Enzymatic Cascades for the Synthesis of Chiral Amines

Lorna J. Hepworth

School of Chemistry

2017

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy (PhD) in the Faculty of Science and Engineering

Table of Contents

42,000 words

The University of Manchester

Lorna J. Hepworth

Doctor of Philosophy

Enzymatic Cascades for the Synthesis of Chiral Amines

September 2017

Abstract

Chiral amines are important chemical building blocks found in a vast array of biologically-active pharmaceutical ingredients, and a sustainable route to their synthesis is of wide appeal. The use of biocatalysis as an alternative to traditional chemical methods boasts many advantageous qualities, such as the ability to catalyse a myriad of highly selective transformations in aqueous environments, and increased atom economies to reduce waste. As such, the primary aim of this study was to develop multi-enzymatic routes towards the synthesis of chiral secondary amine scaffolds. It is becoming increasingly possible to combine several sequential biocatalytic reactions, via *de novo* enzymatic pathways, for the generation of complex value-added compounds from simple precursors. These cascades take advantage of generally compatible reaction conditions, and can be implemented in an *in vivo*, *in vitro* or hybrid *in vitro* / *in vivo* cascade design.

Through the adaptation of the design-build-test-analyse (DBTA) cycle, several multi-enzyme cascades were established for the preparation of chiral pyrrolidine and piperidine compounds, exploiting a range of carboxylic acid reductase (CAR), ω-transaminase (ω-TA), imine reductase (IRED), galactose oxidase (GOase), and alcohol dehydrogenase (ADH) enzyme homologs. In particular, a CAR-TA-IRED single whole cell cascade for the production of chiral secondary amines through the conversion of simple linear keto acids was achieved, with all cofactor requirements being met by the microbial host cell. Cascade optimisation was achieved through sequential parameter investigation, and control over the expression levels of recombinant cascade enzymes was accomplished through gene duplications. An alternate *in vitro* GOase-IRED cascade was developed for the successful synthesis of the valuable drug precursor 3-aminopiperidine.

Amide bond formation remains a key target for biocatalytic processes due to current routes requiring harsh reaction conditions, and as such the ATP-independent *N*-acyltransferase CapW was identified as a potential new addition to the ever-expanding biocatalytic toolbox. Soluble recombinant expression of active CapW in *E. coli* cells was achieved here for the first time, and was verified via a screen for hydrolytic activity against β-lactamase substrate Nitrocefin, paving the way for the production of increased amounts of this protein for future applications such as crystallisation studies.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on Presentation of Theses

Table of Figures

Acknowledgements

First and foremost, I'd like to thank my supervisor Sabine Flitsch for giving me the opportunity to work in her laboratory and pursue aspects of the research that interested me the most, and for always having an open door. I'd like to thank both Sabine and Nick Turner for many fruitful meetings, and for providing support throughout the course of my PhD studies.

Thanks go to Elaine O'Reilly, who supervised me during my MChem project and brought to my attention the iGEM competition. Being involved in the competition taught me a lot about synthetic biology and molecular biology, and so I'd also like to thank Eriko Takano and Rainer Breitling for giving me the opportunity to take part and helping us throughout the entire process. Of course it wouldn't have been so much fun without such a great team to work with, so thanks also go to iGEM Manchester 2013! Specifically Rob, one half of LA's Fiercest Power Couple.

Many people helped me find my footing during the first few months of my PhD (and beyond), and so I'd like to individually thank Jason, Peter and Anthony for making sure I always had someone to turn to if things went wrong (like they're so prone to doing). Thanks also to Sarah and Seb for last minute help when I needed it, and to Sasha for valiantly proofreading some of my thesis.

My four years in the MIB just wouldn't have been the same without them, so I'd like to say a massive thank you to both Shahed and Scott. I don't even know where to begin... so I won't! There are far too many people to name individually, but thanks to Kelly Girona's Hand Models, the Golden Bay past and present, and everyone who was up for going for a pint when something went wrong/right! Special thanks also go to Sinead O'Connor, for getting me through some very long days.

Thank you to my mum, who has always encouraged me to follow my interests and has always had faith in me. Thank you to Kaela for keeping me sane and being so understanding through possibly one of the most chaotic times of my life. I appreciate it all.

And finally I dedicate this thesis to my grandma. She wouldn't have understood a word of it but she'd have wanted a copy anyway.

1 Introduction

1.1 Biocatalysis

In response to a world reliant on dwindling natural resources and energy reserves, and in an age of global warming worsened by vast amounts of pollution, Anastas and Warner drafted a framework in 1998 intended to encourage scientists to strive to develop more resourceful and environmentally-benign chemical processes.[1] The 12 Principles of Green Chemistry are guidelines on how to make a process more efficient, less hazardous to humans and the environment, and more sustainable, and an excellent way to meet these criteria is through the employment of biocatalysts rather than traditional chemical approaches. With the advent of enhanced methods for the discovery and screening of previously uncharacterised enzymes, coupled with gene synthesis becoming faster and cheaper to exploit, the inventory of organic reactions that can be facilitated through the employment of biocatalysts is rapidly increasing.[2–10] The consequence of this is an ever-expanding toolbox of biocatalysts comprised of wildtype enzymes, evolved or engineered enzymes[11–17] and artificial enzymes[18–20] which are implicated in a vast array of chemical transformations. The discovery and characterisation of an increasing number of promiscuous and/or highly selective biocatalysts also lends itself to the design and construction of entirely *de novo* enzymatic pathways for the synthesis of a broad range of high value products.

1.1.1 Enzymatic Cascades

Often heralded as the beginnings of the next great industrial revolution,[21,22] the emergent field of synthetic biology has already taken substantial strides in tack-

ling some of the major challenges currently faced in sustainable energy,[23] medicinal[24] and agricultural research.[25–28] Through the engineering of native biosynthetic pathways, a variety of valuable natural products, such as vanillin[29] and artemisinin,[30] have been successfully obtained in considerable yields, however the full potential of synthetic biology has yet to be harnessed. The exploitation of synthetic biology methodologies to produce pharmaceutically- and industrially-relevant synthetic products is a relatively unexplored landscape for chemical synthesis, and the increasing number of exploitable enzymes in the biocatalytic toolbox presents an opportunity to design unnatural cascades reactions, executed through cell-free *in vitro* systems or built into amenable biological chassis, for the production of these important target compounds.[31–35] In the following section, different approaches for the implementation of *de novo* enzymatic cascade reactions will be discussed.

1.2 Cascade Approaches

Enzymatic cascades are generally implemented in either a (i) *in vitro*, (ii) *in vivo*, or (iii) hybrid *in vitro* / *in vivo* fashion, as depicted in **Figure 1**.[31] Many factors influence the selection between these three cascade approaches, including but not limited to any cofactor requirements of the necessary biocatalysts, the availability of each heterologous enzyme and/or their respective nucleotide sequences, the solubility of substrates and whether or not both substrate and target product can be taken up and released by a microbial cell, and the inherent metabolic stabilities of all substrates, intermediates, and products involved in the cascade process.

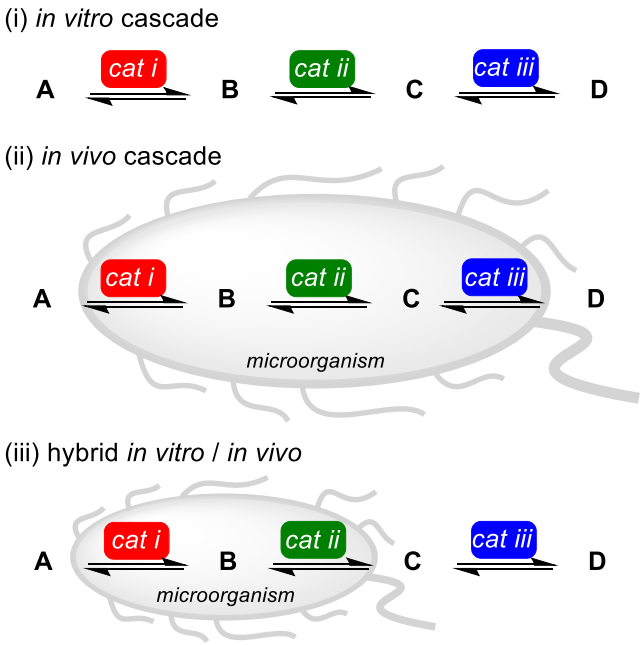


Figure 1 Comparison of (i) *in vitro*, (ii) *in vivo* and (iii) hybrid *in vitro* / *in vivo* cascade reactions. Substrate A is transformed into target product D through the application of highly selective biocatalysts cat i-iii.

1.2.1 *In Vitro* Cascades

The application of an enzymatic cascade in an entirely *in vitro* fashion requires the employment of biocatalysts as either purified enzymes, cell lysates, cell-free extracts or freeze-dried whole cells; i.e. the biocatalysts are not present within living whole cells. The use of an *in vitro* methodology offers select advantages over other modes of cascade implementation, such as the relative ease with which the amounts of each (bio)catalyst needed in the system can be refined to avoid any build-up of substrates and intermediates and therefore maximise cascade product output. Additionally, the use of isolated enzymes removes the possibility of reaction substrates, intermediates, and products being intercepted by other enzymes present within whole cells that could generate unwanted side

products. However, the purification of desired proteins is an expensive and laborious process that can pose significant drawbacks if the cascade reaction is to be applied on a larger scale or in an industrial setting. Crude enzyme preparations, on the other hand, are easier and less expensive to prepare but do not eliminate the potential side reactions catalysed by endogenous proteins of the original whole cell. Another advantage to the use of *in vitro* methodologies is the fact that purified and crude enzyme preparations can generally be more readily stored for longer periods of time prior to being used in biotransformations, whereas living whole cell formulations often require immediate application to maximise substrate turnover before the cells become unviable. One significant drawback to the *in vitro* use of certain biocatalysts arises if one or more enzyme in the system is dependent on a cofactor, such as the common nicotinamide-based cofactors $NAD(P)^+$/NAD(P)H, since these compounds must be provided exogenously alongside a suitable recycling system in the absence of a living whole cell. Additionally, purified and partially purified enzyme preparations may be more prone to denaturation compared to biocatalysts shielded within living or freeze-dried whole cells due to the closer proximity to potentially harsh reaction conditions (such as incompatible pH values or the presence of organic solvents).

1.2.2 *In Vivo* Cascades

The containment of an enzymatic cascade within one or more living host cell presents several advantages over the *in vitro* preparations detailed in Section 1.2.1, yet there have currently been far fewer *in vivo* cascade reactions reported in the scientific literature. Once a whole cell biocatalyst has been prepared it does not generally require further downstream processing (such as protein

purification or cell disruption) prior to use in a biotransformation, thus rendering *in vivo* biocatalysis more economical and straightforward than comparable *in vitro* strategies. Additionally, many cofactor requirements of the cascade enzymes can be successfully met when the proteins remain inside a host cell, since inherent metabolic pathways can provide and recycle most common cofactors (e.g. ATP, NAD(P)H). However, the production and containment of multiple biocatalysts within a single whole cell for use in *in vivo* biocatalysis is not without its drawbacks. Despite the presence of a cell membrane offering some protection and stabilisation of the overexpressed proteins, it can also present issues regarding the transportation of substrates and products into and out of the cell, and may lead to difficulties during the downstream extraction of the desired products.[36] What's more, if the substrate is able to be taken up by the cell, it then becomes at risk of being diverted into complex metabolic pathways within the living system, as do any reaction intermediates and products of the cascade, leading to unwanted side products. The expression of several heterologous biocatalysts in the same host cell often necessitates a compromise on protein expression and cell growth conditions, as enzymes from vastly different sources (e.g. eukaryotes, prokaryotes) are frequently used in conjunction with one another in the construction of non-natural enzymatic pathways. Furthermore, engineering a single whole cell to overproduce multiple recombinant proteins simultaneously may exert a significant metabolic burden on the cell and lead to a slower rate of cell growth, cell toxicity and lower production levels of each target enzyme.[37] Some of the issues raised above can be circumvented by the use of two or more separate whole cells in a single reaction, each containing an individual enzyme of interest.[38–40] The use of distinct host cells containing each cascade protein separately in a mixed-culture reaction permits a higher degree of control when expressing each specific enzyme, and results in lower levels of metabolic burden on each cell.

1.2.3 Hybrid Approaches

In the design and implementation of a *de novo* cascade process it may become apparent that each constituent enzyme of the pathway favours either an *in vitro* or *in vivo* approach, resulting in the recent development of cascade reactions using both cascade modes within the same system.[40–42] These hybrid *in vitro* / *in vivo* approaches harness the power of each individual methodology whilst minimising any of the aforementioned conflicts and drawbacks.

An example of a cascade reaction employing the use of a hybrid biocatalyst system was established by Park *et al.* through the use of a whole cell expressing three cascade enzymes, with a fourth enzyme added sequentially as an *E. coli* cell extract.[41] Renewable fatty acids were the substrate for a cascade composed of an oleate hydratase, an alcohol dehydrogenase (ADH, Section 1.6.5) and one of two Baeyer-Villiger monooxygenases (BVMOs) resulting in the formation of ester intermediates, followed by esterase-catalysed ester hydrolysis to yield α,ω-dicarboxylic acid and ω-hydroxycarboxylic acid products (**Figure 2**).

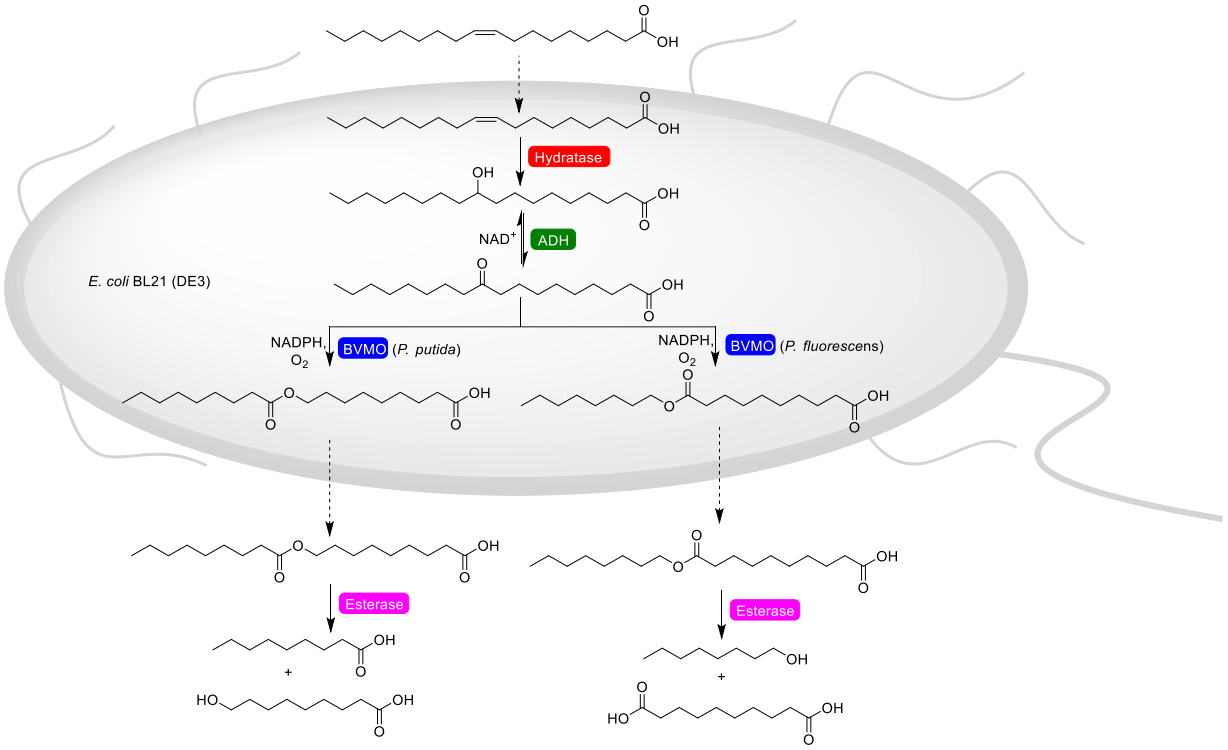


Figure 2 Hybrid *in vivo* / *in vitro* cascade for the conversion of renewable fatty acids to α,ω-dicarboxylic and ω-hydroxycarboxylic acids.

The esterase biocatalyst was added as a separate *in vitro* component to the reaction as the products formed through its reaction were seen to be toxic to living cells. For this reason, the initial hydratase-ADH-BVMO *in vivo* cascade was allowed to reach completion prior to the addition of esterase, and this optimised sequential approach afforded conversions of 60 % to ω-hydroxynonanoic acid product.

1.2.4 Process Optimisation of Enzymatic Cascade Reactions

Many parameters must be considered before a biocatalytic cascade reaction can be successfully implemented in an industrial process, including but not limited to

the optimisation of product concentrations accessible, (optical) purities of the product, the cost of process execution, the stability of the biocatalyst with respect to any necessary reaction conditions (such as organic solvents, elevated temperatures, suboptimal pH values), and the ability to recycle the biocatalyst.[31,43–45] It is possible to account for some of these parameters through the selection of a suitable *in vitro* or *in vivo* system (or a hybrid system composed of both *in vivo* and *in vitro* components), yet process engineering can also be effectively exploited for this purpose. High conversions to a desired product become more difficult to achieve with increasing numbers of steps constituting a cascade reaction, and it is not always possible to optimise each individual step due to the requirement of different enzymes for varying reaction conditions. For this reason it is often more efficient to optimise an entire reaction sequence as a whole, rather than attempting to optimise each reaction step separately. However, it sometimes becomes apparent that a cascade reaction must be implemented in a sequential, rather than a concurrent, fashion, which can be achieved through the temporal separation of cascade enzymes or reagents, for example.[41,46,47]

On the subject of process engineering, Rother *et al.* developed a cascade reaction suited for industrial application, paying particular attention to achieving increased substrate loadings through the use of micro-aqueous systems.[48] A two-enzyme reaction composed of a benzaldehyde lyase (BAL) and an ADH (Section 1.6.5), contained within two separate lyophilised whole cells, was designed for the production of vicinal diols from benzaldehyde and acetaldehyde (**Figure 3**). Many reaction parameters thought to affect the extent of conversion to the desired chiral products were investigated through a concurrent or sequential reaction mode, such as substrate concentrations, variation of buffer conditions and tolerance to organic solvents.

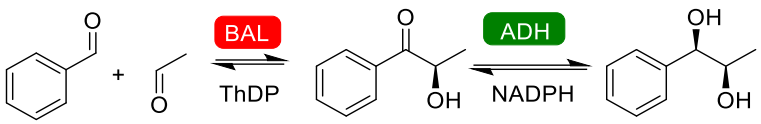


Figure 3 *In vitro* cascade for the conversion of benzaldehyde and acetaldehyde to chiral vicinal diol products.

From a comparison of concurrent and sequential reaction modes, it was indicated that the concurrent cascade led to a significantly higher optimal productivity (327 g L^{-1} day^{-1}), with the sequential reaction resulting in a productivity of just 222 g L^{-1} day^{-1}. The optimisation of this micro-aqueous cascade system afforded a space-time yield that was 1600-fold higher than that seen for purified enzymes in aqueous media, and is of a high enough efficiency to be relevant for industrial applications.

1.3 Expression of Multiple Enzymes in Same Cell

For the *in vivo* application of multi-enzyme cascade reactions, it is usually necessary to overexpress two or more recombinant proteins in the chosen host cell, unless a multi-strain approach has been selected.[38,39,49] In order to achieve the simultaneous expression of several biocatalysts, heterologous enzymes can either be introduced through the co-transformation of multiple separate expression plasmids, or the genes encoding all the cascade enzymes can be contained within a single construct. In the following section, different approaches to the production of multiple enzymes within the same cell will be discussed, followed by various methods that have been developed to control the expression ratios of target proteins for increased efficiencies in cascade reactions.

1.3.1 Methods of Simultaneous Expression

1.3.1.1 Expression Through the Use of Multiple Compatible Plasmids

One way to achieve the overexpression of two or more recombinant proteins in a host cell is through the co-transformation of several individual plasmids, each containing a gene encoding an individual biocatalyst. For the co-transformation of plasmids to be successful, however, the chosen plasmid backbones must be compatible with respect to their origin of replication (ori), and must confer distinct antibiotic resistances to allow for the selection of bacterial colonies containing all of the required vector constructs. This approach is therefore limited by the number of replicons available for exploitation,[32] and so methods for the incorporation of multiple genes on a single plasmid have been developed.

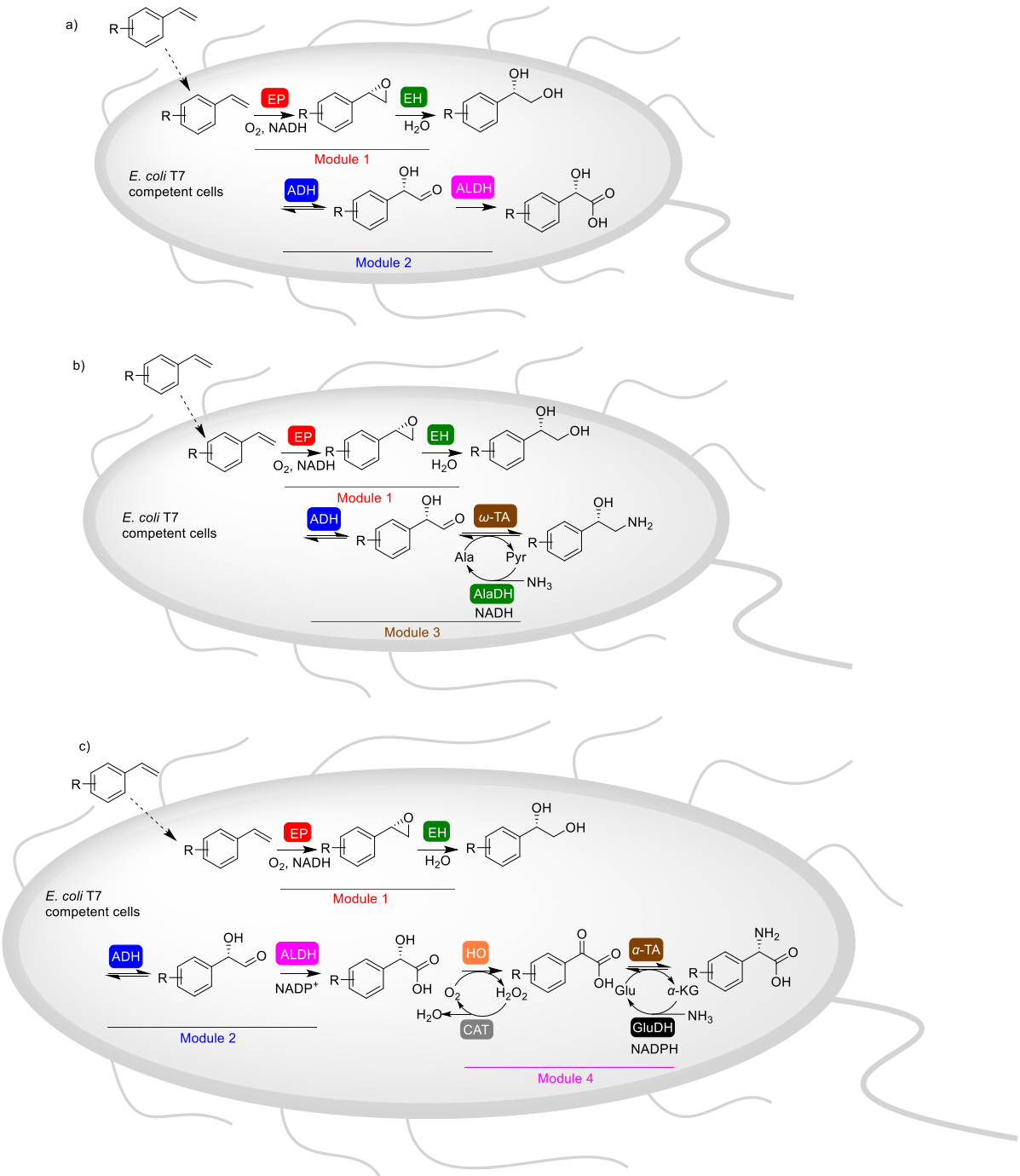
1.3.1.2 Multiple Genes Contained on One Plasmid

1.3.1.2.1 pETDuet-1 Expression System (Novagen)

The pETDuet-1 expression system is a set of vectors containing two multiple cloning sites, developed to enable the incorporation of two separate genes on the same plasmid backbone.[32,50] Each gene contained within a Duet plasmid is controlled by a separate T7 promoter, *lac* operator and ribosome binding site (RBS) through a pseudo-operon design set up, ensuring the adequate regulation of both genes of interest. pETDuet-1 vectors are available with five different origins of replication (ColE1, CloDF13, P15A, ColA and RSF1030) and four antibiotic resistance markers (conferring resistance to ampicillin, streptomycin, chloramphenicol and kanamycin), making them compatible with most widely-used single plasmid systems, as well as with each other. Through a combination

of four pETDuet-1 plasmids, up to eight recombinant proteins can be introduced into a suitable host organism, allowing for the construction of complex multi-enzyme pathways.[51–53] This number of recombinant proteins may be increased through the modular design of artificial operons, as was demonstrated by the Prather group for the synthesis of non-natural alcohols such as 4-methyl-1-pentanol.[52] The constructed *de novo* pathway consisted of ten steps and was completed by the conversion of 4-methylvalerate into 4-methylpentanol through a module comprised of a carboxylic acid reductase from *Nocardia iowensis* (NCAR, Section 1.6.1, Section 2.1 and Chapter 3) and an alcohol dehydrogenase (ADH, Section 1.6.5).

Another recent example of the exploitation of multiple pETDuet-1 vectors for the overproduction of several recombinant enzymes towards the construction of a *de novo* biocatalytic cascade was demonstrated by the Li group, in which three distinct whole cell biocatalysts were created for the conversion of terminal alkene substrates into the corresponding chiral α-hydroxyacids (**Figure 4**a), 1,2-amino alcohols (**Figure 4**b) and α-amino acids (**Figure 4**c).[53]



R = H, 2-F, 3-F, 4-F, 3-Cl, 4-Cl, 3-Br, 4-Br, 3-Me, 4-Me, 3-OMe

Figure 4 *In vivo* cascades contained on several pETDuet-1 vectors for the conversion of terminal alkene to (a.) α-hydroxyacids, (b.) 1,2-amino alcohols and (c.) α-amino acids using a single whole cell biocatalyst.

The whole cell biocatalysts were constructed by containing the required enzymes in four separate modules, each within an individual pETDuet-1 vector: Module 1

was responsible for the conversion of terminal alkene substrate to 1,2-diol product through the use of an epoxidase (EP) and an epoxide hydrolase (EH); Module 2 was responsible for the conversion of 1,2-diol to α-hydroxyacid through the use of an alcohol dehydrogenase (ADH, Section 1.6.5) and an aldehyde dehydrogenase (ALDH); Module 3 was responsible for the conversion of 1,2-diol to 1,2-amino alcohol through the use an ADH, an ω-transaminase (ω-TA, Section 1.6.2) and an alanine dehydrogenase (AlaDH, for the recycling of the L-alanine amine donor), and; Module 4 was responsible for the conversion of α-hydroxyacid to α-amino acid through the use of a hydroxyacid oxidase (HO), an α-transaminase (α-TA), catalase (CAT, for the removal of H_2O_2 by-product) and glutamate dehydrogenase (GluDH, for the recycling of the L-glutamate amine donor) enzymes. These four plasmids were then mixed and matched through the co-transformation of *E. coli* bacterial cells to provide several whole cell systems catalysing a variety of cascade reactions, with the modular nature of the system allowing for the optimisation of differing combinations of expression constructs to achieve a balanced profile of the recombinant enzymes. Employing these whole cell biocatalysts, (*S*)-α-hydroxyacids, (*S*)-amino alcohols, and (*S*)-α-amino acids could be obtained on a preparative scale, giving isolated yields of up to 72 %, 62 %, and 70 %, respectively (98 % ee).

However, certain drawbacks to the use of the pETDuet-1 expression system are evident.[32] It is not possible to modify the order at which the incorporated genes exist once the system has been constructed, and interchanging the genes of interest to assess various homologs of a chosen enzyme is not possible. Additionally, all introduced genes are under the control of an isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible T7 promoter, and so tuning the expression levels of individual cascade proteins cannot be achieved. What's more, the use of multiple plasmids for the assembly of cascade reactions will eventually be limited to the number of compatible vectors available to the

researcher, and so alternative strategies towards expressing more than one gene from the same construct are necessary to overcome this eventuality.

1.3.1.2.2 BioBrick™ Cloning Strategy

A standardised approach to the design and construction of genetic systems was detailed by Knight in 2003,[54] allowing for the rapid and versatile assembly of multi-gene plasmids for the co-expression of several biocatalysts within a single whole cell. The BioBrick™ cloning strategy makes use of modular BioBrick™ parts, nucleotide sequences coding for the expression of biological parts, such as regulatory elements [promoters, terminators and ribosome binding sites (RBS)] or functional enzymes for use in biocatalytic reactions. Each BioBrick™ part is standardised through the presence of a specific prefix (containing the *EcoRI* and *XbaI* restriction sites) and suffix (containing the *SpeI* and *PstI* restriction sites) (**Figure 5**), where endonuclease digestion with the *XbaI* and *SpeI* inner restriction sites leads to complementary overhangs that form a scar sequence upon religation (**Figure 6**).

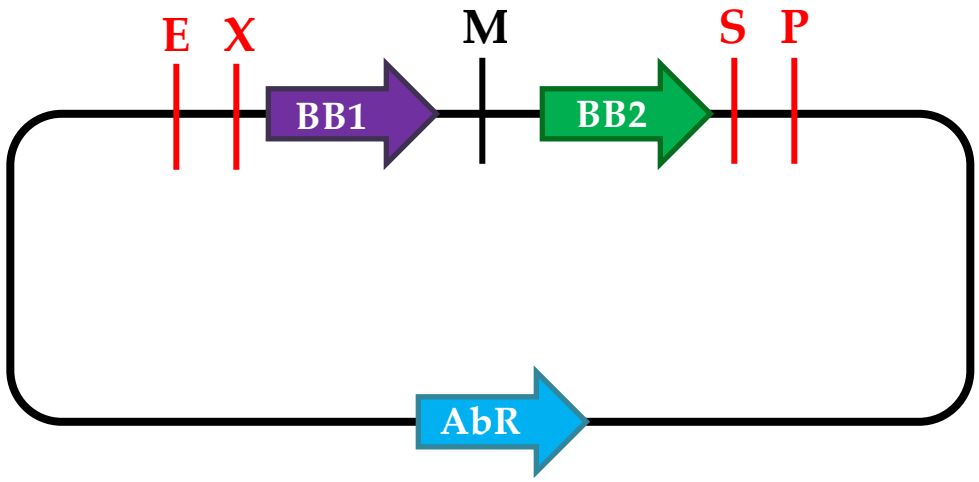


Figure 5 Schematic of a typical plasmid containing two BioBrick™ parts (BB1 and BB2), flanked by the standard BioBrick™ prefix, *EcoRI* (E) and *XbaI* (X), and suffix, *SpeI* (S) and *PstI* (P), with a mixed ligation site (M) from the prior ligation of the two parts (**Figure 6**) . AbR, antibiotic resistance marker.

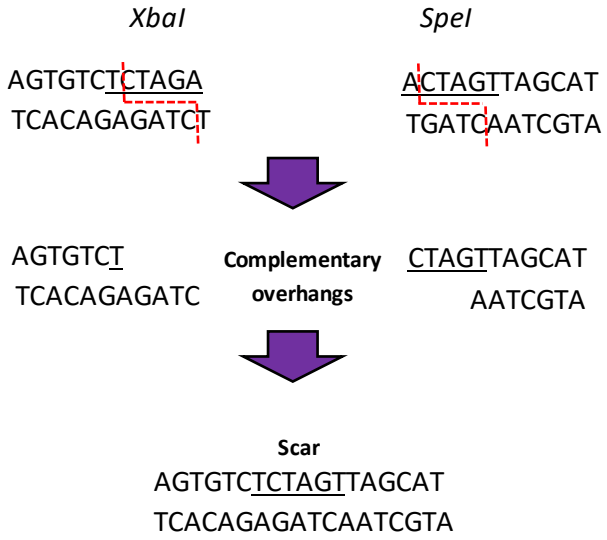


Figure 6 The isocaudomer pair of restriction sites *XbaI* and *SpeI*, contained on the standard BioBrick™ prefix and suffix, respectively. When cut with the appropriate restriction enzymes, complementary overhangs are created that, upon ligation, form a scar site that is no longer recognised by either enzyme.

The ligation of the complementary overhangs created through the digestion of *XbaI* and *SpeI* restriction sites leads to a new BioBrick™ part, enabling the

insertion of a new part either upstream or downstream of the previous part. Through the application of this cloning strategy, a library of compatible genetic parts can be mixed and matched in a "plug-and-play" approach to the construction of multi-gene plasmid systems, and several compatible vectors have been created for use in conjunction with BioBrick™ parts for the recombinant expression of proteins in a variety of organisms.[47,55–60]

For the successful application of the BioBrick™ cloning strategy, it is first necessary to ensure that the four restriction sites (*EcoRI*, *XbaI*, *SpeI*, *PstI*) contained on the standard prefix and suffix are not present within the gene(s) of interest. Once this is the case, iterative rounds of standard cloning can be carried out for the step-by-step construction of a multi-gene plasmid, with the BioBrick™ inserts needing to be purified prior to ligation. This stepwise approach, and the necessary purification of parts, can lead the process to become time-consuming, especially since molecular cloning is not always an efficient process.[32] However, this method of standard cloning has a relatively high success rate, as evidenced by the hundreds of undergraduate students effectively designing and building genetic systems through the exploitation of the BioBrick™ system for the International Genetically Engineered Machine (iGEM) synthetic biology competition that takes place every year.[61] The Registry of Standard Biological Parts is building year-on-year, making a myriad of interesting and useful genetic parts available for use in laboratories worldwide.

Once an appropriate construct has been assembled, containing all of the necessary cascade genes for a successful reaction to take place, it is important to ensure that the expression ratios for each biocatalyst are suitable for efficient conversion to the desired cascade reaction product.

1.3.2 Methods to Control the Expression Levels of Cascade Proteins

1.3.2.1 Optimisation of Genetic Regulatory Parts to Control Cascade Protein Expression Levels Within a Single Cell

It is well known that the use of varying strengths of genetic regulatory elements, such as promoters, terminators and RBSs, can lead to increases or decreases in the soluble expression levels of target recombinant proteins.[32,50,62] The use of strong inducible promoters, for example, is a generally accepted method to increase the dosage of a given protein, and the use of a tightly repressed promoter is beneficial for the expression of particularly toxic proteins. Varying the origin of replication (ori) in the plasmid system can also lend itself to the control of protein expression levels, since plasmids of high copy numbers generally translate to a greater amount of their respective protein(s), whilst lower copy numbers can result in significantly diminished amounts of their target protein(s).

Hauer *et al* employed the use of plasmids of varying copy numbers for the development of a three-enzyme cascade, contained within a single whole cell, for the production of azelaic acid from the substrate linoleic acid (**Figure 7**).[63] The cascade reaction began with the hydroperoxidation of linoleic acid, catalysed by a lipoxygenase (LOX), followed by the hydroperoxide lyase (HPL)-catalysed cleavage of the hydroperoxy group. A third enzyme, an oxidoreductase (ALDH) endogenous to *E. coli*, subsequently catalysed the oxidation of 9-oxononanoic acid to yield azelaic acid as product.

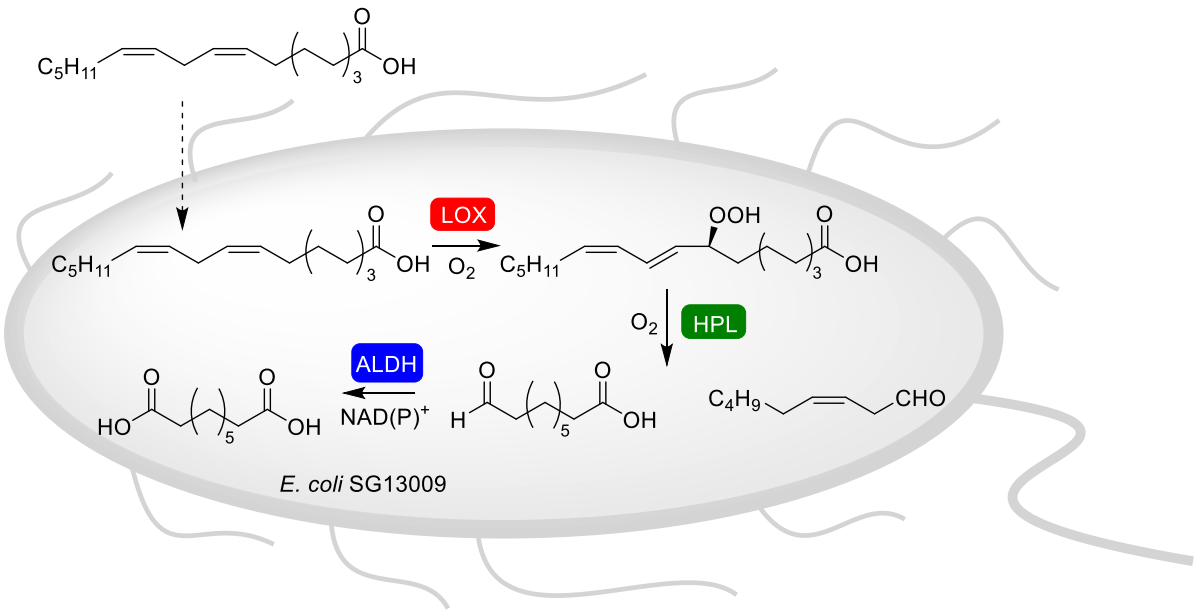


Figure 7 *In vivo* cascade contained on a high copy number plasmid (LOX) and a low copy number plasmid (HPL) for the conversion of linoleic acid to azelaic acid using a single whole cell biocatalyst.

Analysis of the reaction profile suggested that LOX required activation by its product to enable efficient catalysis, making it necessary to carefully balance the expression levels of both LOX and HPL recombinant proteins. This was achieved through the co-expression of a high copy number plasmid containing the LOX gene, and a low copy number plasmid containing the HPL gene, which led to lower soluble expression levels of HPL in comparison to LOX, resulting in adequate levels of the LOX-product being present in the reaction to effectively activate LOX. Further to this, the reaction parameters were optimised through the implementation of a two-phase system employing cyclohexane, which enabled conversions of 26 % to the product azelaic acid using 1 mM linoleic acid substrate.

Finely tuning the expression of a given protein through optimisation of these regulatory elements can, however, be rather time-consuming and laborious, and

so an alternative approach for the tailoring of recombinant cascade protein levels would be advantageous.

1.3.2.2 The Use of Recombinant Protein Scaffolds to Control the Ratios of Cascade Proteins Within a Single Cell

Various ways to control the extent of translation of a target recombinant protein through the optimisation of different genetic parts, such as promoters and origins of replication, have been described in Section 1.3.2.1. Another way in which it is possible to tailor the amounts of multiple enzymes in an enzymatic pathway is through the introduction of a recombinant protein scaffold into the chosen host cell. This technique exploits the bonding between protein-protein interaction domains, contained in a modular fashion within a synthetic protein scaffold, and their specific complementary ligands, which are added to the proteins of interest prior to expression.[64] Varying amounts of protein-protein interaction domains for ligands added to each of three recombinant proteins, namely acetoacetyl-CoA thiolase (AtoB), hydroxymethylglutaryl-CoA synthase (HMGS) and hydroxymethylglutaryl-CoA reductase (HMGR), were used to alter the amounts of each pathway protein in order to increase the production of the metabolite mevalonate from acetyl-CoA (**Figure 8**).

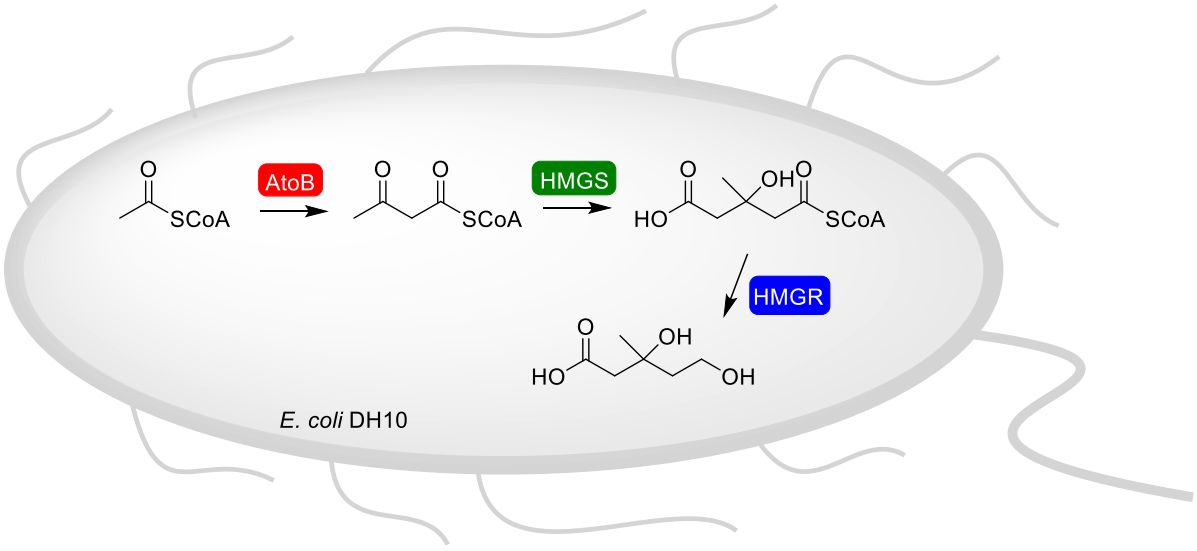


Figure 8 A three-enzyme pathway, constructed recombinantly in *E. coli*, for the production of the metabolite mevalonate from acetyl-CoA. AtoB, acetoacetyl-CoA thiolase; HMGS, hydroxymethylglutaryl-CoA synthase; HMGR, hydroxymethylglutaryl-CoA reductase.

A 77-fold increase in conversion to the product was seen for the system containing the best-performing scaffold design when compared to cells containing no protein scaffold, with the improvements in product titers shown to be very sensitive to the number of domains of each protein. For example, the addition of one extra domain for the binding of HMGS resulted in an increase from 4-fold to 77-fold, which corresponds to the fact that this step of the pathway represents a significant reaction bottleneck. Another advantage of this method for the tailoring of protein production levels is that it allows for the use of lower expressions of the recombinant proteins, since the effective concentration of the protein contained on the scaffold is the factor affecting the efficiency of the enzymatic pathway, rather than the levels of protein actually present within the host cell. Lower recombinant expression of proteins results in a lower metabolic load upon the living system[37] and would be particularly attractive when the expression of proteins that display toxicity to the host cell, or are not very soluble and/or form protein aggregates, is required.

However, the aforementioned approach to regulate the levels of expression of multiple proteins within one host cell is not without its drawbacks. For the utilisation of a synthetic protein scaffold, it is first necessary to introduce a ligand to the proteins of interest. It is, of course, then possible that the addition of a ligand to a target protein may interfere with the proper folding of that protein, and so may affect the functionality of the enzyme. For this reason, it may be necessary to investigate different locations on the protein for their suitability in hosting the necessary ligand. The introduction of a protein scaffold also requires the careful design of such a structure prior to its use, and further optimisation of the scaffold (i.e. the introduction of more/fewer protein-protein interaction domains, the introduction of new domains) may add an element of inefficiency to the process.

1.3.2.3 Gene Duplication for the Amplification of Protein Expression

One way in which living systems may amplify the dosage of a given protein in response to various environmental stimuli is through the effects of gene duplication.[65,66] It is generally accepted that the duplication of genes is the primary source of genotypic diversity, where multiple copies of a gene evolve independently of one another to lead to new functions. Examples of the presence of multiple copies of the same genes have been found in various genomes, and it has been suggested that this may confer a positive selection advantage to the organisms containing them.[65] Multiple copies of genes conferring resistance to hot[67] and cold[68] temperatures, increased uptake of essential nutrients[69] and decreased uptake of toxic heavy metals[70] have all been presented in the literature, and a relationship between the gene copy number of a given protein and its extent of production within a cell has been determined.[71] In complex

metabolic pathways, the presence of a higher copy number for a particular gene product does not necessarily correlate to an increase in the amount of protein produced, since enzyme expression levels are so tightly controlled through negative feedback loops within the organism. However, previous work has suggested that an increase in gene copy number can lead to an increase in protein, though the relationship between the two is not necessarily linear.[71]

Applying a gene duplication approach to the expression of proteins in a *de novo* enzymatic pathway is not expected to be affected by the innate regulation found within living cells, and so it may be possible to exploit this strategy for the tailoring of protein production ratios for a designed biocatalytic cascade reaction. Very recently, Subramanian *et al.* recombinantly introduced a caffeine-degradation pathway into *E. coli* cells for the production of theobromine (**Figure 9**), a compound implicated in the protection of dopaminergic neurons in the brain for the treatment of Parkinson's Disease.[72]

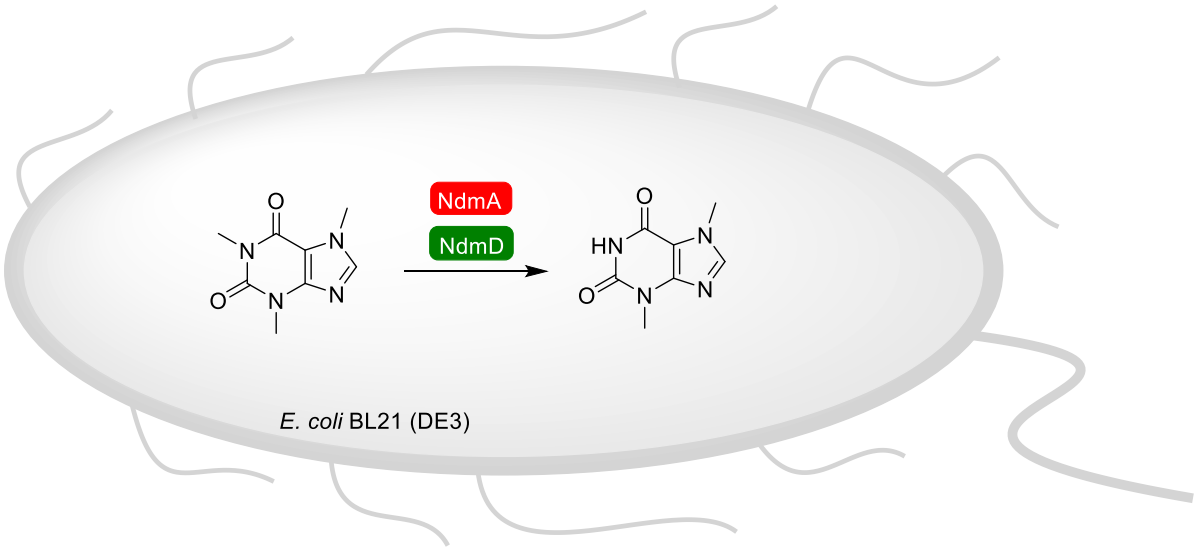


Figure 9 *In vivo* cascade, composed of a monooxygenase (NdmA) and its reductase partner (NdmD), for the *N*-demethylation of caffeine into the product theobromine.

The system employed the use of two enzymes, the monooxygenase NdmA and its reductase partner NdmD, co-expressed on separate plasmids in a series of five whole cell biocatalysts, each containing various iterations of the required enzymes. It was seen that introducing a plasmid hosting three copies of the gene encoding for the reductase protein led to an increased activity of the whole cell biocatalyst, with conversions of up to 98.5 % observed, suggesting that higher levels of NdmD are important for the production of theobromine from caffeine. In another recent example that suggests an increased gene copy number of a particular protein can lead to higher conversions during a biocatalytic reaction, Kroutil *et al.* showed that the introduction of an additional copy of the gene encoding an important recycling enzyme (L-alanine dehydrogenase, AlaDH) for an *in vivo* ω-transaminase (ω-TA) reaction led to an increased catalyst efficiency.[73] This example will be explored further in Section 1.6.2.2.

The exploitation of gene duplication for the optimisation of *de novo* biocatalytic processes has not yet found wide application, with most examples of its use being restricted to the optimisation of flux balance in metabolic processes. For this reason, it would be prudent to explore this concept with respect to the design and construction of novel enzymatic cascade reactions in the future.

1.4 Compounds of Interest

1.4.1 Chiral Secondary Amines

Chiral secondary amines are ubiquitous functionalities found in natural products, and as such are increasingly applied in medicinal and pharmaceutical chemistries to obtain highly desirable bioactive molecules. Many of these interesting chemical building blocks, however, are not readily accessible through natural biosynthetic pathways, or are available in insufficient quantities.[27]

Chemical synthesis of such chiral amines is not trivial and frequently involves the employment of transition metal catalysts,[74,75] the use of high pressures or undesirable solvent systems,[76] and often results in the production of unwanted intermediate racemates.[76] In particular, chiral piperidine-containing compounds and their analogues are implicated in a vast amount of clinical research[74,76] and a facile enzymatic route to these scaffolds is of great interest.

1.4.2 Amides

Despite the widespread occurrence of amide bonds within pharmaceutically-relevant compounds,[77,78] efficient amide synthesis remains one of the most prominent challenges left to the organic chemist. Typically a chemical coupling reagent is used to activate the acid group of an *N*-protected amino acid, with the activated carboxylate then reacting with the amino group of a *C*-protected amino acid (or free amine) in the presence of base to generate a new peptide (or amide) bond. This requirement for the addition of protecting groups and expensive or toxic coupling reagents leads to the chemical synthesis of amides having a poor atom economy, and so alternate (enzymatic) routes to amide bond formation are of great interest.

1.5 Cascade Development Through a Design-Build-Test-Analyse (DBTA) Approach

The task of designing a *de novo* enzymatic pathway is not an insignificant one. An often-cited advantage of biocatalysis over traditional chemical synthesis is the fact that enzymes generally work optimally at physiological conditions,[79] yet

there is no guarantee that a chosen consortium of enzymes will all function well in a given environment. Similarly, the simultaneous production of multiple recombinant enzymes within a single cell may be difficult due to differences in the optimal growth and expression conditions associated with each protein (Section 1.2.2). The enzymes comprising the cascade must also work cooperatively and consecutively in order to successfully transform a substrate into the desired compound, with no cross-talk between reactions. Only when these requirements are met can a successful *in vivo* biocatalytic cascade be constructed.

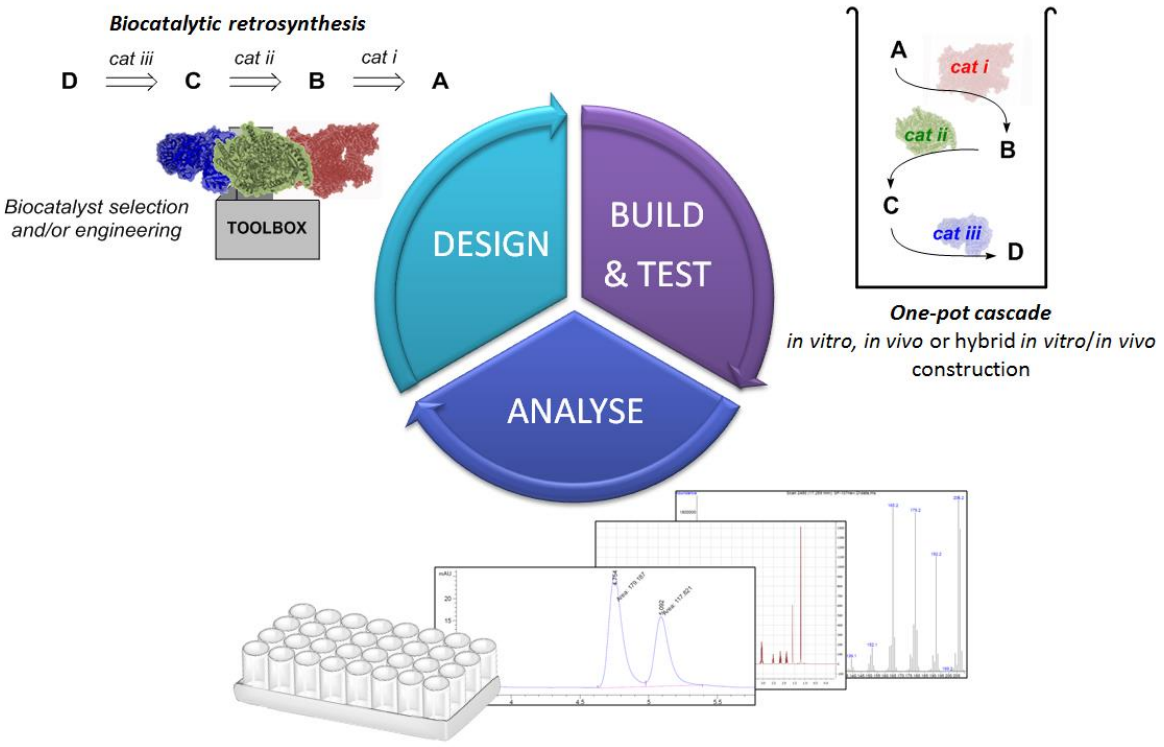


Figure 10 Design-build-test-analyse (DBTA) cycle of multi-enzyme cascade development.

When building genetic circuits or metabolic pathways into living organisms, synthetic biologists implement the design-build-test-analyse (DBTA) cycle to

rapidly assess the viability of their system.[30,80] This approach can be adapted for the design and construction of enzymatic cascade reactions (**Figure 10**), with the following steps taken for constructing novel *in vivo* biocatalytic cascades in particular: (1) the design of an appropriate enzymatic pathway through biocatalytic retrosynthetic analysis (Section 1.5.1),[81,82] (2) the selection of suitable homologs for each enzyme class involved in the proposed cascade, based on differences in substrate scope and/or selectivities, (3) the combinatorial assembly of a library of plasmids containing various mixtures of cascade genes, (4) the introduction of a single (or multiple) plasmid(s) into a suitable chassis and soluble expression of cascade proteins, (5) screening of each whole cell system against appropriate substrate(s) and analysing the biotransformations with a suitable assay. Using this workflow a number of pathway scenarios can be rapidly assessed, without the need for prior isolation and purification of the potential cascade enzymes, and the most successful whole cell biocatalysts can subsequently be taken through to process optimisation (Section 1.2.4 and 3.3).

1.5.1 Biocatalytic Retrosynthesis

As outlined in the above section, the initial stage of developing an enzymatic cascade reaction involves the biocatalytic retrosynthetic analysis of a target scaffold or specific compound.[81] This approach is analogous to the retrosynthesis of molecules, where the systematic disconnection of suitable chemical bonds within target compounds takes place to yield 'synthons'; building blocks that can be utilised for the production of a given chemical structure through logical synthetic reaction steps. The ever-increasing toolbox of biocatalysts, composed of a vast amount of enzymes capable of a myriad of functional group interconversions (FGI) and biotransformations, enables the implementation of

multiple enzymatic reactions in one system for the production of compounds of interest.

In the following section, several enzyme classes suitable for use within biocatalytic cascade reactions will be explored, with examples of their previous application in a number of multistep processes highlighted.

1.6 Biocatalysts of Interest for Enzymatic Cascade Reactions

1.6.1 Carboxylic Acid Reductase (CAR)

1.6.1.1 Introduction to Carboxylic Acid Reductase

Carboxylic acid reductases (CARs) are ATP- and NADPH-dependent enzymes that catalyse the selective reduction of carboxylic acids to the corresponding aldehydes for a broad range of substrates.[8,83–89] CARs are large proteins that are homologous to nonribosomal peptide synthetases (NRPS) and consist of three distinct domains; an N-terminal adenylation domain, a linking peptidyl carrier protein (PCP) domain, and a C-terminal thioester reductase domain. Prior to catalysis the PCP must be posttranslationally modified through the addition of coenzyme A, catalysed by a 4′-phosphopantetheinyl transferase (Sfp), resulting in a 4′-phosphopantetheine (Ppant) prosthetic group attached to a conserved serine residue.[90] The mode of action is proposed to proceed via four mechanistic steps: (1) the carboxylic acid substrate and an equivalent of ATP enter the adenylation domain, where the carboxylate attacks ATP to form an acyl-AMP intermediate through the release of pyrophosphate (PPi), (2) the thiol group of the Ppant arm attacks the acyl-AMP intermediate to tether the acyl substrate and release AMP, (3) the Ppant arm shuttles the tethered acyl intermediate towards the thioester reductase domain, where, (4) the bound acyl-thioester is reduced using NADPH to release aldehyde product, $NADP^+$ and a regenerated Ppant arm (**Figure 11**).[8]

Further reduction of the aldehyde to its corresponding alcohol is not catalysed by CAR, with a recently obtained crystal structure of a CAR PCP-reductase didomain suggesting that the docking of the Ppant prosthetic group in the reductase active site changes the orientation of a key aspartate (Asp) residue, making the conversion to the alcohol impossible.[89]

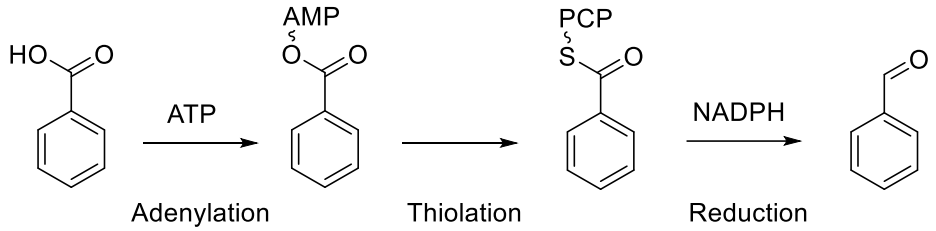


Figure 11 The conversion of CAR substrate benzoic acid to benzaldehyde product, through the adenylation, thiolation and subsequent reduction of the carboxylic acid catalysed by three individual domains.

The conversion of carboxylic acids to aldehydes catalysed by CAR is an extremely thermodynamically favourable reaction due to the hydrolysis of ATP, and as such these enzymes have found broad application for the production of a number of synthetically useful aldehydes,[8,85,88,91] as well as use in multistep biocatalytic cascades.[40,87,91,92] An analysis of the types of substrates generally accepted by CAR enzymes suggests that aromatic carboxylic acids, containing electron donating substituents at the α-position, are converted to the corresponding aldehyde much more readily than when a ring-deactivating group is present, presumably through the delocalisation of electrons into the π-system making the attack on ATP more favourable.[88] In general, substrates where the carboxylate is the only polar group present within the compound are accepted much more readily by CAR, yet activities are still seen for substrates containing an additional carbonyl group or methoxy group, for example.[88] The conversion of more polar substrates was less efficient for the CAR biocatalysts analysed, but

these compounds could still be viable for use in enzymatic cascade reactions since later steps of the process could increase the rate of aldehyde formation catalysed by CAR. Towards this end, CAR has been utilised in both hybrid (Section 1.2.3)[40] and *in vivo*[91–93] (Section 1.2.2) biocatalytic cascade reactions.

1.6.1.2 Application of Carboxylic Acid Reductase in Enzymatic Cascade Reactions

An enzymatic cascade reaction employing a CAR enzyme was developed for the conversion of 3,4-dihydroxyphenylacetic acid to 3-hydroxytyrosol by Winkler *et al.* in 2014.[93] A whole cell biocatalyst was constructed through the recombinant co-expression of NCAR (Section 2.1 and Chapter 3) and Sfp, making use of endogenous enzymes within *E. coli* for the conversion of CAR product aldehyde to the desired alcohol product. A similar approach was taken by Yakunin *et al.* recently,[91] where a cascade composed of CAR in conjunction with an aldo-keto reductase (AKR) catalysed the conversion of 4-hydroxybutanoic acid and adipic acid to 1,4-butanediol and 1,6-hexanediol, respectively. When this cascade was employed through an *in vitro* approach, substrate conversions of up to 90 % were seen, and the construction of a whole cell biocatalyst took advantage of endogenous AKR within the living system for the conversion to the final product diols.

An enzymatic cascade for the conversion of primary alcohols to aldol products, through the use of an alcohol dehydrogenase AlkJ (ADH, Section 1.6.5) and an aldolase Fsa1-A129S, was recently described by Rudroff *et al.*,[92] where the co-expression of NCAR enabled the maintenance of non-toxic levels of intermediate aldehyde within the whole cell (**Figure 12**).

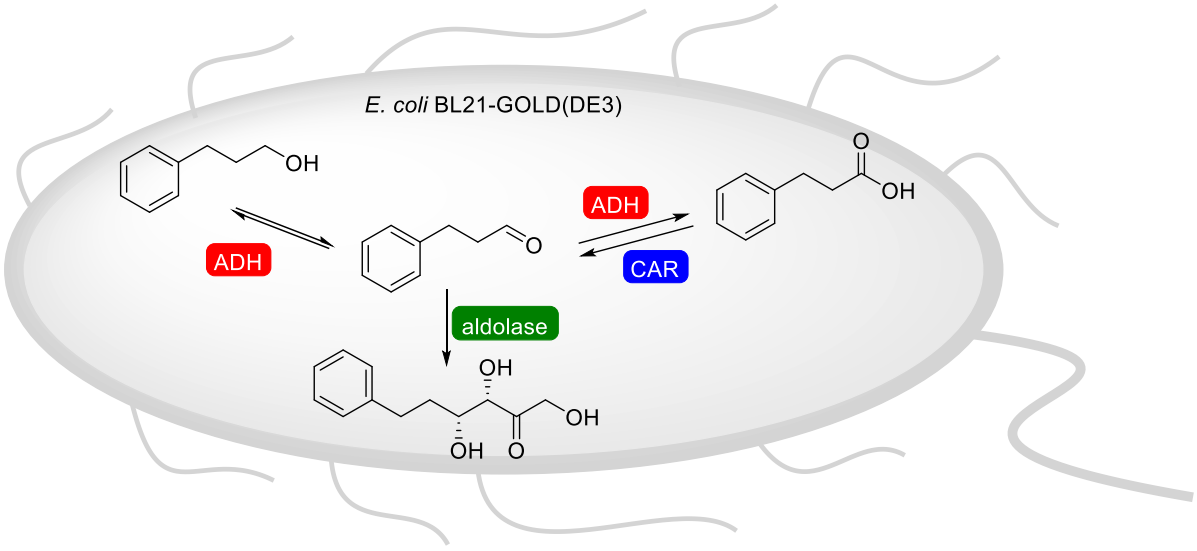


Figure 12 *In vivo* cascade for the conversion of primary alcohol to aldol product, through the use of ADH and aldolase enzymes. NCAR was provided to regenerate any aldehyde intermediate that may have been oxidised to the corresponding carboxylic acid by ADH present in the system.

This four-enzyme system was recombinantly expressed in *E. coli* whole cells through the co-expression of two compatible plasmids, containing NCAR and Sfp, and AlkJ and Fsa1-A129S, in an attempt to limit the plasmid burden exerted on the host cells. However, it was seen that the co-expression of these four genes led to significant metabolic burden on the cell, resulting in low catalyst efficiencies for the final step of the cascade reaction. In an attempt to boost conversion to aldol product, purified aldolase enzyme was subsequently added to the system exogenously and a significant increase to product formation was observed, with 85 % aldol formed after 2 hours when exogenous supplementation of aldolase was provided, compared to 80 % aldol formed after 6 hours for the entirely *in vivo* reaction set-up. This example highlights the need for the effective optimisation of protein expression when multiple recombinant enzymes are contained within a single cell (Section 1.3.2), but also the broad utility of CAR enzymes for a variety of envisaged cascade reactions.

1.6.2 Transaminase (TA)

1.6.2.1 Introduction to Transaminase

Transaminases (TAs), or aminotransferases, catalyse the reversible reductive amination of a carbonyl group within a substrate of interest, through the transfer of an amine group from a sacrificial donor substrate, to yield a chiral amine product.[94–97] These enzymes do not require the presence of cofactors such as NAD(P)H, but rely on the vitamin B_6-derivative cosubstrate pyridoxal 5'-phosphate (PLP), which is linked to an active site lysine residue through an imine bond and facilitates the shuttling of the donated amine group.[95,97] Utilisation of PLP rather than other cofactors commonly found within enzymatic processes is advantageous given the fact that the molecule is regenerated after the reaction is completed, allowing its presence in catalytic amounts to be sufficient for the successful conversion to the desired products. The binding of PLP within the active site also enables very high enantioselectivities of amine product, since the orientation of the substrate binding is dictated by the position of PLP relative to the substrate, making this class of enzymes very attractive for use in biocatalytic processes.[97] However, the equilibrium of the reversible reaction often lies towards carbonyl, rather than amine, production, and so various methods have been employed to shift transaminase reactions towards amine synthesis, including but not limited to: (1) the provision of an excess amount of sacrificial amine donor (e.g. alanine, isopropylamine), but higher concentrations may lead to problems with donor substrate insolubility or inhibition of the transaminase or other enzymes in the system, (2) removal of the reaction product once the reaction has been completed, or removal and/or recycling of the coproduct formed through the amine donor (e.g. the removal of pyruvate through cell metabolism or exogenous systems such as a lactate dehydrogenase (LDH)/glucose dehydrogenase (GDH) system, the evaporation of coproduct

acetone), (3) the use of smart sacrificial diamine donor substrates that can spontaneously cyclise upon donation of an amine group, thus "removing" the coproduct,[98] and, (4) the addition of subsequent reaction steps to build multi-enzyme cascade processes that can pull the transaminase reaction through towards amine production.[97]

The ability to successfully exploit *ω*-transaminases for both aldehyde/ketone formation and amine formation has resulted in the application of these biocatalysts within a number of enzymatic cascade reactions towards the production of a variety of compounds.[95,99–101]

1.6.2.2 Application of Transaminase in Enzymatic Cascade Reactions

The synthesis of chiral 2,5-disubstituted pyrrolidines through the development of a chemoenzymatic cascade was achieved by Turner *et al.* in 2014, through the reductive amination of the least hindered carbonyl group on a 1,4-diketone substrate, catalysed by an *ω*-TA, followed by spontaneous intramolecular cyclisation of the resulting 1,4-aminoketone to give a chiral imine intermediate (Figure 13).[46] Nonselective chemical reduction using $NH_3.BH_3$ then converted the imine into the racemic amine, where a monoamine oxidase (MAO-N) catalysed the regio- and stereoselective oxidation of one enantiomer back to the imine intermediate. Through this process, the enantiomer not selected for by MAO-N accumulated, leading to an optically pure product amine.

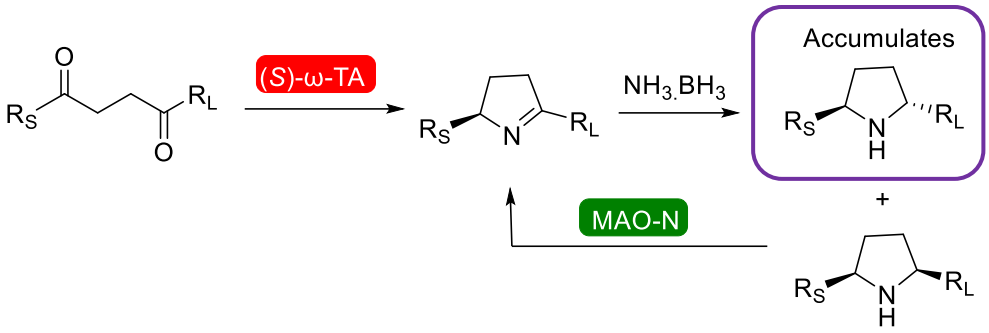


Figure 13 Chemoenzymatic cascade reaction employing an *ω*-TA and a MAO-N for the conversion of 1,4-diketones to the corresponding 2,5-disubstituted pyrrolidines.

Several *ω*-TAs were assessed for their suitability in the envisaged cascade reaction, including the (*S*)-selective ATA-113 variant and the (*R*)-selective ATA-117 variant (Section 2.1 and Chapter 3),[102] and it was observed that ATA-113 could catalyse 91 % (> 99 % ee) conversion to (*S*)-pyrroline whilst ATA-117 only achieved 65 % conversion to the corresponding (*R*)-enantiomer (> 99 % ee). Implementing the full TA-MAO-N cascade using ATA-113 and MAO-N D5 led to conversions and de values of > 99 % for several of the 1,4-diketone substrates screened.

A four-enzyme *in vivo* cascade, hosted within a single whole cell, for the production of enantiopure (*R*)-1-phenylethylamines from 4-substituted ethylbenzenes was constructed by Flitsch *et al.* through the use of P450 monooxygenase, alcohol dehydrogenase (ADH, Section 1.6.5) and *ω*-TA biocatalysts (**Figure 14**).[47] The cascade reaction was initiated by the regioselective oxidation of an unactivated C-H to give racemic benzylic alcohol, which was further oxidised to the ketone through the application of two enantiocomplementary ADH enzymes, and finally chiral benzylic amine product was obtained through the enantioselective reductive amination of intermediate ketone by *ω*-TA ATA-117 (see above and Chapter 3).

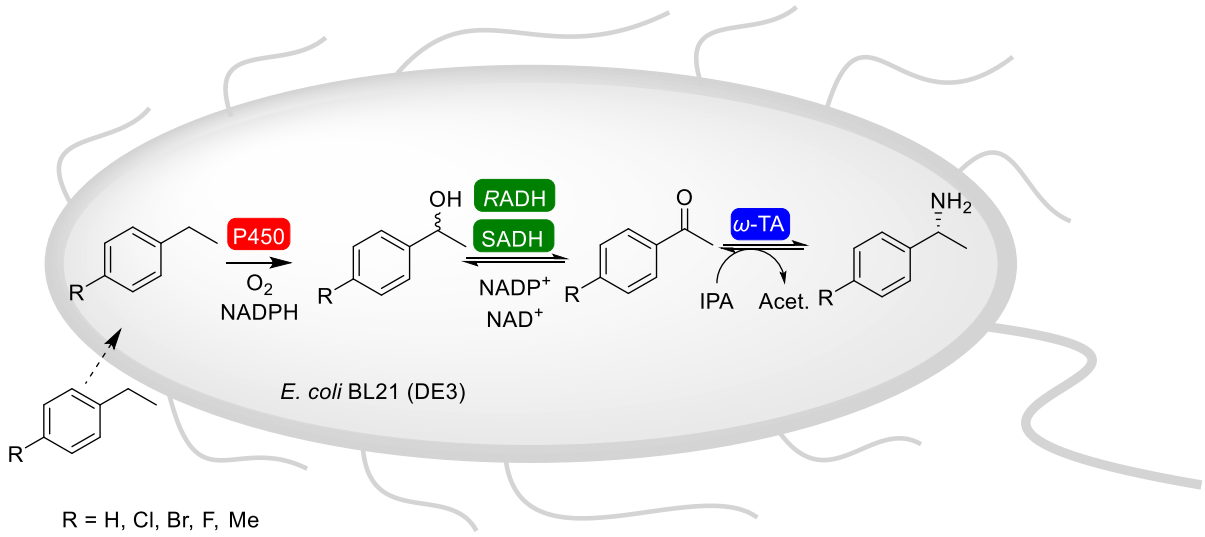


Figure 14 *In vivo* cascade developed for the conversion of ethylbenzenes to chiral benzylic amines through the use of P450, ADH and ω-TA enzymes.

The four cascade genes were expressed recombinantly in *E. coli* host cells through the use of two compatible plasmids. Plasmid one contained both ADHs and ATA-117 in a monocistronic operon design, and was assembled through the utilisation of the BioBrick™ cloning platform (Section 1.3.1.2.2) using pPB01 as the plasmid backbone (Section 2.1.1), and plasmid two contained an individual gene encoding the P450 enzyme to enable increased gene expression control over this potentially toxic protein. The designed cascade reaction sequence necessitated high equivalents of the ω-TA sacrificial amine donor isopropylamine (IPA, 200 eq.) to increase conversions to product amine, due to the fact that the transamination reaction was the final step of the process, and such high concentrations were seen to inhibit the initial P450-catalysed step of the cascade. For this reason a two-step one-pot reaction design was implemented, with addition of IPA after 24 h, resulting in the conversion of ethylbenzenes to their benzylic acid counterparts of up to 26 % and ee values of up to 97.5 %.

An alternative approach to the construction of multi-enzyme processes was developed by Kroutil *et al.*, in which a series of modular whole cell biocatalysts for the efficient asymmetric amination of keto acid and ketone substrates, using one of four ω-TA homologs, was developed.[73] Three modules were designed and constructed for co-expression within *E. coli* host cells: Module 1 consisted of a choice of four ω-TA biocatalysts; Module 2 consisted of a choice of two amine donor recycling enzymes (L-alanine dehydrogenase, AlaDH, or L-glutamate dehydrogenase, GluDH), and; Module 3 contained a choice of three cofactor recycling enzymes (formate dehydrogenase, FDH, glucose dehydrogenase, GDH, or phosphite dehydrogenase, PtDH) (**Figure 15**).

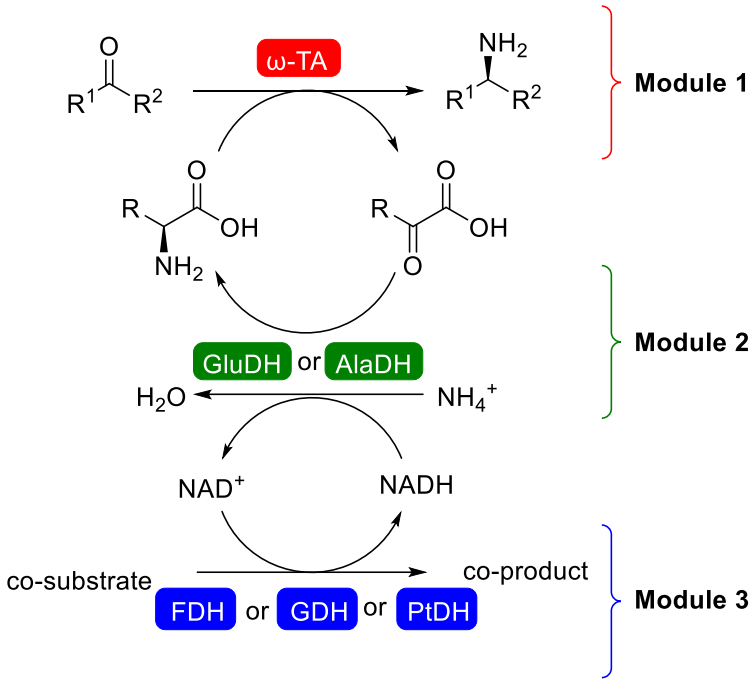


Figure 15 Three modules constructed and expressed within *E. coli* whole cells for the conversion of α-keto acid or prochiral ketone to the corresponding chiral amine.

To enable adequate control over the expression of the individual enzymes, Module 2 was integrated into the host cell genome to give two separate

engineered host cells, one expressing AlaDH and one expressing GluDH, and Module 1 and Module 3 were co-transformed into these host cells via two compatible plasmids to generate a library of potential whole cell biocatalysts. It was observed that the chromosomal integration of the amine donor recycling enzymes did not result in sufficient expression of these proteins, leading to increased amounts (250 mM) of L-alanine or L-glutamate needed for significant conversion to amine product, and so further optimisation of the ratio of enzymes was required. An additional expression plasmid was constructed containing one of the ω-TA homologs and the gene encoding AlaDH (pTrc99A-*avtA-ald*), and was transformed into the AlaDH-containing host cell. This duplication of the *ald* gene (Section 1.3.2.3) resulted in increased conversion to L-isoleucine product, rising from 68 % when one gene copy was present to 84 % when the gene had been duplicated within the system, in addition to requiring lower equivalents of the amine donor. It was possible to increase this conversion further through altering the gene order of the plasmid, with AlaDH-host cells expressing plasmid pTrc99A-*ald-avtA* resulting in 96 % conversion to the product, when compared with the AlaDH-host cells expressing plasmid pTrc99A-*avtA-ald* (84 %). This result suggests that a further optimisation parameter for biocatalytic processes is introduced when a polycistronic operon design is utilised, and so a monocistronic approach to plasmid construction may prove to be more beneficial for the expression of multiple genes from the same vector (Section 2.1). In addition to this extra parameter to be optimised, the described construction of multi-enzyme whole cell biocatalysts for a designed reaction suffers from the fact that a large library for each module must be constructed prior to use, and any additional plasmids required must be assembled from scratch. This approach is laborious and inefficient, and so it would be advantageous to employ a system where the rapid construction of new plasmids is enabled (Section 1.3.1.2.2).

1.6.3 Imine Reductase (IRED)

1.6.3.1 Introduction to Imine Reductase

Imine reductases (IREDs) are NADPH-dependent oxidoreductases that catalyse the reduction of prochiral imines into the corresponding chiral amine products.[103–107] IRED enzymes are dimeric proteins containing a Rossman fold for NADP(H) binding, and it has been suggested that an aspartate or tyrosine residue within the active site of the protein acts as a proton donor during imine reduction.[108] It was previously understood that biosynthetic IREDs exhibited a very narrow substrate scope, until an example described by Nagasawa *et al.* demonstrated the reduction of the synthetic substrate 2-methylpyrroline by two enantiocomplementary IREDs, (*R*)-IRED and (*S*)-IRED (Section 2.1 and Chapter 3), from *Streptomyces* sp. GF3587 and *Streptomyces* sp. GF3546, respectively (**Figure 16**).[107]

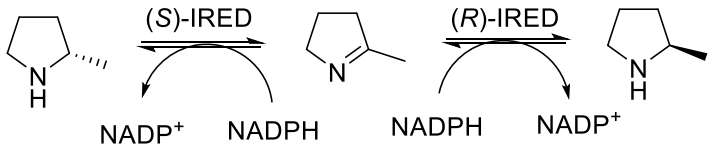


Figure 16 Imine reduction of 2-methylpyrroline to the corresponding chiral amine, catalysed by (*R*)- and (*S*)-IRED.

Following the publication of this example, several more instances of IRED-catalysed imine reductions have been published, with an emphasis on their use in biocatalytic processes[105,106,109] and enzymatic cascade reactions, a particularly attractive application due to the inherent selectivity to reduce C=N double bonds and not carbonyl groups.[40,49,98] Further characterisation of the *Streptomyces* sp.

IREDs, (*R*)- and (*S*)-IRED, was executed by Turner *et al.*, with whole cell biocatalysts expressing the recombinant enzymes capable of catalysing the conversion of a range of aliphatic and aromatic 2-substituted pyrrolines and piperideines (**Figure 17**) with excellent conversions (up to > 98 %) and selectivities (up to > 98 %), and kinetic data revealing an apparent preference for 6-membered ring substrates over the 5- and 7-membered counterparts.[105,106]

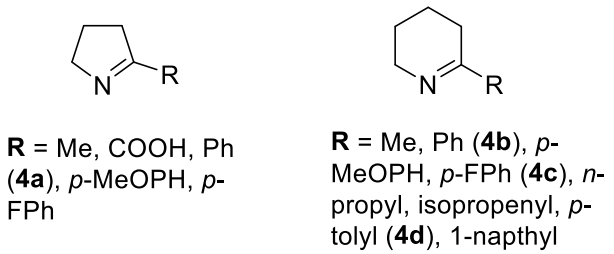


Figure 17 Examples of pyrroline and piperideine substrates successfully converted to the corresponding chiral amines by (*R*)- and (*S*)-IRED.

1.6.3.2 Application of Imine Reductase in Enzymatic Cascade Reactions

Turner *et al.* also developed an enzymatic cascade, composed of an amine oxidase (AO) and an IRED, for the deracemisation of *N*-heterocycles (**Figure 18**).[49] In this example, a mixed culture *in vivo* approach to cascade design was implemented (Section 1.2.2), with an (*R*)-selective amine oxidase 6-hydroxy-D-nicotine oxidase (6-HDNO) variant contained within one *E. coli* whole cell and (*R*)-IRED or (*S*)-IRED contained within a separate *E. coli* whole cell.

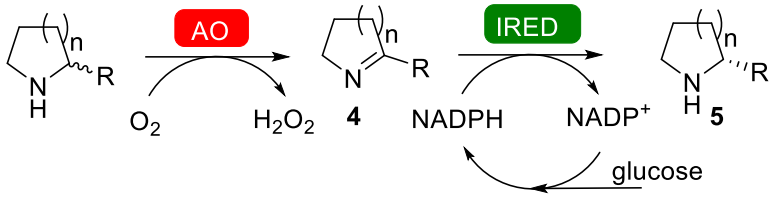


Figure 18 Deracemisation of pyrrolidines and piperidines through an *in vivo* cascade consisting of amine oxidase (AO) and imine reductase (IRED) enzymes, with NADPH cofactor recycled by the whole cell.

A variety of aliphatic and aromatic substituents at the C-2 position were analysed, and the reaction parameters were optimised prior to the preparative-scale synthesis of (*S*)-2-phenylpiperidine **5b** (82 % isolated yield, 99 % ee) using 6-HDNO coupled with (*R*)-IRED. Attempts to co-express both cascade enzymes within the same host cell were unsuccessful, presumably due to differences in the optimal expression conditions of each individual protein (Section 1.2.2), and so additional investigations into tailoring the expression of these enzymes, alongside further reaction engineering, are necessary before the cascade can be successfully implemented in a single cell.

In addition to catalysing imine reduction, it has recently been demonstrated that some IREDs also display reductive amination activity towards a range of ketones using a number of amines.[110–112] For these examples it was shown, through steady-state kinetics, that the enzyme first catalyses imine formation from the ketone substrate and subsequently catalyses imine reduction, allowing access to a wide variety of chiral amine products from achiral starting materials.[110]

The ability of IREDs to catalyse two important chemical reactions, namely imine reduction and reductive amination, render them attractive tools in the construction of biocatalytic cascade reactions.

1.6.4 Galactose Oxidase (GOase)

1.6.4.1 Introduction to Galactose Oxidase

Galactose oxidases (GOases) are single copper metalloenzymes that catalyse the conversion of the C-6 hydroxy functionality of D-galactose to the corresponding aldehyde, recruiting one atom of oxygen from molecular oxygen and concomitantly producing hydrogen peroxide (H_2O_2) as by-product.[113,114] High concentrations of H_2O_2 lead to inactivation of GOase, and so removal of this by-product is paramount in order to successfully exploit these enzymes. The use of a horseradish peroxidase (HRP) enzyme in conjunction with 2,2′-azino-bis-(3-ethylbenzthiazoline-6-sulfonic acid) (ABTS) substrate has been exploited for the removal of H_2O_2,[115] with a more atom efficient approach being the supplementation of biotransformations with catalase enzyme for the effective conversion of H_2O_2 to give H_2O.[115–117] In fact, it was seen that the supplementation of both HRP and catalase to the GOase-catalysed oxidation of methyl-D-galactopyranoside to its corresponding aldehyde increased the conversion from 60 - 70 % when just catalase was used, to 90 % with the HRP/catalase system.[118] The fact that GOase enzymes do not require the use of a cofactor such as NAD(P)H make them an attractive alternative to other aldehyde-forming enzymes within the biocatalytic toolbox, but the wild type enzymes unfortunately display very high selectivities towards the natural substrate D-galactose, limiting their use within *de novo* cascade reactions. This drawback led to the engineering of GOase enzymes from *Fusarium* sp., such as GOase M_{3-5}[116,117,119] and GOase F_2,[117,120] for application with a broader range of non-natural alcohol substrates. These improved mutants have led to the use of GOase in several enzymatic cascade reactions, catalysing the conversion of a range of aryl and alkyl alcohols to the corresponding aldehyde at the expense of O_2.[115–117,121]

1.6.4.2 Application of Galactose Oxidase in Enzymatic Cascade Reactions

Faber *et al.* designed a two-enzyme *in vitro* cascade system composed of a GOase variant from *Fusarium* NRRL 2903 and one of seven *ω*-transaminases (*ω*-TA, Section 1.6.2) for the conversion of benzylic and cinnamic alcohols to their corresponding amine products (**Figure 19**).[115] The reaction employed the use of an L-alanine dehydrogenase (AlaDH) for the recycling of the sacrificial amine donor L-alanine for the transamination reaction, and either a formate dehydrogenase (FDH)/formate or a glucose dehydrogenase (GDH)/glucose system for the recycling of NADH cofactor.

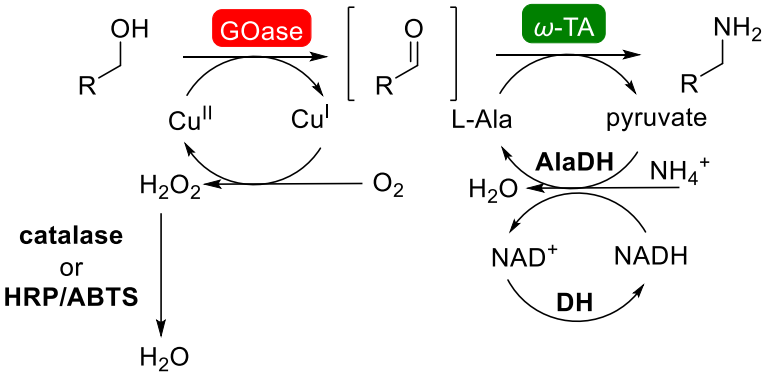


Figure 19 *In vitro* cascade for the GOase-TA-catalysed conversion of primary alcohol into the corresponding amine product, with an L-alanine and NADH cofactor recycling system.

Of the *ω*-TAs screened, the enzyme from *Vibrio fluvialis*, Vf-*ω*-TA, was the most efficient catalyst, with 96 % conversion to benzylamine seen when benzaldehyde was used as the substrate. Lyophilised whole cells of *ω*-TA were then used in conjunction with lyophilised whole cells of GOase, with optimal cascade conversions of > 99 % seen when a GDH/glucose cofactor recycling system was employed. The containment of the biocatalysts within lyophilised whole cells

removed the need for laborious protein purification, yet the use of nonviable cells necessitated an exogenous supplementation of the system with NADH cofactor.

In addition to this, Carnell, Turner *et al.* employed previously evolved GOase variants in the oxidation of various aliphatic and aromatic amino alcohols, followed by subsequent conversion with a xanthine oxidoreductase (XOR) for the production of lactams (**Figure 20**).[117]

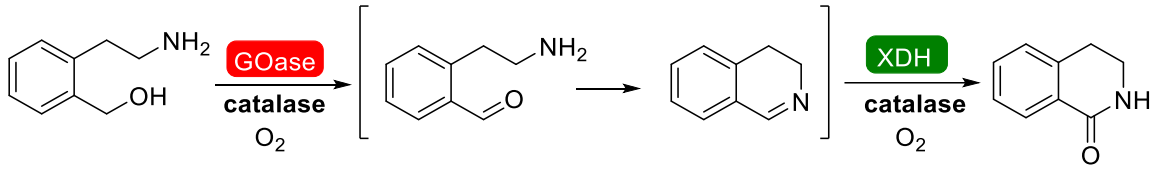


Figure 20 *In vitro* GOase M3-5-XDH mediated synthesis of 3,4-dihydroisoquinolin-1(2H)-one from amino alcohol.

A screen of potential GOase substrates revealed that both of the GOase variants tested, GOase M3-5 and GOase F2, were more efficient when *N*-Cbz-protected aliphatic amino alcohols were used in comparison with the corresponding free amines, with GOase F2 showing a marked preference for longer chain lengths. GOase M3-5 was then used for the one-pot transformation of amino alcohol to 69 % 3,4-dihydroisoquinolin-1(2H)-one product, through a cascade reaction involving xanthine dehydrogenase (XDH) and a spontaneous intramolecular cyclisation between amine and aldehyde functional groups (**Figure 20**). An alternate XOR, periplasmic aldehyde oxidase (PaoABC), was also utilised here to access these lactam products, and a GOase M3-5-PaoABC cascade was reported previously for the oxidation of alcohols to carboxylic acids.[121] This GOase M3-5-PaoABC reaction made use of considerably high substrate loadings (up to 100 mM) for a biocatalytic cascade process and was employed in the preparative-

scale synthesis of the bioplastics precursor furan-2,5-dicarboxylic acid (FDCA) (74 % isolated yield).

The ability for engineered GOase variants to convert relatively high substrate concentrations of a broad range of primary and secondary alcohols to the corresponding aldehyde, without the need for addition of expensive cofactors such as NAD(P)H, renders these biocatalysts an attractive route towards synthetically useful aldehyde products.

1.6.5 Alcohol Dehydrogenase (ADH)

1.6.5.1 Introduction to Alcohol Dehydrogenase

Alcohol dehydrogenases (ADHs) are NAD(P)H-dependent enzymes that are able to catalyse the reversible oxidation of a primary or secondary alcohol to the corresponding aldehyde or ketone product.[122] Due to the reversible nature of this reaction, and the fact that the reaction equilibria lies towards production of alcohol, it is beneficial to have a subsequent reaction step when an alcohol substrate is to be used. Both enantioselective and nonselective ADHs have been used for the generation of ketone and reactive aldehyde intermediates within many biocatalytic cascades, and the following section will highlight several of these systems.[38,41,47,53,92,123,124]

1.6.5.2 Application of Alcohol Dehydrogenase in Enzymatic Cascade Reactions

In an elegant solution to the issue of expensive cofactor requirements, a cascade reported by Turner *et al.* used an ADH coupled with an amine dehydrogenase (AmDH) for the production of chiral amines from the corresponding alcohols through the production of a ketone intermediate (**Figure 21**).[123] The NADH

cofactor generated through the oxidation of alcohol substrate by ADH was utilised in the following reductive amination of intermediate ketone with ammonia, catalysed by AmDH, to obtain a chiral amine product.

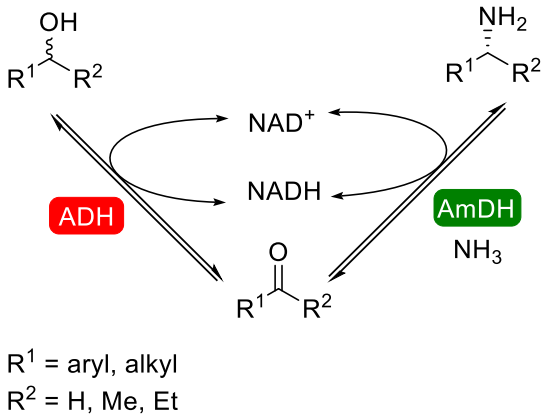


Figure 21 *In vitro* ADH-AmDH hydrogen-borrowing cascade for the production of chiral amines from racemic alcohols.

The recycling of NAD(H) cofactor through a biocatalytic hydrogen-borrowing process led the cascade to be redox neutral, since the hydrogen atom removed in the first step is reinstated in the second. Following optimisation of the reaction parameters, to avoid the depletion of NADH through competing reactions catalysed by endogenous proteins present in the crude lysate biocatalyst preparation, high conversions to chiral amine product were achieved (85 %, > 99 % ee) and preparative-scale syntheses afforded moderate to excellent isolated yields (30 – 91 %) and high selectivities (82 to > 99 % ee). A similar approach was adapted for a cascade involving a nonselective ADH, alongside a reductive aminase (*Asp*RedAm, Section 1.6.3),[110] for the production of a wide range of secondary amine products with only water released as a waste product.[124]

A nonselective ADH was also used in the construction of a three-enzyme cascade, executed through an *in vivo* approach within a single *E. coli* whole cell, for the conversion of simple allylic alcohol substrates into the corresponding chiral lactone products (**Figure 22**).[125] The cascade reaction was initiated through the conversion of an alcohol substrate to the α,β-unsaturated ketone, catalysed by ADH, followed by the selective reduction of the C=C bond by an ene-reductase (ERED), which was finally oxidised further by a Baeyer-Villiger monooxygenase (BVMO) to afford chiral lactone product.

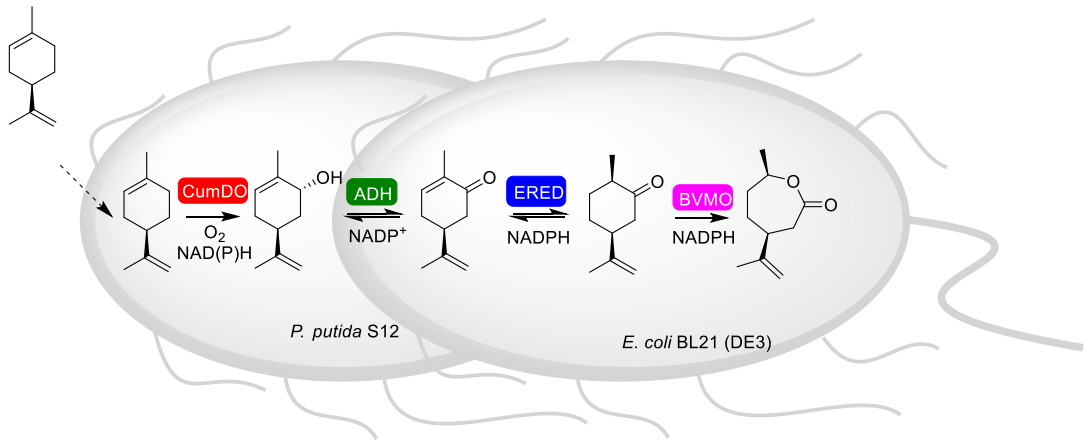


Figure 22 *In vivo* CumDO-ADH-ERED-BVMO cascade for the production of carvolactone from substrate limonene using a mixed culture reaction.

The presence of the ERED step following the ADH-catalysed conversion of alcohol to unsaturated ketone intermediate successfully shifted the equilibrium of the process towards the desired product formation, and thus a potential bottleneck to the designed cascade was mitigated. This ADH-ERED-BVMO cascade was extended further through the incorporation of a fourth enzyme, cumene dioxygenase (CumDO), within a second host cell (*P. putida* S12), to allow the natural product limonene to be the source of the alcohol substrate required

for the previously established three-enzyme cascade (**Figure 22**).[38] Mixed culture biotransformations resulted in around 47 % conversion to the chiral carvolactone product, with further process optimisation increasing the efficiency of this system to almost full conversion to the desired lactone product.

1.6.6 CapW

Most enzymatic routes for the formation of amide bonds (e.g. nonribosomal peptide synthases (NRPS), ATP-grasp enzymes)[126,127] rely on the ATP-dependent activation of one or more substrates. However, recently an enzyme was found within the biosynthetic gene cluster for A-503083 capuramycin-type antibiotics that catalyses the formation of an amide bond between a methyl ester **14a** and L-aminocaprolactam **15a** in the absence of ATP (**Figure 23**).[128] Activation of the acyl substrate occurs via S-adenosylmethionine (AdoMet)-dependent esterification by another enzyme, CapS, to give the methyl ester derivative of A-503083 B. This reaction is followed by an amide-ester exchange reaction catalysed by CapW. The amide bond formation is a thermodynamically favourable reaction, with hydrolysis not seen even after prolonged incubation of CapW with A-503083 B.

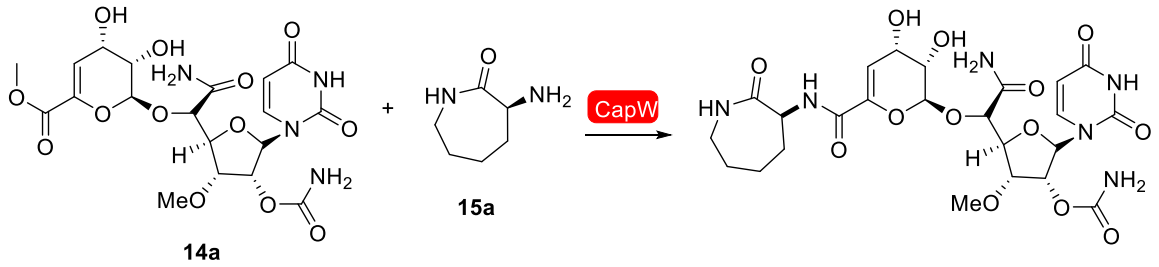


Figure 23 ATP-independent amide bond formation catalysed by the enzyme CapW.

It was proposed that, given the sequence similarity between CapW and the class C β-lactamases, this novel enzyme may catalyse the amide bond formation via an acylenzyme intermediate, as seen with C β-lactamases. This theory was tested through mutation studies, where a serine residue thought to act as an initial nucleophile to form the acylenzyme was mutated. Deletion of the serine residue led to a loss of activity, supporting the theory that CapW proceeds via an acylenzyme intermediate. It is not yet clear whether CapW is a member of a larger family of enzymes responsible for the ATP-independent formation of amide bonds, but genome mining may lead the way to more examples of enzymes using this methyl ester strategy to form amide bonds.

CapW has not yet been explored for its suitability in catalysing amide bond formation using non-natural ester and amine substrates, in part due to the current inability to solubly express large amounts of the protein in standard expression strains such as *E. coli*, yet an ATP-independent route towards amide synthesis in aqueous conditions is a very attractive target for biocatalytic processes.

1.7 Research Project Aims

The overarching goal of this research project was to design and construct novel multi-enzyme cascade reactions for the production of chiral amine scaffolds, through the exploitation of complementary biocatalysts capable of catalysing a range of chemical transformations. In particular, focus was directed at:

(**1**) the design and development of a single whole cell biocatalyst, capable of catalysing the production of chiral pyrrolidine and piperidine building blocks, through the soluble and active co-expression of several enzyme partners and with the host cell meeting all biocatalyst cofactor requirements;

(**2**) the exploration and evaluation of different enzymatic cascade implementation strategies, through the development of *in vivo*, *in vitro* and hybrid *in vitro* / *in vivo* cascade reactions, for the synthesis of a range of chiral piperidine products, and;

(**3**) the development of a potential new candidate for the biocatalytic toolbox, through its soluble expression in standard bacterial production strains and assessment of activity against a range of methyl ester substrates, towards enzymatic amide bond formation in aqueous environments.

In order to achieve these aims, it would first be necessary to select biocatalysts that are compatible with each other. Cascade design was accomplished through the biocatalytic retrosynthetic analysis of target pyrrolidine and piperidine structures, and multiple candidates from each of the enzyme classes described in Section 1.6 were assessed for their suitability in a range of transformations.

Where a whole cell biocatalyst was to be used for the implementation of a cascade reaction, optimisation of the ratios of the recombinantly expressed proteins required for the process would have to first be achieved, in order to avoid potential reaction bottlenecks in the designed cascade systems and to increase flux towards formation of the target products.

Once successful, these biocatalytic transformations were to be executed on a preparative-scale, to assess their future applicability for industrial processes.

2 Generation of Biocatalyst Libraries for Use in Cascade Reactions

The decision to build a particular type of cascade (i.e. *in vitro*, *in vivo*, hybrid *in vitro* / *in vivo*, Section 1.2) calls for the use of different approaches to the design, generation and production of the biocatalysts needed for the reaction to proceed.

If the designed enzymatic cascade is to be implemented as an *in vivo* system in a single host cell, it is beneficial to assemble multi-gene expression constructs containing all of the genes necessary to form an efficient whole cell biocatalyst (Section 1.3). Once the chosen host cell has been transformed with the relevant expression construct, optimisation of soluble protein expression is essential to ensure all cascade proteins are produced to an adequate level under the same conditions (Section 1.3.2).

If the proposed enzymatic cascade is to be executed as an *in vitro* system, on the other hand, each biocatalyst needed for the reaction can be produced individually, meaning that no compromise with respect to protein expression conditions is required. Cascade reactions that are to be employed in an *in vitro* fashion generally make use of either purified protein or crude cell lysates, many of which are commercially available from a number of suppliers.

The enzymatic cascades designed and developed in this body of work included both the *in vivo* and the *in vitro* application of a number of enzymes, and so it was first necessary to generate libraries of biocatalysts for this purpose.

2.1 Preparation of Biocatalysts for Carboxylic Acid Reductase-Transaminase-Imine Reductase (CAR-TA-IRED) Cascade Reactions

The disconnection of chiral piperidines and pyrrolidines suggested that these cyclic architectures could be accessed via linear keto acid substrates transformed by the use of an enzymatic cascade employing carboxylic acid reductase (CAR, Section 1.6.1), ω-transaminase (ω-TA, Section 1.6.2) and imine reductase (IRED, Section 1.6.3) enzymes (**Figure 24**). Due to the various cofactor requirements of the biocatalysts present in this cascade, it was proposed that an entirely *in vivo* methodology would be beneficial for the process, and so it would first be necessary to overexpress all cascade enzymes within the same host cell (see Chapter 3).

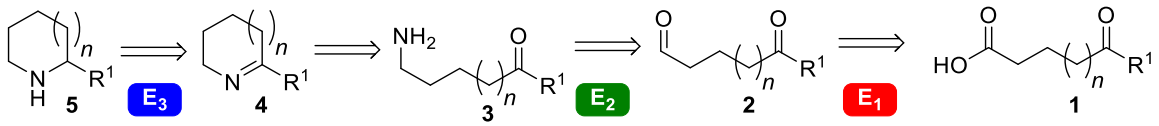


Figure 24 Biocatalytic retrosynthetic analysis of chiral piperidines, suggesting an enzymatic route from keto acid substrate to piperidine product through the use of E_1 (CAR), E_2 (TA) and E_3 (IRED) biocatalysts.

In order to generate a library of constructs for this cascade, two homologs of each enzyme were selected for screening, alongside a 4'-phosphopantetheinyl transferase from *Bacillus subtilis* (*Bs*Sfp)[90] which was needed for the activation of the CAR apoenzyme (Section 1.6.1). CAR from *Mycobacterium marinum* (*mm*CAR/MCAR)[83] and *Nocardia iowensis* (*ni*CAR/NCAR)[129] were selected for the first step of the cascade as both CAR homologs had been previously reported to reduce various medium- and long-chain fatty acids. The ω-TA from *Arthrobacter* sp. (ATA-117)[102] was chosen for the transamination step due to

previous studies suggesting that this enzyme could act on diketone substrates similar to the keto acid substrates intended for investigation in this work.[46] The second ω-TA homolog identified for use in the proposed cascade was a heavily mutated variant of ATA-117, SitATA, that was originally engineered for the production of the bulky antidiabetic drug sitagliptin (Januvia®).[102] For the final step of the designed cascade reaction, two well-characterised enantiocomplementary IRED enzymes, (*R*)-IRED from *Streptomyces* sp. GF3587[107] and (*S*)-IRED from *Streptomyces* sp. GF3546,[105] were selected to provide access to both possible enantiomers of the target piperidine structures. Since some of the chosen biocatalysts (namely ATA-117 and (*R*)- and (*S*)-IRED) had previously been shown to accept similar substrates to those selected for use in this work, and since all chosen biocatalysts have been expressed successfully in *E. coli* whole cells prior to this work, an *in vivo* cascade approach was deemed a viable option.

The nucleotide sequences for each of the seven selected enzymes were used to design seven new monocistronic operons to allow the independent regulation of expression of each gene (Section 6.3.1.3). The seven operons were framed by the BioBrick prefix and suffix restriction sites (Section 1.3.1.2.2) to allow for the construction of several multi-gene plasmids using the BioBrick cloning strategy (Section 1.3.1.2.2), and several purification tags (polyHis, c-Myc, FLAG) were used for separate enzymes to aid downstream protein expression analysis. The seven operons were subsequently codon-optimised for expression in *E. coli* bacterial cells and synthesised by GeneArt (Life Technologies), ready for incorporation into the selected entry vector, pPB01 (**Figure 25**).[47]

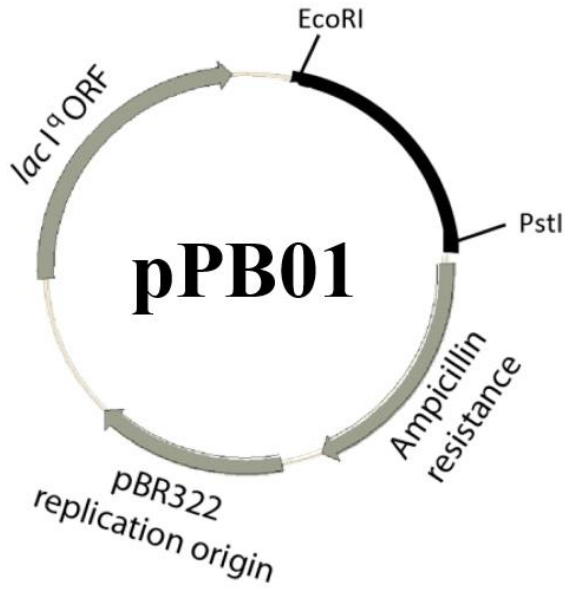


Figure 25 Entry vector, pPB01, devoid of regulatory elements to allow the individual regulation of each separate operon. The presence of the *Eco*RI and *Pst*I restriction sites allow for the introduction of BioBricks into the vector.

2.1.1 BioBrick Cloning Strategy

The envisaged cascade for the production of chiral piperidines from simple keto acid substrates required the co-expression of four different enzymes in a single host cell to allow for the provision and recycling of NADPH and ATP. The construction of a single expression plasmid containing all of the necessary cascade genes was seen as preferential over the use of multiple plasmids encoding the genes for each enzyme separately, so as to avoid the potential conflicts that can arise when several plasmids with different origins of replication (ori) are used in a single cell.

The BioBrick cloning strategy was selected for the modular and sequential addition of several genes into the selected entry vector using standard DNA ligation techniques (Section 1.3.1.2.2 and **Figure 26**).

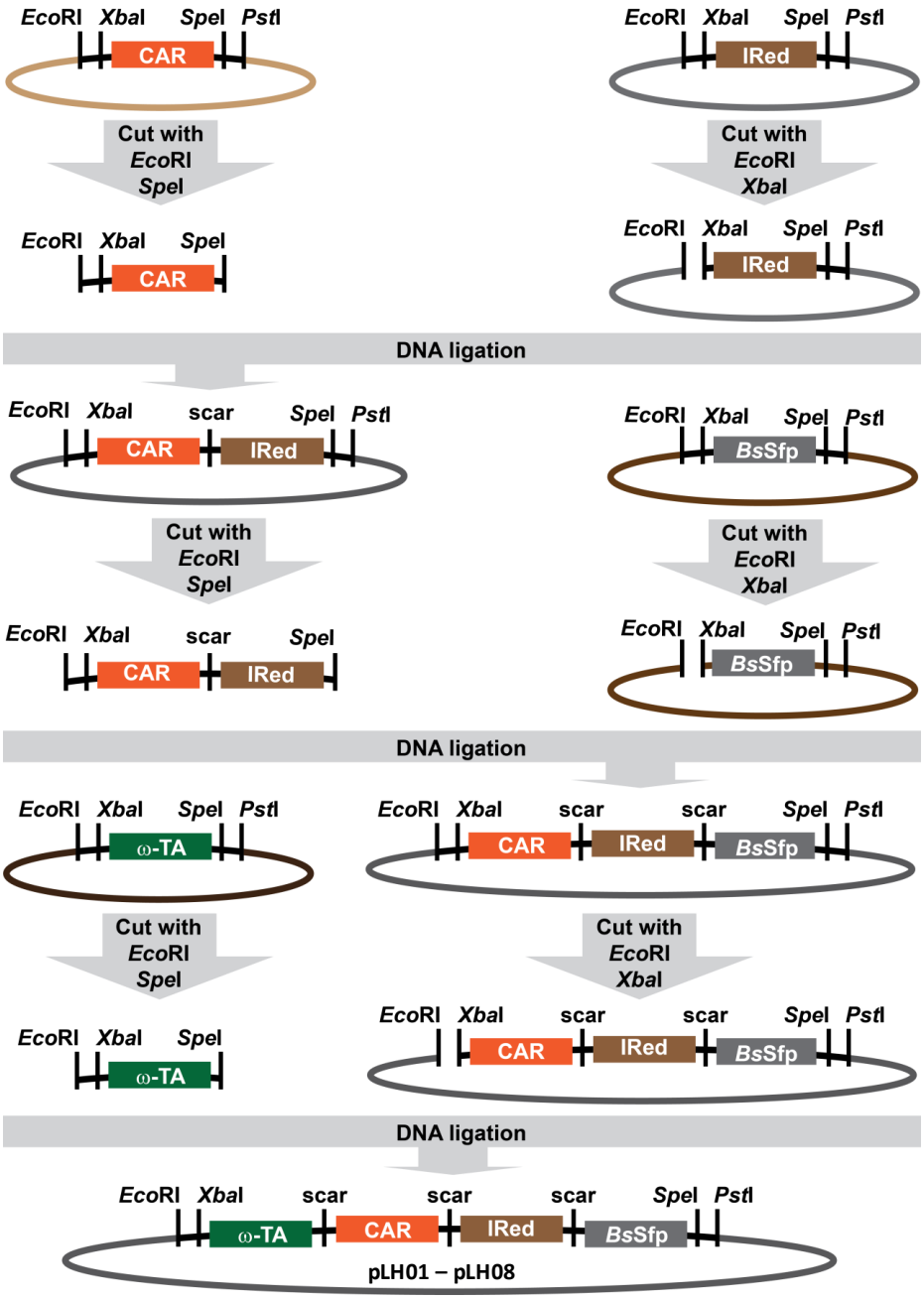


Figure 26 Subcloning of operons into pPB01 using the BioBrick cloning strategy to generate the final expression plasmids pPB01/ω-TA + CAR + IRED + *Bs*Sfp (pLH01 – pLH08). *Eco*RI and *Xba*I (BioBrick prefix) and *Spe*I and *Pst*I (BioBrick suffix) are restriction endonucleases that recognise and cleave specific palindromic DNA sequences. **Scar** denotes the mixed ligation site of *Spe*I and *Xba*I which is no longer recognised by any of these endonucleases.

This method was successfully used in the rapid combinatorial assembly of a library of eight initial expression constructs (pLH01 – pLH08) using the synthesised BioBricks for each of the seven enzymes chosen for this study (**Table 1**).

Construct	**Annotation**
pPB01/ATA-117 + MCAR + (*S*)-IRED +*Bs*Sfp	pLH01
pPB01/ATA-117 + MCAR + (*R*)-IRED +*Bs*Sfp	pLH02
pPB01/ATA-117 + NCAR + (*S*)-IRED +*Bs*Sfp	pLH03
pPB01/ATA-117 + NCAR + (*R*)-IRED +*Bs*Sfp	pLH04
pPB01/SitATA + MCAR + (*S*)-IRED +*Bs*Sfp	pLH05
pPB01/SitATA + MCAR + (*R*)-IRED +*Bs*Sfp	pLH06
pPB01/SitATA + NCAR + (*S*)-IRED +*Bs*Sfp	pLH07
pPB01/SitATA + NCAR + (*R*)-IRED +*Bs*Sfp	pLH08

Table 1 Library of constructs generated in first round of design-build-test-analyse (DBTA) cycle of cascade development.

2.1.2 Soluble Protein Expression in *E. coli*

The eight expression constructs assembled for the application of the envisaged cascade in a single whole cell were then individually expressed in *E. coli* BL21 (DE3) competent cells before undergoing a protein expression trial (e.g. temperature, time length, inducer concentration, cell culture density) to determine the optimal conditions required for the simultaneous production of all

four recombinant enzymes (Section 3.3.1). The expression cultures were then analysed for soluble protein expression via polyacrylamide gel electrophoresis (SDS-PAGE) analysis. Due to a high degree of expression for background proteins, western blot analysis was subsequently exploited to aid visibility of the cascade proteins of interest, an example of which is shown in **Figure 27**.

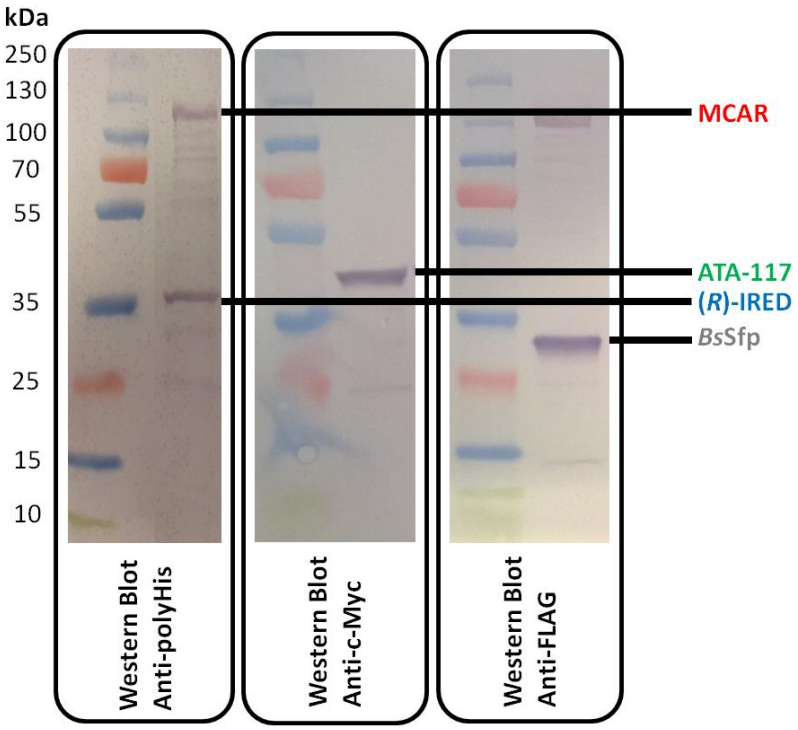


Figure 27 Western blot analysis of protein expression profile of *E. coli* BL21 (DE3) cells harbouring plasmid pPB01/ATA-117 + MCAR + (*R*)-IRED + *Bs*Sfp (pLH02) using Anti-polyHis, Anti-c-Myc IgG and Anti-FLAG antibodies. Analysis shown for the soluble fraction of cell lysate. Optimised expression conditions can be found in Section 6.3.2.1.

2.1.3 Gene Duplication

A second round of cloning was undertaken in order to generate two new expression plasmids; pLH09, a plasmid containing duplicate operons for ATA-117 alongside single copies of MCAR, (*R*)-IRED and *Bs*Sfp operons, and; pLH10, a plasmid containing duplicate operons for (*R*)-IRED alongside single copies of MCAR, ATA-117 and *Bs*Sfp operons (**Figure 28** and **Table 2**). The use of the BioBrick cloning strategy meant that these additional constructs could be assembled easily, since the BioBricks for each extra gene were readily available from the first round of cloning. These extra plasmids were designed in order to analyse the effect increasing the levels of soluble expression of two cascade proteins, ATA-117 and (*R*)-IRED, would have on the efficiency of the cascade reaction (Section 3.5).

This approach was also used to analyse the effect increasing the levels of soluble expression of (*S*)-IRED may have on the efficiency of the CAR-TA-(*S*)-IRED variant of the cascade, generating; pLH11, a plasmid containing duplicate operons for (*S*)-IRED alongside single copies of NCAR, ATA-117 and *Bs*Sfp, and; pLH12, a plasmid containing duplicate operons for (*S*)-IRED alongside single copies of MCAR, ATA-117 and *Bs*Sfp (**Figure 28** and **Table 2**).

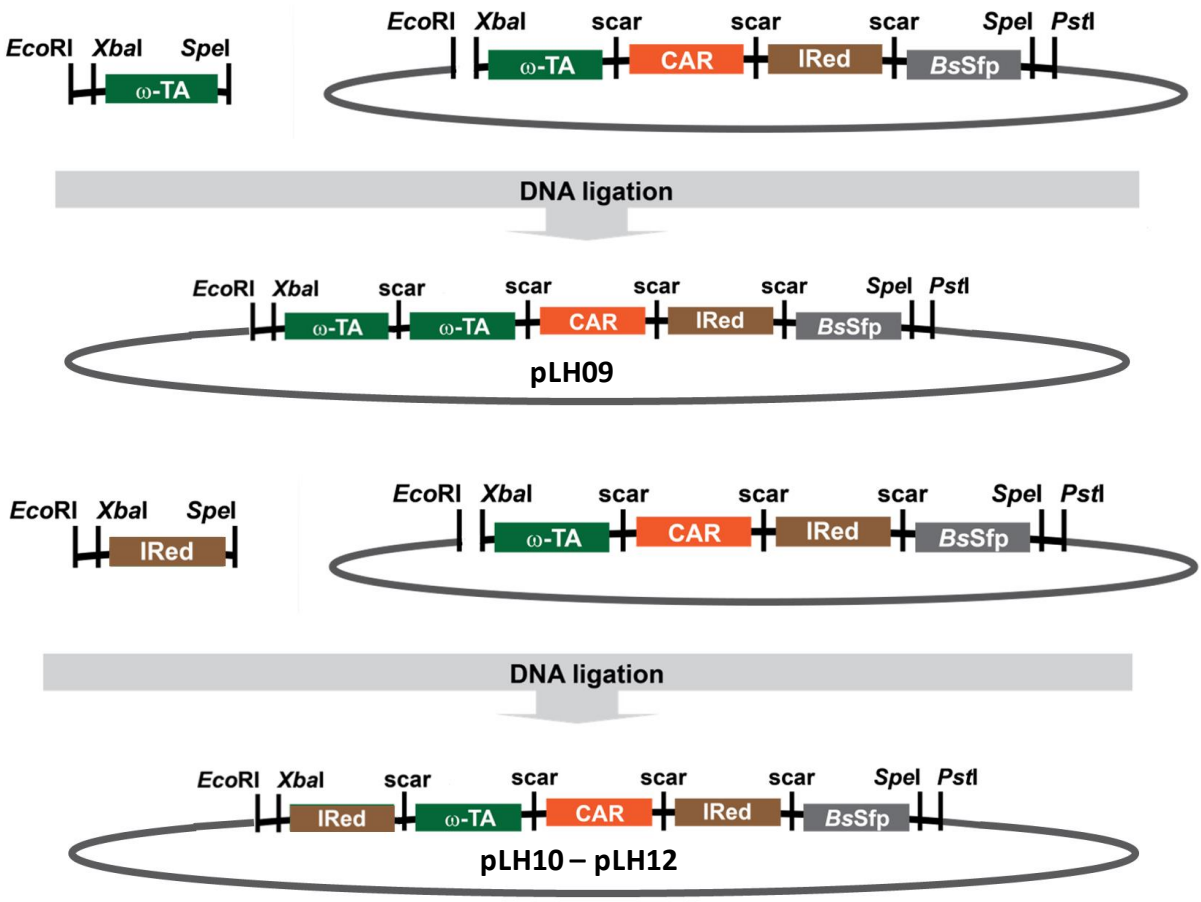


Figure 28 Subcloning of operons into pPB01/ATA-117 + MCAR + (*R*)-IRED + *Bs*Sfp using the BioBrick cloning strategy to generate the four new expression plasmids pPB01/ATA-117 + ATA117 + MCAR+ (*R*)-IRED + *Bs*Sfp (pLH09), pPB01/(*R*)-IRED + ATA117 + MCAR + (*R*)-IRED + *Bs*Sfp (pLH10), pPB01/(*S*)-IRED + ATA117 + NCAR + (*S*)-IRED + *Bs*Sfp (pLH11), pPB01/(*S*)-IRED + ATA117 + MCAR + (*S*)-IRED + *Bs*Sfp (pLH12). *Eco*RI and *Xba*I (BioBrick prefix) and *Spe*I and *Pst*I (BioBrick suffix) are restriction endonucleases that recognise and cleave specific palindromic DNA sequences. **Scar** denotes the mixed ligation site of *Spe*I and *Xba*I which is no longer recognised by any of these endonucleases.

Construct	Annotation
pPB01/ATA-117 + ATA-117 + MCAR + (*R*)-IRED +*Bs*Sfp	pLH09
pPB01/(*R*)-IRED + ATA-117 + MCAR + (*R*)-IRED +*Bs*Sfp	pLH10
pPB01/(*S*)-IRED + ATA-117 + NCAR + (*S*)-IRED +*Bs*Sfp	pLH11
pPB01/(*S*)-IRED + ATA-117 + MCAR + (*S*)-IRED +*Bs*Sfp	pLH12

Table 2 Library of constructs generated in second round of design-build-test-analyse (DBTA) cycle of cascade development.

2.1.4 Comparison of Expression Levels

Each of the four new expression constructs were then individually introduced into separate *E. coli* BL21 (DE3) bacterial cells prior to protein expression using the established conditions found during the initial expression trials for the first eight constructs. Western blot analysis of the soluble fractions for each of these new whole cell biocatalysts revealed that the simple addition of a duplicate gene did, in fact, correspond to increased levels of soluble protein expression for each enzyme (**Figure 29**).

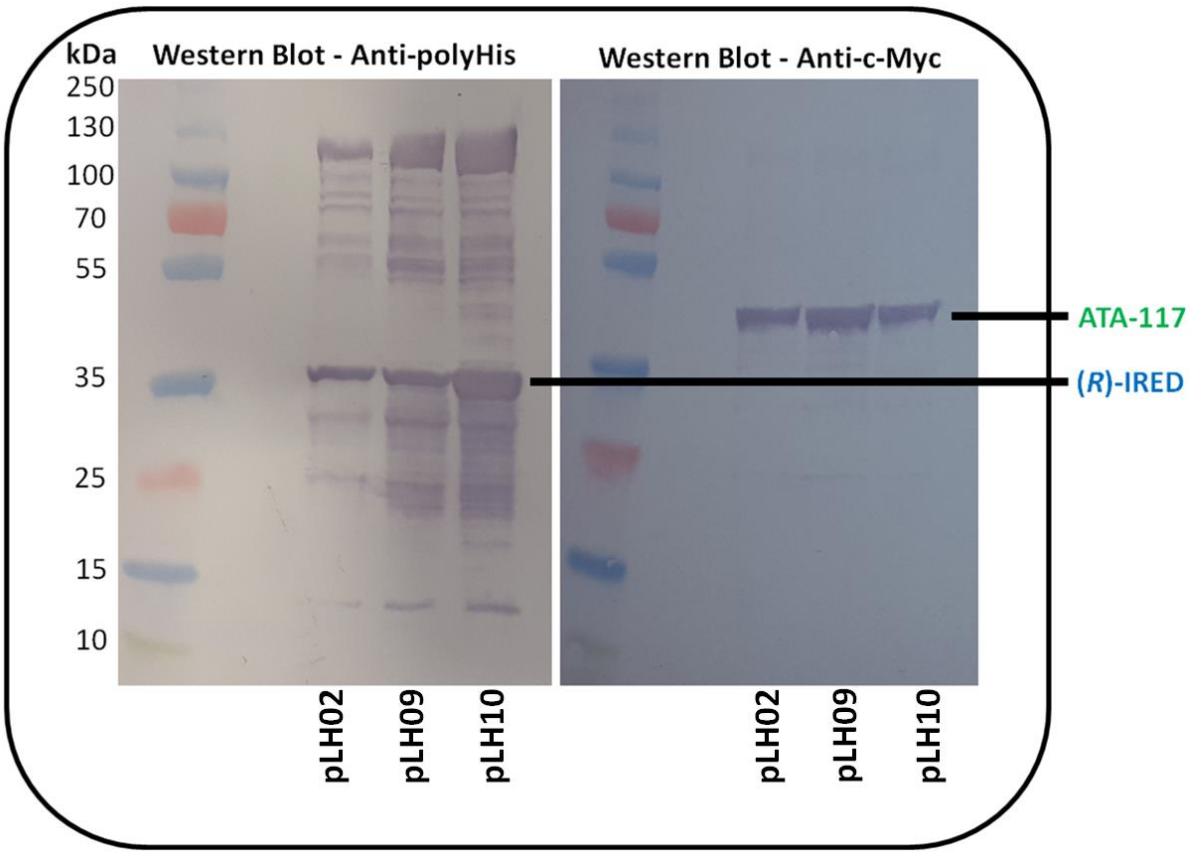


Figure 29 Western blot analysis of protein expression profile of *E. coli* BL21 (DE3) cells harbouring plasmid pPB01/ATA-117 + MCAR + (*R*)-IRED + *Bs*Sfp (pLH02), compared with pPB01/ATA-117 + ATA-117 + MCAR + (*R*)-IRED + *Bs*Sfp (pLH09) and pPB01/(*R*)-IRED + ATA-117 + MCAR + (*R*)-IRED + *Bs*Sfp (pLH10) using Anti-polyHis and Anti-c-Myc IgG antibodies. Increased expression of (*R*)-IRED is seen when pLH10 is used, and increased expression of ATA-117 is seen when pLH09 is used. Analysis shown for the soluble fraction of cell lysate.

2.2 Preparation of Biocatalysts for Carboxylic Acid Reductase-Imine Reductase (CAR-IRED), Alcohol Dehydrogenase-Imine Reductase (ADH-IRED) and Galactose Oxidase-Imine Reductase (GOase-IRED) Cascade Reactions

Biocatalytic retrosynthetic analysis of the drug target molecule 3-aminopiperidine led to the proposal of several enzymatic cascades towards its synthesis. The first cascade envisaged involved the reduction of the carboxylic

acid moiety of ornithine to its corresponding aldehyde by a CAR enzyme, followed by intramolecular cyclisation and subsequent reduction of the imine intermediate by an IRED enzyme (**Figure 30**). To screen the viability of this system, purified CAR homologs from *Mycobacterium smegmatis* (*ms*CAR),[88,89,91] *Tsukamurella paurometabola* (*tp*CAR),[88,89,130] *Nocardia iowensis* (*ni*CAR/NCAR),[40,92,129] *Segniliparus rotundus* (*se*CAR),[89,91,130] *Segniliparus rugosus* (*sr*CAR)[89] were selected to be trialled, alongside purified IRED/RedAm found in *Ajellomyces dermatitidis* (*Ad*RedAm).[110] The novelty of this cascade would favour the use of an *in vitro* approach, as it would first be necessary to assess whether or not ornithine was a substrate for the enzymes before attempting to build a more complex *in vivo* system.

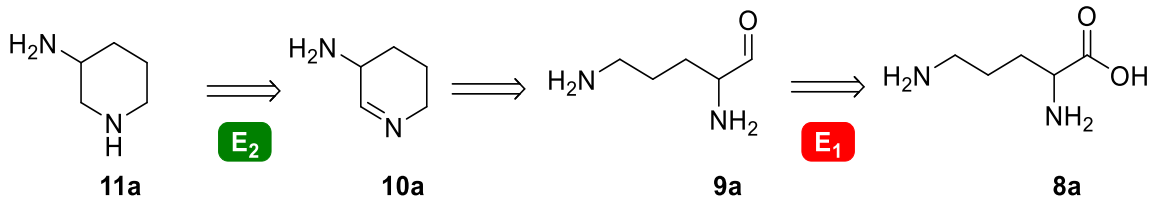


Figure 30 Biocatalytic retrosynthetic analysis of 3-aminopiperidine **11a**, suggesting an enzymatic route using ornithine **8a** as the substrate to the desired product through the use of E_1 (CAR) and E_2 (IRED) biocatalysts.

Next, a cascade composed of alcohol dehydrogenase (ADH) and IRED was put forward, namely the reduction of ornithinol to give the corresponding aldehyde, followed once again by imine reduction catalysed by *Ad*RedAm (**Figure 31**). The ADH enzymes chosen for this initial screen were commercial lysate preparations (ADH101, ADH104, ADH105, ADH112, ADH132, Johnson Matthey) and, as such, would require an *in vitro* methodology for their implementation. Following this, a third cascade reaction was designed with galactose oxidase (GOase) biocatalysts converting the alcohol to its corresponding aldehyde prior to the

imine reduction catalysed by *Ad*RedAm (**Figure 31**). GOase variants $M_{3\text{-}5}$[116,117,121] and F_2[117,119] were selected for this purpose as they have been shown to accept similar amino alcohol substrates in previous work.[117]

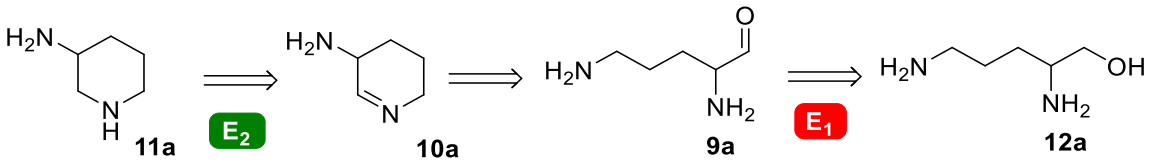


Figure 31 Biocatalytic retrosynthetic analysis of 3-aminopiperidine **11a**, suggesting an enzymatic route using ornithinol **12a** as the substrate to the desired product through the use of E_1 (ADH or GOase) and E_2 (IRED) biocatalysts.

2.2.1 Whole Cell, Purified Enzyme or Crude Cell Lysate Preparations of *Ad*RedAm

Due to the requirement of NADPH cofactor for the imine reduction catalysed by *Ad*RedAm, a whole cell application of this enzyme would be more cost effective than an *in vitro* mode of use. Whole cell reactions utilising IRED enzymes have been reported in the literature and in the following chapter,[40,105,106] and so an analysis of whole cell *Ad*RedAm used in conjunction with purified GOase in a hybrid cascade system (Section 1.2.3), was to be assessed once the cascade reaction had been shown to be functional (Section 4.4).

The use of a crude cell lysate of *Ad*RedAm was also to be assessed for activity in the proposed GOase-IRED cascade (Section 4.4.1.2), since this approach would negate the need for laborious protein isolation and purification. However, NADP(H) cofactor and a suitable exogenous regeneration system would still be required for a successful transformation to the target product.

2.3 Conclusions and Outlook

A variety of enzymatic cascade reactions were envisaged for the production of chiral cyclic amines. The execution of an enzymatic cascade in an *in vitro* approach would first require the isolation and purification of the necessary enzymes prior to use in biotransformations (where commercial enzyme preparations are not available), and would include the supplementation of exogenous cofactors where necessary. However, it can still be beneficial to first assess the viability of a *de novo* biocatalytic process using *in vitro* methodologies prior to attempting the same cascade reaction in an *in vivo* system, to avoid issues related to substrate transportation through cell membranes, for example. Once the success of a designed cascade sequence has been determined, it is then possible to assess its functionality in whole cell systems.

The rapid assembly of a library of multi-gene expression plasmids was facilitated by the employment of the BioBrick™ cloning strategy (Section 1.3.1.2.2). Enzyme candidates of interest for the conversion of carboxylic acid to aldehyde (CAR), aldehyde to amine (*ω*-TA), and reduction of a cyclic imine to the corresponding chiral amine (IRED) were identified, and subsequently converted into BioBrick™ parts to generate a library of compatible single gene operons. These parts were implemented in the combinatorial assembly of a library of cascade expression constructs, each hosting a different variation of enzyme homologs for the proposed CAR-TA-IRED cascade reaction (Chapter 3). The assembled expression constructs were transformed into the common bacterial expression strain *E. coli* BL21 (DE3), and protein expression trials resulted in soluble expression of each individual recombinant enzyme under the same conditions in a single host cell. Further optimisation of the expression constructs, through the duplication of genes suspected of being involved in reaction bottlenecks, was also achieved. This strategy could be expanded further through the generation of new

BioBrick™ parts, encoding alternative homologs of the desired enzyme classes, to investigate the CAR-TA-IRED cascade on a broader range of substrate compounds or with altered enantio- and/or diastereoselectivities.

The generation of BioBrick™ parts for the genes of interest, and ligation into the BioBrick™ compatible entry vector, pPB01, allowed for the efficient construction and optimisation of several cascade expression constructs, yet certain drawbacks to this approach are apparent. Firstly, once a BioBrick™ has been ligated to another, a scar sequence is formed and so it is no longer possible to separate the two parts. This is beneficial in that it effectively creates a new BioBrick™ part for combination with other parts, but it also means that it is not possible to quickly swap a gene out of the expression construct once it has been assembled. If, for example, a newly characterised homolog of a particular enzyme class is to be investigated in a designed cascade system, it would first be necessary to generate a new library of constructs incorporating this new homolog. It is also not currently known where the limit lies with respect to the size of plasmid that can be successfully transformed into bacterial cells. The successive addition of genes into an expression construct may eventually result in a plasmid that is too big to be incorporated into the host organism of choice, but this issue could be circumvented by the use of two or more compatible plasmids (Section 1.3.1.1), each containing multiple enzymes, for the development of longer enzymatic sequences.

3 Enzymatic Cascade Routes to Chiral Piperidines

3.1 Proposed CAR-TA-IRED Cascade for the Conversion of Keto Acids **1** into Chiral Piperidines and Pyrrolidines **5**

Chiral piperidine- and pyrrolidine-containing compounds **5** and their analogues were identified as suitable target product molecules due to their implication in a vast amount of clinical research,[131] and a facile enzymatic route to these scaffolds is of great interest due to the difficulties associated with their chemical synthesis (Section 1.4.1).

In order to design a successful enzymatic cascade towards their synthesis, the desired piperidine structures **5** were first disconnected through a biocatalytic retrosynthetic approach,[81] and it was indicated that a *de novo* enzymatic pathway consisting of carboxylic acid reductase (CAR, Section 1.6.1),[8,83,85,86,129] ω-transaminase (ω-TA, Section 1.6.2)[94,95,132–134] and imine reductase (IRED, Section 1.6.3)[103,105,106,109,111,135,136] enzymes (**Figure 32**) could be developed for the conversion of inexpensive achiral keto acids **1** to the corresponding chiral cyclic secondary amine **5**.[40] The broad applicability of these three enzyme classes for the transformation of a wide range of substrates has been outlined in the literature, thus it was anticipated that a cascade based upon their complementary chemistries could be feasible.

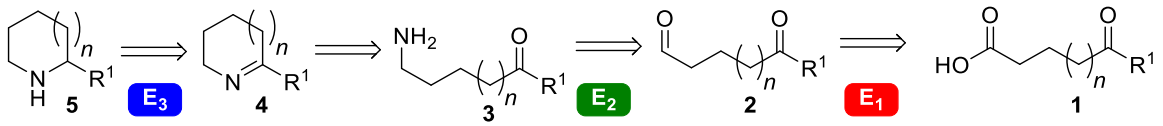


Figure 32 Biocatalytic retrosynthetic analysis of chiral piperidines, suggesting an enzymatic route from keto acid substrate to piperidine product through the use of E_1 (CAR), E_2 (TA) and E_3 (IRED) biocatalysts.

Initiation of the cascade system discussed in the following chapter was via the ATP- and NADPH-dependent reduction of a linear keto acid **1** to its corresponding keto aldehyde **2**, catalysed by the CAR found in *Mycobacterium marinum* (MCAR).[83] Prior to activation of the carboxylic acid, CAR is post-translationally modified with coenzyme A through the actions of a 4′-phosphopantetheinyl transferase from *Bacillus subtilis* (*Bs*Sfp).[90] Another NADPH-dependent enzyme, (*R*)-IRED from *Streptomyces fulvissimus* DSM 40593[107] catalyses the enantioselective conversion of the cyclic imine **4** to the chiral secondary amine end-product **5**. These chemistries were expected to be linked by way of a reductive amination of the intermediate keto aldehyde **2** by the *R*-selective *ω*-transaminase (*ω*-TA) ATA-117 from *Arthrobacter* sp.[102] using pyridoxal 5′-phosphate (PLP) as a cosubstrate and at the expense of a sacrificial amine donor, followed by spontaneous cyclisation of the acyclic amine product **3** to yield cyclic imine **4** (**Figure 33**).

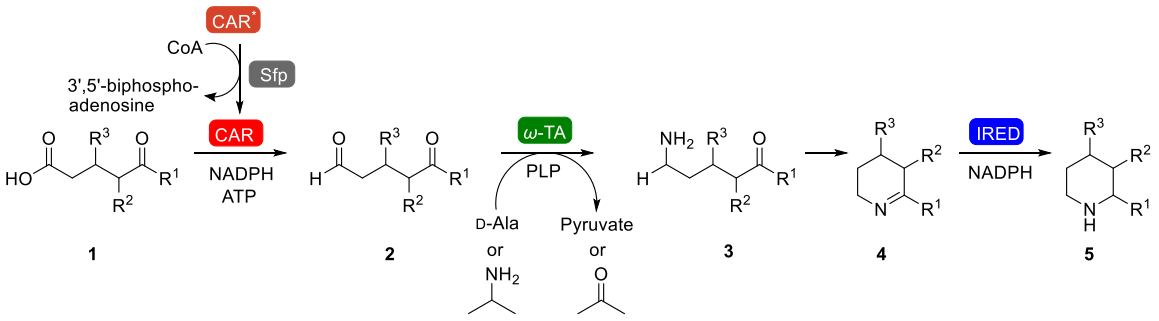


Figure 33 Cascade reaction envisaged for the conversion of keto acid **1** to chiral mono- or disubstituted piperidine **5**, including the posttranslational modification of the CAR biocatalyst by *Bs*Sfp using CoA.

For every successful round of the proposed cascade system, one molecule of ATP and two molecules of NADPH would be required for the conversion of keto acid **1** substrate to the final amine **5** product, in addition to PLP for the *ω*-TA reaction and coenzyme A for the posttranslational modification of CAR. Additionally, due

to the reversibility of the ω-TA reaction, it was predicted that the introduction of an external driving force to shift the equilibrium towards the production of amine **5** would be beneficial. Lactate dehydrogenase (LDH)-mediated removal of pyruvate is commonly used for this purpose.[96,137] The regeneration of cofactors for biocatalytic reactions executed *in vitro* is well-studied,[138,139] for example the use of a glucose dehydrogenase (GDH) system for NAD(P)H regeneration,[40,140] however it is often more practical and cost-effective to perform an enzymatic cascade *in vivo* and harness the inherent cofactor regeneration machinery within the host organism for this purpose. Opting to implement a biocatalytic pathway *in vivo* effectively offers a streamlined route for the production of chiral compounds, abolishing the need to produce and purify multiple proteins and bypassing the requirement for exogenous cofactor supplementation or recycling mechanisms (Section 1.2.2).

In the following chapter, the design and construction of a single plasmid system expressing a non-natural biosynthetic pathway *in vivo*, capable of converting simple keto acid **1** substrates into chiral mono- and disubstituted piperidines and pyrrolidines **5** at the expense of glucose and D/L-alanine, will be discussed. The development of the expression constructs used in this work is described in Chapter 2.

3.2 Initial Screen of CAR-TA-IRED Cascade Against a Library of Eight Expression Constructs

The eight single whole cell biocatalysts, each containing a different combination of genes encoding the potential cascade enzymes (Section 2.1), were prepared for use in biotransformations to assess their competency in converting two linear keto acid substrates, **1a** and **1b**, into the corresponding pyrrolidine and

piperidine structures (**Figure 34**). These two substrates were selected as appropriate model reactions to scrutinise the designed cascade because similar diketones were previously shown to be accepted by a cascade composed of ω-TA enzymes ATA-113 and ATA-117 and a monoamine oxidase (MAO-N) (Section 1.6.2).[46]

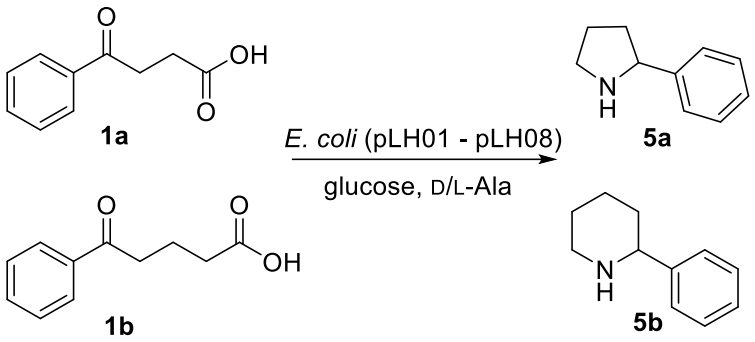


Figure 34 Proposed conversion of keto acids **1a** and **1b** into the corresponding pyrrolidine **5a** and piperidine **5b** through an *in vivo* CAR-TA-IRED cascade reaction.

Each reaction was performed in the presence of glucose (to aid cell growth and enable the cells to regenerate all of the essential cofactors needed for the reaction) and D/L-alanine (known to act as an amine donor for the transaminase step of the cascade process),[137] with all other reaction components envisaged to be provided by the living cell. The results of this initial screen of the whole cell biocatalysts are shown in **Table 3** and **Table 4**.

An analysis of the initial reaction screen revealed that a cascade composed of MCAR (and *Bs*Sfp), ATA-117 and (*R*)-IRED (denoted pLH02) was the most successful system when using keto acid **1b** as substrate to yield 2-phenylpiperidine **5b** (**Table 4**, Entry 2). It was seen that the same cascade system using NCAR instead of MCAR was slightly less effective, suggesting that the

MCAR homolog is more suited to these types of substrates than its NCAR counterpart. When keto acid **1a** was used, all of the constructs performed less effectively, which appears to be due to accumulation of the imine intermediate **4a**, suggesting that both (*S*)- and (*R*)-IRED work better on 6-membered rings in comparison with their 5-membered counterparts (**Table 3**).

Construct	Conversion to imine 4a/%	Conversion to keto alcohol 6a/%	Conversion to amine 5a (%)
pLH01	85	15	n.d.
pLH02	18	62	20
pLH03	88	12	n.d.
pLH04	60	40	n.d.
pLH05	n.d.	100	n.d.
pLH06	5	95	n.d.
pLH07	n.d.	100	n.d.
pLH08	n.d.	100	n.d.

Table 3 Percentage conversions to intermediate imine **4a**, keto alcohol **6a** side product and product amine **5a** for the reaction of 3-benzoylpropionic acid **1a** against all eight cascade expression constructs.

Construct	Conversion to imine 4b/%	Conversion to keto alcohol 6b/%	Conversion to amine 5b (%)
pLH01	11	89	n.d.
pLH02	n.d.	33	67
pLH03	58	35	7
pLH04	n.d.	46	54
pLH05	n.d.	100	n.d.
pLH06	n.d.	100	n.d.
pLH07	n.d.	100	n.d.
pLH08	n.d.	100	n.d.

Table 4 Percentage conversions to intermediate imine **4b**, keto alcohol **6b** side product and product amine **5b** for the reaction of 4-benzoylbutyric acid **1b** against all eight cascade expression constructs.

A comparison of the performances of constructs containing (*S*)-IRED against those containing (*R*)-IRED reveals that (*S*)-IRED is much less effective for the production of amines **5** in these whole cell systems, with much lower conversions occurring when plasmid pLH01 was used in place of pLH02, for example.

For both substrates used, only the cascades utilising ATA-117 for the transamination step were successful. No conversion to the product amines **5** was seen for any reaction using the SitATA variant, which is likely due to the fact that this enzyme was specifically engineered to accept a more bulky, very specific ketone substrate for its original use in the production of the antidiabetic compound sitagliptin (**Figure 35**).[102] The accumulation of keto alcohol **6** intermediate for cascade reactions employing SitATA suggests that, once the keto

aldehyde **2** intermediate is formed through the CAR reaction, it is intercepted by endogenous proteins within *E. coli*, resulting in further reduction taking place before transamination can occur. It is possible that mutating an enzyme so heavily to act on one specific substrate reduces the suitability of that enzyme for application with a broader range of substrates.

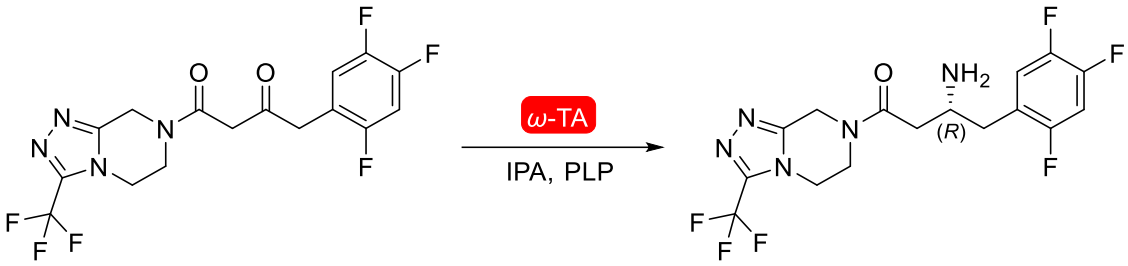


Figure 35 Manufacture of the antidiabetic compound sitagliptin through the employment of an engineered transaminase, using IPA as the amine donor.[102]

The initial screens of the eight cascade constructs against two potential cascade substrates led to the selection of a whole cell biocatalyst containing pLH02, using keto acid **1b** as the substrate, as the model system for the optimisation of the entire cascade process.

3.3 Process Optimisation for the CAR-TA-IRED Cascade

The screen of the initial library of eight constructs assembled in this study suggested that a cascade composed of MCAR, ATA-117 and (*R*)-IRED was the best combination of enzymes to convert the substrate keto acid **1** to its corresponding cyclic amine product **5**. A comparison between 3-benzoylpropionic acid **1a** and 4-benzoylbutyric acid **1b** revealed that the system

was better suited to accept **1b** as the substrate for this pathway, and so a whole cell biocatalyst harbouring construct pLH02 to convert **1b** was chosen as the foundation from which to optimise the proposed *in vivo* cascade system. Factors affecting the extent of soluble protein expression (Section 3.3.1), along with various reaction parameters (Section 3.3.2), were systematically analysed to find the most optimised process conditions, with conversions to 2-phenylpiperidine **5b** product quantified by comparison to a calibration curve using decane as an internal standard.

3.3.1 Optimisation of Protein Expression Conditions to Give Highest Conversion to Product Amine

The heterologous overexpression of a given protein in a specific host cell requires the optimisation of a variety of parameters in order to achieve sufficient levels of soluble and active enzyme. Factors such as expression temperature and time length, as well as the cell density of the microbial culture prior to the induction of protein expression using an optimised concentration of inducer all play a role in how well the recombinant protein is produced in the cell. Indeed, some recombinant proteins may also display toxicity when overexpressed in the standard overproduction strains of bacteria most commonly used in research, and so various specialised strains derived from *E. coli* BL21 (DE3) (e.g. *E. coli* pLysS, *E. coli* C41, *E. coli* C43) have been developed for the expression of particularly difficult proteins (see Chapter 5).

The overexpression of multiple heterologous enzymes within the same whole cell therefore requires compromised expression conditions to ensure that each protein is expressed to a sufficient level to allow the cascade reaction to be successful. The expression constructs developed in this study contained each

cascade enzyme under the same regulatory elements, and so the expression levels of the proteins were successfully controlled through induction with isopropyl β-D-1-thiogalactopyranoside (IPTG) and the optimisation of external expression conditions, rather than by the use of inherently stronger/weaker promoter and terminator pairs, for example.

3.3.1.1 Expression Temperature

An investigation into the effect that different expression temperatures for the cascade proteins, ranging from 20 – 37 °C, may have on the extent of final amine production revealed that lower expression temperatures (20 °C) are required for effective conversion to amine product (**Figure 36**). Higher temperatures, such as 30 °C, still yielded some active protein, but accumulation of the imine **4b** intermediate suggested that (*R*)-IRED is not expressed optimally once higher temperatures are reached. Neither amine **5b** product, imine **4b** intermediate, nor keto alcohol **6b** by-product were detected when protein expression was carried out at 37 °C, suggesting that such a high temperature is not appropriate for the expression of these CAR, TA and IRED proteins.

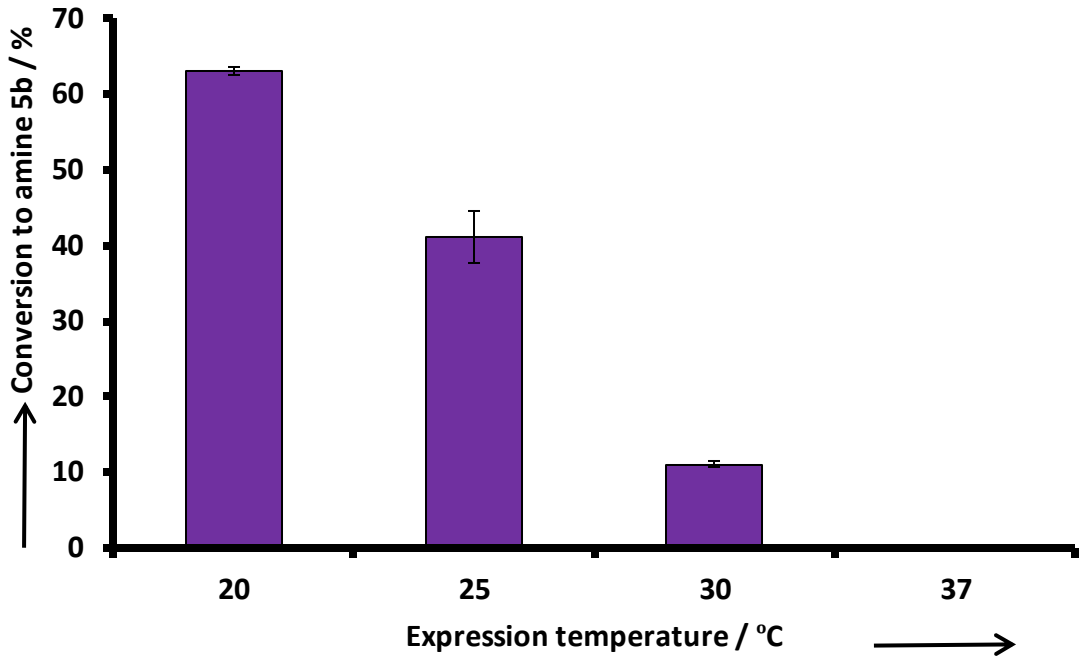


Figure 36 Effect of protein expression temperature on final amine production. *Expression conditions*: OD_{600} 0.6, 0.4 mM IPTG, 16 h expression, 250 rpm. *Reaction conditions*: 5 mM **1b**, 60 mg mL^{-1} whole cell catalyst, 50 mM glucose, 250 mM D/L-alanine, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.1.2 Optical Density

The optimum optical density (OD_{600}) at which protein expression is induced varies from protein to protein, and can have a significant effect on the success of producing soluble and active protein.[141] The effect of the OD_{600} of the *E. coli* BL21 (DE3) cells prior to the expression of the cascade proteins was probed, with OD_{600} values ranging from 0.4 to 1.3 tested. An OD_{600} value of 0.6 was seen to be optimal for the simultaneous expression of the cascade proteins as cells induced at this density resulted in the highest conversion to amine product (**Figure 37**). Higher optical densities still resulted in some conversion to the product but to a lesser extent, and this may be explained by the fact that higher cell densities

typically contain fewer cells that are in the exponential growth stage, meaning that more cells are viable at lower optical densities.

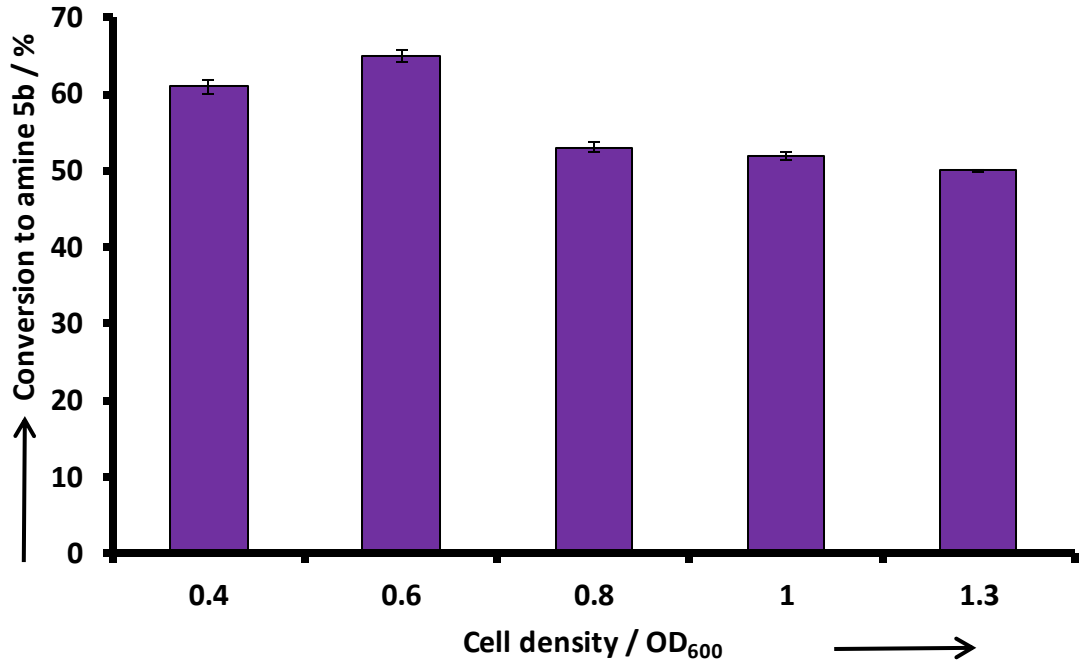


Figure 37 Effect of cell culture density on final amine production. *Expression conditions*: 0.4 mM IPTG, 16 h expression, 20 °C, 250 rpm. *Reaction conditions*: 5 mM **1b**, 60 mg mL^{-1} whole cell catalyst, 50 mM glucose, 250 mM D/L-alanine, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.1.3 Expression Time Length

Next, the effect of the time length of protein expression was investigated, with cell cultures being left to express protein for 2 h, 3.5 h or 16 h before the cells were harvested. It was seen that allowing the cascade proteins to express for only 2 h was not sufficient for optimal protein expression, and that protein expression needed to be carried out for at least 3.5 h to increase the efficiency of the biocatalyst and improve the conversion to product amine (**Figure 38**). It is

sometimes seen that allowing protein expression to continue for longer periods is detrimental to the soluble expression of recombinant proteins due to the potential problems related to toxicity to the host cell or the formation of insoluble protein and inclusion bodies, but leaving the cell culture to express protein for 16 h had no negative effect during this experiment.

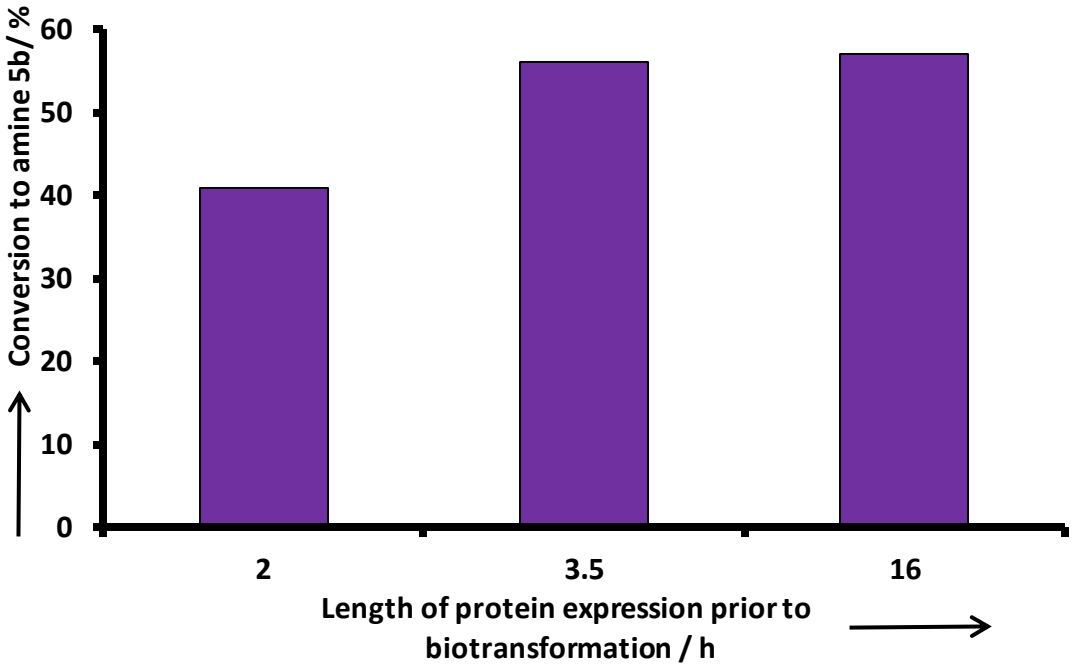


Figure 38 Effect of protein expression time length on final amine production. *Expression conditions*: OD_{600} 0.6, 0.4 mM IPTG, 20 °C, 250 rpm. *Reaction conditions*: 5 mM **1b**, 60 mg mL^{-1} whole cell catalyst, 50 mM glucose, 250 mM D/L-alanine, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.1.4 Inducer Concentration

Due to the fact that each cascade gene was put under the control of a *lac* operon, protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to cell cultures once the desired optical density was

reached. The concentration of the inducer can have an effect on the extent of protein expression achieved, and so varying amounts of IPTG were probed in the expression of the cascade proteins. Concentrations of 0.2 – 0.8 mM IPTG were investigated, with 0.8 mM seen to be the best concentration needed to produce the whole cell biocatalyst demonstrating the highest conversion to amine product (**Figure 39**).

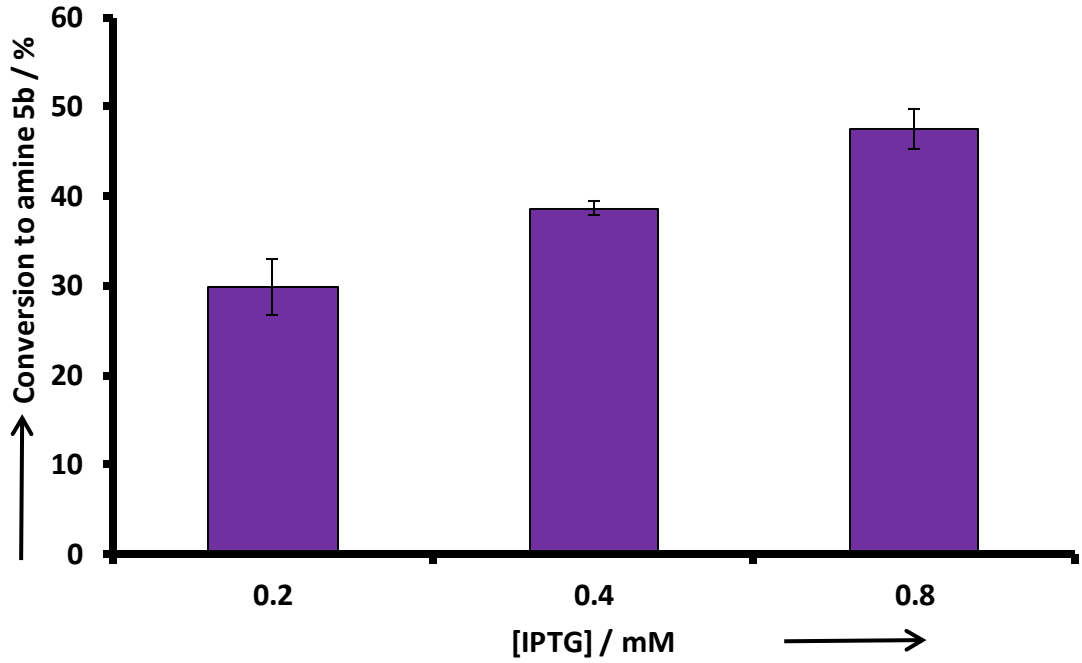


Figure 39 Effect of IPTG concentration used to induce protein expression on final amine production. *Expression conditions*: OD_{600} 0.6, 16 h expression, 20 °C, 250 rpm. *Reaction conditions*: 5 mM **1b**, 60 mg mL^{-1} whole cell catalyst, 50 mM glucose, 250 mM D/L-alanine, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.2 Optimisation of Reaction Parameters to Give Highest Conversion to Product Amine

The use of several enzymes in the same system to catalyse sequential chemical reactions introduces many possible parameters that could be screened for a given cascade process. Each step of the cascade may have different requirements with regards to substrate concentrations, reaction time lengths, and reaction temperatures, for example, and individual enzymes may have additional specific needs with respect to cofactor requirements or donor substrates.

It is therefore necessary to optimise a cascade reaction as a whole system rather than as its individual parts, since there is no guarantee that using several previously optimised enzymes together would work as effectively as when they were tested as isolated processes.

3.3.2.1 Cell Pellet/Supernatant Extraction

For a reaction to be as efficient as possible, as much of the desired product as possible must be able to be extracted from the system. The use of a whole cell biocatalyst can sometimes lead to issues with this extraction, as the reaction products may not be able to leave the cell efficiently.[36] It was therefore necessary to analyse both the supernatant of the reaction once it had been run to completion, as well as the whole cells containing the biocatalysts, to assess whereabouts the amine product was accumulating. Separate extractions of both the cell pellet and the reaction supernatant were carried out after the 24 h reaction period, and analysis of both fractions suggested that little amine product remained within the cell once it had been formed (around 4 %) (**Figure 40**). This result indicated that the desired product freely diffused out of the cells and into

the reaction buffer, leading to more efficient downstream processing than if the product remained contained within the whole cell after the reaction.

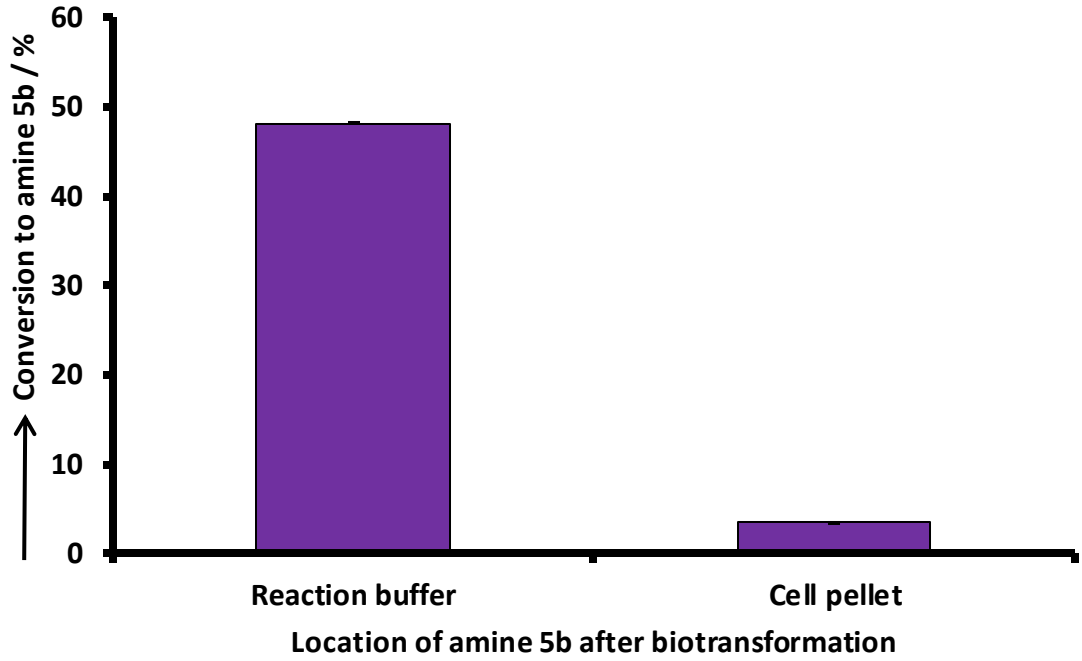


Figure 40 Extraction of reaction buffer and of cell pellet into ethyl acetate after biotransformation. *Expression conditions*: OD_{600} 0.6, 0.4 mM IPTG, 16 h expression, 20 °C, 250 rpm. *Reaction conditions*: 5 mM **1b**, 60 mg mL^{-1} whole cell catalyst, 50 mM glucose, 250 mM D/L-alanine, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.2.2 Choice of Sacrificial Amine Donor

The transaminase used in the cascade reaction, ATA-117, is known to accept both D/L-alanine and isopropylamine (IPA) as the sacrificial amine donor needed for the amination reaction to take place.[47,137] Both of these amine donors offer advantages and disadvantages to their use: D/L-alanine is a naturally occurring amino acid and can be overproduced in living cells, but this may also mean that employing a transaminase in a whole cell reaction would lead to less of the D/L-

alanine being available to the enzyme if it is funnelled off into metabolic processes. IPA is an attractive amine donor for transaminase reactions as it is easily obtainable and the product of its use in transamination reactions is acetone, the volatility of which can drive transamination reactions towards completion. However, IPA has been shown to have adverse effects on some proteins and so may not be the ideal candidate when the process being developed is a cascade reaction composed of enzymes sensitive to its presence.[47] The two amine donors were assessed for their performance in the MCAR-ATA-117-(*R*)-IRED whole cell cascade developed, and both were seen to give comparable conversions to product amine (**Figure 41**).

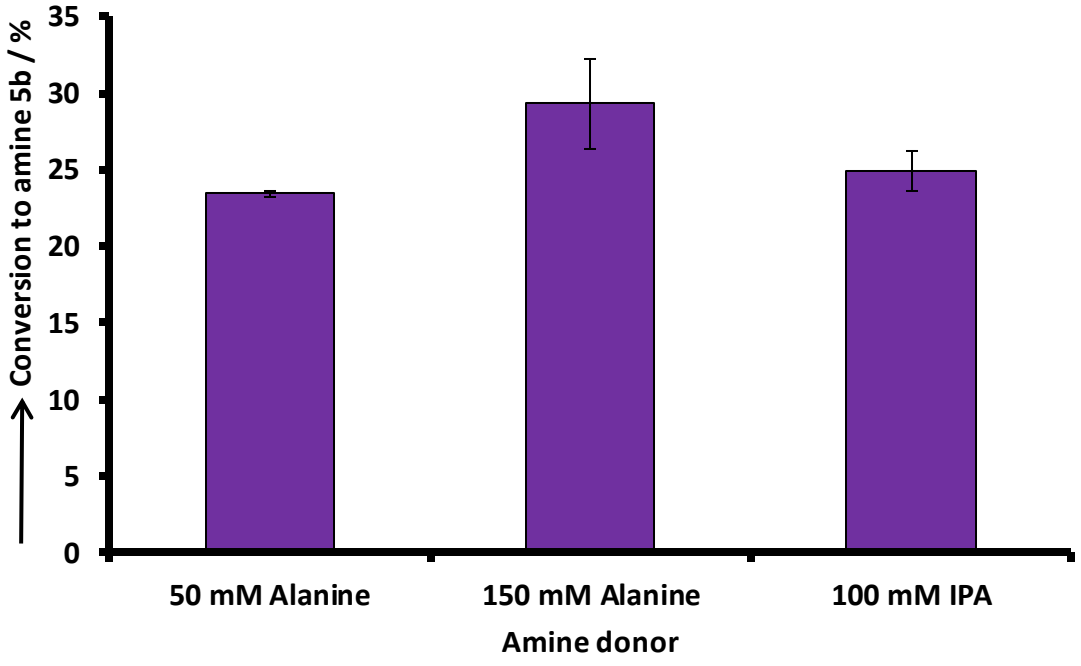


Figure 41 Comparison of D/L-alanine and IPA as amine donor for ATA-117. *Expression conditions*: OD_{600} 0.6, 0.4 mM IPTG, 16 h expression, 20 °C, 250 rpm. *Reaction conditions*: 5 mM **1b**, 60 mg mL^{-1} whole cell catalyst, 50 mM glucose, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.2.3 PLP Concentration

Pyridoxal 5′-phosphate (PLP) is an essential cosubstrate for the reaction catalysed by *ω*-transaminases (Section 1.6.2), and is also found endogenously within living *E. coli* cells.[142] An investigation into whether the amount of PLP provided by the living whole cell biocatalyst was sufficient for the cascade process to be successful was carried out, with concentrations ranging from 0 – 2 mM tested. This experiment indicated that the supplementation of additional PLP to the whole cells resulted in a decrease in the conversion to product amine when compared with biotransformations employed without the addition of exogenous PLP (**Figure 42**). The fact that conversions to product amine decreased as higher concentrations of exogenous PLP were added suggests that the coenzyme may inhibit one of more of the other enzymes comprising this cascade reaction.

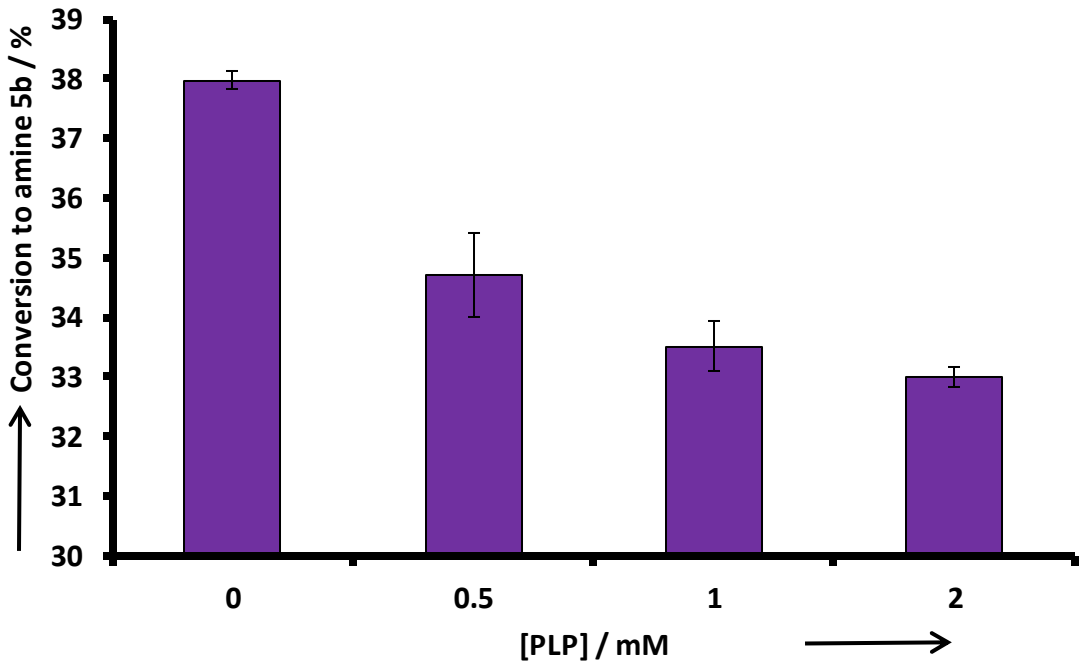


Figure 42 Effect of different concentrations of PLP ranging from 0 - 2 mM on final amine production. *Expression conditions*: OD_{600} 0.6, 0.4 mM IPTG, 16 h expression, 20 °C, 250 rpm. *Reaction conditions*: 5 mM **1b**, 60 mg mL^{-1} whole cell

catalyst, 50 mM glucose, 250 mM D/L-alanine, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.2.4 D/L Alanine Concentration

Despite the fact that both D/L-alanine and IPA were seen to be accepted amine donors for the transaminase included in this cascade reaction (Section 3.3.2.2), D/L-alanine was chosen as the donor to be used for all further experiments due to the fact that the product formed during its use, pyruvate, can be utilised by the living cell in metabolic processes. Therefore, the effect of D/L-alanine concentration on the conversion to product amine was analysed, ranging from 0 – 350 mM D/L-alanine. In general, an increase in the concentration of this amine donor led to an increase in the conversion to the amine product, peaking at 250 mM. If the whole cell biocatalyst was not supplemented with any D/L-alanine, some conversion to the amine product was still seen (around 7 %, presumably due to the transaminase using alanine already contained within the whole cell), but higher levels of D/L-alanine were needed to give optimum conversions to the desired product (**Figure 43**). The highest concentration of D/L-alanine provided to the cells, 350 mM, did not increase the conversion to the amine product further than that seen for a concentration of 250 mM. It is possible that this is due to insufficient removal of pyruvate at higher concentrations of D/L-alanine, or simply an excess of 250 mM is already sufficient for the reaction catalysed by ATA-117.

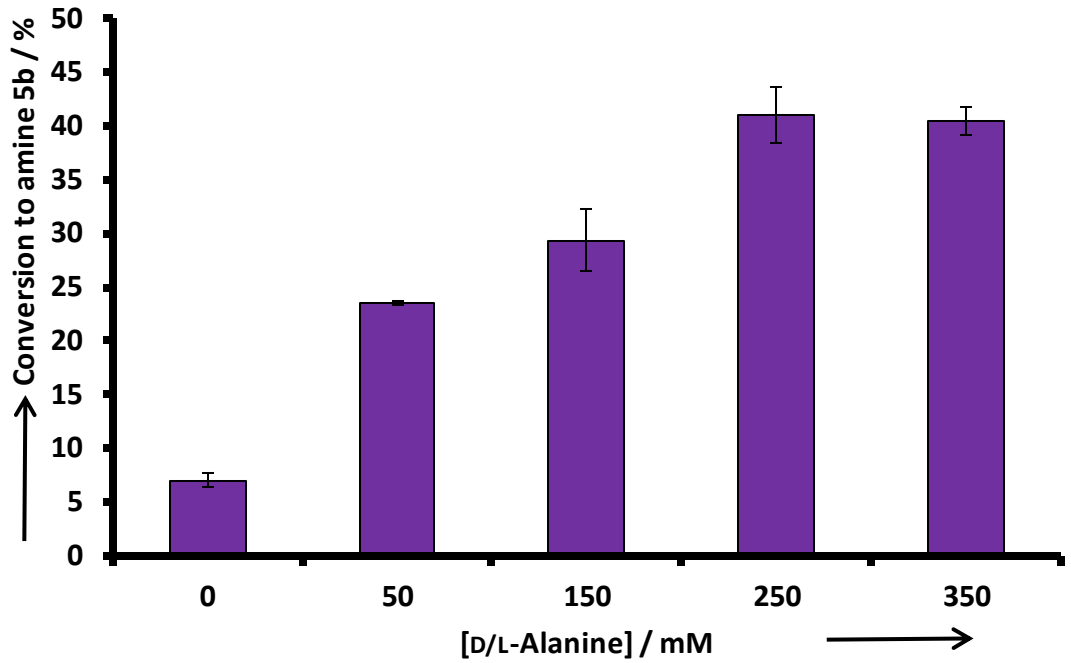


Figure 43 Effect of D/L-alanine concentration ranging from 0 - 350 mM on final amine production. *Expression conditions*: OD_{600} 0.6, 0.4 mM IPTG, 16 h expression, 20 °C, 250 rpm. *Reaction conditions*: 5 mM **1b**, 60 mg mL^{-1} whole cell catalyst, 50 mM glucose, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.2.5 Glucose Concentration

It was expected that, despite the presence of glucose within all respiring whole cells, additional amounts would need to be provided to the whole cell biocatalyst in order to relieve any metabolic burden on the cells and ensure the cascade reaction could proceed effectively. Varying concentrations of glucose, ranging from 0 – 100 mM, were supplemented to the reactions and the effect on conversion to product amine was analysed. It was seen that the provision of a higher concentration of glucose equated to a higher conversion to amine, peaking at 50 mM glucose (**Figure 44**). The absence of exogenous supplementation of glucose was still seen to result in some production of amine (around 12 %), yet the provision of 50 mM glucose lead to the most successful cascade reaction.

Increasing the supplementation to 100 mM glucose did not lead to any further increase in the conversion to product, suggesting that 50 mM is adequate to allow the cells to recycle all necessary cascade cofactors effectively.

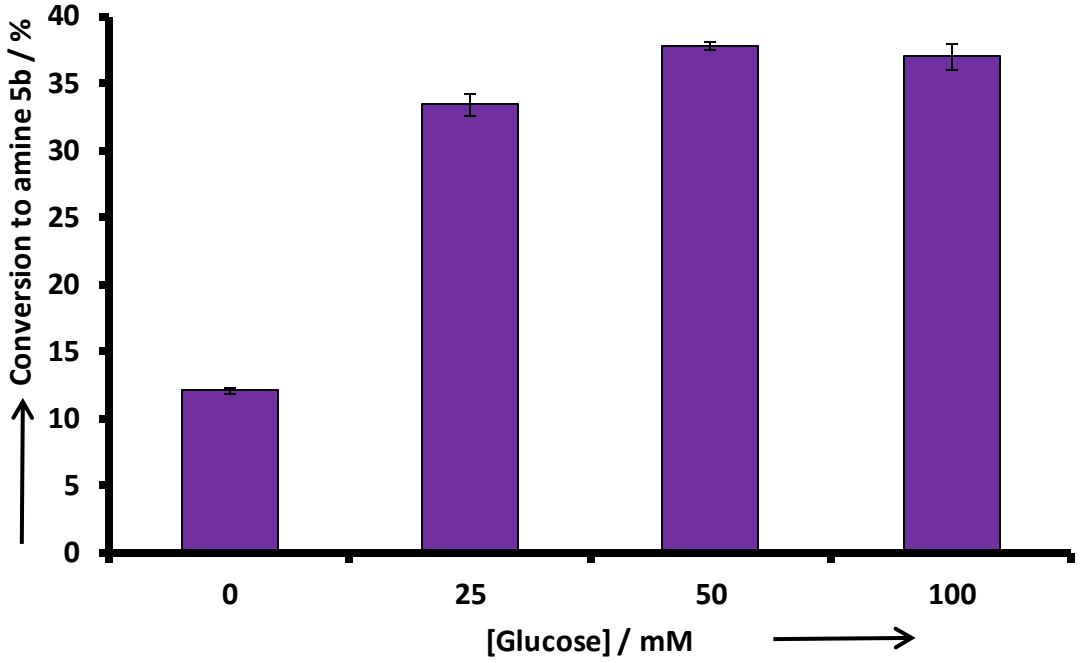


Figure 44 Effect of glucose concentration ranging from 0 - 100 mM on final amine production. *Expression conditions*: OD_{600} 0.6, 0.4 mM IPTG, 16 h expression, 20 °C, 250 rpm. *Reaction conditions*: 5 mM **1b**, 60 mg mL^{-1} whole cell catalyst, 250 mM D/L-alanine, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.2.6 Effect of Substrate Concentration on Substrate Consumption

When developing any reaction it is necessary to probe the substrate concentration tolerances of the process, in order to achieve the most effective system possible. A range of concentrations of the keto acid substrate **1b** were investigated to assess its consumption by the first enzyme in the cascade process, MCAR. It was seen that, up to concentrations of 5 mM, all keto acid **1b** substrate was consumed by

the CAR enzyme after the 24 h reaction window. Concentrations higher than 5 mM resulted in incomplete consumption of the substrate for the same 24 h reaction length (**Figure 45**). Following this, reaction lengths of 48 h were tested to see if a longer time frame would result in further consumption of keto acid for the higher substrate concentrations, however an increase in reaction length was seen to only give rise to a negligible increase in consumption of the keto acid substrate. The fact that an increase in reaction time length did not lead to full conversion of higher substrate concentrations suggests that the whole cell system may lose some activity after the initial 24 h period, a problem that could possibly be overcome by the further addition of glucose at a later time point to keep the cells viable.

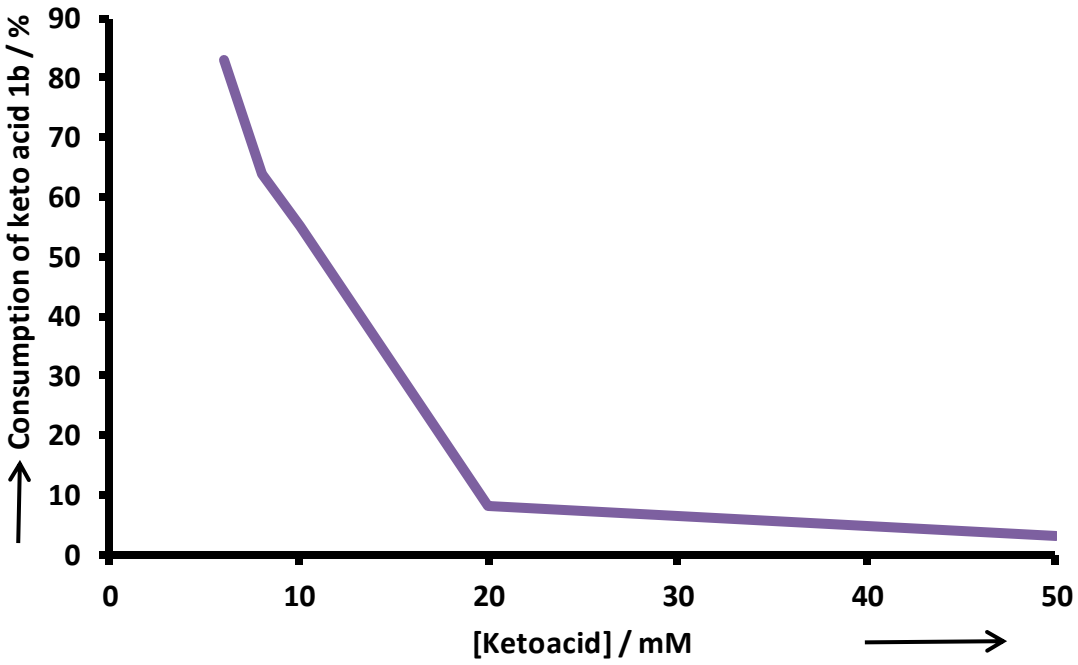


Figure 45 Effect of keto acid **1b** concentration ranging from 0 - 50 mM on substrate consumption. *Expression conditions*: OD_{600} 0.6, 0.4 mM IPTG, 16 h expression, 20 °C, 250 rpm. *Reaction conditions*: 60 mg mL^{-1} whole cell catalyst, 50 mM glucose, 250 mM D/L-alanine, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.2.7 Effect of Substrate Concentration on Product Formation

The range of substrate concentrations described in Section 3.3.2.6 was also used to assess the effect of keto acid **1b** substrate concentration on the efficiency of the entire cascade, with the conversion to product amine **5b** measured for each concentration tested. As can be seen in **Figure 46**, lower substrate concentrations result in the highest percentage conversions to the product **5b**. A substrate concentration of 1 – 3 mM led to comparable conversions to **5b**, but once the substrate concentration exceeds 3 mM, the production of **5b** drops drastically. Concentrations as high as 10 mM keto acid **1b** still gave around 20 % conversions to the desired product, but any value higher than this resulted in insignificant amounts of amine, with 50 mM substrate giving less than 1 % **5b**.

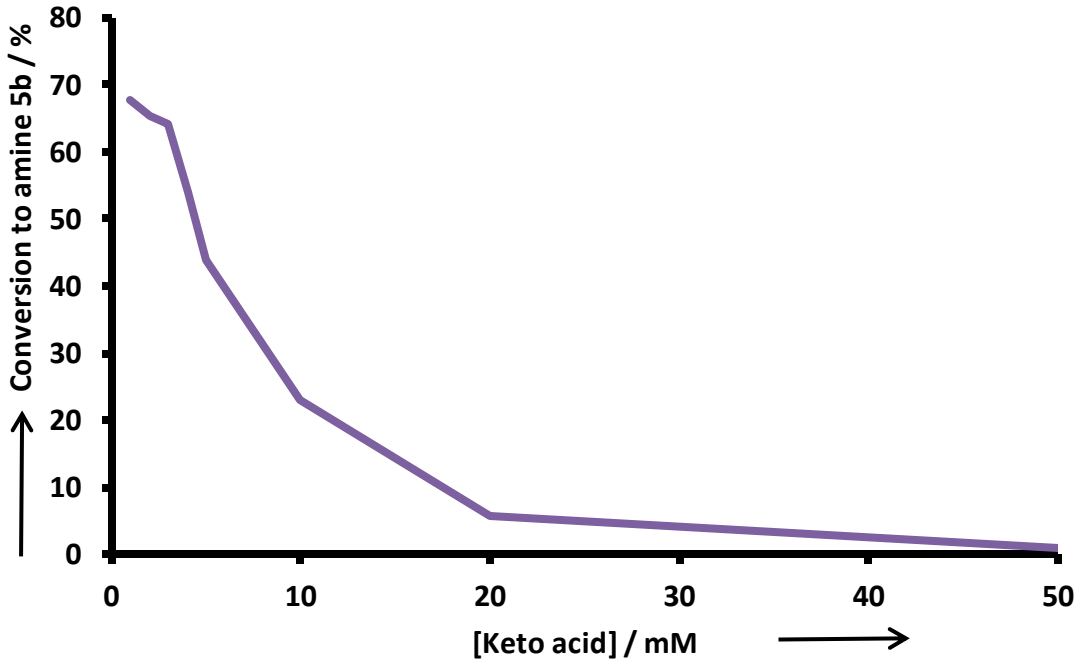


Figure 46 Effect of keto acid **1b** concentration ranging from 0 - 50 mM on final amine production. *Expression conditions*: OD_{600} 0.6, 0.4 mM IPTG, 16 h expression, 20 °C, 250 rpm. *Reaction conditions*: 60 mg mL^{-1} whole cell catalyst, 50 mM glucose, 250 mM D/L-alanine, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.2.8 Wet Cell Concentration

The amount of wet cell mass used in each reaction is inherently linked to the success of a cascade system, since the concentration of whole cells provided for a reaction is essentially the concentration of the biocatalysts present for the process to occur. A range of wet cell concentrations were assessed, ranging from 20 – 240 mg mL^{-1}, for the extent of product formation, and it was seen that a concentration of 40 mg mL^{-1} was the optimum amount of cells needed to maximise product output (**Figure 47**). The higher concentrations of cells tested (120 mg mL^{-1}, 240 mg mL^{-1}) still gave rise to some amine product **5b**, yet the apparent conversions seen were significantly less than those seen for the lower values tested. It is possible that this apparent drop in conversion is actually due to the fact that it becomes increasingly more difficult to successfully extract all of the amine produced when a higher mass of cells is present in the reaction.

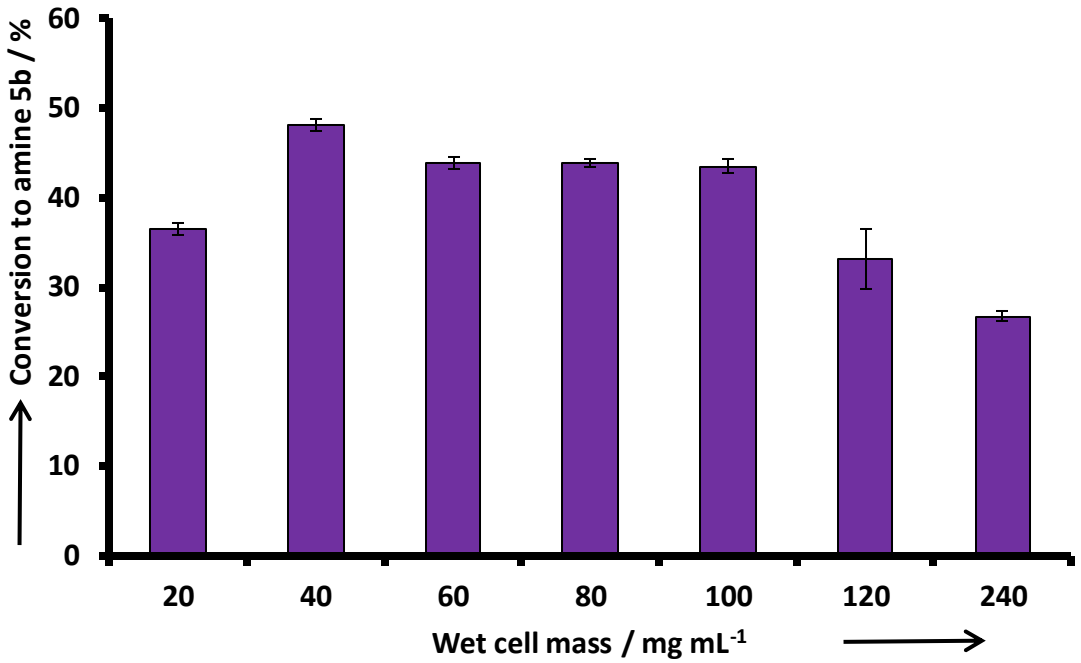


Figure 47 Effect of wet cell mass, ranging from 20 – 240 mg mL^{-1}, on final amine production. *Expression conditions*: OD_{600} 0.6, 0.4 mM IPTG, 16 h expression, 20 °C, 250 rpm. *Reaction conditions*: 5 mM **1b**, 50 mM glucose, 250 mM D/L-alanine, 500 mM NaPi pH 7, 24 h, 30 °C, 250 rpm.

3.3.2.9 Reaction Time Length

A time-point analysis was carried out in order to assess the stage at which the cascade reaction reached the highest conversion to amine product **5b**. A range of time-points were assessed for their conversion to the desired product, from 0.5 to 24 h, and it was seen that the cascade reaction proceeds steadily until it reaches a maximum conversion to amine **5b** after around 9 h (**Figure 48**). Analysis after 24 h and 48 h indicated that no further conversion to amine **5b** had taken place, suggesting that the biocatalyst reaches its maximum conversion to product during a 9 h time frame.

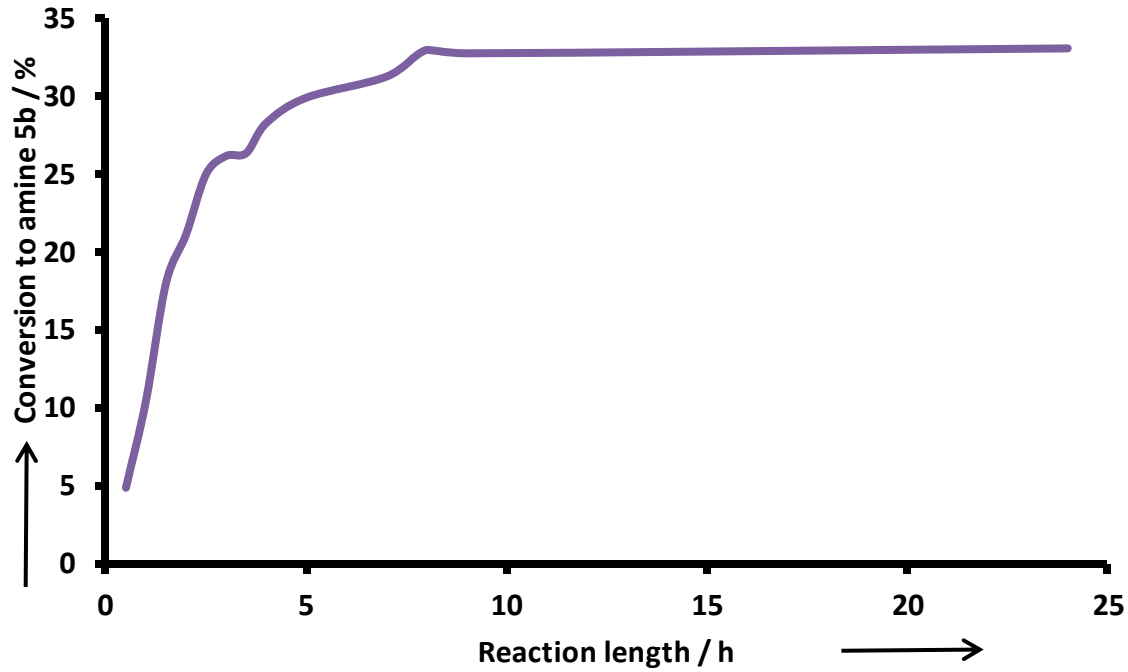
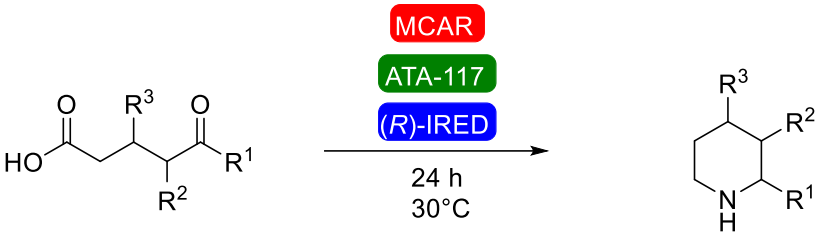


Figure 48 Effect of reaction time length, ranging from 0.5 - 24 h, on final amine production. *Expression conditions*: OD_{600} 0.6, 0.4 mM IPTG, 16 h expression, 20 °C, 250 rpm. *Reaction conditions*: 5 mM **1b**, 60 mg mL^{-1} whole cell catalyst, 50 mM glucose, 250 mM D/L-alanine, 500 mM NaPi pH 7, 30 °C, 250 rpm.

3.4 Additional Substrate Screen for the CAR-TA-IRED Cascade

Once the cascade system had been fine-tuned with respect to protein expression conditions, and various reaction components had been fully optimised to yield the highest conversions to product amine **5b**, the best conditions were applied for the transformation of four additional substrates to assess the applicability of the designed cascade for the synthesis of a range of chiral piperidine structures (**Table 5**).



1b-f

1b R^1 = Ph, R^2 = H, R^3 = H
1c R^1 = *p*FPh, R^2 = H, R^3 = H
1d R^1 = *p*-tolyl, R^2 = H, R^3 = H
1e R^1 = Ph, R^2 = Me, R^3 = H
1f R^1 = Ph, R^2 = H, R^3 = Me

5b-f

5b R^1 = Ph, R^2 = H, R^3 = H
5c R^1 = *p*FPh, R^2 = H, R^3 = H
5d R^1 = *p*-tolyl, R^2 = H, R^3 = H
5e R^1 = Ph, R^2 = Me, R^3 = H
5f R^1 = Ph, R^2 = H, R^3 = Me

Substrate	**Consumption of 1/%**	**Conv. to imine 4/%**	**Conv. to keto alcohol 6/%**	**Conv. to 5/%**	**5 ee/%** (abs. config.)	**5 de/%** (*cis* or *trans*)
1b	>99	-	-	47	30 (*S*)	-
1c	>99	3	47	50	90 (*S*)	-
1d	>99	51	42	7	93 (*S*)	-
1e	>99	n.d.	46	54	30 (*S*)	>98 (*cis*)
1f	>99	24	19	57	-	>98 (*cis*)

Table 5 Synthesis of chiral mono- and disubstituted piperidines through the use of a single whole cell biocatalyst expressing construct pLH02 using optimised expression and reaction conditions. Conversions based on imine:amine:keto alcohol product ratio, except for product **5b** (absolute conversion of which was calculated against a calibration curve using decane as an internal standard). n.d. not detected.

The ee and de values of the amine products **5b-f** observed correspond to selectivities already reported for the IRED-catalysed reduction of imines **4**.[40,106] Interestingly, the racemic substrate **1f** gives rise to racemic amine **5f** with very high *cis*-diastereoselectivity, indicating that both the CAR and TA biocatalysts can convert both possible enantiomers of the keto acid **1** substrate and keto

aldehyde **2** intermediate, respectively, during the course of the cascade reaction. The stereocentre of the imine intermediate **4f** is able to override the inherent selectivity of (*R*)-IRED, meaning that both enantiomers are able to be converted to give both *cis*-diastereomers of the product amine **5f**. It was also seen that racemisation of the pre-existing stereogenic centre on C-4 of keto acid **1e** occurred once the substrate had been added to the cascade, presumably due to imine-enamine tautomerisation (**Figure 49**).[40] This tautomerisation leads to the generation of a chiral imine intermediate which can then be selected for by the IRED, resulting in a dynamic kinetic resolution taking place. In addition to this, (*R*)-IRED has previously been reported to be sensitive to pre-existing chirality in cyclic imine substrates, potentially as a result of steric factors occurring between the incoming NADPH cofactor and any substituents on the substrate ring.[40]

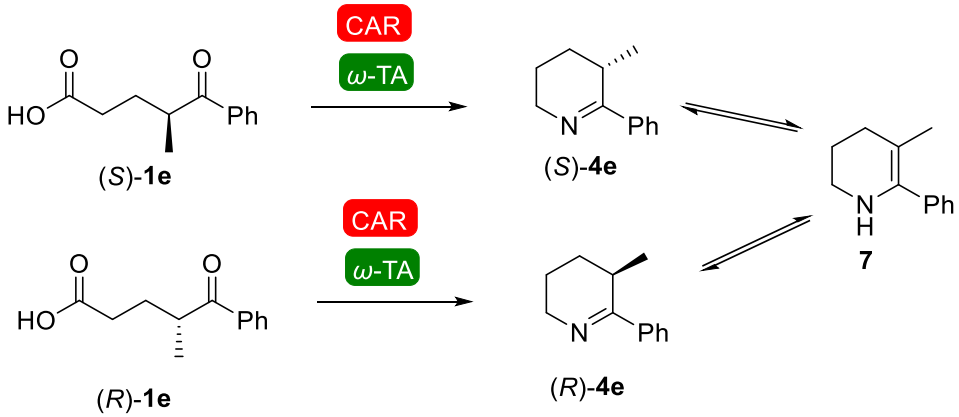


Figure 49 Racemisation of (*S*)- or (*R*)-**1e** by spontaneous imine-enamine tautomerisation.

The first step of the cascade, the conversion of the keto acid **1** substrate into a keto aldehyde **2** intermediate catalysed by CAR, was seen to be highly successful for all substrates investigated for the designed pathway, with no starting material seen after 24 h when 3 mM keto acid **1** was used. The poorest performing substrate of the panel analysed was in the synthesis of 2-(p-4-tolyl)piperidine **5d**,

with a very low conversion to the amine product (7%) seen. Significant amounts of both imine **4d** intermediate and keto alcohol **6d** by-product (51 % and 42 %, respectively) remained after the 24 h reaction period, suggesting that this substrate was not ideal for (*R*)-IRED or ATA-117. A similar scenario was seen with all other substrates tested: the highest conversion to product amine seen was 57 % for the disubstituted piperidine **5f**, with substantial amounts of imine **4f** and keto alcohol **6f** build up observed even for this best-performing reaction. The incomplete conversion of keto acid **1b-f** substrates to the desired end-product suggests that, despite the compounds being very good candidates for the first step of the cascade process, they may not be as well accepted by the subsequent enzymes of the designed reaction pathway.

The accumulation of a keto alcohol **6b-f** by-product for all the substrates tested for this cascade process suggests that, once CAR has catalysed the initial reduction of the keto acid **1** substrate into the corresponding keto aldehyde **2**, endogenous enzymes within the living host cell reduce the compound further to yield keto alcohol **6**. In recent years, the Prather group have constructed a strain of *E. coli* K-12 MG1655 that was engineered to display reduced aromatic aldehyde reduction (denoted RARE) through the deletion of six genes; three genes encoding for aldo-keto reductases and three genes encoding for alcohol dehydrogenases.[143] This engineered strain of *E. coli* has been shown to reduce the extent of the reduction of aldehydes to their corresponding alcohols,[87] and so expression construct pLH02 was transformed into competent *E. coli* K-12 MG1655 RARE cells in an attempt to minimise the build-up of keto alcohol **6** by-product seen for the CAR-TA-IRED cascade reaction described in this chapter. Unfortunately, biotransformations using *E. coli* K-12 MG1655 RARE expressing pLH02 resulted in a similar amount of keto alcohol **6** formation as that seen for *E. coli* BL21 (DE3) expressing pLH02, and so an alternative strategy to avoid the accumulation of this unwanted side product had to be conceived.

3.5 Duplication of Transaminase and Imine Reductase Genes in the Cascade Expression Constructs

Analysis of the GC traces for each substrate tested through the whole cell cascade system suggested that incomplete conversion to the amine **5** product was occurring. For each reaction, evidence of intermediate imine **4** accumulation, as well as keto alcohol **6** by-product, was apparent (**Figure 50**). The presence of unreacted imine **4** intermediate suggests that the piperideine structures formed through prior conversion of the selected keto acid **1** substrates are not ideal candidates for the imine reduction catalysed by the (*R*)-IRED biocatalyst[106] and thus this step represents a reaction bottleneck, whereas the formation of a keto alcohol **6** side product may occur due to the interception of the aldehyde **2** product of the CAR reaction by endogenous alcohol dehydrogenases (ADH) found within living *E. coli* cells.

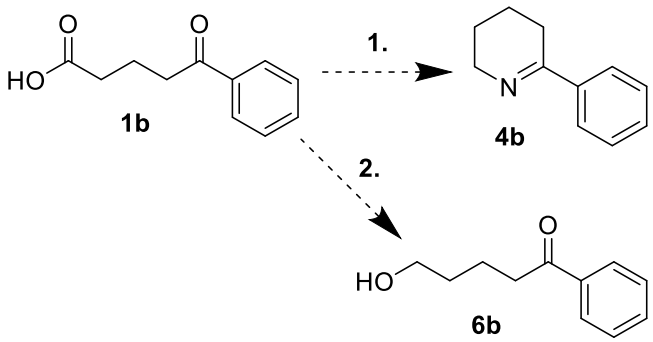


Figure 50 Accumulation of 1. imine intermediate **4b** due to insufficient consumption by (*R*)-IRED and 2. keto alcohol by-product **6b** due to interception of intermediate aldehyde **2b** by endogenous ADH enzymes within *E. coli*.

It was speculated that an increase in the soluble active expression of both (*R*)-IRED and ATA-117 enzymes within the whole cell biocatalyst could circumvent

the accumulation of these unwanted side products and, in turn, increase conversion to the product amine **5**, as an increased presence of IRED may be able to convert the intermediate imine **4** more rapidly, and an increased presence of ATA-117 may be able to utilise more of the aldehyde **2** intermediate before it can be converted into the unwanted keto alcohol **6** by-product. In particular, increasing the efficiency of the IRED step was anticipated to lead to the highest conversion overall, since pulling the imine **4** through to the final amine **5** product was expected to drive the earlier steps of the cascade reaction further to completion.

One way in which nature is known to augment the levels of a particular protein in response to various environmental stimuli is through the process of gene duplication (Section 1.3.2.3).[66,144,145] An increase in gene copy number for a target protein could potentially lead to an increase in its soluble expression, and so it was decided that this approach would be explored in relation to improving the efficiency of the designed cascade process. The fact that the BioBrick genes for both ATA-117 and (*R*)-IRED were readily available to us from the first round of the DBTA cycle (Section 2.1.1) lent itself to the rapid introduction of duplicate TA and IRED genes into the original best-performing plasmid pLH02, resulting in two new expression constructs pLH09 and pLH10 (**Figure 51** and **Table 2**).

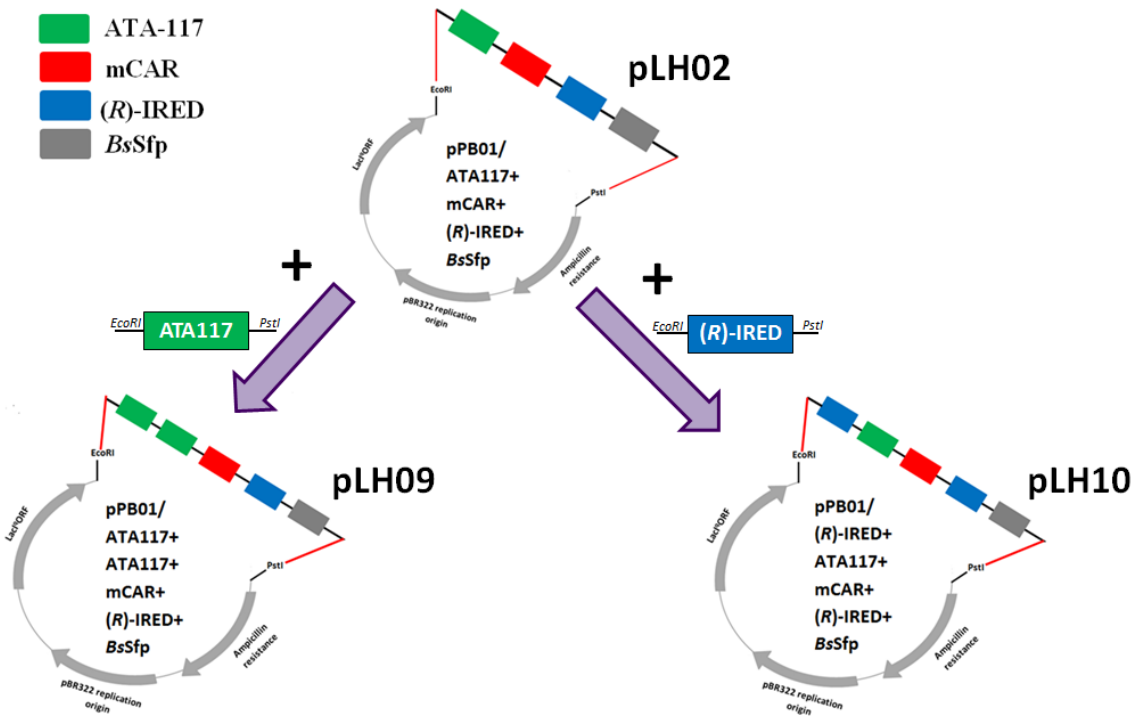


Figure 51 Introduction of duplicate ATA-117 and (*R*)-IRED BioBricks into construct pLH02 to give constructs pLH09 and pLH10.

The two new expression constructs were then transformed into competent *E. coli* cells before being used in whole cell biotransformations with the five chosen keto acid **1b-f** substrates, the results of which are shown in **Table 6**.

Substrate	Plasmid	Conv. to imine 4/%	Conv. to keto alcohol 6/%	Conv. to amine 5/%
1b	pLH02	-	-	47
1b	pLH09	-	-	61
1b	pLH10	-	-	67
1c	pLH02	3	47	50
1c	pLH09	n.d.	23	77
1c	pLH10	n.d.	13	87
1d	pLH02	51	42	7
1d	pLH09	67	24	9
1d	pLH10	21	12	67
1e	pLH02	n.d.	46	54
1e	pLH09	n.d.	25	75
1e	pLH10	11	5	84
1f	pLH02	24	19	57
1f	pLH09	21.	8	71
1f	pLH10	n.d.	7	93

Table 6 Effect of gene duplication on piperidine **5** production when optimised expression and reaction conditions were used. Conversions based on imine:amine:keto alcohol product ratio, except for product **5b** (absolute conversion of which was calculated against a calibration curve using decane as an internal standard). n.d. not detected.

As shown in **Table 6**, when increased expression levels for the bottleneck proteins ATA-117 and (*R*)-IRED were achieved, significant increases in

conversion to the desired amine **5** products were observed. A comparison between the percentage conversions seen when pLH02 was the cascade construct used confirm the expectation that increasing the expression of (*R*)-IRED in particular led to the best set of results. The most notable improvement seen for this set of substrates was in the production of 2-(*p*-tolyl)piperidine **5d**. Analysis of the data collected for this reaction using the initial whole cell biocatalyst expressing pLH02 revealed that imine **4d** accumulation was the major bottleneck for this substrate. The addition of a duplicate ATA-117 gene only increased the percentage conversion to the product amine **5d** from 7 % to 9 %, but the amount of imine **4d** detected in the samples increased from 51 % to 67 %. This increase in imine **4d** suggests that higher expression levels of ATA-117 did indeed lead to an augmented conversion of the intermediate aldehyde **2d** to the intermediate amine **3d**, reducing the amount of **2d** intercepted by endogenous proteins within the whole cell system (a fact that is also demonstrated by the decrease in the presence of keto alcohol **6d** by-product, going from 42 % to 24 % with the addition of an extra copy of the ATA-117 gene). Once the (*R*)-IRED gene was duplicated in the expression construct, however, the percentage conversion to **5d** increased from the initial 7 % to 67 %, with both intermediate imine **4d** and keto alcohol **6d** by-product falling by up to 46 % and 30 %, respectively.

The major reaction bottleneck for the cascade synthesis of 2-(4-fluorophenyl)piperidine **5c** was found to be the conversion of intermediate aldehyde **2c** into the aminoketone intermediate **3c**. When the reaction was catalysed by the initial whole cells expressing plasmid pLH02, conversion to product amine **5c** was 50 %, with 3 % imine **4c** intermediate and 47 % keto alcohol **6c** by-product accumulation. As expected, an increased expression level of ATA-117 in the system when using cells expressing pLH09 resulted in a significant decrease in keto alcohol **6c** production seen for the reaction (from 47 % to 23 %), and this also translated to higher conversions to **5c** (from 50 % to 77 %).

The use of a plasmid containing duplicate genes of (*R*)-IRED enhanced the efficiency of the reaction even further: the amount of keto alcohol **6c** seen dropped to 13 % and conversion to product amine **5c** reached 87 %, confirming that an improvement in the final reaction step catalysed by IRED results in increased efficiencies for earlier steps of the process.

When the disubstituted keto acid substrate **1e** was subjected to the cascade reaction catalysed by whole cells expressing the initial plasmid pLH02, the amount of keto alcohol **6e** side product produced for reactions utilising **1e** was 46 % suggesting that a significant amount of intermediate aldehyde **2e** was being siphoned off into the unwanted side process. The introduction of a duplicate ATA-117 gene reduced the amount of keto alcohol **6e** by-product seen for this substrate to 25 %, and the increased overexpression of (*R*)-IRED improved the efficiency of the process even further (with conversions to keto alcohol **6e** falling as low as 5 %). The highest conversion to piperidine product **5** was seen for the conversion of another disubstituted keto acid **1f** using whole cells expressing pLH10: conversions as high as 93 % were achieved when the duplicate (*R*)-IRED gene was introduced, increasing from an initial 57 %. Once again, the reaction was improved by the presence of higher levels of ATA-117 (from 57 % to 71 %), yet an additional (*R*)-IRED gene was seen to be the most beneficial for the cascade process.

Encouraged by the fact that gene duplication was shown to significantly increase the efficiency of the cascade process for the reaction composed of MCAR, ATA-117 and (*R*)-IRED enzymes, attention was then turned to those initial constructs containing (*S*)-IRED in place of (*R*)-IRED. In the initial screen of the eight expression constructs developed through the first round of the DBTA cycle (Section 3.2), pLH01 and pLH03 showed substantially lower conversions to the amine product **5** than their counterparts containing (*R*)-IRED in place of (*S*)-IRED (pLH02 and pLH04). Duplicate (*S*)-IRED genes were incorporated into pLH01

and pLH03 to yield a further two expression constructs pLH11 and pLH12 (**Table 2**), and competent *E. coli* cells were transformed with these additional plasmids. Biotransformations using these two new whole cell biocatalysts were run with substrate **1b** and, interestingly, percentage conversions to 2-phenylpiperidine **5b** were of comparable values to those achieved when whole cells expressing the "model" construct pLH02 were used. It is plausible that this restoration of activity for constructs containing duplicated levels of (*S*)-IRED is due to the fact (*S*)-IRED is a less efficient catalyst for the reduction of the imine intermediates screened than (*R*)-IRED,[105,106] and so a higher level of soluble protein is required to be present before significant conversions to the product amine **5** are observed. It is also possible that the expression of the (*S*)-IRED enzyme is negatively affected by the simultaneous expression of other recombinant proteins in the same host cell. To assess this premise, a whole cell biocatalyst expressing a plasmid containing just the (*S*)-IRED gene (pPB01/(*S*)-IRED) was developed and screened for the reduction of a known IRED substrate, 1-methyl-3,4-dihydroisoquinoline **4g**, to its corresponding amine **5g** (**Figure 52**).

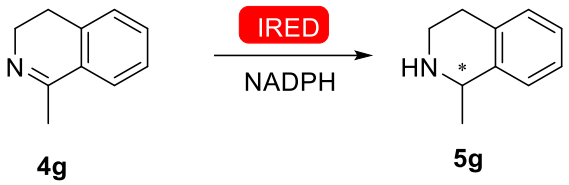


Figure 52 IRED-catalysed reduction of 1-methyl-3,4-dihydroisoquinoline **4g** to the corresponding tetrahydroisoquinoline **5g**.

The whole cell biocatalysts expressing plasmids pLH01 and pLH03 were also subjected to this analysis, and a comparison of the conversions to amine **5g** showed that the imine reduction was much worse when the multi-gene constructs were used. This result suggests that when the (*S*)-IRED biocatalyst is expressed in conjunction with other recombinant proteins in the same system,

significant metabolic burden hinders its adequate expression. For this reason, the introduction of a duplicate (*S*)-IRED gene into the expression constructs would lead to enhanced activity with respect to amine production due to an increase in the amount of soluble biocatalyst present in the cell.

3.6 Preparative-Scale Syntheses

In order to demonstrate the scalability of the designed and developed cascade process detailed in the previous sections, two keto acid substrates (**1b** and **1f**) were selected for preparative-scale reactions using the best plasmid system optimised through this work (a whole cell biocatalyst expressing pLH10). The optimised reaction conditions found during extensive process development (Section 3.3) were used in the synthesis of piperidine products (*S*)-**5b** and (±)-*cis*-**5f** , affording 70 mg of (*S*)-**5b** (58 % isolated yield, 30 % ee) and 50 mg of (±)-*cis*-**5f** (59 % isolated yield, >98 % de) and allowing successful isolation and characterisation of these cascade products (**Figure 53**). The optimised reaction conditions used for preparative scale syntheses can be found in Section 6.5.1.

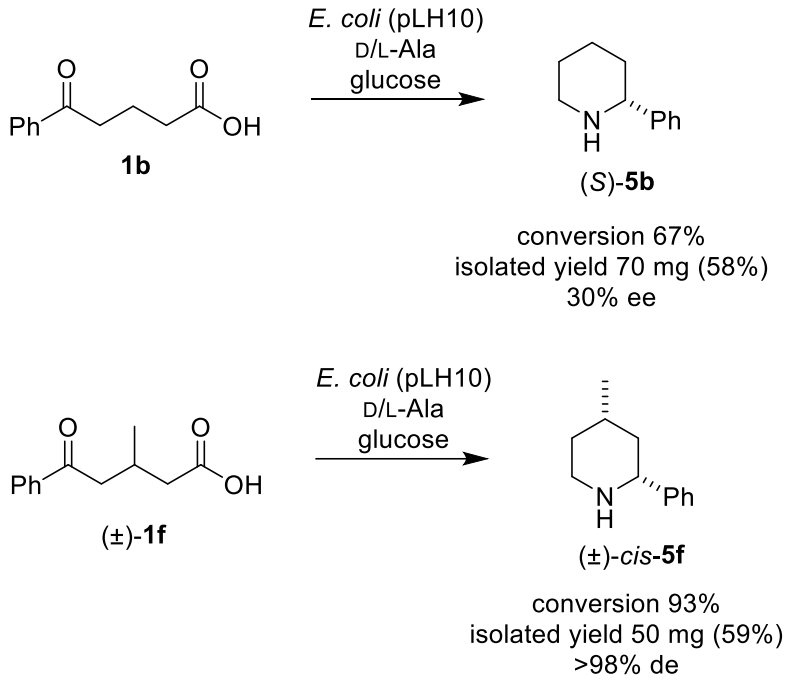


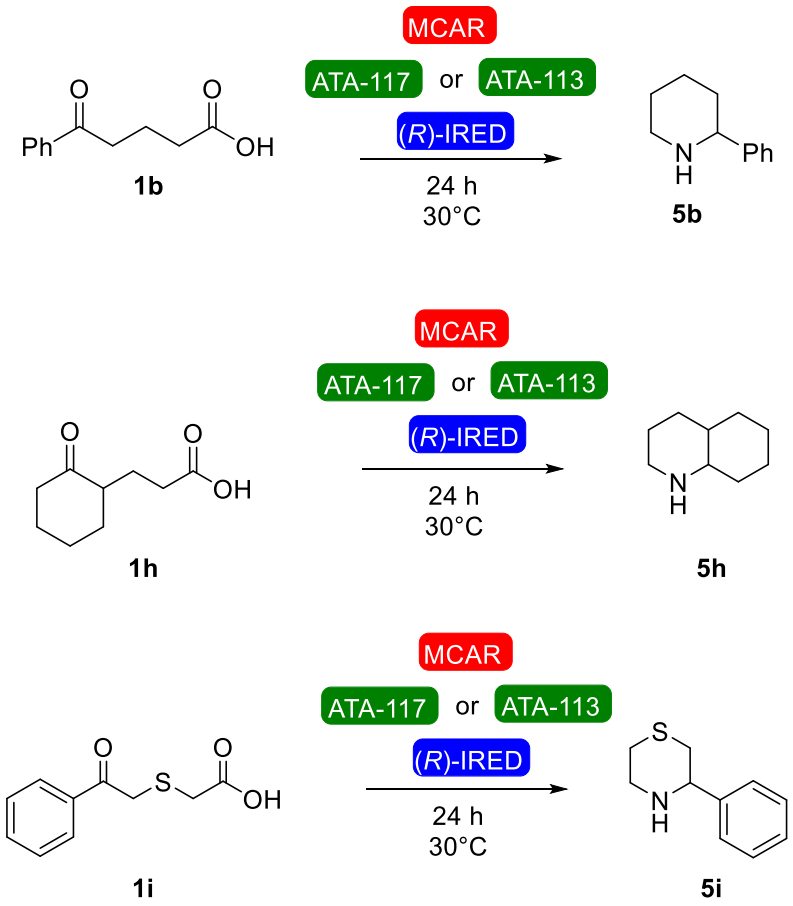
Figure 53 Isolated yields from preparative-scale biotransformations performed using *E. coli* whole cells harbouring pLH10.

The scaled up reactions were employed on 100 mg of substrate, and following the reactions using GC-MS suggested that the conversions to product amine were the same as those encountered for comparable analytical-scale reactions. The discrepancy between the conversions suggested through GC analysis and the isolated yields obtained may be explained by the fact that some material is inevitably lost during the extraction and purification of the product.

The success in executing these preparative-scale reactions demonstrates that the cascade process developed is a viable route to accessing chiral piperidines at scale. Employing the cascade reaction in a whole cell system is a more attractive alternative to employing the same cascade steps in an *in vitro* fashion, as no laborious protein purification is required and no expensive exogenous cofactors must be added for the cascade to proceed successfully.

3.7 Hybrid Cascade vs. In Vivo Cascade

The designed cascade process, comprised of CAR, *ω*-TA and IRED enzymes for the production of chiral piperidines, can also be implemented in a hybrid cascade mode (Section 1.2.3), as well as the *in vivo* system described thus far.[40] The application of CAR/*Bs*Sfp and IRED biocatalysts in separate whole host cells proffers some of the benefits of the *in vivo* application of enzymes (e.g. cofactor provision by the living cell, increased stability of the recombinant protein, no protein purification required), and also allows for individual expression conditions for each enzyme (in contrast to the compromised expression conditions needed to be reached when multiple recombinant proteins are being expressed in the same whole cell). Both ATA-117 and ATA-113 transaminases are available commercially as crude lysates provided by Codexis, and their *in vitro* application provides a simple way to tailor the amount of this biocatalyst present in the system. Biotransformations employing MCAR, (*R*)-IRED and either ATA-117 or ATA-113, both in hybrid reaction set ups as well as the *in vivo* reaction catalysed by whole cells expressing pLH02, were compared for the conversion of the model substrate **1b** as well as two new, more complex keto acid substrates **1h** and **1i** (**Table 7**).



Cascade System	Conv. to 5b/%	Conv. to 5h/%	Conv. to 5i/%
in vivo (ATA-117 + MCAR + (R)-IRED)[a]	30	14	8
Hybrid (ATA-117 + MCAR + (R)-IRED)[b]	26	13	8
Hybrid (ATA-113 + MCAR + (R)-IRED)[b]	58	30	30

Table 7 Synthesis of amines **5b**, **5h** and **5i** through the use of either an *in vivo* or a hybrid cascade system using 5 mM of keto acid **1** substrate. Absolute conversions to product amines **5** were calculated against calibration curves using decane as internal standard. [a]See Section 6.5.1 for reaction conditions; [b]See Section 6.5.2 for reaction conditions.

It was initially hypothesised that a single whole cell, expressing all of the enzymes needed for the cascade reaction to proceed, would be less efficient in the production of the desired amine **5** compounds when compared with the same

cascade process implemented in a hybrid fashion. This conclusion was based on the fact that it was not certain that all cofactor requirements could be efficiently met by the living cell, and that the levels of each biocatalyst present in the cell may not be sufficient for optimum conversions to the desired products. However, the results shown in **Table 7** demonstrate that the main contributor to the success of the reaction was the transaminase homolog used. When ATA-117 was used, the conversions to all three amine **5** products analysed were very similar, whether the reactions were executed in a hybrid system or in an *in vivo* system. The same reactions were also carried out using ATA-113 for the transaminase step, at which point the conversions to amine **5** product were significantly higher than those seen when ATA-117 was used. These results correspond with previous comparisons between the two homologs (Section 1.6.2.2),[46] and suggest that the cascade process could be improved by utilisation of ATA-113 in place of ATA-117 in the whole cell reactions described in this chapter, which is unfortunately not currently possible as the nucleotide sequence for ATA-113 is not available.

The execution of these cascade reactions in a hybrid fashion necessitates the employment of an external cofactor regeneration system (glucose dehydrogenase (GDH),[40,140] glucose) and the provision of a sufficient driving force for the reductive amination reaction (the removal of pyruvate through the actions of lactate dehydrogenase (LDH),[96,137] D/L-alanine as amine donor) due to the *in vitro* application of the transaminase biocatalyst, as well as the necessary addition of expensive cofactors (NAD^+, PLP). When the whole cell biocatalyst containing one of the expression constructs is used, however, the need for this exogenous supplementation is abolished as the living cell provides and regenerates all of the necessary cofactors (**Figure 54**). The relative ease of implementing these cascade processes using the *in vivo* system allows for the rapid screening of further, more complex cascade substrates, as demonstrated in **Table 7** for the production of

bicyclic amine **5h** and thiomorpholine **5i**. When it is confirmed that a new substrate is transformed through all of the cascade steps and gives the desired and expected products, the same process can be operated in a hybrid mode to allow the use of alternative transaminase enzymes, such as the cascades utilising ATA-113 detailed in this section, potentially leading to higher conversions to the product amine and higher substrate loading capabilities.

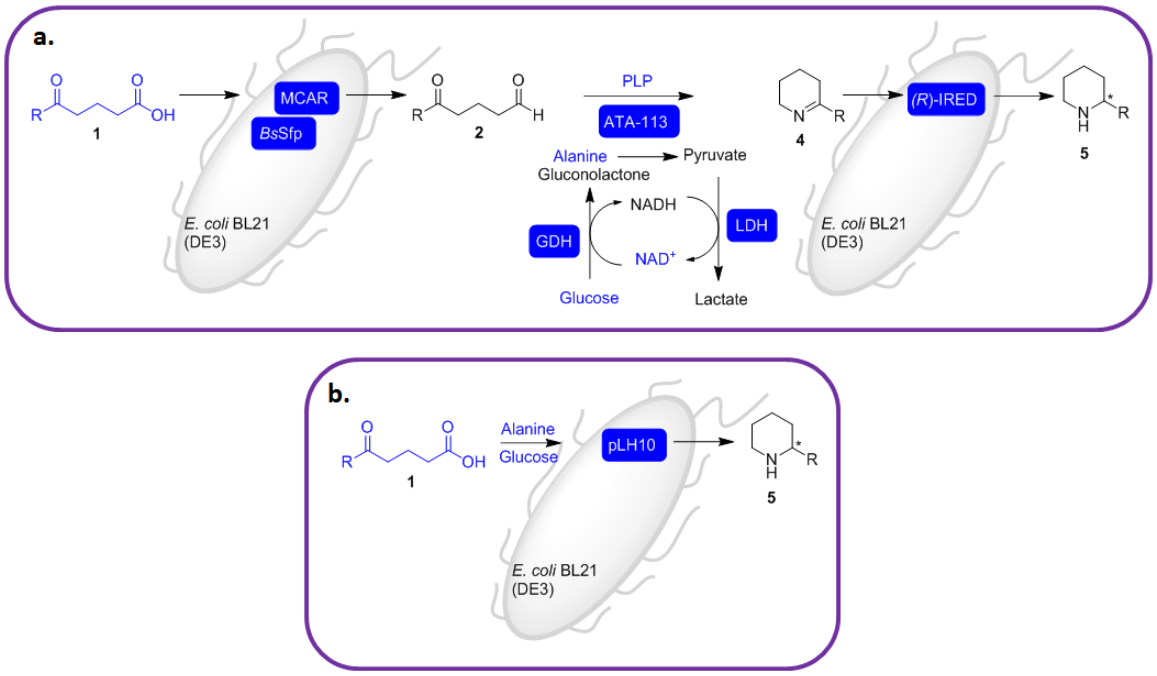


Figure 54 Comparison of requirements to successfully implement a. a multi-component one-pot hybrid CAR-TA-IRED cascade reaction and b. a single whole cell *in vivo* cascade. Items highlighted in blue represent the reaction components supplied to facilitate the cascade reaction.

3.8 Conclusions and Outlook

A design-build-test-analyse (DBTA, Section 1.5) approach to cascade design was used in the development of single cell enzymatic cascade reactions, for the conversion of simple linear keto acid substrates **1** into chiral mono- and

disubstituted pyrrolidines and piperidines **5**. A library of BioBrick™ DNA parts corresponding to homologs of CAR, Sfp, ω-TA, and IRED enzymes was built and used to assemble a library of multi-gene expression constructs for incorporation into *E. coli* BL21 (DE3) bacterial cells. An initial screen of the whole cell biocatalysts generated through this approach, against two selected keto acid substrates **1a**-**b**, suggested that a cascade reaction using MCAR and *Bs*Sfp, ATA-117 and (*R*)-IRED biocatalysts was an efficient route towards chiral 2-phenylpiperidine **5b**. Corresponding cascade reactions utilising the enantiocomplementary (*S*)-IRED did not result in as high conversions to the target products, and biotransformations for the production of 2-phenylpyrrolidine **5a** were less successful than those concerning the 6-membered ring counterpart **5b**. The most successful system, whole cells expressing novel plasmid pLH02 (pPB01/ATA-117 + MCAR + (*R*)-IRED + *Bs*Sfp) for the transformation of **1b** to **5b**, was subsequently optimised with respect to various biocatalyst expression and reaction parameters, including expression temperature and time length, inducer concentration, reaction time length, ω-transaminase amine partner choice and concentration, whole cell biocatalyst concentration and substrate concentration. Optimised reaction conditions were used for a screen of additional keto acid substrates **1c**-**f**, with conversions to product **5** of up to 57 % and selectivities of up to 90 % ee and > 98 % de. Analysis of the product profiles achieved through the cascade reactions revealed that significant accumulation of imine intermediate **4** and keto alcohol by-product **6** was occurring for all keto acid **1b**-**f** substrates investigated. The build-up of these compounds suggested that reaction bottlenecks involving the ω-TA and IRED biocatalysts were leading to lower conversions to the desired amine products **6**, and so a second round of the DBTA cycle was undertaken. Additional expression plasmids harbouring gene duplications of either ATA-117 or (*R*)-IRED genes were efficiently assembled through BioBrick™ cloning resulted in increased conversions to **6** of up to 93 %, and subsequently used for the rapid screening of

two additional keto acid substrates **1g**-**h**. This exploitation of a gene duplication approach, for the optimisation of expression levels of recombinant cascade proteins for a *de novo* enzymatic pathway, had not been reported prior to this work. Nevertheless, this methodology was shown to be a versatile and effective modular way of editing expression constructs for the tailoring of protein expression levels, and could be applied in the future for the construction of a broad range of cascades using BioBrick™ parts and compatible vectors.

This single whole cell cascade system was consequently compared against a hybrid *in vitro* / *in vivo* reaction design, and it was observed that the two reaction setups resulted in comparable conversions to product amine **5** when the same biocatalyst homologs were used. However, it was seen that increased substrate loadings could be tolerated when a hybrid approach was taken to cascade implementation,[40] suggesting that the *in vivo* system is less suited to industrial-scale processes at the current stage of its development. On the other hand, the containment of all cascade biocatalysts within a single host cell offered a streamlined execution of the cascade reactions, removing the need to supplement biotransformations with cofactors, cofactor recycling systems and *ω*-transaminase co-product removal systems, which would also be attractive for industrial applications.

The success of the whole cell CAR-TA-IRED cascade relies on high equivalents of a sacrificial amine donor, such as IPA or alanine, to drive the transaminase reaction towards amination of the keto aldehyde **2** intermediate. Despite the fact that the concentrations of alanine and IPA required for the described cascade system to function optimally were lower than those previously reported for other systems,[47,146] increased efficiencies could potentially be achieved through the co-expression of an alanine recycling enzyme, such as AlaDH,[53,73] to theoretically reduce the amount of D/L-alanine needed to be added to the reactions.

Since the start of this project, numerous examples of novel CAR,[88,89,91,130] TA[98] and IRED[110,112] homologs have been described in the literature, demonstrating different substrate preferences, reaction chemistries and/or enantio-/diastereoselectivities. Using the BioBrick™ approach detailed here, it will be possible to design and construct new cascade expression plasmids for the generation of additional whole cell biocatalysts capable of catalysing a broader range of cascade reactions.

4 Enzymatic Cascade Routes to 3-Aminopiperidine **11a**

4.1 Introduction

In the previous chapter, an enzymatic route toward the production of several chiral piperidines was established through the use of a CAR-TA-IRED cascade system. During the development of this work, we became interested in the synthesis of a particular target piperidine product, 3-aminopiperidine **11a**. Both enantiomers of this compound are implicated in the assembly of a vast amount of bioactive compounds, with (*R*)-3-aminopiperidine being a key intermediate in the production of potent and selective dipeptidyl peptidase IV (DPP-4) inhibitors for the treatment of type 2 diabetes such as linagliptin (Tradgenta®) and alogliptin (Vipidia®)(**Figure 55**),[147–149] and (*S*)-3-aminopiperidine being an important building block in the construction of anticancer agents such as checkpoint kinase (CK1) inhibitors (**Figure 55**).[150]

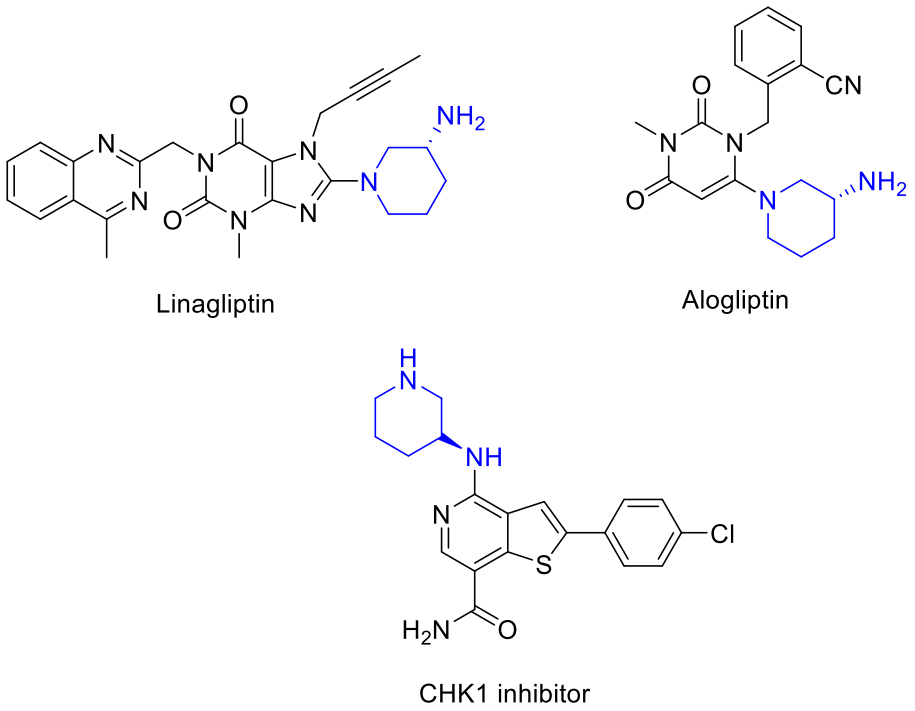


Figure 55 The 3-aminopiperidine **11a** motif is found in a wide range of bioactive compounds, for example DPP-4 inhibitors (linagliptin and alogliptin), and checkpoint kinase (CHK1) inhibitors.

There are many drawbacks to the chemical synthesis of 3-aminopiperidine **11a** (Section 1.4.1), and so currently there is scope for the development of an efficient and scalable route to its production. The aim of this project, then, was to exploit the ever-expanding biocatalytic toolbox to design and construct a *de novo* enzymatic pathway towards the sustainable synthesis of 3-aminopiperidine **11a**.

4.2 Proposed CAR-IRED Cascade for the Conversion of L-Ornithine **8a** to 3-Aminopiperidine **11a**

The first step in the development of a novel enzymatic cascade system is the biocatalytic retrosynthetic analysis of the target structure in order to assess the possible routes to its manufacture (Section 1.5.1). Disconnecting 3-

aminopiperidine **11a** through this approach suggested that the desired product could be accessed through the conversion of the nonessential amino acid ornithine **8a**, via a two-step reaction catalysed first by a carboxylic acid reductase (CAR) enzyme, followed by further reduction through the actions of an imine reductase (IRED) enzyme (**Figure 56**).

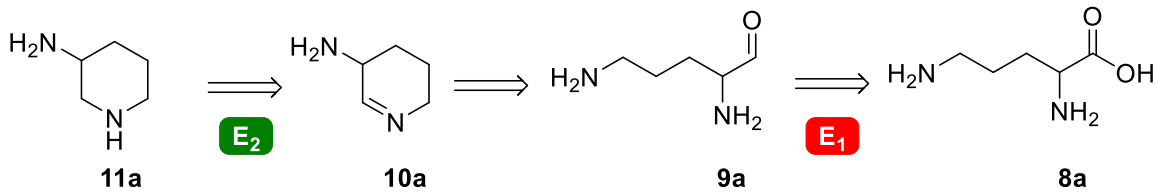


Figure 56 Biocatalytic retrosynthetic analysis of 3-aminopiperidine **11a**, suggesting an enzymatic route using ornithine **8a** as the substrate to the desired product through the use of E_1 (CAR) and E_2 (IRED) biocatalysts.

The use of L-ornithine **8a** as the starting material for the proposed cascade is particularly advantageous owing to its relative abundance and its central role in the urea cycle, and overproduction strains of both bacteria[151] and yeast[152] have been engineered previously to allow it to be obtained using cheap raw materials as a sole carbon source, for example through the sustainable fermentation of sucrose or molasses.[151]

Therefore, an *in vitro* cascade composed of a range of purified CAR homologs and a purified IRED/RedAm obtained from *A. dermatitidis* (*Ad*RedAm) (Section 2.2) was designed, built and screened for the synthesis of 3-aminopiperidine **11a** from the substrate L-ornithine **8a**. The application of this cascade in an *in vitro* fashion necessitated the supplementation of several additional reaction components to ensure the desired process could function successfully: $NADP^+$, GDH (Codexis, CDX-901) and glucose were added to recycle the NADPH

cofactor required for both the CAR and the IRED reactions, and ATP and $MgCl_2$ were provided for the CAR reaction to proceed (**Figure 57**).

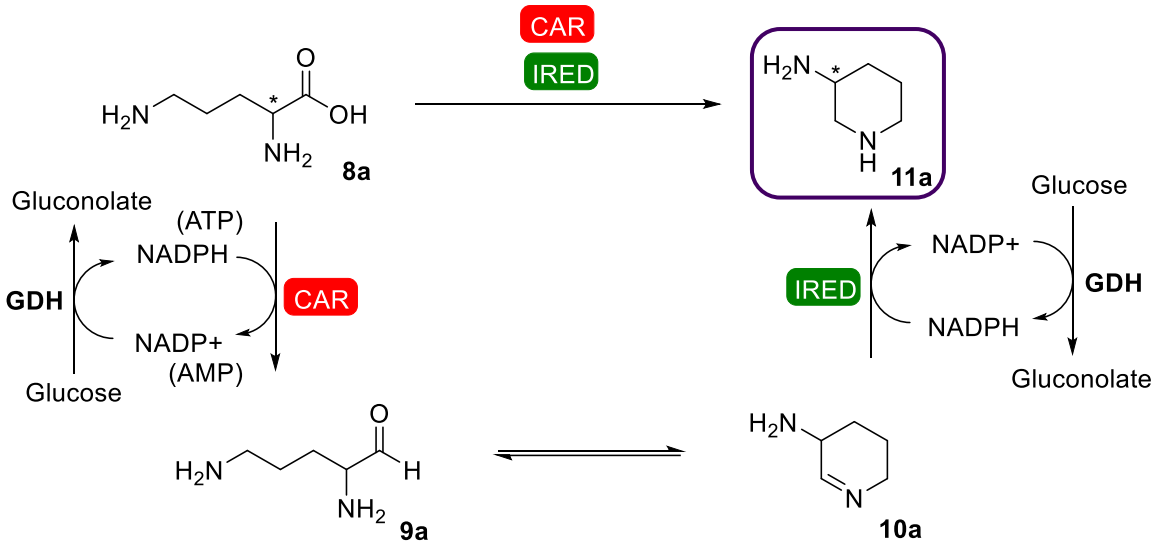


Figure 57 Enzymatic cascade proposed for the production of 3-aminopiperidine **11a**. GDH cofactor recycling system needs to be implemented to regenerate the NADPH cofactor required for the successful application of both CAR and IRED biocatalysts.

4.2.1 Chemical Synthesis of Cbz-Orn-OH **8d** and Boc-Orn-OH **8e**

Initial cascade reactions were executed using L-ornithine as the substrate, with a range of CAR homologs being used in conjunction with *Ad*RedAm. Extraction of the reactions into DCM (Section 6.4) followed by GC analysis indicated that, unfortunately, no conversion to 3-aminopiperidine **11a** was achieved. In fact, no aldehyde **9a** or imine **10a** intermediates were detected for the reactions either, suggesting that ornithine **8a** is not a suitable substrate for the initial CAR-catalysed step of the cascade reaction.

Previous studies have shown that CAR enzymes do not function effectively on substrates that contain a polar moiety, such as an amine group, at the alpha position relative to the carbonyl.[85] It was therefore decided that Boc- and Cbz-protected **8a** should be synthesised and tested for activity in the proposed cascade reaction. Both doubly protected Boc-Orn(Cbz)-OH **8b** and Cbz-Orn(Boc)-OH **8c** are commercially available, so selective deprotection of either the terminal Boc group or the terminal Cbz group resulted in mono-protected ornithine **8** at the alpha position, with the terminal amine remaining free (**Figure 58** and Section 6.2).

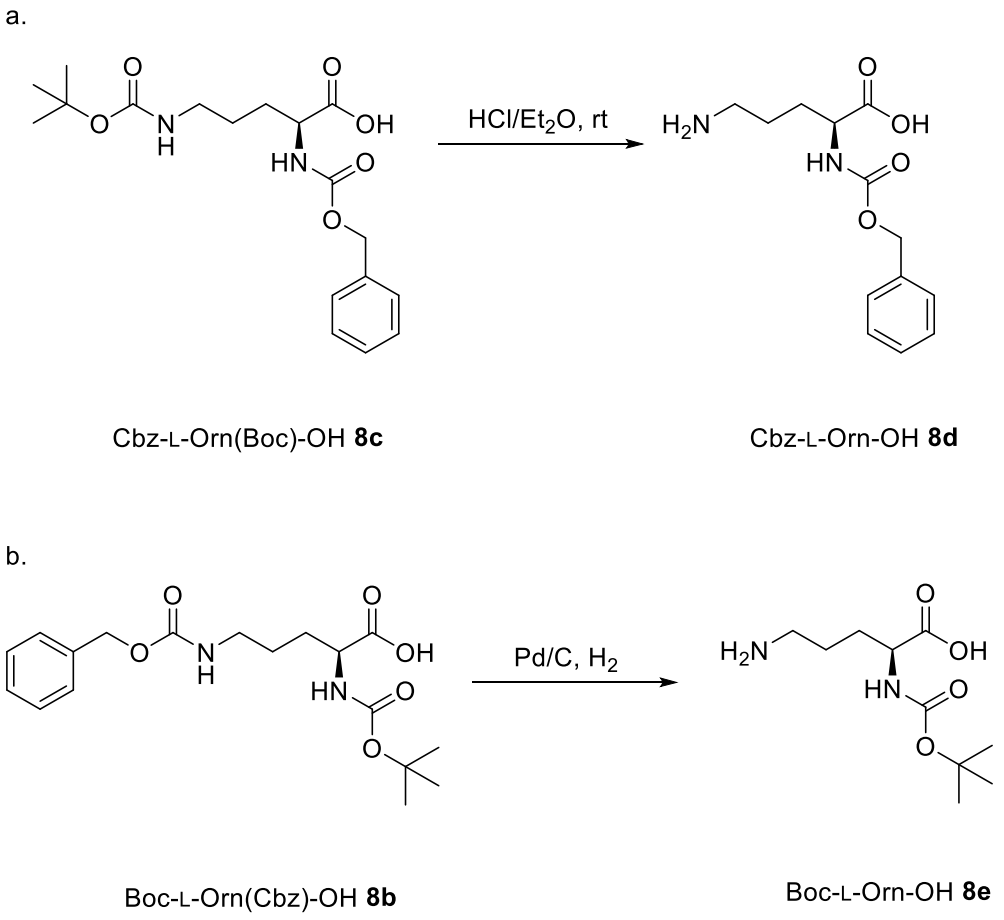


Figure 58 Deprotection of a. Cbz-Orn(Boc)-OH **8c** to give substrate Cbz-Orn-OH **8d** and b. Boc-Orn(Cbz)-OH **8b** to give substrate Boc-Orn-OH **8e**.

4.2.2 Results Using Boc- and Cbz-Protected Ornithine **8** as Substrate in the Proposed CAR-IRED Cascade

Once Cbz-Orn-OH **8d** and Boc-Orn-OH **8e** had been synthesised and purified, they were tested as substrates in the envisaged CAR-IRED cascade for the production of 3-aminopiperidine **11d-e**. Analysis of the reaction extractions revealed that the first enzyme of the proposed cascade, CAR, was not recognising the provided substrates, as no conversion to product amine **11** was seen, along with no evidence of aldehyde **9** or imine **10** production. It is possible that CAR is failing to act on these mono-protected substrates due to the presence of the free terminal amine contained on the ornithine **8** molecule (which is required for the proposed intramolecular cyclisation to intermediate imine **10** during a later step in the designed cascade reaction).

These negatives results suggested that, currently, carboxylic acid reductase is not a suitable enzyme for the conversion of ornithine **8** into its corresponding aldehyde **9**. For this reason, alternative cascade routes were required for the synthesis of 3-aminopiperidine **11**.

4.3 Proposed ADH-IRED Cascade for the Conversion of L-Ornithinol **12a** to 3-Aminopiperidine **11**

An alternative route towards 3-aminopiperidine **11** was developed through further biocatalytic retrosynthetic analysis, whereby the corresponding alcohol of ornithine **8** could be used as the substrate and be converted to an intermediate aldehyde **9** through the employment of an alcohol dehydrogenase (ADH) enzyme. The second step of the cascade would then, once again, be the

intramolecular cyclisation of the amino aldehyde to give an imine intermediate **10**, which could be reduced through the actions of *Ad*RedAm (**Figure 59**).

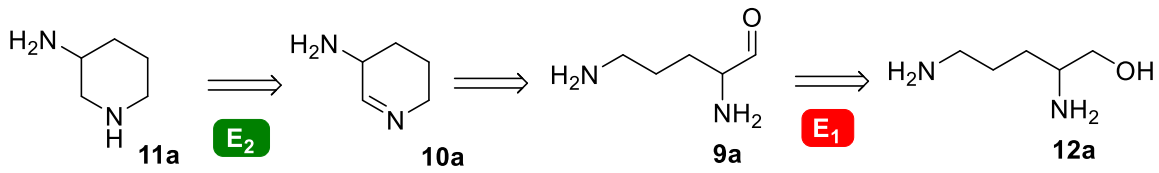


Figure 59 Biocatalytic retrosynthetic analysis of 3-aminopiperidine **11**, suggesting an enzymatic route using ornithinol **12** as the substrate to the desired product through the use of E_1 (ADH) and E_2 (IRED) biocatalysts.

Once again, all reactions had to be supplemented with $NADP^+$, glucose and GDH for the regeneration of the NADPH cofactor, as well as NAD^+ for ADH homologs reliant on this specific cofactor.

4.3.1 Chemical Synthesis of Mono-Protected Ornithinol **12a** from Mono-Protected Ornithine **8**

The mono-protected substrates of choice for the proposed ADH-IRED reaction, Cbz-L-ornithinol **12d** and Boc-L-ornithinol **12e**, are not commercially available and so must first be chemically synthesised in a two-step process from di-protected Cbz-Orn(Boc)-OH **8c** and Boc-Orn(Cbz)-OH **8b**, respectively. The first step of the process is the reduction of the carboxylic acid group of L-ornithine to its corresponding alcohol, followed by removal of the terminal protecting group to yield free terminal amine (**Figure 60**).

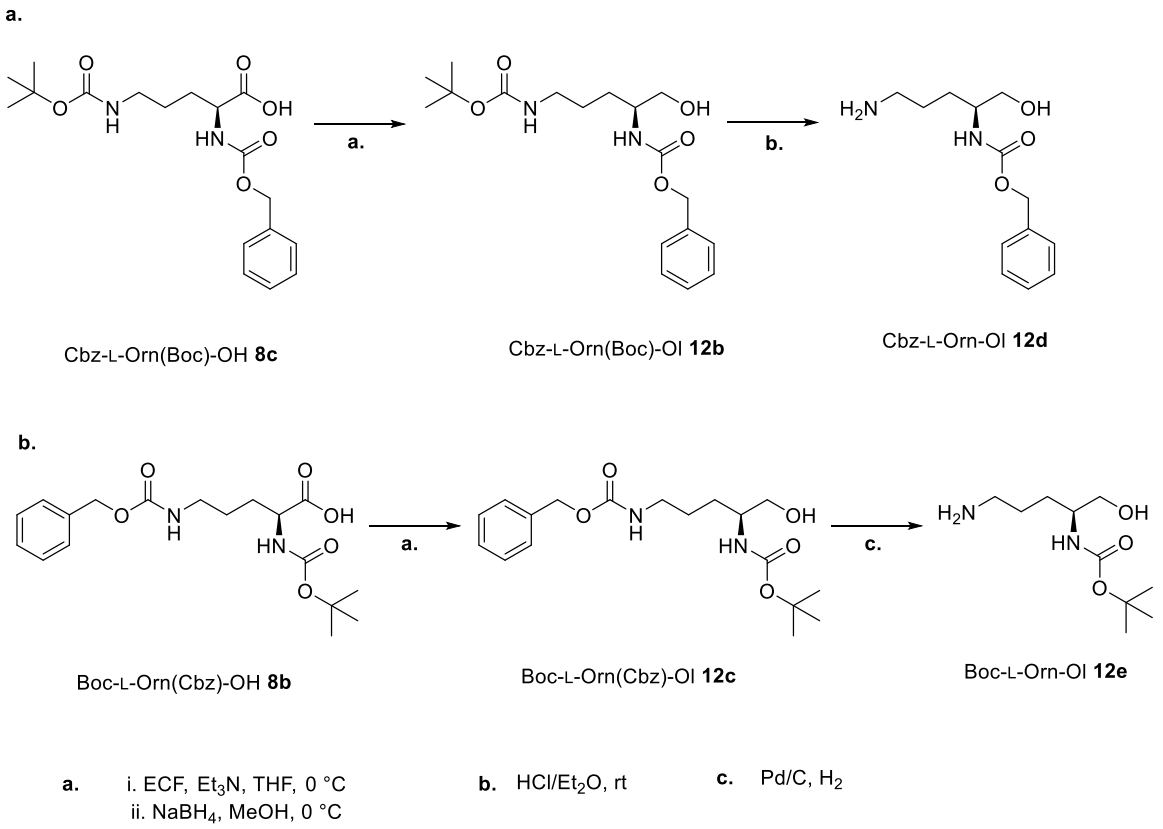


Figure 60 Reduction and deprotection of a. Cbz-Orn(Boc)-OH **8c** to give substrate Cbz-Orn-Ol **12d** and b. Boc-Orn(Cbz)-OH **8b** to give substrate Boc-Orn-Ol **12e**.

4.3.2 Results Using Boc- and Cbz-Protected L-Ornithinol **12** as Substrate in the Proposed ADH-IRED Cascade

The ADH-IRED cascade, suggested through the biocatalytic retrosynthetic disconnection of 3-aminopiperidine, was investigated using five commercial crude lysate preparations of ADH from Johnson Matthey (Section 2.2) in conjunction with the IRED *Ad*RedAm. The amino alcohol substrates Boc- and Cbz-L-ornithinol (**12e** and **12d**, respectively) were analysed for the conversion to the desired piperidine product. However, as with the CAR-IRED cascade tested previously, no amounts of product or intermediate aldehyde or imine were detected, suggesting that the selected substrates were not suitable to be

consumed by the ADH enzymes utilised during this process. It was therefore necessary to design an alternative cascade approach for the synthesis of the selected product.

4.4 Proposed GOase-IRED Cascade for the Conversion of L-Ornithinol **12** to 3-Aminopiperidine **11**

A similar disconnection to that seen in Section 4.3 revealed that another potential route to the synthesis of 3-aminopiperidine may be possible by the enzymatic oxidation of protected L-ornithinol to its corresponding aldehyde by a galactose oxidase (GOase) enzyme, with subsequent imine reduction catalysed by the selected IRED from *A. dermititidis* (**Figure 61**).

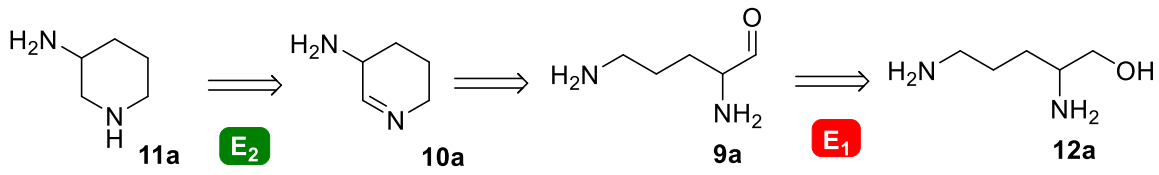


Figure 61 Biocatalytic retrosynthetic analysis of 3-aminopiperidine **11a**, suggesting an enzymatic route using ornithinol **12a** as the substrate to the desired product through the use of E_1 (GOase) and E_2 (IRED) biocatalysts.

Previous work from the Turner group indicated that two GOase variants in particular, GOase $M_{3\text{-}5}$[116,117,119] and GOase F_2,[117,120] are able to act on a range of *N*-Cbz-protected and free amino alcohol substrates (Section 1.6.4.2)[117]. For this reason, both of these GOase variants were selected for incorporation into the designed cascade process, since the intended mono-protected L-ornithinol substrates contain both a protected amine group alongside the free terminal

amine required for the spontaneous intramolecular cyclisation reaction during a later reaction step of the cascade (**Figure 62**).

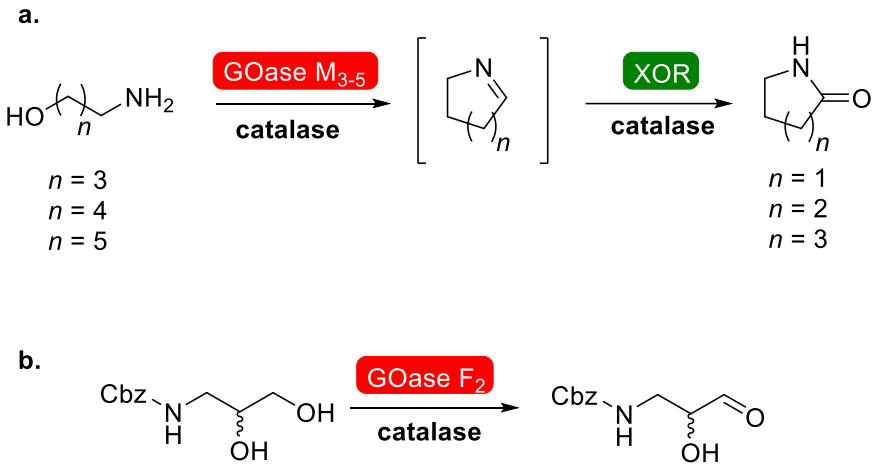


Figure 62 Both GOase $M_{3\text{-}5}$ and GOase F_2 variants have been shown to act on a range of (a.) free and (b.) *N*-Cbz-protected amino alcohols for the production of the corresponding amino aldehyde compounds.[117]

For the successful *in vitro* application of the GOase-IRED cascade described, several additional components need to be supplied to the system to enable the desired process to operate effectively. As detailed in Section 4.2, the use of a purified IRED enzyme requires the provision of the required $NADP^+$ cofactor, alongside a suitable regeneration system such as GDH and glucose. In addition to this, two extra enzymes, catalase and horseradish peroxidase (HRP) must be present in the reaction vessel when a GOase biocatalyst is used, to aid removal of the hydrogen peroxide (H_2O_2) by-product which can lead to inactivation of GOase and other enzymes.[114] A typical reaction set up for the intended GOase-IRED cascade is shown in **Figure 63**.

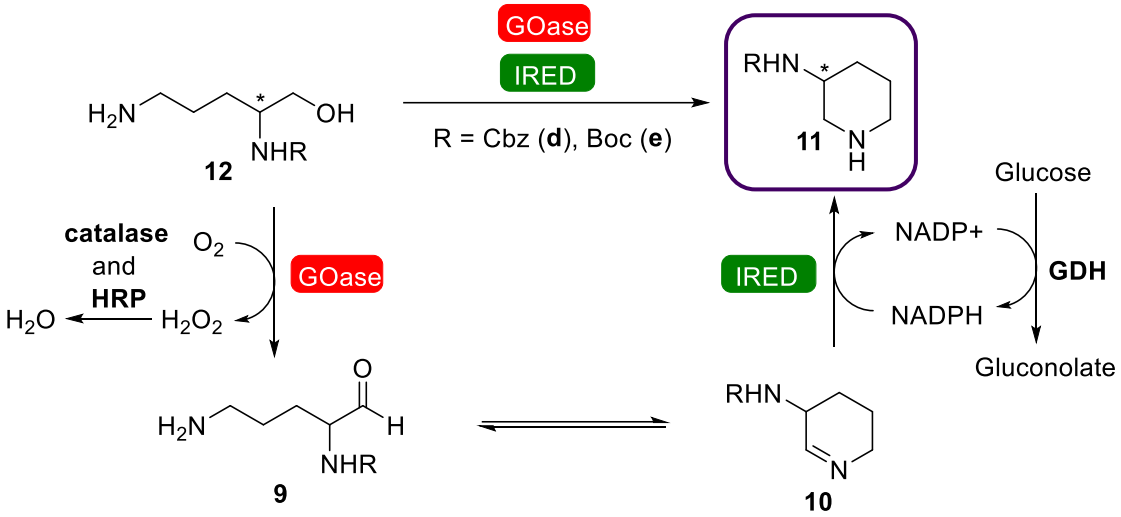


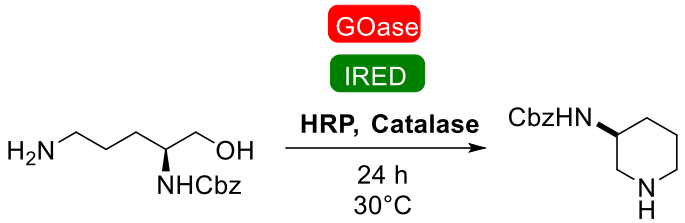
Figure 63 Enzymatic cascade proposed for the production of 3-aminopiperidine **11**. HRP and catalase must be supplemented to the reactions to enable the GOase enzyme to function effectively; GDH cofactor recycling system needs to be implemented to regenerate the NADPH cofactor required for the successful application of the IRED biocatalyst.

4.4.1 Results Using Boc- and Cbz-protected L-Ornithinol as Substrate in the Proposed GOase-IRED Cascade

4.4.1.1 Initial Results for the Production of Mono-Protected 3-Aminopiperidine **11** Using the Proposed GOase-IRED Cascade

To assess the functionality of the proposed GOase-IRED cascade for the production of the valuable pharmaceutical building block 3-aminopiperidine **11**, reactions containing either purified GOase $M_{3\text{-}5}$ or GOase F_2 and *Ad*RedAm were executed using either Boc-L-ornithinol or Cbz-L-ornithinol as the starting material (Section 6.5.5). Unfortunately, for reactions using Boc-L-ornithinol as the substrate, no conversion to product amine, nor aldehyde or imine intermediate was seen, suggesting that this Boc-protected compound is not a suitable substrate for either of the GOase variants screened. However, extraction of the reactions of Cbz-L-ornithinol into DCM (Section 6.4) followed by GC analysis revealed the appearance of a peak with the same retention time as that seen for an authentic

analytical standard of Cbz-3-aminopiperidine. Both GOase (M_{3-5})-*Ad*RedAm and GOase (F_2)-*Ad*RedAm cascade variants provided access to the target piperidine product, with the GOase M_{3-5} variant performing significantly better than its F_2 counterpart (**Table 8**). This result is in agreement with previously reported comparisons between the two GOase variants, where M_{3-5} regularly outperformed F_2 for the conversion of *N*-Cbz-protected amino alcohols.[117]



12d **11d**

Cascade Variant	Conversion to 11d/%
GOase (M_{3-5})-*Ad*RedAm	54
GOase (F_2)-*Ad*RedAm	20

Table 8 Synthesis of 3-aminopiperidine through the use of a GOase-IRED *in vitro* cascade reaction. Absolute conversion to product **11d** was calculated against a calibration curve using 1-phenylethanol as an internal standard (Section 6.4.1.1.4).

Following the success of this cascade reaction, a number of substrate concentrations were subsequently analysed for their percentage conversion to 3-aminopiperidine, in an attempt to optimise the process and to ensure that it was as efficient as possible. The results of this experiment are shown in **Figure 64**. It was seen that lower concentrations of the Cbz-L-ornithinol substrate gave rise to higher percentage conversions to the target product **11d**, with the conversion

dropping dramatically once higher concentrations of substrate **12d** were utilised. This indicates that, in order to achieve the highest conversions to **11d** possible, a substrate concentration of 1 mM should be used. However, some conversion to Cbz-3-aminopiperidine **11d** is still seen when higher substrate loadings such as 10 mM are used, suggesting that, with further optimisation, the GOase-IRED cascade could be operated at higher substrate concentrations quite efficiently.

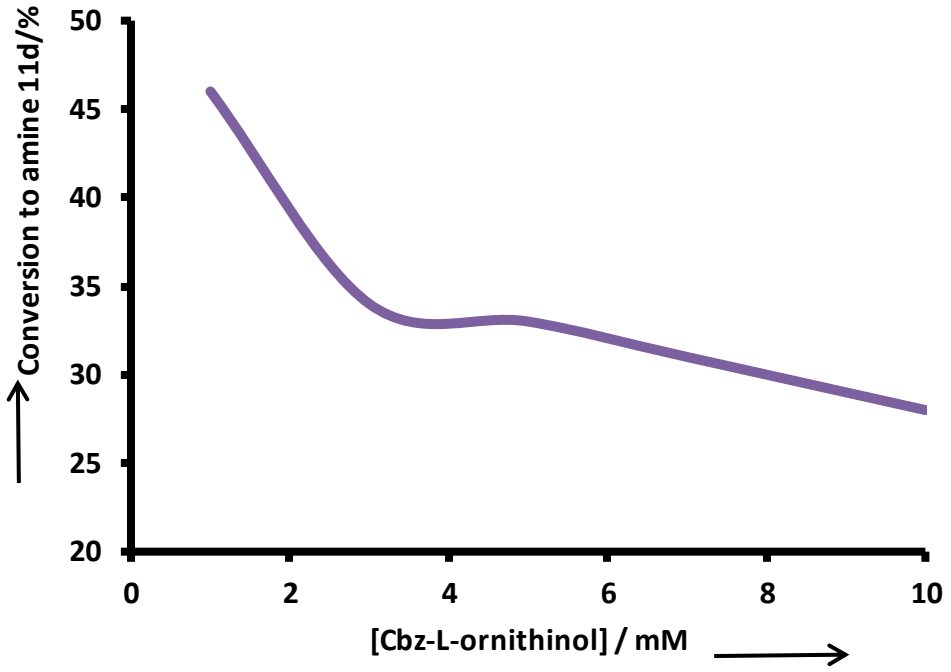


Figure 64 Effect of Cbz-L-ornithinol **12d** concentration ranging from 0 – 10 mM on Cbz-3-aminopiperidine **11d** production.

Following the investigation into the effect that an increased substrate loading would have on the GOase-IRED cascade, the effect of altering the concentrations of glucose and GDH present in the reaction was assessed. Both GDH and glucose are necessary for the successful implementation of the designed cascade process, since their absence would lead to the rapid depletion of the NADP(H) cofactor required for the reduction of the intermediate imine by *Ad*RedAm. The results of

varying the amounts of these essential reaction components are shown in **Figure 65** and **Figure 66**. As expected, it was shown that a lack of glucose provided to the reaction process leads to no conversion to Cbz-3-aminopiperidine **11d**, whilst a concentration of 25 mM appears to be a suitable amount to yield the reaction product. Increasing the levels of glucose beyond 25 mM was not seen to be beneficial to the overall reaction process, with slightly lower conversions to the expected product **11d** seen when 50 mM or higher was used. Similarly, neglecting to provide the *in vitro* cascade reaction with the required GDH enzyme necessary for the regeneration of the NADP(H) cofactor led to only minimal conversions to the desired product (around 6 %). Since the biotransformations were supplemented with $NADP^+$ rather than the NADPH required for the IRED step of the cascade, this 6 % conversion to product in the absence of any recycling system was presumably due to the use of residual NADPH cofactor within the IRED active site. Providing a concentration of GDH as low as 0.2 mg mL^{-1} provided access to 54 % of the expected amine product, with concentrations of 0.5 – 1 mg mL^{-1} resulting in slightly lower conversions to **11d**.

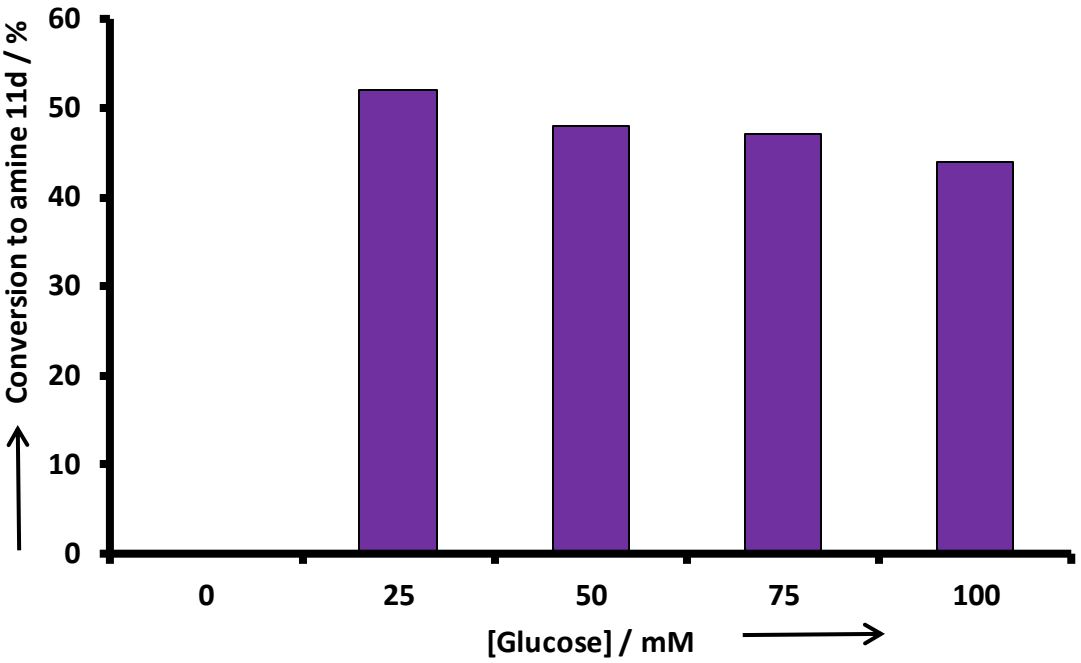


Figure 65 Effect of glucose concentration ranging from 0 - 100 mM on Cbz-3-aminopiperidine **11d** production.

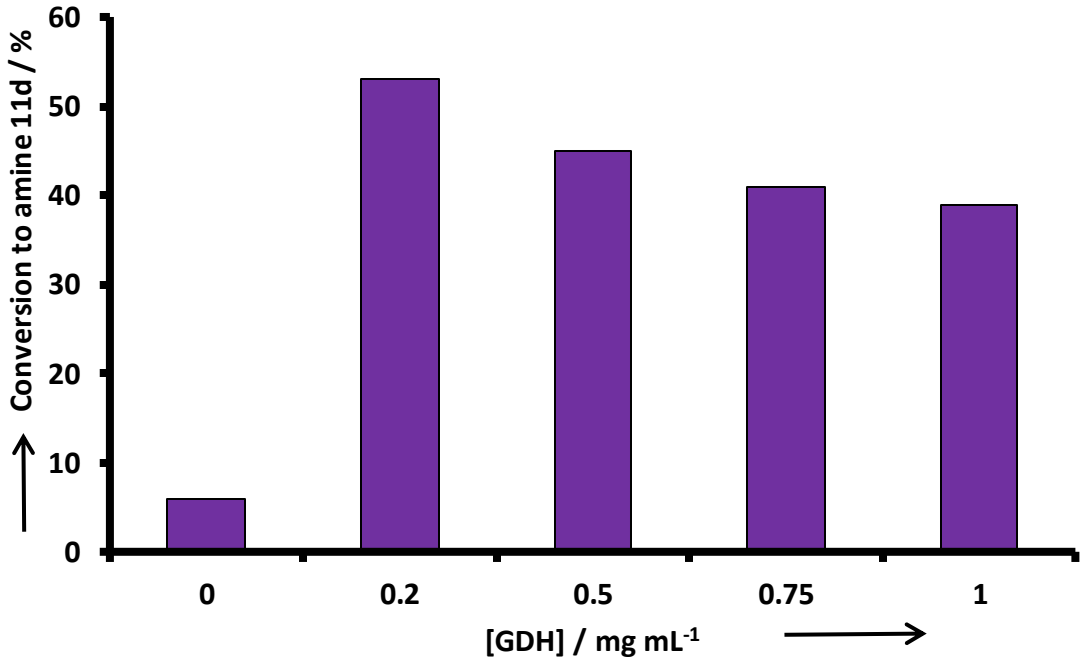


Figure 66 Effect of GDH concentration ranging from 0 - 1 mg mL^{-1} on Cbz-3-aminopiperidine **11d** production.

This preliminary result, demonstrating that a cascade composed of GOase and IRED biocatalysts can successfully generate the desired Cbz-3-aminopiperidine **11d** structure through the conversion of the less complex, linear Cbz-L-ornithinol **12d** substrate, is a good starting point for the development of an efficient and sustainable route to the synthesis of this valuable molecule. However, as with the CAR-TA-IRED cascade described in Chapter 3, there are many reaction parameters implicated in the success of this reaction that must be suitably optimised. **Figure 67** highlights the different reaction components that must be added to each reaction to facilitate the successful function of both cascade enzymes.

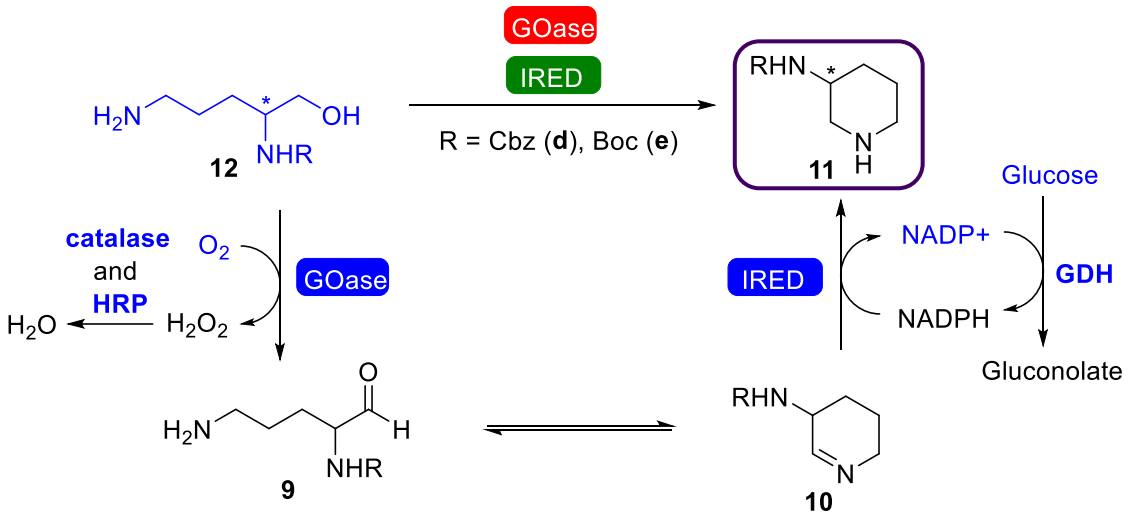


Figure 67 Enzymatic cascade proposed for the production of 3-aminopiperidine **11**. Items highlighted in blue represent the reaction components supplied to facilitate the cascade reaction.

Optimisation of the concentration of each of these reaction components, as seen with the external reaction conditions optimised for the CAR-TA-IRED cascade in Chapter 3 (Section 3.3.2), would lead to the highest possible conversion to the desired product of this cascade reaction. Owing to the large number of additional

parameters that would have to be explored due to the *in vitro* nature of the designed enzymatic pathway (e.g. $NADP^+$ concentration), a more streamlined alternative for the enhancement of the biocatalytic process would be beneficial (e.g. a bioinformatics approach through the use of Design of Experiments).[153] Once the entire system had been suitably optimised, the cascade reaction could be implemented on a larger scale to allow for the isolation and full characterisation of the obtained product.

4.4.1.2 A Comparison Between an In Vitro Application and a Hybrid Application for the Proposed GOase-IRED Cascade

The implementation of a chosen biocatalyst as either an *in vivo* component or an *in vitro* component to the designed cascade reaction carries with it certain benefits and drawbacks, dependent on the cascade mode used (Section 1.2). In the initial screening of the GOase-IRED cascade, purified enzyme for both GOase and *Ad*RedAm was used. However, in an attempt to circumvent the requirement for laborious protein purification, a crude lysate preparation of *Ad*RedAm was assessed to see whether it retained sufficient activity for the reduction of the imine intermediate. A comparison between reactions executed using purified *Ad*RedAm against reactions utilising the crude lysate suggested that the IRED remained stable and active when prepared in either fashion, but the conversions to Cbz-3-aminopiperidine **11d** for reactions using the crude lysate preparations were significantly lower than those seen when purified enzyme was used (**Figure 68**).

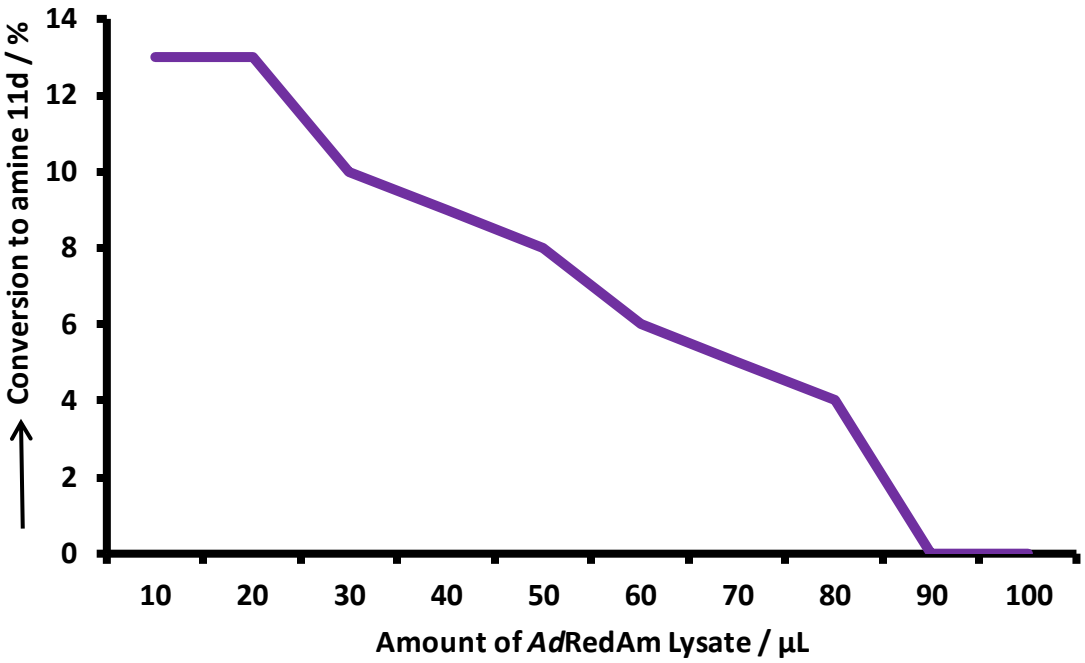


Figure 68 Effect of the amount of *Ad*RedAm lysate, ranging from 0 - 100 µL, on Cbz-3-aminopiperidine **11d** production.

The results depicted in **Figure 68** suggest that increasing the amount of *Ad*RedAm crude lysate in the reaction leads to a decrease in the conversion to **11d**, and the conversions seen are drastically lower than when a purified preparation of *Ad*RedAm is used as the IRED biocatalyst (from 54 % when purified *Ad*RedAm is used to 13 % when 10 µL of *Ad*RedAm lysate is used). For this reason, it may first be necessary to improve the biotransformations utilising a crude lysate preparation of *Ad*RedAm by investigating the optimal concentrations of essential components, such as the GDH/glucose cofactor recycling system, before the IRED is applied in this fashion for the established GOase-IRED cascade. Previous studies have also shown that GOase can be added to a biotransformation as lyophilised whole cells,[115] suggesting that it is not necessary to purify the protein prior to use. If the conversions to amine product can be improved through optimisation of biotransformations employing *Ad*RedAm as a lysate, this leads to the possibility of co-expressing both enzymes

within the same cell prior to lyophilisation, resulting in a more efficient and streamlined biocatalyst preparation.

It has also been shown previously, in this work (Chapter 3) and elsewhere,[40,49,105,106] that IRED enzymes can be employed within a living whole cell. When the imine reduction catalysed by an IRED enzyme is contained within a host cell, it is no longer necessary to supplement the reactions with $NADP^+$/NADPH or a cofactor regeneration system since the inherent machinery within the cell meets these needs. A hybrid reaction set up, comprised of purified GOase enzyme and *Ad*RedAm contained within a whole cell, was therefore investigated for its production of Cbz-3-aminopiperidine. Unfortunately these reactions did not yield any production of **11d**, suggesting that the imine intermediate may not be able to efficiently traverse the cell membrane in order to come into contact with the IRED biocatalyst.

4.4.1.3 Assessing the Chirality of the 3-Aminopiperidine Cascade Product

In the current work, the substrate shown to be accepted by both enzymes in the established cascade was the L-enantiomer of Cbz-ornithinol. Since the substrate has pre-existing chirality, it may be expected that this chirality would be carried through into the resulting product of the cascade reaction, meaning that the conversion of Cbz-L-ornithinol would provide access to Cbz-(*S*)-3-aminopiperidine, whilst the same cascade process using Cbz-D-ornithinol would generate the corresponding Cbz-(*R*)-3-aminopiperidine compound. However, as shown in Section 3.4, it is possible for the intermediate imine to tautomerise into its corresponding enamine **13d** prior to being reduced by the IRED biocatalyst present in the reaction, which would racemise any pre-existing chirality of the substrate (**Figure 69**).

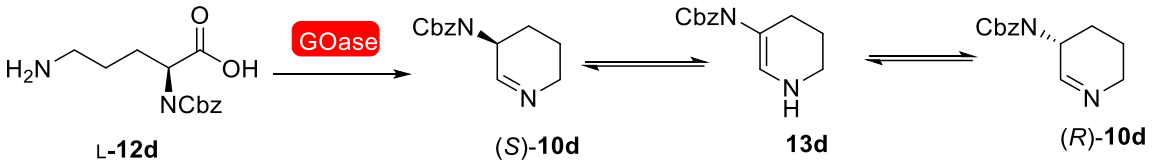


Figure 69 Potential racemisation of Cbz-L-ornithinol **12d** by spontaneous imine-enamine tautomerisation.

If this imine-enamine tautomerisation takes place, it is then possible for both enantiomers of the imine intermediate to be made available for the final step of the cascade reaction, meaning that the chirality of the given product would be dependent on the inherent selectivity of the IRED biocatalyst present rather than the enantiomer of the substrate used. For this reason, it is necessary to analyse the chirality of the resulting product of the cascade reaction, through either comparison with optically pure authentic standards using an appropriate analytical tool, such as GC equipped with a chiral column, or through optical rotation measurements once the reaction has been achieved on a preparative-scale and the product has been successfully isolated. Owing to the time constraints of this project, the evaluation of the chirality of the developed GOase-IRED cascade has not yet been executed.

4.5 Conclusions and Outlook

Three possible *de novo* multi-enzyme biocatalytic routes to the synthesis of the important drug precursor scaffold, 3-aminopiperidine **11**, were conceived through biocatalytic retrosynthesis of the target compound. Cascades comprised of carboxylic acid reductase (CAR, Section 1.6.1) and imine reductase (IRED, Section 1.6.3), alcohol dehydrogenase (ADH, Section 1.6.5) and IRED and

galactose oxidase (GOase, Section 1.6.4) and IRED were assessed for the ability to produce **11** from nonessential amino acid L-ornithine **8** (CAR-IRED) or its corresponding alcohol, L-ornithinol **12** (ADH-IRED and GOase-IRED).

Initial screens using L-ornithine with Cbz- or Boc-protection of the α-amine, **8d** and **8e**, respectively, as substrate for a cascade using one of five CAR homologs and the IRED from *Ajellomyces dermatitidis* (*Ad*RedAm) showed no detectable conversion of the substrate by CAR. Previous work has suggested that CAR enzymes do not work optimally on substrates containing polar constituents in addition to the carboxylate group,[88,154] which may explain the lack of conversion seen for **8d** and **8e** here. The adenylation-peptidyl carrier protein (adenylation-PCP) didomains of two CAR homologs studied for this cascade, NCAR and *sr*CAR, have recently been successfully crystallised, leading to the elucidation of their structures.[89] Computational modelling of L-ornithine **8**, and its two protected derivatives **8d** and **8e**, may therefore shed light on how this substrate interacts within the active site of the adenylation domain of CAR, and could implicate specific residues that could be mutated in an attempt to engineer CAR variants to convert **8** into the corresponding aldehyde. A CAR-IRED method of producing 3-aminopiperidine **11** would be beneficial due to the fact that both enzymes have previously been shown to be functional as whole cell biocatalysts, alongside use as isolated enzymes. A whole cell expressing both proteins would remove the need to add the essential cofactors required for the reaction, (2)NADPH and ATP, as the inherent regeneration machinery within the host could be used towards this purpose.

A cascade composed of ADH-IRED was screened for the conversion of Cbz- or Boc-L-ornithinol **12d** and **12e** to yield mono-protected 3-aminopiperidine **11d** or **11e**, respectively. A range of commercial crude lysate preparations of ADH were employed in an *in vitro* cascade with purified *Ad*RedAm, but no conversion of the substrate was detected. GOase variants $M_{3\text{-}5}$ and F_2 have previously been shown

to convert a range of aliphatic and aromatic amino alcohols to the corresponding amino aldehydes,[117] and so an alternative cascade for the conversion of **12d** and **12e** to yield the desired amine product was proposed. Mono-protected L-ornithinol **12d-e** was screened against a cascade consisting of one of the two GOase variants, and *Ad*RedAm, and the production of Cbz-3-aminopiperidine **11d** was clearly detectable through GC analysis, yet no conversion of Boc-L-ornithinol **12e** was seen to occur. The novel GOase M3-5-*Ad*RedAm cascade variant was the highest performing of the two systems investigated, and so initial optimisation experiments with respect to substrate concentration, *Ad*RedAm enzyme preparation (purified, crude lysate or whole cell), glucose concentration and GDH recycling enzyme concentration were carried out.

The substrate for the GOase-IRED cascade, Cbz-L-ornithinol **12d**, was provided as an optically pure preparation, and it was expected that this pre-existing chirality would carry through to the cyclic amine product. However, as seen in Chapter 3 for the conversion of keto acid **1e** to the corresponding piperidine **5e**, it may be possible that the intermediate imine formed after the GOase-catalysed production of aldehyde can tautomerise with its enamine, thus racemising this stereocentre and providing *Ad*RedAm access to both enantiomers of imine **10d**. For this reason, it is necessary to develop a chiral analytical method for the separation of the possible enantiomers of Cbz-3-aminopiperidine **11d**, and, subsequently, compare the biotransformation products against authentic chemical standards to determine the selectivity, if any, of the cascade reaction. Towards this same objective, it would also be interesting to provide Cbz-D-ornithinol **12d** as substrate for the cascade reaction, to assess whether or not the same product enantiomer(s) is produced.

Once the designed GOase-IRED cascade for the conversion of Cbz-L-ornithinol **12d** has been sufficiently optimised to obtain the best conversions to Cbz-3-aminopiperidine **11d**, the most favourable conditions should be used for a

preparative-scale synthesis to enable isolation and characterisation of the product obtained.

5 Biocatalytic Routes to Amide Bond Formation

5.1 Introduction

Amide bond functionality has been identified in a large number of pharmaceutically-relevant and biologically active compounds,[77,78] yet the current chemical routes towards the synthesis of amides have several drawbacks, including the need for high temperatures and long reaction times, and the use of strong base catalysts (Section 1.4.2). For this reason, a sustainable enzymatic route to amide formation presents itself as an attractive target for addition to the biocatalytic toolbox at a chemist's disposal. Current enzymatic methods for the formation of amide bonds are either dependent on stoichiometric amounts of ATP as cofactor and display very narrow substrate specificities,[127,155,156] or require low water environments to improve the synthesis/hydrolysis (S/H) ratio to drive the process towards amide bond formation,[157] which all limit their applicability for biocatalytic processes. However, an interesting candidate was identified in the literature that is capable of catalysing the ATP-independent formation of an amide bond between a methyl ester and a free amine partner in aqueous environments.[128,158] The *N*-acyltransferase CapW (Section 1.6.6) demonstrated the ability to ligate the ester substrate **14a** with L-aminocaprolactam **15a** (**Figure 70**) *in vivo* for the synthesis of the capuramycin-type antibiotic compound A-503083 B. Another enzyme, the nonribosomal peptide synthetase CapU, was responsible for the cyclisation of L-lysine to yield the amine partner L-aminocaprolactam **15a** at the expense of ATP.

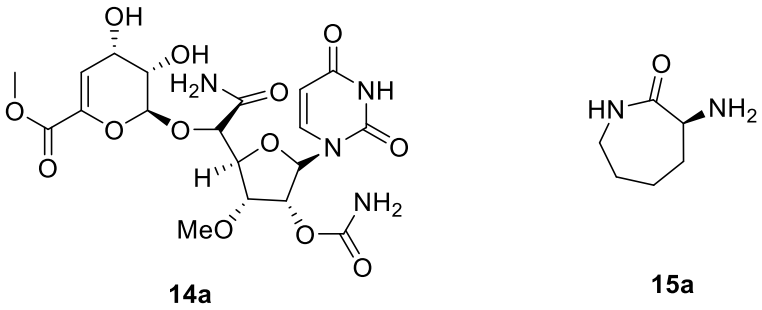


Figure 70 The natural methyl ester **14a** and amine **15a** substrates of CapW.

Another enzymatic strategy reported by Flitsch *et al.* recently was the formation of amide bonds through the employment of a CAR (Section 1.6.1) enzyme.[130] CAR is a modular protein consisting of three distinct domains: the (1) adenylation domain, responsible for the selection and activation of carboxylic acid substrates, a (2) PCP domain to tether the acyl substrate and transfer it to the (3) reductase domain, which catalyses the formation of aldehyde through reduction using NADPH.

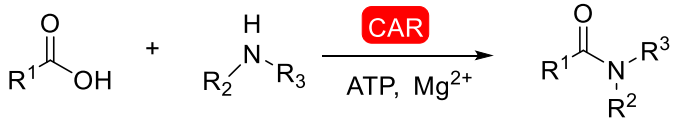


Figure 71 Amide bond formation catalysed by intercepting adenylated acyl substrate of CAR using amine as an external nucleophile.

It was observed that, in the absence of the NADPH cofactor required to reduce the tethered acyl substrate at the reductase domain, the acyl-AMP intermediate formed through the actions of the adenylation domain could be intercepted and react with an external nucleophile, such as a primary or secondary amine, resulting in the formation of an amide bond (**Figure 71**). This expanded reaction scope of CAR enzymes led to the possibility that carboxylic acid substrates containing a terminal nucleophile, such as an amine or hydroxyl group, could be

used in the production of lactam and lactone products, through the spontaneous intramolecular cyclisation made possible via the activated acyl-AMP. If this cyclisation reaction could be catalysed by CAR then an alternative route towards the production of the natural amine donor of CapW, L-aminocaprolactam **15a**, could be provided enzymatically, prior to subsequent CapW-catalysed linkage to a methyl ester substrate through amide bond formation.

CapW has not yet been explored with respect to its ability to catalyse amide bond formation between L-aminocaprolactam **15a** and non-natural methyl ester substrates, in large part due to the fact that it has not currently been solubly expressed in standard bacterial expression strains such as *E. coli*. Therefore, the main aim of this study was to obtain soluble recombinant expression of the CapW protein, prior to screening the enzyme for amide bond forming activity against a range of methyl ester substrates and its natural amine donor L-aminocaprolactam **15a**. Alongside this, initial studies into screening CAR enzymes for the cyclisation of simple carboxylic acids containing terminal amine groups were performed with the target of generating lactam products.

5.2 Introduction to ATP-Independent *N*-Transacylase CapW

CapW is an *N*-transacylase that catalyses amide bond formation without relying on ATP-activation of substrates.[128] The fact that this enzyme proceeds in the absence of ATP, in addition to the fact that it demonstrates very low hydrolytic activity with respect to the amide bond present in its natural product, suggests that CapW could be an ideal biocatalyst for amide bond formation *in vitro* under aqueous conditions. For CapW to realise its full potential, its substrate scope must first be investigated. Van Lanen *et al.* reported that a number of expression conditions had been screened in an attempt to achieve soluble expression of

CapW in *E. coli*, yet these experiments failed and so CapW was ultimately expressed in *Streptomyces lividans* TK6431. To effectively explore the possibility of exploiting CapW for the synthesis of a range of amide products, it was first necessary to obtain soluble expression of CapW in the more genetically tractable host organism *E. coli*. The ability to efficiently obtain large amounts of this protein would also be beneficial in downstream processing and in potential further studies, such as mutagenesis.

5.2.1 Optimisation of Soluble Expression of CapW

To date, attempts to achieve soluble heterologous expression of CapW in standard bacterial expression strains, such as *E. coli* BL21 (DE3), have been unsuccessful. Van Lanen *et al*. reported that a variety of expression conditions in *E. coli* whole cells had been tried yet all yielded only insoluble protein.[128] For this reason, expression trials for soluble CapW expression in *E. coli* were undertaken. The gene for CapW (His-tagged at the N-terminal for protein purification) was codon-optimised for expression in *E. coli*, ligated into entry vector pRSET, and transformed into a variety of *E. coli* strains to assess soluble protein expression.

pRSET/CapW was initially transformed into *E. coli* BL21 (DE3) pLysS competent cells and a range of different expression times and temperatures were screened in LB media. Soluble and insoluble protein expression was analysed using western blotting as analysis using SDS-PAGE was often inconclusive. This trial was unsuccessful; clear overexpression of CapW was demonstrated but unfortunately only insoluble protein was obtained at every time length and temperature investigated. pRSET/CapW was then transformed into ArcticExpress competent cells, a strain derived from BL21-Gold competent cells that have been engineered to increase soluble heterologous protein expression for problematic proteins.

Expression temperatures ranging from 16 °C – 30 °C were utilised for a variety of time lengths (1 h, 2 h, 4 h, o/n). ArcticExpress cells contain two heterologous cold-adapted chaperonins (Cpn10 and Cpn60 from *Oleispira antarctica*) that are expected to improve protein refolding activities at lower temperatures,[159] and so it was expected that more soluble protein would be produced when lower expression temperatures were used. However, no soluble protein expression was observed at 16 °C and a slight improvement was seen at 25 °C (2 h expression). The levels of expression seen using western blotting were not adequate, and so an alternative method was needed.

It has been reported that supplementing bacterial growth media with solubility enhancers such as sorbitol and glycyl betaine (betaine) can encourage the expression of soluble protein.[160] A high concentration of sorbitol outside of the cell creates great osmotic pressure, leading to the uptake of betaine (an osmolyte) through the cell membrane. Once within the cell it is thought that betaine acts as a chemical chaperone to the overexpressed protein, resulting in the properly folded protein being more energetically stable than unfolded protein.[161] Both pRSET/CapW in *E. coli* BL21(DE3) pLysS and in ArcticExpress competent cells were grown in LB media supplemented with 1 M sorbitol and 2.5 mM betaine. Again, different expression temperatures (20 °C, 23 °C, 25 °C) and expression times (1 h, 2 h, 4 h, o/n) were investigated. It was seen that, in LB supplemented with 1 M sorbitol and 2.5 mM betaine, the expression of soluble protein was increased significantly when compared to cultures grown in solely LB media. The best expression conditions found during the sorbitol/betaine screen was pRSET/CapW in *E. coli* BL21(DE3) pLysS at 25 °C for 1 h (**Figure 72**).

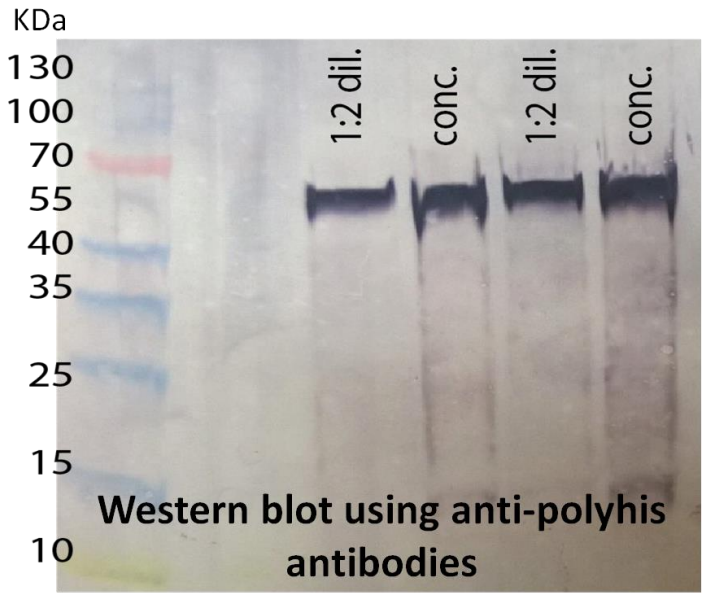


Figure 72 Western blot showing soluble expression of CapW (band at 47 kDa) in *E. coli* BL21(DE3) pLysS cells. LB with 1 M sorbitol, 2.5 mM betaine. 25 °C for 1 h.

One problem noted when using the sorbitol/betaine growth media was the abnormally slow growth rate of the cells. Cells grown in the presence of 1 M sorbitol and 2.5 mM betaine took on average twice as long to reach OD_{600} 0.6 as cells grown in just LB. This slower growth resulted in lower efficiency when using the supplemented growth media, and so the issue of high concentrations of sorbitol was addressed. The amount of soluble expression of CapW was assessed under altering concentrations of sorbitol in the growth media, from 0 M sorbitol to 1 M sorbitol. It was seen that lower concentrations of sorbitol (< 0.25 M) did not appear to increase the amount of soluble protein expression, yet 0.5 M sorbitol gave a comparable amount of soluble expression to 1 M sorbitol (**Figure 73**). The growth rate of *E. coli* BL21(DE3) pLysS in the presence of the different sorbitol concentrations was also measured (**Figure 74**). As was expected, increasing the amount of sorbitol in the growth media led to a decrease in the growth rate of the cells. This is likely due to the cells experiencing osmotic shock when being introduced to the high concentrations of sorbitol in the growth

media. *E. coli* BL21(DE3) pLysS cells harbouring pRSET/CapW took around 2 h longer to reach OD_{600} 0.6 in 1 M sorbitol than they did in 0.5 M sorbitol, with the extent of soluble protein expression apparently similar for both concentrations. It was also found that *E. coli* BL21(DE3) pLysS cells harbouring empty pET16-b vector (to confer antibiotic resistance) used as a negative control reached OD_{600} 0.6 on average 2 h earlier than cells harbouring pRSET/CapW, both grown in the presence of 1 M sorbitol. This indicates that the slow growth rate for the CapW cells was not solely due to the presence of a high concentration of sorbitol within the growth media, and suggests that high concentrations of CapW may be toxic to the host cell.

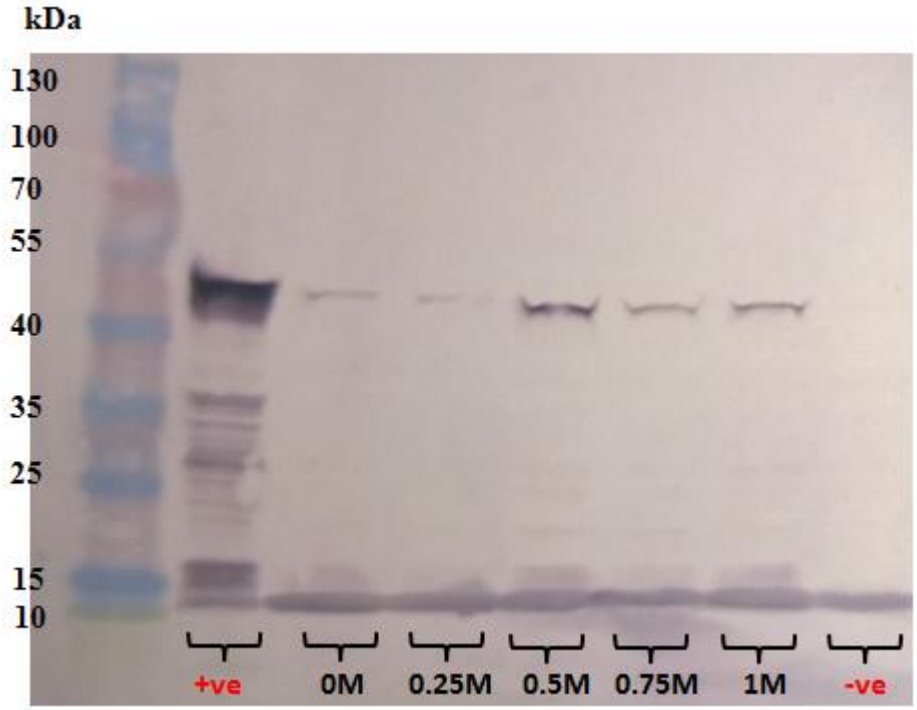


Figure 73 Western blot of soluble expression of CapW (47 kDa) under differing concentrations of sorbitol (+ve = positive control of CapW expression, -ve = negative control of empty pET16-b vector).

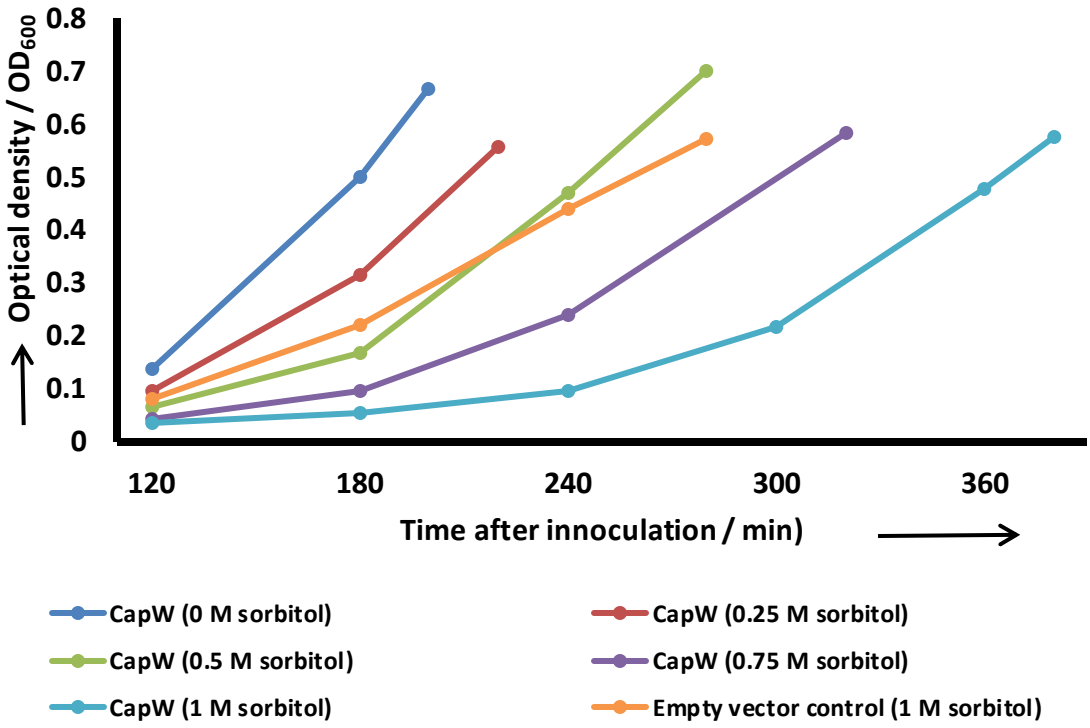


Figure 74 Effect of sorbitol concentration on culture growth rate (*E.coli* BL21 (DE3) pLysS).

OverExpress C41 (DE3) and C43 (DE3) chemically competent cells are *E. coli* BL21 (DE3) mutant strains that have been engineered to aid expression of toxic proteins. For this reason, pRSET/CapW was transformed into both strains and the soluble expression of CapW was assessed in LB, LB supplemented with 0.5 M sorbitol and LB supplemented with 1 M sorbitol. It was seen that soluble CapW was expressed in C41 (DE3) cells under all three conditions, yet no soluble CapW expression was observed for C43 (DE3) (**Figure 75**). In fact, soluble expression appeared to be highest in C41 (DE3) cells when the growth media was not supplemented with sorbitol/betaine. An expression trial of differing temperatures (20 °C, 25 °C, 30 °C, 37 °C) and time lengths (2 h, 4 h, o/n) for C41 (DE3) cells harbouring pRSET/CapW followed. From this trial, 4 different conditions were selected and scaled up (**Figure 76**). Inducing expression at 25 °C and allowing

expression to continue for 4 h was considered the best method to produce soluble CapW in C41 (DE3) cells.

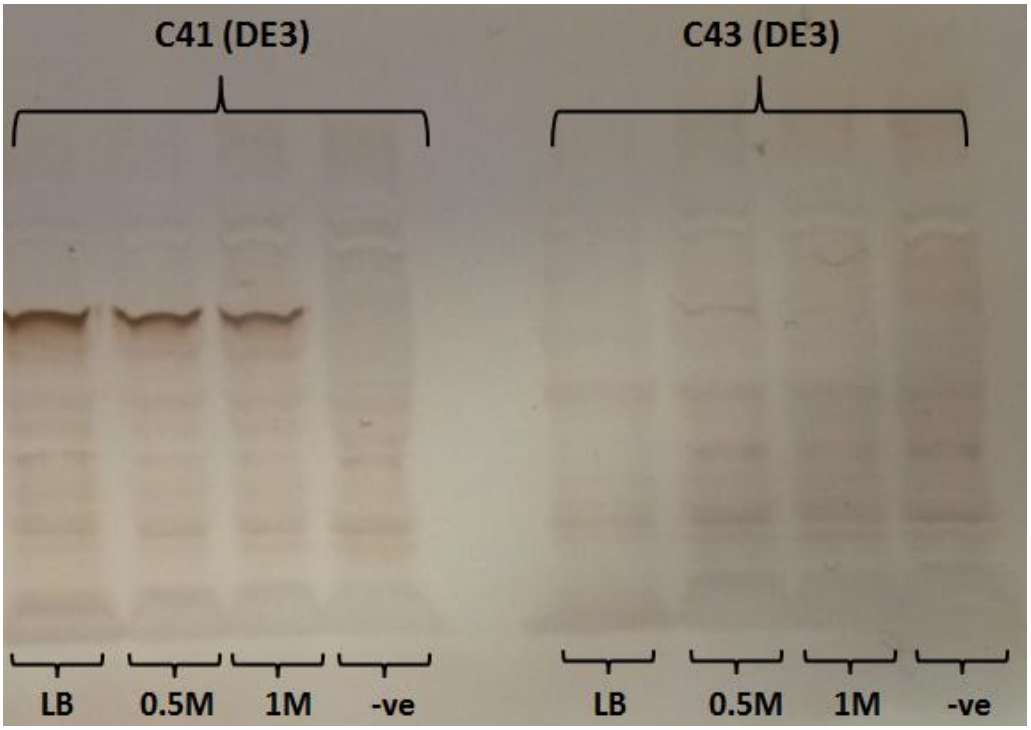


Figure 75 Western blot of soluble fractions from C41 (DE3) and C43 (DE3) expression trial in LB, LB and 0.5 M sorbitol, LB and 1 M sorbitol (-ve = negative control of cells harbouring empty pET16-b vector).

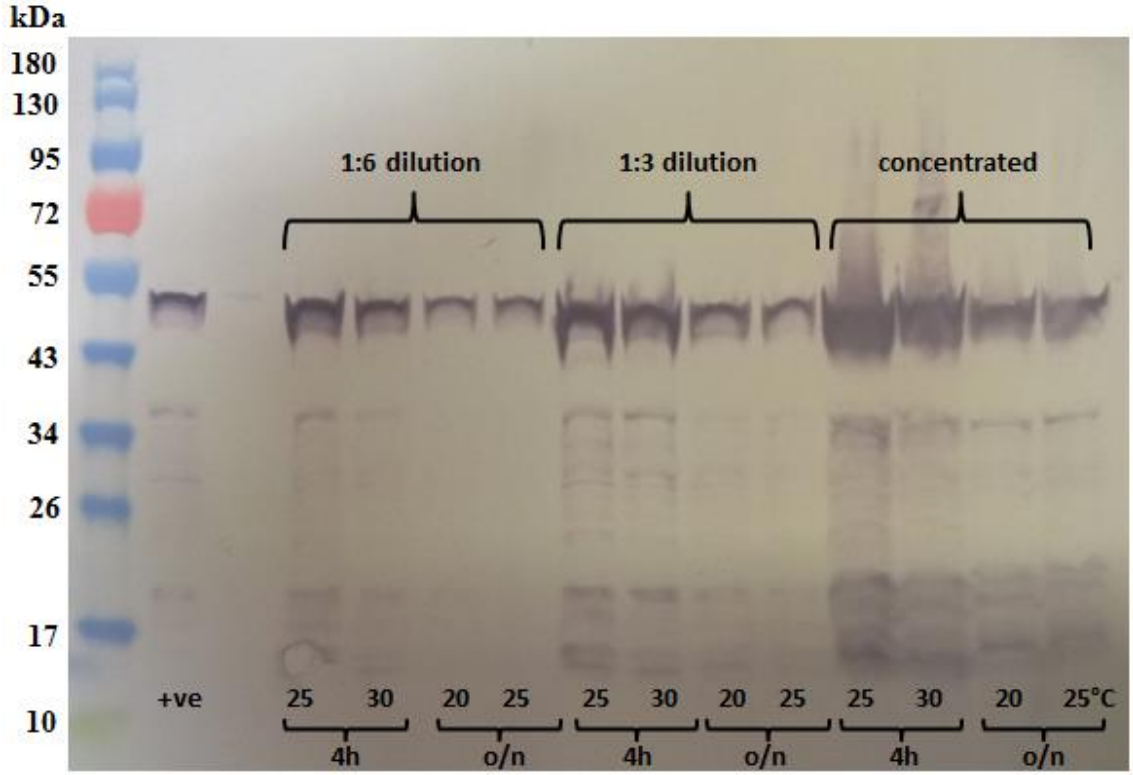


Figure 76 Western blot of soluble fractions from further C41 (DE3) expression trial (+ve = positive control of best CapW expression for 1 M sorbitol/2.5 mM betaine conditions) (CapW at 47 kDa).

5.2.2 Screening CapW for Activity Against a Range of Non-Natural Methyl Ester Substrates

5.2.2.1 Introduction

The natural acyl donor substrate **14a** for CapW is a complex metabolite that is not commercially available (**Figure 70** and **Figure 80**). This was problematic as it meant that the reaction known to be catalysed by CapW could not be used to rapidly verify whether the recombinantly expressed protein was active or not. Therefore, an alternative highly sensitive and high-throughput method of analysis was required to screen CapW for activity against a wide range of acyl donor substrates (**14b-t**, Section 5.2.2.2). A method was developed using mass spectrometry to allow swift detection of trace amounts of amide product. This

was achieved by analysing seven authentic amide standards **16a-g** corresponding to seven of the potential products that biotransformations using CapW could yield through the coupling of L-aminocaprolactam **15a** with the corresponding acyl donor (**Figure 77**).

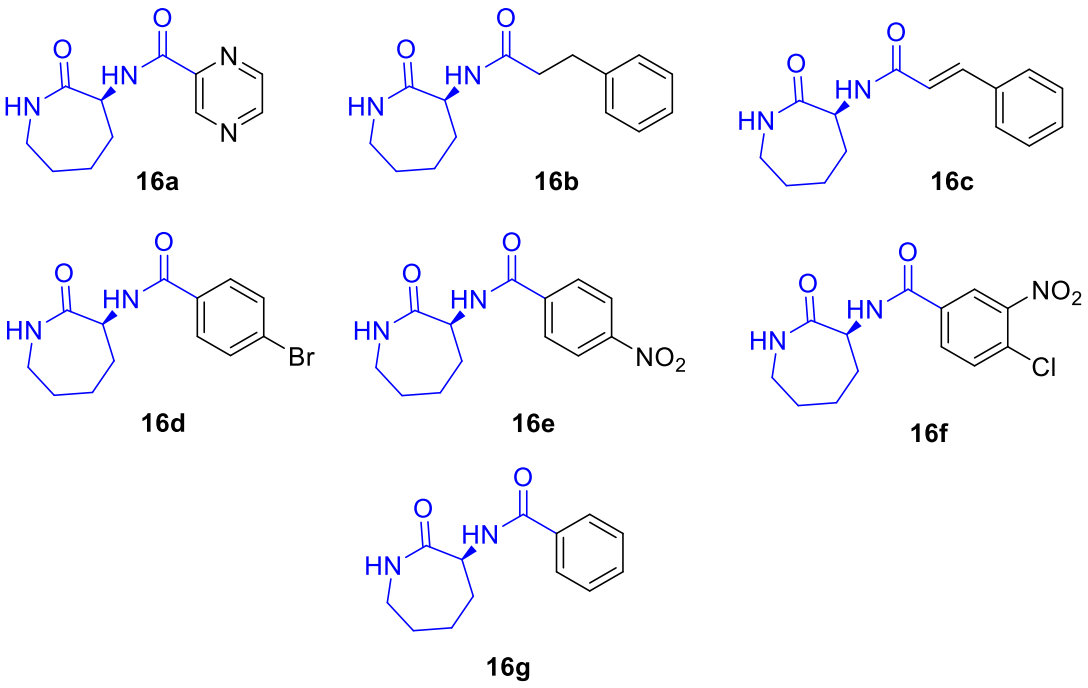


Figure 77 Authentic amide standards **16a-g** used during mass spectrometry analysis (MSA), with the characteristic fragment from L-aminocaprolactam **15a** coupling to an acyl donor highlighted in blue.

The fragmentation patterns of these seven compounds were analysed and two fragments of *m/z* 112 and 84 were found to be common for all seven compounds (**Figure 78**). This allowed rapid screening of a variety of substrates for the presence of these two fragment ions, and once detected the *m/z* and retention time of the corresponding amide standard would be able to be compared to the sample of interest to determine more thoroughly whether the expected amide had been produced.

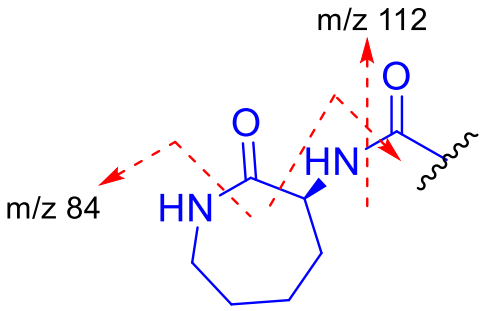


Figure 78 Collision-induced dissociation of the seven amide standards **16a-g** (**Figure 77)** allowed two characteristic *m/z* fragments to be identified (*m/z* 84 and *m/z* 112). These fragments could then be selected for when analysing biotransformation samples, alongside specifying the *m/z* of the expected product.

5.2.2.2 Methyl Ester Substrates Screened

As previously stated, the methyl ester substrate **14a** (**Figure 80**) of CapW is not readily obtainable, yet the amine partner L-aminocaprolactam **15a** is. For this reason, L-aminocaprolactam and a range of commercial methyl ester substrates (**14b-t, Figure 79**) were screened for activity against CapW and using mass spectrometry for an indication of amide product formation.

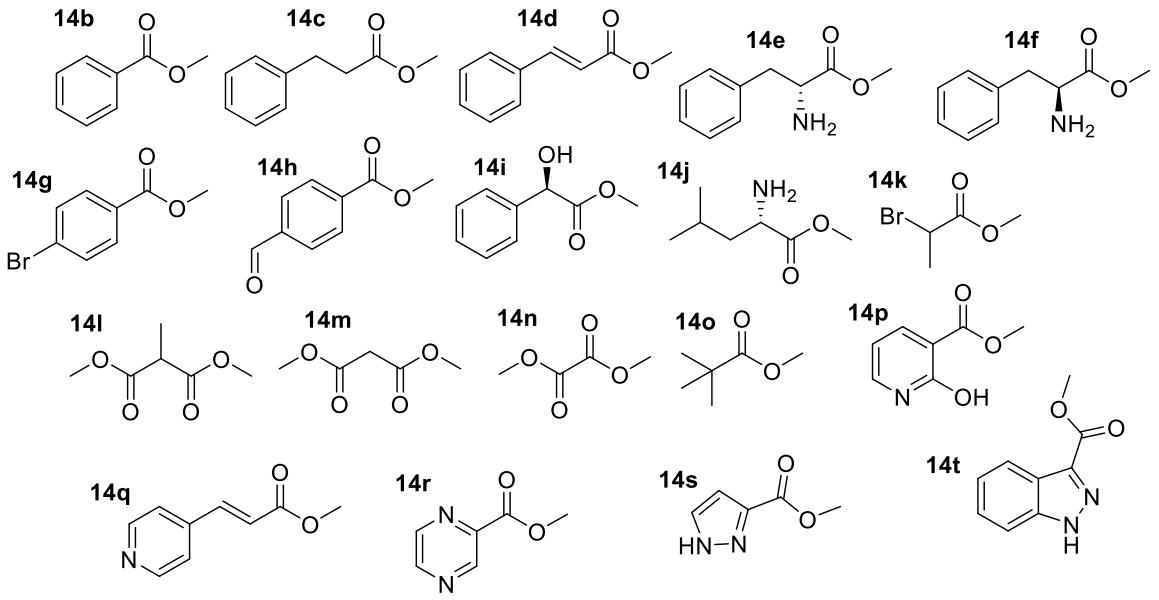


Figure 79 Range of aliphatic and aromatic methyl esters **14b-t** screened against CapW for amide bond formation.

CapW failed to catalyse the coupling of the 19 methyl ester compounds seen in **Figure 79** with L-aminocaprolactam **15a**. The failure of these reactions could be for several reasons. Firstly, the activity of CapW expressed following the methods described in Section 5.2.1 could not be determined because its natural acyl substrate **14a** (**Figure 80**) was not readily available. This means that, since we could not assess the enzyme for its known reaction, we do not know whether or not the soluble protein is properly folded. It may be that CapW expressed in *E. coli* in this study is relatively soluble but not active towards its natural substrates, which would then open up the possibility that the failed reactions reported here are false negatives due to inactivated protein. If the protein used in these reactions was inactive, the panel of methyl ester substrates investigated in this study should not yet be ruled out as potential substrates accepted by active CapW.

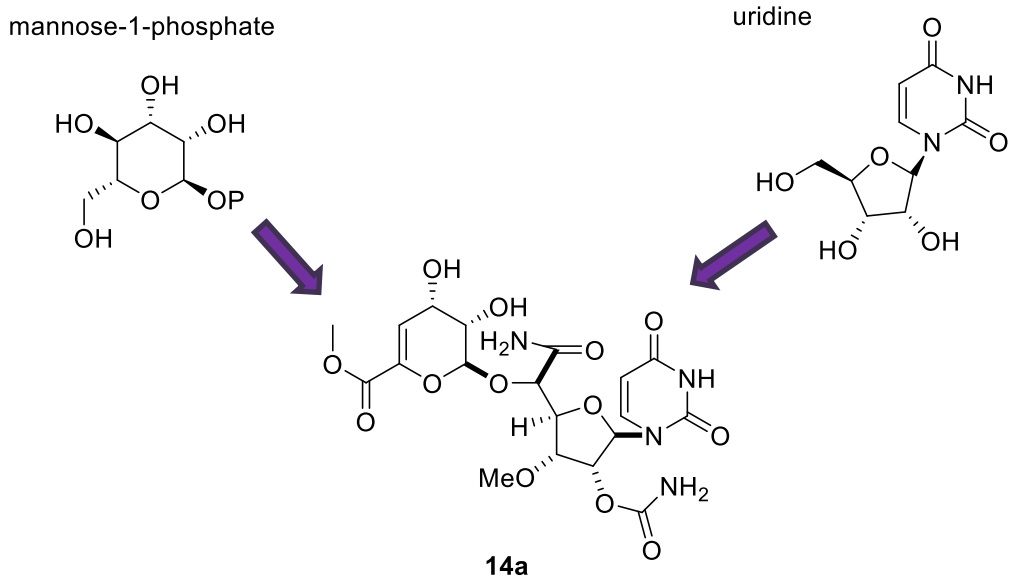


Figure 80 The natural acyl donor substrate **14a** of CapW, highlighting two precursors to its formation.

However, it is also possible that the expression trials (Section 5.2.1) yielded soluble and active CapW protein, and that the methyl esters **14b-t** screened in this initial study are simply not recognised by the enzyme. The acyl substrate **14a** known to be accepted by CapW is shown in **Figure 80**. Two notable precursors to this substrate are mannose and the nucleoside uridine. The two hydroxyl groups derived from mannose-1-phosphate in close proximity to the methyl ester involved in the amide formation of the natural reaction of CapW may play a significant role (via hydrogen bonding, for example) in substrate binding within the active site of the enzyme. It is possible that CapW only accepts substrates that are more hydrophilic than the substrates investigated in this study. Another possibility is that the presence of the uridine derivative found within the natural acyl substrate **14a** is essential for substrate binding. The active site of CapW may contain a nucleotide-binding site which is used to dock the substrate in position, and therefore substrates lacking this motif would not be accepted by the enzyme.

In order to elucidate whether or not the soluble expression of CapW achieved in the heterologous *E. coli* expression strain (demonstrated in Section 5.2.1) yielded active protein, an alternative method of screening, independent of amide bond forming activity, was required.

5.2.2.2.1 Hydrolytic Activity Assay of CapW Using Nitrocefin

Based on sequence similarity CapW was originally proposed to be a putative class C β-lactamase, an enzyme family known to catalyse the hydrolysis of a wide range of β-lactam rings.[128] Despite the fact that CapW was later shown to demonstrate very low levels of hydrolytic activity with respect to the amide bond it is now known to form, it was proposed here that the enzyme may display some residual β-lactamase activity. A routine screen for the presence of β-lactamases employs the use of the cephalosporin Nitrocefin which, upon hydrolysis of the lactam ring, exhibits a colour change from yellow to red (**Figure 81**).[162]

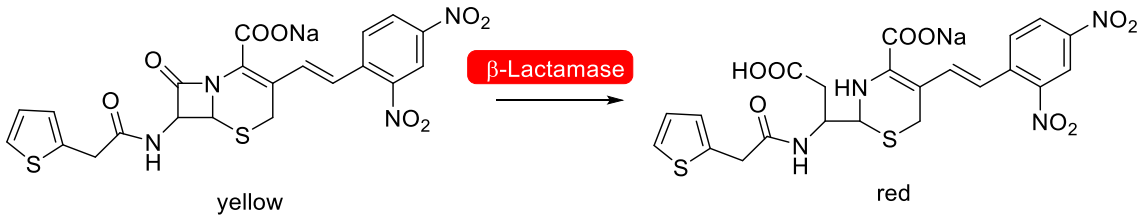


Figure 81 Hydrolysis of β-lactamase substrate Nitrocefin results in a characteristic yellow to red colour change, indicating the presence of an active β-lactamase enzyme.

Plasmid pRSET/CapW contains the gene encoding a β-lactamase to confer resistance to ampicillin and act as a positive selection marker, and so first the CapW gene was subcloned into an alternate kanamycin-resistant expression vector, pET24a. Whole cells expressing empty pRSET, empty pET24a,

pRSET/CapW and pET24a/CapW were then used in an initial screen to assess whether or not CapW demonstrated the ability to hydrolyse β-lactam rings. An equal amount of Nitrocefin was then added to Eppendorf tubes containing aliquots of each of the four cell cultures, and colour change was assessed over a period of 30 minutes (**Figure 82**).

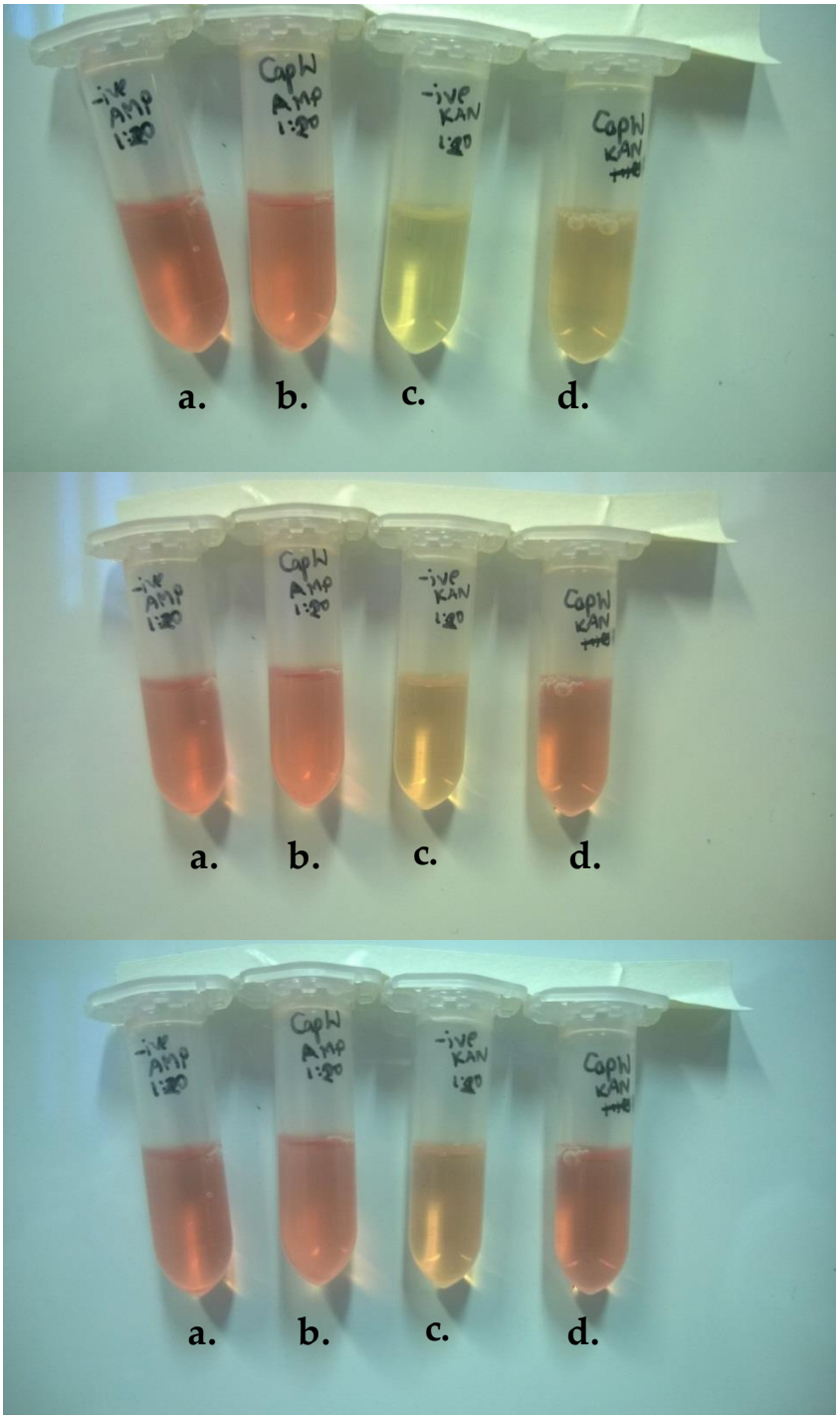


Figure 82 Hydrolysis of Nitrocefin assay after (top) 5 minutes, (middle) 15 minutes and (bottom) 30 minutes. a. Cells containing empty pRSET vector; b. Cells containing pRSET/CapW; c. Cells containing empty pET24a vector; d. Cells containing pET24a/CapW.

As was expected, the two cell cultures containing the β-lactamase gene via an ampicillin resistant vector rapidly developed a red colour upon addition of Nitrocefin, indicating the present of β-lactamase activity. The cell culture containing pET24a/CapW, and so lacking in β-lactamase gene, also developed the red colouring, albeit at a slower rate than the cultures expressing β-lactamase. A comparison between the cell culture containing empty pET24a and the cell culture containing pET24a/CapW suggested that CapW does demonstrate some hydrolytic activity towards the β-lactam ring of the Nitrocefin substrate. The fact that the pET24a/CapW culture displayed increased activity towards the cephalosporin compared to its negative control indicated that the CapW soluble expression obtained in Section 5.2.1 yielded properly folded, and therefore active, protein.

This observation suggesting that CapW was indeed correctly folded and active implies that the methyl esters **14b**-**t** screened for amide bond formation in Section 5.2.2.2 are not suitable candidates for CapW, and so alternative substrates more closely related to the natural acyl donor **14a** should be assessed. However, the indication that properly folded protein was obtained through the expression trials detailed in Section 5.2.1 enables future studies, such as crystallisation of the protein to determine its structure and substrate recognition motifs, to be investigated.

5.3 Preliminary Screen for Cyclisation Reactions Using CAR Adenylation Domain

Recent work by Flitsch *et al.* has demonstrated that activated acyl-AMP intermediates produced by the adenylation domain of a CAR enzyme can be used as substrates in amide bond formation when an amine nucleophile is provided to the system in place of NADPH.[130] Following on from this, it has

been proposed that carboxylic acids containing a terminal amine group may be able to spontaneously cyclise upon activation using ATP, as shown in **Figure 83**.

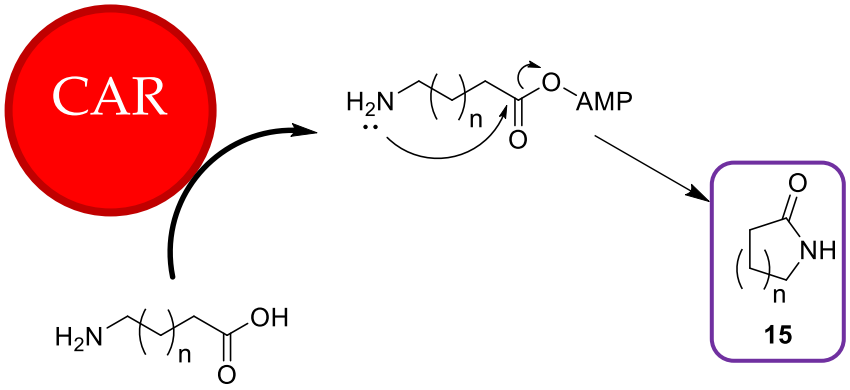


Figure 83 Proposed route to access lactam products through the exploitation of a CAR adenylation domain providing an acyl-AMP intermediate.

Through this spontaneous intramolecular cyclisation reaction, it may be possible to access a range of lactam rings, due to the relatively broad substrate scope displayed by CAR enzymes,[88,91,154] including the cyclisation of L-lysine to yield the natural amine donor substrate for CapW, L-aminocaprolactam **15a**.

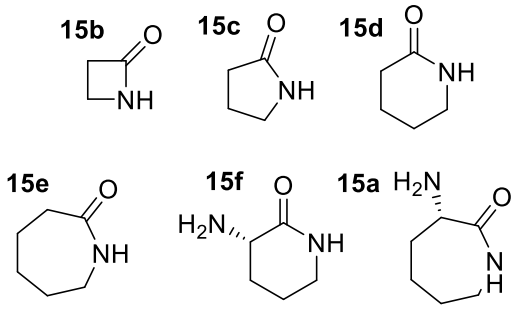


Figure 84 Anticipated lactam products **15a-f** from the activation of carboxylic acids containing a terminal amine group.

To investigate whether or not the acyl-AMP intermediates, produced through the activation of carboxylic acid substrates catalysed by CAR, can undergo spontaneous cyclisation, a panel of linear carboxylic acids of increasing chain lengths were screened for the production of their corresponding lactam **15a-f** (**Figure 84**). A general GC method was developed for the detection of these six compounds, and biotransformations employing either purified *ms*CAR, *tp*CAR, NCAR, *se*CAR or *sr*CAR (Section 2.2) along with ATP and a source of Mg^{2+} ions were extracted after 24 h. Unfortunately, no conversion to product lactam **15a-f** was observed via GC for any of the substrates assayed in this preliminary screen, suggesting that either the reaction conditions were not suitable to encourage cyclisation, or that the substrates were not recognised by the CAR enzymes. For this reason, future experiments must focus on varying certain reaction parameters, such as buffer pH, temperature and reaction time lengths, to accurately gauge whether or not these substrates are suitable candidates for adenylation mediated by CAR.

5.4 Conclusions and Outlook

Soluble recombinant expression of the ATP-independent *N*-acyltransferase CapW had not previously been reported in *E. coli*. A series of expression trials were executed, with various parameters such as expression temperature, expression time length, solubility enhancers, and host strain being assessed with respect to the extent of soluble protein expression. It was observed that the solubility enhancers, sorbitol and betaine, led to a significantly enhanced amount of soluble protein, but at the expense of bacterial culture growth rate. Eventually, acceptable levels of soluble CapW expression were achieved for the first time through the use of C41 (DE3) cells harbouring pRSET/CapW, with cells being left to express protein for 4 h at 25 °C. The known methyl ester substrate **14a** used by

CapW for the *N*-acylation of L-aminocaprolactam **15a** was not readily obtainable, and so CapW was screened for amide bond forming activity against a panel of commercially available methyl ester substrates **14b**-**t**. Since a positive control reaction using the natural reaction of CapW was not available, the collision-induced dissociation (CID) fragmentation patterns of a range of seven authentic amide standards **16a**-**g** were determined through mass spectrometry, and two characteristic *m*/*z* values were identified. Biotransformations using methyl ester **14b**-**t** substrates were studied for the appearance of these two peaks, but no conversion to amide product was detected for any of the reactions tested. In an effort to determine whether the soluble protein achieved through the expression trials was properly folded and active, the hydrolytic activity of CapW against known β-lactamase substrate Nitrocefin was assessed. CapW was subcloned into a kanamycin-resistant vector and whole cells containing either an empty vector control or CapW were studied for the characteristic yellow to red colour change upon the addition of Nitrocefin. This investigation revealed that CapW demonstrates some hydrolytic activity towards the β-lactam ring of Nitrocefin, suggesting that the protein was properly folded during its soluble expression in C41 (DE3) cells. It is not currently possible to ascertain whether or not CapW is able to be exploited for the production of a wider range of amide bond-containing products, but this study suggested that the methyl ester substrates **14b**-**t** screened were not recognised by the enzyme. This was potentially due to the fact that the natural acyl donor substrate of CapW, methyl ester **14a**, is a complex molecule derived from mannose and uridine, leading to the possibility that these motifs must be present on a substrate prior to an effective reaction utilising CapW.

It was recently reported that the acyl-AMP intermediates produced from the adenylation of carboxylic acids by CAR adenylation domains can, in the absence of NADPH, be intercepted by a nucleophile to achieve amide bond formation.[130]

Preliminary studies to investigate whether it was also possible to encourage head-to-tail cyclisation of a carboxylic acid containing a terminal amine moiety were thus undertaken, with six suitable carboxylic acid substrates against five CAR homologs being evaluated. No production of any of the expected lactam products **15a-f** was detected upon GC analysis, suggesting that these substrates may not be suitable candidates for CAR. Following the recent elucidation of a CAR adenylation-PCP didomain,[89] it will now be possible to computationally model potential CAR substrates prior to biotransformations and determine the likelihood of adenylation. However, the lack of success with this initial study may also be due to the reaction conditions used for the biotransformations. It is possible that the use of a higher pH may encourage spontaneous cyclisation by ensuring that the terminal amine is sufficiently deprotonated, for example. For this reason, additional experiments with respect to the reaction setup should be executed before a conclusive result can be obtained.

6 Materials and Methods

6.1 Materials, Equipment and General Analytical Methods

6.1.1 Chemicals

Commercial reagents and solvents were purchased from Sigma Aldrich, Alfa Aesar, TCI Chemicals, Bachem or Fluorochem and used without further purification.

6.1.2 Equipment and General Analytical Methods

NMR spectra were recorded using a Bruker Avance 400 spectrometer with chemical shifts reported in ppm relative to residual protic solvent signals ($CHCl_3$ in $CDCl_3$, ^1H = 7.27; $CDCl_3$, ^{13}C = 77.0; CHD_2OD in CD_3OD, ^1H = 3.31; CD_3OD, ^{13}C = 49.0; CHD_2SOCD_3 in $(CD_3)_2SO$, ^1H = 2.50; $(CD_3)_2SO$, ^{13}C = 39.52).The coupling constants (*J*) are quoted in Hz to the nearest 0.1 Hz. Signal multiplicities are assigned as singlet (s), doublet (d), triplet (t), quartet (q), quintet (quint), sextet (sxt), multiplet (m), broad (br) or a combination of the above. Low resolution mass spectrometry (MS) was performed on a HP-6890 GC connected to a HP5973 MS detector.

Normal phase HPLC was performed on an Agilent system equipped with a G1379A degasser, G1312A binary pump, a G1367A well plate autosampler unit, a G1316A temperature controlled column compartment and a G1315C diode array detector. CHIRALPAK®IC analytical column was purchased from Daicel (Osaka, Japan). The column possesses dimensions of 250 mm length, 4.6 mm diameter, 5 μm particle size. An injection volume of 10 μL was used and chromatograms were monitored at 265 nm.

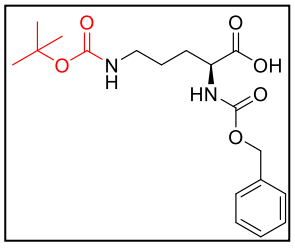
Gas chromatography was performed on an Agilent 6850 GC system using an Agilent Technologies 6850 Series Auto Sampler for injections and equipped with a flame ionisation detector (FID) for detection, using a 25 m CP-Chirasil-DEX CB column with 0.25 mm inner diameter and 0.25 μm film thickness (Agilent, Santa Clara, CA, USA). Helium was used as the carrier gas (1.2 mL min^{-1}). Derivatisation of samples for chiral GC-FID analysis was achieved using acetic anhydride and an excess of triethylamine at room temperature, where stated.

GCMS analysis was performed on a HP-6890 Series GC coupled to a HP5973 MS detector, EI positive mode with helium as the carrier gas.

Analysis and determination of ee for amines **5b-f** based on previously reported HPLC or GC-FID analysis on a chiral stationary phase.[40,105,106]

6.2 Synthesis of Substrates for CAR-IRED, ADH-IRED and GOase-IRED Cascade Reactions

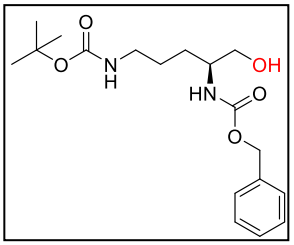
6.2.1 Synthesis of N^{α}-Cbz-N^{δ}-Boc-L-ornithine, Cbz-L-Orn(Boc)-OH, **8c**



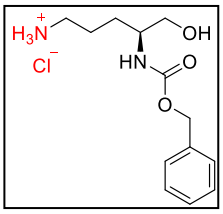
To a solution of N^{α}-Cbz-L-ornithine **8d** (2.0 g, 7.51 mmol) in H_2O and THF (200 mL, 1:1), NaOH (3 mL of 10 M stock, 0.015 M) was added, followed by boc anhydride (1.64 g, 7.51 mmol). The reaction mixture was stirred at room temperature overnight, acidified to pH 2.0 using 1 M HCl and extracted into EtOAc (3 x 100 mL). The organic layers were combined, dried over anhydrous $MgSO_4$ and the solvent removed under reduced pressure to yield the title compound **8c** in 94.5 % yield (2.75 g, white solid): **^{1}H-NMR** δH (400 MHz, $CDCl_3$) 7.37-7.32 (m, 5H), 7.00 (s, 1H), 5.12 (s, 2H), 4.51 (br s, 1H), 3.21-3.01 (m, 2H), 1.78-

1.65 (m, 2H), 1.62-1.37 (m, 11H); **^{13}C-NMR δC** (100 MHz $CDCl_3$) 175.5, 156.4, 136.3, 128.5, 128.2, 128.1, 80.0, 67.2, 53.6, 40.1, 29.9, 28.7, 25.9.

6.2.2 Synthesis of N^{α}-Cbz-L-ornithinol, Cbz-L-Orn-Ol, **12d**



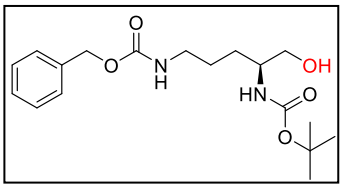
To a solution of N^{α}-Cbz-N^{δ}-Boc-L-ornithine **8c** (1.5 g, 4.09 mmol) in dry THF (30 mL), triethylamine (570 µL, 1 eq.) was added at room temperature under nitrogen. Ethyl chloroformate (390 µL, 1 eq.) was then added dropwise at 0 °C. After 30 m, $NaBH_4$ (464 mg, 3 eq.) and MeOH (45 mL, slow addition) were added successively at 0 °C. The reaction mixture was stirred for further 1 h at room temperature, and then concentrated under reduced pressure. The residue was re-dissolved in a mixture of H_2O/EtOAc (50 mL, 1:1), the organic phase was sequentially washed with brine (30 mL), dried over $MgSO_4$, filtered, and concentrated under reduced pressure to yield a colourless gel (91 % crude yield). The crude product (1.31 g) was purified over silica using flash chromatography (DCM/MeOH: 98/2), to give the compound **12b** in 71% yield (1 g, white solid): **^{1}H-NMR** δH (400 MHz, $CDCl_3$) 7.35-7.30 (m, 5H), 5.10 (br s, 1H), 5.09 (s, 2H), 4.63 (br s, 1H), 3.70-3.69 (m, 1H), 3.66-3.59 (m, 2H), 3.11 (s, 2H), 2.18 (br s, 1H), 1.56-1.43 (m, 13H); **^{13}C-NMR δC** (100 MHz $CDCl_3$) 156.8, 156.3, 136.5, 128.7, 128.3, 128.2, 67.0, 65.3, 53.0, 40.4, 28.5, 26.9.



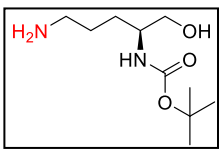
Nα-Cbz-Nδ-Boc-L-ornithinol hydrochloride **12b** (1 g, 2.84 mmol) was suspended in HCl (2 M in Et_2O, 30 mL) at room temperature, followed by the addition of DCM (15 mL). The reaction mixture was stirred overnight at room temperature, and subsequently filtered off, washed with Et_2O (2 x 10 mL), and dried under vacuum to give the title compound **12d** in 84% yield (685 mg, white powder): **^{1}H-**

NMR δH (400 MHz, CD_3OD) 7.46-7.40 (m, 5H), 5.10 (s, 2H), 3.65-3.62 (m, 1H), 3.59-3.50 (m, 2H), 3.02-2.96 (m, 2H), 1.75-1.60 (m, 3H), 1.47-1.42 (m, 1H); **^{13}C-NMR δC** (100 MHz CD_3OD) 158.4, 136.5, 128.8, 128.3, 127.6, 66.9, 63.6, 52.3, 39.1, 27.4, 23.3.

6.2.3 Synthesis of N^{α}-Boc-L-ornithinol, Boc-L-Orn-Ol, **12e**



To a solution of N^{α}-Boc-N^{δ}-Cbz-L-ornithine **8b** (1.5 g, 4.09 mmol) in dry THF (30 mL), triethylamine (570 µL, 1 eq.) was added at room temperature under nitrogen. Ethyl chloroformate (390 µL, 1 eq.) was then added dropwise at 0 °C. After 30 m, $NaBH_4$ (464 mg, 3 eq.) and MeOH (45 mL, slow addition) were added successively at 0 °C. The reaction mixture was stirred for further 1 h at room temperature, and then concentrated under reduced pressure. The residue was re-dissolved in a mixture of H_2O/EtOAc (50 mL, 1:1), the organic phase was sequentially washed with brine (30 mL), dried over $MgSO_4$, filtered, and concentrated under reduced pressure to quantitatively yield a colourless amorphous solid. The crude product (1.45 g) was purified over silica using flash chromatography (DCM/MeOH: 98/2), to give the title compound **12c** in 84% yield (1.44 g, colourless gel): **^{1}H-NMR** δH (400 MHz, $CDCl_3$) 7.35-7.30 (m, 5H), 5.10 (br s, 1H), 5.09 (s, 2H), 4.63 (br s, 1H), 3.70-3.69 (m, 1H), 3.66-3.59 (m, 2H), 3.11 (s, 2H), 2.18 (br s, 1H), 1.56-1.43 (m, 13H); **^{13}C-NMR δC** (100 MHz $CDCl_3$) 156.8, 156.3, 136.5, 128.7, 128.3, 128.2, 67.0, 65.3, 53.0, 40.4, 28.5, 26.9.



N^{α}-Boc-N^{δ}-Cbz-L-ornithinol **12c** (1.2 g, 3.41 mmol) was dissolved in dry MeOH (25 mL) under nitrogen. Pd/C (100 mg) was added, and the reaction was stirred overnight at room temperature under hydrogen. The product was then filtered through Celite®, washed with

MeOH (3 x 20 mL), and the solvent was removed under reduced pressure to yield the title compound **12e** in quantitative yield (747 mg, colourless gel): **^{1}H-NMR** δH (400 MHz, $CDCl_3$) 3.61-3.51 (m, 2H), 2.82-2.75 (app br s, 1H), 2.75-2.69 (m, 1H), 1.61-1.38 (m, 4H), 1.42 (m, 9H); **^{13}C-NMR δC** (100 MHz $CDCl_3$) 156.3, 79.5, 64.8, 52.7, 41.5, 28.8, 28.7, 28.4.

6.3 Molecular Biology, Biocatalyst Preparation and Protein Expression

6.3.1 Nucleotide Sequences

6.3.1.1 Entry Vector pPB01

GGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAA
TGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCA
ACGCAATTAATGTGAGTTAGCGCGAATTGATCTGGTTTGACAGCTTATCA
TCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGC
TGTGGTATGGCTGTGCAGGTCGTAAATCACTGCATAATTCGTGTCGCTCA
AGGCGCACTCCCGTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGTT
CTGGCAAATATTCTGAAGAATTCTGTTGACAATTAATCATCCGGCTCGTAT
AATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCGCC
GCTGAGAAAAAGCGAAGCGGCACTGCTCTTTAACAATTTATCAGACAAT
CTGTGTGGGCACTCGACCGGAATTATCGATTAACTTTATTATTAAAAATTA
AAGAGGTATATATTAATGTATCGATTAAATAAGGAGGAATAAACCATGG
GGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGA
CAGCAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCGATGGG
GATCCGAGCTCGAGATCTGCAGCTGGTACCATATGGGAATTCGAAGCTTG
GCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCA
GAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCG
CGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAG

CGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAG

GCATCAAATAAAACGAAAGGCTCAGTCGAAAGCTGCAGCCTTTCGTTTTA

TCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGA

GCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGAC

GCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGAC

GGATGGCCTTTTTGCGTTTCTACAAACTCTTTTTGTTTATTTTTCTAAATAC

ATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAA

TAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCT

TATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACG

CTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTT

ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCC

GAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGC

GGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATAC

ACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCAT

CTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCAT

GAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCG

AAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCT

TGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGT

GACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAA

CTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATG

GAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGG

CTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTA

TCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATC

TACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCG

CTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTT

TACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGG

ATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGT

GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATC

TTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA

ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC

TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTC

CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACC

GCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGG

CGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATA

AGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTT

GGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGA

GAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTA

AGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGA

AACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG

CGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGC

CAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCA

CATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCC

TTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCG

AGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTT

ACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAAT

CTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGT

GACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGC

CCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACC

GTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACG

CGCGAGGCAGCAGATCAATTCGCGCGCGAAGGCGAAGCGGCATGCATTT

ACGTTGACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAG

CGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGT

TATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCG

TGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGA

AGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAA

CTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGC

CCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATC

AACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGA

AGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGG
CTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGC
TGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACC
CATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGG
AGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCA
TTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTC
ACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTG
CCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTT
CCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCG
CGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGG
GATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTCAACCACC
ATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCT
GCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCT
CACTGGTGAAAAGAAAAACCACCCT

6.3.1.2 BioBrick Prefix and Suffix

BioBrick prefix DNA sequence:

GAATTCGAGCAGTGTCTCTAGAAGAGACGTAC
EcoRI *XbaI*

BioBrick suffix DNA sequence:

ACTGGGCCTTTCGTTTTATCTGACTAGTTAGCATCGTTCACTGCAG
SpeI *PstI*

6.3.1.3 Operon Designs

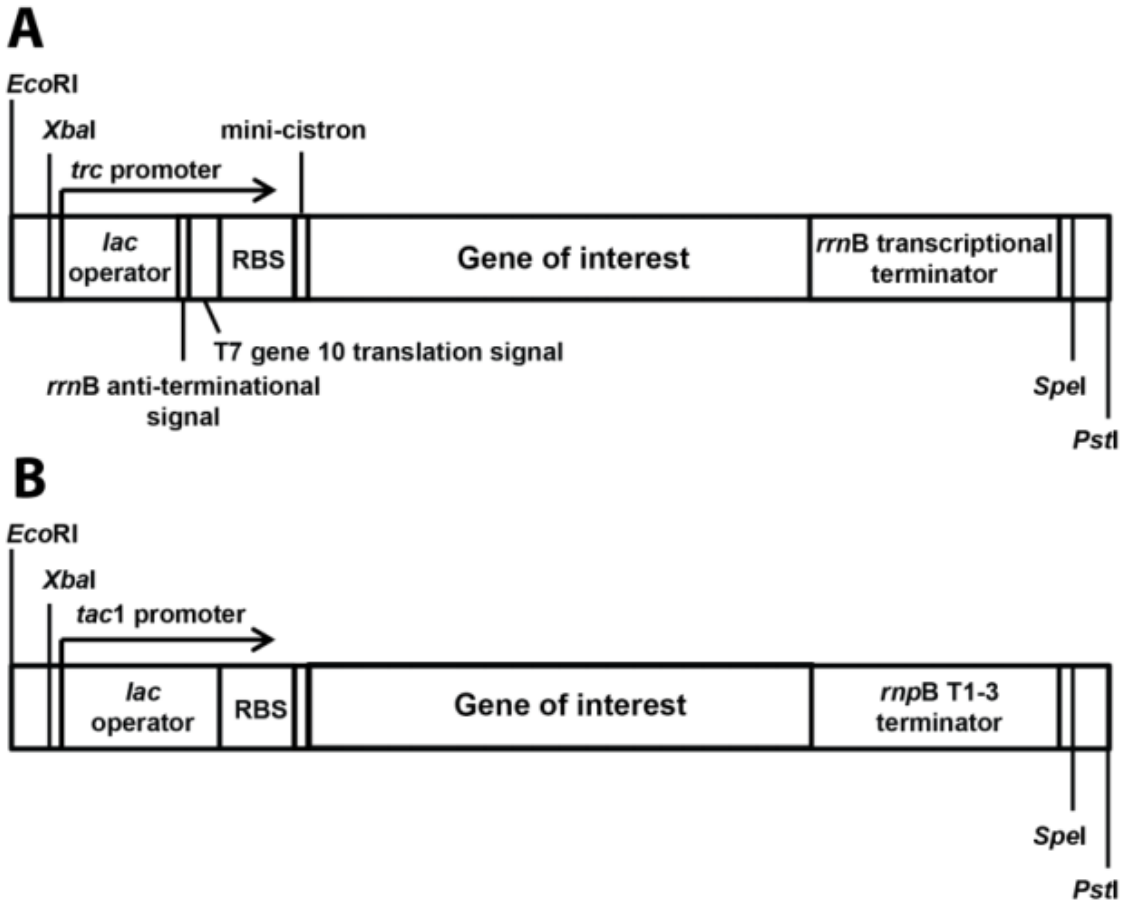


Figure 85 Operon designs for the genes used in this study. A) Operon design used for ATA-117, BsSfp. B) Operon design used for MCAR, (R)-IRED.

6.3.1.4 Operons and Genes

6.3.1.4.1 MCAR

MCAR operon sequence:

GAATTCGAGCAGTGTCTCTAGAAGAGACGTACGAGCTGTTGACAATTAAT

CATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACA

GGAAACAGAATTGTAGAGGAGATATACATATGCACCATCATCATCATCA

TTCTTCTGGTAGCCCGATTACCCGTGAAGAACGTCTGGAACGTCGTATTCA

GGATCTGTATGCGAACGATCCGCAGTTCGCAGCAGCCAAACCGGCGACC

GCGATTACCGCGGCGATTGAACGTCCGGGTCTGCCGCTGCCGCAGATCAT

CGAAACGGTGATGACCGGCTATGCGGATCGTCCGGCACTGGCACAACGT

AGCGTGGAATTTGTGACCGATGCGGGCACCGGTCATACCACCCTGCGTCT

GCTGCCGCATTTTGAAACCATTAGCTATGGCGAACTGTGGGATCGTATTA

GCGCGCTGGCCGATGTTCTGAGCACCGAACAGACCGTGAAACCGGGCGA

TCGTGTGTGCCTGCTGGGCTTTAACAGCGTGGATTATGCGACCATTGATAT

GACCCTGGCACGTCTGGGTGCTGTCGCTGTCCCGCTGCAAACCTCTGCTG

CGATTACCCAGCTGCAACCGATTGTGGCGGAAACCCAGCCGACCATGAT

TGCGGCGAGCGTGGATGCCCTGGCCGATGCGACCGAACTGGCACTGAGT

GGTCAAACGGCTACGCGTGTGCTGGTGTTTGATCATCATCGTCAGGTGGA

TGCGCATCGTGCGGCGGTTGAAAGCGCGCGTGAACGTCTGGCCGGTAGC

GCGGTGGTTGAAACCCTGGCCGAAGCGATTGCGCGTGGTGATGTGCCGCG

TGGTGCGAGCGCGGGTAGCGCACCGGGCACCGATGTGAGCGATGATAGC

CTGGCCCTGCTGATTTATACCTCTGGTAGTACGGGTGCGCCGAAAGGCGC

CATGTATCCGCGTCGTAACGTGGCGACCTTTTGGCGTAAACGTACCTGGTT

TGAAGGCGGCTATGAACCGAGCATTACCCTGAACTTTATGCCGATGAGCC

ATGTGATGGGCCGTCAGATTCTGTATGGCACCCTGTGCAACGGCGGCACC

GCGTATTTTGTGGCGAAAAGCGATCTGAGCACCCTGTTTGAAGATCTGGC

CCTGGTGCGTCCGACCGAACTGACCTTCGTCCCGCGTGTTTGGGATATGGT

GTTCGATGAATTTCAGAGCGAAGTGGATCGTCGTCTGGTGGATGGCGCGG

ATCGTGTTGCGCTGGAAGCGCAGGTGAAAGCGGAAATTCGTAACGATGT

GCTGGGCGGTCGTTATACCTCTGCTCTGACGGGTTCTGCTCCGATTAGCGA

TGAAATGAAAGCGTGGGTGGAAGAACTGCTGGATATGCATCTGGTGGAA

GGCTATGGCAGCACCGAAGCGGGCATGATTCTGATTGATGGCGCGATTCG

TCGTCCGGCGGTGCTGGATTATAAACTGGTGGATGTTCCGGATCTGGGCT

ATTTTCTGACCGATCGTCCGCATCCGCGTGGCGAACTGCTGGTGAAAACC

GATAGCCTGTTTCCGGGCTATTATCAGCGTGCGGAAGTGACCGCGGATGT

GTTTGATGCGGATGGCTTTTATCGCACCGGCGATATTATGGCGGAAGTGG

GCCCGGAACAGTTTGTGTATCTGGATCGTCGTAACAACGTGCTGAAACTG

AGCCAGGGCGAATTTGTTACCGTGAGCAAACTGGAAGCGGTGTTTGGCG

ATAGCCCGCTGGTGCGTCAGATTTATATTTATGGCAACAGCGCGCGTGCG

TATCTGCTGGCCGTGATTGTGCCGACCCAGGAAGCGCTGGACGCGGTCCC

GGTTGAAGAACTGAAAGCGCGTCTGGGTGACTCTCTGCAAGAAGTGGCG

AAAGCGGCGGGTCTGCAAAGCTATGAAATTCCGCGCGATTTTATTATCGA

AACCACCCCGTGGACCCTGGAAAACGGCCTGCTGACGGGTATTCGTAAA

CTGGCCCGTCCGCAGCTGAAAAAACATTATGGTGAACTGCTGGAACAAA

TTTATACCGATCTGGCCCACGGCCAGGCGGATGAACTGCGTAGCCTGCGT

CAGAGCGGTGCGGATGCGCCGGTGCTGGTGACCGTTTGTCGTGCGGCTGC

GGCTCTGCTGGGTGGTAGCGCGAGCGATGTGCAGCCGGATGCGCATTTCA

CCGATCTGGGTGGTGATAGCCTGAGCGCCCTGAGCTTTACCAACCTGCTG

CATGAAATCTTTGATATTGAAGTGCCGGTGGGCGTGATTGTGAGCCCGGC

GAACGATCTGCAAGCGCTGGCCGATTATGTGGAAGCGGCGCGTAAACCG

GGTAGCAGCCGTCCGACCTTTGCGAGCGTGCATGGCGCGAGCAACGGCC

AGGTGACCGAAGTGCATGCGGGCGATCTGAGCCTGGATAAATTTATTGAT

GCGGCGACCCTGGCCGAAGCCCCGCGTCTGCCGGCTGCAAATACCCAGG

TGCGTACCGTGCTGCTGACCGGTGCGACCGGCTTTCTGGGCCGTTACCTG

GCCCTGGAATGGCTGGAACGTATGGATCTGGTTGATGGCAAACTGATTTG

CCTGGTGCGTGCCAAAAGCGATACCGAAGCGCGTGCGCGTCTGGATAAA

ACCTTTGATAGCGGCGATCCGGAACTGCTGGCCCATTATCGTGCGCTGGC

CGGCGATCATCTGGAAGTGCTGGCCGGTGATAAAGGCGAAGCGGATCTG

GGCCTGGATCGTCAGACCTGGCAACGCCTGGCAGATACCGTGGATCTGAT

TGTTGACCCGGCTGCCCTGGTGAATCATGTGCTGCCGTATAGCCAGCTGTT

TGGCCCGAATGCGCTGGGCACCGCTGAACTGCTGCGCCTGGCTCTGACCA

GCAAAATTAAACCGTATAGCTACACCAGCACCATTGGCGTGGCGGATCA

GATTCCGCCGAGCGCGTTTACCGAAGATGCGGATATTCGTGTGATTAGCG

CGACCCGTGCGGTGGATGATAGCTATGCGAACGGCTATAGCAACAGCAA
ATGGGCGGGTGAAGTGCTGCTGCGTGAAGCGCATGATCTGTGCGGTCTGC
CGGTGGCGGTGTTTCGTTGCGATATGATCCTGGCAGACACGACCTGGGCG
GGTCAGCTGAACGTGCCGGATATGTTTACCCGTATGATTCTGTCTCTGGCA
GCTACGGGTATCGCACCGGGTAGCTTTTATGAACTGGCCGCGGATGGTGC
GCGTCAGCGTGCGCATTATGATGGCCTGCCGGTGGAATTTATTGCGGAAG
CGATTAGCACCCTGGGCGCGCAGAGCCAGGATGGCTTTCATACCTATCAT
GTGATGAATCCGTATGATGATGGCATTGGCCTGGATGAATTTGTGGATTG
GCTGAACGAAAGCGGCTGCCCGATTCAGCGTATTGCGGATTATGGCGATT
GGCTGCAACGTTTTGAAACCGCGCTGCGCGCTCTGCCGGATCGTCAGCGT
CATAGCAGCCTGCTGCCGCTGCTGCATAACTATCGTCAGCCGGAACGTCC
GGTGCGTGGTAGCATTGCGCCGACCGATCGCTTTCGTGCGGCCGTGCAGG
AAGCGAAAATTGGCCCGGATAAAGATATTCCGCATGTGGGTGCGCCGAT
TATTGTGAAATATGTGAGCGATCTGCGCCTGCTGGGCCTGCTGTAACCGG
CTTATCGGTCAGTTTCACCTGATTTACGTAAAAACCCGCTTCGGCGGGTTT
TTGCTTTTGGAGGGGCAGAAAGATGAATGACTGTCCACGACGCTATACCC
AAAAGAAAGCGGCTTATCGGTCAGTTTCACCTGGTTTACGTAAAAACCCG
CTTCGGCGGGTTTTTGCTTTTGGAGGGGCAGAAAGATGAATGACTGTCCA
CGACACTATACCCAAAAGAAAGCGGCTTATCGGTCAGTTTCACCTGTTTT
ACGTAAAAACCCGCTTCGGCGGGTTTTTACTTTTGGAGGGGCAGAAAGAT
GAATGACTGTCCACGACACTATACCCAAAAGAAAACTGGGCCTTTCGTTT
TATCTGACTAGTTAGCATCGTTCACTGCAG

MCAR open reading frame (ORF) sequence [gene with N-terminal His_6 tag]:

ATGCACCATCATCATCATCATTCTTCTGGTAGCCCGATTACCCGTGAAGA
ACGTCTGGAACGTCGTATTCAGGATCTGTATGCGAACGATCCGCAGTTCG
CAGCAGCCAAACCGGCGACCGCGATTACCGCGGCGATTGAACGTCCGGG

TCTGCCGCTGCCGCAGATCATCGAAACGGTGATGACCGGCTATGCGGATC

GTCCGGCACTGGCACAACGTAGCGTGGAATTTGTGACCGATGCGGGCAC

CGGTCATACCACCCTGCGTCTGCTGCCGCATTTTGAAACCATTAGCTATGG

CGAACTGTGGGATCGTATTAGCGCGCTGGCCGATGTTCTGAGCACCGAAC

AGACCGTGAAACCGGGCGATCGTGTGTGCCTGCTGGGCTTTAACAGCGTG

GATTATGCGACCATTGATATGACCCTGGCACGTCTGGGTGCTGTCGCTGTC

CCGCTGCAAACCTCTGCTGCGATTACCCAGCTGCAACCGATTGTGGCGGA

AACCCAGCCGACCATGATTGCGGCGAGCGTGGATGCCCTGGCCGATGCG

ACCGAACTGGCACTGAGTGGTCAAACGGCTACGCGTGTGCTGGTGTTTGA

TCATCATCGTCAGGTGGATGCGCATCGTGCGGCGGTTGAAAGCGCGCGTG

AACGTCTGGCCGGTAGCGCGGTGGTTGAAACCCTGGCCGAAGCGATTGC

GCGTGGTGATGTGCCGCGTGGTGCGAGCGCGGGTAGCGCACCGGGCACC

GATGTGAGCGATGATAGCCTGGCCCTGCTGATTTATACCTCTGGTAGTAC

GGGTGCGCCGAAAGGCGCCATGTATCCGCGTCGTAACGTGGCGACCTTTT

GGCGTAAACGTACCTGGTTTGAAGGCGGCTATGAACCGAGCATTACCCTG

AACTTTATGCCGATGAGCCATGTGATGGGCCGTCAGATTCTGTATGGCAC

CCTGTGCAACGGCGGCACCGCGTATTTTGTGGCGAAAAGCGATCTGAGCA

CCCTGTTTGAAGATCTGGCCCTGGTGCGTCCGACCGAACTGACCTTCGTCC

CGCGTGTTTGGGATATGGTGTTCGATGAATTTCAGAGCGAAGTGGATCGT

CGTCTGGTGGATGGCGCGGATCGTGTTGCGCTGGAAGCGCAGGTGAAAG

CGGAAATTCGTAACGATGTGCTGGGCGGTCGTTATACCTCTGCTCTGACG

GGTTCTGCTCCGATTAGCGATGAAATGAAAGCGTGGGTGGAAGAACTGCT

GGATATGCATCTGGTGGAAGGCTATGGCAGCACCGAAGCGGGCATGATT

CTGATTGATGGCGCGATTCGTCGTCCGGCGGTGCTGGATTATAAACTGGT

GGATGTTCCGGATCTGGGCTATTTTCTGACCGATCGTCCGCATCCGCGTGG

CGAACTGCTGGTGAAAACCGATAGCCTGTTTCCGGGCTATTATCAGCGTG

CGGAAGTGACCGCGGATGTGTTTGATGCGGATGGCTTTTATCGCACCGGC

GATATTATGGCGGAAGTGGGCCCGGAACAGTTTGTGTATCTGGATCGTCG

TAACAACGTGCTGAAACTGAGCCAGGGCGAATTTGTTACCGTGAGCAAA

CTGGAAGCGGTGTTTGGCGATAGCCCGCTGGTGCGTCAGATTTATATTTAT
GGCAACAGCGCGCGTGCGTATCTGCTGGCCGTGATTGTGCCGACCCAGGA
AGCGCTGGACGCGGTCCCGGTTGAAGAACTGAAAGCGCGTCTGGGTGAC
TCTCTGCAAGAAGTGGCGAAAGCGGCGGGTCTGCAAAGCTATGAAATTC
CGCGCGATTTTATTATCGAAACCACCCCGTGGACCCTGGAAAACGGCCTG
CTGACGGGTATTCGTAAACTGGCCCGTCCGCAGCTGAAAAAACATTATGG
TGAACTGCTGGAACAAATTTATACCGATCTGGCCCACGGCCAGGCGGAT
GAACTGCGTAGCCTGCGTCAGAGCGGTGCGGATGCGCCGGTGCTGGTGA
CCGTTTGTCGTGCGGCTGCGGCTCTGCTGGGTGGTAGCGCGAGCGATGTG
CAGCCGGATGCGCATTTCACCGATCTGGGTGGTGATAGCCTGAGCGCCCT
GAGCTTTACCAACCTGCTGCATGAAATCTTTGATATTGAAGTGCCGGTGG
GCGTGATTGTGAGCCCGGCGAACGATCTGCAAGCGCTGGCCGATTATGTG
GAAGCGGCGCGTAAACCGGGTAGCAGCCGTCCGACCTTTGCGAGCGTGC
ATGGCGCGAGCAACGGCCAGGTGACCGAAGTGCATGCGGGCGATCTGAG
CCTGGATAAATTTATTGATGCGGCGACCCTGGCCGAAGCCCCGCGTCTGC
CGGCTGCAAATACCCAGGTGCGTACCGTGCTGCTGACCGGTGCGACCGG
CTTTCTGGGCCGTTACCTGGCCCTGGAATGGCTGGAACGTATGGATCTGGT
TGATGGCAAACTGATTTGCCTGGTGCGTGCCAAAAGCGATACCGAAGCG
CGTGCGCGTCTGGATAAAACCTTTGATAGCGGCGATCCGGAACTGCTGGC
CCATTATCGTGCGCTGGCCGGCGATCATCTGGAAGTGCTGGCCGGTGATA
AAGGCGAAGCGGATCTGGGCCTGGATCGTCAGACCTGGCAACGCCTGGC
AGATACCGTGGATCTGATTGTTGACCCGGCTGCCCTGGTGAATCATGTGC
TGCCGTATAGCCAGCTGTTTGGCCCGAATGCGCTGGGCACCGCTGAACTG
CTGCGCCTGGCTCTGACCAGCAAAATTAAACCGTATAGCTACACCAGCAC
CATTGGCGTGGCGGATCAGATTCCGCCGAGCGCGTTTACCGAAGATGCGG
ATATTCGTGTGATTAGCGCGACCCGTGCGGTGGATGATAGCTATGCGAAC
GGCTATAGCAACAGCAAATGGGCGGGTGAAGTGCTGCTGCGTGAAGCGC
ATGATCTGTGCGGTCTGCCGGTGGCGGTGTTTCGTTGCGATATGATCCTGG
CAGACACGACCTGGGCGGGTCAGCTGAACGTGCCGGATATGTTTACCCGT

ATGATTCTGTCTCTGGCAGCTACGGGTATCGCACCGGGTAGCTTTTATGAA
CTGGCCGCGGATGGTGCGCGTCAGCGTGCGCATTATGATGGCCTGCCGGT
GGAATTTATTGCGGAAGCGATTAGCACCCTGGGCGCGCAGAGCCAGGAT
GGCTTTCATACCTATCATGTGATGAATCCGTATGATGATGGCATTGGCCTG
GATGAATTTGTGGATTGGCTGAACGAAAGCGGCTGCCCGATTCAGCGTAT
TGCGGATTATGGCGATTGGCTGCAACGTTTTGAAACCGCGCTGCGCGCTC
TGCCGGATCGTCAGCGTCATAGCAGCCTGCTGCCGCTGCTGCATAACTAT
CGTCAGCCGGAACGTCCGGTGCGTGGTAGCATTGCGCCGACCGATCGCTT
TCGTGCGGCCGTGCAGGAAGCGAAAATTGGCCCGGATAAAGATATTCCG
CATGTGGGTGCGCCGATTATTGTGAAATATGTGAGCGATCTGCGCCTGCT
GGGCCTGCTGTAA

6.3.1.4.2 NCAR

NCAR operon sequence:

GAATTCGAGCAGTGTCTCTAGAAGAGACGTACGAGCTGTTGACAATTAAT
CATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACA
GGAAACAGAATTGTAGAGGAGATATACATATGCACCATCATCATCATCA
TTCTTCTGGTGCGGTGGATAGCCCGGATGAACGTCTGCAACGTCGTATTG
CGCAGCTGTTTGCGGAAGATGAACAGGTGAAAGCAGCACGCCCGCTGGA
AGCGGTTAGCGCAGCGGTGAGCGCACCGGGTATGCGTCTGGCCCAGATT
GCGGCGACCGTGATGGCGGGCTATGCGGATCGTCCGGCAGCGGGTCAGC
GTGCGTTTGAACTGAACACCGATGATGCGACCGGCCGTACCAGCCTGCGT
CTGCTGCCGCGTTTTGAAACCATTACCTATCGTGAACTGTGGCAGCGTGTG
GGTGAAGTTGCGGCAGCGTGGCATCACGATCCGGAAAATCCGCTGCGTG
CGGGCGATTTTGTGGCGCTGCTGGGCTTTACCAGCATTGATTATGCGACCC
TGGATCTGGCCGATATTCATCTGGGCGCGGTGACCGTTCCGCTGCAAGCG

AGCGCAGCAGTCAGCCAACTGATTGCGATTCTGACCGAAACGAGTCCGC

GCCTGCTGGCATCTACCCCGGAACATCTGGATGCGGCGGTGGAATGTCTG

CTGGCAGGTACGACGCCGGAACGCCTGGTGGTGTTTGATTATCATCCGGA

AGATGATGATCAGCGTGCGGCGTTTGAAAGCGCGCGTCGTCGTCTGGCCG

ATGCGGGCAGCCTGGTGATTGTGGAAACCCTGGATGCGGTGCGTGCGCGT

GGTCGTGATCTGCCGGCTGCTCCGCTGTTTGTGCCGGATACCGATGATGAT

CCGCTGGCCCTGCTGATTTATACCTCTGGTAGCACGGGTACGCCGAAAGG

CGCCATGTATACCAACCGCCTGGCAGCAACGATGTGGCAAGGTAACAGC

ATGCTGCAAGGCAATAGCCAGCGTGTGGGCATTAACCTGAACTATATGCC

GATGAGCCATATTGCGGGCCGTATTAGCCTGTTTGGCGTGCTGGCCCGTG

GTGGCACCGCGTATTTTGCGGCGAAAAGCGATATGAGCACCCTGTTTGAA

GATATTGGCCTGGTGCGTCCGACCGAAATTTTTTTTGTGCCGCGTGTGTGC

GATATGGTGTTTCAGCGTTATCAGAGCGAACTGGATCGTCGTAGCGTGGC

GGGTGCGGATCTGGATACCCTGGATCGTGAAGTGAAAGCGGATCTGCGTC

AGAACTATCTGGGCGGTCGTTTTCTGGTGGCGGTGGTGGGTAGCGCACCG

CTGGCCGCGGAAATGAAAACCTTTATGGAAAGCGTGCTGGATCTGCCGCT

GCATGATGGCTATGGCAGCACCGAAGCGGGTGCGAGCGTGCTGCTGGAT

AACCAGATTCAGCGTCCGCCGGTGCTGGATTATAAACTGGTGGACGTCCC

GGAACTGGGCTATTTTCGTACCGATCGTCCGCATCCGCGTGGCGAACTGC

TGCTGAAAGCGGAAACCACCATTCCGGGCTATTATAAACGTCCGGAAGT

GACCGCGGAAATTTTTGATGAAGATGGCTTCTATAAAACCGGCGATATTG

TGGCGGAACTGGAACATGATCGTCTGGTGTATGTGGATCGTCGCAACAAC

GTGCTGAAACTGAGCCAGGGCGAATTTGTGACCGTGGCGCATCTGGAAG

CGGTGTTTGCGAGCAGCCCGCTGATTCGTCAGATTTTTATCTACGGCTCTA

GTGAACGCTCTTATCTGCTGGCAGTGATTGTGCCGACCGATGATGCCCTG

CGTGGCCGTGATACCGCGACCCTGAAAAGCGCGCTGGCCGAAAGCATTC

AGCGTATTGCGAAAGATGCGAACCTGCAACCGTATGAAATTCCGCGTGAT

TTTCTGATTGAAACCGAACCGTTCACCATTGCGAACGGCCTGCTGTCTGG

CATTGCGAAACTGCTGCGTCCGAACCTGAAAGAACGTTATGGCGCGCAG

CTGGAACAAATGTATACCGATCTGGCCACCGGCCAGGCGGATGAACTGC

TGGCCCTGCGTCGTGAAGCGGCGGATCTGCCGGTTCTGGAAACCGTTAGC

CGTGCGGCGAAAGCCATGCTGGGTGTGGCGAGCGCGGATATGCGTCCGG

ATGCGCATTTTACCGATCTGGGCGGCGATAGCCTGAGCGCCCTGAGCTTT

AGCAACCTGCTGCATGAAATTTTTGGCGTGGAAGTGCCGGTGGGTGTGGT

TGTGAGCCCGGCAAACGAACTGCGTGACCTGGCCAACTATATTGAAGCG

GAACGTAACAGCGGCGCGAAACGTCCGACCTTTACCAGCGTGCATGGCG

GCGGTAGCGAAATTCGTGCGGCCGATCTGACCCTGGATAAATTTATTGAT

GCGCGTACCCTGGCCGCAGCGGATAGCATTCCGCATGCACCGGTTCCGGC

ACAGACCGTCCTGCTGACGGGCGCAAATGGCTATCTGGGCCGTTTTCTGT

GCCTGGAATGGCTGGAACGTCTGGATAAAACCGGTGGCACCCTGATTTGC

GTGGTGCGTGGCAGCGATGCGGCGGCAGCCCGTAAACGCCTGGATAGCG

CGTTTGATAGCGGCGATCCGGGCCTGCTGGAACATTATCAGCAGCTGGCC

GCACGCACCCTGGAAGTTCTGGCCGGTGATATTGGCGATCCGAACCTGGG

CCTGGATGATGCCACCTGGCAGCGTCTGGCCGAAACCGTGGATCTGATTG

TGCACCCGGCTGCTCTGGTGAATCATGTGCTGCCGTATACCCAGCTGTTTG

GCCCGAACGTTGTGGGCACCGCGGAAATCGTTCGTCTGGCTATTACCGCG

CGTCGTAAACCGGTGACCTATCTGAGCACCGTGGGCGTGGCGGATCAGGT

TGATCCGGCGGAATATCAGGAAGATAGCGACGTCCGCGAAATGAGCGCA

GTCCGCGTCGTTCGCGAAAGTTATGCAAACGGTTATGGTAACAGCAAATG

GGCGGGTGAAGTGCTGCTGCGTGAAGCGCATGATCTGTGCGGTCTGCCGG

TGGCGGTGTTTCGTAGCGATATGATTCTGGCCCATAGCCGTTATGCGGGCC

AGCTGAACGTGCAGGATGTGTTTACCCGTCTGATTCTGAGCCTGGTGGCG

ACCGGCATTGCGCCGTATAGCTTTTATCGCACCGATGCGGATGGCAACCG

TCAGCGTGCGCATTATGATGGCCTGCCGGCGGATTTTACCGCAGCAGCTA

TCACCGCACTGGGCATTCAGGCGACCGAAGGCTTTCGTACCTATGATGTG

CTGAATCCGTATGATGATGGCATTAGCCTGGATGAATTTGTCGATTGGCTG

GTCGAATCTGGTCACCCGATTCAGCGCATTACCGATTATAGCGATTGGTTT

CACCGCTTTGAAACCGCGATTCGTGCGCTGCCGGAAAAACAGCGTCAGG

CGAGCGTTCTGCCGCTGCTGGATGCGTATCGTAATCCGTGTCCGGCGGTC
CGTGGTGCAATTCTGCCGGCGAAAGAATTTCAGGCGGCGGTGCAGACCG
CGAAAATTGGCCCGGAACAGGATATTCCGCATCTGAGCGCACCGCTGATT
GATAAATATGTGAGCGATCTGGAACTGCTGCAACTGCTGTAACCGGCTTA
TCGGTCAGTTTCACCTGATTTACGTAAAAACCCGCTTCGGCGGGTTTTTGC
TTTTGGAGGGGCAGAAAGATGAATGACTGTCCACGACGCTATACCCAAA
AGAAAGCGGCTTATCGGTCAGTTTCACCTGGTTTACGTAAAAACCCGCTT
CGGCGGGTTTTTGCTTTTGGAGGGGCAGAAAGATGAATGACTGTCCACGA
CACTATACCCAAAAGAAAGCGGCTTATCGGTCAGTTTCACCTGTTTTACG
TAAAAACCCGCTTCGGCGGGTTTTTACTTTTGGAGGGGCAGAAAGATGAA
TGACTGTCCACGACACTATACCCAAAAGAAAACTGGGCCTTTCGTTTTAT
CTGACTAGTTAGCATCGTTCACTGCAG

NCAR open reading frame (ORF) sequence [gene with N-terminal His_6 tag]:

ATGCACCATCATCATCATCATTCTTCTGGTGCGGTGGATAGCCCGGATGA
ACGTCTGCAACGTCGTATTGCGCAGCTGTTTGCGGAAGATGAACAGGTGA
AAGCAGCACGCCCGCTGGAAGCGGTTAGCGCAGCGGTGAGCGCACCGGG
TATGCGTCTGGCCCAGATTGCGGCGACCGTGATGGCGGGCTATGCGGATC
GTCCGGCAGCGGGTCAGCGTGCGTTTGAACTGAACACCGATGATGCGAC
CGGCCGTACCAGCCTGCGTCTGCTGCCGCGTTTTGAAACCATTACCTATCG
TGAACTGTGGCAGCGTGTGGGTGAAGTTGCGGCAGCGTGGCATCACGATC
CGGAAAATCCGCTGCGTGCGGGCGATTTTGTGGCGCTGCTGGGCTTTACC
AGCATTGATTATGCGACCCTGGATCTGGCCGATATTCATCTGGGCGCGGT
GACCGTTCCGCTGCAAGCGAGCGCAGCAGTCAGCCAACTGATTGCGATTC
TGACCGAAACGAGTCCGCGCCTGCTGGCATCTACCCCGGAACATCTGGAT
GCGGCGGTGGAATGTCTGCTGGCAGGTACGACGCCGGAACGCCTGGTGG
TGTTTGATTATCATCCGGAAGATGATGATCAGCGTGCGGCGTTTGAAAGC

GCGCGTCGTCGTCTGGCCGATGCGGGCAGCCTGGTGATTGTGGAAACCCT

GGATGCGGTGCGTGCGCGTGGTCGTGATCTGCCGGCTGCTCCGCTGTTTGT

GCCGGATACCGATGATGATCCGCTGGCCCTGCTGATTTATACCTCTGGTA

GCACGGGTACGCCGAAAGGCGCCATGTATACCAACCGCCTGGCAGCAAC

GATGTGGCAAGGTAACAGCATGCTGCAAGGCAATAGCCAGCGTGTGGGC

ATTAACCTGAACTATATGCCGATGAGCCATATTGCGGGCCGTATTAGCCT

GTTTGGCGTGCTGGCCCGTGGTGGCACCGCGTATTTTGCGGCGAAAAGCG

ATATGAGCACCCTGTTTGAAGATATTGGCCTGGTGCGTCCGACCGAAATT

TTTTTTGTGCCGCGTGTGTGCGATATGGTGTTTCAGCGTTATCAGAGCGAA

CTGGATCGTCGTAGCGTGGCGGGTGCGGATCTGGATACCCTGGATCGTGA

AGTGAAAGCGGATCTGCGTCAGAACTATCTGGGCGGTCGTTTTCTGGTGG

CGGTGGTGGGTAGCGCACCGCTGGCCGCGGAAATGAAAACCTTTATGGA

AAGCGTGCTGGATCTGCCGCTGCATGATGGCTATGGCAGCACCGAAGCG

GGTGCGAGCGTGCTGCTGGATAACCAGATTCAGCGTCCGCCGGTGCTGGA

TTATAAACTGGTGGACGTCCCGGAACTGGGCTATTTTCGTACCGATCGTCC

GCATCCGCGTGGCGAACTGCTGCTGAAAGCGGAAACCACCATTCCGGGC

TATTATAAACGTCCGGAAGTGACCGCGGAAATTTTTGATGAAGATGGCTT

CTATAAAACCGGCGATATTGTGGCGGAACTGGAACATGATCGTCTGGTGT

ATGTGGATCGTCGCAACAACGTGCTGAAACTGAGCCAGGGCGAATTTGT

GACCGTGGCGCATCTGGAAGCGGTGTTTGCGAGCAGCCCGCTGATTCGTC

AGATTTTTATCTACGGCTCTAGTGAACGCTCTTATCTGCTGGCAGTGATTG

TGCCGACCGATGATGCCCTGCGTGGCCGTGATACCGCGACCCTGAAAAG

CGCGCTGGCCGAAAGCATTCAGCGTATTGCGAAAGATGCGAACCTGCAA

CCGTATGAAATTCCGCGTGATTTTCTGATTGAAACCGAACCGTTCACCATT

GCGAACGGCCTGCTGTCTGGCATTGCGAAACTGCTGCGTCCGAACCTGAA

AGAACGTTATGGCGCGCAGCTGGAACAAATGTATACCGATCTGGCCACC

GGCCAGGCGGATGAACTGCTGGCCCTGCGTCGTGAAGCGGCGGATCTGC

CGGTTCTGGAAACCGTTAGCCGTGCGGCGAAAGCCATGCTGGGTGTGGCG

AGCGCGGATATGCGTCCGGATGCGCATTTTACCGATCTGGGCGGCGATAG

CCTGAGCGCCCTGAGCTTTAGCAACCTGCTGCATGAAATTTTTGGCGTGG

AAGTGCCGGTGGGTGTGGTTGTGAGCCCGGCAAACGAACTGCGTGACCT

GGCCAACTATATTGAAGCGGAACGTAACAGCGGCGCGAAACGTCCGACC

TTTACCAGCGTGCATGGCGGCGGTAGCGAAATTCGTGCGGCCGATCTGAC

CCTGGATAAATTTATTGATGCGCGTACCCTGGCCGCAGCGGATAGCATTC

CGCATGCACCGGTTCCGGCACAGACCGTCCTGCTGACGGGCGCAAATGG

CTATCTGGGCCGTTTTCTGTGCCTGGAATGGCTGGAACGTCTGGATAAAA

CCGGTGGCACCCTGATTTGCGTGGTGCGTGGCAGCGATGCGGCGGCAGCC

CGTAAACGCCTGGATAGCGCGTTTGATAGCGGCGATCCGGGCCTGCTGGA

ACATTATCAGCAGCTGGCCGCACGCACCCTGGAAGTTCTGGCCGGTGATA

TTGGCGATCCGAACCTGGGCCTGGATGATGCCACCTGGCAGCGTCTGGCC

GAAACCGTGGATCTGATTGTGCACCCGGCTGCTCTGGTGAATCATGTGCT

GCCGTATACCCAGCTGTTTGGCCCGAACGTTGTGGGCACCGCGGAAATCG

TTCGTCTGGCTATTACCGCGCGTCGTAAACCGGTGACCTATCTGAGCACC

GTGGGCGTGGCGGATCAGGTTGATCCGGCGGAATATCAGGAAGATAGCG

ACGTCCGCGAAATGAGCGCAGTCCGCGTCGTTCGCGAAAGTTATGCAAA

CGGTTATGGTAACAGCAAATGGGCGGGTGAAGTGCTGCTGCGTGAAGCG

CATGATCTGTGCGGTCTGCCGGTGGCGGTGTTTCGTAGCGATATGATTCTG

GCCCATAGCCGTTATGCGGGCCAGCTGAACGTGCAGGATGTGTTTACCCG

TCTGATTCTGAGCCTGGTGGCGACCGGCATTGCGCCGTATAGCTTTTATCG

CACCGATGCGGATGGCAACCGTCAGCGTGCGCATTATGATGGCCTGCCGG

CGGATTTTACCGCAGCAGCTATCACCGCACTGGGCATTCAGGCGACCGAA

GGCTTTCGTACCTATGATGTGCTGAATCCGTATGATGATGGCATTAGCCTG

GATGAATTTGTCGATTGGCTGGTCGAATCTGGTCACCCGATTCAGCGCATT

ACCGATTATAGCGATTGGTTTCACCGCTTTGAAACCGCGATTCGTGCGCTG

CCGGAAAAACAGCGTCAGGCGAGCGTTCTGCCGCTGCTGGATGCGTATC

GTAATCCGTGTCCGGCGGTCCGTGGTGCAATTCTGCCGGCGAAAGAATTT

CAGGCGGCGGTGCAGACCGCGAAAATTGGCCCGGAACAGGATATTCCGC

ATCTGAGCGCACCGCTGATTGATAAATATGTGAGCGATCTGGAACTGCTG
CAACTGCTGTAA

6.3.1.4.3 *Bs*Sfp

***Bs*Sfp operon sequence:**

GAATTCGAGCAGTGTCTCTAGAAGAGACGTACTGTTGACAATTAATCATC
CGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGA
AACAGCGCCGCTGAGAAAAAGCGAAGCGGCACTGCTCTTTAACAATTTA
TCAGACAATCTGTGTGGGCACTCGACCGGAATTATCGATTAACTTTATTAT
TAAAAATTAAAGAGGTATATATTAATGTATCGATTAAATAAGGAGGAAT
AAACCATGGGTAAAATCTACGGGATCTATATGGATCGCCCCCTCAGCCAG
GAAGAGAATGAACGCTTTATGTCATTTATTAGTCCCGAGAAACGTGAAAA
GTGCCGCCGCTTTTACCATAAAGAAGATGCACATCGTACCCTGTTAGGTG
ATGTGTTAGTCCGGTCGGTGATTAGTCGCCAGTACCAATTGGATAAAAGC
GATATCCGTTTTTCCACCCAGGAATATGGCAAACCGTGCATTCCGGATCT
CCCTGACGCTCATTTTAACATTTCTCATTCAGGTCGCTGGGTCATCTGCGC
ATTTGACTCTCAGCCAATTGGCATTGATATTGAAAAAACAAAACCCATTT
CCCTGGAGATTGCAAAACGTTTCTTTGCTAAAACGGAGTATAGCGATCTG
CTGGCGAAAGACAAAGACGAACAGACGGATTACTTCTATCATTTGTGGA
GCATGAAGGAAAGTTTTATCAAGCAAGAGGGAAAGGGCCTTTCTCTTCCT
CTGGATTCTTTCAGCGTCCGTCTTCACCAGGACGGCCAAGTTAGTATCGA
ACTCCCAGACTCACACAGTCCGTGTTATATCAAGACGTATGAGGTCGATC
CGGGTTATAAAATGGCGGTATGTGCCGCACATCCAGATTTTCCGGAGGAT
ATTACAATGGTGTCGTATGAGGAGTTGCTGCGCGGTAGCGGCGATTATAA
GGACGATGATGATAAATAATACAGATTAAATCAGAACGCAGAAGCGGTC
TGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGAC
CCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGG

GGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAA
AGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGACTAGTTAGCATCG
TTCACTGCAG

***Bs*Sfp open reading frame (ORF) sequence [gene with C-terminal FLAG tag]:**

ATGGGTAAAATCTACGGGATCTATATGGATCGCCCCCTCAGCCAGGAAG
AGAATGAACGCTTTATGTCATTTATTAGTCCCGAGAAACGTGAAAAGTGC
CGCCGCTTTTACCATAAAGAAGATGCACATCGTACCCTGTTAGGTGATGT
GTTAGTCCGGTCGGTGATTAGTCGCCAGTACCAATTGGATAAAAGCGATA
TCCGTTTTTCCACCCAGGAATATGGCAAACCGTGCATTCCGGATCTCCCTG
ACGCTCATTTTAACATTTCTCATTCAGGTCGCTGGGTCATCTGCGCATTTG
ACTCTCAGCCAATTGGCATTGATATTGAAAAAACAAAACCCATTTCCCTG
GAGATTGCAAAACGTTTCTTTGCTAAAACGGAGTATAGCGATCTGCTGGC
GAAAGACAAAGACGAACAGACGGATTACTTCTATCATTTGTGGAGCATG
AAGGAAAGTTTTATCAAGCAAGAGGGAAAGGGCCTTTCTCTTCCTCTGGA
TTCTTTCAGCGTCCGTCTTCACCAGGACGGCCAAGTTAGTATCGAACTCCC
AGACTCACACAGTCCGTGTTATATCAAGACGTATGAGGTCGATCCGGGTT
ATAAAATGGCGGTATGTGCCGCACATCCAGATTTTCCGGAGGATATTACA
ATGGTGTCGTATGAGGAGTTGCTGCGCGGTAGCGGCGATTATAAGGACG
ATGATGATAAATAA

6.3.1.4.4 ATA-117

ATA-117 operon sequence:

GAATTCGAGCAGTGTCTCTAGAAGAGACGTACTGTTGACAATTAATCATC
CGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGA
AACAGCGCCGCTGAGAAAAAGCGAAGCGGCACTGCTCTTTAACAATTTA

TCAGACAATCTGTGTGGGCACTCGACCGGAATTATCGATTAACTTTATTAT

TAAAAATTAAAGAGGTATATATTAATGTATCGATTAAATAAGGAGGAAT

AAACCATGGAACAAAAACTTATTTCTGAAGAAGATCTGACCAGCGAAAT

TGTTTATACCCATGATACCGGTCTGGATTATATCACCTATAGCGATTATGA

ACTGGATCCGGCAAATCCGCTGGCAGGCGGTGCAGCATGGATTGAAGGT

GCATTTGTTCCGCCTAGCGAAGCACGTATTAGCATTTTTGATCAGGGCTAT

CTGCATAGTGATGTTACCTATACCGTGTTTCATGTGTGGAATGGTAATGCA

TTTCGCCTGGATGATCATATTGAACGTCTGTTTAGCAATGCAGAAAGCAT

GCGTATTATTCCGCCTCTGACCCAGGATGAAGTTAAAGAAATTGCACTGG

AACTGGTTGCAAAAACCGAACTGCGTGAAGCATTTGTTAGCGTTAGCATT

ACCCGTGGTTATAGCAGCACACCGGGTGAACGTGATATTACCAAACATC

GTCCGCAGGTTTATATGTATGCAGTTCCGTATCAGTGGATTGTTCCGTTTG

ATCGTATTCGTGATGGTGTTCATGCAATGGTTGCACAGAGCGTTCGTCGTA

CACCGCGTAGCAGCATTGATCCGCAGGTTAAAAATTTTCAGTGGGGTGAT

CTGATTCGTGCAGTTCAAGAAACCCATGATCGTGGTTTTGAAGCACCGCT

GCTGCTGGATGGTGATGGTCTGCTGGCCGAAGGTAGCGGTTTTAATGTTGT

TGTGATTAAAGATGGTGTGGTTCGTAGTCCGGGTCGTGCAGCACTGCCTG

GTATTACCCGTAAAACCGTTCTGGAAATTGCAGAAAGCCTGGGTCATGAA

GCAATTCTGGCAGATATTACCCTGGCAGAACTGCTGGATGCAGATGAAGT

TCTGGGTTGTACCACCGCAGGCGGTGTTTGGCCGTTTGTTAGCGTGGATGG

TAATCCGATTTCAGATGGTGTTCCGGGTCCGATTACCCAGAGCATTATTCG

TCGTTATTGGGAACTGAATGTTGAAAGCAGCAGCCTGCTGACACCGGTTC

AGTATAACTGGTCTCACCCGCAGTTCGAAAAATAATACAGATTAAATCAG

AACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGC

GGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGC

GCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGG

CATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTAT

CTGACTAGTTAGCATCGTTCACTGCAG

ATA-117 open reading frame (ORF) sequence [gene with c-Myc epitope and C-terminal Strep II tag]:

ATGGAACAAAAACTTATTTCTGAAGAAGATCTGACCAGCGAAATTGTTTA
TACCCATGATACCGGTCTGGATTATATCACCTATAGCGATTATGAACTGG
ATCCGGCAAATCCGCTGGCAGGCGGTGCAGCATGGATTGAAGGTGCATTT
GTTCCGCCTAGCGAAGCACGTATTAGCATTTTTGATCAGGGCTATCTGCAT
AGTGATGTTACCTATACCGTGTTTCATGTGTGGAATGGTAATGCATTTCGC
CTGGATGATCATATTGAACGTCTGTTTAGCAATGCAGAAAGCATGCGTAT
TATTCCGCCTCTGACCCAGGATGAAGTTAAAGAAATTGCACTGGAACTGG
TTGCAAAAACCGAACTGCGTGAAGCATTTGTTAGCGTTAGCATTACCCGT
GGTTATAGCAGCACACCGGGTGAACGTGATATTACCAAACATCGTCCGC
AGGTTTATATGTATGCAGTTCCGTATCAGTGGATTGTTCCGTTTGATCGTA
TTCGTGATGGTGTTCATGCAATGGTTGCACAGAGCGTTCGTCGTACACCG
CGTAGCAGCATTGATCCGCAGGTTAAAAATTTTCAGTGGGGTGATCTGAT
TCGTGCAGTTCAAGAAACCCATGATCGTGGTTTTGAAGCACCGCTGCTGC
TGGATGGTGATGGTCTGCTGGCCGAAGGTAGCGGTTTTAATGTTGTTGTGA
TTAAAGATGGTGTGGTTCGTAGTCCGGGTCGTGCAGCACTGCCTGGTATT
ACCCGTAAAACCGTTCTGGAAATTGCAGAAAGCCTGGGTCATGAAGCAA
TTCTGGCAGATATTACCCTGGCAGAACTGCTGGATGCAGATGAAGTTCTG
GGTTGTACCACCGCAGGCGGTGTTTGGCCGTTTGTTAGCGTGGATGGTAA
TCCGATTTCAGATGGTGTTCCGGGTCCGATTACCCAGAGCATTATTCGTCG
TTATTGGGAACTGAATGTTGAAAGCAGCAGCCTGCTGACACCGGTTCAGT
ATAACTGGTCTCACCCGCAGTTCGAAAAATAA

6.3.1.4.5 SitATA

SitATA operon sequence:

GAATTCGAGCAGTGTCTCTAGAAGAGACGTACCCGCGAAATTAATACGA

CTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTTGAAATAAT

TTTGTTTAACTTTAAGAAGGAGATATACCATGGGCCATCATCATCATCATC

ATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTATGGCATTTAGC

GCAGATACACCGGAAATTGTTTATACCCATGATACCGGTCTGGATTATAT

CACCTATAGCGATTATGAACTGGACCCTGCAAATCCGCTGGCAGGCGGTG

CAGCATGGATTGAAGGTGCATTTGTTCCGCCTAGCGAAGCACGTATTAGC

ATTTTTGATCAGGGCTTTTATACCAGTGATGCAACCTATACCACCTTTCAT

GTGTGGAATGGTAATGCATTTCGTCTGGGTGATCATATTGAACGTCTGTTT

AGCAATGCCGAAAGCATTCGTCTGATTCCGCCTCTGACCCAGGATGAAGT

TAAAGAAATTGCACTGGAACTGGTTGCAAAAACCGAACTGCGTGAAGCA

ATGGTTACCGTTACCATTACCCGTGGTTATAGCAGCACCCCGTTTGAACGT

GATATTACCAAACATCGTCCGCAGGTTTATATGAGCGCATGTCCGTATCA

GTGGATTGTTCCGTTTGATCGTATTCGTGATGGTGTTCATCTGATGGTTGCA

CAGAGCGTTCGTCGTACACCGCGTAGCAGCATTGATCCGCAGGTTAAAA

ATTTTCAGTGGGGTGATCTGATTCGTGCAATTCAGGAAACCCACGATCGC

GGTTTTGAACTGCCGCTGCTGCTGGATTGTGATAATCTGCTGGCCGAAGGT

CCGGGTTTTAATGTTGTTGTTATTAAAGATGGCGTGGTTCGTAGTCCGGGT

CGTGCAGCACTGCCTGGTATTACCCGTAAAACCGTTCTGGAAATTGCAGA

AAGCCTGGGTCATGAAGCAATTCTGGCAGATATTACACCGGCAGAACTGT

ATGATGCAGATGAAGTTCTGGGTTGTAGCACCGGTGGTGGTGTTTGGCCG

TTTGTTAGCGTTGATGGTAATAGCATTAGTGATGGCGTTCCGGGTCCGGTT

ACCCAGAGCATTATTCGTCGTTATTGGGAACTGAATGTTGAACCGAGCAG

CCTGCTGACACCGGTTCAGTATTAACAAAGCCCGAAAGGAAGCTGAGTT

GGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTA

AACGGGTCTTGAGGGGTTTTTTGACTGGGCCTTTCGTTTTATCTGACTAGTT
AGCATCGTTCACTGCAG

SitATA open reading frame (ORF) sequence [gene with N-terminal His_6 tag]:

ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATAT
CGAAGGTCGTATGGCATTTAGCGCAGATACACCGGAAATTGTTTATACCC
ATGATACCGGTCTGGATTATATCACCTATAGCGATTATGAACTGGACCCT
GCAAATCCGCTGGCAGGCGGTGCAGCATGGATTGAAGGTGCATTTGTTCC
GCCTAGCGAAGCACGTATTAGCATTTTTGATCAGGGCTTTTATACCAGTG
ATGCAACCTATACCACCTTTCATGTGTGGAATGGTAATGCATTTCGTCTGG
GTGATCATATTGAACGTCTGTTTAGCAATGCCGAAAGCATTCGTCTGATTC
CGCCTCTGACCCAGGATGAAGTTAAAGAAATTGCACTGGAACTGGTTGCA
AAAACCGAACTGCGTGAAGCAATGGTTACCGTTACCATTACCCGTGGTTA
TAGCAGCACCCCGTTTGAACGTGATATTACCAAACATCGTCCGCAGGTTT
ATATGAGCGCATGTCCGTATCAGTGGATTGTTCCGTTTGATCGTATTCGTG
ATGGTGTTCATCTGATGGTTGCACAGAGCGTTCGTCGTACACCGCGTAGC
AGCATTGATCCGCAGGTTAAAAATTTTCAGTGGGGTGATCTGATTCGTGC
AATTCAGGAAACCCACGATCGCGGTTTTGAACTGCCGCTGCTGCTGGATT
GTGATAATCTGCTGGCCGAAGGTCCGGGTTTTAATGTTGTTGTTATTAAAG
ATGGCGTGGTTCGTAGTCCGGGTCGTGCAGCACTGCCTGGTATTACCCGT
AAAACCGTTCTGGAAATTGCAGAAAGCCTGGGTCATGAAGCAATTCTGG
CAGATATTACACCGGCAGAACTGTATGATGCAGATGAAGTTCTGGGTTGT
AGCACCGGTGGTGGTGTTTGGCCGTTTGTTAGCGTTGATGGTAATAGCATT
AGTGATGGCGTTCCGGGTCCGGTTACCCAGAGCATTATTCGTCGTTATTGG
GAACTGAATGTTGAACCGAGCAGCCTGCTGACACCGGTTCAGTATTAA

6.3.1.4.6 (*R*)-IRED

(*R*)-IRED operon sequence:

GAATTCGAGCAGTGTCTCTAGAAGAGACGTACGAGCTGTTGACAATTAAT

CATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACA

GGAAACAGAATTGTAGAGGAGATATACATATGGGCAGCAGCCATCATCA

TCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGGCGAC

AATCGTACGCCGGTCACGGTTATCGGTCTGGGTCTGATGGGTCAAGCACT

GGCAGCAGCATTTCTGGAAGCAGGTCACACCACGACCGTGTGGAACCGT

AGCGCGGGTAAAGCCGAACAGCTGGTTTCTCAGGGTGCGGTTCAGGCCG

CAACCCCGGCAGATGCTGTTGCAGCTTCAGAACTGGTGGTTGTCTGCCTG

TCGACCTATGATAACATGCATGACGTCATTGGTAGTCTGGGCGAATCCCT

GCGTGGTAAAGTCATCGTGAATCTGACGAGCGGTAGCTCTGATCAGGGTC

GTGAAACCGCCGCATGGGCAGAAAAACAGGGTGTTGAATACCTGGACGG

CGCAATTATGATCACGCCGCCGGGTATTGGCACGGAAACCGCAGTCCTGT

TTTATGCTGGTACCCAGTCTGTGTTCGAAAAATACGAACCGGCTCTGAAA

CTGCTGGGCGGTGGCACGACCTATCTGGGTACCGATCATGGCATGCCGGC

CCTGTACGACGTGTCACTGCTGGGTCTGATGTGGGGCACGCTGAACTCGT

TTCTGCATGGCGTGGCAGTGGTTGAAACCGCGGGTGTTGGCGCCCAGCAA

TTTCTGCCGTGGGCACACATGTGGCTGGAAGCTATTAAAATGTTCACCGC

GGATTATGCAGCTCAAATCGATGCGGGTGACGGCAAATTCCCGGCAAAT

GACGCTACGCTGGAAACCCACCTGGCGGCCCTGAAACATCTGGTTCACG

AATCAGAAGCGCTGGGCATTGATGCCGAACTGCCGAAATACAGTGAAGC

GCTGATGGAACGCGTGATCTCCCAGGGTCACGCTAAAAACAGCTATGCG

GCAGTCCTGAAAGCCTTCCGTAAACCGTCCGAATAACCGGCTTATCGGTC

AGTTTCACCTGATTTACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGG

AGGGGCAGAAAGATGAATGACTGTCCACGACGCTATACCCAAAAGAAA

GCGGCTTATCGGTCAGTTTCACCTGGTTTACGTAAAAACCCGCTTCGGCG

GGTTTTTGCTTTTGGAGGGGCAGAAAGATGAATGACTGTCCACGACACTA
TACCCAAAAGAAAGCGGCTTATCGGTCAGTTTCACCTGTTTTACGTAAAA
ACCCGCTTCGGCGGGTTTTTACTTTTGGAGGGGCAGAAAGATGAATGACT
GTCCACGACACTATACCCAAAAGAAAACTGGGCCTTTCGTTTTATCTGAC
TAGTTAGCATCGTTCACTGCAG

_(R)_-IRED open reading frame (ORF) sequence [gene with N-terminal His_6 tag]:

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGC
GCGGCAGCCATATGGGCGACAATCGTACGCCGGTCACGGTTATCGGTCTG
GGTCTGATGGGTCAAGCACTGGCAGCAGCATTTCTGGAAGCAGGTCACA
CCACGACCGTGTGGAACCGTAGCGCGGGTAAAGCCGAACAGCTGGTTTC
TCAGGGTGCGGTTCAGGCCGCAACCCCGGCAGATGCTGTTGCAGCTTCAG
AACTGGTGGTTGTCTGCCTGTCGACCTATGATAACATGCATGACGTCATTG
GTAGTCTGGGCGAATCCCTGCGTGGTAAAGTCATCGTGAATCTGACGAGC
GGTAGCTCTGATCAGGGTCGTGAAACCGCCGCATGGGCAGAAAAACAGG
GTGTTGAATACCTGGACGGCGCAATTATGATCACGCCGCCGGGTATTGGC
ACGGAAACCGCAGTCCTGTTTTATGCTGGTACCCAGTCTGTGTTCGAAAA
ATACGAACCGGCTCTGAAACTGCTGGGCGGTGGCACGACCTATCTGGGTA
CCGATCATGGCATGCCGGCCCTGTACGACGTGTCACTGCTGGGTCTGATG
TGGGGCACGCTGAACTCGTTTCTGCATGGCGTGGCAGTGGTTGAAACCGC
GGGTGTTGGCGCCCAGCAATTTCTGCCGTGGGCACACATGTGGCTGGAAG
CTATTAAAATGTTCACCGCGGATTATGCAGCTCAAATCGATGCGGGTGAC
GGCAAATTCCCGGCAAATGACGCTACGCTGGAAACCCACCTGGCGGCCC
TGAAACATCTGGTTCACGAATCAGAAGCGCTGGGCATTGATGCCGAACTG
CCGAAATACAGTGAAGCGCTGATGGAACGCGTGATCTCCCAGGGTCACG
CTAAAAACAGCTATGCGGCAGTCCTGAAAGCCTTCCGTAAACCGTCCGA
ATAA

6.3.1.4.7 (*S*)-IRED

(*S*)-IRED operon sequence:

GAATTCGAGCAGTGTCTCTAGAAGAGACGTACGAGCTGTTGACAATTAAT

CATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACA

GGAAACAGAATTGTAGAGGAGATATACATATGGGCAGCAGCCATCATCA

TCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGAGCAAA

CAGTCAGTTACGGTGATTGGTCTGGGTCCGATGGGTCAAGCGATGGTCAA

TACCTTTCTGGATAATGGTCACGAAGTGACCGTGTGGAACCGTACGGCGT

CAAAAGCAGAAGCTCTGGTGGCGCGCGGCGCAGTTCTGGCACCGACCGT

CGAAGATGCTCTGAGCGCGAATGAACTGATTGTTCTGTCTCTGACCGATT

ATGACGCCGTGTACGCAATCCTGGAACCGGTTACGGGCTCACTGTCGGGT

AAAGTGATTGCAAACCTGAGCTCTGATACCCCGGACAAAGCGCGTGAAG

CGGCCAAATGGGCAGCTAAACATGGTGCGAAACATCTGACCGGCGGTGT

GCAGGTTCCGCCGCCGCTGATCGGCAAACCGGAAAGTTCCACCTATTACT

CCGGTCCGAAAGATGTTTTTGACGCCCATGAAGATACCCTGAAAGTCCTG

ACGAACGCCGATTATCGTGGTGAAGATGCAGGTCTGGCCGCAATGTATTA

CCAGGCGCAAATGACCATTTTCTGGACCACGATGCTGAGCTATTACCAGA

CGCTGGCTCTGGGCCAAGCGAATGGTGTTAGTGCTAAAGAACTGCTGCCG

TATGCCACCATGATGACGTCCATGATGCCGCATTTTCTGGAACTGTATGCT

CAGCACGTCGATTCTGCGGACTATCCGGGTGATGTGGACCGTCTGGCGAT

GGGCGCAGCTTCAGTCGATCACGTGCTGCATACCCACCAAGATGCGGGT

GTTAGCACCGTCCTGCCGGCCGCAGTGGCCGAAATCTTCAAAGCCGGTAT

GGAAAAAGGCTTTGCTGAAAATTCGTTCTCCTCTCTGATTGAAGTCCTGA

AAAAACCGGCAGTGTAACCGGCTTATCGGTCAGTTTCACCTGATTTACGT

AAAAACCCGCTTCGGCGGGTTTTTGCTTTTGGAGGGGCAGAAAGATGAAT

GACTGTCCACGACGCTATACCCAAAAGAAAGCGGCTTATCGGTCAGTTTC

ACCTGGTTTACGTAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGGAGGGGC

AGAAAGATGAATGACTGTCCACGACACTATACCCAAAAGAAAGCGGCTT
ATCGGTCAGTTTCACCTGTTTTACGTAAAAACCCGCTTCGGCGGGTTTTTA
CTTTTGGAGGGGCAGAAAGATGAATGACTGTCCACGACACTATACCCAA
AAGAAAACTGGGCCTTTCGTTTTATCTGACTAGTTAGCATCGTTCACTGCA
G

(*S*)-IRED open reading frame (ORF) sequence [gene with N-terminal His_6 tag]:

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGC
GCGGCAGCCATATGAGCAAACAGTCAGTTACGGTGATTGGTCTGGGTCCG
ATGGGTCAAGCGATGGTCAATACCTTTCTGGATAATGGTCACGAAGTGAC
CGTGTGGAACCGTACGGCGTCAAAAGCAGAAGCTCTGGTGGCGCGCGGC
GCAGTTCTGGCACCGACCGTCGAAGATGCTCTGAGCGCGAATGAACTGAT
TGTTCTGTCTCTGACCGATTATGACGCCGTGTACGCAATCCTGGAACCGGT
TACGGGCTCACTGTCGGGTAAAGTGATTGCAAACCTGAGCTCTGATACCC
CGGACAAAGCGCGTGAAGCGGCCAAATGGGCAGCTAAACATGGTGCGA
AACATCTGACCGGCGGTGTGCAGGTTCCGCCGCCGCTGATCGGCAAACC
GGAAAGTTCCACCTATTACTCCGGTCCGAAAGATGTTTTTGACGCCCATG
AAGATACCCTGAAAGTCCTGACGAACGCCGATTATCGTGGTGAAGATGC
AGGTCTGGCCGCAATGTATTACCAGGCGCAAATGACCATTTTCTGGACCA
CGATGCTGAGCTATTACCAGACGCTGGCTCTGGGCCAAGCGAATGGTGTT
AGTGCTAAAGAACTGCTGCCGTATGCCACCATGATGACGTCCATGATGCC
GCATTTTCTGGAACTGTATGCTCAGCACGTCGATTCTGCGGACTATCCGGG
TGATGTGGACCGTCTGGCGATGGGCGCAGCTTCAGTCGATCACGTGCTGC
ATACCCACCAAGATGCGGGTGTTAGCACCGTCCTGCCGGCCGCAGTGGCC
GAAATCTTCAAAGCCGGTATGGAAAAAGGCTTTGCTGAAAATTCGTTCTC
CTCTCTGATTGAAGTCCTGAAAAAACCGGCAGTGTAA

6.3.1.4.8 ***Ad*RedAm open reading frame (ORF) sequence**

ATGGCAAATAGTCCGGTTAGCGTTTTTGGTCTGGGTGCAATGGGCACCGC
ACTGGCAACCCAGTTTCTGCGTAAAGGTCATAAAACCACCGTTTGGAATC
GTACACCGGCAAAAGCACAGCCGCTGATTGCAATTGGTGCAAGCCATGC
ACCGACCATTGATAGCGCAGCAGCAGCAAGCAGCCTGCTGATTATTTGTC
AGCTGGATAAAGCAAGCGTTATGCAGACCCTGCAGCAGGCACCGACCGC
ATGGGCAGCAAAAACCATTGTTGATCTGACCAATGGCACACCGGCACAT
GCACGTGAAACCGCAGATTGGGCACTGGCACATGGTGCACGTTATATTCA
TGGTGGTATTATGGCAGTGCCGTTTATGATTGGTCAGCCGGATGCAATGA
TTCTGTATAGCGGTCCTGCCGAAGTTTTTGAAGGTGTTAAAGATACCCTGA
GCGTTCTGGGCACCAATACCTATGTTGGTGAAGATGTTGGTCTGGCAAGC
CTGCATGATCTGGCACTGCTGAGCGGTATGTATGGTCTGTTTAGCGGTTTT
ACCCATGCAGTTGCACTGGTTCAGAGCGCAAATATTCCGGCAGCAGGTTT
TGTTGCAACCCAGCTGATTCCGTGGCTGACCGCAATGACCCAGCATCTGA
ATCTGCTGGCCACCCAGGTTGATGAAAAAGATTATGGTGATGGTGGTAGC
AGCCTGGATATGCAGGCAAAAGCCGCACCGAATATTCTGGAAGCAAGCC
AGGCACAGGGTGTTAGCGTTGAACTGATTCAGCCGATCTTTAAACTGATT
GAACGTCGTGTTGAAGAAGGTAAAGGTAGCGAAGGTCTGGCAGCACTGG
TTGGTATGATTATGAAAGGTGGCACCAAAGATAGCGTGTAA

6.3.1.4.9 **CapW open reading frame (ORF) sequence [gene with N-terminal His_6 tag]**

ATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGT
GGACAGCAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCCGA
TGACCCGTAGCGGTTTTAGCGCAAAACGTCTGATTCGTGTTCGTGATGTTC

TGGAACGTCGTGTTGATGCAGGTCATAGTCCGGGTGCAGTTGCAGTTGTT
GCACGTCATGGTGAAGTTCATATTGAAGCAGTTGGTAATCTGGCATTTGA
AGGTGCAGGTAGCACCACCCCGATGGCAGCAGATACCATTTGTCGTATTG
CAAGCATGACCAAAAGCATTGTTGCAGCATGTGCAATGACCCTGGTTGA
AGATTGTACCGTTCGTCTGAATGATCCGGTTGATGAACTGCTGCCGGAAC
TGGCAAATATGACCGTTCTGGCAGATCCGAATGGTCCGCTGGATGATACC
GTTCCGATGAAACGTCCGATTACCCTGCGTGATCTGCTGACCTGTCGTCTG
GGCACCGGCACCATTCTGGCAGAACCGGGTACAGTTCCGATTTGTGATGC
ACTGAATGCAATTGAACGTGGTGATGGTCCTGGTGCACCGGAACCGAGTC
CGGATGAATGGCTGCGTCGTCTGGGTACACTGCCGCTGGTTCATCAGCCT
GGTGAACGTTGGATGTATAACATTGGTCCGACCGCAACCGGTCTGCTGAT
TGCACGTGCCACCGGTATGCCGCTGGAAGATGCAATGCGTGAACGTATTT
GTGAACCGCTGGGTATGAAAGATACCACCTTTGGTCTGGGTGATGCAAGC
GTTAGCCGTGTTGCAGCAGCATATGTTCATGATGATGCAACCGGTGAACT
GGTTGAAGAAGGTAGCGACGGTTTTGGTAAACGTCCGCCTACCTTTAAAA
GCGCAGCAAGCGGTCTGGTTAGCACCGCAGAAGATTGTCTGGCATTTGCA
AGCGCACTGCTGGCAAATGGCACCCATCGTGGTGAACGTGTTCTGAGCCG
TCCGAGCGTTAGCCTGATGACCAGCGATCATCTGCGTCCGACCCAGAAAG
CAGCAAGCCGTTGGATTTGGCCTCTGGAACATTTTGCAGATTTTGGTTGGG
GTTTTGGTATGGAAGTTCGTACCCGTCGTACCAATCTGGGTCCGAGCGTTG
GTAGCTATGGTTGGTATGGTAAATATGGCACCGCATGGTCAAATGATCCG
GCAGAAGGTATGACCACCATTTTTATGATTCAGCGTGCAGAAGGTCTGGA
ACTGCCGACCCTGGTGGATTTTTGGACCGCAACCTATCAGGCAATTGATG
ATTAA

6.3.2 Biocatalyst Preparation

6.3.2.1 Preparation of Whole Cell Biocatalyst (CAR-TA-IRED)

Chemically competent *E. coli* BL21 (DE3) expression cells were transformed with plasmid pLH02, pLH09 or pLH10 following the manufacturer's protocol (NEB). Pre-cultures were inoculated using either a single colony picked from a transformation plate or from a glycerol stock and grown in LB media (supplemented with 100 μg mL^{-1} ampicillin) for 16 h at 37 °C. 400 mL LB media (supplemented with 100 μg mL^{-1} ampicillin) was inoculated with pre-culture (1:100 dilution) and grown at 37 °C until an optical density (OD_{600}) of 0.6 was reached. Flasks were then cooled for 20 min at 20 °C before induction of protein expression using isopropyl-β-D-1-thiogalactopyranoside (IPTG, 0.8 mM). Protein expression at 20 °C and 250 rpm was stopped after 16 h by centrifugation (3200 rcf, 4 °C, 20 min). Wet cell mass was recorded and resuspended in 500 mM sodium phosphate buffer (pH 7.0) to a final concentration of 40 mg mL^{-1} cells, ready to be used in biotransformation reactions.

6.3.2.2 Preparation of GOase (F2 and M3-5 Variants)

GOase F_2 and GOase $M_{3\text{-}5}$ biocatalysts were prepared and purified as described previously by the Turner group.[119,120]

6.3.2.3 Preparation of CAR

Chemically competent *E. coli* BL21 (DE3) expression cells were co-transformed with a plasmid containing one of five CAR genes and Sfp from *Bacillus subtilis* (*Bs*Sfp) following the manufacturer's protocol (NEB). Pre-cultures were inoculated using either a single colony picked from a transformation plate or from a glycerol stock and grown in LB media (supplemented with 50 μg mL^{-1} kanamycin and 50 μg mL^{-1} streptomycin) for 16 h at 37 °C. 700 mL of autoinduction medium LB broth base including trace elements (Formedium) (supplemented with 50 μg mL^{-1} kanamycin and 50 μg mL^{-1} streptomycin) was inoculated with pre-culture (1:100 dilution) and grown at 30 °C for 48 h. Protein expression was stopped after 16 h via centrifugation (3200 rcf, 4 °C, 20 min). Cells were resuspended in 50 mM Tris.HCl pH 7.5, 10 mM imidazole containing 2 mg/mL lysozyme (200 mg pellet per mL buffer) and incubated at 30 °C for 30 min. Cells were then lysed by ultrasonication (Section 6.3.3.2) followed by centrifugation (39 000*g*, 30 mins, 4 °C) to remove insoluble debris. CAR enzymes were purified using HisTrap™ Fast Flow Crude 5 mL columns (GE Healthcare) and eluted over an imidazole gradient (100 mM Tris.HCl pH 7.5, 25 mM imidazole and 100 mM Tris.HCl pH 7.5, 500 mM imidazole). Fractions were analysed using SDS-PAGE followed by western blot analysis, and fractions containing desired purified protein were exchanged into 50 mM Tris.HCl pH 7.5 and concentration was carried out using Vivaspin 20 centrifugal concentrators with a MWCO of 30 000 Da.

6.3.2.4 Preparation of CapW

6.3.2.4.1 General Preparation of CapW

pRSET/CapW with N-terminal polyHistidine-tag (His-tag) for purification was transformed into *E. coli* BL21(DE3) pLysS, OverExpress C41 (DE3), OverExpress C43 (DE3) and ArcticExpress chemically competent cells.

For growth, a single colony of cells containing pRSET/CapW was picked from overnight agar plates and used to inoculate LB media (10 mL - 50 mL) containing ampicillin (100 μg mL^{-1}) followed by overnight growth at 37 °C. Overnight culture (1 mL - 10 mL) was used to inoculate a larger culture (100 mL - 1000 mL) of LB or LB supplemented with sorbitol (1 M) and betaine (2.5 mM) (LB B/S) containing ampicillin (100 μg mL^{-1}), which was grown at 37 °C until optical density (OD_{600}) 0.6 was reached. Protein expression was induced using IPTG (1 mM) and stopped after 4 h, when the wet cell mass was recorded and resuspended in 500 mM sodium phosphate buffer (pH 7.0) to a final concentration of 40 mg mL^{-1} cells, ready to be used in biotransformation reactions. Alternatively, the cells were lysed using ultrasonication or cell disruption (Sections 6.3.3.2 and 6.3.3.3) and used in biotransformations as cell-free lysate. For purification of CapW, a nickel purification column was equilibrated with potassium phosphate buffer (50 mM, pH 8.0, NaCl (150 mM)), and the cell lysate supernatant was loaded onto the column. The column was washed with buffer (40 mL) and the flow-through kept. Column was eluted with imidazole of varying concentrations (15 mM, 50 mM, 150 mM, 250 mM) and NaCl (500 mM), with each fraction being collected and saved. Fractions were analysed using SDS-PAGE followed by western blot analysis, and fractions containing desired purified protein were dialysed overnight to remove imidazole.

6.3.2.4.2 Expression Trials for Production of CapW

E. coli (pRSET/CapW) cultures in LB or LB B/S of OD_{600} 0.6 were equilibrated to a range of temperatures (20 °C, 23 °C, 25 °C, 30 °C, 37 °C) and induced with isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.4 mM) to start protein expression. An aliquot of culture was also taken before addition of IPTG to assess the degree of expression prior to induction. Expression was allowed to proceed for various lengths of time (1 h, 2 h, 4 h, o/n) before cells were harvested via centrifugation and lysed using bacterial protein extraction reagent (B-PER, Thermo Scientific). Lysed samples were analysed for soluble and insoluble protein expression using SDS-PAGE followed by western blot analysis.

6.3.3 Cell Lysis

6.3.3.1 Chemical Lysis

Cells were harvested via centrifugation following expression, with the cell pellet resuspended in B-PER solution (4 mL g^{-1}) containing DNaseI (2 μl mL^{-1}), lysozyme (2 μl mL^{-1}) and protease inhibitor. The suspension was incubated at room temperature for 15 m followed by centrifugation (2000 rcf, 4 °C, 10 m) to remove cell debris. Supernatant was centrifuged again at high speed (20000 rcf, 4 °C, 20 m) to separate soluble protein from insoluble protein.

6.3.3.2 Ultrasonication Lysis

Cells were harvested via centrifugation following expression, with the cell pellet resuspended in reaction buffer. The suspension was incubated at room temperature for 30 m, followed by ultrasonication on ice (20 s on, 20 s off, 20 cycles for *Ad*RedAm and CAR, 30 s on, 45 s off, 15 cycles for CapW). The resulting solution was centrifuged (2000 rcf, 4 °C, 10 m) to remove cell debris. Supernatant was centrifuged again at high speed (20000 rcf, 4 °C, 20 m) to separate soluble protein from insoluble protein.

6.3.3.3 Liquid Homogenisation Lysis

Cells were harvested via centrifugation following expression, with the cell pellet resuspended in potassium phosphate buffer (50 mM, pH 8.0) containing NaCl (150 mM), glycerol (10% v/v) and protease inhibitor. The suspension (> 10 mL) was incubated at room temperature for 15 m, followed by lysis of the cells using the Cell Disruptor (25 kpsi, Constant Cell Disruption Systems). The resulting solution was centrifuged (2000 rcf, 4 °C, 10 m) to remove cell debris. Supernatant was centrifuged again at high speed (20000 rcf, 4 °C, 20 m) to separate soluble protein from insoluble protein.

6.3.4 Protein Analysis

Soluble and insoluble fractions of protein were boiled in Laemmli buffer (1:2 dilution) for 10 m prior to sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis. Completed gels were stained with Instant

Blue to visualise or used for western blot analysis (15 V for 50 m, monoclonal anti-polyHistidine antibody produced in mouse (1:3000 dilution), monoclonal anti-c-Myc antibody produced in mouse (1:4000 dilution) or anti-FLAG antibody produced in rabbit (1:4000 dilution), SIGMAFAST BCIP/NBT substrate tablets).

6.4 Analytical Methods and Sample Preparation

6.4.1 Analytical Methods

6.4.1.1 Gas Chromatography Methods for Conversion and ee and de Determination

Chiral GC equipped with a 25 m CP-Chirasil-DEX CB column with 0.25 mm inner diameter and 0.25 μm film thickness (Agilent, Santa Clara, CA, USA) was used for the detection of all imine **4** intermediates and piperidine **5** products, and for the determination of **5b, 5h** and **5i** production using a calibration curve (Sections 6.4.1.1.1, 6.4.1.1.2 and 6.4.1.1.3). Determination of conversion and ee for **5b**: injector temperature 200°C, detector temperature 250 °C. Method: 90 °C - 200 °C, 4 °C min^{-1}, hold at 200 °C for 5 min. Determination of ee for **5e**-**f**: injector temperature 200°C, detector temperature 250 °C. Method: 50 °C - 200 °C, 5 °C min^{-1}, hold at 200 °C for 2 min. Chiral GC equipped with a 30 m HP-1MS column with 0.32 mm inner diameter and 0.25 μm film thickness (Agilent, Santa Clara, CA, USA) was used for the detection of product **11d**: injector temperature 200°C, detector temperature 250 °C. Method: 120 °C - 250 °C, 5 °C min^{-1}, hold at 250 °C for 5 min.

GCMS equipped with a HP-1 MS column (Agilent, 30.0 m x 0.32 mm x 0.25 μm) was used for the detection of all keto alcohol and imine intermediates, and piperidine products (method: 50 °C – 175 °C at a rate of 5 °C min^{-1}, followed by 175 °C – 250 °C at a rate of 10 °C min^{-1}, inlet temperature 270 °C).

Percentage conversion of keto acid to imine, keto alcohol and amine was determined based on integration of these corresponding peaks (after comparison with authentic commercial or chemically obtained standards), except for amine **5b**, **5h**, **5i** and **11d,** where amine production was calculated by integration of its peak with decane as an external standard (Sections 6.4.1.1.1, 6.4.1.1.2, 6.4.1.1.3 and 6.4.1.1.4). Retention times for the imine **4**, keto alcohol **6** and amine **5** products were reported previously.[40]

6.4.1.1.1 Calibration Curve for the Determination of Conversion to Product Using Compound 5b

Reactions of compound **1b** were extracted in the presence of external standard decane (1 mg mL^{-1}) to enable the calculation of conversion to product **5b** by use of a calibration curve of decane standard vs. **5b** authentic standard.

The calibration curve was constructed using 1 mg mL^{-1} decane against varying concentrations of **5b** authentic standard in triplicate.

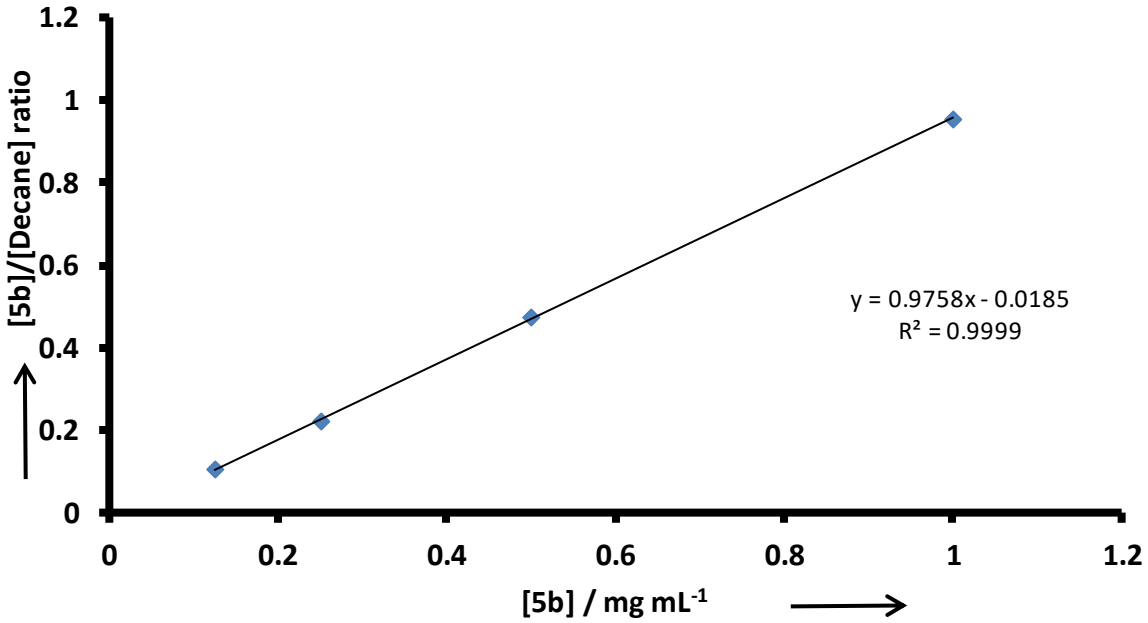


Figure 86 Calibration curve used to determine the conversion to product amine **4a** for the reactions of substrate **1b.**

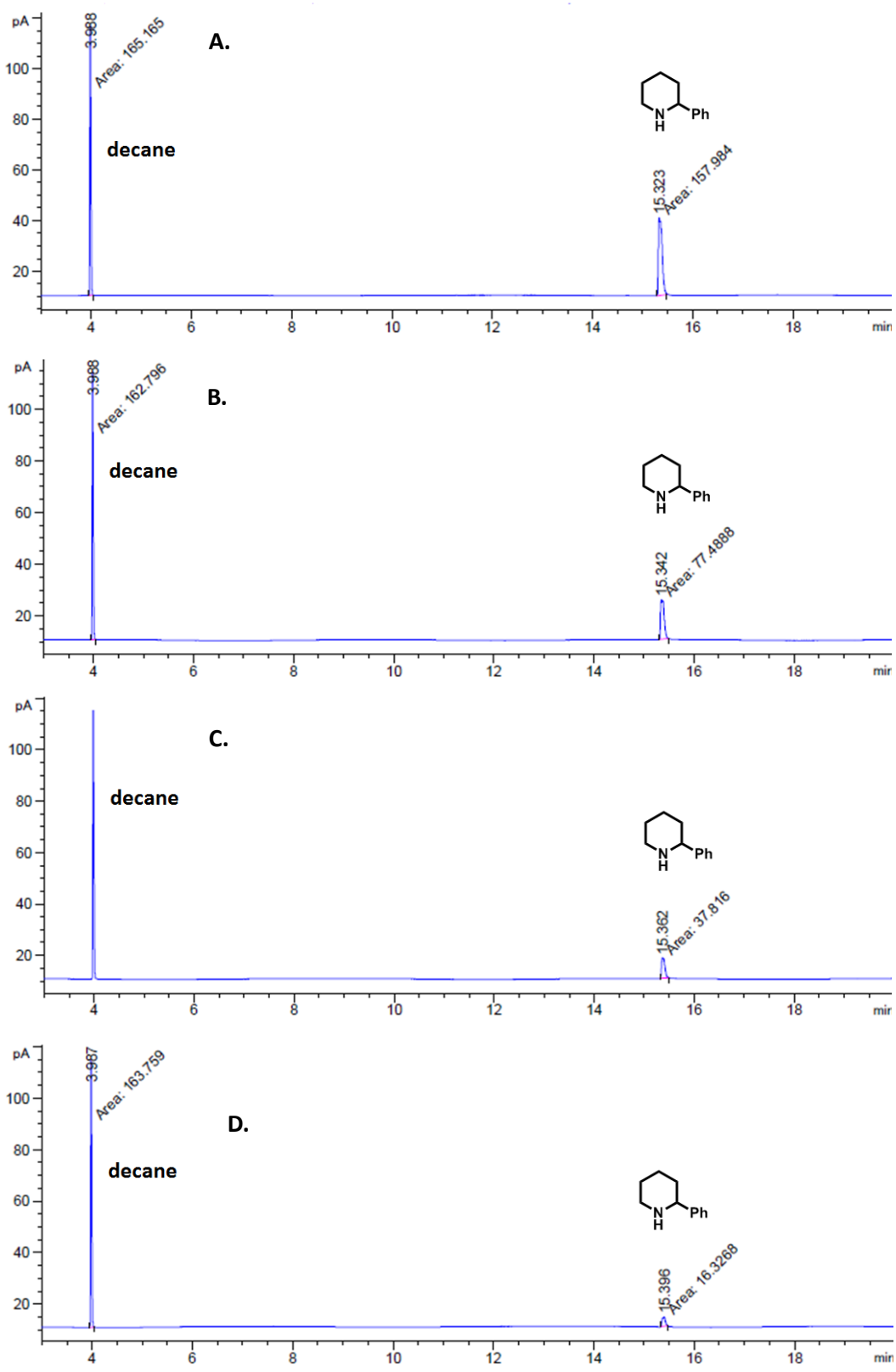


Figure 87 GC chromatograms used for the construction of the calibration curve used to determine the conversion to product amine **5b** for reactions using substrate **1b**. A) 1 mg mL^{-1} **5b**. B) 0.5 mg mL^{-1} **5b**. C) 0.25 mg mL^{-1} **5b**. D) 0.125 mg mL^{-1} **5b**. (CP-Chirasil-DEX CB (Agilent, 25.0 m x 0.25 mm x 0.25 μm), injector temperature 200°C, detector temperature 250 °C . Method: 90 °C - 200 °C, 4 °C min^{-1}, hold at 200 °C for 5 min; carrier gas helium.).

6.4.1.1.2 Calibration Curve for the Determination of Conversion to Product Using Compound **5h**

Reactions of compound **1h** were extracted in the presence of external standard decane (1 mg mL^{-1}) to enable the calculation of conversion to product **5h** by use of a calibration curve of decane standard vs. **5h** authentic standard.

The calibration curve was constructed using 1 mg mL^{-1} decane against varying concentrations of **5h** authentic standard in triplicate.

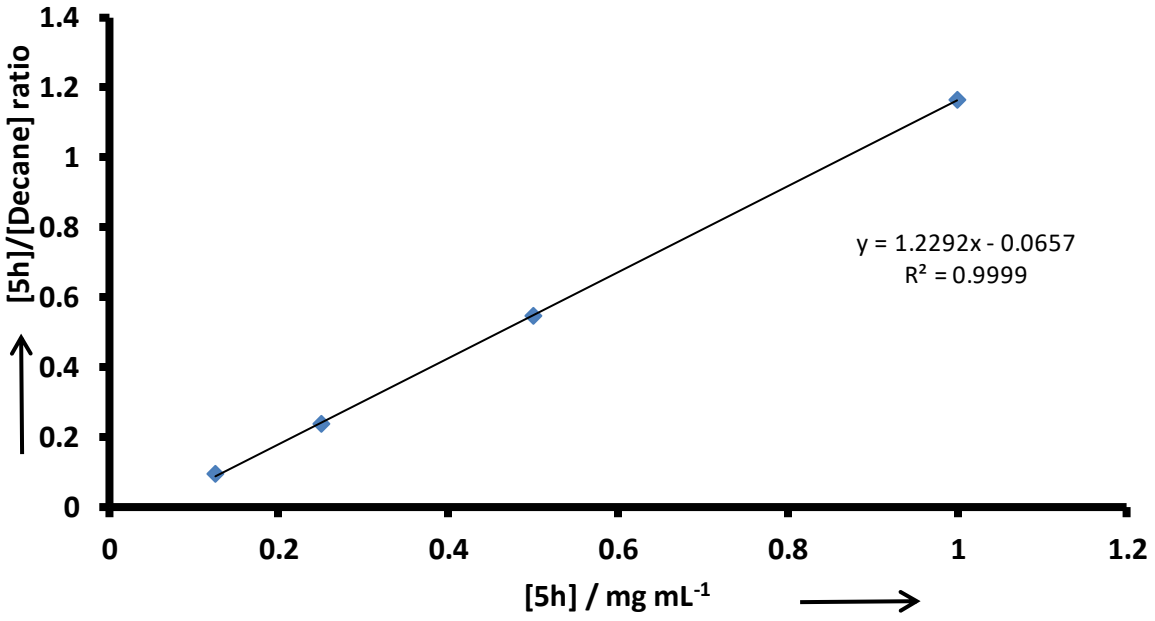


Figure 88 Calibration curve used to determine the conversion to product amine **5h** for the reactions of substrate **1h**.

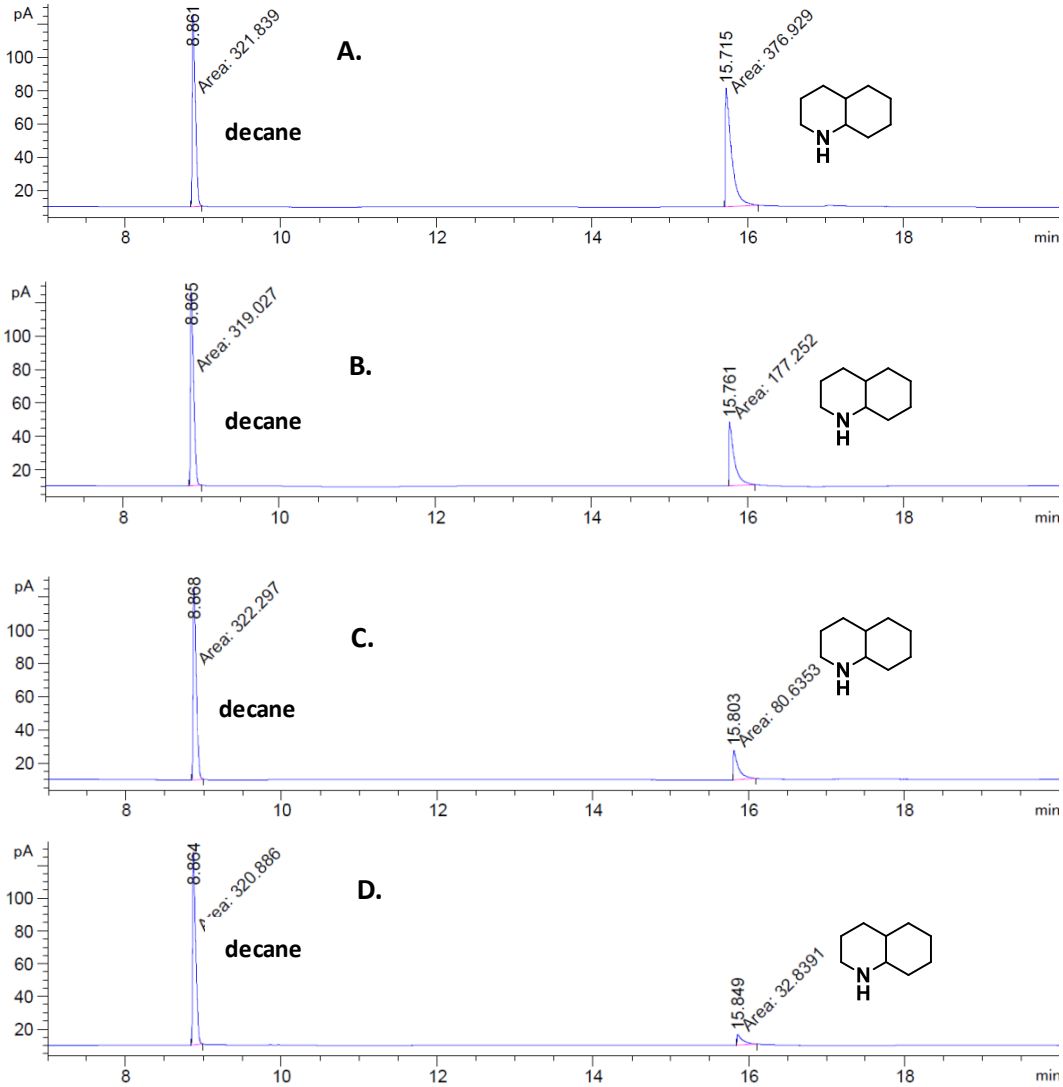


Figure 89 GC chromatograms used for the construction of the calibration curve used to determine the conversion to product amine **5h** for reactions using substrate **1g**. A) 1 mg mL^{-1} **5h**. B) 0.5 mg mL^{-1} **5h**. C) 0.25 mg mL^{-1} **5h**. D) 0.125 mg mL^{-1} **5h**. (CP-Chirasil-DEX CB (Agilent, 25.0 m x 0.25 mm x 0.25 μm), injector temperature 200°C, detector temperature 250 °C . Method: 90 °C - 200 °C, 4 °C min^{-1}, hold at 200 °C for 5 min; carrier gas helium.).

6.4.1.1.3 Calibration Curve for the Determination of Conversion to Product Using Compound **5i**

Reactions of compound **1i** were extracted in the presence of external standard decane (1 mg mL^{-1}) to enable the calculation of conversion to product **5i** by use of a calibration curve of decane standard vs. **5i** authentic standard.

The calibration curve was constructed using 1 mg mL^{-1} decane against varying concentrations of **5i** authentic standard in triplicate.

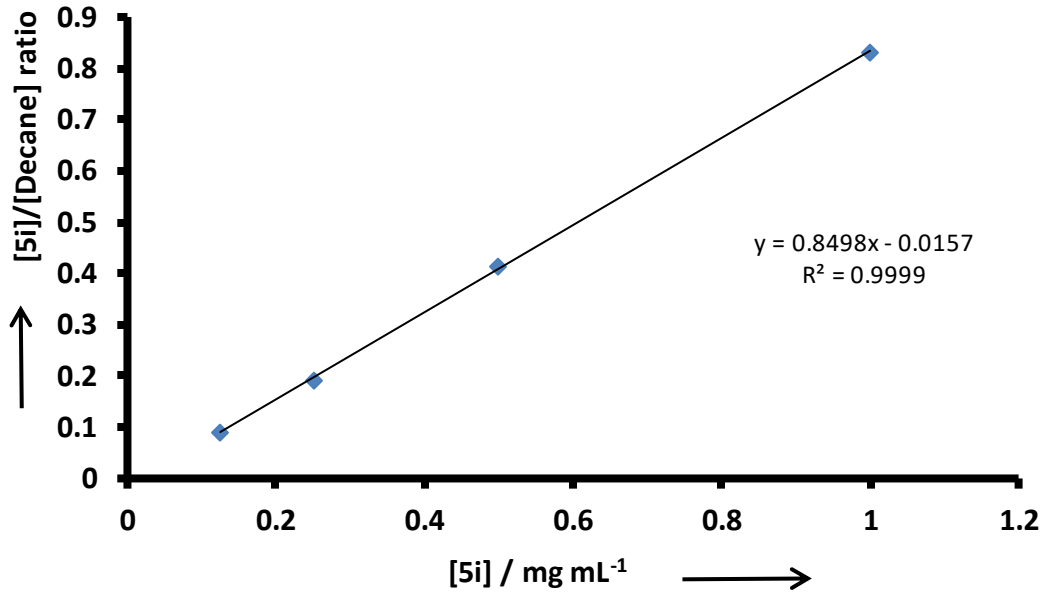


Figure 90 Calibration curve used to determine the conversion to product amine **5i** for the reactions of substrate **1i**.

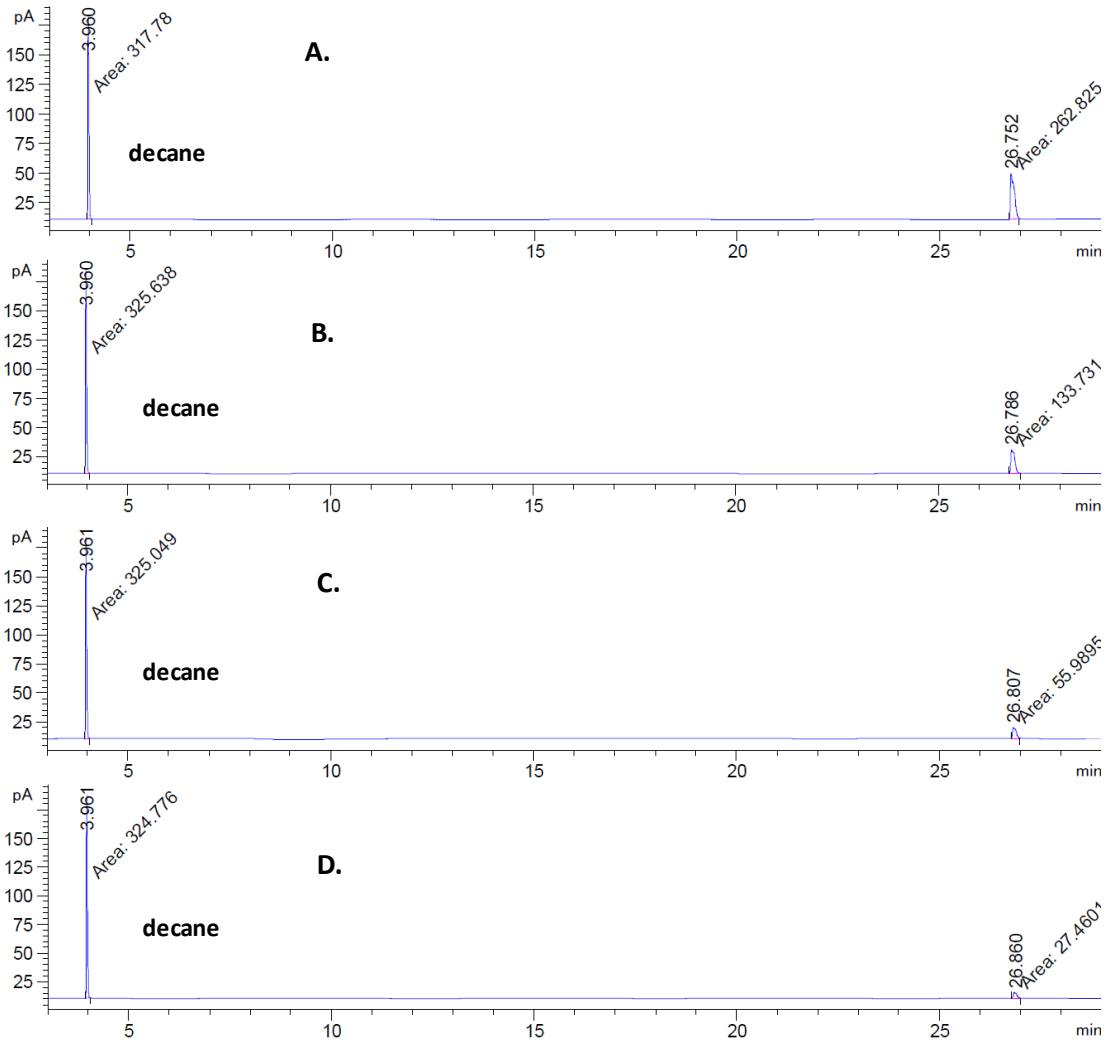


Figure 91 GC chromatograms used for the construction of the calibration curve used to determine the conversion to product amine **5i** for reactions using substrate **1i**. A) 1 mg mL^{-1} **5i**. B) 0.5 mg mL^{-1} **5i**. C) 0.25 mg mL^{-1} **5i**. D) 0.125 mg mL^{-1} **5i**. (CP-Chirasil-DEX CB (Agilent, 25.0 m x 0.25 mm x 0.25 μm), injector temperature 200°C, detector temperature 250 °C . Method: 90 °C - 300 °C, 3 °C min^{-1}, hold at 300 °C for 5 min; carrier gas helium.).

6.4.1.1.4 Calibration Curve for the Determination of Conversion to Product Using Compound **11d**

Reactions of compound **12d** were extracted in the presence of external standard 1-phenylethanol (1 mg mL^{-1}) to enable the calculation of conversion to product **11d** by use of a calibration curve of decane standard vs. **11d** authentic standard.

The calibration curve was constructed using 1 mg mL^{-1} 1-phenylethanol against varying concentrations of **11d** authentic standard in triplicate.

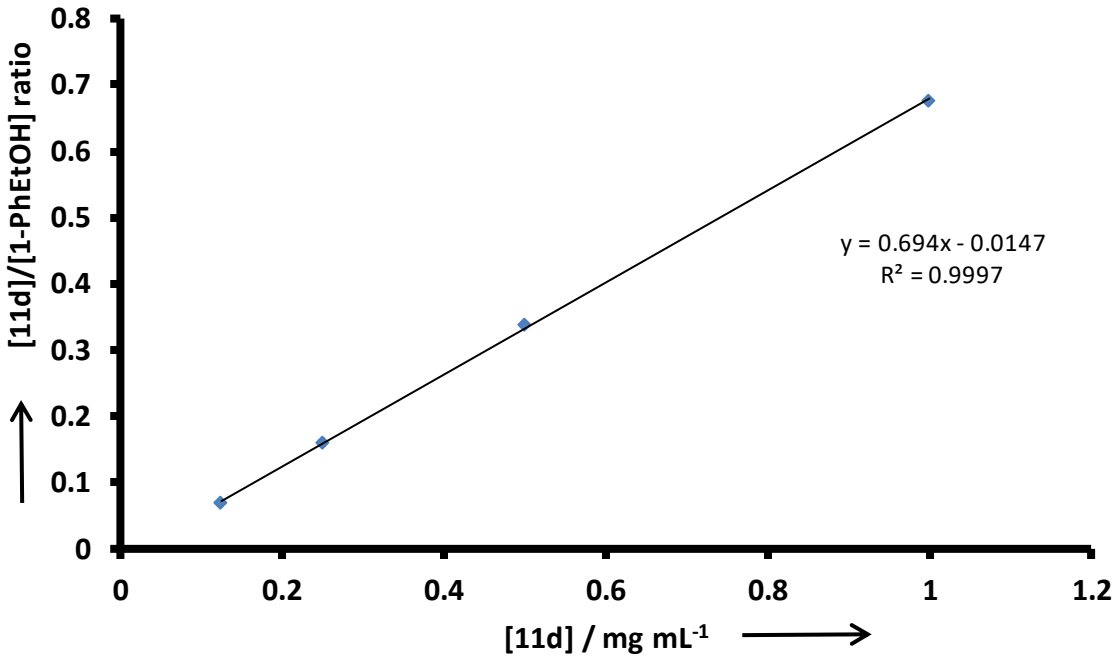


Figure 92 Calibration curve used to determine the conversion to product amine **11d** for the reactions of substrate **12d**.

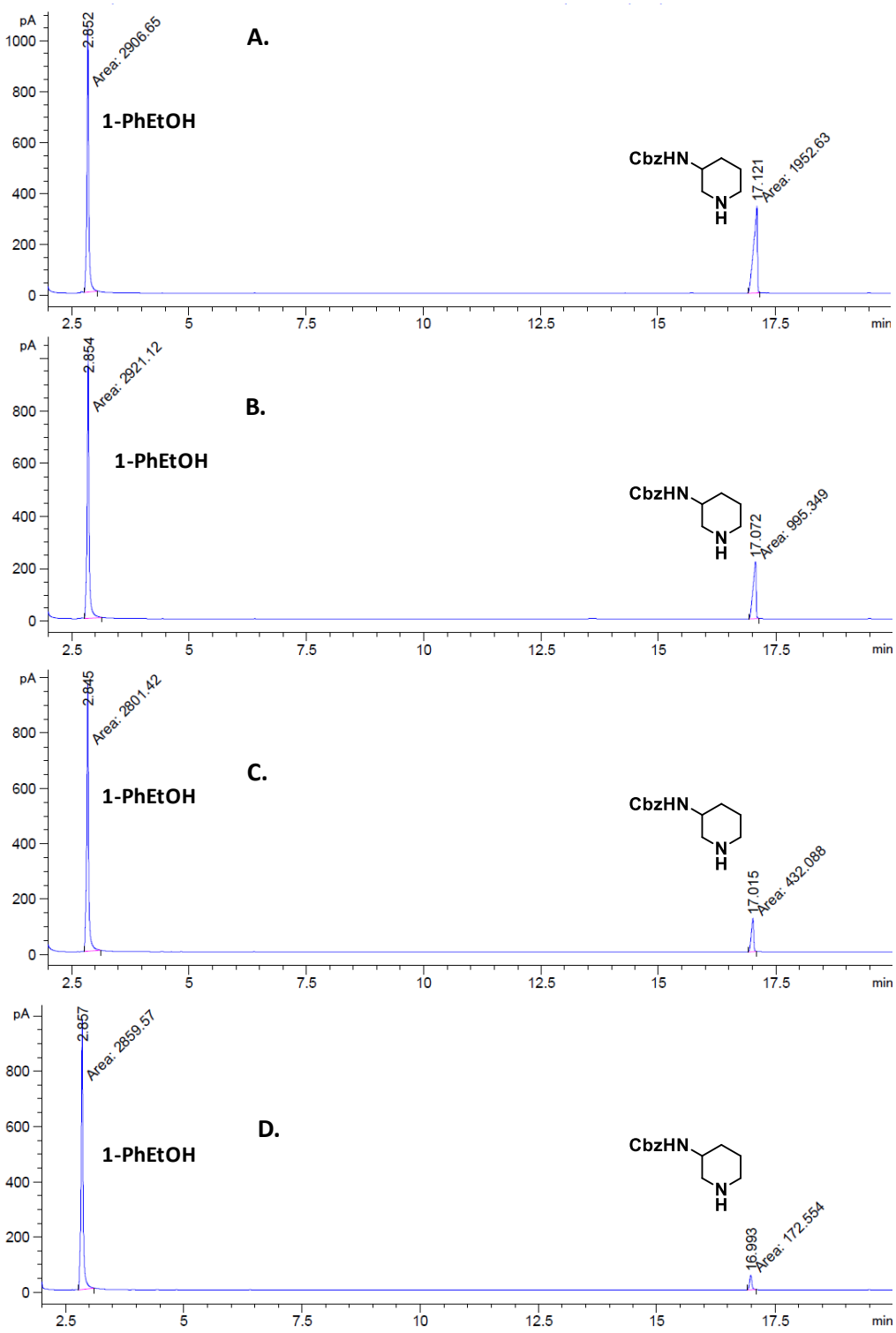


Figure 93 GC chromatograms used for the construction of the calibration curve used to determine the conversion to product amine **11d** for reactions using substrate **12d**. A) 1 mg mL^{-1} **11d**. B) 0.5 mg mL^{-1} **11d**. C) 0.25 mg mL^{-1} **11d**. D) 0.125 mg mL^{-1} **11d**. (HP-1MS (Agilent, 30.0 m x 0.32 mm x 0.25 μm), injector temperature 200°C, detector temperature 250 °C . Method: 120 °C - 250 °C, 5 °C min^{-1}, hold at 250 °C for 5 min; carrier gas helium.).

6.4.1.2 High Performance Liquid Chromatography Methods for Consumption of Keto Acid Substrates **1b**-**f** and ee Determination

HPLC equipped with a 250 mm x 4.6 mm, 5 μm CHIRALPAK®IC column was used for the analysis of keto acid substrate consumption. Solvent: n-hexane/isopropanol/trifluoroacetic acid (90/10/0.1), 1 mL min^{-1}, 265 nm. Retention times for keto acids **1b**-**f** using this method were reported previously.[40]

Determination of ee for **5c**-**d** was carried out by normal phase HPLC using a Daicel CHIRALPAK®IC column (250 mm × 4.6 mm, 5 μm), solvent: n-hexane/isopropanol/diethylamine = 98/2/0.1 (**5c**) or 90/10/0.1 (**5d**), 1 mL/min, 265 nm.

6.4.2 Extraction Protocol and Sample Preparation

For extraction of the amine products **5** and **11** (and keto aldehyde **2** and **9**, imine **3** and **10** and keto alcohol **6** intermediates), reactions were quenched after 24 h with 500 μL dichloromethane (DCM, containing 1 mg mL^{-1} decane external standard for reactions using substrate **1b**, **h** and **i**, and 1 mg mL^{-1} 1-PhEtOH for substrate **12d**) and basified to pH 12.0 using sodium hydroxide (NaOH, 10M). Samples were vortexed (2 m), separated using a benchtop centrifuge (16000 rcf, 5 m), and the organic layer dried over anhydrous magnesium sulfate ($MgSO_4$) before being transferred to a clean vial ready for analysis by GC-FID on a chiral column and GCMS.

For analysis of keto acid **1** (starting material) consumption, samples were extracted as before but acidified to pH 4.0 using hydrochloric acid (HCl, 37%). Analysis of keto acid was performed using HPLC.

To determine the ee of some the chiral amine products **5** by GC-FID on a chiral column, derivatisation of the sample was required. In these cases, the samples were extracted as described above and acetic anhydride (5 μL) and triethylamine (10 μL) were added. The mixture was vortexed (2 m) before the addition of H_2O (100 μL). The mixture was vortexed again and separated using a benchtop centrifuge (16000 rcf, 5 m), with the organic phase then dried over anhydrous $MgSO_4$ and transferred to a clean vial ready for analysis.

Prior to organic extraction, whole cell CapW biotransformation samples using substrates **14b-t** were lysed using a sonication bath for 20 m and then centrifuged to remove cell debris. Reactions were extracted twice with DCM (2 x 800 μL), DCM was removed completely and the resulting solution was redissolved in acetonitrile/water (150/150 μL). Samples were subsequently analysed using LESA-MS or UPLC-MS (Orbitrap) (Dr. Cunyu Yan).

6.5 Biotransformation Conditions

6.5.1 Optimised CAR-TA-IRED Whole Cell Cascade Reactions

For analytical-scale biotransformations, 3 mM keto acid substrate, 50 mM glucose, 250 mM D/L-alanine, 40 mg mL^{-1} *E. coli* BL21 (DE3) cells containing plasmid pLH02, pLH09 or pLH10 in 500 mM sodium phosphate (pH 7.0) reaction buffer (300 μL) was added to a 1.7 mL Eppendorf tube, with a final volume of 500 μL. Reactions were then incubated at 30 °C with shaking (250 rpm) for 24 h to allow full consumption of starting material. Reactions were then extracted following the procedure in Section 6.4 and analysed using GC-FID.

6.5.2 CAR-TA-IRED Hybrid Cascade Reactions

Biotransformations containing 5 mM keto acid substrate, 75 mg mL^{-1} MCAR wet whole cells, 50 mg mL^{-1} (*R*)-IRED wet whole cells, 2.5 mg mL^{-1} ATA-113 or ATA-117, 1 mg mL^{-1} GDH (CDX-901), 0.5 mg mL^{-1} LDH (LDH-103), 250 mM racemic D/L-alanine, 100 mM glucose, 1.5 mM, NAD+ and 1 mM PLP in 500 mM pH 7.0 sodium phosphate buffer and 1% v/v DMSO (from addition of substrate as a solution in DMSO) were incubated at 30°C, 250 rpm for 24 h. Reactions were then extracted following the procedure in Section 6.4 and analysed using GC-FID.

6.5.3 CAR-IRED

Biotransformations containing 3 mM substrate, 0.5 mg mL^{-1} CAR (variable), 1 mg mL^{-1} *Ad*RedAm, 1 mg mL^{-1} GDH (CDX-901), 100 mM glucose, 6 mM ATP, 1.5 mM NADP+ and 60 mM $MgCl_2$ in 100 mM pH 7.5 sodium phosphate buffer up to a total volume of 500 μL (in 2 mL Eppendorf tubes) were incubated at 30°C, 250 rpm for 24 h. Reactions were then extracted following the procedure in Section 6.4 and analysed using GC-FID.

6.5.4 ADH-IRED

Biotransformations containing 3 mM substrate, 0.5 mg mL^{-1} ADH (variable), 1 mg mL^{-1} *Ad*RedAm, 0.5 mg mL^{-1} GDH (CDX-901), 25 mM glucose, 1 mM NAD^+ or $NADP^+$ in 100 mM pH 7.5 sodium phosphate buffer up to a total volume of 500 μL (in 2 mL Eppendorf tubes) were incubated at 30°C, 250 rpm for 24 h.

Reactions were then extracted following the procedure in Section 6.4 and analysed using GC-FID.

6.5.5 GOase-IRED

Biotransformations containing 3 mM substrate, 0.064 mg mL^{-1} HRP, 0.08 mg mL^{-1} catalase, 0.25 mg mL^{-1} GOase (F_2 or M_{3-5}), 1 mg mL^{-1} *Ad*RedAm, 0.2 mg mL^{-1} GDH (CDX-901), 25 mM glucose, 0.5 mM NADP+ in 100 mM pH 7.5 sodium phosphate buffer up to a total volume of 500 μL (in 2 mL Eppendorf tubes) were incubated at 30°C, 250 rpm for 24 h. Reactions were then extracted following the procedure in Section 6.4 and analysed using GC-FID.

6.5.6 CAR Cyclisation

Biotransformations containing 5 mM substrate, 0.1 mg mL^{-1} CAR (variable), 10 mM ATP and 100 mM $MgCl_2$ in 100 mM pH 7.5 sodium phosphate buffer up to a total volume of 500 μL (in 2 mL Eppendorf tubes) were incubated at 25 °C, 170 rpm for 24 h. Reactions were then extracted following the procedure in Section 6.4 and analysed using GC-FID.

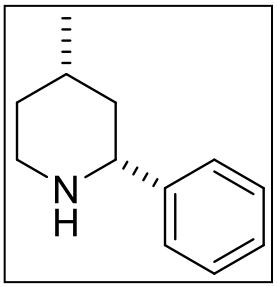
6.5.7 CapW

Biotransformations using CapW were attempted both with whole cells and with cell-free extract (prepared following the procedure in Section 6.3.2.4). For whole cell preparations, cell pellets were resuspended in potassium phosphate buffer

(50 mM, pH 8.0) prior to reaction. Methyl ester substrates **14b**-**t** were dissolved in reaction buffer where possible, or alternatively DMSO (5% final reaction concentration). 100 μL or 500 μL reactions containing ester **14b**-**t** (20 mM), L-aminocaprolactam **15a** (4 mM) and either whole cell suspension or cell-free extract of CapW were incubated with shaking overnight at 30 °C. Reactions were then extracted following the procedure in Section 6.4 and analysed using LESA-MS or UPLC-MS (Orbitrap) (Dr. Cunyu Yan).

6.5.8 Preparative-Scale Biotransformations

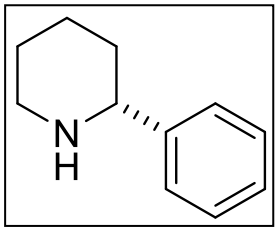
6.5.8.1 4-methyl-2-phenylpiperidine (CAR-TA-IRED), (±)-*cis*-**5f**



Following the optimised expression and reaction conditions described in Section 6.5.1, preparative-scale synthesis of (±)-*cis*-4-methyl-2-phenylpiperidine, (±)-*cis*-**5f**, was achieved through conversion of **1f** (3 mM, 100 mg, 0.49 mmol) using cells harbouring pLH10 in a 500 mL baffled flask. After 24 h the biotransformation was basified to pH 12.0 using 10 M NaOH and extracted twice into EtOAc. The crude extract was subjected to further purification by dissolving the residue into EtOAc (20 mL) and extracting the amine product into 1 M HCl (3 x 20 mL). The aqueous layers were then combined, basified with 10 M NaOH to pH 12.0 and the product extracted into EtOAc (4 x 25 mL). The organic layers were then combined, dried over anhydrous $MgSO_4$ and the solvent removed under reduced pressure to yield (±)-*cis*-**5f** (50 mg, 0.29 mmol, 59%, de >98%) as a yellow oil: **^{1}H-NMR** δH (400 MHz, $CDCl_3$) 7.41-7.30 (m, 4H), 7.29-7.23 (m, 1H), 3.62 (dd, *J* = 11.3, 2.5 Hz, 1H), 3.22 (ddd, *J* = 11.6, 4.1, 2.3 Hz, 1H), 2.82 (app td, *J* = 12.0, 2.5 Hz, 1H), 2.15 (br s, 1H), 1.85-1.77 (m, 1H), 1.74-1.58 (m, 2H), 1.30-1.15 (m, 2H), 0.97 (d, *J* = 6.4 Hz, 3H) ; **^{13}C-NMR δC** (100 MHz $CDCl_3$) 145.2,

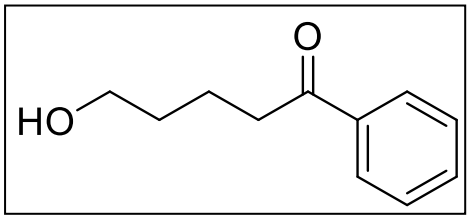
128.4, 127.1, 126.7, 61.9, 47.3, 43.5, 34.4, 32.0, 22.5. **MS** *m*/*z* 175[M^+]. Data consistent with literature values.[40]

6.5.8.2 2-phenylpiperidine (CAR-TA-IRED), (*S*)-**5b**



Following the optimised expression and reaction conditions described in Section 6.5.1, preparative-scale synthesis of (*S*)-2-phenylpiperidine, (*S*)-**5b**, was achieved through conversion of **1b** (3 mM, 144 mg, 0.75 mmol) using cells harbouring pLH10 in a 500 mL baffled flask. After 24 h the biotransformation was basified to pH 12.0 using 10 M NaOH and extracted twice into EtOAc. The crude extract was subjected to further purification by dissolving the residue into EtOAc (20 mL) and extracting the amine product into 1 M HCl (3 x 20 mL). The aqueous layers were then combined, basified with 10 M NaOH to pH 12.0 and the product extracted into EtOAc (4 x 25 mL). The organic layers were then combined, dried over anhydrous $MgSO_4$ and the solvent removed under reduced pressure to yield (*S*)-**5b** (70 mg, 0.43 mmol, 58%, ee 30%) as a yellow oil: **^{1}H-NMR** δH (400 MHz, $CDCl_3$) 7.42-7.30 (m, 4H), 7.30-7.23 (m, 1H), 3.70-3.58 (m, 1H), 3.24-3.15 (m, 1H), 2.80 (td, *J* = 11.6, 3.0 Hz, 1H), 1.95-1.87 (m, 2H), 1.86-1.77 (m, 1H), 1.72-1.65 (m, 1H), 1.65-1.45 (m, 3H); **^{13}C-NMR** δC (100 MHz $CDCl_3$) 145.3, 128.4, 127.1, 126.7, 62.3, 47.7, 34.8, 25.8, 25.4; **MS** *m*/*z* 161 [M^+]. Data consistent with literature values.[163]

6.5.8.3 Keto alcohol by-product (CAR-TA-IRED), **6b**



An adaptation of the reaction conditions outlined in Section 6.5.1 was used for the preparative-scale synthesis of keto alcohol intermediate 5-hydroxy-1-phenyl-1-pentanone, through conversion of **1b** (5 mM, 75 mg, 0.46 mmol) using glucose (50 mM) and cells harbouring plasmid pPB01/MCAR + *Bs*Sfp in a 250 mL baffled flask. After 24 h the biotransformation was basified to pH 12.0 using 10 M NaOH and extracted twice into EtOAc. The organic layers were then dried over $MgSO_4$ and the solvent removed under reduced pressure to yield 5-hydroxy-1-phenyl-1-pentanone **6b** (63 mg, 0.35 mmol, 91 %) as a yellow oil: **^{1}H-NMR** δH (400 MHz, CDCl3) 8.00-7.94 (m, 2H), 7.60-7.55 (m, 1H), 7.51-7.44 (m, 2H), 3.69 (t, *J* = 6.4 Hz, 2H), 3.05 (t, *J* = 7.1 Hz, 2H) 2.63 (2H, s), 1.91-1.81 (m, 2H), 1.72-1.63 (m , 2H); **^{13}C-NMR δC** (100 MHz $CDCl_3$) 200.5, 136.9, 133.0, 128.6, 128.0, 62.4, 38.1, 32.3, 20.1; **MS** *m*/*z* 160 [M - H_2O]$^+$. Data consistent with literature values.[164]

7 References

[1] P. T. Anastas, J. C. Warner, *Green Chemistry: Theory and Practice*, Oxford University Press, New York, **1998**.

[2] B. M. Nestl, S. C. Hammer, B. A. Nebel, B. Hauer, *Angew. Chemie - Int. Ed.* **2014**, *53*, 3070–3095.

[3] C. M. Clouthier, J. N. Pelletier, *Chem. Soc. Rev.* **2012**, *41*, 1585–1605.

[4] Y. Ni, E. Fernández-Fueyo, A. G. Baraibar, R. Ullrich, M. Hofrichter, H. Yanase, M. Alcalde, W. J. H. van Berkel, F. Hollmann, *Angew. Chemie - Int. Ed.* **2016**, *55*, 798–801.

[5] B. Wu, W. Szymański, G. G. Wybenga, M. M. Heberling, S. Bartsch, S. de Wildeman, G. J. Poelarends, B. L. Feringa, B. W. Dijkstra, D. B. Janssen, *Angew. Chemie - Int. Ed.* **2012**, *51*, 482–486.

[6] H. S. Toogood, N. S. Scrutton, *Curr. Opin. Chem. Biol.* **2014**, *19*, 107–115.

[7] P. S. Coelho, E. M. Brustad, A. Kannan, F. H. Arnold, *Science* **2013**, *339*, 307–310.

[8] K. Napora-Wijata, G. A. Strohmeier, M. Winkler, *Biotechnol. J.* **2014**, *9*, 822–843.

[9] V. Gotor-Fernández, E. Busto, V. Gotor, *Adv. Synth. Catal.* **2006**, *348*, 797–812.

[10] K. Faber, W.-D. Fessner, N. J. Turner, Eds. *Biocatalysis in Organic Synthesis 1-3*, *Science of Synthesis*, Georg Thieme Verlag, Stuttgart, Germany, **2015**.

[11] H. Renata, Z. J. Wang, F. H. Arnold, *Angew. Chemie - Int. Ed.* **2015**, *54*, 3351–3367.

[12] J. C. Moore, F. H. Arnold, *Nat. Biotechnol.* **1996**, *14*, 458–467.

[13] K. B. Otte, B. Hauer, *Curr. Opin. Biotechnol.* **2015**, *35*, 16–22.

[14] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185–194.

[15] R. G. Rosenthal, B. Vögeli, N. Quade, G. Capitani, P. Kiefer, J. A. Vorholt, M.-O. Ebert, T. J. Erb, *Nat. Chem. Biol.* **2015**, *11*, 398–402.

[16] M. J. Abrahamson, E. Vázquez-Figueroa, N. B. Woodall, J. C. Moore, A. S. Bommarius, *Angew. Chemie - Int. Ed.* **2012**, *51*, 3969–3972.

[17] N. J. Turner, *Nat. Chem. Biol.* **2009**, *5*, 567–573.

[18] L. Giger, S. Caner, R. Obexer, P. Kast, D. Baker, N. Ban, D. Hilvert, *Nat. Chem. Biol.* **2013**, *9*, 494–8.

[19] C. Letondor, N. Humbert, T. R. Ward, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 4683–4687.

[20] D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, et al., *Nature* **2008**, *453*, 190–5.

[21] G. Mukunda, K. A. Oye, S. C. Mohr, *Polit. Life Sci.* **2009**, *28*, 2–26.

[22] M. Schmidt, *Syst. Synth. Biol.* **2008**, *2*, 1–6.

[23] C. A. Rabinovitch-Deere, J. W. K. Oliver, G. M. Rodriguez, S. Atsumi, *Chem. Rev.* **2013**, *113*, 4611–4632.

[24] W. Weber, M. Fussenegger, *Nat. Rev. Genet.* **2012**, *13*, 21–35.

[25] C. Rogers, G. E. D. Oldroyd, *J. Exp. Bot.* **2014**, *65*, 1939–1946.

[26] X. Wang, J. G. Yang, L. Chen, J. L. Wang, Q. Cheng, R. Dixon, Y. P. Wang, *PLoS One* **2013**, *8*, DOI 10.1371/journal.pone.0068677.

[27] A. S. Khalil, J. J. Collins, *Nat. Rev.* **2010**, *11*, 367–379.

[28] D. E. Cameron, C. J. Bashor, J. J. Collins, *Nat. Publ. Gr.* **2014**, *12*, 381–390.

[29] A. R. Brochado, C. Matos, B. L. Møller, J. Hansen, U. H. Mortensen, K. R. Patil, *Microb. Cell Fact.* **2010**, *9*, 84.

[30] C. J. Paddon, J. D. Keasling, *Nat. Rev. Microbiol.* **2014**, *12*, 355–367.

[31] S. P. France, L. J. Hepworth, N. J. Turner, S. L. Flitsch, *ACS Catal.* **2017**, *7*, 710–724.

[32] T. Bayer, S. Milker, T. Wiesinger, F. Rudroff, M. D. Mihovilovic, *Adv. Synth. Catal.* **2015**, *357*, 1587–1618.

[33] J. Muschiol, C. Peters, N. Oberleitner, M. D. Mihovilovic, *Chem. Commun.* **2015**, *51*, 5798–5811.

[34] C. Schmidt-Dannert, F. Lopez-Gallego, *Microb. Biotechnol.* **2016**, *9*, 601–609.

[35] V. Köhler, N. J. Turner, *Chem. Commun.* **2015**, *51*, 450–64.

[36] R. R. Chen, *Appl. Microbiol. Biotechnol.* **2007**, *74*, 730–738.

[37] B. R. Glick, *Biotechnol. Adv.* **1995**, *13*, 247–261.

[38] N. Oberleitner, A. K. Ressmann, K. Bica, P. Gärtner, M. W. Fraaije, U. Bornscheuer, F. Rudroff, M. Mihovilovic, *Green Chem.* **2017**, *19*, 367–371.

[39] J. Liu, Z. Li, *ACS Catal.* **2013**, *3*, 908–911.

[40] S. P. France, S. Hussain, A. M. Hill, L. J. Hepworth, R. M. Howard, K. R. Mulholland, S. L. Flitsch, N. J. Turner, *ACS Catal.* **2016**, *6*, 3753–3759.

[41] J.-W. Song, E.-Y. Jeon, D.-H. Song, H.-Y. Jang, U. T. Bornscheuer, D.-K. Oh, J.-B. Park, *Angew. Chemie - Int. Ed.* **2013**, *125*, 2534–2537.

[42] E. Liardo, N. Ríos-Lombardía, F. Morís, J. González-Sabín, F. Rebolledo, *Org. Lett.* **2016**, *18*, 3366–3369.

[43] P. A. Santacoloma, J. M. Woodley, in *Cascade Biocatalysis: Integrating Stereoselective and Environmentally Friendly Reactions* (Eds.: S. Riva, W.-D. Fessner), Wiley-VCH Verlag, **2014**, pp. 231–248.

[44] J. Wachtmeister, D. Rother, *Curr. Opin. Biotechnol.* **2016**, *42*, 169–177.

[45] I. Ardao, E. T. Hwand, A.-P. Zeng, *Adv. Biochem. Eng. Biotechnol.* **2013**, *137*, 153–184.

[46] E. O'Reilly, C. Iglesias, D. Ghislieri, J. Hopwood, J. L. Galman, R. C. Lloyd, N. J. Turner, *Angew. Chemie - Int. Ed.* **2014**, *53*, 2447–2450.

[47] P. Both, H. Busch, P. P. Kelly, F. G. Mutti, N. J. Turner, S. L. Flitsch, *Angew. Chemie - Int. Ed.* **2016**, *55*, 1511–1513.

[48] A. Jakoblinnert, D. Rother, *Green Chem.* **2014**, *16*, 3472–3482.

[49] R. S. Heath, M. Pontini, S. Hussain, N. J. Turner, *ChemCatChem* **2016**, *8*, 117–120.

[50] G. L. Rosano, E. A. Ceccarelli, *Front. Microbiol.* **2014**, *5*, 1–17.

[51] H. Tseng, K. L. J. Prather, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17925–17930.

[52] M. J. Sheppard, A. M. Kunjapur, S. J. Wenck, K. L. J. Prather, *Nat. Commun.* **2014**, *5*, 1–10.

[53] S. Wu, Y. Zhou, T. Wang, H.-P. Too, D. I. C. Wang, Z. Li, *Nat. Commun.* **2016**, *7*, 1–13.

[54] T. Knight, *Idempotent Vector Design for Standard Assembly of Biobricks*, **2003**.

[55] J. C. Anderson, J. E. Dueber, M. Leguia, G. C. Wu, J. A. Goler, A. P. Arkin, J. D. Keasling, *J. Biol. Eng.* **2010**, *4*, 1–12.

[56] J. Ellinger, C. Schmidt-Dannert, *Plasmid 90* **2017**, *90*, 1–4.

[57] J. E. Vick, E. T. Johnson, S. Choudhary, S. E. Bloch, F. Lopez-Gallego, P. Srivastava, I. B. Tikh, G. T. Wawrzyn, C. Schmidt-Dannert, *Appl. Microbiol. Biotechnol.* **2011**, *92*, 1275–1286.

[58] M. Kang, J. Lee, Y. Um, T. S. Lee, M. Bott, S. J. Park, H. M. Woo, *Appl. Microbiol. Biotechnol.* **2014**, *98*, 5991–6002.

[59] I. B. Tikh, M. Held, C. Schmidt-Dannert, *Appl. Microbiol. Biotechnol.* **2014**, *98*, 3111–3119.

[60] S. Liu, L. Ruban, Y. Wang, Y. Zhou, D. N. Nesbeth, *Heliyon* **2017**, *3*, 1–20, DOI:10.1016/j.heliyon.2017.e00238.

[61] C. D. Smolke, *Nat. Biotechnol.* **2009**, *27*, 1099–1102.

[62] G. Hannig, S. C. Makrides, *Trends Biotechnol.* **1998**, *16*, 54–60.

[63] K. B. Otte, J. Kittelberger, M. Kirtz, B. M. Nestl, B. Hauer, *ChemCatChem* **2014**, *6*, 1003–1009.

[64] J. E. Dueber, G. C. Wu, G. R. Malmirchegini, T. S. Moon, C. J. Petzold, A. V Ullal, K. L. J. Prather, J. D. Keasling, *Nat. Biotechnol.* **2009**, *27*, 753–761.

[65] F. A. Kondrashov, *Proc. R. Soc. B* **2012**, DOI 10.1098/rspb.2012.1108.

[66] K. T. Elliott, L. E. Cuff, E. L. Neidle, *Future Microbiol.* **2013**, *8*, 887–99.

[67] D. Christ, J. W. Chin, *Protein Eng. Des. Sel.* **2008**, *21*, 121–125.

[68] Z. Chen, C.-H. C. Cheng, J. Zhang, L. Cao, L. Chen, L. Zhou, Y. Jin, H. Ye, C. Deng, Z. Dai, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 12944–12949.

[69] C. J. Brown, K. M. Todd, R. F. Rosenzweig, *Mol. Biol. Evol.* **2017**, *15*, 931–942.

[70] T. Von Rozycki, D. H. Nies, *Antonie Van Leeuwenhoek* **2009**, *96*, 115–139.

[71] B. E. Stranger, M. S. Forrest, M. Dunning, C. E. Ingle, C. Beazley, N. Thorne, R. Redon, C. P. Bird, A. De Grassi, C. Lee, et al., *Science* **2007**, *315*, 848–854.

[72] K. H. R. Algharrawi, R. M. Summers, M. Subramanian, *Biocatal. Agric. Biotechnol.* **2017**, *11*, 153–160.

[73] J. E. Farnberger, E. Lorenz, N. Richter, V. F. Wendisch, W. Kroutil, *Microb. Cell Fact.* **2017**, *16*, 1–17.

[74] I. Nakamura, Y. Yamamoto, *Chem. Rev.* **2004**, *104*, 2127–2198.

[75] P. Ferraboschi, M. De Mieri, P. Grisenti, M. Lotz, U. Nettekoven, *Tetrahedron: Asymmetry* **2011**, *22*, 1626–1631.

[76] C. Y. Legault, B. Charette, *J. Am. Chem. Soc.* **2005**, *127*, 8966–8967.

[77] V. R. Pattabiraman, J. W. Bode, *Nature* **2011**, *480*, 471–9.

[78] J. S. Carey, D. Laffan, C. Thomson, M. T. Williams, *Org. Biomol. Chem.* **2006**, *4*, 2337–2347.

[79] J. Muschiol, C. Peters, N. Oberleitner, M. D. Mihovilovic, U. T. Bornscheuer, F. Rudroff, *Chem. Commun.* **2015**, *51*, 5798–5811.

[80] J. E. Bailey, *Science* **1991**, *252*, 1668–1675.

[81] N. J. Turner, E. O'Reilly, *Nat. Chem. Biol.* **2013**, *9*, 285–288.

[82] A. P. Green, N. J. Turner, *Perspect. Sci.* **2016**, *9*, 42–48.

[83] M. K. Akhtar, N. J. Turner, P. R. Jones, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 87–92.

[84] A. He, T. Li, L. Daniels, I. Fotheringham, J. P. N. Rosazza, *Appli. Environ. Microbiol.* **2004**, *70*, 1874–1881.

[85] M. Moura, D. Pertusi, S. Lenzini, N. Bhan, L. J. Broadbelt, K. E. J. Tyo, *Biotechnol. Bioeng.* **2016**, *113*, 944–952.

[86] A. S. Lamm, P. Venkitasubramanian, J. P. N. Rosazza, in *Biocatalysis in Organic Synthesis 2, Science of Synthesis* (Eds.: K. Faber, W.-D. Fessner, N.J. Turner), Georg Thieme Verlag, Stuttgart, Germany, **2015**, pp. 459–478.

[87] D. Schwendenwein, G. Fiume, H. Weber, F. Rudroff, M. Winkler, *Adv. Synth. Catal.* **2016**, *358*, 3414–3421.

[88] W. Finnigan, A. Thomas, H. Cromas, B. Gough, R. Snajdrova, J. Adams, J. A. Littlechild, N. Harmer, *ChemCatChem* **2016,** *9*, 1005–1017.

[89] D. Gahloth, M. S. Dunstan, D. Quaglia, E. Klumbys, M. P. Lockhart-Cairns, A. M. Hill, S. R. Derrington, N. S. Scrutton, N. J. Turner, D. Leys, *Nat. Chem. Biol.* **2017**, *13*, 975–984.

[90] L. E. N. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber, C. T. Walsh, *Biochemistry* **1998**, *2960*, 1585–1595.

[91] A. N. Khusnutdinova, R. Flick, A. Popovic, G. Brown, A. Tchigvintsev, B. Nocek, K. Correia, J. C. Joo, R. Mahadevan, A. F. Yakunin, *Biotechnol. J.* **2017**, *12*, 1600751.

[92] T. Bayer, S. Milker, T. Wiesinger, M. Winkler, M. D. Mihovilovic, F. Rudroff, *ChemCatChem* **2017**, *9*, 2919–2923.

[93] K. Napora-Wijata, K. Robins, A. Osorio-Lozada, M. Winkler, *ChemCatChem* **2014**, *6*, 1089–1095.

[94] M. S. Malik, E. Park, J. Shin, *Appl. Microbiol. Biotechnol.* **2012**, *94*, 1163–1171.

[95] R. C. Simon, N. Richter, E. Busto, W. Kroutil, *ACS Catal.* **2014**, *4*, 129–143.

[96] J. Shin, B. Kim, *Biotechnol. Bioeng.* **1999**, *65*, 206–211.

[97] F. Guo, P. Berglund, *Green Chem.* **2017**, *19*, 333–360.

[98] I. Slabu, J. L. Galman, N. J. Weise, R. C. Lloyd, N. J. Turner, *ChemCatChem* **2016**, *8*, 1038–1042.

[99] D. Monti, M. C. Forchin, M. Crotti, F. Parmeggiani, F. G. Gatti, E. Brenna, S. Riva, *ChemCatChem* **2015**, *7*, 3106–3109.

[100] H. Land, P. Hendil-Forssell, M. Martinelle, P. Berglund, *Catal. Sci. Technol.* **2016**, *6*, 2897–2900.

[101] T. Sehl, H. C. Hailes, J. M. Ward, R. Wardenga, E. Von Lieres, H. Offermann, R. Westphal, M. Pohl, D. Rother, *Angew. Chemie - Int. Ed.* **2013**, *52*, 6772–6775.

[102] C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, et al., *Science* **2010**, *329*, 305–309.

[103] G. Grogan, N. J. Turner, *Chem. Eur. J.* **2016**, *22*, 1900–1907.

[104] J. Mangas-Sanchez, S. P. France, S. L. Montgomery, G. A. Aleku, H. Man, M. Sharma, J. I. Ramsden, G. Grogan, N. J. Turner, *Curr. Opin. Chem. Biol.* **2017**, *37*, 19–25.

[105] F. Leipold, S. Hussain, D. Ghislieri, N. J. Turner, *ChemCatChem* **2013**, *5*, 3505–3508.

[106] S. Hussain, F. Leipold, H. Man, E. Wells, S. P. France, K. R. Mulholland, G. Grogan, N. J. Turner, *ChemCatChem* **2015**, *7*, 579–583.

[107] K. Mitsukura, M. Suzuki, K. Tada, T. Yoshida, T. Nagasawa, *Org. Biomol. Chem.* **2010**, *8*, 4533–5.

[108] M. Rodríguez-Mata, A. Frank, E. Wells, F. Leipold, N. J. Turner, S. Hart, J. P. Turkenburg, G. Grogan, *ChemBioChem* **2013**, *14*, 1372–1379.

[109] K. Mitsukura, T. Kuramoto, T. Yoshida, N. Kimoto, H. Yamamoto, T. Nagasawa, *Appl. Microbiol. Biotechnol.* **2013**, *97*, 8079–8086.

[110] G. A. Aleku, S. P. France, H. Man, J. Mangas-Sanchez, S. L. Montgomery, M. Sharma, F. Leipold, S. Hussain, G. Grogan, N. J. Turner, *Nat. Chem.* **2017**, *9*, 961–969.

[111] P. N. Scheller, M. Lenz, S. C. Hammer, B. Hauer, B. M. Nestl, *ChemCatChem* **2015**, *7*, 3239–3242.

[112] S. P. France, R. M. Howard, J. Steflik, N. J. Weise, J. Mangas-Sanchez, S. L. Montgomery, R. Crook, R. Kumar, N. J. Turner, *ChemCatChem* **2018**, *10*, 510-514.

[113] A. Siebum, A. van Wijk, R. Schoevaart, T. Kieboom, *J. Mol. Catal. B Enzym.* **2006**, *41*, 141–145.

[114] K. Parikka, E. Master, M. Tenkanen, *J. Mol. Catal. B Enzym.* **2015**, *120*, 47–59.

[115] M. Fuchs, K. Tauber, J. Sattler, H. Lechner, J. Pfeffer, W. Kroutil, K. Faber, *RSC Adv.* **2012**, *2*, 6262–6265.

[116] B. Bechi, S. Herter, S. Mckenna, C. Riley, S. Leimkühler, N. J. Turner, A. J. Carnell, *Green Chem.* **2014**, *16*, 4524–4529.

[117] S. Herter, S. M. Mckenna, A. R. Frazer, S. Leimkühler, A. J. Carnell, N. J. Turner, *ChemCatChem* **2015**, *7*, 2313–2317.

[118] K. Parikka, M. Tenkanen, *Carbohydr. Res.* **2009**, *344*, 14–20.

[119] F. Escalettes, N. J. Turner, *ChemBioChem* **2008**, *9*, 857–860.

[120] J. B. Rannes, A. Ioannou, S. C. Willies, G. Grogan, C. Behrens, S. L. Flitsch, N. J. Turner, *J. Am. Chem. Soc.* **2011**, *133*, 8436–8439.

[121] S. M. Mckenna, S. Leimkühler, S. Herter, N. J. Turner, A. J. Carnell, *Green Chem.* **2015**, *17*, 3271–3275.

[122] W. Kroutil, H. Mang, K. Edegger, K. Faber, *Curr. Opin. Chem. Biol.* **2004**, *8*, 120–126.

[123] F. G. Mutti, T. Knaus, N. S. Scrutton, M. Breuer, N. J. Turner, *Science* **2015**, *349*, 1525–1529.

[124] S. L. Montgomery, J. Mangas-Sanchez, M. P. Thompson, G. A. Aleku, B. Dominguez, N. J. Turner, *Angew. Chemie - Int. Ed.* **2017**, *129*, 10627–10630.

[125] N. Oberleitner, C. Peters, J. Muschiol, M. Kadow, S. Saß, T. Bayer, P. Schaaf, N. Iqbal, F. Rudroff, M. D. Mihovilovic, et al., *ChemCatChem* **2013**, *5*, 3524–

3528.

[126] R. Finking, M. A. Marahiel, *Annu. Rev. Microbiol.* **2004**, *58*, 453–88.

[127] M. V Fawaz, M. E. Topper, S. M. Firestine, *Bioorg. Chem.* **2011**, *39*, 185–91.

[128] M. Funabashi, Z. Yang, K. Nonaka, M. Hosobuchi, Y. Fujita, T. Shibata, X. Chi, S. G. Van Lanen, *Nat. Chem. Biol.* **2010**, *6*, 581–586.

[129] A. He, T. Li, L. Daniels, I. Fotheringham, J. P. N. Rosazza, *Appl. Environ. Microbiol.* **2004**, *70*, 1874–1881.

[130] A. J. L. Wood, N. J. Weise, J. D. Frampton, M. S. Dunstan, M. A. Hollas, S. R. Derrington, R. C. Lloyd, D. Quaglia, F. Parmeggiani, N. J. Turner, et al., *Angew. Chemie - Int. Ed.* **2017**, DOI: 10.1002/anie.201707918.

[131] L. N. Gautam, Y. Su, N. G. Akhmedov, J. L. Petersen, X. Shi, *Org. Biomol. Chem.* **2014**, *12*, 6384–8.

[132] S. Mathew, H. Yun, *ACS Catal.* **2012**, *2*, 993–1001.

[133] D. Koszelewski, K. Tauber, K. Faber, W. Kroutil, *Trends Biotechnol.* **2010**, *28*, 324–332.

[134] C. E. Paul, M. Rodriguez-Mata, E. Busto, I. Lavandera, V. Gotor-Fernandez, V. Gotor, S. Garcia-Cerrada, J. Mendiola, O. De Frutos, I. Collado, *Org. Process Res. Dev.* **2014**, *18*, 788–792.

[135] D. Wetzl, M. Berrera, N. Sandon, D. Fishlock, M. Ebeling, M. Müller, S. Hanlon, B. Wirz, H. Iding, *ChemBioChem* **2015**, *16*, 1749–1756.

[136] P. N. Scheller, S. Fademrecht, S. Hofelzer, J. Pleiss, F. Leipold, N. J. Turner, B. M. Nestl, B. Hauer, *ChemBioChem* **2014**, *15*, 2201–2204.

[137] D. Koszelewski, I. Lavandera, D. Clay, D. Rozzell, W. Kroutil, *Adv. Synth. Catal.* **2008**, *350*, 2761–2766.

[138] H. Zhao, W. A. Van Der Donk, *Curr. Opin. Biotechnol.* **2003**, *14*, 583–589.

[139] R. D. Woodyer, T. W. Johannes, H. Zhao, in *Enzym. Technol.* (Eds.: A. Pandey, C. Webb, C.R. Soccol, C. Larroche), Springer, New Delhi, **2006**, pp. 85–104.

[140] C. Wong, D. G. Drueckhammer, H. M. Sweers, *J. Am. Chem. Soc.* **1985**, *107*, 4028–4031.

[141] B. C. Joseph, S. Pichaimuthu, S. Srimeenakshi, M. Murthy, K. Selvakumar, G. M, S. R. Manjunath, *J. Cell. Sci. Ther.* **2015**, *6*, 1–7.

[142] J. Kim, J. P. Kershner, Y. Novikov, R. K. Shoemaker, S. D. Copley, *Mol. Syst. Biol.* **2010**, *6*, 436.

[143] A. M. Kunjapur, Y. Tarasova, K. L. J. Prather, *J. Am. Chem. Soc.* **2014**, *136*, 11644–11654.

[144] K. T. Elliott, E. L. Neidle, *IUBMB Life* **2011**, *63*, 1075–1080.

[145] F. A. Kondrashov, *Proc. R. Soc. B* **2012**, *279*, 5048–5057.

[146] M. D. Truppo, J. D. Rozzell, J. C. Moore, N. J. Turner, *Org. Biomol. Chem.* **2009**, *7*, 395–398.

[147] J. M. Cox, B. Harper, A. Mastracchio, B. Leiting, R. S. Roy, R. A. Patel, J. K. Wu, K. A. Lyons, H. He, S. Xu, et al., *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4579–4583.

[148] J. M. Sutton, D. E. Clark, S. J. Dunsdon, G. Fenton, A. Fillmore, N. V Harris, C. Higgs, C. A. Hurley, S. L. Krintel, R. E. Mackenzie, et al., *Bioorg. Med. Chem. Lett.* **2012**, *22*, 1464–1468.

[149] Z. Lai, C. Li, J. Liu, L. Kong, X. Wen, H. Sun, *Eur. J. Med. Chem.* **2014**, *83*, 547–560.

[150] L. Zhao, Y. Zhang, C. Dai, T. Guzi, D. Wiswell, W. Seghezzi, D. Parry, T. Fischmann, M. A. Siddiqui, *Bioorg. Med. Chem. Lett.* **2010**, *20*, 7216–7221.

[151] Y. Zhang, Y. Bu, J. Liu, *Folia Microbiol.* **2015**, *60*, 393–398.

[152] J. Qin, Y. J. Zhou, A. Krivoruchko, M. Huang, L. Liu, S. Khoomrung, V. Siewers, B. Jiang, J. Nielsen, *Nat. Commun.* **2015**, *6*, 1–11.

[153] S. A. Weissman, N. G. Anderson, *Org. Process Res. Dev.* **2015**, *19*, 1605–1633.

[154] M. Moura, D. Pertusi, S. Lenzini, N. Bhan, L. J. Broadbelt, K. E. J. Tyo, *Biotechnol. Bioeng.* **2016**, *113*, 944–952.

[155] C. T. Walsh, *Science* **2004**, *303*, 1805–10.

[156] L. Heide, *Nat. Prod. Rep.* **2009**, *26*, 1241–50.

[157] V. Gotor, V. Gotor-Fernández, E. Busto, in *Compr. Chirality*, **2012**, pp. 101–121.

[158] X. Liu, Y. Jin, W. Cai, K. D. Green, A. Goswami, S. Garneau-Tsodikova, K. Nonaka, S. Baba, M. Funabashi, Z. Yang, et al., *Org. Biomol. Chem.* **2016**, *14*, 3956–3962.

[159] D. Hartinger, S. Heinl, H. E. Schwartz, R. Grabherr, G. Schatzmayr, D.

Haltrich, W.-D. Moll, *Microb. Cell Fact.* **2010**, *9*, 1–14.

[160] J. R. Blackwell, R. Horgan, *FEBS Lett.* **1991**, *295*, 10–2.

[161] N. Oganesyan, I. Ankoudinova, S.-H. Kim, R. Kim, *Protein Expr. Purif.* **2007**, *52*, 280–285.

[162] K. Montgomery, L. Raymundo, W. L. Drew, *J. Clin. Microbiol.* **1979**, *9*, 205–207.

[163] H. Prokopcova, S. D. Bergman, K. Aelvoet, V. Smout, W. Herrebout, B. Van Der Veken, L. Meerpoel, B. U. W. Maes, *Chem. - A Eur. J.* **2010**, *16*, 13063–13067.

[164] J. Xie, L. Guo, X. Yang, L. Wang, Q. Zhou, *Org. Lett.* **2012**, *14*, 4758–4761.