGGRK
ELLDRIAPRIVNNRELAGYLVSEPGAGSDVKAIRTKAVKDGNDWVINGTKC
WATNGPIASFYSCLCRTKDDKGVQGYSFFLVERNTPGLSVGKIEHKMGMR
GSQTSEVILEDVRVPAENLLGELNNGFKLAMKDFDMSRPAIAAQALGISEG
AFAQMETYSRERYTFGKPLCEHGMITQIIADSAALIEAGRGLIYQAADLYDK
GKKNTKLASMAKFFMGDAAVKITTDAIQVFGGYGYTHDYPVERMFRDAKLT
QIFEGANQIQRIVVAREIRDEQSKHHHHHH-

**Figure 8.24. The polypeptide sequence of ChCoaDH expressed from *chcoADH* (pQR2798).**

The start Methionine (M) is in bold font. The C-terminal His-tag is highlighted in teal.

## 8.25. The nucleotide sequence of *fcbC* (pQR2799)

**ATG**CATCGCACTTCGAATGGTTCTCACGCAACTGGTGGTAATTTGCCTG
ACGTCGCTAGCCACTACCCGGTCGCCTACGAGCAGACGCTGGATGGTA
CCGTTGGCTTTGTGATTGACGAAATGACCCCGGAGCGTGCAACCGCGT
CCGTCGAGGTTACGGATACGCTGCGTCAACGCTGGGGTCTGGTCCACG
GTGGCGCGTATTGTGCGCTGGCAGAAATGCTGGCGACCGAAGCAACC
GTGGCCGTTGTTCATGAGAAAGGCATGATGGCCGTGGGCCAGAGCAAC
CATACCAGCTTTTTCCGTCCGGTGAAGGAAGGCCACGTGCGCGCAGAG
GCTGTGCGCATCCACGCGGGCAGCACGACCTGGTTCTGGGATGTTAGC
CTGCGTGACGACGCCGGTCGTCTGTGCGCGGTCAGCTCCATGAGCATC
GCGGTTCGTCCGCGTCGTGATGGTGGTAGCCACCACCACCATCATCAT
TAA

**Figure 8.25. The nucleotide sequence of *fcbC* encoded within the pD451-SR vector (pQR2799).**

The open reading frame is highlighted in grey. The start codon (ATG) is in bold font and the stop codon is underlined. The C-terminal His-tag is highlighted in teal.

## 8.26. The polypeptide sequence expressed from *fcbC* (pQR2799)

**M**HRTSNGSHATGGNLPDVASHYPVAYEQTLDGTVGFVIDEMTPERATASV
EVTDTLRQRWGLVHGGAYCALAEMLATEATVAVVHEKGMMAVGQSNHTS
FFRPVKEGHVRAEAVRIHAGSTTWFWDVSLRDDAGRLCAVSSMSIAVRPR
RDGGSHHHHHH-

**Figure 8.26. The polypeptide sequence of FcbC expressed from *fcbC* (pQR2799).**

The start Methionine (M) is in bold font. The C-terminal His-tag is highlighted in teal.

## 8.27. The nucleotide sequence of *pa2801* (pQR2800)

AGGAGAGAATTC**ATG**GCTGACAGACAATTGCTACACACGGCCCATATC
CCGGTGCGCTGGGGCGACATGGACAGCTACGGGCACGTCAACAACAC
CCTCTATTTCCAGTACCTGGAAGAGGCCCGGGTGGCCTGGTTCGAAAC
CCTCGGCATCGACCTGGAAGGCGCTGCCGAGGGGCCGGTCGTGCTGC
AAAGCCTGCACACCTACCTGAAGCCGGTGGTGCATCCGGCCACCGTGG
TGGTGGAACTGTACGCCGGCAGGCTGGGCACCAGCAGCCTGGTACTG
GAACATCGCCTGCACACCCTGGAAGATCCGCAAGGCACCTATGGCGAA
GGCCACTGCAAGCTGGTCTGGGTGCGCCACGCGGAAAACCGTTCGAC
GCCCGTGCCGGACAGCATACGCGCCGCGATCGCCTGAACTAGT

**Figure 8.27. The nucleotide sequence of *pa2801* encoded within pET29a (pQR2800).**

The open reading frame is highlighted in grey. The start codon (ATG) is in bold font and the stop codon is underlined. The C-terminal His-tag is highlighted in teal. The restriction sites (*Eco*RI and *Spe*I) are highlighted in turquoise.

## 8.28. The polypeptide sequence expressed from *pa2801* (pQR2800)

**M**ADRQLLHTAHIPVRWGDMDSYGHVNNTLYFQYLEEARVAWFETLGIDLE
GAAEGPVVLQSLHTYLKPVVHPATVVVELYAGRLGTSSLVLEHRLHTLEDP
QGTYGEGHCKLVWVRHAENRSTPVPDSIRAAIA-

**Figure 8.28. The polypeptide sequence of PA2801 expressed from *pa2801* encoded within pET29a (pQR2800).**

The start Methionine (M) is shown in bold font.

## 8.29. The nucleotide sequence of *rpaL* (pQR2801)

AGGAGATATACAT**ATG**AGCAAAAGCCTGATTGATCTGATTAGCATTCTG
GATCTGGAACCGCTGGAAGTGAACCTGTTTCGCGGCACCAGCCCGCAG
ACCAGCTGGCAGCGCGTGTTTGGCGGCCAGGTGATTGGCCAGGCGAT
GGTGGCGGGCTGCCGCACCGTGGAAAACCGCCTGCCGCATAGCCTGC
ATTGCTATTTTATTCTGCCGGGCGATCCGGCGGTGCCGATTATTTATGA
AGTGGAACGCCTGCGCGATGGCAAAAGCTATACCACCCGCCGCGTGAC
CGCGATTCAGCATGGCCAGGCGATTTTTAGCCTGATGATGAGCTTTCAT
GATGATGAAGAAACCGAATTTGATCATCAGGATAAAATGCCGGATGTGC
CGCCGCCGGAAGCGCTGAGCGCGGAAGAAATTGTGAAACAGCCGTTTT

TTAAAGAAATGCCGGATTTTATTAAACGCTATTATGAAAGCGATCGCCC
GATTGAACTGCGCCCGGTGGAACTGAGCCGCTATTTTGGCCAGAAAAT
TGAAGATGGCCGCATTCATGTGTGGATTCGCACCGCGGCGAAACTGCC
GGATGATCCGGCGCTGCACATGTGCGCGCTGGCGTATGCGAGCGATTT
TAGCCTGCTGGATGCGGTGATGGCGCGCTATGGCCGCACCCTGTTTGA
TAAACGCATGATGCCGGCGAGCCTGGATCATGCGATGTGGTTTCATCG
CCCGTTTCGCGCGGATGAATGGCTGCTGTATGTGCAGGATAGCCCGAG
CGCGCAGAGCGGCCGCGGCCTGACCCGCGGCATGATTTATAAAGCGG
ATGGCACCCTGGTGGCGAGCGTGGCGCAGGAAGGCAGCGTGCGCCAG
CGCCGCGATCTGCCGCGCACCTAATAGCTCGAG

**Figure 8.29. The nucleotide sequence of *rpaL* encoded within pET29a (pQR2801).**

The open reading frame is highlighted in grey. The start codon (ATG) is in bold font and the stop codons are underlined. The restriction sites (*Nde*I and *Xho*I) are highlighted in turquoise.

## 8.30. The polypeptide sequence expressed from *rpaL* (pQR2801)

**M**SKSLIDLISILDLEPLEVNLFRGTSPQTSWQRVFGGQVIGQAMVAGCRTV
ENRLPHSLHCYFILPGDPAVPIIYEVERLRDGKSYTTRRVTAIQHGQAIFSLM
MSFHDDEETEFDHQDKMPDVPPPEALSAEEIVKQPFFKEMPDFIKRYYESD
RPIELRPVELSRYFGQKIEDGRIHVWIRTAAKLPDDPALHMCALAYASDFSL
LDAVMARYGRTLFDKRMMPASLDHAMWFHRPFRADEWLLYVQDSPSAQS
GRGLTRGMIYKADGTLVASVAQEGSVRQRRDLPRT--

**Figure 8.30. The polypeptide sequence of RpaL expressed from *rpaL* encoded within pET29a (pQR2801).**

The start Methionine (M) is shown in bold font.

## 8.31. The nucleotide sequence of *ydbB* (pQR2802)

AGGAGAGAATTC**ATG**ATCTGGAAACGCCATTTAACGCTCGACGAACTGA
ACGCCACCAGCGATAACACAATGGTGGCGCATCTGGGAATTGTGTATA
CCCGTCTGGGCGATGATGTGCTGGAAGCCGAAATGCCGGTTGATACCC
GTACTCATCAGCCGTTCGGTTTACTACATGGCGGCGCGTCGGCGGCGC
TGGCGGAAACGCTGGGATCGATGGCCGGATTTATGATGACCCGCGACG
GACAGTGTGTGGTAGGCACAGAACTTAATGCAACACACCATCGCCCGG
TGTCTGAGGGAAAGGTACGCGGCGTCTGCCAGCCGCTGCATCTTGGTC
GGCAAAATCAGAGCTGGGAAATCGTCGTTTTCGATGAACAGGGGCGGC
GTTGCTGCACTTGTCGGCTGGGTACGGCAGTTTTGGGATGAACTAGT

**Figure 8.31. The nucleotide sequence of *ybdB* encoded within pET29a (pQR2802).**

The open reading frame is highlighted in grey. The start codon (ATG) is in bold font and the stop codon is underlined. The restriction sites (*Eco*RI and *Spe*I) are highlighted in turquoise.

## 8.32. The polypeptide sequence expressed from *ydbB* (pQR2802)

**M**IWKRHLTLDELNATSDNTMVAHLGIVYTRLGDDVLEAEMPVDTRTHQPFG
LLHGGASAALAETLGSMAGFMMTRDGQCVVGTELNATHHRPVSEGKVRG
VCQPLHLGRQNQSWEIVVFDEQGRRCCTCRLGTAVLG-

**Figure 8.32. The polypeptide sequence of YbdB expressed from *ybdB* encoded within pET29a (pQR2802).**

The start Methionine (M) is shown in bold font.

## 8.33. The nucleotide sequence of *aliA* (pQR2803)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCAT**ATG**GAATTCGATGCCGTATTGCTG
CCGGCGCGTCGGGCCGAGAGCATCGCCCGCGGGTTCTGGCACGACC
GCACCATCAACGACGATCTCGATGCCTGCGTCGCCGCCTGCCCGGACA
AGACCGCCCTGACGGCGGTGCGTCTCGACGGCGGTCAGCCGCGCCGT
TTCAGCTATCGCGAACTCGCAATGCTCGCTGATCGCGTCGCGGTCGGG
CTGTCGCGGCTCGGGGTCGCGCGCAACGATGTGGTGGCGATGCAACT
ACCGAACTGGTGGCAGTTCAGCATTCTCTACCTCGCCTGTTCGCGCATC
GGCGCGGTGCTCAATCCCTTGATGCCGATCTTCCGCGAGCGCGAATTG
TCGTTCATGCTGAAGCACGGCGAGGCCAAGGTGCTCGTCGTTCCGAAG
ACCTTCCGCAATTTCGATCACGAAGCCATGGCGCGCGGGCTGCAGCCG

TCTTTGCCGGCGCTGCAGCGGATCGTGGTGGTCGACGGCGGCGGGGC
AGATGATTTCAACGCGCTACTGACCATCCCGGAATGGGAGCAGCAGCC
GGACGCGCAGAAGATCCTGAGCAGCAACCGGCCGGGTCCGGACGACA
TCACGCAACTGATCTACACCTCGGGGACCACGGGCGAGCCCAAGGGC
GTGATGCACAGCGCCAACACGCTGATGGCCAATATCATTCCCTATGCC
GAGCGGCTCCGGCTCGGCGCAGACGACGTCATCCTGATGGCCTCGCC
GATGGCGCATCAGACCGGCTTCATGTACGGGCTGATGCTGCCGATCAT
GCTGAAGGCGAGCGCCGTGCTCCAGGACGTCTGGGACCCGGCCAAGG
CGGCCGAACTGATCCGGGCGGAGCAAGTGAGCTTCACCATGGCCTCC
ACCCCGTTCCTCACCGACCTCACCCGCGTCGTGAAGGAGAGCGCGCA
GCCGGTTCCGAGCCTCAAGACCTTCCTCTGCGCCGGCGCGCCGATTCC
GGGGCCGCTGGTCGAACAGGCGCAGATCGGCCTCGGCGCCAAGATCG
TTTCTGCTTGGGGGATGACGGAGAACGGCGCCGTCACTCTGATCAAGC
TCGACGATGACGACAGACTGGCCTCGACCACCGACGGCTGCGCGCTG
CCGGGCGTCGAGGTGAGGGTGGTCGACGGCGACGGCAACGCGTTGC
CGGCCGGCCAGATAGGTCGCCTCGTGGTGCGGGCGTGCTCGAATTTC
GGCGGCTATCTGAAACGCCCGCACTGGAACGCCACCGACGCTGAGGG
CTGGTTCGACACCGGCGATCTCGCCTACATGACCGCCGAGGGCTACAT
CCGGATCAGCGGCCGCAGCAAGGACGTGATCATTCGCGGCGGCGAGA
ACATCCCGGTCGTCGAAGTCGAAGCCCTGCTCTACAAGCATCCTGCTG
TCGCGCAGGTCGCCATCGTCGCCTATCCCGATGAGCGCCTCGGCGAG
CGCGCCTGCGCCGTGGTGGTTCCCAGGAGCGGCGGGGCGATCGATTT
CCCAGCGATGGTCGAGTTTCTGAAATCTCAGAAACTCGCGGTCCAGTA
CATTCCTGAACGGCTGATCGTTCGTGAAGCCATGCCCGCGACGCCTTC
CGGCAAGATTCAGAAATTCCGCCTGCGCGAGATGCTGCAGCAGAACGA
CCTTTAGCTCGAG

**Figure 8.33. The nucleotide sequence of *aliA* encoded within pET28a (pQR2803).**

The open reading frame is highlighted in grey. The start codons (ATG) are in bold font and the stop codon is underlined. The restriction sites (*Nde*I and *Xho*I) are highlighted in turquoise. The ribosome binding site is highlighted in red and the N-terminal His-tag is in turquoise.

## 8.34. The polypeptide sequence expressed from *aliA* (pQR2803)

**M**GSSHHHHHHSSGLVPRGSHMEFDAVLLPARRAESIARGFWHDRTINDDL
DACVAACPDKTALTAVRLDGGQPRRFSYRELAMLADRVAVGLSRLGVARN
DVVAMQLPNWWQFSILYLACSRIGAVLNPLMPIFRERELSFMLKHGEAKVL
VVPKTFRNFDHEAMARGLQPSLPALQRIVVVDGGGADDFNALLTIPEWEQ
QPDAQKILSSNRPGPDDITQLIYTSGTTGEPKGVMHSANTLMANIIPYAERL
RLGADDVILMASPMAHQTGFMYGLMLPIMLKASAVLQDVWDPAKAAELIRA
EQVSFTMASTPFLTDLTRVVKESAQPVPSLKTFLCAGAPIPGPLVEQAQIGL
GAKIVSAWGMTENGAVTLIKLDDDDRLASTTDGCALPGVEVRVVDGDGNA
LPAGQIGRLVVRACSNFGGYLKRPHWNATDAEGWFDTGDLAYMTAEGYIR
ISGRSKDVIIRGGENIPVVEVEALLYKHPAVAQVAIVAYPDERLGERACAVVV
PRSGGAIDFPAMVEFLKSQKLAVQYIPERLIVREAMPATPSGKIQKFRLREM
LQQNDL-

**Figure 8.34. The polypeptide sequence of AliA expressed from *aliA* encoded within pET28a (pQR2803).**

The start Methionine (M) is in bold font and the N-terminal His-tag is highlighted in turquoise.

## 8.35. The nucleotide sequence of *rpaL*$_{His}$ (pQR2804)

AGGAGATATACC**ATG**GGCAGCAGCCATCATCATCATCATCACAGCAGC
GGCCTGGTGCCGCGCGGCAGCCAT**ATG**AGCAAAAGCCTGATTGATCTG
ATTAGCATTCTGGATCTGGAACCGCTGGAAGTGAACCTGTTTCGCGGCA
CCAGCCCGCAGACCAGCTGGCAGCGCGTGTTTGGCGGCCAGGTGATT
GGCCAGGCGATGGTGGCGGGCTGCCGCACCGTGGAAAACCGCCTGCC
GCATAGCCTGCATTGCTATTTTATTCTGCCGGGCGATCCGGCGGTGCC
GATTATTTATGAAGTGGAACGCCTGCGCGATGGCAAAAGCTATACCACC
CGCCGCGTGACCGCGATTCAGCATGGCCAGGCGATTTTTAGCCTGATG
ATGAGCTTTCATGATGATGAAGAAACCGAATTTGATCATCAGGATAAAAT
GCCGGATGTGCCGCCGCCGGAAGCGCTGAGCGCGGAAGAAATTGTGA
AACAGCCGTTTTTTAAAGAAATGCCGGATTTTATTAAACGCTATTATGAA
AGCGATCGCCCGATTGAACTGCGCCCGGTGGAACTGAGCCGCTATTTT
GGCCAGAAAATTGAAGATGGCCGCATTCATGTGTGGATTCGCACCGCG
GCGAAACTGCCGGATGATCCGGCGCTGCACATGTGCGCGCTGGCGTA
TGCGAGCGATTTTAGCCTGCTGGATGCGGTGATGGCGCGCTATGGCCG
CACCCTGTTTGATAAACGCATGATGCCGGCGAGCCTGGATCATGCGAT
GTGGTTTCATCGCCCGTTTCGCGCGGATGAATGGCTGCTGTATGTGCA
GGATAGCCCGAGCGCGCAGAGCGGCCGCGGCCTGACCCGCGGCATG
ATTTATAAAGCGGATGGCACCCTGGTGGCGAGCGTGGCGCAGGAAGG
CAGCGTGCGCCAGCGCCGCGATCTGCCGCGCACCTAATAGCTCGAG

**Figure 8.35. The nucleotide sequence of *rpaL*$_{His}$ encoded within pET28a (pQR2804).**

The open reading frame is highlighted in grey. The start codons (ATG) are in bold font and the stop codons are underlined. The restriction sites (*Nde*I and *Xho*I) are highlighted in turquoise. The ribosome binding site is highlighted in red and the N-terminal His-tag is in turquoise.

## 8.36. The polypeptide sequence expressed from *rpaL*$_{His}$ (pQR2804)

**M**GSSHHHHHHSSGLVPRGSHMSKSLIDLISILDLEPLEVNLFRGTSPQTSW
QRVFGGQVIGQAMVAGCRTVENRLPHSLHCYFILPGDPAVPIIYEVERLRD
GKSYTTRRVTAIQHGQAIFSLMMSFHDDEETEFDHQDKMPDVPPPEALSA
EEIVKQPFFKEMPDFIKRYYESDRPIELRPVELSRYFGQKIEDGRIHVWIRTA
AKLPDDPALHMCALAYASDFSLLDAVMARYGRTLFDKRMMPASLDHAMW
FHRPFRADEWLLYVQDSPSAQSGRGLTRGMIYKADGTLVASVAQEGSVR
QRRDLPRT--

**Figure 8.36. The polypeptide sequence of RpaL expressed from *rpaL*$_{His}$ encoded within pET28a (pQR2804).**

The start Methionine (M) is in bold font and the N-terminal His-tag is highlighted in turquoise.

## 8.37. The nucleotide sequence of *chcB* from *S. rishiriensis*

GGAGGGGAAAACGCC**ATG**GTCGACACCGTGCTCTACGAGGTGAGCAA
CGGGCTCGCGACGATCACGCTGAACCGCCCCGAGGCGATGAACGCGC
TGAACATCGCGACCAAGGTCGCCCTCCGGGACGCGGTGCGGTCCGCG
GCCGACGACGACGCCGTACGGGCCGTGCTGCTGACCGCGGCCGGCG
ACCGGGCGTTCTGCGTGGGGCAGGACCTCAAGGAGCACATCGGACTG
CTGGCCGAAGGCTCGAAGCAGGTCATGAGCACGGTCGCGGAGCACTA
CAACCCGATCGTGCGGGCGCTGACGGAGGCCCCGAAGCCGGTCGTCG
CCGCGGTGAACGGGGTCGCCGCGGGCGCCGGCCTCGGTTTCGCGCT
GGCCGCGGACTACCGCGTCGTGTCCGACACGGCGTCCTTCAACACCTC
GTTCGCGGGGGTCGCGCTCACCGCCGACTCGGGCGTCTCGTGGACGC
TGCCCCGGGTGATCGGTCCGGGCCGGGCCGCCGACCTGCTGCTCTTC
CCGCGCAGCATCAAGGCCCAGGAGGCCTACGAGCTGGGCATCGCGAG
CCGGGTCGTGCCGGCGGACGCGCTGCGCGCGGAGGCGGAGAAGGTG
GCCCGCGCCCTCGCCGCGGGGCCGACGGTGGCCTATGCGGCGATCAA
GGAGGCCGTCGCCTACGGGTCGACGCACACCCTCTCCGAGACGCTGG
ACAAGGAGGACGAGCTGCAGACGCGGGCCGGGGCCTCCGAGGACCA
CGCGATCGCGGTGCGGGCGTTCGTCCACAAGGAGACGCCGGAGTACC
TGGGCCGCTGA

**Figure 8.37. The nucleotide sequence of *chcB* from the genome of *S. rishiriensis*.**

The start codon has been highlighted in bold text and the stop codon underlined. A proposed ribosome binding site has been highlighted in red.

## 8.38. The 16S rRNA nucleotide sequence identified within the genome of *S. rishiriensis*

CATTCACGGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTG
CTTAACACATGCAAGTCGAACGATGAACCACTTCGGTGGGGATTAGTG
GCGAACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACA
AGCCCTGGAAACGGGGTCTAATACCGGATAACACTTCCACTCGCATGG
GTGGAGGTTAAAAGCTCCGGCGGTGAAGGATGAGCCCGCGGCCTATC
AGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAGCCG
GCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGAC
TCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCT
GATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACC
TCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTACCTGCAGAAGAAGCG
CCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGC
GTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTCTGTCGCG
TCGGATGTGAAAGCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGG
GCAGACTAGAGTGTGGTAGGGGAGATCGGAATTCCTGGTGTAGCGGTG

AAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGATCTCTG
GGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGAT
TAGATACCCTGGTAGTCCACGCCGTAAACGGTGGGAACTAGGTGTTGG
CGACATTCCACGTCGTCGGTGCCGCAGCTAACGCATTAAGTTCCCCGC
CTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGG
CCCGCACAAGCAGCGGAGCATGTGGCTTAATTCGACGCAACGCGAAGA
ACCTTACCAAGGCTTGACATACACCGGAAACGGCCAGAGATGGTCGCC
CCCTTGTGGTCGGTGTACAGGTGGTGCATGGCTGTCGTCAGCTCGTGT
CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTCTGT
GTTGCCAGCATGCCCTTCGGGGTGATGGGGACTCACAGGAGACTGCC
GGGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAGTCATCATGCCC
CTTATGTCTTGGGCTGCACACGTGCTACAATGGCCGGTACAAAGAGCT
GCGAAACCGTGAGGTGGAGCGAATCTCAAAAAGCCGGTCTCAGTTCGG
ATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTTGCTAGTAATCG
CAGATCAGCATTGCTGCGGTGAATACGTTCCCGGGCCTTGTACACACC
GCCCGTCACGTCACGAAAGTCGGTAACACCCGAAGCCGGTGGCCCAA
CCCCTTGTGGGAGGGAGCTGTCGAAGGTGGGACTGGCGATTGGGACG
AAGTCGTAACAAGGTAGCCGTACCGGAAGGTGCGGCTGGATCACCTCC
TTTCT

**Figure 8.38. The 16s rRNA nucleotide sequence from the genome of *S. rishiriensis.***

The 3' region has been underlined, this region was used to identify complementarity with the proposed ribosome binding sites of all genes identified from *S. rishiriensis*.

**Table 8.1. A table of transformed strains.**

| Plasmid | Organism |
|---|---|
| pQR2786 | *Escherichia coli BL21 (DE3)* |
| pQR2786 + pQ-Tf2 | *Escherichia coli BL21 (DE3)* |
| pQR2786 + pTF16 | *Escherichia coli BL21 (DE3)* |
| pQR2786 + pGro7 | *Escherichia coli BL21 (DE3)* |
| pQR2786 + pKJE7 | *Escherichia coli BL21 (DE3)* |
| pQR2786 + pQ-KJE8 | *Escherichia coli BL21 (DE3)* |
| pQR2787 | *Escherichia coli BL21 (DE3)* |
| pQR2788 | *Escherichia coli BL21 (DE3)* |
| pQR2789 | *Escherichia coli BL21 (DE3)* |
| pQR2790 | *Escherichia coli BL21 (DE3)* |
| pQR2790 + pQ-Tf2 | *Escherichia coli BL21 (DE3)* |
| pQR2790 + pTF16 | *Escherichia coli BL21 (DE3)* |
| pQR2790 + pGro7 | *Escherichia coli BL21 (DE3)* |
| pQR2790 + pKJE7 | *Escherichia coli BL21 (DE3)* |
| pQR2790 + pQ-KJE8 | *Escherichia coli BL21 (DE3)* |
| pQR2791 | *Escherichia coli BL21 (DE3)* |
| pQR2792 | *Escherichia coli BL21 (DE3)* |
| pQR2793 | *Escherichia coli BL21 (DE3)* |
| pQR2794 | *Escherichia coli BL21 (DE3)* |
| pQR2795 | *Escherichia coli BL21 (DE3)* |
| pQR2796 | *Escherichia coli BL21 (DE3)* |
| pQR2797 | *Escherichia coli BL21 (DE3)* |
| pQR2798 | *Escherichia coli BL21 (DE3)* |
| pQR2799 | *Escherichia coli BL21 (DE3)* |
| pQR2800 | *Escherichia coli BL21 (DE3)* |
| pQR2801 | *Escherichia coli BL21 (DE3)* |
| pQR2804 | *Escherichia coli BL21 (DE3)* |
| pQR2802 | *Escherichia coli BL21 (DE3)* |
| pQR2803 | *Escherichia coli BL21 (DE3)* |
| pQR2786 | Escherichia coli TOP10 |
| pQR2787 | Escherichia coli TOP10 |
| pQR2788 | Escherichia coli TOP10 |
| pQR2789 | Escherichia coli TOP10 |
| pQR2790 | Escherichia coli TOP10 |
| pQR2791 | Escherichia coli TOP10 |
| pQR2792 | Escherichia coli TOP10 |
| pQR2793 | Escherichia coli TOP10 |
| pQR2794 | Escherichia coli TOP10 |
| pQR2795 | Escherichia coli TOP10 |
| pQR2796 | Escherichia coli TOP10 |
| pQR2797 | Escherichia coli TOP10 |
| pQR2798 | Escherichia coli TOP10 |
| pQR2799 | Escherichia coli TOP10 |
| pQR2800 | Escherichia coli TOP10 |
| pQR2801 | Escherichia coli TOP10 |
| pQR2804 | Escherichia coli TOP10 |
| pQR2802 | Escherichia coli TOP10 |
| pQR2803 | Escherichia coli TOP10 |

# 9. Chapter 9. Bibliography


Andreozzi, S. *et al.* (2016) 'Identification of metabolic engineering targets for the enhancement of 1,4-butanediol production in recombinant E. coli using large-scale kinetic models', *Metabolic Engineering*, 35, pp. 148–159. doi: 10.1016/j.ymben.2016.01.009.

Atkinson, M. R. *et al.* (2003) 'Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli', *Cell*, 113(5), pp. 597–607. doi: 10.1016/S0092-8674(03)00346-5.

Averesch, N. J. H. and Krömer, J. O. (2018) 'Metabolic engineering of the shikimate pathway for production of aromatics and derived compounds-Present and future strain construction strategies', *Frontiers in Bioengineering and Biotechnology*, 6(32), pp. 1–19. doi: 10.3389/fbioe.2018.00032.

Barth, T. (2011) *Biaxially stretched polyamide film*, *Biaxial Stretching of Film*. Woodhead Publishing Limited. doi: 10.1533/9780857092953.1.125.

Berti, P. J. and Chindemi, P. (2009) 'Catalytic residues and an electrostatic sandwich that promote enolpyruvyl shikimate 3-phosphate synthase (AroA) catalysis', *Biochemistry*, 48(17), pp. 3699–3707. doi: 10.1021/bi802251s.

Bervoets, I. and Charlier, D. (2019) 'Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology', *FEMS microbiology reviews*, 43(3), pp. 304–339. doi: 10.1093/femsre/fuz001.

Bonner, W. M. and Bloch, K. (1972) 'Purification and properties of fatty acyl thioesterase I from Escherichia coli.', *Journal of Biological Chemistry*, 247(10), pp. 3123–3133.

Brockman, I. M. and Prather, K. L. J. (2015) 'Dynamic metabolic engineering: New strategies for developing responsive cell factories', *Biotechnology Journal*, 10(9), pp. 1360–1369. doi: 10.1002/biot.201400422.

Cameron, D. E., Bashor, C. J. and Collins, J. J. (2014) 'A brief history of synthetic biology.', *Nature reviews. Microbiology*, 12(5), pp. 381–90. doi: 10.1038/nrmicro3239.

Cantu, D. C., Chen, Y. and Reilly, P. J. (2010) 'Thioesterases: A new perspective based on their primary and tertiary structures', *Protein Science*, 19(7), pp. 1281–1295. doi: 10.1002/pro.417.

Carothers, W. H. (1938) 'Synthetic fiber'. United states of america: United states patent office.

Carraher, C. E. (1978) 'Synthesis of caprolactam and Nylon 6', *Journal of Chemical Education*. American Chemical Society, 55(1), p. 51. doi: 10.1021/ed055p51.

Carrillo, N. and Ceccarelli, E. A. (2003) 'Open questions in ferredoxin-NADP+reductase catalytic mechanism', *European Journal of Biochemistry*, 270(9), pp. 1900–1915. doi: 10.1046/j.1432-1033.2003.03566.x.

Chen, S. *et al.* (1999) 'Biosynthesis of ansatrienin (mycotrienin) and naphthomycin: Identification and analysis of two separate biosynthetic gene clusters in Streptomyces collinus Tu 1892', *European Journal of Biochemistry*, 261(1), pp. 98–107. doi: 10.1046/j.1432-1327.1999.00244.x.

Chu, C. (2013) 'Materials for absorbable and nonabsorbable surgical sutures', in King, M. W., Gupta, B. S., and Guidoin, R. (eds) *Biotextiles as Medical Implants*. Cambridge: Woodhead Publishing (Woodhead Publishing Series in Textiles), pp. 275–334. doi: https://doi.org/10.1533/9780857095602.2.275.

Clayden, J., Greeves, N. and Warren, S. G. (2012) 'Alkylation of enolates', in Clayden, J., Greeves, N., and Warren, S. G. (eds) *Organic Chemistry*. 2nd editio. New York: Oxford University Press, pp. 596–597. doi: 10.1007/s00897010513a.

Collins, J. J., Gardner, T. S. and Cantor, C. R. (2000) 'Construction of a genetic toggle switch in Escherichia coli', *Nature*, 403(6767), pp. 339–342. doi: 10.1038/35002131.

Connelly, N. G. and Geiger, W. E. (1996) 'Chemical redox agents for organometallic chemistry', *Chemical Reviews*, 96(2), pp. 877–910. doi: 10.1021/cr940053x.

Cropp, T. a, Wilson, D. J. and Reynolds, K. a (2000) 'Identification of a cyclohexylcarbonyl CoA biosynthetic gene cluster and application in the production of doramectin.', *Nature biotechnology*, 18(9), pp. 980–983. doi: 10.1038/79479.

Culler, S. (2016) 'A Bioengineering Platform to Industrialize Biotechnology', *Chemical Engineering Progress; SBE SUPPLEMENT: Synthetic Biology*, 112(9), pp. 42–51.

Curran, K. A. and Alper, H. S. (2012) 'Expanding the chemical palate of cells by combining systems biology and metabolic engineering', *Metabolic Engineering*, 14(4), pp. 289–297. doi: 10.1016/j.ymben.2012.04.006.

Desborough, M. J. R. and Keeling, D. M. (2017) 'The aspirin story – from willow to wonder drug', *British Journal of Haematology*, 177(5), pp. 674–683. doi: 10.1111/bjh.14520.

Deshayes, G. *et al.* (2008) 'Activation of the hydrolytic polymerization of ε-caprolactam by ester functions: Straightforward route to aliphatic polyesteramides', *Reactive and Functional Polymers*, 68(9), pp. 1392–1407. doi: 10.1016/j.reactfunctpolym.2008.06.020.

Dillon, S. C. and Bateman, A. (2004) 'The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases', *BMC Bioinf.*, 5, p. 109. doi: 10.1186/1471-2105-5-109.

Dros, A. B. *et al.* (2015) 'Hexamethylenediamine (HMDA) from fossil- vs. bio-based routes: an economic and life cycle assessment comparative study', *Green Chemistry*, 17(10), pp. 4760–4772. doi: 10.1039/c5gc01549a.

Van Duuren, J. B. J. H. *et al.* (2011) 'A limited LCA of bio-adipic acid:

Manufacturing the nylon-6,6 precursor adipic acid using the benzoic acid degradation pathway from different feedstocks', *Biotechnology and Bioengineering*, 108(6), pp. 1298–1306. doi: 10.1002/bit.23074.

Egland, P. G. *et al.* (1997) 'A cluster of bacterial genes for anaerobic benzene ring biodegradation.', *Proceedings of the National Academy of Sciences*, 94(12), pp. 6484–6489. doi: 10.1073/pnas.94.12.6484.

Eklöf, J. M. *et al.* (2009) 'The crystal structure of the outer membrane lipoprotein YbhC from Escherichia coli sheds new light on the phylogeny of carbohydrate esterase family 8', *Proteins: Structure, Function and Bioinformatics*, 76(4), pp. 1029–1036. doi: 10.1002/prot.22453.

Ellman, G. L. (1959) 'Tissue sulfhydryl groups', *Archives of Biochemistry and Biophysics*, 82(1), pp. 70–77. doi: 10.1016/0003-9861(59)90090-6.

EPA (2014) *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2014*. Washington, DC. Available at: http://www3.epa.gov/climatechange/emissions/usinventoryreport.html.

EPA (2018) *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2016*. Washington, DC. Available at: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks.

Fang, X. *et al.* (2002) 'Ring-opening polymerization of ε-caprolactam and ε-caprolactone via microwave irradiation', *Journal of Polymer Science, Part A: Polymer Chemistry*, 40(14), pp. 2264–2275. doi: 10.1002/pola.10306.

Flynn, J. H. (2002) 'Chapter 14 - Polymer degradation', in Cheng, S. Z. D. (ed.) *Applications to Polymers and Plastics*. Amsterdam: Elsevier Science B.V., pp. 587–651. doi: https://doi.org/10.1016/S1573-4374(02)80017-6.

Frost, J. (2005) 'Synthesis of caprolactam from lysine', *WO Patent 2005/123669*. Available at: https://www.patentimages.storage.googleapis.com/4d/0e/81/2fda73c8aca775/US7399855.pdf.

Garcia Villaluenga, J. P. and Tabe-Mohammadi, A. (2000) 'A review on the separation of benzene/cyclohexane mixtures by pervaporation processes', *Journal of Membrane Science*, 169(2), pp. 159–174. doi: 10.1016/S0376-7388(99)00337-3.

Ghosh, S., Chisti, Y. and Banerjee, U. C. (2012) 'Production of shikimic acid', *Biotechnology Advances*, 30(6), pp. 1425–1431. doi: 10.1016/j.biotechadv.2012.03.001.

Godbey, W. T. and Godbey, W. T. (2014) 'Genes: The Blueprints for Proteins', in *An Introduction to Biotechnology*. Cambridge: Woodhead Publishing, pp. 65–105. doi: 10.1016/B978-1-907568-28-2.00004-6.

Gonzalez, C. F. *et al.* (2012) 'Structure and activity of the Pseudomonas aeruginosa hotdog-fold thioesterases PA5202 and PA2801', *Biochemical Journal*, 444(3), pp. 445–455. doi: 10.1042/BJ20112032.

Grand view research (2019) *Nylon 6 & 66 Market Size, Share & Trends*

*Analysis Report By Product (Nylon 6, Nylon 66), By Application (Engineering Plastics, Automotive, Textiles, Electrical & Electronics, Others), And Segment Forecasts, 2019 - 2025*. San Francisco. Available at: https://www.grandviewresearch.com/industry-analysis/nylon-6-6-market.

Hale, V. *et al.* (2007) 'Microbially derived artemisinin: A biotechnology solution to the global problem of access to affordable antimalarial drugs', *American Journal of Tropical Medicine and Hygiene*, 77(SUPPL. 6), pp. 198–202. Available at: https://www.ncbi.nlm.nih.gov/books/NBK1717/.

Harwood, C. S. *et al.* (1998) 'Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway', *FEMS Microbiology Reviews*, 22(5), pp. 439–458. doi: 10.1111/j.1574-6976.1998.tb00380.x.

Hirakawa, H. *et al.* (2015) 'BadR and BadM proteins transcriptionally regulate two operons needed for anaerobic benzoate degradation by Rhodopseudomonas palustris', *Applied and Environmental Microbiology*, 81(13), pp. 4253–4262. doi: 10.1128/AEM.00377-15.

Hoffmann, A., Bukau, B. and Kramer, G. (2010) 'Structure and function of the molecular chaperone Trigger Factor', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1803(6), pp. 650–661. doi: 10.1016/j.bbamcr.2010.01.017.

Howard, T. P. *et al.* (2013) 'Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in Escherichia coli.', *Proceedings of the National Academy of Sciences of the United States of America*, 110(19), pp. 7636–7641. doi: 10.1073/pnas.1215966110.

Hwang, K. C. and Sagadevan, A. (2014) 'One-pot room-temperature conversion of cyclohexane to adipic acid by ozone and UV light.', *Science*, 346(6216), pp. 1495–1498. doi: 10.1126/science.1259684.

International Energy Agency (2018) *CO2 emissions from fuel combustion 2018 Highlights*. Available at: https://webstore.iea.org/co2-emissions-from-fuel-combustion-2018-highlights.

Jackson, D. (2017) 'Building a Synthetic Pathway For Nylon precursor Biosynthesis'. PhD Thesis, University College London, London.

Jullesson, D. *et al.* (2015) 'Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals', *Biotechnology Advances*, 33(7), pp. 1395–1402. doi: 10.1016/j.biotechadv.2015.02.011.

Kameda, A. *et al.* (2001) 'A novel ATP regeneration system using polyphosphate-AMP phosphotransferase and polyphosphate kinase', *Journal of Bioscience and Bioengineering*. Elsevier, 91(6), pp. 557–563. doi: 10.1016/S1389-1723(01)80173-0.

Keifer, D. H. (2000) *The Establishment of Modern Polymer Science By Wallace H. Carothers*, *ACS Division of the History of Chemistry and The Office of Communications*. Edited by D. H. Keifer. Wilmington, Delaware: American Chemical Society. Available at: http://www.acs.org/content/acs/en/education/whatischemistry/landmarks/carotherspolymers.html.

Kogure, T. *et al.* (2016) ‘Metabolic engineering of Corynebacterium glutamicum for shikimate overproduction by growth-arrested cell reaction’, *Metabolic Engineering*, 38, pp. 204–216. doi: 10.1016/j.ymben.2016.08.005.

Korman, T. P. *et al.* (2010) ‘Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis’, *Proceedings of the National Academy of Sciences*, 107(14), pp. 6246–6251. doi: 10.1073/pnas.0913531107.

Kumar, V., Bhalla, A. and Rathore, A. S. (2014) ‘Design of experiments applications in bioprocessing: Concepts and approach’, *Biotechnology Progress*, 30(1), pp. 86–99. doi: 10.1002/btpr.1821.

Kung, J. W. *et al.* (2013) ‘Cyclohexanecarboxyl-coenzyme a (CoA) and cyclohex-1-ene-1-carboxyl-CoA dehydrogenases, two enzymes involved in the fermentation of benzoate and crotonate in syntrophus aciditrophicus’, *Journal of Bacteriology*, 195(14), pp. 3193–3200. doi: 10.1128/JB.00322-13.

Kung, S. H. *et al.* (2018) ‘Approaches and recent developments for the commercial production of semi-synthetic artemisinin’, *Frontiers in Plant Science*, 9(87), pp. 1–7. doi: 10.3389/fpls.2018.00087.

Kuver, J., Xu, Y. and Gibson, J. (1995) ‘Metabolism of cyclohexane carboxylic acid by the photosynthetic bacterium Rhodopseudomonas palustris’, *Archives of Microbiology*, 164(5), pp. 337–345. doi: 10.1007/bf02529980.

Lapkin, A. A. *et al.* (2010) ‘Screening of new solvents for artemisinin extraction process using ab initio methodology’, *Green Chemistry*, 12(2). doi: 10.1039/b922001a.

Latham, J. A. *et al.* (2014) ‘Divergence of substrate specificity and function in the Escherichia coli Hotdog-fold Thioesterase Paralogs YdiI and YbdB’, *Biochemistry*, 53(29), pp. 4775–4787. doi: 10.1021/bi500333m.

Laursen, B. S. *et al.* (2005) ‘Initiation of Protein Synthesis in Bacteria’, *Microbiology and Molecular Biology Reviews*, 69(1), pp. 101–123. doi: 10.1128/mmbr.69.1.101-123.2005.

Lechner, A., Brunk, E. and Keasling, J. D. (2016) ‘The need for integrated approaches in metabolic engineering’, *Cold Spring Harbor Perspectives in Biology*, 8(11). doi: 10.1101/cshperspect.a023903.

Lenfant, N. *et al.* (2013) ‘ESTHER, the database of the α/β-hydrolase fold superfamily of proteins: tools to explore diversity of functions’, *Nucleic Acids Research*. Oxford University Press, 41(D1), pp. D423–D429. doi: 10.1093/nar/gks1154.

Li, J. *et al.* (2000) ‘Crystal structure of the Escherichia coli thioesterase II, a homolog of the human Nef binding enzyme.’, *Nature structural biology*, 7(7), pp. 555–559. doi: 10.1038/76776.

Li, L. *et al.* (2014) ‘Reducing nitrous oxide emissions to mitigate climate change and protect the ozone layer’, *Environmental Science and Technology*, 48(9), pp. 5290–5297. doi: 10.1021/es404728s.

Listwan, P. *et al.* (2010) 'The optimization of in vitro high-throughput chemical lysis of Escherichia coli. Application to ACP domain of the polyketide synthase ppsC from Mycobacterium tuberculosis', *Journal of Structural and Functional Genomics*, 11(1), pp. 41–49. doi: 10.1007/s10969-009-9077-8.

Lourenco, F. R. (2012) 'Simple Estimation of Uncertainty in the Quantification of Cefazolin by HPLC and Bioassay', *Journal of Chromatography & Separation Techniques*, 3(8). doi: 10.4172/2157-7064.1000153.

Lu, X. *et al.* (2019) 'Impact of bacterial chaperonin GroEL–GroES on bacteriorhodopsin folding and membrane integration', *Biophysics Reports*, 5(3), pp. 133–144. doi: 10.1007/s41048-019-0090-6.

Mann, S. and Chen, Y. P. P. (2010) 'Bacterial genomic G + C composition-eliciting environmental adaptation', *Genomics*, 95(1), pp. 7–15. doi: 10.1016/j.ygeno.2009.09.002.

Martin, J. (1998) 'Protein folding assisted by the GroEL/GroES chaperonin system.', *Biochemistry (Mosc)*, 53(4), pp. 374–381.

McKeen, L. W. (2014) '6 - Polyamides (Nylons)', in McKeen, L. W. (ed.) *The Effect of Temperature and other Factors on Plastics and Elastomers*. Third edit. William Andrew Publishing, pp. 233–340. doi: 10.1016/B978-0-323-31016-1.00006-4.

McMahon, M. D. and Prathera, K. L. J. (2014) 'Functional screening and in vitro analysis reveal thioesterases with enhanced substrate specificity profiles that improve short-chain fatty acid production in Escherichia coli', *Applied and Environmental Microbiology*, 80(3), pp. 1042–1050. doi: 10.1128/AEM.03303-13.

McNeil, M. B. *et al.* (2014) 'The succinate dehydrogenase assembly factor, SdhE, is required for the flavinylation and activation of fumarate reductase in bacteria', *FEBS Letters*, 588(3), pp. 414–421. doi: 10.1016/j.febslet.2013.12.019.

McNeil, M. B. and Fineran, P. C. (2013) 'The conserved RGxxE motif of the bacterial fad assembly factor sdhe is required for succinate dehydrogenase flavinylation and activity', *Biochemistry*, 52(43), pp. 7628–7640. doi: 10.1021/bi401006a.

Moore, B. S. *et al.* (1993) 'Biosynthetic Studies on Ansatrienin A. Formation of the Cyclohexanecarboxylic Acid Moiety', *Journal of the American Chemical Society*, 115, pp. 5254–5266. doi: 10.1021/ja00065a042.

Morgan-Kiss, R. M. and Cronan, J. E. (2004) 'The Escherichia coli fadK (ydiD) gene encodes an anerobically regulated short chain acyl-CoA synthetase', *Journal of Biological Chemistry*, 279(36), pp. 37324–37333. doi: 10.1074/jbc.M405233200.

Nguyen, N. *et al.* (2010) 'Design and analysis of a robust genetic Muller C-element.', *Journal of theoretical biology*. Elsevier, 264(2), pp. 174–87. doi: 10.1016/j.jtbi.2009.10.026.

Nicolaou, K. C. (2018) 'The Emergence and Evolution of Organic Synthesis and Why It is Important to Sustain It as an Advancing Art and Science for Its Own Sake', *Israel Journal of Chemistry*, 58(1–2), pp. 104–113. doi: 10.1002/ijch.201700121.

Paddon, C. J. and Keasling, J. D. (2014) 'Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development', *Nat Rev Micro*, 12(5), pp. 355–367. doi: 10.1038/nrmicro3240.

Padgette, S. R. *et al.* (1991) 'Site-directed mutagenesis of a conserved region of the 5-enolpyruvylshikimate-3-phosphate synthase active site', *Journal of Biological Chemistry*, 266, pp. 22364–22369.

Pasini, M. *et al.* (2016) 'Using promoter libraries to reduce metabolic burden due to plasmid-encoded proteins in recombinant Escherichia coli', *New Biotechnology*, 33(1), pp. 78–90. doi: 10.1016/j.nbt.2015.08.003.

Patton, S. M., Cropp, T. A. and Reynolds, K. A. (2000) 'A novel delta(3),delta(2)-enoyl-CoA isomerase involved in the biosynthesis of the cyclohexanecarboxylic acid-derived moiety of the polyketide ansatrienin A', *Biochemistry*, 39(25), pp. 7595–7604. doi: 10.1021/bi0005714.

Pelletier, D. A. and Harwood, C. S. (2000) '2-hydroxycyclohexanecarboxyl coenzyme a dehydrogenase, an enzyme characteristic of the anaerobic benzoate degradation pathway used by Rhodopseudomonas palustris', *Journal of Bacteriology*, 182(10), pp. 2753–2760. doi: 10.1128/JB.182.10.2753-2760.2000.

Peñate, I. Q. *et al.* (2012) 'Clean synthesis of adipic acid from cyclohexene in microemulsions with stearyl dimethyl benzyl ammonium chloride as surfactant: From the laboratory to bench scale', *Chemical Engineering Journal*, 200–202, pp. 357–364. doi: 10.1016/j.cej.2012.06.041.

Perret, A. *et al.* (2011) 'A novel acyl-CoA beta-transaminase characterized from a metagenome', *PLoS ONE*, 6(8). doi: 10.1371/journal.pone.0022918.

Peter, D. M. *et al.* (2016) 'A chemo-enzymatic road map to the synthesis of CoA esters', *Molecules*, 21(4), p. 517. doi: 10.3390/molecules21040517.

Pidugu, L. S. *et al.* (2009) 'Analysis of proteins with the "hot dog" fold: Prediction of function and identification of catalytic residues of hypothetical proteins', *BMC Structural Biology*, 9(37). doi: 10.1186/1472-6807-9-37.

Pieracci, J. P. *et al.* (2018) 'Chapter 9- Industry Review of Cell Separation and Product Harvesting Methods', in *Biopharmaceutical Processing: Development, Design, and Implementation of Manufacturing Processes*. Elsevier, pp. 165–206. doi: 10.1016/B978-0-08-100623-8.00009-8.

Polen, T., Spelberg, M. and Bott, M. (2013) 'Toward biotechnological production of adipic acid and precursors from biorenewables', *Journal of Biotechnology*, 167(2), pp. 75–84. doi: 10.1016/j.jbiotec.2012.07.008.

Qu, X., Lei, C. and Liu, W. (2011) 'Transcriptome mining of active biosynthetic pathways and their associated products in Streptomyces flaveolus.', *Angewandte Chemie (International ed. in English)*, 50(41), pp.

9651–4. doi: 10.1002/anie.201103085.

Quan, J. and Tian, J. (2011) ‘Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries’, *Nature Protocols*, 6(2), pp. 242–251. doi: 10.1007/978-1-62703-764-8_8.

Raemakers-Franken, P. C. *et al.* (2005) ‘Biochemical synthesis of 6-amino caproic acid’. European Patent Office Patent no. EP2194139A1. Available at: https://patentimages.storage.googleapis.com/f7/de/fa/fa942536355e4d/EP2194139A1.pdf.

Rase, H. (2000) ‘12.2 Hydrogenation of aromatic rings’, in *Handbook of Commercial Catalysts*. Boca Raton, Florida: CRC Press LLC, pp. 122–125. doi: 10.1201/b21367-6.

Rather, L. J. *et al.* (2010) ‘Coenzyme A-dependent aerobic metabolism of benzoate via epoxide formation’, *Journal of Biological Chemistry*, 285(27), pp. 20615–20624. doi: 10.1074/jbc.M110.124156.

Ren, S. *et al.* (2009) ‘Clean synthesis of adipic acid catalyzed by complexes derived from heteropoly acid and glycine’, *Catalysis Communications*, 10(5), pp. 464–467. doi: 10.1016/j.catcom.2008.10.013.

Rogers, J. K. *et al.* (2015) ‘Synthetic biosensors for precise gene control and real-time monitoring of metabolites’, *Nucleic Acids Research*, 43(15), pp. 7648–7660. doi: 10.1093/nar/gkv616.

Saba, J. *et al.* (2019) ‘The elemental mechanism of transcriptional pausing’, *eLife*, 8. doi: 10.7554/eLife.40981.

Dos Santos, A. M. *et al.* (2017) ‘Unraveling the Addition-Elimination Mechanism of EPSP Synthase through Computer Modeling’, *Journal of Physical Chemistry B*, 121(37), pp. 8626–8637. doi: 10.1021/acs.jpcb.7b05063.

Shiba, T. *et al.* (2005) ‘Polyphosphate:AMP phosphotransferase as a polyphosphate-dependent nucleoside monophosphate kinase in Acinetobacter johnsonii 210A’, *Journal of Bacteriology*, 187(5), pp. 1859–1865. doi: 10.1128/JB.187.5.1859-1865.2005.

Shimizu, B. I. (2014) ‘2-Oxoglutarate-dependent dioxygenases in the biosynthesis of simple coumarins’, *Frontiers in Plant Science*, 5(549). doi: 10.3389/fpls.2014.00549.

Shimizu, K. (2013) ‘Metabolic Regulation of a Bacterial Cell System with Emphasis on Escherichia coli Metabolism’, *ISRN Biochemistry*, 2013. doi: 10.1155/2013/645983.

Shuttleworth, W. A. *et al.* (1999) ‘Site-directed mutagenesis of putative active site residues of 5- enolpyruvylshikimate-3-phosphate synthase’, *Biochemistry*, 38(1), pp. 296–302. doi: 10.1021/bi9815142.

Singh, R. (2011) ‘Facts , Growth , and Opportunities in Industrial Biotechnology Abstract ’:, 15(1), pp. 175–179. doi: 10.1021/op100312a.

Singhal, K. *et al.* (2015) ‘The Trigger Factor Chaperone Encapsulates and

Stabilizes Partial Folds of Substrate Proteins', *PLoS Computational Biology*, 11(10). doi: 10.1371/journal.pcbi.1004444.

Smith, J. K. and Hounshell, D. a (1985) 'Wallace h. Carothers and fundamental research at du pont.', *Science (New York, N.Y.)*, 229(4712), pp. 436–42. doi: 10.1126/science.229.4712.436.

Sneader, W. (2000) 'The discovery of aspirin: A reappraisal', *British Medical Journal*, 321(7276), pp. 1591–1594. doi: 10.1136/bmj.321.7276.1591.

Sofeo, N. *et al.* (2019) 'Altering the Substrate Specificity of Acetyl-CoA Synthetase by Rational Mutagenesis of the Carboxylate Binding Pocket', *ACS Synthetic Biology*, 8(6), pp. 1325–1336. doi: 10.1021/acssynbio.9b00008.

Song, F. *et al.* (2012) 'The catalytic mechanism of the hotdog-fold enzyme superfamily 4-hydroxybenzoyl-coa thioesterase from arthrobacter sp. Strain SU', *Biochemistry*, 51(35), pp. 7000–7016. doi: 10.1021/bi301059m.

Song, Y. N. *et al.* (2015) 'New ansamycin derivatives generated by simultaneous mutasynthesis', *Organic Letters*, 17(3), pp. 556–559. doi: 10.1021/ol5035639.

Starbird, C. A. *et al.* (2015) 'Flavoenzymes: Covalent versus Noncovalent', *eLS*. (Major Reference Works), pp. 1–11. doi: doi:10.1002/9780470015902.a0026073.

Starbird, C. A. *et al.* (2017) 'Structural and biochemical analyses reveal insights into covalent flavinylation of the Escherichia coli Complex II homolog quinol:fumarate reductase', *Journal of Biological Chemistry*, 292(31), pp. 12921–12933. doi: 10.1074/jbc.M117.795120.

Strauss, E. (2010) '7.11 - Coenzyme A Biosynthesis and Enzymology', in Liu, H.-W. (Ben) and Mander, L. (eds) *Comprehensive Natural Products II*. Oxford: Elsevier, pp. 351–410. doi: https://doi.org/10.1016/B978-008045382-8.00141-6.

Sugita, M. *et al.* (1982) 'Studies on mycotrienin antibiotics, a novel class of ansamycins: I. taxonomy, fermentation, isolation and properties of mycotrienins I and II', *The Journal of Antibiotics*, 35(11), pp. 1460–1466. doi: 10.7164/antibiotics.35.1460.

Sutherland, C. A., Nicolau, D. P. and Kuti, J. L. (2011) 'Development of an HPLC Method for the Determination of Anidulafungin in Human Plasma and Saline', *Journal of Chromatographic Science*, 49(5), pp. 397–400. doi: 10.1093/chromsci/49.5.397.

Teufel, R. *et al.* (2010) 'Bacterial phenylalanine and phenylacetate catabolic pathway revealed', *Proceedings of the National Academy of Sciences of the United States of America*, 107(32), pp. 14390–14395. doi: 10.1073/pnas.1005399107.

The Intergovernmental Panel on Climate Change (IPCC). (2014) *Climate Change 2014: Mitigation of Climate Change*, *Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate*

*Change*. doi: 10.1017/CBO9781107415416.

Turk, S. C. H. J. *et al.* (2016) 'Metabolic Engineering toward Sustainable Production of Nylon-6', *ACS Synthetic Biology*, 5(1), pp. 65–73. doi: 10.1021/acssynbio.5b00129.

Ueda, K. *et al.* (1997) 'Synthesis of High Molecular Weight Nylon 6 by Anionic Polymerization of ε-Caprolactam. Mechanism and Kinetics', *Polym J*, 29(7), pp. 568–573. doi: 10.1295/polymj.29.568.

Verma, D. K. and Des Tombe, K. (2002) 'Benzene in gasoline and crude oil: Occupational and environmental implications', *American Industrial Hygiene Association Journal*, 63(2), pp. 225–230. doi: 10.1080/15428110208984708.

Van De Vyver, S. and Román-Leshkov, Y. (2013) 'Emerging catalytic processes for the production of adipic acid', *Catalysis Science and Technology*, 3(6), pp. 1465–1479. doi: 10.1039/c3cy20728e.

Wallace, K. K. *et al.* (1995) 'Purification of Crotonyl- CoA Reductase from Streptomyces collinus and Cloning, Sequencing and Expression of the Corresponding Gene in Escherichia coli', *European Journal of Biochemistry*, 233(3), pp. 954–962. doi: 10.1111/j.1432-1033.1995.954_3.x.

Wang, P. *et al.* (1996) 'Cloning and characterization of the gene encoding 1-cyclohexenylcarbonyl coenzyme A reductase from Streptomyces collinus', *J Bacteriol.*, 178(23), pp. 6873–6881. doi: 10.1128/jb.178.23.6873-6881.1996.

Webb, M. E. and Smith, A. G. (2011) 'Pantothenate Biosynthesis in Higher Plants', in Fabrice Rébeillé, R. D. (ed.) *Advances in Botanical Research*. Cambridge: Academic Press, pp. 203–255. doi: 10.1016/B978-0-12-386479-6.00001-9.

Weckbecker, A. and Hummel, W. (2005) 'Glucose Dehydrogenase for the Regeneration of NADPH and NADH', *Microbial Enzymes and Biotransformations*. Edited by J. L. Barredo. Totowa, NJ: Humana Press, 17, pp. 225–238. doi: 10.1385/1-59259-846-3:225.

Wei, J., Kang, H. W. and Cohen, D. E. (2009) 'Thioesterase superfamily member 2 (Them2)/acyl-CoA thioesterase 13 (Acot13): a homotetrameric hotdog fold thioesterase with selectivity for long-chain fatty acyl-CoAs', *Biochemical Journal*, 421(2), pp. 311–322. doi: 10.1042/BJ20090039.

Welham, R. D. (1963) 'The Early History of the Synthetic Dye Industry*', *Journal of the Society of Dyers and Colourists*, 79(6), pp. 229–237. doi: 10.1111/j.1478-4408.1963.tb02549.x.

Wu, R. *et al.* (2014) 'Structure and Catalysis in the Escherichia coli Hotdog-fold Thioesterase Paralogs YdiI and YbdB', *Biochemistry*. American Chemical Society, 53(29), pp. 4788–4805. doi: 10.1021/bi500334v.

Yilmaz, B., Asci, A. and Erdem, A. F. (2014) 'HPLC method for naproxen determination in human plasma and its application to a pharmacokinetic study in Turkey', *Journal of Chromatographic Science*, 52(7), pp. 584–589. doi: 10.1093/chromsci/bmt080.

Yim, H. *et al.* (2011) 'Metabolic engineering of Escherichia coli for direct

production of 1,4-butanediol', *Nature Chemical Biology*, 22(7), pp. 445–452. doi: 10.1038/nchembio.580.

Yuzawa, S. *et al.* (2018) 'Chapter Fourteen - Commodity Chemicals From Engineered Modular Type I Polyketide Synthases', in Scrutton, N. (ed.) *Enzymes in Synthetic Biology*. Academic Press (Methods in Enzymology), pp. 393–415. doi: https://doi.org/10.1016/bs.mie.2018.04.027.

Zaar, A. *et al.* (2004) 'New enzymes involved in aerobic benzoate metabolism in Azoarcus evansii', *Molecular Microbiology*, 54(1), pp. 223–238. doi: 10.1111/j.1365-2958.2004.04263.x.

Zhang, J. and Deutscher, M. P. (1992) 'A uridine-rich sequence required for translation of prokaryotic mRNA', *Proceedings of the National Academy of Sciences of the United States of America*, 89(7), pp. 2605–2609. doi: 10.1073/pnas.89.7.2605.

Zhang, Y. *et al.* (2018) 'Ultrasonic assisted extraction of artemisinin from: Artemisia Annua L. using monoether-based solvents', *Green Chemistry*, 20(3), pp. 713–723. doi: 10.1039/c7gc03191b.

Zhuang, Z. *et al.* (2003) 'Characterization of the 4-Hydroxybenzoyl-Coenzyme A Thioesterase from Arthrobacter sp. Strain SU', *Applied and Environmental Microbiology*. American Society for Microbiology, 69(5), pp. 2707–2711. doi: 10.1128/AEM.69.5.2707-2711.2003.

Zong, B. *et al.* (2017) 'Green Production Technology of the Monomer of Nylon-6: Caprolactam', *Engineering*, 3(3), pp. 379–384. doi: https://doi.org/10.1016/J.ENG.2017.03.003.