Department of Chemical and Environmental Engineering

Understanding carbon flux of a heterologous metabolic pathway in *Escherichia coli* for sustainable methyl methacrylate biosynthesis

Clemency Tilley

Supervisors: Alex Conradie and Ian Kerr

Submitted to the University of Nottingham for the degree of Doctor of Philosophy

May 2022

Abstract

Three thousand kilotons of methyl methacrylate (MMA) are produced each year. The polymer of MMA, also known as Perspex®, has become ubiquitous in applications from construction to medical technologies since it was first synthesised in 1936. Presently, MMA is exclusively manufactured *via* chemical synthesis from petrochemical feedstocks, 96% of which is acetone. However, in recent years biosynthesis of an MMA precursor, butyl methacrylate (BMA), has been demonstrated in *Escherichia coli*. This is characterised by the up-regulation of 2-ketoisovalerate production, followed by expression of heterologous branched-chain β-ketoacid dehydrogenase, acyl-CoA oxidase (ACX), and alcohol acyltransferase (AAT). Decoupling industrial MMA production from the petrochemical industry by achieving industrial fermentation of BMA would provide a green alternative to Perspex® at a time where a shift towards low carbon technologies is increasingly supported politically, economically, and societally.

Microbial fermentation is a powerful platform for sustainable resource production to displace current fossil fuel-based manufacturing practices and has gradually gained traction in recent years. In this project, I worked to understand and improve the industrial viability of BMA biosynthesis in *E. coli* strain BW25113. Despite the assumption that product toxicity would be the primary hindrance to high BMA productivity, the maximum titres that were synthesised from a series of BMA-resistant *E. coli* mutants were 1.4 mM from biotransformation and 0.13 mM in logarithmic growth, well below the IC_{50} for BMA in *E. coli*. In examining BMA producer strains, I determined that extraneous (off target) butyl esters are formed during BMA biosynthesis: butyl acetate, butyl isobutyrate (BIB), butyl propionate, and butyl isovalerate being the most prevalent. This led me to develop a series of experiments to identify and circumvent bottlenecks, by assessing carbon flux through individual stages of BMA biosynthesis. I identified a production bottleneck at the final two steps in synthesis. These concern the oxidation of isobutyryl-CoA to methacrylyl-CoA by acyl-CoA oxidase 4 from *Arabidopsis thaliana* (AtACX4) and methacrylyl-CoA conversion to BMA as catalysed by a mutant alcohol acyltransferase (AATm4). I

determined that AtACX4 has a low K_i of 32.8 µM for its product methacrylyl-CoA, which prevents adequate intracellular methacrylyl-CoA accumulation to ensure sufficient specificity and activity from AATm4. I used bioinformatics to identify phylogenetically related but diverse ACX and AAT enzymes to replace AtACX4 and AATm4. I followed this with Golden Gate assembly to subsequently generate an ACX-AATm4 variant library, and 8 combinatorial ACX-AAT libraries. To facilitate screening of such genomic diversity I took advantage of a novel BMA-reactive fluorescent probe to develop a semi-quantitative screening approach. Using this plate-based screen, I isolated two new ACX4s with comparable activity to ACX4 from *A. thaliana*, from *Zea mays* and *Vigna radiata*, as well as identifying an ACX3 from *Spinacea oleracea* with improved ester selectivity.

During this project I determined that the highest product titre achievable using the existing BMA production pathway was 0.175 mM. Furthermore, de-bottlenecking experiments revealed a significant carbon flux hold up at methacrylyl-CoA: AATm4 preferentially produces off target esters with a ratio of 1:74 BMA:BIB, whilst failing to compete with endogenous esterase isobutyryl-CoA consumption, at a scale of 63 mM isobutyric acid as compared to 0.026 mM BMA. My *in vitro* assay work on AtACX4 revealed it is substantially inhibited by low methacrylyl-CoA concentrations. My subsequent bioinformatics and screening approaches resulted in the identification of 6 novel ACX4 enzymes active on isobutyryl-CoA, as well as an ACX3 enzyme. I also developed and implemented a Golden Gate assembly system capable of more efficiently swapping in alternative oxidases and transferases into the BMA pathway. To date, the industrial target of 2 g L^{-1} h^{-1} BMA has not yet been achieved, but understanding of the limitations on carbon flux through the BMA pathway in *E. coli* has been expanded. This provides guidance for future engineering towards industrial bioproduction of BMA.

Acknowledgements

First of all, I would like to thank my supervisors, Ian Kerr and Alex Conradie for their academic input and supervision, as well as my bench supervisor Carolina Paiva and my previous primary supervisor Gill Stephens. Secondly, I would like to thank the BBSRC and Mitsubishi Chemical UK for funding this project, as well as several of the team from Mitsubishi Chemical for providing guidance and support throughout my PhD. Particular thanks go to Graham Eastham, David Johnson, Kenji Oki, Mark Reynolds, Steve Cartman, and Mark Waugh.

I would also like to thank the members of Sustainable Process Technologies, past and present, who have trained me in the lab and provided useful insights to further my work. In particular, Andrew Yiakoumetti, Luca Rossoni, Victoria Outram, Floriane Martins, Sophie Vaud, Sean Craig, Amy Pearson, Steve Hall, and Harry Padley. Thank you also to Jeni Luckett for her help in accessing and operating the IVIS Spectrum.

Last, but not least, thank you to my friends and family for their support and encouragement throughout my PhD, also to Manchee for moral support and an excuse to go outside both during lockdown and the writing up period.

Context statement

The principal supervisor for this project changed over the period 2018-2019 due to the phased retirement of Professor Gill Stephens. Professor Alex Conradie became the principal supervisor from that point onwards.

Due to the COVID-19 pandemic access to laboratory facilities and in-person supervision was not possible for a period of six months from March 2020 to August 2020. After this period, limited access was permitted to laboratory facilities at either alternative weeks or reduced hours until July 2021. No access to the IVIS Spectrum was authorised between March 2020 and February 2021.

Table of Contents

Table of Tables

Table of Figures

Abbreviations

13PD	1,3-propanediol
2-KIV	2-ketoisovalerate
3-HIBA	3-hydroxyisobutyrate
3-HIBH	3-hydroxyisobutyryl-CoA hydrolase
AAT	Alcohol acetyltransferase
ACAD	Acyl-CoA dehydrogenase
ACN	Acetonitrile
ACT	4-hydroxybenzoyl-CoA thioesterase
ACX	Acyl-CoA oxidase
ALE	Adaptive lab evolution
ATL	Alcohol acyltransferase library
AvAAT	Avocado alcohol acetyltransferase
BA	Butyl acetate
BAAT	Banana alcohol acetyltransferase
BCKD	Branched chain ß-ketoacid dehydrogenase
BIB	Butyl isobutyrate
BIV	Butyl isovalerate
BLAST	Basic Local Alignment Search Tool
BMA	Butyl methacrylate
BPI	Butyl propionate
BRENDA	BRaunschweig ENzyme Database
BVMO	Baeyer-Villiger monooxygenase
DTZ	Diaryl tetrazole
ECH	Enoyl-CoA hydratase
ETF	Electron transferring flavoprotein
ETFQ	Electron transferring flavoprotein Q oxidoreductase

FA	Fatty acid
FAD	Flavin adenine dinucleotide
GD	Glutaryl-CoA dehydrogenase
HA	Hexyl acetate
HADH	Hemiacetal dehydrogenase
HIB	Hexyl isobutyrate
HMA	Hexyl methacrylate
IBA	Isobutyric acid
IBBE	Industrial biotechnology and bioenergy
IB-CoA	Isobutyryl-CoA
IBD	Isobutyryl-CoA dehydrogenase
ISPR	*In situ* product recovery
IVD	Isovaleryl-CoA dehydrogenase
IVIS	*in vivo* imaging system
LC	Long-chain
LCAD	Long chain acyl-CoA dehydrogenase
MAAT	Apple alcohol acyl transferase
MBP	Maltose-binding protein
MC	Medium chain
MCAD	Medium chain acyl-CoA dehydrogenase
M-CoA	Methacrylyl-CoA
MD	Molecular dynamics
MMA	Methyl methacrylate
MSA	Multiple sequence alignment
PAAT	Pear alcohol acetyltransferase
PHB	Polyhydroxy butyrate
PLA	Polylactic acid
PMMA	(Poly) methyl methacrylate

PZ-BMA	Pyrazoline-butyl methacrylate conjugate
RNAP	RNA polymerase
SAAT	Strawberry alcohol acetyltransferase
SBCAD	Short, branched chain acyl-CoA dehydrogenase
SC	Short chain
SCAD	Short chain acyl-CoA dehydrogenase
SPA	Solid phase assay
SSM	Single-site mutation
TFA	Trifluoroacetic acid
VAAT	Wild strawberry alcohol acetyltransferase
VLCAD	Very long chain acyl-CoA dehydrogenase
WS	Wax synthase
WT	Wild type

Chapter 1: Literature review

1.1 Industrial biotechnology and bioenergy

1.1.1 Industrial biotechnology in the current market

Oil production forms the basis of our industrialised society, and in recent years the debate about transitioning to low-carbon energy has culminated in predictions of a future 'peak oil' or 'peak demand', after which oil production will steadily decline [1, 2]. This peak is expected to occur within the next 10-15 years [3]. Concerns about the insecurity of petrochemical resources, increasing political tensions, as well as steadily growing fuel costs and disruptive price fluctuations, also create financial incentive to gradually decrease our dependence on petrochemical resources [2]. Industrial biotechnology has potential to stabilise, and even reduce, the complexity and running costs of petrochemical-based processes.[4] This is already having impact in the replacement of common fuels with products such as bioethanol [5], and industrial biotechnology is projected to provide £4.5 billion p/a in added value to the UK economy through both direct use and downstream industry [6].

The term 'biotechnology' was first coined in 1919, and comprises the production of useful raw materials from living organisms [7]. However, the practice of biotechnology has existed for thousands of years, encompassing a vast range of applications from fermentation with brewer's yeast, to preserving milk using acid fermentation [8, 9]. Biotechnology was initially concerned with production of primary and secondary metabolites from carefully curated process conditions to increase yield, such as production of penicillin and amino acids from *Penicillum notatum* and *Clostridium glutamicum*, respectively. Biotechnology has more recently repurposed enzymes for synthesizing large biomolecules and chemicals through rational design [9, 10]. Examples include isobutanol from *Escherichia coli*, artemisinin from *Saccharomyces cerevisiae*, and bioprocessing poly (3-hydroxybutyrate) from glucose in *Cupriavidus necator* [11].

Industrial biotechnology and bioenergy (IBBE), or white biotechnology, is a subdivision of biotechnology with huge potential to mitigate an impending oil crisis, by enabling the production of industrially useful materials through biological processes in large-scale fermentation [12]. Development of IBBE has been facilitated by a number of key technological advances such as recombinant DNA technology, microbiological understanding, and fermentation technology and protein purification [13]. These approaches allow efficient redirection of cellular resources. More recently, introducing synthetic biology, adaptive evolution, systems biology, bioinformatics, and fast, cheap DNA sequencing have facilitated the design of biocatalysts to fit process specifications [10, 14]. For example, 1,3-propanediol (13PD), which is a building block for many industrial polymers, and was chosen as a target by DuPont. A combination of synthetic biology and systems biology was used to produce *E. coli* capable of fermentation from glucose to 13PD. Today a number of biofuels and commodity chemicals are produced from fermentation on a commercial scale with bioengineered microorganisms, including isobutene from Global Bioenergies, isobutanol from Gevo, ethanol from Lallemand, and a variety including 1,4-butanediol, scleroglucan polysaccharides and ß-carotene from BASF [15-17].

Bioprocesses have the potential to be developed to synthesize a variety of target chemicals which drop directly into existing production chains. They can use petrochemical feedstocks to improve upon the target purity, or carry out fully consolidated bioprocessing, and can often provide unparalleled enantio- and regio-selectivity [18, 19]. Bioprocesses such as this are sustainable, are far less likely to produce hazardous by-products, are extremely selective, and ultimately have potential to uncouple manufacturing from the oil industry and increase the atom efficiency from feedstock to product [12, 18, 20].

In the UK, the private sector has invested £564 million into synthetic biology and biotechnology between 2009-2016, compared to £300 million from the UK government [21]. Private industry investment in biotechnology has grown rapidly in the

years since, recorded at over £1 billion in 2019, while public investment continues to lag behind [22]. However large, established industry has been slow to take up bio-based products, largely because of the decrease in profit, e.g. bio-polyethene and bio-polypropylene sell at -30% profit compared to petrochemical-based plastics [23], regulatory constraints, and the inconvenience of any disruption to established supply chains. Despite the potential for a bio-based materials industry, the commercialised sector currently makes up just 3% of the entire UK bioeconomy, while just 11% of private investment in biotechnology goes to industrial purposes [22, 24]. The majority of this is populated by bio-based, non-biodegradable polyethylene, polyethylene tetra phthalate or polyamides, which themselves only make up 1% of the 320 million tonnes of plastic manufactured annually [25].

It also remains challenging to successfully commercialise IBBE processes, particularly manufacturing non-native chemicals [14, 26]. In order for biocatalysts to be used in this way, industry needs to also persuade end users that bio-based chemicals are advantageous in terms of economic viability, as 'green' chemicals that are neither cheaper nor provide advantageous properties in the final product will only appeal to a subset of environmentally conscious consumers. Four key criteria must therefore be met: (a) price, which is usually at a premium for bioproducts and must be the same or lower than petrochemically-sourced products, (b) perception, or demonstrating that a bioproduct is of superior purity or quality to justify a higher price, and (c) management, whereby the supply of bio-based products needs to be reliable to the point where it justifies disrupting an industry's embedded supply chains [27]. Therefore, high yields and the ability to develop cheap but reliable fermentation and separation processes in demonstration and pilot scale bioreactors is vital [18].

1.1.2 Commercialising bioprocesses for industrially useful chemicals

Existing bioprocesses typically enhance production of native compounds, such as organic acids and amino acids, as this is lower risk and can reach near theoretical yields in some cases, such as succinate overproduction from *E. coli* [28]. The most

successful commercialised IBBE processes all produce short-chain alcohols, particularly n-butanol, 13PD, polylactic acid (PLA) and (poly)hydroxybutyrate (PHB) [28]. One large manufacturing site, owned by DuPont, produces 3.5 g L^{-1} h^{-1} of 13PD at up to 99.7% purity through *E. coli* fermentation of glucose [15], using a $100 million plant in Tennessee [29]. 13PD is naturally produced in many microbes, but even non-natural products can be commercially bioprocessed, for example 1,4-butanediol (14BDO) [30]. Genomatica first commercialised production of 14BDO from *E. coli* and can now produce 3 g L^{-1} h^{-1} of 14BDO [4, 31]. Verdezyme also commercially synthesizes dodecanedioic acid from plant oil fermentation by yeast, in place of butadiene [29]. In contrast, commercialising a bioprocess for xenobiotic chemicals is a relatively novel area of industrial biotechnology. The design of novel bioprocessing pathways in biocatalysts traditionally follows three key stages of process design to develop successfully scaled-up bioprocesses: upstream, midstream and downstream (Figure 1-1) [32].

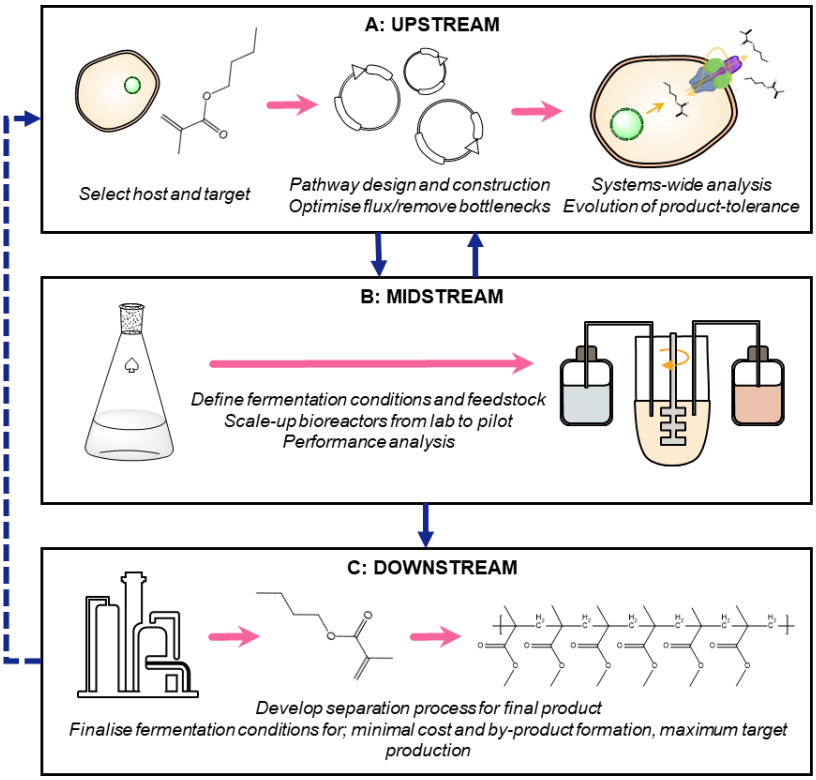


Figure 1-1: Summary of the industrial bioprocess engineering workflow

Upstream process design (Figure 1-1A) involves selecting microbial chassis, identifying a chemical target, and characterising putative metabolic pathways, whether native or heterologous, to achieve target synthesis with as few steps from central metabolism as possible [19]. The selected chassis is ideally a well-characterised microorganism with a number of tools and methods for synthetic biology already available [33]. *E. coli* is commonly used due to its genetic tractability, well-characterised genome and rapid doubling time [32]. However, non-traditional hosts have been commercialised with non-xenobiotic products, such as glyceric acid production to 101.8 g L^{-1} and 136.5 g L^{-1} produced from *Gluconobacter rateurii* and *Acetobacter tropicalis*, respectively [34], as well as 124.2 g L^{-1} citric acid from *Yarrowia lipolytica* [35].

Next, candidate enzymes are identified that have broad specificity and produce molecules with a similar structure to the target chemical. Identified enzymes can be transplanted and repurposed from non-industrially useful organisms [36]. For example, in isobutanol biocatalysts, the host *ilvIHCD* (Ehrlich pathway) was linked to heterologous α-ketoisovalerate decarboxylase (*kivD*) from *Lactococcus lactis*. *kivD* was identified after a family of broad substrate 2-ketoacid decarboxylases were overexpressed in *E. coli* and screened for activity with 2-ketoisovalerate using gas chromatography-mass spectrometry (GC-MS) [37].

Once a biosynthetic pathway towards the target chemical is assembled, biocatalysts usually need to be tuned to increase flux and tackle metabolic bottlenecks, which remains a significant challenge in many synthetic bioprocesses. For a process using a sugar-based carbon source, which most commercialised bioprocesses do, flux must generally be improved sufficiently to achieve at least an 80% yield to make the process industrially feasible [28]. Perhaps the most common method used to improve production titres is via systematic metabolic engineering, which can be time consuming and may show varying degrees of success. For example, D-pantothenic acid titres were increased from 0.49 g L^{-1} to 28.45 g L^{-1} during *E. coli* fermentation,

using a rational approach varying promotor strength, co-factor pools, and knocking down genes using CRISPR-Cas9 [38].

A similar approach in agmatine biosynthesis resulted in industrially relevant titres of 40.43 g L^{-1} from 1.26 g L^{-1}, also in *E. coli* [39]. Other successful approaches also include adaptive evolution, protein engineering and, more recently, fermentative co-cultures; several rosmarinic acid *E. coli* production strains were co-cultured with selectivity for alternative carbon sources, and achieved a 38-fold increase in productivity [40]. Many of these approaches rely on knowledge of where key carbon flux hold-ups are within a novel production pathway.

In fact, identifying bottlenecks in synthetic metabolic pathways is often vital in targeting metabolic engineering to where the greatest impact will be seen. This can be achieved *in vitro* by individually isolating low enzyme activity or subpar substrate preference, promoters and ribosome binding site variation can be used to alter expression, and 'omics approaches can build a full picture of the pathway [28]. Used synergistically, these methods can accurately identify carbon flux hold-ups, as well as themselves improving final titres: for example, a library of 3125 strains with different promoter strengths for the *vioA*-*vioE* genes was generated to synthesize violacein culminating in varying expression of these genes. Screening the most promising library fraction produced strains with 62% improvement in violacein production compared to systems using just strong promoters [41]. Despite the variety of methods available, this stage of bioprocess commercialisation is far from exact, and often requires multiple generations of low producing strains, as well as slow, iterative improvements in flux.

Tuning metabolic pathways is frequently concomitant with evolving tolerance to toxic targets, particularly when heterologous molecules are produced [42]. Isobutanol, which is biosynthesised on the large scale, has a negative effect on the growth of wild-type (WT) *E. coli*, *S. cerevisiae* and *Bacillus subtilis* at concentrations of 50 mM [43]. To even be considered for industrial scale-up, most organisms will need to

tolerate roughly 50 g L^{-1} product, or 2 g L^{-1} h^{-1}; indeed, productivity as high as 3.5 g L^{-1} h^{-1} was reported for 13PD and succinate production plants [28]. During ethanol fermentation at titres of 50 g L^{-1} product inhibition was identified as the primary limitation on yields, and a biphasic hexadecane system was previously used to increase yields of isobutyl acetate to 80% [44-46]. Even naturally robust host organisms will almost certainly require improved tolerance, or *in situ* product recovery (ISPR) will be required to continuously remove the target compound from fermentation to prevent end product inhibition [28, 47]. A recent example of host tolerance engineering being used in conjunction with ISPR to improve end product titres was in the case of cadaverine production from *E. coli* [48]. In this instance, random mutagenesis was carried out on *E. coli* expressing lysine decarboxylases for cadaverine production. Three cadaverine tolerance proteins were identified: HokD, PhnI and PuuR. Overexpression of these genes, teamed with increased cadaverine export, resulted in increased cadaverine titres from 1.48 g L^{-1} h^{-1} to 58.7 g L^{-1} h^{-1} from fed-batch fermentation.

Development of biocatalysts often requires a back and forth between upstream processing, and midstream fermentation [32] (Figure 1-1B) in order to co-develop bioprocesses with fermentation optimisation, and the scale-up of bioreactor volumes [32]. Midstream process development facilitates the real-world scale up to >10,000 L, usually using batch and fed-batch fermentation, with a transition to continuous becoming increasingly viable [49]. Fed-batch fermentation is the traditional, and most common, process used in a large scale for white biotechnology [50]. During fed-batch fermentation all required nutrients are provided in the initial culture medium, and once these are consumed a feed is initiated to provide a new supply of nutrients, allow culture growth to continue [51]. Continuous bioprocessing however, which involves the continuous feeding and removal of nutrients and waste, theoretically reduces product heterogeneity and system downtime. The primary hindrance to utilisation of continuous fermentations on an industrial scale is the need to develop biocatalysts that can maintain a stable productivity throughout the lengthy process lifetime [50, 52].

Downstream process development traditionally occurs late stage in the design pipeline, once a high flux, high titre, and otherwise industrially feasible process has been developed (Figure 1-1C). Factors including separation cost-effectiveness, feedstock availability and price, as well as robustness of cell lines, must also be considered before a biocatalyst can be implemented on a commercial scale [27].

The key issue with this traditional design approach is that downstream separation, which accounts for a sizeable portion of the final bioprocess cost [53], is not often considered at the upstream stages. This occasionally results in late changes to the target chemical [54, 55]. Every additional purification step reduces recovery yield, which itself is limited most often by sensitivity of the biocatalyst to low concentrations of the non-natural product [56]. An under-adopted approach is therefore to tackle biocatalysis in a more synergistic way, where downstream separation efficacy is given particular bias. For example, gaseous hydrocarbon butadiene is initially presented as a difficult target for microbial biocatalysis due to its low occurrence in nature and cellular toxicity, but can very easily be extracted from fermentation broth with gas stripping, as has been patented by Genomatica and Versalis [57, 58].

A significant number of chemical targets have proved much harder to scale-up to the required level because of unfavourable process economics: this includes styrene bioproduction in *E. coli*, which only reached 0.26 g L^{-1} [59]. Production of bio-isoprene by Genecor (DuPont), also from *E. coli*, reached similarly low titres (1.83 g L^{-1}) [57]. Amyris' attempts to biosynthesize trans-ß-farnesene as a new alternative building block replacing butadiene and isoprene were similarly unsuccessful because scaling-up to the required 40 million litres *per annum* fermentation volume proved unsuccessful [60]. In 2017 they had yet to implement a process despite several patents for the manufacture of bio-based farnesene derived products from sugar fermentation by yeast [61].

Bio-based (poly) methyl methacrylate (PMMA) is another challenging bioprocess target. It is an acrylic polymer formed from free radical polymerisation of methyl

methacrylate (MMA) (Figure 1-2). Due to the varied applications of its polymer, the global market for MMA is now worth approximately $6 billion *per annum*. A large portion of the 3 million tonnes manufactured each year is sourced from MMA produced by industry leader Mitsubishi Chemical UK (MCUK, formerly Lucite International) [54].

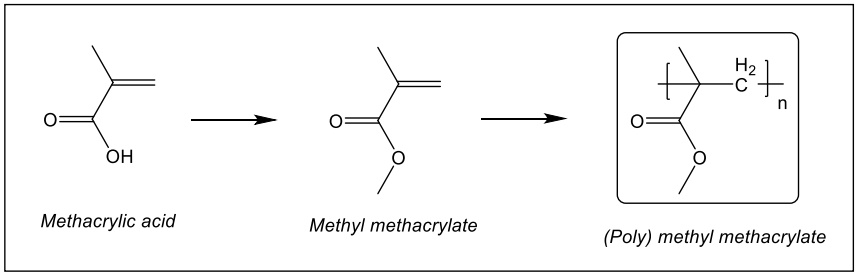


Figure 1-2: (Poly) methyl methacrylate biosynthesis

Both MCUK and Genomatica have patented bioprocesses yielding bio-based MMA intermediates such as methacrylic acid (MAA), citramalate and mesaconic acid [62, 63]. Bioproduction of these precursors is achievable in practice, and several metabolic pathways have been successfully developed. However, implementations of these processes is hindered by either low yields, high product toxicity, or uneconomical separation costs [19, 54, 64].

1.2 Manufacture of acrylic materials

1.2.1 Market and uses of PMMA

Acrylics are thermoplastic polymers which have been manufactured commercially ever since their initial use as the windshields in Spitfire aircraft in 1936 [65]. The most common form of acrylic, PMMA, was first trademarked under the iconic brand Perspex® by ICI in 1934 [66]. After World War II, instead of falling out of common use, manufacturers took advantage of their sizable production capacity for PMMA by devising numerous additional ways it could be applied, initially within construction and manufacturing of corrugated roofing and car headlights, before moving on to consumer goods during the 50s and 60s [65].

These advances were aided by the fact that PMMA possesses a number of beneficial properties. It is durable, transparent, resistant to UV light, less hazardous than glass on breakage, and is lightweight and easily malleable [65]. Varying the proportion of different methacrylate esters in the polymer, with increasing waxiness and malleability as more long chain esters are introduced, also allows further control of the end product properties [67]. Commercial applications of PMMA have been continually adapted to contemporary requirements, such as manufacturing radiation dosimeters [68], with more recent developments capitalising on PMMA's biocompatibility for crafting bone cements [69], and facilitating targeted drug delivery [70].

1.2.2 Petrochemical-based synthesis of MMA and MAA

There are three classes of approach used to manufacture either MMA directly, or an MAA precursor. These are the C2, C3, or C4 routes [66]. However, the principal method employed to synthesize MMA on a commercial scale is the acetone cyanohydrin (ACH) process (a C3 route), as described by ICI in 1931 [65, 71].

The ACH process (Figure 1-3A), begins by reacting acetone with hydrogen cyanide over a basic catalyst to generate acetone cyanohydrin. This is then reacted with excess sulphuric acid to produce methacrylamide sulphate *via* an unstable intermediate. Methacrylamide sulphate can then be hydrolysed to MAA or MMA [66]. The ACH process has an 80-90% yield, but it produces 1.5x more ammonium bisulphate as a by-product than MMA [64]. Sulphuric acid can be regenerated from the ammonium bisulphate, but this must then be stored safely before recycling into the ACH process. Other MMA production routes were initially developed purely to avoid the cost of sulphuric acid regeneration [19].

The other C3 and C4 processes (Figure 1-3C, D) struggle to compete economically with the ACH process. Both produce MAA instead of MMA, thereby adding an esterifying step from MAA to MMA. The C3 process requires reacting propylene with an acid to produce isobutyric acid (Figure 1-3C) [66]. Usually hydrofluoric acid is used,

which itself is highly corrosive. Isobutyric acid is then passed over an Fe-P or Mo catalyst at high temperatures to make MAA, which must be distilled from a mixture with methyl isobutyrate [19]. In the C4 route, isobutylene is oxidised to MAA via two oxidation steps at 300-420°C, over first a mixed MoO_3, Bi_2O_3, or Fe_3O_4 catalyst, then a $H_3Mo_{12}O_{40}P_{72}$ catalyst [64]. This process only achieves yields of 65-75%, with contaminated water produced as a by-product (Figure 1-3D) [64].

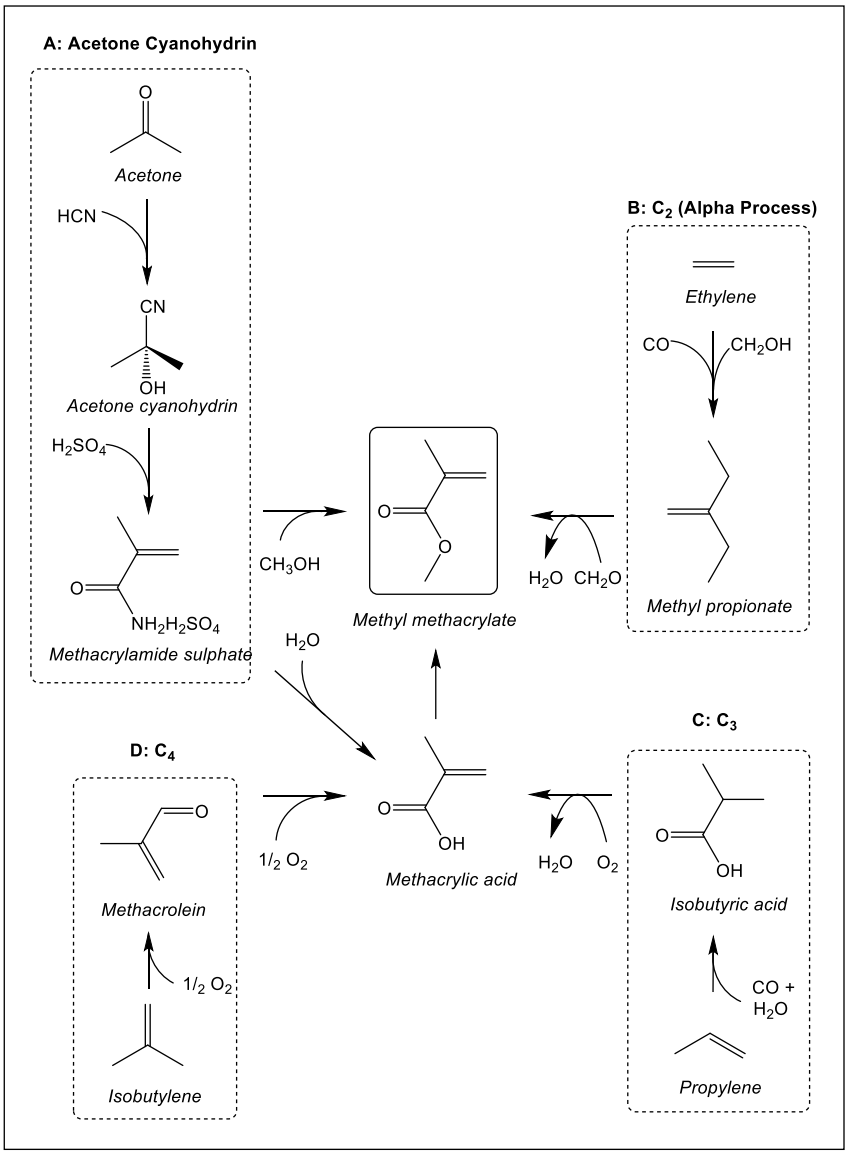


Figure 1-3: Chemical synthesis of methyl methacrylate

The C2 routes to MMA initiate from ethylene as a starting material. The Alpha process (Figure 1-3B) is one example, created by MCUK, and is the second most extensively commercialised method of synthesising MMA, after the ACH process. The first

dedicated $230 million Alpha plant was completed in 2008, and is capable of producing 120,000 tonnes of MMA a year [72]. The Alpha process uses an extremely selective palladium catalyst to produce MMA *via* methyl propionate [73]. Low volumes of complex by-products are formed [64].

Common to all commercial manufacture of MMA or MAA is a reliance, to a lesser or greater extent, on petrochemical feedstocks. In addition to the environmental impact of petrochemicals, many of the associated manufacturing processes use hazardous chemicals that increase risk to health and safety and incur expensive storage and removal costs. For example, the ACH process uses hydrogen cyanide, which is poisonous, and excess, corrosive sulphuric acid (Figure 1-3A). This contrasts with the mild, aqueous process conditions usually required for chemical production through microbial fermentation. As such, there is plenty of cause and incentive for the PMMA manufacturing industry to produce bio-based 'green' Perspex®.

1.3 Development of an industrial biocatalyst for bio-based acrylates

1.3.1 Methacrylic acid and citramalate synthesis

Earliest approaches to synthesise bio-based MMA focussed on producing two targets from metabolically engineered *E. coli*; citramalate and MAA [54]. Initially this was achieved by introducing citramalate synthase into the biocatalyst to form citramalate, which could then be extracted and chemically converted to MAA (Figure 1-4) [64]. Fed-batch fermentation for citramalate produced high yields of 80-90 g L^{-1}, and this process was patented [62, 74]. However, during citramalate synthesis, carbon flux also diverted to acetone, acetate, pyruvate, carbon monoxide and acetaldehyde. Citramalate is also highly soluble in aqueous fermentation conditions, requiring a hot pressurised water process to extract the final product, which led to unfeasible process economics [72]. Direct biosynthesis of MAA was the next target (Figure 1-4A) as it could reduce the number of downstream chemical steps, as well as potentially reducing processing costs compared to citramalate conversion [54].

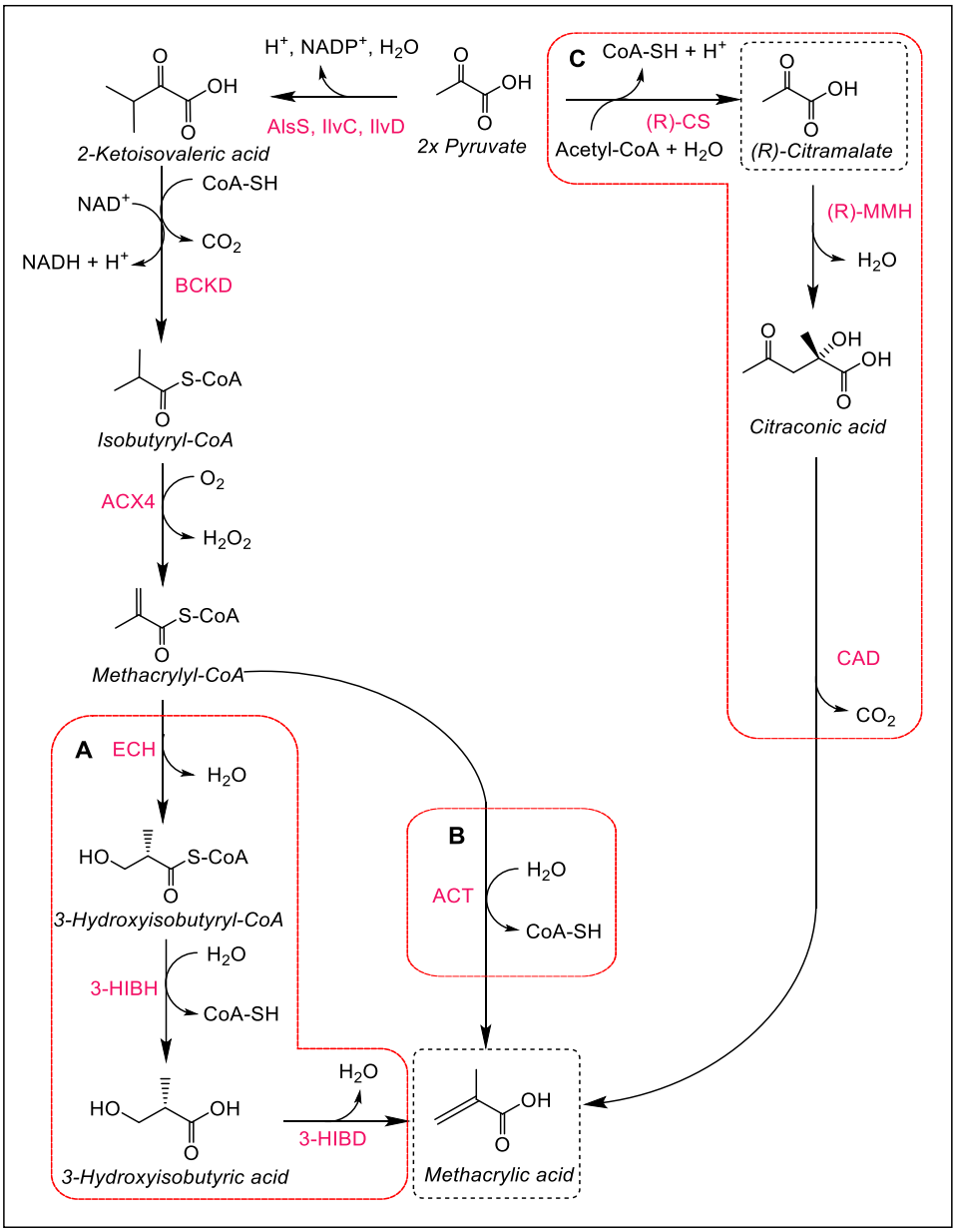


Figure 1-4: Metabolic engineering routes for bio-based methyl methacrylate.

A: 3-HIBA route B: Thioesterase route and C: Citramalate route. BCKD = Branched-chain α-ketoacid dehydrogenase (*Pseudomonas aeruginosa*) EC 1.2.4.4, ACX4 = Acyl-CoA oxidase (*Arabidopsis thaliana*) EC 1.3.3.6, ECH = Enoyl-CoA hydratase (*Escherichia coli*) EC 4.2.1.17, 3-HIBH = 3-Hydroxyisobutyryl CoA hydratase (*E. coli*) EC 3.1.2.4, 3-HIBD = 3-hydroxyisobutyryl-CoA dehydratase (*E. coli*) EC 1.1.1.35, ACT = Acyl-CoA thioesterase (*E. coli*) EC 3.1.2.2, (R)-CS = (R)-Citramalate synthase (*Methanocaldococcus jannaschii*) EC 2.3.1.182, (R)-MMH = (R)-2-Methylmalate hydrolase, and CAD = Citraconate dehydratase.

Initially this was attempted by overproducing 2-KIV in *E. coli* using the often reported AlsS (acetolactate synthase, *B. subtilis*, EC 2.2.1.6), IlvC (acetohydroxylacid isomeroreductase, *E. coli*, EC 1.1.1.86), IlvD (dihydroxyacid dehydratase, *E. coli*, EC 4.2.1.9) pathway from pyruvate [19]. This, combined with heterologous expression of

an α-ketoacid dehydrogenase (BCKD, *Pseudomonas aeruginosa*, EC 1.2.4.4) and acyl-CoA oxidase (AtACX4, *Arabidopsis thaliana*, EC 1.3.3.6), produced the branched chain acyl-CoA methacrylyl-CoA (M-CoA). Then, M-CoA could either be directly converted to MAA by a 4-hydroxybenzoyl-CoA thioesterase (ACT) (Figure 1-4B), or *via* the high flux formation of 3-hydroxyisobutyric acid (3-HIBA), which is subsequently dehydrated to MAA [62, 75, 76]. Direct formation of MAA from citramalate is also possible *via* (R)-2-methylmalate hydrolase and a citraconate dehydrogenase. Unfortunately, as with citramalate, separating MAA from cultures proved too expensive for a cost effective process [54].

After the separation and yield issues arising with citramalate and MAA, the next logical step was to consider tailoring product choice to the requirements for easy separation. Bioproduction of methacrylate esters is a promising solution [67]. Some methacrylate esters have low solubility limits, and may phase separate from water during fermentation, thereby bypassing the extraction problems experienced with citramalate, 3-HIBA, and MAA [77]. This would also bring the endpoint of biosynthesis a step closer to MMA. Methacrylate esters require minimal chemo-catalytic steps to produce the final polymer, where a single transesterification reaction can be used to reach MMA [74]. MCUK has patented a metabolic process similar to the 3-HIBA and thioesterase pathways, in which the final step is catalysed by alcohol acyltransferase (AATm4, MCUK) to form methacrylate ester [62].

1.3.2 Methacrylate esters as a production target, toxicity, and occurrence

From a traditional bioprocess design perspective, methacrylate esters are unlikely targets. They have high toxicity to microorganisms, unfavourable process thermodynamics in cells, and few analogues endogenously synthesized in any organism [54, 77]. However, from a more holistic viewpoint, the simple separation of methacrylate esters and single transesterification step to produce MMA make them an attractive end target. In this case, there are two questions: (a) Is it possible to design a

metabolic pathway towards methacrylate esters? (b) If so, which ester should be produced?

Living organisms do not naturally produce methacrylate esters, and even short chain (SC) esters rarely [78]. The alcohol moiety of methacrylate esters affects their physical properties; SC esters such as MMA are water soluble and volatile, medium chain (MC), such as ethyl methacrylate (EMA) and *n*-butyl methacrylate (BMA), and long chain (LC) esters, as n-dodecyl methacrylate, become increasingly waxy, malleable, and decrease in solubility [67]. Aliphatic LC esters are produced naturally in plants and some microorganisms [79]. In contrast, the transfer of a free methyl group required to form MMA is rarely observed in any organism, making direct MMA biosynthesis unlikely [80]. As such, the MC methacrylate esters from EMA to isopropanyl methacrylate (iPMA) offer a prospective solution to the product extraction problems seen with citramalate, 3-HIBA and MAA. These esters are relatively volatile and modestly insoluble, which may allow easier product separation without preventing transport across the cell membrane.

Zoe Disley (ZD) grew *E. coli* MG1655 on plates containing 20% (v/v) MMA, EMA, BMA and iPMA for 72 hours, BMA resistant strains appeared [54]. Four colonies of these BMA-resistant mutants were isolated and sequenced [77, 81]. High tolerance to BMA developed, despite its low inhibitory concentration (IC_{50}) of 0.07-0.11 g L^{-1}. In contrast, the other methacrylate esters all have IC_{50} values greater than 1.18 g L^{-1} [54]. Interestingly, BMA also has the lowest solubility limit of all the methacrylate esters; 0.37 g L^{-1} at 37°C. For comparison iPMA has the second lowest solubility limit, at 5.90 g L^{-1} [77]. Indeed, extracting ester into a separate phase, or gas stripping, could prevent the product reaching critical concentrations. The relative evolved tolerance to BMA, as well as its low solubility, made it the focus of efforts to produce biosynthesised methacrylate ester.

The mechanism by which BMA causes toxicity to cells is not yet fully understood. However, organic solvents generally exhibit similar properties when it comes to their

interaction with biocatalysts. Many intercalate into cell membranes, causing disruption [82]. There is a toxicity threshold above which the disruption caused by an organic solvent to energy generation, membrane transport and protein denaturation, leads to a massive stress response and cell death [82]. Cell sensitivity to organic solvents, and toxicity of the solvent itself, is quantified using the logarithm of the partition coefficient, in 1-octanol and water ($LogP_{o/w}$) [83]. Values on this scale between 0.7 and 4 are generally considered to be toxic, and many industrially significant esters sit somewhere within this range: ethyl acetate $LogP_{o/w}$ =0.73, butyl acetate $LogP_{o/w}$ = 1.78 and ethyl hexanoate $LogP_{o/w}$ = 2.40 [43]. The $LogP_{o/w}$ of BMA is between 2.29-2.60, the highest $LogP_{o/w}$ of the methacrylate esters MMA, EMA, iPMA and BMA [84]. However, the tolerance limit of *E. coli* in complex medium appears to be a $LogP_{o/w}$ of 4.1, below which many solvents inhibit growth [77].

In addition to the $LogP_{o/w}$, more recent work has confirmed the adverse interaction of BMA with *E. coli*-like artificial cell membranes [85].During these experiments it was shown that BMA most likely intercalates between the phospholipid head groups in the lipid bilayer, thus uncoupling 'collective' motion in the membrane. Additionally, at BMA concentrations over 80%, it was discovered that the saturated phospholipids flip, which would likely kill cells. Interestingly, BMA appeared to have less of an effect on unsaturated lipids, leading to the conclusion that increasing the proportion of unsaturated lipids in host cell membranes may directly improve BMA tolerance.

1.3.3 Engineering strains with product-resistance to butyl methacrylate

Within the group of product-resistant mutants isolated by ZD, mutations in two genes, translating into proteins AcrR and SoxR, conveyed product-tolerance both individually and synergistically [81]. Deletion mutant strains Δ*acrR* and Δ*soxR* were able to reach the same OD_{600} as WT *E. coli* does in the absence of BMA [81].

AcrR (Figure 1-5A) is a repressor for *acrAB*, which forms the inner membrane (IM) and periplasmic components of the AcrAB-TolC efflux pump [82]. AcrAB-TolC is

already known to confer resistance to several antibiotics, as well as tolerance to some alkanes, heptane, nonane and octane [82, 86]. Efflux pumps often demonstrate broad substrate specificity and can increase product titres by removing xenobiotics as they form, thus reducing feedback inhibition of synthetic circuits [42]. Therefore, mutations in the AcrAB-TolC operon that may increase its expression are not unexpected, although previous studies have shown that increasing expression of efflux pumps alone may not be enough to reliably confer resistance [87].

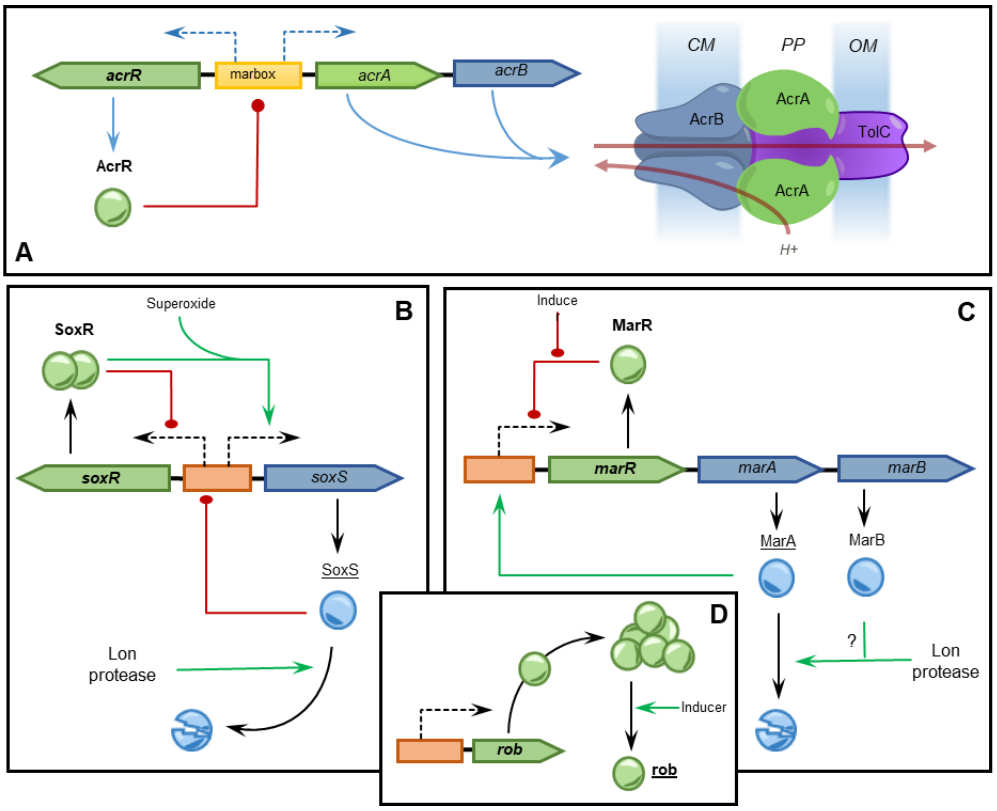


Figure 1-5: Regulation of genes affected by tolerance evolution.

Genes and proteins in bold differ from wild type in one or more BMA resistant mutants. Proteins underlined are global regulators that bind marbox promoters and alter expression of the corresponding genes. Arrows in red indicate repression, arrows in green indicate induction. CM = cytoplasmic membrane, PP = periplasm, OM = outer membrane.

SoxR is a transcriptional regulator (Figure 1-5B) exhibiting negative autoregulation and it activates *soxS* transcription as part of the cellular defence against redox cycling agents, as well by extension superoxide [88, 89]. SoxRS is one set of operons known as the Sox-Mar-Rob regulon, which controls the expression of many genes involved in cellular stress response. SoxS, along with MarA and Rob, bind to conserved

sequences in promoter regions known as ‘marboxes’ to regulate gene expression (Figure 1-5A). Operons that increase expression in response to SoxS include *acrAB*, *sodA* and *fldAB* [88].

Two strategies were pursued to develop tolerance to exogenous BMA in *E. coli* BW25113. Adaptive laboratory evolution (ALE) to generate more resistant strains was investigated by Russel Menchavez (RNM). Laura Martins (LM) used a rational engineering approach to introduce BMA resistance into *E. coli* based upon transcriptomics data from the *E. coli* MG1655 mutants produced by ZD [81, 90]. Over the course of 48 dilutions in batch ALE, BMA concentrations were ramped up to 20% (v/v), after which chemostat cultures were held at a constant dilution rate of 0.46 h^{-1}, and 0.54 h^{-1} in a second round of ALE, at 20% (v/v) BMA throughout [90]. This yielded 12 distinct 20% (v/v) BMA tolerant mutants of *E. coli* BW25113 [90].

All 12 ALE mutants acquired mutations in at least one gene relating to the Sox-Mar-Rob regulon. Either *soxR*, *acrR*, *marR* or *rob*. MarR represses expression of genes *marA* and *marB*, and MarA up- and down- regulates a number of genes in a similar manner to SoxS (Figure 1-5C) [88]. MarB is a small protein which indirectly represses *marRAB* [91]. Rob also regulates the expression of genes with marbox promoters. Rob is constitutively expressed but remains aggregated and inactive until a decanoate inducer is present (Figure 1-5D) [88]. Interestingly, BMA-tolerant mutants with *rob* mutations achieved the uppermost maximum biomass concentrations in 20% (v/v) BMA, the highest reaching 1.036 g L^{-1}, compared to 1.436 g L^{-1} of wild-type in absence of BMA [90]. Three of the isolated mutants had acquired mutations in *acrB* [92]. In addition, a myriad of other mutations were identifiable: *ilvN*, a gene coding for an isozyme for IlvIH, which catalyses first step of endogenous valine synthesis in *E. coli* [93]. Also *rpoB*, which forms the ß subunit of RNA polymerase, *ompR*, which regulates expression of porins OmpF and OmpC in response to osmotic stress, and *phoP* which controls the cellular response to Mg^{2+} starvation and is implicated in the *E. coli* acid stress response [94].

Evolving or engineering tolerance in microorganisms is a complex process. There is a large reservoir of information about the *E. coli* genome Despite this, the multifactorial nature of most stress and tolerance mechanisms in *E. coli*, and all organisms, makes targeted improvements difficult to achieve. Mutation data from BMA tolerance experiments suggests that a number of stress responses are affected by BMA toxicity, including osmotic, acid, and oxidative stress responses. Experimental work expressing AcrAB/TolC related proteins demonstrates that increasing the efflux of BMA from cells also increases viability, as is observed with many solvent targets [95, 96]. On-going work continues to look directly at the effect of BMA on *E. coli* cell membranes, and at fermentation to produce hexyl methacrylate (HMA), as a 'less toxic' alternative to BMA. Biosynthesis of BMA is still in its development stage; however, product resistance will eventually become a significant consideration of any high titre industrial bioprocess.

1.4 Metabolic engineering for the biosynthesis of butyl methacrylate

1.4.1 Ester synthesis in industry and nature

Esters have been produced industrially for many years, and there remains a large market for their synthesis: the global flavour and fragrance ester market was $16.6 billion in 2012, while butyl acetate and ethyl acetate are produced on a scale of 89-107 million Kg and 69-91 million Kg *per annum*, respectively [46, 97]. SC and MC esters are useful as solvents, plasticisers and as drop-ins for biofuels, as is the case for butyl butyrate. MC and SC fatty acyl esters such as ethyl acetate and isoamyl acetate are important materials for the food, flavour, and fragrance industries [43]. Industrial ester production has traditionally been carried out by chemical synthesis using Fisher-Speier esterification, a process developed in 1895 [98]. In this process alcohols and carboxylic acids derived from petrochemical sources are condensed with an acid catalyst at high temperatures. There are several well-established limitations to this method. First, the ΔG for ester hydrolysis is -5 kcal mol^{-1}, which necessitates a high process temperature and acid catalyst to prevent the reverse reaction taking

place [46]. Water is released during the condensation reaction which must be continually removed, as it inhibits the acid catalyst. Finally, the reaction equilibrium during Fisher-Speier esterification causes a decrease in reaction rate with time, only ever leading to incomplete conversion to the final ester [43].

Direct extraction from natural sources is not necessarily a profitable alternative, however. This field has primarily focussed on wax ester extraction from plants, such as sugar cane and jojoba [79, 99]. Both processes produce oils with a prohibitive cost and relatively low availability, but are highly dependent on seasonal growth, intensive farming methods and restrictive geographical factors [100, 101].

Although cultivation and extraction are not the most efficient approach to bulk ester production, an extensive collection of diverse esters are naturally produced by both plants and microbes. These are often complex ester cocktails as expressed from many fruits, varying from anywhere between 100-350 volatile esters during ripening. While these organisms may produce many esters simultaneously, they often form several characteristic volatile products: such as ethyl isovalerate and ethyl hexanoate from *Geotricum* sp. (yeast), hexyl esters and butyl acetate from ‘Pink Lady’ apples, and isoamyl acetate from bananas [97]. In contrast, some microorganisms, which have been explored as industrial biocatalysts, can produce high concentrations of a single ester, such as synthesis of ethyl acetate by *Kluyveromyces marxianus*, and in *Euglena gracilis*, where wax ester can make up 60% dry cell weight [43, 102]. Availability of mechanisms producing a wide range of esters is clearly present, however natural ester formation lacks the specificity and capacity required to support an industrial process without additional metabolic engineering.

Microbes that produce the fruity aromas characteristic of ester formation were first isolated as early as the 1800s, but knowledge about the purpose of these esters in microorganisms remains somewhat obscure [103]. This is largely because ester composition in most organisms is too diverse to attribute a single function [43]. Small C4-C12 volatile esters are posited to protect against pathogens in plants. Conversely,

dimycocerosate esters in the *Mycobacterium tuberculosis* and *Mycobacterium leprae* outer membrane lipids may contribute to virulence [43, 46, 104]. Wax esters appear to facilitate intracellular storage in yeasts or are formed as a product of anaerobic fermentation [43]. Often ester formation in microorganisms is linked to growth in sub-optimal conditions. An example of this is ethyl acetate production in *Pichia anomala* to repress the growth of competing microorganisms, whereas in other organisms ethyl acetate is used to combat challenging nutrient limitations [43].

Perhaps unsurprisingly, due to the wide variety of esters occurring in nature, organisms can synthesise esters though several mechanisms. The four primary synthetic routes are *via* esterases, hemiacetal dehydrogenases (HADHs), Baeyer-Villiger monooxygenases (BVMOs) and alcohol acyltransferases (AATs). Less common mechanisms are S-adenosyl methionine dependant-O-methyltransferases and polyketide associated proteins [43]. There is a disproportionate amount of knowledge about the different ester forming enzymes, which leads to biases in literature and industry favouring esterase use in research, in particular lipase.

Esterases are ubiquitous enzymes that can catalyse the reaction between an alcohol and an acid to form esters. As their name might suggest, ester synthesis is actually the reverse reaction carried out by esterases; under aqueous conditions the esterase reaction has a positive ΔG [105]. In order to synthesise esters instead, the requirements are similar to chemical ester synthesis, in that reaction conditions must be non-aqueous, using organic solvents or very high reagent concentrations in order to favour the reverse reaction [43]. All esterases have an α/β hydrolase fold and a catalytic triad consisting of a histidine residue, and either a serine, aspartate, or glutamate residue. Lipases, which are a subset of esterases, have an additional lid domain that covers their active site under aqueous conditions and a generally higher optimum pH than other esterases. Lipases are used as purified, immobilised catalysts for industrial processes [97, 103].

Hemiacetal dehydrogenation occurs when an alcohol and an aldehyde react together spontaneously to form a hemiacetal, which is then usually converted to an ester by HADH [43]. Strictly speaking, HADH enzymes do not exist, as the hemiacetal dehydrogenation reaction is actually a side activity of many alcohol dehydrogenases. HADH reactions are usually attributed to the synthesis of methyl formate in *Pichia methanolica*, *Candida boidinii* and *S. cerevisiae* [43, 106]. The majority of literature concerning HADH is focussed solely on 'methyl formate synthase' by the involvement of the first discovered HADH, a class III alcohol dehydrogenase.

BVMOs are ubiquitous, NAD(P)H dependant flavoenzymes that directly convert ketones into esters [43]. The reaction is analogous to secondary alcohol oxidation, playing a role in secondary metabolite synthesis and unconventional carbon utilisation. For example, growth on alkanes, ketones, and cyclic alcohols. More metabolic engineering has been carried out using BVMOs than HADH, including the successful cloning of cyclopentanone 1,2-monooxygenase from *Pseudomonas* sp., and *acmA*/*acmB* from *Gordonia* sp. into *E. coli* to produce methyl acetate [107, 108]. ChnB, which is a cyclohexane 1,2-monooxygenase, has been successfully expressed in *E. coli* multiple times, exhibiting activity over a wide range of different ketone substrates [107].

Another broadly specific class of enzymes, the AATs, transfers the acyl group from acyl-CoAs to alcohols to form esters [109]. AATs comprise a large and diverse class of enzymes; each capable of producing a plethora of esters, due to broad substrate ranges. This makes AATs ideal enzymes for tailoring to specific ester products, although structural data linking AAT sequences to their specificity is patchy [97]. By extension substrate promiscuity is an anticipated challenge for integrating AATs into an industrial process. Plenty of research has focussed on the purification and expression of novel AAT enzymes, but kinetic characterisation of alcohol and acyl-CoA substrates for AAT is relatively limited (Table 1-1). The most studied AATs are AftA/B wax ester synthases from *Acinetobacter baylyi* and Atf1/2 from *S. cerevisiae*

[43]. These enzymes play a key role in attracting insects in plants, detoxifying MC fatty ethyl esters from fatty acid synthesis and removing excess acetyl-CoA [110].

CoA substrate	Alcohol substrate																															
	ethanol	butanol	octanol	benzyl alcohol	2-phenylethanol	hexanol	cinnamyl alcohol	methanol	isoamyl alcohol	propanol	geraniol	3-hexen-1-ol	heptanol	isobutanol	Naphthalene	pentanol	3-methyl-1-butanol	2-hexen-1-ol	nonanol	eugenol	isopentanol	decanol	farnesol	2-ethoxyethanol	furfuryl alcohol	2-methyl-1-butanol	linalool	nerol	perilla alcohol	phenylethanol	dodecanol
acetyl	11	10	10	14	12	8	10	6	8	6	7	8	8	5	5	3	1	3	3	3	1	2	2	2	2	1	2	2	2	2	
palmitoyl	1	1	1	1	1	2		1	2		1											1	1								2
benzoyl	3	3	2	2		1		1		1				1							1										
hexanoyl	4	2			1	1	1	1									1														
anthraniloyl	1	1	1	1		1		1		1				1							1										
butanoyl	5	1					1	1																							
propanoyl	1	1			1	1											1														
propionyl		2					1																								
malonyl			1	1																											
coumaryl			1	1																											
caffeoyl			1	1																											
2-acetyl		2																													
2-propionyl		2																													
octanoyl	2																														

Table 1-1: Alcohol/-CoA substrate combinations trialled in previous studies.

Number of individual studies investigating each alcohol/-CoA combination experimentally indicated in boxes. Cells are coloured according to frequency, where ■ is greatest and ■ is least. Data obtained as a literature survey using Web of Science and SciFinder.

AATs can be broadly divided into three families: wax synthases (WS), esterase-like, and BAHD-like [99]. The esterase-like family bears some resemblance to esterases, but in this case the ΔG for ester formation (-7.5 kcal mol^{-1}) is lower than that for hydrolysis (-5 kcal mol^{-1}), thereby providing enough energy upon release of the free CoA-SH to form ester products at an ambient temperature and under aqueous conditions [46]. A comparison of AAT and esterase activity in *Neurospora* sp. demonstrated that, in normal conditions, AAT was responsible for 48.5 ppm ester, while esterases only produced 1.05 ppm [97]. In this study, the activity of a *Neurospora* sp. AAT was compared against an esterase from the same organism. In *Neurospora* sp. ATCC 46892 ethyl hexanoate is the most abundant ester, contributing to both flavour and fragrance. The AAT and esterase were obtained from cell free extracts and assayed in a solution containing 659 mM ethanol, and 73 mM of either hexanoyl-CoA or hexanoic acid for 1 hour. Gas chromatography analysis of the ester products demonstrated that the AAT synthesised roughly 50x the concentration of ethyl hexanoate produced by the esterase. AAT also did not produce acetate esters

when provided with ethanol and acetyl-CoA, while esterase preferentially synthesised ethyl acetate as opposed to MC or LC esters [111].

Evidently, several possible routes are available for methacrylate ester biosynthesis. Metabolic engineering of HADH for ester production has not yet been reported and would require high concentrations of toxic aldehydes to accumulate. BVMOs depend upon high microbial supplies of ketones, a relatively uncommon metabolite. Esterases and lipases have been used extensively for transesterification or reverse-esterase activity in almost non-aqueous conditions, this is not conducive to a large-scale microbial fermentation. In contrast, AATs already dominate metabolic engineering efforts for ester biosynthesis, and use substrates that are readily produced by many organisms. Previous engineering efforts have shown some success with a range of AAT produced esters, with yields between 0.5-2 gL^{-1} of acetate esters produced by Atf1 in *E. coli* in combination with overproduction of alcohols ethanol, isopropanol, isobutanol and 1-butanol [43, 46]. MCUK selected AATs as the class of enzymes to investigate catalysis of the last step in BMA synthesis.

1.5 Designing a metabolic pathway to biosynthesize methacrylate esters

With an ester selected and BMA-resistance demonstrated in *E. coli*, the remaining question is by what route can we synthesise BMA directly? During MAA biosynthesis, early-stage valine synthesis was upregulated in *E. coli*, increasing 2-KIV production (Figure 1-6A). 2-KIV then feeds directly into a heterologous high flux pathway to M-CoA, which is converted to MAA by an acyl-CoA thioesterase [19, 112]. A benefit of using AAT to catalyse the final step of BMA biosynthesis is that we can take advantage of this pre-existing pathway to produce M-CoA, which becomes a substrate for AAT. Similar approaches have previously been used to achieve ester synthesis. In the case of isobutyl and isoamyl acetate synthesis, one study overproduced 2-KIV in a Δ*poxB* Δ*ldh* knockout strain, while also expressing AAT, and achieved a titre of 36 g L^{-1} isobutyl acetate after 72 hours in a 1.3 L reactor [113].

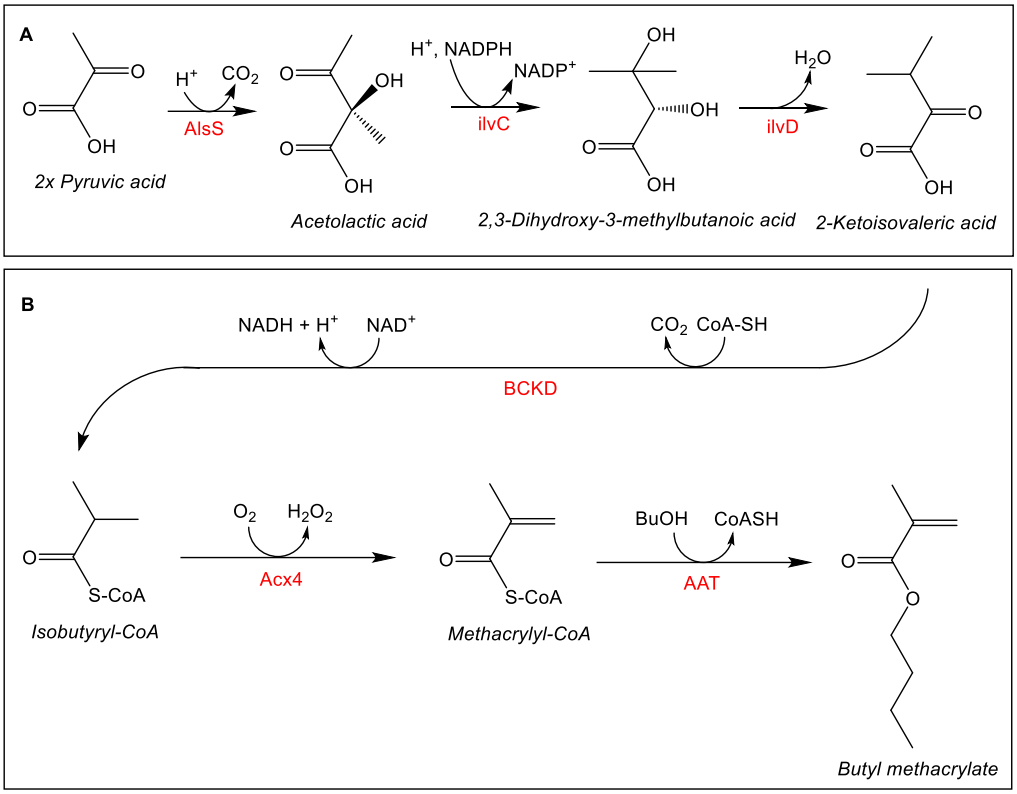


Figure 1-6: Microbial synthesis of butyl methacrylate.

A: Stage of BMA synthesis overexpressing the endogenous production of 2-ketoisovalerate, and B: Introduction of dehydrogenase, oxidase, and a transferase to form two non-native substrates and BMA. AlsS = acetolactate synthase (*Bacillus subtilis*) EC 2.2.1.6, *ilvC* = acetohydroxyacid isomeroreductase (*Escherichia coli*) EC 1.1.1.86, *ilvD* = dihydroxyacid dehydratase (*E. coli*) EC 4.2.1.9, BCKD = branched-chain ketoacid dehydrogenase (*Pseudomonas aeruginosa*) EC 1.2.4.4, ACX4 = acyl-CoA oxidase (*Arabidopsis thaliana*) EC 1.3.3.6, and AAT = alcohol acyltransferase m4 (*Malus pumila*) (MCUK).

1.5.1 Up-regulation of 2-ketoisovalerate and formation of IB-CoA

Overproduction of 2-KIV has been demonstrated many times previously. Carbon flux from pyruvate to 2-KIV forms a key stage in valine synthesis. Amino acid synthetic pathways are highly active in living organisms, with 2 million tons of glutamate, lysine and threonine produced *per annum* by industrial fermentation [114]. As a result, plenty of data is available concerning achieving high flux to the intermediate 2-ketoacids. In the pathway overexpressed for 2-KIV upregulation during BMA biosynthesis, acetolactate synthase (AlsS) from *B. subtilis* catalyses the conversion of 2 molecules of pyruvate to acetolactic acid, as with *E. coli* IlvIH [115]. AlsS has a higher pyruvate specificity than IlvIH [37]. Acetolactic acid is reduced and dehydrated to 2-KIV by

overexpressed IlvC and IlvD, respectively, both of which are endogenous to *E. coli* [115]. This approach has been used to achieve isobutyrate titres of 90 g L^{-1} in *E. coli*. This demonstrates that high flux from pyruvate to 2-KIV is possible on an industrially relevant scale [115].

Commonly, overproduced isobutyrate has been used for downstream isobutanol biosynthesis. This has been done by introducing genes *kivD* and *yqhD*, a 2-ketoacid decarboxylase and an alcohol dehydrogenase, respectively. KivD transforms 2-KIV first to isobutyraldehyde and then YqhD converts isobutyraldehyde to isobutanol [113]. To produce esters, acetyl-CoA is usually the source of acyl-CoA in any metabolic engineering efforts; however, for methacrylate ester production, AAT requires a branched chain acyl-CoA to provide the acrylate moiety of the final product.

Branched-chain acyl-CoAs occur as intermediates higher alcohol biosynthesis and during branched-chain amino acid degradation. *E. coli* does not normally produce SC acyl-CoA intermediates, and their presence can inhibit valuable metabolite formation [116]. However, other organisms such as *Streptomyces avermitilis* and *P. aeruginosa* use branched chain ketoacid dehydrogenases (BCKD)s to convert α-ketoisovalerate, α-ketomethylvalerate, or α-ketoisocaproate to isobutyryl-CoA as part of the valine, isoleucine or leucine degradation II pathways, respectively [116, 117]. The reaction mechanism is similar to that carried out by pyruvate dehydrogenase to form acetyl-CoA [43]. BCKD was previously expressed in *E. coli* to produce isovaleryl-CoA, 3-methylvaleryl-CoA, and isobutyryl-CoA. Similarly, BCKD from *S. avermitilis* was expressed in *E. coli*, in combination with chloramphenicol acetyltransferase, AlsS, IlvC and IlvD to produce 80.77 nmol g^{-1} wet cell weight isovaleryl-CoA [116]. During methacrylate ester biosynthesis, 2-KIV acts as a substrate for a heterologous *P. aeruginosa* BCKD (Figure 1-6B). This catalyses an irreversible reaction generating isobutyryl-CoA (IB-CoA) [54].

1.5.2 Oxidation of IB-CoA to M-CoA catalysed by an oxygen-dependent acyl-CoA oxidase

Isobutyryl-CoA can be converted to methacrylyl-CoA (M-CoA) by either an acyl-CoA oxidase (ACX), or acyl-CoA dehydrogenase (ACAD) [118]. This is the rate determining step of valine degradation in Pseudomonads and *B. subtilis* [112, 119]. Acyl-CoA thioesters are important intermediates in central metabolism, predominantly in fatty acid ß oxidation and branched chain amino acid catabolism [120]. During fatty acid oxidation, fatty acyl-CoAs of varying chain lengths are broken down to acetyl-CoA, feeding primarily into the citric acid cycle. The rate limiting step of this process is α,ß-dehydrogenation of the acyl-CoA thioester substrate into a trans-Δ_2-enoyl-CoA [119], which can be catalysed by ACAD in prokaryotes and mitochondria, or by ACX in peroxisomes [119, 121]. ACADs and ACXs are both members of the same flavoprotein superfamily but differ in structure and mechanism of $FADH_2$ re-oxidation (Figure 1-7).

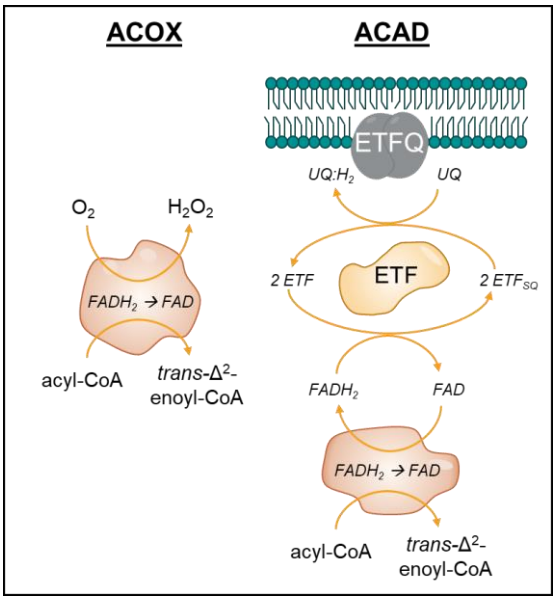


Figure 1-7: FAD re-oxidation in acyl-CoA oxidases and acyl-CoA dehydrogenases.

Mechanisms of action of acyl-CoA oxidases (ACX, left) and acyl-CoA dehydrogenases (ACD, right). ETF = electrotransferring flavoprotein, UQ = ubiquinol, UQ:H_2 = ubiquinone, SQ = semiquinol, ETFQ = membrane-bound electron transferring flavoprotein.

ACX enzymes are homodimers, usually 75 kDa in size, with one non-covalently bound FAD per subunit [121, 122]. ACX binds molecular O_2 while the substrate is present,

and directly oxidises $FADH_2$ to produce H_2O_2 [121]. This simple mechanism of FAD regeneration means that ACX enzymes are normally higher flux than ACADs. ACXs have been identified in many plants and yeast peroxisomes, but never in Gram negative bacteria to date. ACX isoforms differ in chain length specificity for acyl-CoA substrates, and are responsible for the bulk of ß-oxidation in plants [123].

The substrate specificity of ACX isoforms overlap slightly, with ACX activity in *A. thaliana* being performed by 5 different enzymes: ACX1, ACX2, ACX3, ACX4 and ACX5 [124]. This allows plants to metabolize all acyl-CoAs from short- to long- chain in their peroxisomes [125]. In contrast, mammals possess no short chain ACX. Therefore, they can only perform ß-oxidation with acyl-CoA chains longer than C8 in their peroxisomes, after which acyl-CoAs must be transported to the mitochondria for further shortening by ACADs [125, 126]. Thus, native ACX activity on IB-CoA as a substrate has only been demonstrated in plant ACX enzymes, and only two to date. The first ACX identified with IB-CoA activity was in *A. thaliana*, where ACX4 (AtACX4) was purified and assayed using butyryl-, hexanoyl- and IB-CoA. AtACX4 bears less structural similarity to ACX1-3 and ACX5 from *A. thaliana* than it does to mitochondrial ACADs (30%), and is also smaller than the other ACXs, at 50 kDa instead of 75 kDa [118].

The highest recorded activity of AtACX4 was at 25 units mg^{-1} for hexanoyl-CoA, but an activity of 2.5 units mg^{-1} was also measured for IB-CoA [125]. Since this initial study, only one other ACX has been reported with experimentally verified activity using IB-CoA, from *Vigna radiata* [127]. In this instance, ^{14}C-labelled IB-CoA was used to measure the evolution of $^{14}CO_2$ from a coupled dehydrogenase reaction. IB-CoA was formed at a rate of 35 pmol H_2O_2 s^{-1} mg^{-1}, which was only 2% of the maximal activity observed with a palmitoyl-CoA substrate [127]. *In vitro* enzyme assays and work carried out by Andrew Yiakoumetti (AY) prior to this project also demonstrated the non-native activity of AtACX4 from *A. thaliana* using IB-CoA [19, 125].

During α,ß-dehydrogenation by ACADs, a hydride ion is transferred to the N5 atom of FAD, which is buried between the α and ß subunits of the enzyme [122]. After a reaction has taken place, the reduced FAD transfers an electron to another protein, electron transferring flavoprotein (ETF), which in turn is oxidised by the membrane-bound ETF Q oxidoreductase (ETFQ). ETFQ uses ubiquinone as the terminal electron acceptor, and produces ubiquinol, thus linking ACAD FAD oxidation into the respiratory chain (Figure 1-7) [119]. Unlike oxidases, several IB-CoA specific ACADs have been identified in Gram negative bacteria. In fact, ACADs can be split into nine classes based upon their role in the cell and substrate specificity (Table 1-2). Most ACADs are homotetramers, roughly 43 kDa in size, with one FAD bound *per* 2 monomer subunits [121]. Very long chain ACADs (VLCAD1/2) are the exception [121]. Short chain (SCAD), medium chain (MCAD), long chain (LCAD) and VLCAD are specific to linear chain acyl-CoA substrates because of their role in breaking down a range of fatty acid chains towards acetyl-CoA [128]. Of the branched-chain amino acid degradation ACADs, IB-CoA dehydrogenase (IBD) uses IB-CoA as its native substrate during valine degradation [129]. Several studies have also demonstrated activity of isovaleryl-CoA dehydrogenase (IVD) with IB-CoA, and it has been reported that there is a degree of overlap of substrates between ACADs [122].

Name	Pathway	Native substrate(s)
SCAD	ß-oxidation	C_4-C_6
MCAD	ß-oxidation	C_4-C_{16}
LCAD	ß-oxidation	C_{10}-C_{18}
VLCAD1	ß-oxidation	>C_{16}
ACAD9/VLCAD2	ß-oxidation	>C_{16}
IVD	Leucine degradation	Isovaleryl-CoA
SBCAD	Isoleucine degradation	(S)-2-
IBD	Valine degradation	Isobutyryl-CoA
GD	Lysine and tryptophan	Glutaryl-CoA

Table 1-2: Summary of classes of acyl-CoA dehydrogenases.

Despite the challenge posed by co-expressing ACAD, ETF and ETFQ in *E. coli*, ACAD conversion of IB-CoA to M-CoA is a reaction natively carried out in several

identified bacteria. This includes during valine metabolism in *Aspergillus nidulans*, *Pseudomonas putida* and *Streptomyces* sp. [112, 129, 130], as well as during other metabolic processes such as glyoxylate regeneration in *Methylobacterium extorquens* [131]. IB-CoA is a natural substrate for IBD and is readily utilised by some IVDs. This presents the possibility that higher flux through from IB-CoA to M-CoA may be achieved by replacing ACX4 with a dehydrogenase.

The complexity of FAD regeneration in ACADs has meant that despite individual expression of components of the cofactor regeneration system in heterologous organisms, co-expression of ETF subunits A and B, ETFQ and ACAD heterologously has not yet been achieved. Due to the intricacy of identifying and implementing such a system, ACX enzymes were chosen to convert IB-CoA to M-CoA for methacrylate ester production. AtACX4 was selected for this purpose and cloned into *E. coli* by AY in 2014 [19].

1.5.3 Alcohol acyltransferases

AATs are responsible for most ester production in plants, yeast, fungi, and some bacteria. Fruit AATs, such as those found in apple, banana, melon and especially strawberry have been widely investigated [132, 133]. Whilst the most well studied bacterial AAT is AftA, a wax synthase from *A. baylyi* [43, 134]. Although AATs are known to transfer the acyl moiety from acyl-CoA's to various alcohols, relatively little is known about their precise mechanism of action, and importantly what factors determine acyl-CoA and alcohol specificity across different AATs [132]. AATs generally have a broad specificity for both substrates, and many studies struggle to reach a consensus on measured specificity for individual AATs [43, 135]. However, many AATs preferentially utilise acetyl-CoA, perhaps due to its relative abundance in both prokaryotic and eukaryotic cells. This makes selecting an AAT specifically to produce methacrylate esters a challenging task.

Of the previous work done to express AATs heterologously, the majority were subsequently purified for sequencing and attempted structural analysis. Microbial AATs, in particular Atf1, have usually been cloned into microbial hosts to produce acetate esters [99]. However, some studies have looked at co-producing alcohol, acetyl-CoA and heterologously expressed AATs *in vivo*. A screen of Atf1, Atf2, LuxE (*Cardamine brewerii*), BPBT (*C. brewerii*) and strawberry AAT (SAAT, *Fragaria x ananassa*) investigated co-production of isobutanol and isoamyl acetate. Researchers achieved a production titre of 2.14 g L^{-1} isoamyl acetate using Atf1 [99], with minimal formation through LuxE, BPBT and SAAT. Atf1 demonstrated similarly high yields in another study using a 120 hour batch culture, which produced 19.7 g L^{-1} [136].

Although most bioprocess-focussed research has been carried out using microbial AATs, the majority of genetically characterised AAT sequences are from plants. This is likely because the volatile esters that contribute to the aroma and taste of many fruits are often significant for food and fragrance industries. Banana, apple, papaya, and melon AATs have been extensively screened for activity on various alcohol substrates, and a small number of acyl-CoAs [135, 137, 138]. Of particular note are SAAT and wild strawberry AAT (VAAT, *Fragaria vesca*). SAAT and VAAT have both demonstrated some flexibility to acyl-CoA substrates and have been screened against microbial AATs [135]. Another study compared SAAT and *Malus domestica* AAT1 (MdAAT) expressed in *Clostridium acetobutylicum* for formation of butyl butyrate, using intracellular butanol. The enzymes formed >30 mg L^{-1} and >45 mg L^{-1} butyl butyrate, respectively [139].

MCUK tested several AATs to identify enzymes with activity on M-CoA and butanol, for use in BMA biosynthesis. They investigated the yield of methacrylate esters from the flesh of several fruits (Table 1-3). The highest BMA yield (13%) was obtained from banana AAT (BanAAT). Yields between 1% - 1.5% were recorded for strawberry and wild strawberry AATs (SAAT and VAAT), MpAAT, papaya AAT (PAAT) and avocado AAT (AvAAT). After purification, MpAAT, SAAT and VAAT all showed activity *in vitro*

with butanol and M-CoA. The highest activity, 481.5 mU g-1, was observed with MpAAT, at the expense of a lack of selectivity for BMA over BA (1.41 U g^{-1}). No data was reported for AvAAT and PAAT, and the BanAAT purified for assays was inactive, and no wax synthase enzymes were assayed (MCUK, unpublished). MdAAT was identified as the best candidate for BMA biosynthesis. A mutant AAT, AATm4, was subsequently produced using active site targeted mutagenesis with a substrate selectivity of 212:68 acetate: methacrylate. More recently MCUK produced another notable MdAAT variant, AATm10j, which demonstrates comparable acetyl-CoA and M-CoA concentrations (MCUK, unpublished).

Source	BMA from pulp (%)	AAT	BA (mU g^{-1})	BMA (mU g^{-1})
Banana	13	BanAAT	0	0
Strawberry	0.4	VAAT	1770	22.8
		SAAT	4990	10.2
Kiwi	0.5			
Lemon	0			
Apple	1	MpAAT	1410	22.8
Tomato	0			
Melon	0.6			
Pear	1			
Lime	0			
Papaya	1.5			
Avocado	1.5			
Grape	0			
Blueberry	0.4			
Yeast		Atf1	4240	0
		Atf2	6750	0

Table 1-3: Screening fruit pulp for BMA synthesis activity

indicates instances where no data was collected. indicates the AAT selected for eventual creation of AATm4.BA = butyl acetate, AAT = alcohol acyltransferase. Data obtained by MCUK.

1.6 Early implementation of BMA synthesis in WT and BMA resistant *E. coli*

AATm4 catalysed formation of BMA was implemented in *E. coli* BW25113 (AY, unpublished), using the pathway shown in Figure 1-6. BMA biosynthesis can be introduced into *E. coli* using two plasmids developed by MCUK, AY and an industrial collaborator, Ingenza (Edinburgh, UK). The first plasmid is pKIV, which overexpresses

alsS, ilvC, and *ilvD*. This upregulates host production of 2-KIV. pKIV also carries a *katE* gene. KatE is a catalase which may reduce H_2O_2 inhibition on cells in the event of high AtACX4 activity [140]. A second plasmid, pBAD-MMA050, was created by MCUK. This carries heterologous genes *bkdA1, bkdA2, bkdB* and *ipdV*, to express BCKD. pBAD-MMA050 also encodes *aatm4* and *atacx4* genes. Strains of *E. coli* expressing both pKIV and pBAD-MMA050 can synthesise BMA, or other methacrylate esters depending on the alcohol provided.

Previous data using this system for BMA production varies widely in both wild type (WT) and BMA-resistant *E. coli*, using a series of mutants generated by RNM. 0.15 mmol L^{-1} BMA was produced from WT *E. coli* after 24 h, whilst in **RNM-18** (the fastest growing RNM mutant) over 1.40 mmol L^{-1} was synthesized in the same time frame. One *soxR* mutant from ZD/LM failed to produce more than 0.05 mmol L^{-1} at 4 h, and completely halted BMA production by 24 h [54]. These early results suggested that a lack of endogenous BMA tolerance may hinder product yields during BMA biosynthesis. Although BMA production improved in **RNM-18** compared to from WT *E. coli*, BMA titres remained low enough to raise the additional question of whether flux to BMA was sufficient even when toxicity was reduced. Both sets of mutants were evolved only to resist exogenous BMA, further complicating the issue. This is customary practice in tolerance engineering, with the justification that most xenobiotic chemicals exhibit microbial toxicity through membrane interaction [141].

The question therefore remained: Does poor product-tolerance or low pathway flux limit BMA formation in *E. coli*? Increased BMA titres from some resistant mutants notwithstanding, BMA concentrations were still far lower than the ideal goal of 2 g L^{-1} h^{-1} required for profitable commercial fermentation [54]. It also remained to be seen whether poor BMA titres arose from other causes, such as the high metabolic burden of BMA synthesis [142], inefficient flux through the heterologously expressed enzymes in the pathway, or how health of the product-resistant *E. coli* may be impaired compared to WT *E. coli*.

1.7 Aims and objectives

The overarching aim of this project was to identify and begin to find solutions to bottlenecks in BMA biosynthesis in *E. coli*, answering the question of whether BMA toxicity drives a significant loss in product formation under the existing production platform. Preliminary data gathered by Ingenza, as outlined in this literature review, had indicated that BMA titres were several fold too low to be considered industrially viable. Complimentary work carried out by RNM, ZD, & LM had both demonstrated the toxicity of BMA to *E. coli*, as well as identifying mutations which might confer product-resistance.

Therefore my first objective was to assess BMA production from the 'best' form of the existing expression system, using pKIV and pBAD-MMA050 to produce BMA from *E. coli* BW25113. This was then compared against the productivity of the same two-plasmid system from four of the **RNM-*n*** series of BMA-resistant mutants in order to determine to what extent BMA toxicity may reduce production titres.

In the second phase of the project, I investigated the cause of the low, sub-toxic BMA productivity identified in the first objective. I did this by constructing a series of analogous *E. coli* strains which isolated carbon flux through individual stages of BMA production, from 2-KIV to BMA, as well as to a high yield alternative product, 3-hydroxyisobutyrate. Once I had isolated a flux-limiting step, I tested my hypothesis for its cause by developing an *in vitro* assay for AtACX4.

Once a bottleneck had been identified, my primary research aim became to investigate a way in which the information learnt from my bottleneck identification experiments could be applied to the purpose of increasing carbon flux to BMA. Therefore my next objective was to find a way to assess the selected enzyme libraries for IB-CoA or M-CoA activity. I addressed this by adapting a BMA-sensitive fluorescent screen, originally developed by Ingenza and MCUK, for use with an *in vitro* imaging system spectrum.

The following objective was to identify enzymes to assess using this novel screening method: I did this by targeting both AtACX4 and AATm4 with a bioinformatic search for alternative enzymes. The search was based on existing literature, particularly the challenging features of AATs as outlined in this literature review, as well as phylogeny, and work carried out by MCUK in their process of selecting apple AAT for the original BMA pathway.

The final objective in the project was to use the *in vitro* imaging system screen to investigate enzymes in the ACX and AAT libraries, as well as to design and implement a way in which to assemble the libraries combinatorially into a series of BMA production variant strains.

Chapter 2: Materials

2.1 Preparation of common stock solutions, buffers, and reagents

2.1.1 Isopropyl-β-D-thiogalactopyranoside (IPTG)

IPTG stocks were prepared to a concentration of 0.5 M by dissolving 5 g IPTG (Melford) in 42 mL deionised water (dH_2O). This was filter sterilised with a 0.2 µm filter under aseptic conditions and stored at -20°C in 1.5 mL microcentrifuge tubes.

2.1.2 Flavin adenine dinucleotide (FAD)

FAD 100 µM stocks were prepared by dissolving 4.15 mg FAD disodium salt hydrate (Sigma) in 50 mL dH_2O. This was filter sterilised with a 0.2 µm filter under aseptic conditions and stored at -20°C in 1.5 mL microcentrifuge tubes for up to 1 week.

2.1.3 Carbenicillin

100 mg mL^{-1} carbenicillin stocks were prepared by dissolving 5g carbenicillin disodium salt (Melford) in 44.8 mL dH_2O. Stocks were 0.2 µM filter sterilised into 1.5 mL microcentrifuge tubes and stored at -20°C.

2.1.4 Chloramphenicol

34 mg mL^{-1} chloramphenicol stocks were prepared by dissolving 1 g chloramphenicol (Sigma) in 29.4 mL dH_2O. Stocks were 0.2 µM filter sterilised into 1.5 mL microcentrifuge tubes and stored at -20°C.

2.1.5 Phosphate buffer

For 1 L 250 mM phosphate buffer at pH 7, 13.57 g sodium phosphate dibasic (Sigma) and 24.18 g sodium phosphate monobasic dihydrate (Sigma) were dissolved in 160 mL dH_2O. The pH was adjusted to 7-8.5 (as required) using concentrated HCl or NaOH. This was subsequently made up to 200 mL with more dH_2O and autoclave sterilised.

2.1.6 Diaryltetrazole (mDTZ) probe

To prepare a 300 mM stock solution, 5.43 g mDTZ (MCUK) was dissolved in 500 µL dimethyl sulfoxide (DMSO) and stored for up to 3 months at 4°C in either a LightSafe™ or foil-wrapped 1.5 mL microcentrifuge tube.

2.1.7 Glucose and glycerol solutions

Glucose stock solutions were prepared by weighing out 25 g solid glucose (Merck) and dissolving it in dH_2O, to a final volume of 100 mL, and a concentration of 25% (w/v). This solution was then autoclave sterilised before use. Glycerol stocks were also prepared to a final 25% (w/v) concentration by pipetting 25 g glycerol (Merck) into a beaker, before making this solution up to 100 mL with dH_2O and autoclave sterilising.

2.2 Media

2.2.1 Lysogeny Broth (LB) media

LB media was prepared to a concentration of 25 g L^{-1} using LB broth (Fisher Scientific) [143]. This was supplemented with 1% (w/v) glucose, or 1% (w/v) glycerol, from 20% (w/v) stock solutions where indicated. Antibiotics carbenicillin, chloramphenicol, spectinomycin, or tetracycline were added where specified to final concentrations of 50 µg mL^{-1}, 34 µg mL^{-1}, 34 µg mL^{-1}, and 12 µg mL^{-1}, respectively. Volumes of LB greater than 50 mL were autoclave sterilised and allowed to cool before the addition of carbon source and/or antibiotic under aseptic conditions. Volumes of 50 mL or smaller were filter sterilised using a 0.2 µm syringe filter and additional supplements added immediately.

2.2.2 Super Optimal broth with Catabolite repression (SOC) media

To prepare 1 L SOC [144, 145], 2.5 mL sterile 1 M KCl (74 g L^{-1}) (Sigma), 20 g tryptone (Sigma) and 5 g yeast extract (Sigma) were added to 500 mL dH_2O and adjusted to pH 7 using concentrated NaOH (Sigma). This solution was made up to

970 mL with more dH_2O and autoclave sterilised. Stock solutions of 1 M $MgCl_2$ (203.3 g L^{-1}) (Sigma) and 1 M glucose (180 g L^{-1}) were also prepared with dH_2O and autoclave sterilised. To make the final SOC solution, 10 mL 1 M $MgCl_2$ and 20 mL 1 M glucose were added to the 970 mL sterile solution. 50 mL aliquots were prepared and stored at 4°C until required.

2.2.3 LUND media

1 L LUND media (Ingenza) was prepared using 200 mL 5x LUND salt solution (10 g L^{-1} $(NH_4)_2$ SO_4, 73 g L^{-1} K_2HPO_4, 18 g L^{-1} $NaH_2PO_4.2H_2O$, 2.5 g L^{-1} $(NH_4)_2$ citrate), 40 mL 25% (w/v) glucose or 25% (v/v) glycerol solution, 40 mL yeast extract (50 g L^{-1}) (only where specified), 20 mL $MgSO_4$ solution (1 M), 2 mL LUND trace element (TE) solution (0.5 g L^{-1} $CaCl_2.2H_2O$, 10.03 g L^{-1} $FeCl_3$, 0.18 g L^{-1} $ZnSO_4.7H_2O$, 0.16 g L^{-1} $CuSO_4.5H_2O$, 0.15 g L^{-1} $MnSO_4.H_2O$, 0.18 g L^{-1} $CoCl_2.6H_2O$, 22.3 g L^{-1} $Na_2EDTA.2H_2O$), antibiotic, and 736 mL dH_2O. All solutions were autoclaved before use, and preparation of LUND media carried out under aseptic conditions.

2.2.4 Biotransformation (BT) media

To prepare 1 L BT medium, 200 mL 5x BT solution (73 g L^{-1} KH_2PO_4, 18 g L^{-1} $NaH_2PO_4.2H_2O$, 10 mL 1 M $MgSO_4$, 10 mL LUND TE solution), 35.75 g sodium-2-ketoisovalerate (Sigma), 10 mL glycerol (20% (w/v)) were made up to 900 mL with dH_2O, and the pH adjusted to 7. dH_2O was added to a final volume of 1 L, which was filter sterilised by vacuum filter funnel under aseptic conditions.

2.2.5 LB agar

Agar plates were prepared with 25 g L^{-1} LB media (Fisher Scientific) and 20 g L^{-1} agar (Miller). This was sterilised in the autoclave and subsequently kept in a 50°C water bath for 1 hour, or until the solution cooled to 50-60°C. Before pouring, molten LB agar was supplemented with either carbenicillin (50 µg mL^{-1}), chloramphenicol (34 µg mL^{-1}), spectinomycin (34 µg mL^{-1}), or tetracycline (12 µg mL^{-1}), where required. For

plates containing 1% (w/v) glucose or glycerol, LB agar was returned to the 50°C water bath for 10 min after mixing. Antibiotics and alcohols were added just prior to pouring as required. Plates were poured and left to cool with lids off under a laminar flow hood. After 30 min the lids were replaced, and dried plates flipped. Plates were stored wrapped in cling film at 4°C and were pre-warmed in a 37°C New Brunswick Scientific Innova 40 incubator shaker before use.

2.2.6 Sucrose counter-selection (SC) agar

SC agar plates were prepared with 5 g L^{-1} yeast extract, 10 g L^{-1} tryptone and 20 gL^{-1} agar. This was autoclave sterilised and subsequently kept in the water bath at 50°C for 1 hour. Antibiotics were added where indicated. Plates were poured under a laminar flow hood and allowed to cool for 30 min with the lids off before the plates were flipped upside down and stored at 4°C.

2.2.7 Solid phase assay (SPA) agarose

SPA agarose plates contain 0.04 M Triton X-100 (Sigma), 0.06 M 1-butanol, 0.04 M sodium-2-ketoisovalerate, 0.055 M phosphate buffer, 1.5% agarose (Sigma) and 48 mM DTZ probe. To prepare 200 mL SPA buffer 4.39 mL 1-butanol, 4.42 g sodium 2-ketoisovalerate and 175.61 mL or 250 mM phosphate buffer, pH 7 were mixed on a stirrer plate. Stirring was reduced to 150-200 rpm and 20 mL Triton X-100™ slowly added to prevent bubble formation. This solution was placed in a 60°C water bath until the agarose solution reached the correct temperature. 12 g agarose was dissolved in 600 mL dH_2O and heated until molten. This solution was then left on a heated stirrer plate at 300-500 rpm until the temperature reached 60-65°C. At this point 200 mL SPA buffer was added slowly. 128 µL DTZ (300 mM) was added and allowed to mix just before pouring. 20 mL SPA agarose was poured per plate using a 50 mL serological pipette. Plates were allowed to cool on the bench, before storage wrapped in foil at 4°C.

2.3 Strains

2.3.1 *E. coli* stocks

2.3.1.1 BW25113 cells

In most instances, *E. coli* BW25113 was the strain in use. WT *E. coli* BW25113 cryostocks were prepared from a stock provided by Luca Rossoni. BW25113 mutants were generated by the following co-workers: *E. coli* BW25113 Δ*ldhA* Δ*infA* :: KanR (Andrew Yiakoumetti, AY). *E. coli* BW25113 Δ*tesB* Δ*yciA* (Ingenza), *E. coli* BW25113 RNM-2 (Russel Menchavez, RM), *E. coli* BW25113 RNM-3 (RM), *E. coli* BW25113 RNM-18 (RM), and *E. coli* BW25113 RNM-19 (RM).

2.3.1.2 DH5α, BL21(DE3) and NEB5α cells

E. coli DH5α stocks were prepared from a WT strain supplied by AY. *E. coli* BL21(DE3) stocks were prepared from a WT strain provided by Maria-Letizia Cassioli. NEB5α High Efficiency cells were obtained from NEB (C2987H).

2.4 Table of plasmids

Name	Parent plasmid	Genotype (excl. selection)	Selection
pSTV28	Standard plasmid	*lacZα*	CamR
pJET1.2	Standard plasmid	*eco47I/T7*	AmpR
pBAD24	Standard plasmid	*araC*	AmpR
pET20b (+)	Standard plasmid	*n/a*	AmpR
pCL1	Standard plasmid	*n/a*	AmpR
pMMA050 (Ingenza)	pTRC-99a	*bkdA1, bkdA2, bkdB, ipdV, acx4, aatm4*	AmpR
pMMA050-MBP-ACX4 (Ingenza)	pTRC-99a	*bkdA1, bkdA2, bkdB, ipdV, mbp-acx4, aatm4*	AmpR
pMMA126 (Ingenza)	pMMA050	*bkdA1, bkdA2, bkdB, ipdV, acx4, ech, hch*	AmpR
pBAD-MMA050	pBAD24, pMMA050	*bkdA1, bkdA2, bkdB, ipdV, acx4, aatm4*	AmpR
pHIBA (CT/AY)	pBAD-MMA050$_{corrected}$(mACX4)	*acx4, ech, hch*	AmpR
pHIBA-0.3 (AY)	pHIBA	*bkdA1, bkdA2, bkdB, ipdV, acx4, ech, hch*	AmpR
pHIBA-3 (AY)	pHIBA	*bkdA1, bkdA2, bkdB, ipdV, acx4, ech, hch, infA*	AmpR, InfA
pSC101_SoxR_AcrR (LM)	pSC101	*soxR acrR*	TetR

Name	Parent plasmid	Genotype (excl. selection)	Selection
pKIV (AY)	pSTV28	*alsS, ilvC, ilvD*	CamR
$pKIV_{ara}$ (Ingenza)	pKIV	*alsS, ilvC, ilvD, katE*	SpecR
pKIVrha (Ingenza)	pKIV	*alsS, ilvC, ilvD, katE*	SpecR
$pBAD\text{-}MMA050_{corrected}$	pBAD-MMA050	*bkdA1, bkdA2, bkdB, ipdV, acx4, aatm4*	AmpR
pBAD-MMA050(mACX4)	pBAD-MMA050, pMMA050-MBP-ACX4	*bkdA1, bkdA2, bkdB, ipdV, mbp-acx4, aatm4*	AmpR
$pBAD\text{-}MMA050_{corrected}$(mACX4)	$pBAD\text{-}MMA050_{corrected}$, pMMA050-MBP-ACX4	*aatm4, mbp-acx4, bkdA1, bkdA2, bkdB, ipdV*	AmpR
pCAN-1	pHIBA-3, pMMA126, pSC101	*acx4, ech, hch*	TetR
pCAN-2	pHIBA-0.3, pJET_AAT-ACX4	*bkdA1, bkdA2, bkdB, ipdV, acx4, aatm4*	AmpR
pCAN-3	pHIBA-0.3, pJET_AAT	*bkdA1, bkdA2, bkdB, ipdV, aatm4*	AmpR
pCAN-4	pHIBA-3, pJET_AAT-ACX4	*bkdA1, bkdA2, bkdB, ipdV, acx4, aatm4, infA*	AmpR, InfA
pCAN-5	pHIBA-3, pJET_AAT	*bkdA1, bkdA2, bkdB, ipdV, aatm4, infA*	AmpR, InfA
pJET_AAT	pJET1.2, $pBAD\text{-}MMA050_{corrected}$(mACX4)	*aatm4*	AmpR
pJET_AAT_ACX4	PJET1.2, $pBAD\text{-}MMA050_{corrected}$(mACX4)	*acx4, aatm4*	AmpR
pOX-2_ACX4	pET20b(+),$pBAD\text{-}MMA050_{corrected}$	*acx4*	AmpR
pOX-2_MBP-ACX4	pET20b(+),$pBAD\text{-}MMA050_{corrected}$(mACX4)	*mbp-acx4*	AmpR
pOX-2_ACX4-HIS	pET20b(+),$pBAD\text{-}MMA050_{corrected}$	*acx4*	AmpR
pOX-2_MBP-ACX4-HIS	pET20b(+),$pBAD\text{-}MMA050_{corrected}$(mACX4)	*mbp-acx4*	AmpR
pGGV4	pHIBA-3, pCAN-4, pEX18-Gm	*bkdA1, bkdA2, bkdB, ipdV, sacB*	AmpR, InfA
pOX-3_ACX#1 (Twist)	pET-21(+)	*so-acx1*	AmpR
pOX-3_ACX#2 (Twist)	pET-21(+)	*zm-acx1*	AmpR
pOX-3_ACX#3 (Twist)	pET-21(+)	*pa-acx4*	AmpR
pOX-3_ACX#4 (Twist)	pET-21(+)	*so-sovf-b*	AmpR
pOX-3_ACX#5 (Twist)	pET-21(+)	*zm-acx1(2)*	AmpR
pOX-3_ACX#6 (Twist)	pET-21(+)	*as-acx4*	AmpR
pOX-3_ACX#7 (Twist)	pET-21(+)	*so-acx4*	AmpR
pOX-3_ACX#8 (Twist)	pET-21(+)	*zm-acx4*	AmpR
pOX-3_ACX#9 (Twist)	pET-21(+)	*cm-acx4*	AmpR
pOX-3_ACX#10 (Twist)	pET-21(+)	*vr-acx1(2)*	AmpR
pOX-3_ACX#11 (Twist)	pET-21(+)	*yl-acx3(1)*	AmpR
pOX-3_ACX#12 (Twist)	pET-21(+)	*vr-acx3(2)*	AmpR
pOX-3_ACX#13 (Twist)	pET-21(+)	*ct-acx4*	AmpR
pOX-3_ACX#14 (Twist)	pET-21(+)	*vr-acx4 x2*	AmpR
pOX-3_ACX#15 (Twist)	pET-21(+)	*pa-acx4*	AmpR

Name	Parent plasmid	Genotype (excl. selection)	Selection
pOX-3_ACX#16 (Twist)	pET-21(+)	*vr-acx4 x1*	AmpR
pOX-3_ACX#17 (Twist)	pET-21(+)	*ah-acx4*	AmpR
pOX-3_ACX#18 (Twist)	pET-21(+)	*zm-acx1(1)*	AmpR
pOX-3_ACX#19 (Twist)	pET-21(+)	*to-acx4*	AmpR
pOX-3_ACX#20 (Twist)	pET-21(+)	*gn-acx4*	AmpR
pCAN-6_AAT#1 (GU)	pUC57-Kan	*at1*	KanR
pCAN-6_AAT#2 (GU)	pUC57-Kan	*at9*	KanR
pCAN-6_AAT#3 (GU)	pUC57-Kan	*caat1*	KanR
pCAN-6_AAT#4 (GU)	pUC57-Kan	*cer2*	KanR
pCAN-6_AAT#5 (GU)	pUC57-Kan	*chat*	KanR
pCAN-6_AAT#6 (GU)	pUC57-Kan	*cm-aat1*	KanR
pCAN-6_AAT#7 (GU)	pUC57-Kan	*cm-aat4*	KanR
pCAN-6_AAT#8 (GU)	pUC57-Kan	*dat*	KanR
pCAN-6_AAT#9 (GU)	pUC57-Kan	*dbbt*	KanR
pCAN-6_AAT#10 (GU)	pUC57-Kan	*dk-aat1*	KanR
pCAN-6_AAT#11 (GU)	pUC57-Kan	*eeb1*	KanR
pCAN-6_AAT#12 (GU)	pUC57-Kan	*eht1*	KanR
pCAN-6_AAT#13 (GU)	pUC57-Kan	*ej-aat1*	KanR
pCAN-6_AAT#14 (GU)	pUC57-Kan	*fc-aat1*	KanR
pCAN-6_AAT#15 (GU)	pUC57-Kan	*la-aat1*	KanR
pCAN-6_AAT#16 (GU)	pUC57-Kan	*ma-aat*	KanR
pCAN-6_AAT#17 (GU)	pUC57-Kan	*man-aat*	KanR
pCAN-6_AAT#18 (GU)	pUC57-Kan	*mb-aat*	KanR
pCAN-6_AAT#19 (GU)	pUC57-Kan	*pc-aat*	KanR
pCAN-6_AAT#20 (GU)	pUC57-Kan	*ph-bebt1*	KanR
pCAN-6_AAT#21 (GU)	pUC57-Kan	*pu-aat*	KanR
pCAN-6_AAT#22 (GU)	pUC57-Kan	*rh-aat1*	KanR
pCAN-6_AAT#23 (GU)	pUC57-Kan	*saat*	KanR
pCAN-6_AAT#24 (GU)	pUC57-Kan	*vp-aat1*	KanR
pCAN-6_AAT#25 (GU)	pUC57-Kan	*vs*	KanR
pCAN-6_AAT#26 (GU)	pUC57-Kan	*ws*	KanR
pCAN-7_ACX#1	pGGV4, pCAN-4, pOX-3_ACX#1	*bkdA1, bkdA2, bkdB, ipdV, aatm4, so-acx1*	AmpR, InfA
pCAN-7_ACX#2	pGGV4, pCAN-4, pOX-3_ACX#2	*bkdA1, bkdA2, bkdB, ipdV, aatm4, zm-acx1*	AmpR, InfA

Name	Parent plasmid	Genotype (excl. selection)	Selection
pCAN-7_ACX#3	pGGV4, pCAN-4, pOX-3_ACX#3	*bkdA1, bkdA2, bkdB, ipdV, aatm4, pa-acx4*	AmpR, InfA
pCAN-7_ACX#4	pGGV4, pCAN-4, pOX-3_ACX#4	*bkdA1, bkdA2, bkdB, ipdV, aatm4, so-sovf-b*	AmpR, InfA
pCAN-7_ACX#5	pGGV4, pCAN-4, pOX-3_ACX#5	*bkdA1, bkdA2, bkdB, ipdV, aatm4, zm-acx1(2)*	AmpR, InfA
pCAN-7_ACX#6	pGGV4, pCAN-4, pOX-3_ACX#6	*bkdA1, bkdA2, bkdB, ipdV, aatm4, as-acx4*	AmpR, InfA
pCAN-7_ACX#7	pGGV4, pCAN-4, pOX-3_ACX#7	*bkdA1, bkdA2, bkdB, ipdV, aatm4, so-acx4*	AmpR, InfA
pCAN-7_ACX#9	pGGV4, pCAN-4, pOX-3_ACX#9	*bkdA1, bkdA2, bkdB, ipdV, aatm4, cm-acx4*	AmpR, InfA
pCAN-7_ACX#10	pGGV4, pCAN-4, pOX-3_ACX#10	*bkdA1, bkdA2, bkdB, ipdV, aatm4, vr-acx1(2)*	AmpR, InfA
pCAN-7_ACX#13	pGGV4, pCAN-4, pOX-3_ACX#13	*bkdA1, bkdA2, bkdB, ipdV, aatm4, ct-acx4*	AmpR, InfA
pCAN-7_ACX#14	pGGV4, pCAN-4, pOX-3_ACX#14	*bkdA1, bkdA2, bkdB, ipdV, aatm4, vr-acx4 x2*	AmpR, InfA
pCAN-7_ACX#15	pGGV4, pCAN-4, pOX-3_ACX#15	*bkdA1, bkdA2, bkdB, ipdV, aatm4, pa-acx4*	AmpR, InfA
pCAN-7_ACX#16	pGGV4, pCAN-4, pOX-3_ACX#16	*bkdA1, bkdA2, bkdB, ipdV, aatm4, vr-acx4 x1*	AmpR, InfA
pCAN-7_ACX#17	pGGV4, pCAN-4, pOX-3_ACX#17	*bkdA1, bkdA2, bkdB, ipdV, aatm4, ah-acx4*	AmpR, InfA
pCAN-7_ACX#18	pGGV4, pCAN-4, pOX-3_ACX#18	*bkdA1, bkdA2, bkdB, ipdV, aatm4, zm-acx1(1)*	AmpR, InfA
pCAN-7_ACX#19	pGGV4, pCAN-4, pOX-3_ACX#19	*bkdA1, bkdA2, bkdB, ipdV, aatm4, to-acx4*	AmpR, InfA
pCAN-7_ACX#20	pGGV4, pCAN-4, pOX-3_ACX#20	*bkdA1, bkdA2, bkdB, ipdV, aatm4, gn-acx4*	AmpR, InfA
pCAN-7_AtACX4-ATL	pGGV4, pCAN-4, pCAN-6 (all)	*bkdA1, bkdA2, bkdB, ipdV, at-acx4,* ***aat****	AmpR, InfA
pCAN-7_ACX#3-ATL	pGGV4, pOX-3_ACX#3, pCAN-6 (all)	*bkdA1, bkdA2, bkdB, ipdV, pa-acx4,* ***aat****	AmpR, InfA
pCAN-7_ACX#4-ATL	pGGV4, pOX-3_ACX#4, pCAN-6 (all)	*bkdA1, bkdA2, bkdB, ipdV, so-sovf-b,* ***aat****	AmpR, InfA
pCAN-7_ACX#7-ATL	pGGV4, pOX-3_ACX#7, pCAN-6 (all)	*bkdA1, bkdA2, bkdB, ipdV, so-acx4,* ***aat****	AmpR, InfA
pCAN-7_ACX#14-ATL	pGGV4, pOX-3_ACX#14, pCAN-6 (all)	*bkdA1, bkdA2, bkdB, ipdV, vr-acx4-x2,* ***aat****	AmpR, InfA
pCAN-7_ACX#15-ATL	pGGV4, pOX-3_ACX#15, pCAN-6 (all)	*bkdA1, bkdA2, bkdB, ipdV, pa-acx4,* ***aat****	AmpR, InfA
pCAN-7_ACX#17-ATL	pGGV4, pOX-3_ACX#17, pCAN-6 (all)	*bkdA1, bkdA2, bkdB, ipdV, ah-acx4,* ***aat****	AmpR, InfA
pCAN-7_ACX#19-ATL	pGGV4, pOX-3_ACX#19, pCAN-6 (all)	*bkdA1, bkdA2, bkdB, ipdV, to-acx4,* ***aat****	AmpR, InfA
pCAN-7_ACX#20-ATL	pGGV4, pOX-3_ACX#20, pCAN-6 (all)	*bkdA1, bkdA2, bkdB, ipdV, gn-acx4,* ***aat****	AmpR, InfA
pCAN-7_ACX#7-AAT#16	pGGV4, pOX-3_ACX#7, pCAN-6_AAT#16	*bkdA1, bkdA2, bkdB, ipdV, so-acx4, ma-aat*	AmpR, InfA

Table 2-1: Table of plasmids used during this work.

Plasmids in shaded blue rows are commercially obtainable plasmids that were available in the lab. Plasmids labelled “Ingenza” were provided by Ingenza, plasmids labelled “JS”,” AY,” or ”LM” were prepared by Jennifer Spencer, Andrew Yiakoumetti or Laura Martins, respectively. Plasmids labelled “Twist” were synthesised by Twist Biosciences, and plasmids labelled “GU” were synthesised by Gene Universal. * = Genes shown in **bold** vary as part of AAT variant library.

2.5 Table of primers

Name	Sequence (5' to 3')
(CimN)22.1.SEQ3	GGAGACCCCACACTACCATC
ALT_Fwd_1	ATTTGCCTGGCGGCAGTGAATTCAAG
ALT_Fwd_2	GGCAGTGAATTCAAGGAGATATACCATGAC
ALT_Rev_1	AAAACAGCCAAGCTTGCATGGTCGAC
ALT_Rev_2	GCCAAGCTTGCATGGTCG
AS-CHK-F1	GCACCATCCACGCCCATCC
AS-CHK-F2	CTAGAGTCGACCTGCAGGTATTATCTCGAGC
AS-CHK-R	CCATTCGCCAATCCGGATATAGTTCCTCC
GG-AAT_F	CTCGAGTAAACTAGTTTTG
GG-AAT_R	TCTCAGTGGTGTATCTTAAG
GG-ACX_F	GGGAATTCTTAACTTTAAGAAGG
GG-ACX_R	CAAAACTAGTTTACTCGAG
GG-FIX-F	GTTTCCCGTTGAATATGGC
GG-FIX-R	TTTGTGATGCTCGTCAGG
GU-VECTOR2-F	CTCCCGCGTCTCACTAGTTTTGTTTAACTTTAAGAAGG
GU-VECTOR2-R	TTTGCTGGCCTTTTGCTCCGTCTCCTTAAG
GU-VECTOR-F	AAACCTCTGACACATGCAGCTC
GU-VECTOR-R	AACGCGGCCTTTTTACGG
LIB-COLPCR-CEN_F	GAAGGAGATATACCATGTCC
LIBRARY_COLPCR_F	CGAGCGTCAATTGTCTGATTCG
LIBRARY_COLPCR_R	CCAATCCGGATATAGTTCCTCC
OG-AAT_F	GAGTAAACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGAAAAGCTTTTCTGTAC
OG-AAT_F-NEW	GAGTAACGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGAAAAGCTTTTCTGTAC
OG-AAT_R	TTTTGCTCCGTCTCCTTAAGTTACTGGCTGG
OG-ACX4-F	GGATCCCGTCTCGAATTCTTTGTTTAACTTTAAGAAGG
OG-ACX4-R	TTATATGGCGTCTCACTAGTTTACTCGAGTTACAGGCGAGAACG
pCAN-1_pCON F	CCTGAATGGAATGGCCGTCGTTTTACAACG
pCAN-1_pCON R	TTATTTCTAGAGGGGCCTTCGCTAGCGTCG
pCAN-1_pMMA126 F	CTAGCGAAGGCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGG
pCAN-1_pMMA126 R	CAGAGTCCCTGCAGGATCAGAGCC
pCAN-1_pSC101 F	TTTTTATTGGTGAGAATCGTAAAGCCCGTGACC

Name	Sequence (5’ to 3’)
pCAN-1_pSC101 R	ACGGCCATTCCATTCAGGTCGAGGTGG
pCAN-1_SP F	TCCTGCAGGGACTCTGGGGTTCGAGAGCTCGCTTGGACTCCTGTTGATAGATC CAGTAATGACCTCAGAACTCC
pCAN-1_SP R	TTTACGATTCTCACCAATAAAAAACGCCCGGCGGCAACCGAGCGTTCTGAACA AATCCAGATGGAGTTCTGAGGTCATTACTGG
pCAN-2_AAT-FRAG-F	GAATTCTTTGTTTAACTTTAAGAAGGAGATATACC
pCAN-2_AAT-FRAG-R	CTTAAGTTACTGGCTGGTGCTAC
pCAN-2_ACX4-HiFi-F	CTTTAAGAAGGAGATATACCATGGCTGTCCTGTCAAGC
pCAN-2_ACX4-HiFi-R	AGTACAGAAAAGCTTTTCATGAAGTTCCTCACTGGATCAG
pCAN-3_AAT-HiFi-F	ATGAAAAGCTTTTCTGTACTC
pCAN-3_AAT-HiFi-R	GGTATATCTCCTTCTTAAAGTTAAAC
pGGV4_gBLOCK_fwd	AGGAAGCGGAAGAGGGCCTGATGCGGTATTTTCTCCTTACGC
pGGV4_gBLOCK_rev	ACCTAGGACTGAGCTAGCTGTCAAGCGGCCGCGCATTTCGCGGGATCGAGATC
pGGV4_pEX18-GM_fwd	TAGGTATAATGCTAGCACCCGTTTTTTTGGGAATTCGAGACGCACATATACCT GCCGTTC
pGGV4_pEX18-GM_rev	GGTGTATCTTAAGGAGACGAGAGTGCACCATAATCG
pGGV4_pHIBA3_fwd	GCACTCTCGTCTCCTTAAGATACACCACTGAGATCCGGCTG
pGGV4_pHIBA3_rev	TACCGCATCAGGCCCTCTTCCGCTTCCTCGCTCAC
pGGV4_SEQ1	AGTTCCTGAGTTCGATTCGTCC
pGGV4_SEQ2	AAACGCACGGCTGAGTTAGC
pGGV4_SEQ3	GAAAGGAGGAACTATATCCGG
pGGV4_SEQ4	ATCACTCAGGGTCAATGC
pOX-1_gBLOCK_Fwd	TGGCGGCAGTGAATTCAAGG
pOX-1_gBLOCK_Rev	AAAACAGCCAAGCTTGCATGG
pOX-1_SEQ1	CTAGCAGGAGGAATTCAAGG
pOX-1_SEQ2	CGTTCGTAGTTCCTCTTCG
pOX-1_SEQ3	GTACAATGTACGTGGGAAGG
pOX-1_SEQ4	AATCTTCTCTCATCCGCC
pOX-1_SEQ5	AACTGCGTAGTTTGGACG
pOX-1_SEQ6	CTCGCATCCTTATGAAGACG
pOX-1_SEQ7	CCTTCTGGTAGATGGTAAGG
pOX-1_SEQ8	AGTGGATATAACGGCTTCG
pOX-2_ACX4_HIS_R	TATTTAATATACTCGAGGCCCAGGCGAGAACGGGTAG
pOX-2_ACX4_NdeI_F	GCGATTAGACTACATATGGCTGTCCTGTCAAGC
pOX-2_MBP_NdeI_F	GCGAGCGGACGCCATATGAAAATCGAAGAAGGTAAACTGG
pOX-2_MBP+ACX4_XhoI_R	GTAATACTCGAGTTACAGGCGAGAACGGGTAG

Name	Sequence (5’ to 3’)
pSacB_Fwd	TAATGCTAGCACCCGTTTTTTTGGGAATGAATTCGAGACGCACATATACC
pSacB_Rev	ATCTCAGTGGTGTATCTTAAGGAGACGTTATTTGTTAAC
SACB_REV	GACTCTCGTTTGGATTGC
SEQ-1	CGCAACTCTCTACTGTTTCTCCATAC
SEQ-10	GATGATGAACCTCTCCTC
SEQ-11	GAAGCGCAACCACGCATCC
SEQ-18	ATCACCCTGAGCAGCCTCG
SEQ-19	GGTGATCTCCTCCACCGAG
SEQ-2	GAACGTGAACTGCTGTTCGCACG
SEQ-20	GCCTGGTGGCGAGCTTC
SEQ-21	ACATGTTCTTTCCTGCGTTATCC
SEQ-21	ACATGTTCTTTCCTGCGTTATCC
SEQ-24	CGCTCATGAGACAATAACCC
SEQ-25	CGAACTACTTACTCTAGCTTCCC
SEQ-26	GAGTTTTCGTTCCACTGAGC
SEQ-3	TGGAGCGTTTCCAACAGGAGC
SEQ-34	CTCATGTTTGACAGCTTATCATCG
SEQ-4	CAACGGTCAGAAACGTTGGATTG
SEQ-5	CCCGTTCTCGCCTGTAAACTAG
SEQ-57	ATCCTCCCGACAACACAG
SEQ-58	CATATGCACAGATGAAAACGGTG
SEQ-59	CTCTGGGGTTCGAGAGC
SEQ-6	GATCAAGGGCGACACGCGC
SEQ-60	AGCGGAATTTACAGAGGGTC
SEQ-61	TGAGTTTAAAAGGCTTAACCAATGG
SEQ-62 (pCAN-1_SEQ1)	TCTTTGTGAGTCCATGCG
SEQ-63 (pCAN-1_SEQ2)	TGTTCCGCCTGGAATACG
SEQ-64 (pCAN-1_SEQ3)	AGCGAACTGAATGTCACG
SEQ-65 (pCAN-1_SEQ4)	TGATGACGAGAACTGTGG
SEQ-66 (pCAN-1_SEQ5)	GGATCTGAGGTTCTTATGGC
SEQ-67 (pCAN-1_SEQ6)	AAAAGGTGCGGGTTACTGG
SEQ-68 (pCAN-1_SEQ7)	GTATGGAAGTTCCTCGCTCC
SEQ-69 (pCAN-1_SEQ8)	GCATCGTAGACGCTGAGG

Name	Sequence (5’ to 3’)
SEQ-7	GCGCTCGGCGATGGACATC
SEQ-8	GCCGAGATCATCGACCTGC
SEQ-9	TGCAGTTCGTGCAGGGCAG

Table 2-2: Table of primers used during this work.

All primers were synthesised by Integrated DNA Technologies (IDT). Primers were obtained freeze-dried in tubes containing 25 nmol DNA (<60 bp) or 100 nmol DNA (=>60 bp). Freeze-dried primers from IDT were re-suspended to 100 µM concentration using DI water, and aliquoted to 10 µM dilutions for molecular biology work.

Chapter 3: Methods

3.1 Molecular Biology

3.1.1 Preparation of purified plasmid DNA

Plasmid DNA was isolated from a 10 mL overnight cell culture using either a QIAprep Spin Miniprep Kit (Qiagen 27104) or a Monarch® Plasmid Miniprep Kit (NEBT1010S). DNA was eluted into sterile molecular biology water (mb-H_2O), and DNA concentrations quantified using a BioDrop µLITE. Where required, DNA precipitation was used to increase the concentration of purified DNA, using the sodium acetate method as outlined in the GenElute™ kit protocol.

3.1.2 Restriction digestion

Enzymes for restriction digestion were used according to product specifications to generate linear DNA fragments from purified plasmid DNA, although the incubation time was extended to 4+ hours to ensure complete digestion of fragments. Enzymes used were FastDigest *Not*I (ThermoFisher FD0595), FastDigest *Xba*I (ThermoFisher FD0685), FastDigest *Mau*BI (ThermoFisher FD2084), FastDigest *Eco*RI (ThermoFisher FD0275), *Spe*I (ThermoFisher ER1252), FastDigest *Sal*I (ThermoFisher FD0644), FastDigest *Nde*I (ThermoFisher FD0583), FastDigest *Bsp*TI (*Afl*II) (ThermoFisher FD0834), *Esp*3I (NEBR0734S), *Not*I (NEB R3189) and *Sbf*I-HF (NEB R3642S). DNA fragments from restriction digests were isolated by gel extraction.

3.1.3 Golden Gate assembly

Golden Gate assembly was used to generate the pCAN-7 AAT and ACX plasmid libraries for solid phase BMA screening. 20 fmol of pGGV4 was mixed with 40 fmol each of the AAT and ACX donor plasmids, as well as 1 µL T4 DNA Ligase (NEB M0202S), 1 µL *Esp*3I (NEB R0734S), and 2 µL 10x T4 DNA Ligase Buffer (NEB M0202S). The solution was made up to 20 µL using sterile mb-H_2O. The solution was

mixed and centrifuged briefly, conditions for the assembly were as follows: 30 digestion-ligation cycles of 37°C for 5 min followed by 16°C for 5 min, followed by an extra digest step at 37°C for 5 min, then heat inactivation at 65°C for 20 min. 2 µL of the assembly mix was immediately transformed into chemically competent *E. coli* DH5α or NEB5α cells.

3.1.4 Gel electrophoresis and gel extraction

Enzyme digests and PCR products were run on 1% agarose DNA gel in a 1x TAE buffer containing either 10 mg mL^{-1} ethidium bromide (Merck E1510) or SYBR™ Safe DNA gel stain (Invitrogen S33102). Colony PCR products were dyed with JumpStart REDTaq ReadyMix (Sigma P1107) so 10 µL product was loaded directly into wells. Amplified PCR products for gel extraction were in a 50 µL volume, to which 10 µL 6x TriTrack DNA Loading Dye (ThermoFisher R1161) was mixed and the 60 µL sample loaded into wells. Digests with ThermoFisher enzymes contained FastDigest® Green Buffer (ThermoFisher B72) and so were loaded onto gels directly. 10 µL 6x Purple Gel Loading Dye (NEB 7025) was added to digests using restriction enzymes from NEB. DNA fragments for gel extraction were isolated using either Zymoclean™ Gel DNA Recovery kit (Nordic Biolabs D4001), GenElute™ Gel Extraction kit (Sigma NA1111) or Monarch® DNA Gel Extraction kit (NEB T1020S).

3.1.5 Ligation

Ligation was used to construct pHIBA, pGG, pCAN, and pOX plasmid derivatives. T4 DNA Ligase (NEB M0202S) or Hi-T4™ DNA Ligase (NEB M2622S) and 10x T4 ligase buffer was used in all instances, with a vector:insert molar ratio of either 3:1 or 6:1. Ligation reaction were incubated at room temperature overnight, before heat inactivation for 10 min at 65°C, and transformation into electrocompetent cells.

3.1.6 PCR

Colony PCR was used to detect the presence of all plasmids in *E. coli* BW25113, BW25113 Δ*ldhA* Δ*infA*::*kanR*, DH5α and BL21(DE3). Single colonies were picked from plates using a pipette tip. Colonies were re-plated onto fresh LB agar plates by touching the pipette tip to the surface and were added to 20 µL dH_2O. 2 µL of the colony in water was transferred to a PCR mix containing JumpStart REDTaq ReadyMix (Sigma P1107) (10 µL), and 1.5 µL of each 10 µM primer. This solution was made up to 20 µL with dH_2O. PCR conditions were as follows: 94°C for 5 min, followed by 30 cycles of melting at 94°C for 30 seconds, primer annealing at 57°C for between 1.5 - 3.0 min, and elongation at 72°C for 3 min. At the end of 30 cycles, elongation at 72°C continued for an additional 5 min, after which samples were held at 4°C. Four-part HiFi DNA assembly was used to construct pHIBA and pCAN-1, as well as the Golden Gate vector strain. 1 µL template DNA, 0.5 µL Q5 DNA polymerase (NEB), 10 µL Q5 buffer (NEB), 1 µL dNTPs, and 2.5 µL each of 10 µM primers were mixed and made up to a 20 µL solution with dH_2O. PCR conditions were as follows: 98°C for 30 seconds, followed by 5 cycles of 98°C for 10 seconds, annealing of overlaps for 10 seconds, 72°C for 3 min. Then followed 25 cycles of melting at 98°C for 15 seconds, annealing at 65-72°C for 20 seconds - 3.0 min, and elongation at 72°C for 2 min. 0.1 pmols of each amplified DNA fragment total 0.4 pmol) were mixed and made up to 10 µL. To this 10 µL of NEBuilder® HiFi DNA Assembly Master Mix (NEB) was added. The HiFi mix was then incubated for 1 hour at 50°C, after which 2 µL of chilled product was transformed into chemically competent *E. coli* NEB5α

3.1.7 Transformations

3.1.7.1 Electrocompetent cell preparation

500 µL of overnight cell culture was added to 50 mL LB in a 250 mL baffled flask and incubated at 37°C and 250 rpm until OD_{600} reached 0.6-0.8. The culture was then divided into two 50 mL Falcon tubes and immediately placed on ice for 1 hour. After an hour, the tubes were centrifuged for 10 min at 5,000 rpm and 4°C, then

immediately returned to ice. The supernatant was gently poured off and 10 mL sterile dH_2O added to re-suspend pellet. After re-suspension, the volume was made up to 50 mL with more sterile dH_2O. Cells were centrifuged and re-suspended this way three times. After the last wash step, cells were re-suspended in 10 mL 20% (v/v) sterile glycerol and centrifuged as before. Most of the remaining supernatant was poured off, and the cells resuspended in the remainder. Electrocompetent cells were then stored as 40 µL aliquots in 1.5 mL microcentrifuge tubes at in a -80°C DW-86L828J freezer (Haier Biomedical).

3.1.7.2 Chemically competent cell preparation

E. coli DH5α cells were plated onto LB agar and incubated overnight at 37°C. A single colony was picked and inoculated into 10 mL LB in a 50 mL Falcon tube. After 12-15 hours the cells were sub-cultured into 500 mL of LB in a 2 L baffled flask. At an OD_{600} 0.35-0.40 the culture was immediately placed on ice and chilled for 30 min, swirling occasionally. The culture was then split into five in 50 mL Falcon tubes, and centrifuged at 4000 rpm and 4°C for 10 min. The supernatant was decanted and the cell pellet resuspended in a total of 100 mL cold $MgCl_2$. Cells were harvested again at 3000 rpm and 4°C for 10 min, and the supernatant removed. This time the pellets were resuspended in 200 mL $CaCl_2$. This suspension was incubated on ice for 20 min, before cells were harvested once again at 3000 rpm and 4°C for 15 min. After removing supernatant, the pellets were then resuspended and combined together into a total of 50 mL 85 mM $CaCl_2$, 15% glycerol. This cell suspension was centrifuged for a final times at 2100 rpm and 4°C for 15 min, after which the supernatant was removed. The pellet was then resuspended in 2 mL 85 mM $CaCl_2$, 15% glycerol. The competent cells were split into 50 µL aliquots in chilled 1.5 mL microcentrifuge tubes which were either used immediately or stored at -80°C.

3.1.7.3 Electroporation

Transformation by electroporation was used to introduce pBAD-MMA050, pCAN-2, pCAN-3, pCAN pSC101_SoxR_AcrR, pHIBA-0.3, pHIBA-3, and all pOX-1, pOX-2, and pOX-3 plasmids into *E. coli* BW25113. Electroporation was also used to transform pCAN-4, pCAN-5, pSC101_SoxR_AcrR and pHIBA-3 into *E. coli* BW25113 Δ*ldhA* Δ*infA*::*kanR*. All transformations used electrocompetent cells prepared in the lab according to the Datsenko method [146], using 10% glycerol, which were stored at -80°C. Electroporation was carried out using 0.2 cm electroporation cuvettes. Aliquots of electrocompetent cells were defrosted slowly on ice and 1-3 µL (~50 ng) plasmid DNA added. Cells were electroporated using a GenePulser XcellTM (BioRad) at 2.5 kV, 25 µF and 200 Ω, before immediate addition of 950 µL SOC medium. 50 µL of cells were plated on an LB agar plate with corresponding antibiotic selection. The remaining cells were centrifuged at 1300 rpm for 1 min. Supernatant was discarded and cells re-suspended in the remaining liquid. All of this remaining solution was plated on a second LB agar or SCA selection plate.

3.1.7.4 Heat shock

Heat shock transformation was used to introduce HiFi assembled pCAN-1 and pHIBA, and pJET intermediate plasmids for constructing pCAN-2-5 and all pOX-1 plasmids into *E. coli* DH5α. Chemically competent *E. coli* DH5α cells were obtained from Jennifer Spencer (JS) or high efficiency NEB5α competent cells (NEB). Aliquots of chemically competent cells were fully defrosted on ice and 2-5 µL DNA added. Cells were incubated on ice for 5 min, then at 42°C for 20 seconds, then on ice for a further 2 min. 950 mL SOC medium was added, and cells were then incubated at 37°C and 250 rpm for 1 hour. 50 µL of cells were plated on an LB agar plate with corresponding antibiotic selection. The remaining cells were centrifuged at 13000 rpm for 1 min. Supernatant was discarded and cells re-suspended in the remaining liquid. All of this remaining solution was plated on a second LB agar plate.

3.2 Cell growth, culture, and maintenance

3.2.1 Cryostock preparation and storage

Initial wild-type *E. coli* BW25113 cells were obtained from cryostocks made by LR from cells provided by Ingenza. *E. coli* BW25113 Δ*ldhA* Δ*infA*::*kanR* knockouts were prepared by AY. *E. coli* BL21(DE3) and DH5α cells were prepared as chemically competent cells by MC and JS, respectively. Subsequent cryostocks were prepared by picking a single colony using a sterile plastic loop and swirling it in 10 mL LB, plus glucose and carbenicillin, chloramphenicol, tetracycline or spectinomycin as appropriate, in a sterile 50 mL Falcon tube. The cultures were incubated at 37°C and 250 rpm until an OD_{600} between 0.6-1.0 was reached. 150 µL cells were then added to 850 µL sterile 80% glycerol solution in a cryovial under a laminar flow hood. This solution was vortexed briefly and immediately stored at -80°C. To inoculate from these cryostocks, tubes were removed from -80°C and stored on ice until a small amount of the cells defrosted. 30 µL of cryostock was then added to LB media in a baffled flask and was incubated overnight at 37°C and 250 rpm.

3.3 Protein expression and purification

3.3.1 Protein expression

Pre-cultures were grown in 10-12 mL LB in a 50 mL Falcon tube at 37°C and 250 rpm overnight. This was sub-cultured to OD_{600} 0.1 in 500 mL LB in a 2 L baffled flask, which was returned to the same conditions for incubation. At OD_{600} 0.6-1.0 protein expression was induced with 1 mM IPTG, and the temperature lowered to 30°C. Cells were then left to grow for 12-15 hours. After this, the 500 mL culture was divided into two 250 mL vessels for the Avanti G-26 XP Centrifuge (Beckman Coulter).

3.3.2 Cell harvesting

Cells were harvested at 8,000 rpm for 15 min at 4°C using a JLA 10,500 rotor. The supernatant was then removed, and cells re-suspended in the remaining LB, before

pooling cells together in a 50 mL Falcon tube. These were then harvested once more using a 5810R centrifuge (Eppendorf) at 8,000 rpm for 15 min at 4°C. After harvesting, cell pellets were either flash frozen and stored at -80°C, or immediately lysed for purification.

3.3.3 Lysis by sonication

Throughout sonication all samples were kept on ice or at 4°C. Initially, cell pellets from protein expression samples were resuspended in 10 mL chilled binding buffer (2 mg 0.02 L^{-1} lysozyme (Sigma), 2 µL 0.02 L^{-1} benzonase (R) endonuclease (Merck), ¼ tab 0.02 L^{-1} cOmplete protease inhibitor cocktail (Roche)) and were agitated on ice for 30 min. Samples were then sonicated in the Falcon tube for 10x 30 second intervals, with 30 seconds pause, using a One-Shot Cell Disruptor (Constant System Ltd). Lysed cells were centrifuged at 23,000 rpm in a Beckmann Floor Centrifuge at 4°C for 20 min. The lysate was then collected and kept on ice for loading onto a pre-prepared Ni-NTA column.

3.3.4 Preparation of crude SDS-PAGE samples

Samples for SDS-PAGE for non-purified proteins were prepared using Bugbuster reagent. Cells were grown overnight at 37°C and 250 rpm in 10-12 mL LB in a 50 mL Falcon tube. 1-2 mL of this culture was collected in a 1.5 mL microcentrifuge tube and centrifuged at 13,000 rpm for 2 min. Supernatant was removed and a volume of Bugbuster master mix (100 µL Bugbuster, 20 µL protease inhibitor, 1 µL R lysozyme, 0.1 µL benzonase nuclease, 879 µL PBS buffer) was added to resuspend the pellet. Volumes of master mix to add were calculated using the following equation:

$$\frac{\left(\frac{OD600}{0.2}\right) \times 45}{2} = volume\ of\ Bugbuster$$

Samples were incubated at room temperature for 20 min at 150 rpm, before centrifuging at 13,000 rpm for 5 min. The supernatant (soluble fraction) was removed

and stored on ice or at -20°C until needed. The cell pellet (insoluble fraction) was re-suspended in an equal volume of PBS buffer and stored as with the soluble fractions. Protein content was visualised using NuPAGE SDS. 10 µL of sample was transferred to a fresh 0.2 mL PCR tube, and a 1:1 volume of NuPAGE SDS buffer added (5 µL 4X NuPAGE sample buffer, 2 µL denaturing solution, 3 µL dH_2O). Samples were boiled for 10 min, briefly centrifuged, and 8-10 µL loaded onto gel. The gel ran at 150 V for 50 min, before visualising protein bands in water for 15 min, Coomassie blue (0.1% (v/v), 50% (v/v) methanol, 10% (v/v) acetic acid) for 1 hour, and in water overnight.

3.3.5 His-trap protein purification

Binding buffer A (pH 7.5, 20 mM HEPES, free acid, 500 mM NaCl, 10 µM FAD Na_2 salt, 20 mM imidazole), elution buffer B (pH 7.5, 20 mM HEPES, free acid, 500 mM NaCl, 10 µM FAD Na_2 salt, 500 mM imidazole) and storage buffer C (pH 7.5, 20 mM HEPES, free acid, 150 mM NaCl, 10 µM FAD Na_2 salt) were prepared in dH_2O and stored at 4°C. 5 mL of lysed cells in Buffer A were loaded onto a HisTrap FF crude 5 mL crude column pre-packed with Ni-Sepharose fast-flow resin (GE Healthcare), pre-charged with $NiSO_4$ (0.1 M) following manufacturer's instructions, on an AKTA fast protein liquid chromatography system (GE Healthcare). Protein was purified by step purification at 2.5 mL min^{-1} and 0.3-0.4 mPa. All fractions containing protein were collected and the purity of each checked using SDS-PAGE.

3.3.6 Analytical SDS-PAGE

Fractions from protein purification were mixed in a 1:1 ratio with 10 µL β-mercaptoethanol (Sigma) and boiled in a thermocycler for 10 min. 15 µL of the denatured samples were loaded onto a BioRad 12% Tris-glycine SDS gel and run at 200 V for 35 min, against PageRuler™ Unstained Protein Ladder. Gels were then visualised using InstantBlue™ for 15 min, before resolving overnight in dH_2O.

3.3.7 Buffer exchange

Vivaspin 20 (10,000 MWCO) columns (Sartorius Stedim Biotech) were loaded with 20 mL storage buffer (Buffer C). This was spun for 30 min at 6,000g and 4°C or until all storage buffer passed through the column. Flow through was then discarded and protein fractions added to the column, topped up to 20 mL with Buffer C. This was centrifuged as before and topped up to 20 mL with Buffer C three times. 1-2 mL Buffer C was added to the column after the final spin and purified protein was stored in 500 µL aliquots at -20°C until required.

3.3.8 Bradford determination of purified protein concentration

To check the concentration of purified protein fractions, a Bradford assay with bovine serum albumin (BSA) concentrations (Sigma) between 0-2 mg mL^{-1} was used. Samples were prepared in a 96-well microtiter plate (Corning inc.) to a total volume of 10 µL per sample. 300 µL Bradford reagent (BioRad laboratories inc. (UK)) was added to each well, and the microtiter plate was left away from UV light for 5-10 min. Absorbance values were read using a FluOstar Optima (BMG lab tech) plate reader.

3.4 Synthetic and analytical chemistry

3.4.1 M-CoA synthesis and purification

Before starting the M-CoA synthesis reaction, 30 mg Coenzyme A trilithium salt (CAS: 85-61-0) was dissolved in 2 mL of 100 mM sodium phosphate buffer, pH 8.5 in a 10 mL glass vial kept on ice. 12.3 µL methacrylic anhydride was added to begin the reaction, and the vial vortexed immediately. The reaction was incubated on ice for 30 min, vortexing again for 10 seconds every 2 min. The reaction was then acidified to pH 3.5 with 5 M HCl. Methacrylic acid and unreacted methacrylic anhydride were removed by vortexing the reaction sample four times 1:1 with water saturated diethyl ether. The M-CoA product was purified using RP-HPLC. 75 µL of sample was injected at a time and purified on a Zorbax Eclipse C18 column with an initial flowthrough of 98% trifluoroacetic acid (TFA) (0.1% v/v) and 2% acetonitrile (ACN) (90% (v/v) ACN,

0.1% (v/v) TFA) at a rate of 1 mL min^{-1}. The concentration of ACN solution was increased to 15% over 40 min. The M-CoA peak was detected by fluorescence at 215 nm at 25 min, and flow-through collected directly from the fluorometer into an ice-cold glass vial. ACN in the purified sample was removed by rotary evaporation at 21°C and 3 kPa, after which the M-CoA solution was neutralised with dilute NaOH. Finally, water was removed from the sample by flash freezing the sample in a 50 mL round bottomed flask, and overnight freeze-drying. TFA was removed by re-dissolving the freeze-dried sample in 10 mL nuclease-free dH_2O and repeating the flash freezing-lyophilisation process, this was done again with 5 mL, and then 1 mL nuclease-free dH_2O. Purity of the final M-CoA solution was checked using mass spectrometry and NMR. NMR spectra ran at 298 K using a Bruker AV(III)400, AV400.DPX400 (400 MHz ^{1}H frequency, 100 MHz ^{13}C frequency). Chemical shifts are quoted in parts per million (ppm), referenced to the residual deuterated solvent quoted as internal standards. Coupling constant (j) are in Hz. Multiplicity of the signals is abbreviated as follows: s, singlet; d, doublet; t, triplet; q, quartet; dd, doublet of doublet; dt, doublet of triplet; m, multiplet; br, broad. MS was recorded using a Bruker MicroTOF 61 mass spectrometer using electrospray ionisation (ESI). Concentration of M-CoA was determined using the extinction coefficient and absorbance at 280 nm using a UV mini 1240 UV-Vis spectrophotometer (Shimadzu) equipped with a CPS-100 temperature controller. Purified M-CoA was stored in 50 µL aliquots at -20°C.

3.4.2 GC-MS

Gas Chromatography – Mass spectrometry (GC-MS) (Agilent 7890A and 5975C) samples were prepared from ethyl acetate (EtAc) extraction. 10 mL supernatant was vortexed with 10 mL EtAc for 1 min in glass vials with PTFE lids. Once left to settle, two 1 mL samples were taken from the organic phase for analytical duplicates in GC-MS. One GC-MS run lasted 27 min. 1 µL of each sample was injected with 10 split into the inlet, at 280°C, 9.4 psi and 24.2 mL min^{-1}. The column (Agilent 19091S-433), 30 m x 20 µm x 0.25 µm, was heated to a maximum temperature of 300°C for 5 min,

ramping from 45°C in 20°C intervals. Oven temperature held at 300°C for a further 10 min. Flow through the column was 1.197 mL min^{-1}, at a velocity of 39.78 cm sec^{-1}. Each sample cycle began and ended with four injections of pure ethanol to wash the column.

3.4.3 HPLC

High Performance Liquid Chromatography (HPLC) samples were also prepared in duplicate. Isocratic HPLC was performed using a Rezex ROA Organic Acid column H+ (8%), 300 mL x 7.8 mL, with 0.01 N sulphuric acid flowing through at a rate of 0.500 mL min^{-1} and 60 bar pressure. Samples were detected by refractometric detection (RID).

3.4.4 Oxygen electrode assays

Before starting the assay, 40 mL assay solution containing 0.8 mg mL^{-1} purified ACX, 50 mM HEPES, pH 7.5, 10 µM FAD, and between 0-500 µM M-CoA, was prepared in a chilled 50 mL Falcon tube. The Oxygraph (Hansatech) was calibrated by running it at 100% oxygen saturation for 5 min using a pump to bubble air through the solution, followed by 0% oxygen saturation introduced by adding sodium bisulphite (Thermo Scientific). The assay chamber was then washed three times with 1 mL dH_2O. To start the assay, 1 mL assay solution was added to Oxygraph, with the stirrer set to 3 and temperature at 30°C, and the oxygen concentration allowed to stabilise for 3 min. Once the oxygen concentration became stable, 10 mM IB-CoA solution was added to the electrode to a final concentration of 0-1 mM. The reaction was allowed to proceed until either the oxygen concentration had equilibrated, or until 5 min had passed, as only initial change in oxygen concentration was required for determination of reaction kinetics.

Chapter 4: BMA formation from product resistant *E. coli*

4.1 Introduction

The overarching belief in our consortium in late 2017 was that BMA toxicity significantly inhibited its own production within an *E. coli* host chassis. This hypothesis arose from the low IC_{50} value that was experimentally determined by ZD, at 0.07-0.11 g L^{-1} (0.5-0.8 mM) for *E. coli* exposed to exogenous BMA [77]. A commercially relevant 2 g L^{-1} h^{-1} or 14 mM h^{-1} productivity is required for MCUK to feasibly scale-up for an industrial bio-MMA process. An IC_{50} of 0.07-0.11 g L^{-1} would therefore reduce cell viability well before a concentration of BMA approaching commercial viability could be achieved in a bioreactor. To combat this hypothesised toxicity bottleneck, product-resistant strains of *E. coli* BW25113 were evolved and characterised in an attempt to bypass this potential BMA toxicity problem. The two major avenues of investigation were the characterisation of a SoxR AcrR mutant strain using transcriptomics by LM and the iterative generation of new BMA resistant strains in batch and chemostat cultures by RNM using adaptive lab evolution (ALE). Both sets of product-resistant candidate strains had only been briefly surveyed for BMA productivity.

12 novel strains were generated from the ALE experiments, which possessed a number of mutations when compared to wild-type (WT) *E. coli* BW25113. Most of these changes occurred within stress response and global regulatory genes, perhaps as would be expected (Table 4-1). Four strains, **RNM-2**, **RNM-3**, **RNM-18,** and **RNM-19** exhibited improved growth characteristics, in particular **RNM-18** and **RNM-19**, which grew at a similar rate to WT *E. coli* BW25113 (in the absence of BMA). **RNM-18** also produced relatively high BMA titres in biotransformations, reaching 1.50 mM BMA after 6 hours (Table 4-2). I selected these four mutants in this project to compare against a WT BMA producer strain. Therefore **RNM-2**, **RNM-3**, **RNM-18** and **RNM-19** were assessed for BMA productivity, growth, and selectivity towards BMA. Comparing these mutants against a WT *E. coli* BMA producer allowed exploration of whether

BMA toxicity was indeed preventing higher production titres, an effect which may be mitigated by introducing BMA biosynthesis into product-resistant host strain instead [90, 147].

RNM-18 was isolated from serial batch culture. It contains a mutation in *rob*, as well as mutations in the genes coding for heat shock protein DnaK [82], inner membrane protein YohJ [148], and in an intergenic region between global transcriptional regulator *lrp* and the coding region of thioredoxin reductase *trxB* [149, 150]. Although no knockouts were found in **RNM-18**, four of the changes reported were frameshift mutations, which may have a similar effect to a knockout in completely deactivating the gene. These frameshifts were in *rpoC*, *acrR*, *clsA*, and *cra*. For *rpoC*, which produces RNA polymerase (RNAP) subunit ß' [151], this mutation would put well over half the remaining amino acid sequence out of frame, most likely rendering the β' subunit inactive. The frameshift in *acrR* occurs at position 191 of 215 amino acids. Similarly, a frameshift mutation at A448 in *clsA* will affect only the remaining 38 amino acids. The frameshift in *cra* will affect a greater proportion of the translated protein, with the shift occurring at 270 out of 334 amino acids. Although we do not yet know whether any of the frameshifted proteins will remain functional, it can be inferred that some modified function may remain for those where the mutation affects only the C-terminus of the protein, as is the case for AcrR and to some extent ClsA. Although the location of the RNAP ß' subunit mutation suggests it is most likely inactive, the fact that the mutation is not lethal to *E. coli* demonstrates that the cells must still be able to translate DNA in some capacity.

RNM-2 and **RNM-3** were also isolated from batch culture. Of the mutations for which location data was provided, **RNM-2** contains 4 known missense mutations, and one large frameshift. The **RNM-2** missense mutations occurred in *ygbK*, which is putatively involved in flagellar motility [152], *rpoC*, *ilvN*, and *rob* [151]. All of these mutations occur well within the coding regions of the affected genes. Whilst the *rob* mutation R156H converts the positively charged arginine into a relatively similar amino acid, the

ilvN C14Y mutation changes the small, uncharged cysteine residue into the much larger, hydrophobic tryptophan, potentially with larger consequences for IlvN function if this occurs near the active site. Both the *rob* and *ilvN* mutations are also present in **RNM-3**. *lpxM*, a myristoyl transferase which catalyses the terminal step in lipid A synthesis [153], is subject to a frameshift mutation in **RNM-2**, and potentially putting 20% of the amino acid sequence out of frame. Similarly, **RNM-3** contains a frameshift mutation in *icd*, at amino acid 298 of 741. *Icd* codes for an isocitrate dehydrogenase, which catalyses the formation of α-ketoglutarate from isocitrate using $NADP^+$ [154].

Several of the mutations reported for **RNM-2** unfortunately were not provided along with their locations, only with gene annotations, and so it is not possible to potential effects of the alterations in *pinE, acrR, stfE* or *stfP*. Although interestingly *pinE*, *stfE,* and *stfP* are all knocked out in **RNM-3.** Indeed, many knockouts are present in the genome of **RNM-3**, some of which affect large clusters of genes. For example, 3 genes relating to fructose utilisation were deleted in **RNM-3** (*fruA, fruB* and *fruK*) [155] and 3 relating to pseudouridine synthesis (*psuG, psuK* and *psuT*). While the deletion of fructose utilisation genes may be more obscure, the *psu* genes affect the flexibility and conformation of RNA, and therefore may have wider-ranging effects on global gene expression in *E. coli* when deleted [156].*[155]*

Although somewhat unlikely, no knockouts were reported in **RNM-2** or **RNM-19** (Table 4-1B), however this is likely only because of an omission in the data provided. **RNM-19** was isolated from chemostat culture and contains the same annotations for *acrR, pinE, stfE* and *stfP* that were identified but no details given for **RNM-2** and **RNM-3**. Unique to **RNM-19** are missense mutations in *rpoC, acrB, marR* and *ompR*, all of which are genes which relate to the *E. coli* stress response [157]. The *rpoC* mutation occurs at position 1075, 332 amino acids away from the C terminal tail of the protein. This position lies between an α-helix and a β-sheet region of the secondary structure, and replaces an arginine residue with cysteine, a substitution which may interfere with the secondary protein structure in some way as Cys is responsible for disulphide

bridge formation. The *marR* V84G mutation maintains similar sidechain properties at that position to the structure prior to mutations, and the R15S mutation in *ompR* replaces positively charged arginine with the un-charged, polar serine. As previously mentioned, no knockouts were reported for this BMA resistant strain.

A: MUTATIONS						B: KNOCKOUTS		
Gene	Mutation	Strain				Gene	Strain	
		RNM-2	RNM-3	RNM-18	RNM-19		RNM-3	RNM-18
rpoC	K215fs			x		*mcrA*	Δ	
	L361R	x				*tfaE*	Δ	
	R1075C				x	*aaaE*	Δ	
acrR	n/a	x	x		x	*beeE*	Δ	
	A191fs			x		*croE*	Δ	
acrB	T379I				x	*fruA*		Δ
	V448L		x			*fruB*		Δ
rob	R156H	x	x			*fruK*		Δ
	A70T			x		*intE*	Δ	
clsA	A448fs			x		*jayE*	Δ	
cpxA	P177Q			x		*lit*	Δ	
cra	I270fs			x		*oweE*	Δ	
creA	V85V			x		*pinE*	Δ	
dnaK	V377G			x		*psuG*		Δ
icd	D298fs		x			*psuK*		Δ
ilvN	C41Y	x	x			*psuT*		Δ
lpxM	L259fs	x				*setB*		Δ
marR	V84G				x	*stfE*	Δ	
ompR	R15S				x	*stfP*	Δ	
ompT	n/a	x				*tfaP*	Δ	
ompX	n/a			x		*xisE*	Δ	
opgH	R95P			x		*ymfD*	Δ	
phoP	L11F		x			*ymfE*	Δ	
pinE	n/a	x			x	*ymfI*	Δ	
pitA	n/a					*ymfJ*	Δ	
stfE	n/a	x			x	*ymfL*	Δ	
stfP	n/a	x			x	*ymfM*	Δ	
ygbK	A294E	x				*ymfQ*	Δ	
yohJ	L109R			x		*ymfR*	Δ	

Table 4-1: Genetic variation in BMA resistant strains RNM-2**,** -3**,** -18 **and** -19**.**

A: Genes present but mutated in **RNM-2**, **RNM-3**, **RNM-18** and **RNM-19**. Blue text indicates genes which possess mutations in more than one BMA resistant strain. fs = frameshift, mutations marked in highlighted rows. For mutations labelled as "n/a" no location was provided for mutation. B: Genes knocked out in **RNM-3** and **RNM-18**. No knockouts were reported for **RNM-2** and **RNM-19**. Data obtained by RNM using strains grown in up to 20% (v/v) BMA in sequential batch and chemostat cultures (RNM, unpublished).

A biotransformation carried out by Ingenza prior to this project demonstrated that **RNM-2**, **RNM-3**, **RNM-18**, and **RNM-19** produced higher BMA titres than the other ALE mutants (Table 4-2). During this experiment, BMA-resistant cells expressing the BMA pathway using plasmid pMMA050 were harvested from an overnight culture in LB and then re-suspended to OD_{600} 50 with 0.1 M sodium phosphate buffer. The re-suspended solution was supplemented with 40 mM 2-KIV and 37 mM 1-butanol in a sealed 250 mL Schott bottle [147]. BMA synthesis was then allowed to proceed at 30°C and 250 rpm for 24 hours. Results from this biotransformation showed **RNM-18** produced 1.4 mM BMA after 24 hours. This was as compared to WT *E. coli* BW25113 pKIV pMMA050, which produced 0.15 mM BMA within the same time frame. Although BMA formation from all four mutants in Table 4-2 was improved, none of these changes were particularly dramatic. This could signify either that evolving BMA-resistant mutants to resist extracellular BMA is not sufficient, or that even when product toxicity is relieved carbon flux was still too low to further increase production.

Strain	Max biomass concentration		Growth rate		Lag time	BMA
	gL^{-1}	%	h^{-1}	%	h	mM
RNM-2	0.651 ± 0.052	45.3	0.532 ± 0.028	73.1	6	0.35
RNM-3	0.790 ± 0.076	55.0	0.566 ± 0.010	77.7	4	0.37
RNM-18	1.036 ± 0.046	72.1	0.709 ± 0.015	97.4	5	1.40
RNM-19	0.653 ± 0.023	45.5	0.707 ± 0.007	97.1	5	0.20

Table 4-2: Characteristics of BMA resistant strains RNM-2**,** -3**,** -18 **and** -19**.**

Cell growth and concentration in presence of 20% (v/v) BMA (extracellular). % values represent biomass concentration or growth rate of mutants as compared to wild-type *E. coli* with no butyl methacrylate (BMA) present. Growth data obtained by RNM (unpublished). Production data obtained by Ingenza (unpublished).

Another factor likely to contribute to low BMA titres was the potential formation of alternate esters, produced directly by the action of AATm4. It is well known that AAT enzymes have very broad substrate ranges [158]. Therefore, it was expected that AATm4 would not exclusively synthesise BMA, particularly as acetyl-CoA and IB-CoA are also present in producer cells. Indeed, the most abundant ester produced by the original apple AAT from which AATm4 was derived is butyl acetate (BA) [159]. Even

the modified AATm4 used in the BMA pathway preferentially synthesises BA *in vitro*, with an activity ratio of 1410:22.8 U mg-1 for BA: BMA (MCUK, unpublished). BA was thus expected to form a significant draw for AAT activity during BMA formation *in vivo*, particularly as acetyl-CoA is present in relatively high intracellular concentrations of 20-400 µM for the purpose of normal metabolism in *E. coli* [160]. IB-CoA is produced by BCKD during BMA biosynthesis, and may also act as a substrate for AATm4 to produce butyl isobutyrate (BIB). Off-target BA and BIB production detract from BMA formation, and make it challenging to determine the cause of low titres without also checking for additional butyl esters in analytical samples.

However, as well as the various BMA resistant mutants available to compare, there were also several different expression systems in use to collectively allow heterologous BMA formation, none of which had been compared for BMA productivity. Therefore, before any resistant strains were assessed, or by-products identified, an expression vector had to be selected from available permutations of the BMA biosynthetic pathway. As previously described, the metabolic pathway from pyruvate to BMA was expressed in *E. coli* using a two-plasmid system. The first plasmid (Table 4-3, pKIV1-5) component expressed variations on the genes *alsS, ilvC* and *ilvD* to convert pyruvate to 2-KIV. The second plasmids were methacrylate ester producing (Table 4-3, pMAE-1-4) and were generally used to express *aatm4, at-acx4, bkdA1, bkdA2, bkdB* and *ipdV* to produce BMA from 2-KIV. Five versions the first plasmid, pKIV1-5 were available and four of the pMAE plasmid. This small library of plasmid variations were assembled into WT *E. coli* BW25113 for direct comparison.

pKIV-1 contains the genes as listed above, while pKIV-2-5 express an altered version of IlvC which was modified to utilise NADH as a cofactor instead of NADPH (IlvC*). IlvC* was generated in anticipation of the anaerobic conditions under which industrial scale fermentation generally operates: In this environment, *E. coli* carries out glycolysis instead and as such only the NADH reducing equivalent will be available [161]. pKIV-3-5 are all spectinomycin, instead of chloramphenicol, resistant. pKIV-4

and pKIV-5 use arabinose and rhamnose, respectively, as inducers instead of IPTG. As with pKIV-1, pMAE-1 expresses the genes as listed above. pMAE-2 and pMAE-4 contain a correction of a C→T point mutation which occurred in the gene *bkdB*. pMAE-3 and pMAE-4 express a maltose binding protein fused version of AtACX4 (MBP-ACX4). MBP-AtACX4 was produced by Ingenza, and demonstrates favourable solubility when compared to AtACX4. None of these variations in the basic pKIV and pMAE plasmids had been examined for effects on BMA formation and cell viability at the commencement of this project. Strains expressing these plasmids were named to reflect which iteration of pKIV and pMAE are present. For example, the WT strain containing pKIV-1 and pMAE-3 is named **WT_K1.M3**.

Plasmid	Genotype	Description
pKIV-1	*alsS, ilvC, ilvD, katE, camR*	
pKIV-2	*alsS, ilvC*, ilvD, katE, camR*	From pKIV-1, IlvC uses NADH
pKIV-3	*alsS, ilvC*, ilvD, katE, specR*	From pKIV-2, *specR* replaces *camR*
pKIV-4	*alsS, ilvC*, ilvD, katE, specR*	From pKIV-3, arabinose promoter replaces *lacO*
pKIV-5	*alsS, ilvC*, ilvD, katE, specR*	From pKIV-3, rhamnose promoter replaces *lacO*
pMAE-1	*aatm4, at-acx4, bkdA1, bkdA2*T, bkdB, ipdV*	
pMAE-2	*aatm4, at-acx4, bkdA1, bkdA2*C, bkdB, ipdV*	From pMAE-1, C→T point mutation corrected
pMAE-3	*aatm4, mbp-at-acx4, bkdA1, bkdA2*T, bkdB, ipdV*	From pMAE-1, *mbp-at-acx4* in place of *at-acx4*
pMAE-4	*aatm4, mbp-at-acx4, bkdA1, bkdA2*C, bkdB, ipdV*	From pMAE-2, *mbp-at-acx4* in place of *at-acx4*

Table 4-3: List of plasmids for two-plasmid production of BMA

AlsS = acetolactate synthase (*Bacillus subtilis*) EC 2.2.1.6, ilvC = acetohydroxyacid isomeroreductase (*Escherichia coli*) EC 1.1.1.86, ilvD = dihydroxyacid dehydratase (*E. coli*) EC 4.2.1.9, BCKD = branched-chain ketoacid dehydrogenase (*Pseudomonas aeruginosa*) EC 1.2.4.4, AtACX4 = acyl-CoA oxidase (*Arabidopsis thaliana*) EC 1.3.3.6, and AATm4 = alcohol acyltransferase m4 (*Malus pumila*) (MCUK).

At the beginning of this project the question was therefore: Is carbon flux through the BMA pathway sufficient to supply inhibitory concentrations of BMA within *E. coli*? As an extension to this, is the current bottleneck in BMA formation toxicity or carbon flux efficiency? To investigate this, the existing BMA formation pathway from both WT *E. coli* BW25113 and from the mutants produced by RNM were compared. In doing so, I searched for formation of off-target butyl esters. Ialso briefly investigated the synthesis

of a less volatile ester product, hexyl methacrylate (HMA). The IC_{50} for HMA in *E. coli* has to date not been reported, but hydrophobicity of HMA's longer hexyl- chain compared to BMA may increase interactions between the ester and *E. coli* cell membrane. This hypothesis presumes similar carbon flux to HMA as is found with BMA but may allow further comparison of carbon flux through the methacrylate ester pathway, whilst reducing product loss due to volatility.

In this chapter the **RNM-2**, **RNM-3**, **RNM-18**, and **RNM-19** strains with no plasmid were created by RNM. pKIV-1, pMAE-1, and pMAE-2 were received from LR. pKIV-2, pKIV-3, pKIV-4 and pKIV-5 were received from Ingenza. I generated the plasmids pMAE-3 and pMAE-4. I assembled the pKIV/pMAE plasmid library in WT *E. coli*. LR carried out the flask tests to determine BMA productivity from these strains. All subsequent strain assemblies and flask tests were carried out by me.

4.2 Results

4.2.1 Assembly of a library of BMA biosynthesis variants

Because several iterations of pKIV (Figure 4-1A) and pMAE (Figure 4-1B) were available to facilitate BMA formation from *E. coli* it was necessary to generate a library encompassing all combinations of both plasmids into the WT strain. BMA production was assayed from each strain, including the empty pBAD plasmid and pSTV28 to act as negative controls for pMAE and pKIV, respectively (Table 4-3). pBAD and pSTV28 were used as they are the plasmids from which both pMAE and pKIV were originally adapted (Table 2-1). I assembled and sequenced pMAE-3 and pMAE-4, then transformed all pathway combinations into *E. coli* BW25113 using electroporation to generate a library of 30 strains.

MBP-AtACX4 was provided by Ingenza on the plasmid pMMA050. The first stage of library assembly therefore required replacing *at-acx4* with *mbp-at-acx4* in both pMAE-1 and pMAE-2, to make pMAE-3 and pMAE-4, respectively. pMMA050, pMAE-1 and pMAE-2 have *NotI* and *XbaI* restriction sites flanking the region of the plasmids

carrying genes for AATm4 and AtACX4/MBP-AtACX4. Double digest with both restriction enzymes was unsuccessful, so the vectors and MBP-AtACX4 insert fragments were generated by sequential reactions with first *XbaI* and then *NotI*. These fragments were then ligated together using T4 DNA ligase to produce pMAE-3 and pMAE-4 (Figure 4-1B).

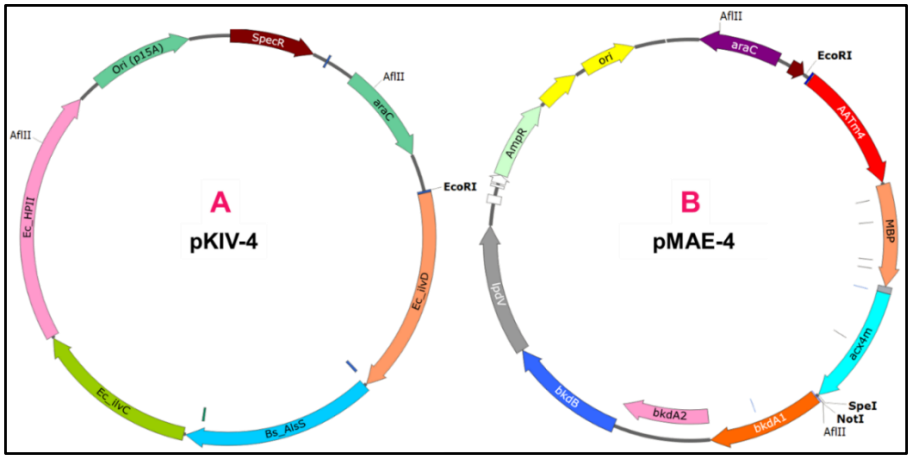


Figure 4-1: Plasmid maps of pKIV-4 and pMAE-4.

(A) pKIV-4 upregulates 2-ketoisovalerate formation, and (B) pMAE-4 produces BMA. *AlsS* = acetolactate synthase (*Bacillus subtilis*) EC 2.2.1.6, *ilvC* = acetohydroxyacid isomeroreductase (*Escherichia coli*) EC 1.1.1.86, ilvD = dihydroxyacid dehydratase (*E. coli*) EC 4.2.1.9, *bkdA1, bkdA2, bkdB, ipdV* (BCKD) = branched-chain ketoacid dehydrogenase (*Pseudomonas aeruginosa*) EC 1.2.4.4, *acx4m* = acyl-CoA oxidase (*Arabidopsis thaliana*) EC 1.3.3.6 (Ingenza), and *aatm4* = alcohol acyltransferase m4 (*Malus pumila*) (MCUK). *specR* = spectinomycin resistance, *araC* = arabinose promoter, *ampR* = ampicillin/carbenicillin resistance.

Finally, I transformed electrocompetent WT *E. coli* BW25113 cells first with pKIV variants, followed by pMAE variants. Colony PCR confirmed plasmid uptake in all strains. LR tested the library I generated using a combination of flask tests and fed-batch fermentation (data not available). Although many of the library strains produced largely equivalent results, experiments showed that a combination of pKIV-4 and pMAE-4 resulted in the highest BMA selectivity, with BMA as the major product of fermentation. As a result of the outcomes from fed-batch fermentation, I selected pKIV-4 and pMAE-4 for the production pathway. These plasmids were successfully transformed into the selected product-resistant mutants *via* electroporation, which was confirmed by colony PCR (Figure 4-2).

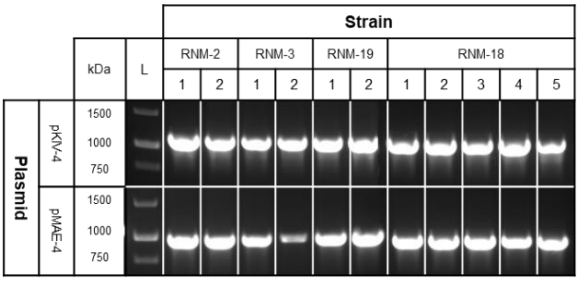


Figure 4-2: Colony PCR of pKIV4/pMAE4 strains

Colony PCR reactions were loaded onto a 1% agarose gel in TAE buffer, which ran at 80 V for 40 min. 1 kb plus DNA ladder.

Using pKIV-4 and pMAE-4, I assembled mutant BMA production strains using **RNM-2**, **RNM-3**, **RNM-18,** and **RNM-19**. I selected these strains because, as previously shown (Table 4-2), they represent a range of different BMA productivities, maximum biomass concentrations, and lag times from the previous work carried out by Ingenza and RNM.

4.2.2 Flask fermentation for production analysis of BMA variant strains

I cultured **WT_K4.M4** overnight with 1% glucose at 37°C and 250 rpm. After 12-15 h the cells were sub-cultured into 500 mL sealed filter flasks with PTFE tubing clamped shut on the side arm. Each overnight culture was diluted to OD_{600} 0.1 in 65 mL LUND medium with 1% glycerol and 0.77 g L^{-1} yeast extract. Pathway expression was induced at OD_{600} 0.6 by addition of 0.02% (w/v) arabinose. I added 1-butanol, to allow AATm4 to begin forming BMA, at 1 h intervals to a total concentration of 15 mM. Although butanol concentrations of up to 1 M have been used *in vitro* to assay AAT activity [162], the majority of studies use concentrations of butanol at between 2 mM [137, 163] – 20 mM [135]. 15 mM was selected as a middle-high butanol concentration in order to maximise BMA production without compromising cell viability. I carried out a growth screen using WT *E. coli* BW25113. In this screen, cells were grown to mid-exponential phase, after which they were sub-cultured to OD_{600} 0.1 in a microtiter plate with concentrations of 0 mM to 20 mM butanol (BuOH). OD_{600} was then monitored over the course of 20 h, until all cultures appeared to have reached a stationary phase (Figure 4-3). No notable change in growth was observed over this

period, and so adding 15 mM BuOH to BMA producer cells was not expected to have a toxic effect on *E. coli*.

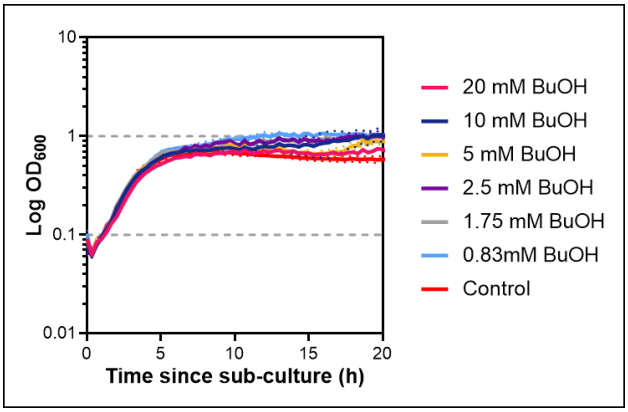


Figure 4-3: Growth of *E. coli* in the prescence of extracellular butanol

WT *E. coli* BW25113 was grown from OD_{600} 0.1 in a microtiter plate with wells containing ■ = 0 mM, ■ = 0.83 mM, ■ = 1.75 mM, ■ = 2.50 mM, ■ = 5 mM, ■ = 10 mM, or ■ = 20 mM butanol. The OD_{600} was monitored using a Bioscreen C (Thermo Labsystems) which read the microtiter plate once every 5 min over a period of 20 hours. Mean values of the triplicate dataset are shown as ▬ with error indicated as ●●●.

At 0 h, 5 h and 24 h, samples were collected *via* the PTFE side arm by inverting the sealed flasks and inserting a 5 mL syringe. These samples were then extracted 1:1 into ethyl acetate to identify the products using GC-MS (Figure 4-4). Over the course of 24 h after induction, as measured at 0 h, 5 h and 24 h, **WT_K4.M4** cells continued to grow, reaching an OD_{600} of 3.03. A total of 0.314 mM ester was detected in the supernatant samples after 24 hours (Figure 4-4A). Of this, 56% was BMA, making BMA the major ester product from flask culture. The BMA concentration at 5 hours was 0.061 mM, just under half of the total concentration at 24 hours, 0.175 mM. This corresponds to a productivity of 1.68 µM h^{-1} OD_{600}^{-1} (1.03 x 10^{-3} g L^{-1} h^{-1}), which is almost 2000x lower than the target productivity of 2 $gL^{-1}h^{-1}$ (Figure 4-4B). The error in this reading was also particularly high, likely due to the volatility of BMA causing variations in the concentration remaining in solution.

Several other peaks were identified in the GC-MS trace. I identified the potential identities of these peaks using the predicted structures library on the GC-MS software. I then ran standards of the selected esters using GC-MS to verify peak identities. The

unknowns had retention times of 5.1 min, 7.8 min and 9.0 min, which corresponded to butyl acetate (BA), butyl isobutyrate (BIB) and butyl isovalerate (BIV) respectively. I then ran calibrations to determine the concentrations of these off-target esters using standards made up to concentrations between 0.3 mM – 10 mM (Supplementary Figure S 3). Low concentrations of BA and BIV were detected, at 0.092 mM and 0.002 mM after 5 hours, and 0.016 mM and 0.015 mM after 24 hours, respectively. 34% of total ester concentration after 24 hours was BIB. At 5 hours the concentration of BIB was similar to that of BA and BIV at 0.013 mM. This concentration of BIB had increased 8-fold after 24 hours, to a final value of 0.108 mM. Correspondingly the specific productivity of BMA and BIB, 1.68 µM h^{-1} OD_{600}^{-1} and 1.18 µM h^{-1} OD_{600}^{-1}, were 5-12 times greater than the specific productivities for BA and BIV (Figure 4-4B). No other esters were detected to significant levels using GC-MS.

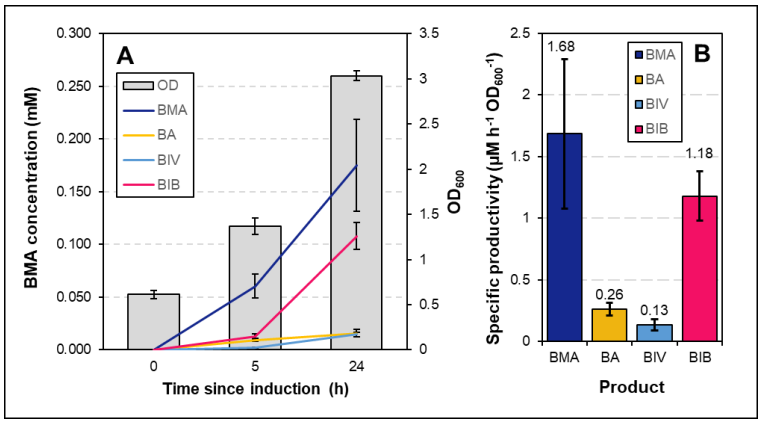


Figure 4-4: Flask characterisation of BMA biosynthesis from *E. coli*.

WT *E. coli* transformed with pKIV-4 and pMAE-4 was induced to synthesize esters following addition of butanol. Concentrations of four esters = BMA, = BA (butyl acetate), = BIV (butyl isovalerate), = BIB (butyl isobutyrate) and = OD_{600} were determined at two time points after induction (A) and specific productivity determined (B). Error was calculated as standard error of the mean of a triplicate dataset.

4.2.3 Extra- and intracellular HMA formation from WT_K4.M4

At this point I investigated HMA formation. If BMA formation is indeed limiting its own production, then we would expect to see much higher product titres from a less prohibitive product. I therefore set up flask cultures of **WT_K4.M4** as for the BMA formation assays. Only one 24 h sample was collected, both because a sufficient

volume of cells were required to prepare lysed samples and because HMA is less volatile than BMA, reducing the risk that ester would be lost from the culture before the 24 h mark. To form HMA, 1-hexanol was provided instead of the 1-butanol for BMA formation.

Maximum OD_{600} was reduced in the HMA producing cells compared to the OD_{600} of 3.03 observed for BMA producer cells after 24 h. I observed the same difference using *E. coli* **RNM-18** (Figure 4-5A). In both cases, the OD_{600} of HMA producer cells did not exceed 1.0 over the course of 24 hours. I carried out a simple growth assay to check whether this growth inhibition effect might result from hexanol toxicity. WT *E. coli* BW25113 was grown to mid-exponential phase, after which the cells were sub-cultured to OD_{600} 0.1 in a microtiter plate with concentrations of 0 mM to 20 mM hexanol (HeOH), as tested with BuOH in 4.2.2. OD_{600} was then monitored over the course of 20 hours, until all cultures appeared to have reached a stationary phase. Between 0-5 mM HeOH no change in growth was observed (Figure 4-5B). At 10 mM HeOH an increase in lag phase from roughly 1 hour to over 10 hours was observed, with the cells eventually reaching the same OD_{600} as control cultures after 18 hours. At 20 mM HeOH no growth was observed over the 20 hour period.

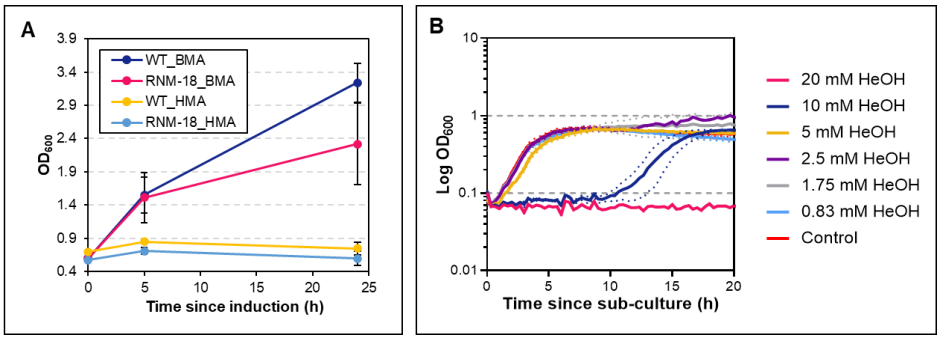


Figure 4-5: Effect of butanol and hexanol addition on the growth of *E. coli* BW25113.

A: OD_{600} values measured from samples taken for GC-MS analysis during BMA/HMA production experiments. **WT** — and **RNM-18** —, and HMA producing **WT** — and **RNM-18** —. B&C: WT *E. coli* was grown from OD_{600} 0.1 in a microtiter plate with wells containing = 0 mM, = 0.83 mM, = 1.75 mM, = 2.50 mM, = 5 mM, = 10 mM, or = 20 mM hexanol. The OD_{600} was monitored using a Bioscreen C (Thermo Labsystems) which read the microtiter plate once every 5 min over a period of 20 hours. Mean values of the triplicate dataset are shown as — with error indicated as ●●●.

At 24 h after hexanol addition, I collected supernatant samples and extracted directly into 1:1 ethyl acetate as before to measure the extracellular HMA concentration. HMA is more hydrophobic than BMA, so I also lysed the HMA producer cells to release intracellular products under the assumption that a significant quantity of HMA may be retained within *E. coli*. To prepare these intracellular samples, I pelleted and resuspended cell culture in fresh LB twice before adding lysozyme, benzonase nuclease and protease inhibitor to resuspended cells. This was in an effort to remove as much of the extracellular HMA from samples before lysis, which I carried out using sonication. Harvested cell extract was then also extracted into ethyl acetate for GC-MS analysis (Figure 4-6).

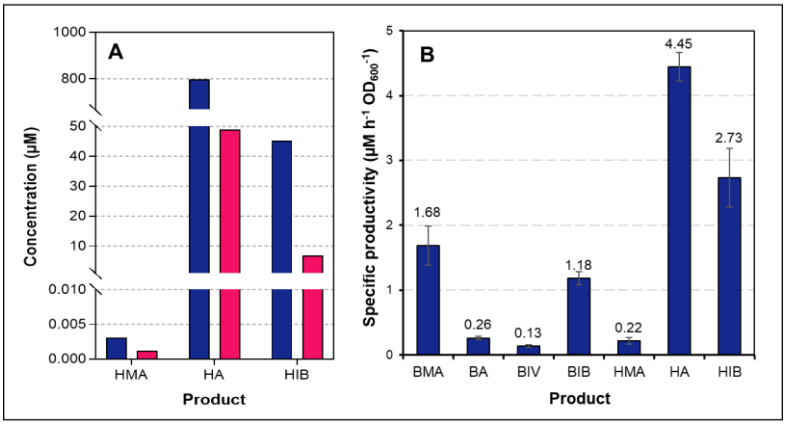


Figure 4-6: Butyl- and hexyl- ester formation from *E. coli*.

WT *E. coli* transformed with pKIV-4 and pMAE-4 was induced to form either butyl- or hexyl- esters, after the addition of butanol or hexanol, respectively. Ester concentration was measured in the ■ = extracellular, and ■ = intracellular fractions at 24 hours after induction (a) and specific productivity was calculated (b). HMA – hexyl methacrylate, HA = hexyl acetate, HIB = hexyl isobutyrate, BA = butyl acetate, BIV = butyl isovalerate, BIB = butyl isobutyrate. Error was calculated as the standard error of the mean of a triplicate dataset.

The total ester detected in both phases was 0.900 mM, of which 0.841 mM was in the extracellular sample (Figure 4-6A). Intracellular ester concentrations were 15 times lower than for the extracellular, assuming no extracellular ester remained in the lysed samples. Of the total ester in both samples, 88.5% was extracellular hexyl acetate (HA). Hexyl isobutyrate (HIB) was the second most abundant ester, making up 5.3% of the extracellular ester and 12% of the intracellular ester. No equivalent ester to BIV was detected. When compared to BMA production (Figure 4-6B), the specific

productivity of HA formation was significantly higher than all other recorded ester titres, followed by HIB. Indeed, product specificity for HA was 4.45 µM h^{-1} OD_{600}^{-1} compared to 2.73 µM h^{-1} OD_{600}^{-1} for HIB and 1.69 µM h^{-1} OD_{600}^{-1} for BMA. Specificity for HMA as a proportion of total ester compared to BMA was reduced from 56% of the total to 2.9%.

4.2.4 Flask characterisation of BMA-resistant *E. coli* mutants

Flask characterisation of the four selected mutant strain expressing the BMA pathway; **RNM-2_K4.M4**, **RNM-3_K4.M4**, **RNM-18_K4.M4** and **RNM-19_K4.M4** was carried out under conditions which were identical to those we used during the **WT_K4.M4** flask tests. Pre-culture stages of the tests were longer due to the reduced growth rates of the mutants, **RNM-2** in particular (Figure 4-7).

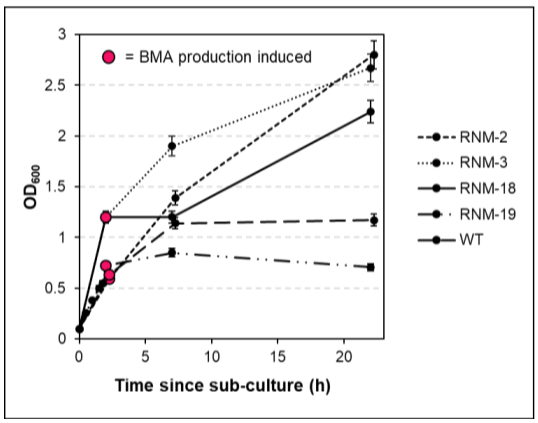


Figure 4-7: Growth of BMA resistant producer strains

Growth of the **WT_K4.M4, RNM-2_K4.M4**, **RNM-3_K4.M4**, **RNM-18_K4.M4,** and **RNM-19_K4.M4** strains before and after BMA production induced over a 24 hour period. Measurements collected on and after induction taken from 0, 5 and 24 h BMA production samples. Data collected in triplicate and error bars represent standard error of the mean.

To combat this, seed culture start times were staggered to reduce the disparity between cultures. In addition to BMA, peaks corresponding to BA, BIV and BIB were observed in all the RNM mutants. Despite a longer lag phase, **RNM-2_K4.M4** reached the highest OD_{600} of 3.52 after 24 hours. **RNM-3_K4.M4** and **RNM-18_K4.M4** showed lower growth, at OD_{600} 2.62 and 1.97 after 24 hours. **RNM-19_K4.M4** appeared to

have a limited maximum cell concentration, with an OD_{600} of 0.82 after 5 hours, which only increased by 0.15 to 0.97 at the conclusion of the experiment.

RNM-2_K4.M4 produced more BIB than BMA, of 0.073 mM ester 0.039 mM (53%) was BIB (Figure 4-8A). The BMA concentration was 0.026 mM, just under 7 times lower than the concentration of BMA formed by **WT_K4.M4**. Minimal BMA was produced between 5 hour and 24 hour samples, increasing by only 0.009 mM over the 19 hour period. Conversely, BIB production continued after 5 hours, with 82% of the total BIB formed between the 5 hour and 24 hour time points. Little BIV and BA were detected, at 0.005 mM and 0.003 mM, respectively. Low BIV and BA concentrations were also recorded from **RNM-3_K4.M4**, at 0.002 mM and 0.001 mM. **RNM-3_K4.M4** formed a total of only 0.0374 mM ester, of which similar concentrations of BMA and BIB were produced, each making up roughly 46% of the total. Marginally more BMA was formed, at 0.01 mM more than BIB concentration (Figure 4-8B).

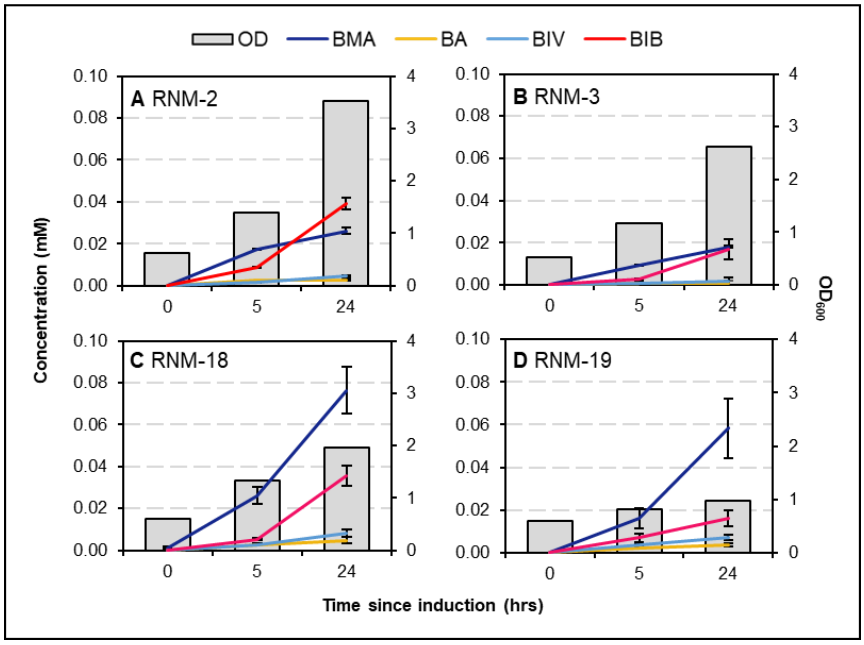


Figure 4-8: Flask characterisation of BMA production from product-resistant mutants.

RNM-2 (a), **RNM-3** (b), **RNM-18** (c) and **RNM-19** (d) expressing pKIV-4 and pMAE-4 were induced to produce esters in the presence of butanol. Several ester concentrations = BMA, = BA, = BIV, = BIB, and the = OD_{600} were recorded at 5 hours and 24 hours after induction. Error bars are calculated as standard error of a triplicate dataset.

Highest BMA concentrations were observed from the **RNM-18_K4.M4** mutant (Figure 4-8C). This produced 0.026 mM BMA after 5 hours and 0.076 mM BMA after 24 hours, making up 61% of the 0.124 mM total ester. This translated to just under a three times increase in BMA concentration between 5 and 24 hours. 0.036 mM BIB was detected in the 24 hour sample, a similar concentration to the concentration of BIB produced in the **RNM-2_K4.M4** strain. Once again only small concentrations of BA and BIV were detected at both time points, reaching a maximum of 0.005 mM and 0.008 mM respectively after 24 hours.

RNM-19_K4.M4 grew slowest post-induction but was the second highest RNM mutant producer of BMA, forming 0.016 mM after 5 hours and 0.058 mM after 24 hours (Figure 4-8D). This made up 68% of the 0.085 mM total ester. A similar increase to **RNM-18_K4.M4** in BMA concentration was observed between 5 and 24 hours, with the concentration of BMA increasing 3.6 times during that period. BIB formation from **RNM-19_K4.M4** was the lowest from all tested strains, including the WT. After 24 hours the BIB concentration in samples was 0.016 mM, making up 19% of the total ester. As with all other strains, BA and BIV were the least abundant products, and were detected at 0.004 mM and 0.007 mM concentrations, respectively.

When specific productivity was calculated, **RNM-19_K4.M4** showed low overall growth and similar BMA production to **RNM-18_K4.M4** (0.051 mM) and consequently, specific productivity of that strain was comparable, and in some cases higher, than from **WT_K4.M4** (Figure 4-9). **RNM-19_K4.M4** grew to half the OD_{600} of **RNM-18_K4.M4** and yet produced BMA concentrations to >90% the concentration from **RNM-18_K4/M4**. **RNM-19_K4.M4** also simultaneously produced less BA and less BIB than WT *per* hour *per* OD_{600}, although statistically identical concentrations of both compared to **RNM-18_K4.M4**. As expected, specific productivity of both BA and BIV was low compared to BIB and BMA for all strains. It also appeared that both **RNM-18_K4.M4** and **RNM-19_K4.M4** had a lower selectivity towards BIB compared to **WT_K4.M4**.

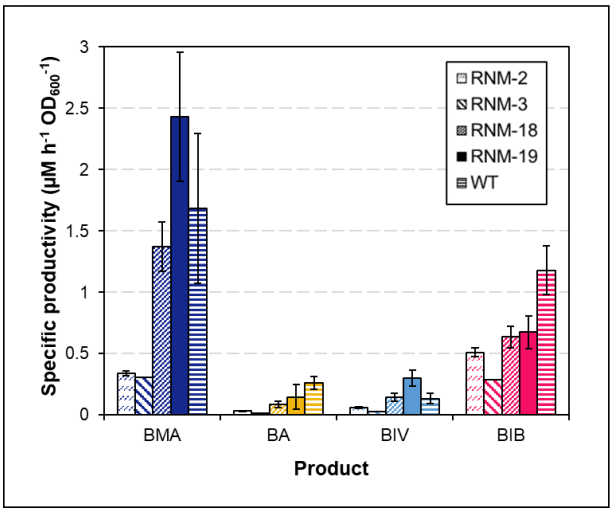


Figure 4-9: Specific productivity of BMA resistant mutants.

RNM-2, **RNM-3**, **RNM-18**, **RNM-19**, and **WT** transformed with pKIV-4 and pMAE-4 were induced to synthesize esters in the presence of butanol. Specific productivity was calculated from a sample taken 24 hours after induction. = BMA (butyl methacrylate), = BA (butyl acetate), = BIV (butyl isovalerate), = BIB (butyl isobutyrate). Error bars represent the standard error of the mean from a triplicate dataset.

4.3 Discussion

4.3.1 BMA formation from product resistant mutants

Perhaps what was most clear from this series of experiments was that increasing extracellular resistance to BMA in *E. coli* did not create a concurrent increase in BMA formation, or particularly improve ester selectivity. The highest BMA titre from these flask tests was 0.175 mM, produced by **WT_K4.M4**. This was almost 2-fold greater than the production recorded from the best performing mutant, **RNM-18_K4.M4**, which made 0.076 mM BMA.

BMA formation from **RNM-2_K4.M4** and **RNM-3_K4.M4** was particularly low, at 0.026 mM and 0.018 mM, respectively. This is perhaps unsurprising for **RNM-3**, considering the large number of knockouts in this strain will undoubtedly interfere with normal cellular function, as well as the ability of *E. coli* to grow under harsher growth conditions (Table 4-1). For example, a large number of the **RNM-3** knockouts occur in the *ymfM* and *oweE*, which are associated with the cellular response to DNA damage [164]. Biotransformation data from Ingenza also indicated that **RNM-2** and **RNM-3**

(0.35 mM and 0.37 mM) produced less BMA than **RNM-18** (1.4 mM), which in their experiments was also the highest producer. The biotransformation experiments carried out by Ingenza used cells that were resuspended to OD_{600} 50. This perhaps goes some way towards explaining the disparity between the BMA titres measured from flask cultures and those recorded from biotransformations.

Clearly in all cases the mutant strains, although able to grow effectively in the presence of 20% (v/v) exogenous BMA, are likely handicapped by the number of mutations accumulated in several genes associated with key processes. This includes the mutations in *rpoC* (K215fs/L361R/R1075C), *acrR* (unknown location/A191fs) and *rob* (R156H/A70T) which were present in most, if not all, of the tested mutants. Of these Rob in particular is a global stress response regulator, the mutation of which will likely have more far-reaching effects than simply increasing product-resistance, as was observed in the ALE work that produced the RNM mutants.

Interestingly, **RNM-19_K4.M4** produced more BMA than **RNM-2_K4.M4** and **RNM-3_K4.M4** (0.058 mM), despite biotransformation experiments previously showing the lowest production of all four mutants. In addition, **RNM-19_K4.M4** selectivity was favourable even when compared to the WT (Figure 4-9). **RNM-19_K4.M4** produced 2.4 µM h^{-1} OD_{600}^{-1} BMA, which was almost 1 µM h^{-1} OD_{600}^{-1} higher than the productivity of the WT. When the ratio of BIB:BMA was compared for both strains, a clear improvement can be observed; where **WT_K4.M4** is 1:1.4 and **RMN-19_K4.M4** is 1:3.6. The low OD_{600} of **RNM-19_K4.M4** throughout the experiment coupled with better selectivity contrasts with the faster growth and poor selectivity of **WT_K4.M4**. Indeed, the selectivity towards BMA as opposed to BIB is a marked improvement upon the WT data. Only two affected genes are unique to **RNM-19**, of the RNM strains tested here. Those both contain missense mutations, one in *marR* (V84G) and the other in *ompR* (R15S). As previously mentioned, MarR represses transcription of the global regulator MarA, as well as transcription of *marB,* which represses *marRAB* [88, 91]. One hypothesis may be that the mutations present in *marR* allow **RNM-19** a

reduced repression of its stress response systems. OmpR is also a regulator, repressing the expression of membrane porins OmpF and OmpC [94]. Increased production of OmpF and OmpC may increase the cells ability to rapidly remove BMA, and other toxic by-products, into the supernatant. However, many of the genes mutated in **RNM-19** are also affected identically in other RNM strains. Further work will need to be done to determine precisely what causes such a clear improvement in product specificity.

RNM-18_K4.M4, the highest producing mutant strain, contains mutations in *rpoC* (K215fs), *acrR* (A191fs) and *rob* (A70T). Although changes in these genes are common to the other RNM strains, these particular mutations are only present in **RNM-18**. As previously mentioned, it is somewhat surprising that **RNM-18** is able to survive despite what appears to be a large disruption to the β-subunit of RNAP. However, this strain also has several unique mutations in a collection of additional genes. One that is of particular interest is the frameshift mutation of *clsA*, A448fs. *ClsA* encodes the protein cardiolipin synthase which has been associated with increased resistance to 3,4-dihydroxybutyl-1-phosphonate when mutated [165]. Mutations in this gene also increased the lag time and maximum cell concentrations of *E. coli* JM109. Unlike the *rpoC* K215fs mutation, *clsA* A448fs only affects the C-terminus of ClsA, perhaps allowing the protein to retain some function. Similarly, DnaK, which has a V377G mutation in **RNM-18**, is a GroES/GroEL chaperone protein during the heat stress response that has been linked with increased exogenous butanol tolerance in *C. acetobutylicum* [33] and PHB tolerance in recombinant *E. coli* [97, 166]. V377 is located near to the substrate binding domain of DnaK, and so the **RNM-18** *dnaK* V377G mutation may have some impact on DnaK function [167]. However, several mutations are also linked to increased export through expression of membrane pumps; such as those in *ompX* (unknown location)and *yohJ* (L109R). It is difficult to speculate on the impact of the *ompX* mutation without further information, but the substitution in L109R *yohJ* replaces non-polar Leucine with a large, charged

Arginine side chain on the long 'tail' domain of the protein C-terminus, possibly impacting cellular localisation.

Despite a collection of mutations within genes that have previously been associated with increasing tolerance, **RNM-18_K4.M4** BMA production reached a maximum of only 43% of the BMA formation observed from **WT_K4.M4** *E. coli*. It is possible, as is clear with other RNM strains, that the introduction of many mutations within essential genes linked to core cellular responses, such as *rob* A70T, *dnaK* V377G, and *cra* I270fs, may inherently impair carbon flux in *E. coli*. As such, it is not possible to definitively conclude that increased BMA tolerance is detrimental to BMA formation. Further work must be done to identify which mutations are vital to tolerance, and which can be removed to enhance cellular functionality before the yield of BMA from tolerant strains could reasonably be expected to exceed that of a less metabolically compromised WT *E. coli* strain.

Maximum ester concentrations from each strain were far below the concentrations required to calculate a complete carbon balance for the experiment. It is likely that other by-products were also being synthesised that were not detected by the GC-MS, or were perhaps too volatile to remain in the supernatant during sample preparation. Previous work done by Ingenza and AY revealed that a major by-product during BMA synthesis appears to be isobutyric acid (IBA), present in higher concentrations than is seen cumulatively of all ester products investigated in these experiments. It was encouraging however that the ester specificity of the strains tested here favoured BMA over BA, BIB and BIV.

4.3.2 BMA synthesis toxicity in WT and product-resistant mutants

BMA's IC_{50} in *E. coli* is 0.07-0.11 gL^{-1}, or 0.49-0.77 mM. The IC_{50} value is of course dependant on the efficiency of the export system removing BMA from cytoplasm. Many BMA-resistant mutants possess mutations in the proteins or the expression systems controlling the AcrAB/TolC export system. This suggests that these pumps

may be responsible for the removal of BMA from the periplasm. If this is the case, then an internal BMA toxicity limit may be reached, despite the observation of low concentrations of BMA outside the cells. Additionally, the IC_{50} does not account for the cumulative effect of toxicity of the butanol fed into the system, in addition to the by-product esters, each of which may have some level of toxicity in their own right. Indeed, 0.31 mM total of BMA, BA, BIB and BIV were formed by **WT_K4/M4**, which is not far from the lower limit of BMA's IC_{50} 0.49 mM, although of course this value will differ depending on the ester.

I replicated the BMA flask test set up, forming HMA instead, aiming to identify a simple solution to avoiding product toxicity and increasing product titres. However, HMA specific productivity was at least 4-fold lower than BMA productivity in WT *E. coli*. Exceedingly high HA concentrations also indicate that AATm4 is more specific for the acetate ester, rather than methacrylate, when hexanol is the alcohol. This is not only an issue for HMA titres, but would also pose an increasing problem as process volume grows, where acetate itself will be toxic to cells. In addition to this, hexanol is toxic to *E. coli* at lower concentrations than butanol, so further experiments to produce HMA would require some level of tolerance engineering to alleviate hexanol toxicity. Although HMA formation may avoid issues of product toxicity, problems with substrate toxicity and AATm4 specificity for hexyl esters limit its benefit. Hexanol toxicity will clearly also be a major limitation on HMA production. In addition to the results presented here, Ingenza reported on the production of HMA from biotransformations at OD_{600} 25, and discovered globules of ester built up inside cells to 20% of cell volume, leading to eventual cell death (MCUK conference, 2018). Once a higher flux production pathway is developed, to the extent that BMA toxicity becomes an immediate problem, it may be useful to investigate alternative methacrylate esters.

Regardless of the potential for higher toxicity both internally and from the by-product formation during BMA synthesis, it was clear that reduced sensitivity to BMA did not improve BMA formation. This indicated that toxicity of the pathway components,

although it may affect producer cells, was not the key limitation affecting BMA production. In order to assess whether the RNM strains do confer an advantage, and resistance to internal BMA, expression and induction of the biosynthetic pathway would need to be significantly inhibiting the growth and function of the WT *E. coli*.

Of course, it was not known precisely what proportion of BMA remained inside the cells during these experiments [147]. Unfortunately the conditions required to lyse the cells as used for HMA extraction precluded their use to lyse cells containing BMA, as it is likely the majority of the BMA released would be lost into the gas phase during sonication. Despite this, even if 50% of BMA remains inside the cells, WT *E. coli* would still be producing concentrations of BMA that are well below the toxic threshold. It is more likely, therefore, that flux through the production pathway is the limiting factor in total BMA concentrations, as opposed to growth inhibition from product toxicity.

4.3.3 Metabolic causes of carbon flux limitation

Although it is clear that product toxicity cannot be completely ruled out as an issue for BMA biosynthesis, these data showed that a large issue is product specificity within producer strains. Regardless of BMA toxicity, as production increases these problems will only compound, resulting in impure products that will add expensive additional stages to downstream processing. Another problem with prevalent by-product formation is that it is likely an indicator of carbon “hold-ups” within the metabolic pathway that has been introduced.

Many of the by-products formed by BMA production strains likely arise from AAT substrate promiscuity [159]. The modified AAT used in these experiments, AATm4, was altered to increase selectivity towards methacrylyl-CoA (Figure 4-10A). This did not prevent the formation of relatively large concentrations of BIB, by the action of AATm4 with IB-CoA and butanol (Figure 4-10B). This could either be the result of AATm4 possessing a higher affinity for IB-CoA or slow activity of AtACX4 could be

facilitating an accumulation of IB-CoA, which in turn favours increasing production of BIB. Of course, it is likely to be a combination of both factors which causes BIB to be formed as well as BMA. Acetyl-CoA is a central metabolite in *E. coli* metabolism. As such, an approximate intracellular concentration of 0.1 nmol (mg dry wt)$^{-1}$ acyl-CoA is available for AATm4 to utilise at stationary phase, allowing the formation of BA (Figure 4-10C) [160].

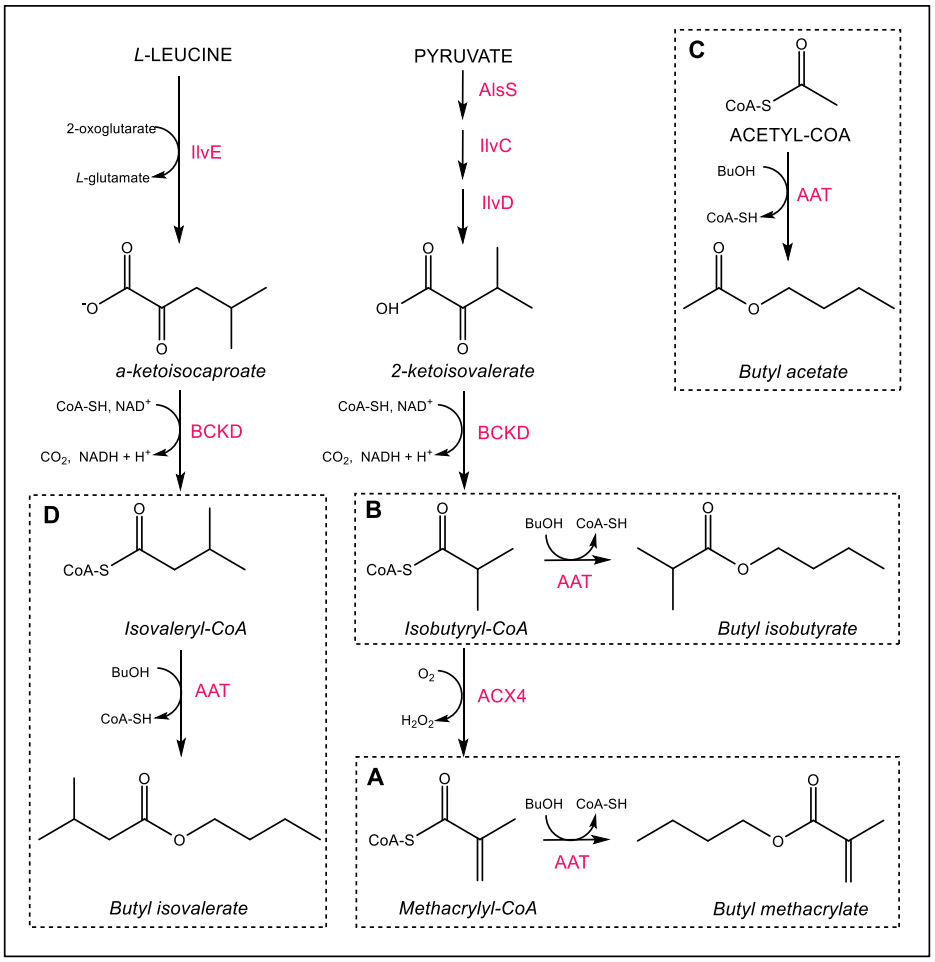


Figure 4-10: Formation of ester by-products during BMA biosynthesis.

Enzymes are shown in PINK. A: Shows desired production of BMA, B,C & D show formation of off-target esters. IlvE = branched-chain amino acid aminotransferase (*Escherichia coli*) EC 2.1.6.42, AlsS = acetolactate synthase (*Bacillus subtilis*) EC 2.2.1.6, IlvC = acetohydroxyacid isomeroreductase (*Escherichia coli*) EC 1.1.1.86, IlvD = dihydroxyacid dehydratase (*E. coli*) EC 4.2.1.9, BCKD = branched-chain ketoacid dehydrogenase (*Pseudomonas aeruginosa*) EC 1.2.4.4, ACX4 = acyl-CoA oxidase (*Arabidopsis thaliana*) EC 1.3.3.6, and AAT = alcohol acyltransferase m4 (*Malus pumila*) (MCUK).

The source of isovaleryl-CoA for BIV synthesis is less clear. A possible explanation is that BCKD may use α-ketoisocaproate, which is an intermediate produced by IlvE during *L*-Leucine degradation in *E. coli* [116]. Indeed, BCKD from *S. avermitilis* was previously used to upregulate the production of isovaleryl-CoA from α-ketoisocaproate in *E. coli* [116].

Not investigated in these experiments were the acid and alcohols isobutyric acid, isobutanol and isobutanal. Excess 2-KIV may be diverted by host enzymes such as pyruvate oxidase (PoxB) to isobutyric acid, and *via* generic decarboxylase and dehydrogenases to isobutanal and isobutanol, respectively [168]. However this explanation is less likely as data from Ingenza using ΔPoxB *E. coli* did not exhibit any significant reduction in their isobutanol formation. As a result of this various cross reactivity, all strains produce a cocktail of by-products due to the non-specific activity of a number of enzymes, primarily AATm4.

4.3.4 Carbon flux and not toxicity limits the formation of bio-BMA

The objective of these experiments was to clearly define whether BMA toxicity is limiting our maximum product titres. However what was clear from the product concentrations and by-product characterisation was that the limitations on BMA production were more widespread than simply the challenge of product toxicity. In order to improve BMA biosynthesis in a targeted way, a more detailed unpicking of the synthesis pathway was required.

The variety and prevalence of AATm4-synthesised ester by-products in the latter half of the pathway suggested that a flux problem occurs between 2-KIV and BMA. This is perhaps to be expected, considering that the upstream pathway is the fairly well-known up-regulation of *E. coli* valine synthesis. In contrast, BCKD, AtACX4 and AATm4 are all exogenous enzymes, two of which are from plants and not microorganisms, and had not been expressed in *E. coli* prior to the introduction of BMA biosynthesis.

Chapter 5: Carbon flux bottlenecks during BMA synthesis

5.1 Introduction

5.1.1 Implications of 3-hydroxyisobutyryl-CoA productivity for BMA formation

At this point, carbon flux towards the bioproduction of BMA from a pKIV-4/pMAE-4 strain was too low to sustain an economically viable industrial process, or even to allow robust investigation of the toxicity effect of intracellular BMA production on *E. coli.* This conclusion was evidenced by the BMA production observed from **WT_K4.M4** in shake flask experiments, reaching only 1.4 mM as a maximum BMA titre, while production of BMA from the **RNM-*n*** library of BMA-resistant mutants had no positive effect on total titre. If this is to be improved, a deeper understanding of carbon flux through the BMA pathway was required, in particular the latter half of BMA biosynthesis, during which a surfeit of off-target butyl- esters are formed. A more complete analysis of BCKD to AATm4 activity was therefore sought to help identify the current key limitations on BMA metabolism.

3-hydroxyisobutyryl-CoA (3-HIBA) production experiments carried out by Andrew Yiakoumetti (AY) demonstrated that very high carbon flux could be achieved from 2-KIV to 3-HIBA. The biosynthetic pathway for 3-HIBA is similar to the BMA pathway (Figure 5-1) [169, 170]. As with BMA formation, 2-KIV can either be supplied to the biocatalyst, or overproduced by the introduction of *alsS*, *ilvC* and *ilvD.* BCKD then converts 2-KIV to IB-CoA, with AtACX4 forming M-CoA. After M-CoA formation, in place of AATm4, an enzyme called enoyl-CoA hydratase (ECH) is introduced into *E. coli* which can rapidly convert M-CoA to 3-hydroxyisobutyryl-CoA. 3-hydroxyisobutyryl-CoA is subsequently hydrolysed to 3-HIBA by 3-hydroxyisobutyrate dehydrogenase (3-HIBH) [13]. AY cloned ECH and 3-HIBH from *P.aeruginosa* were into *E. coli* BW25113, on a constitutive plasmid also expressing BCKD and AtACX4. The strain containing this expression system, **HIBA-3**, produced 12 g L^{-1} 3-HIBA (AY, unpublished). As 3-HIBA production is only achievable through the production of IB-

CoA and its oxidation to M-CoA, this result demonstrated that under some conditions AtACX4 and BCKD facilitate high flux through a non-native metabolic pathway.

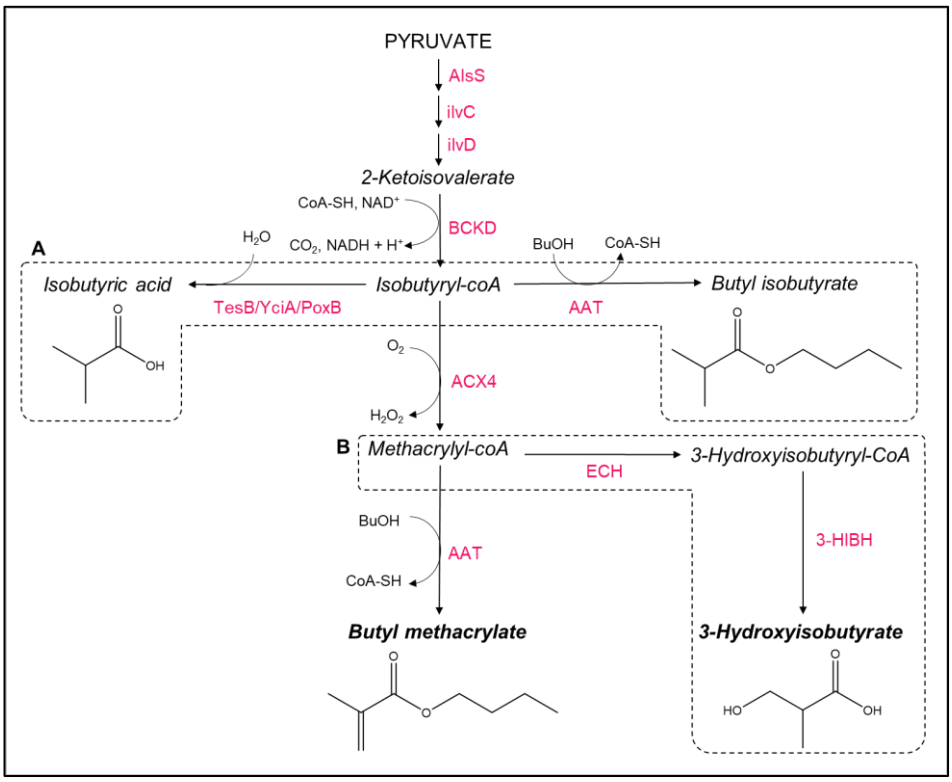


Figure 5-1: Alternative isobutyryl- and methacrylyl-CoA utilisation in BMA and 3-HIBA biocatalysts.

A: By-products isobutyric acid and butyl isobutyrate produced by AAT and native thioesterases compete with ACX4 for isobutyryl-CoA utilisation. B: Introduction of ECH and HCH in a AAT negative strain facilitates the formation of 3-hydroxyisobutyrate *via* methacrylyl-CoA and 3-hydroxyisobutyryl-CoA. Enzymes are shown in PINK. Target product from each pathway is shown in **bold**. BCKD = branched-chain ketoacid dehydrogenase (*Pseudomonas aeruginosa*) EC 1.2.4.4, ACX4 = acyl-CoA oxidase (*Arabidopsis thaliana*) EC 1.3.3.6, and AAT = alcohol acyltransferase m4 (*Malus pumila*) (MCUK), ECH = Enoyl-CoA hydratase (*Escherichia coli*) EC 4.2.1.17, 3-HIBH = 3-Hydroxyisobutyryl CoA hydratase (*E. coli*) EC 3.1.2.4.

When considering the BMA and 3-HIBA pathways in *E. coli* together, two main differences are present: AATm4 is removed in the 3-HIBA strain, and ECH/3-HIBH are introduced. The differences between the ECH and AAT enzymes can go some way towards justifying the relative activities of 3-HIBA and BMA metabolism. AATs are promiscuous enzymes, usually active on an unpredictable and broad spectrum of both acyl-CoA and alcohol substrates. This was reflected in the formation of BA, BIB and BIV observed in Chapter 4:4.2.4. The K_M values for the acyl-CoA substrates of AATs usually limit their activity rate, and the overwhelming majority of kinetic AAT data has

only been recorded for acetyl-CoA substrates, with which many AATs are most active, including apple AAT, whether from *M. pumila* (MpAAT) or *M. domestica* (MdAAT). Both MdAAT and MpAAT have been shown to preferentially form hexyl-, 2-methylbutyl, and acetate esters [138, 171]. With this in mind, the closest K_M values for BMA are for MpAAT producing butyl acetate. These are 2.7 mM for butanol and 0.11 mM for acetyl-CoA [138]. No K_M data exists for apple AATs with M-CoA as a substrate, and with perhaps the closest match, butyryl-CoA, MpAAT was inactive. However, no apple AAT K_M lower than 2.7 mM has been reported for less favourable substrates than hexanol, 2-methylbutanol or acetyl-CoA. So for the purposes of this project the K_M for AATm4 was assumed to be at least 2.7 mM with M-CoA [138, 172].

ECH meanwhile is a well-characterised enzyme, with a low K_M towards most substrates with which it demonstrates activity (Table 5-1). The closest substrate to M-CoA for which kinetic data is available for ECH is crotonyl-CoA. The average K_M value obtained for this substrate is 29 µM, three times lower than the range that can be expected for AATm4 [173-176] (Table 5-1).

Organism	Substrate	K_M (µM)	V_{max} (mmol min^{-1} mg^{-1})	pH	°C	Ref
Bos taurus	Crotonyl-CoA	20	$3.5x10^6$ mol min^{-1} mol^{-1}	7.4	25	[173]
	Hexadecenoyl-CoA	500	-	7.4	25	[173]
	Crotonyl-CoA	-	$7.3x10^5$ mol min^{-1} mol^{-1}	7.5	25	[177]
	Penta-2,4-dienoyl-CoA	24	-	7.5	30	[174]
	Crotonyl-CoA	35	-	7.5	30	[174]
	trans-2-hexenoyl-S-CoA	25	-	7.4	25	[173]
	Methacrylyl-CoA	Activity reported - no values given		7.4	25	[173]
C. acetobutylicum	Crotonyl-CoA	30	149	8.0	25	[175]
	Hexenoyl-CoA	130	0.89	8.0	25	[175]
M. smegmatis	Crotonyl-CoA	82	2.49	7.5	27	[178]
	Decenoyl-CoA	91	1.04	7.5	27	[178]
	Hexadecenoyl-CoA	105	0.15	7.5	27	[178]
Sus scrofa	Crotonyl-CoA	13	1.67	8.0	25	[176]
	Hexenoyl-CoA	29	1.28	8.0	25	[176]
	Octenoyl-CoA	29	0.91	8.0	25	[176]
	Decenoyl-CoA	29	0.54	8.0	25	[176]
	Dodecanoyl-CoA	30	0.16	8.0	25	[176]
E. coli	Crotonyl-CoA	50	-	8.0		[179]
	2-decenoyl-CoA	8	-	8.0		[179]

Table 5-1: Experimentally determined activities of enoyl-CoA hydratase enzymes

Kinetic values reported from literature searching entries recorded on the BRaunscweig Enzyme Database for EC 4.2.1.17. Where cells contain "-" no data was reported for this value.

It is also possible that high 3-HIBA titres were a result of changing the expression construct: pHIBA-3 constitutively expresses all pathway genes and reorders them compared to the original plasmid pMAE-4. It was therefore insufficient to hypothesize that poor BMA titres were solely due to inherently low activity of AtACX4, and as such, dissecting the latter half of the BMA pathway potentially allowed us a more complete understanding of why biosynthesis of methacrylate esters did not reach titres on the scale of 3-HIBA production.

5.1.2 Designing an approach to identify late-stage BMA formation bottlenecks

As of this point, several factors potentially limited yield from the biocatalyst (Figure 5-2A). Toxicity was the most obvious cause, as BMA was previously shown to limit growth of *E. coli* at low concentrations, and NMR data suggests it is capable of intercalating into the *E. coli* cell membrane [85]. However, BMA toxicity was not the only concern. Other esters, especially BA and BIB, are themselves likely to be toxic to *E. coli* [180]. Indeed, a 2017 study investigated the inhibitory effects of 32 esters on *E. coli* MG1655. BA reduced growth rate and maximum cell density below 1 g L^{-1} concentrations. Similar esters such as butyl butyrate and butyl propionate also reduced cell growth and a correlation was found between reduced polarity and increasing *E. coli* toxicity [180]. However, the by-product formation and sub-toxic BMA titres obtained during our shake flask experiments demonstrate that there is likely a bottleneck during BMA formation other than simply the toxicity of the final product.

We therefore hypothesised that a metabolic bottleneck occurred in the latter half of BMA synthesis. As evidenced by high 3-HIBA formation seen during the experiments carried out by AY, the final stages of biosynthesis, catalyzed by AtACX4 and AATm4, were the most likely cause . The literature supported this hypothesis, and several likelihoods were possible. For one, M-CoA is known to readily react with available sulfhydryl groups (-SH) through Michael addition reactions (Figure 5-2B) [181]. This occurs spontaneously under aqueous conditions, and can occur between M-CoA and free cysteine, cysteine residues or glutathione (GSH) to form an M-CoA-GSH adduct

[182]. Not only will the spontaneous removal of M-CoA reduce its availability as a substrate for AAT, M-CoA-GSH adduct formation would also decrease the intracellular availability of GSH, further reducing cell viability.

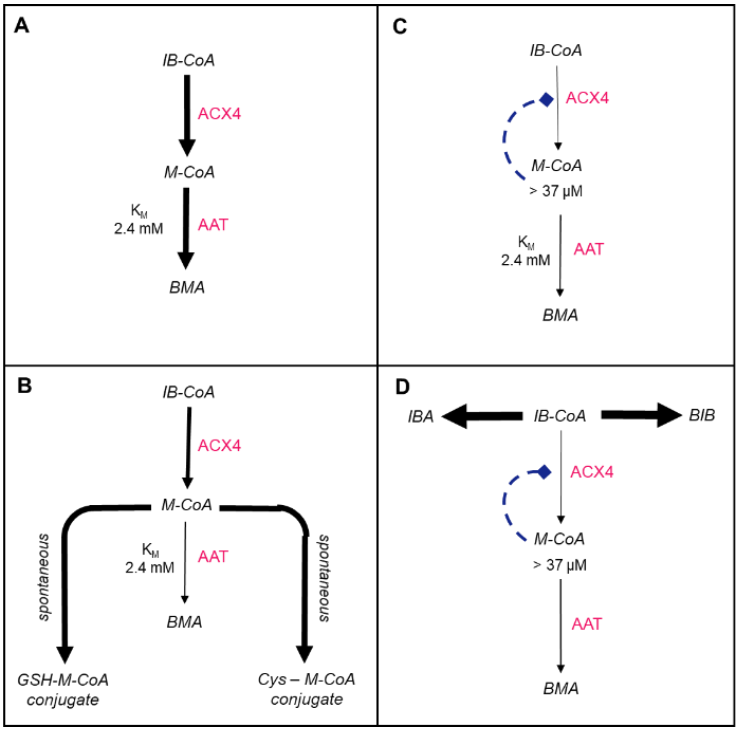


Figure 5-2: Methacrylyl-CoA bottlenecks during BMA biocatalysis.

A: Carbon flux from IB-CoA to BMA is limited by the high K_M of AATm4, B: which leads to an accumulation of M-CoA, favouring its spontaneous reaction with intracellular sulfhydryl groups, C: as well as inhibiting the activity of the ACX4 that produces it. D: The inhibition of ACX4 activity causes IB-CoA to also accumulate, favouring off-target by-product formation. Enzymes shown in PINK. **- - -** indicates product inhibition. IB-CoA = isobutyryl-CoA, M-CoA = methacrylyl-CoA, GSH-M-CoA = glutathione-methacrylyl-CoA conjugate, IBA = isobutyric acid, BIB = butyl isobutyrate. ACX4 = acyl-CoA oxidase (*Arabidopsis thaliana*) EC 1.3.3.6, and AAT = alcohol acyltransferase m4 (*Malus pumila*) (MCUK).

Another possible explanation linked to M-CoA production was that accumulation, not draw-off, of M-CoA may also limit BMA formation. This is because when ECH replaces AATm4 for 3-HIBA formation 12 gL^{-1} of 3-HIBA was produced (AY, unpublished). Compared to AAT, ECH rapidly converts M-CoA to 3-hydroxyisobutyryl-CoA, with a K_M of ~29 µM [130, 178, 183]. It is therefore likely that in the 3-HIBA biocatalysts M-CoA is removed by ECH as quickly as it is produced by AtACX4. In contrast, AAT from *M. pumila* has a K_M of roughly 2.7 mM and as such would require several fold higher M-CoA concentrations before it could achieve a comparable rate of

reaction to ECH [171]. Another hypothesis was therefore that as AATm4 cannot remove M-CoA as efficiently as ECH, M-CoA may accumulate to higher concentrations in BMA biocatalysts. This would have a detrimental effect on both cellular toxicity and AtACX4 activity, where ACX4 may become product inhibited (Figure 5-2C).

In the event that AtACX4 activity was either reduced or insufficient during BMA biocatalysis, we would expect to see an increase in the by-products formed from IB-CoA and other upstream intermediates (Figure 5-2D). As previously discussed, BIB is one such example, formed in a reaction catalyzed by the non-specific activity of AAT. However, several fermentations and shake flask experiments have also demonstrated that isobutyric acid (IBA) is a significant product from *E. coli* expressing BCKD, both in BMA and 3-HIBA metabolism. Although it is not yet verified which native enzymes are primarily responsible for the formation of IBA, it is believed to largely be the result of endogenous thioesterases or oxidases, in particular TesB, YciA and PoxB (Ingenza). However, unlike BIB, which is formed by an enzyme expressed as part of BMA biosynthesis, IBA draws-off IB-CoA by the action of enzymes native to *E. coli*. As with the possibility of M-CoA's spontaneous reactivity, IBA formation is likely to occur much more rapidly than AATm4 driven BIB formation. While not as affected by AtACX4 activity, as high concentrations of IBA were observed from 3-HIBA biocatalysts, this may limit the IB-CoA availability, and thus M-CoA, for AtACX4 and AATm4 to utilise, respectively. Another native metabolic drain common within many metabolically engineered pathways is also the carbon source draw-off from pyruvate to central metabolism. In some cases, this challenge is combated by reducing pyruvate dehydrogenase expression in the host genome [31, 39, 184].

The upstream section of the BMA pathway, from IlvC to BCKD, has been much more extensively studied than the latter. Similar versions of this pathway, which overexpresses a native *E. coli* lysine synthesis pathway, have been successfully used to produce high titres of isobutanol [37, 185, 186]. The same is not true for the two terminal stages of BMA synthesis: As implied by the high K_M value for AAT, it is

expected that activity through the penultimate stage of BMA synthesis will be less than optimal [138]. AAT enzymes are also known to accept a broad substrate range, often with a marked preference for acetyl-CoA, as opposed to the longer branched chain M-CoA [43, 109, 135, 162, 187]. This is evidenced by the formation of BIB, BA and BIV in BMA biosynthesis, which are most likely synthesised through non-specific AAT activity. Although AtACX4 is capable of facilitating g L^{-1} production titres, it is likely that the slow and non-specific activity of AAT will have the added effect of limiting flux through AtACX4. If M-CoA concentrations are too high, AtACX4 may be product inhibited. If this hypothesis proves true it poses a particular challenge, as product inhibition is usually resolved by low throughput screening of alternative enzymes, and it is possible the M-CoA pool in *E. coli* needs to remain at a low concentration to prevent product inhibition, with the knock-on effect of greatly reducing AATm4 activity and specificity.

Therefore, in this chapter I developed a series of strains to isolate each individual step of BMA biosynthesis between 2-KIV and BMA (Table 5-2). These strains were then tested using a biotransformation approach in place of the original flask test approach used in the previous chapter, allowing a switch to using a fully defined medium.

Product analysis of the strains identified using this biotransformation approach was used to identify the metabolic bottlenecks in late-stage BMA synthesis.

Shorthand	Full name	Predicted flux	Author
HIBA-3	*E. coli* Δ*infA* Δ*ldhA* :: KanR pHIBA-3	3-HIBA	AY
CAN-4	*E. coli* Δ*infA* Δ*ldhA* :: KanR pCAN-4	BMA	CT
CAN-5	*E. coli* Δ*infA* Δ*ldhA* :: KanR pCAN-5	BIB, IBA	CT
CAN-4R	*E. coli* Δ*infA* Δ*ldhA* :: KanR pCAN-4, pSC101_AcrR_SoxR	BMA	CT (LM pSC101)
CAN-5R	*E. coli* Δ*infA* Δ*ldhA* :: KanR pCAN-5, pSC101_AcrR_SoxR	BIB, IBA	CT (LM pSC101)

Table 5-2: List of strains used in de-bottlenecking experiments

In this chapter, AY provided the plasmid 3-HIBA, after my initial M-CoA synthesis subsequent samples were provided by MCUK, pSC101_AcrR_SoxRa was provided

by LM. Construction of CAN plasmids, biotransformations, protein purification and AtACX assays were carried out by me.

5.2 Results

5.2.1 Constructing BMA, BIB and IBA producing strains

AY prepared a plasmid, pHIBA-3, which contained all the genes required for formation of 3-HIBA from 2-KIV: BCKD (*bkdA1, bkdA2, bkdB, ipdV*), *at-acx4*, *ech* and *hch* (Figure 5-1). This was assembled into a pET20b+ vector backbone, in parallel with the introduction of a constitutive Anderson series promoter and *infA*, using HiFi assembly. pHIBA-3 was transformed into *E. coli* Δ*ldhA* Δ*infA*:: KanR, where the complementation of essential gene *infA* stabilises the plasmid in the cell. This gave the strain **HIBA-3**.

To facilitate direct comparison with **HIBA-3** and high yield 3-HIBA formation, I constructed two new plasmids using pHIBA-3 as the template: pCAN-4 and pCAN-5. pCAN-4 would contain *aatm4* in place of *ech* and *hch* to allow BMA formation, while in pCAN-5 *atacx4, ech* and *hch* would be removed, with the assumption this plasmid would produce BIB and IBA.

The plasmids pHIBA-3, pCAN-4 and pCAN-5 all contain the same vector backbone. Within this system, BCKD is under control of the J23104 promoter which is a medium strength Anderson series promoter, while expression of *atacx4, aatm4, ech* and *hch* is controlled by the J23119 promoter, the strongest of the Anderson series [188]. As availability of IB-CoA was not predicted to limit BMA formation, these promoters were chosen in order try and balance metabolite concentrations, increasing *atacx4, aatm4, ech* and *hch* expression relative to *bkdA1, bkdA2, bkdB*, and *ipdV*. The RBS used to initiate transcription of *atacx4, aatm4*, *ech* and *hch* is a well-known RBS from T7 phage as described by Olins *et al*. [189]. The same RBS is used for BCKD expression.

To prepare these plasmids, *aatm4* was first amplified from pMAE-4 (Figure 5-3A, pink). The PCR product was subsequently blunt end ligated into pJET1.2 to make the

pJET_AAT intermediate. An *atacx4* fragment was likewise amplified from pHIBA-3 (Figure 5-3A, light blue) and inserted before *aatm4* in pJET_AAT using HiFi assembly. This made the pJET_AAT_ACX4 intermediate. Both pJET_AAT_ACX4 and pJET_AAT were then digested using *Afl*II and *Eco*RI, as was the pHIBA-3 vector to give pJET_AAT_ACX4 (*Afl*II/*Eco*RI), pJET_AAT (*Afl*II/*Eco*RI), and pHIBA-3 (*Afl*II/*Eco*RI), respectively.

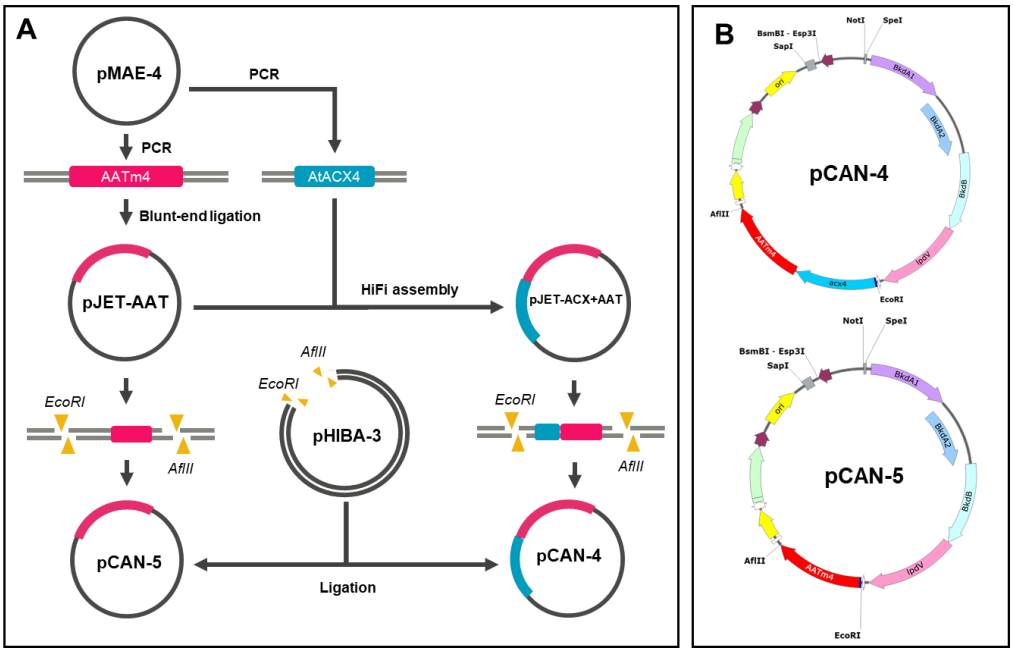


Figure 5-3: Assembling two plasmids for constitutive BIB and BMA formation

A: Plasmid assembly strategy. *aatm4* is in pink, *atacx4* is in blue. Restriction cut sites shown as yellow triangles. B: pCAN-4 and pCAN-5 plasmid maps. *infA* = translation initiation factor 1 (*Escherichia coli*) EC 3.6.5.3, BCKD = branched-chain ketoacid dehydrogenase (*Pseudomonas aeruginosa*) EC 1.2.4.4, ACX4 = acyl-CoA oxidase (*Arabidopsis thaliana*) EC 1.3.3.6, and AAT = alcohol acyltransferase m4 (*Malus pumila*) (MCUK). Plasmids designed and figure produced using Snapgene 5.0.8.

Ligation of the pJET_AAT_ACX4 (*Afl*II/*Eco*RI) fragment into pHIBA-3 (*Afl*II/*Eco*RI) made pCAN-4 (Figure 5-3B, pCAN-4). Ligation of pJET_AAT (*Afl*II/*Eco*RI) into pHIBA-3 (*Afl*II/*Eco*RI) produced pCAN-5 (Figure 5-3B, pCAN-5).pCAN-4 and pCAN-5 were used to transform into *E. coli* BW25113 Δ*ldhA* Δ*infA*:: KanR, to produce *E. coli* **CAN-4** and *E.coli* **CAN-5**, for the production of BMA or BIB and IBA, respectively. pSC101_AcrR_SoxR, previously made by LM, was co-transformed with pCAN-4 and pCAN-5 to result in the 'BMA resistant' producer strains **CAN-4R** and **CAN-5R** (Figure 5-4).

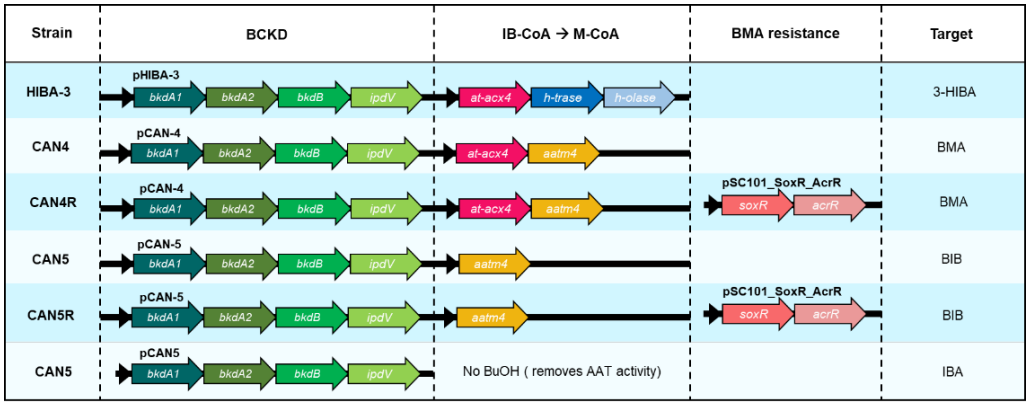


Figure 5-4: Utilisation strategy for de-bottlenecking production strains.

Heterologous genes expressed in each of our de-bottlenecking library strains. BCKD = branched-chain ketoacid dehydrogenase (*Pseudomonas aeruginosa*) EC 1.2.4.4 expressed using *bkdA1, bkdA2, bkdB,* and *ipdV*, *at-acx4* = acyl-CoA oxidase (*Arabidopsis thaliana*) EC 1.3.3.6, *aatm4* = alcohol acyltransferase m4 (*Malus pumila*) (MCUK), *h-trase* = Enoyl-CoA hydratase (*Escherichia coli*) EC 4.2.1.17, *h-olase* = 3-Hydroxyisobutyryl CoA hydratase (*E. coli*) EC 3.1.2.4.

Using these newly constructed plasmids in combination with pHIBA-3 would allow me to a number of producer strains, as outlined by Table 5-2 and Figure 5-4, whereby productivity through individual BMA biosynthesis stages can be isolated and examined. In this set up, **HIBA3** demonstrates maximal AtACX4 activity, **CAN4** (with and without pSC101_SoxR_AcrR) can be used to compare ECH from **HIBA3** against AATm4, and **CAN5** can show AATm4 activity isolated from AtACX4. Additionally, assaying **CAN5** without adding any butanol will essentially 'remove' AATm4, thereby giving a representation of flux through BCKD. This can be achieved by quantifying the IBA produced due to endogenous thioesterase activity in *E. coli*.

5.2.2 Formation of 3-HIBA from CAN strain background

I initially assayed **HIBA-3** for 3-HIBA and IBA formation. **HIBA-3** was grown overnight in 10 mL LUND medium with 1% glycerol at 37°C and 250 rpm. Cells were then sub-cultured to OD_{600} 0.1 in 100 mL LUND with 1% glycerol in 500 mL shake flasks. These cells grew to OD_{600} 1.8 - 2.1, and were spun down and re-suspended to OD_{600} 15 in BT medium with 0.2% glycerol, 30.0 g L^{-1} (217 mM) 2-KIV and 5.00 mM butanol in a sealed 250 mL flask. **HIBA-3** consumed all but 0.58% of the 2-KIV added to the medium. Over 48 hours 16.5 g L^{-1} (159 mM) 3-HIBA was formed by **HIBA-3**, demonstrating a productivity of 0.34 g L^{-1} h^{-1} (Figure 5-5). 5.00 g L^{-1} (56.6 mM) of IBA

was also formed by **HIBA-3**. Thus, the yield of 3-HIBA product was 61.2%. 5.00 mM butanol was added to the BT medium at the start of the experiment as a control for a strain without AAT. After 48 hours, all of this butanol remained in the medium.

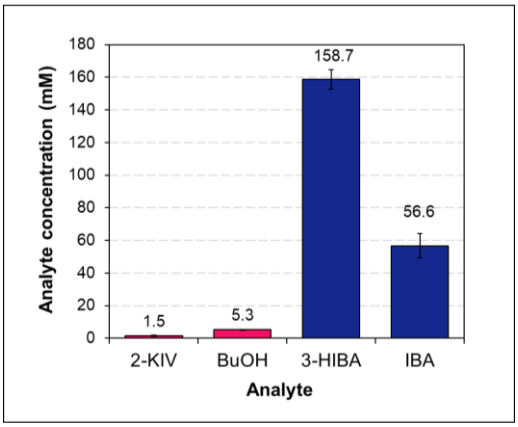


Figure 5-5: Product analysis of 3-HIBA **strain.**

HIBA-3 was provided with 217 mM 2-KIV to induce production of 3-HIBA, in biotransformation conditions at an OD_{600} of 15. Samples were collected at 48 hours after re-suspension and RP-HPLC used to determine ■ = remaining concentration of substrates, ■ = concentration of products formed by **HIBA-3**. 2-KIV = 2-ketoisovalerate, BuOH = 1-butanol, 3-HIBA = 3-hydroxyisobutyrate, IBA = isobutyric acid.

5.2.3 Formation of BMA from CAN-4 strain

CAN-4 was assayed for BMA formation under the same conditions as **HIBA-3**. When 5 mM butanol was added to the BT medium, 0.006 mM BMA was formed (Figure 5-6A). As with **HIBA-3**, IBA was formed in high quantities, with 63.3 mM detected in 48 hour samples. 24% of the butanol remained after48 hours, although only 2.66 mM total ester was detected by GC-MS. This accounts for just 70% of the butanol 'consumed' during the biotransformation. Increasing butanol concentration to 15.0 mM resulted in a negligible change in BMA concentration; 0.004 mM was detected, and IBA concentration remained high at 55.9 mM (Figure 5-6B). 75% of the butanol remained in the medium at the end of this experiment, and again only 70% of the butanol consumed during the experiment could be mass balanced with the total ester concentration detected, which in this case was 2.56 mM. 57% of the 2-KIV feed was retained in the media. BIB was the most abundant ester product, with concentrations

ranging from 1.67-2.20 mM, whilst a new ester, butyl propionate (BPI) was detected at concentrations of 0.22-0.86 mM.

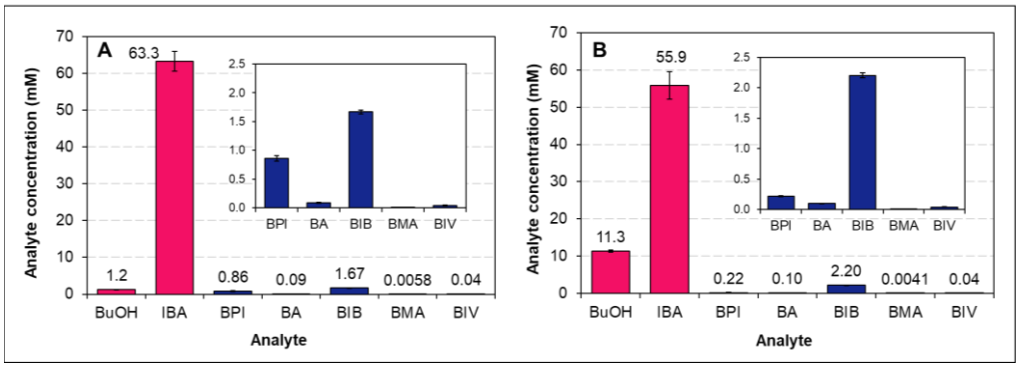


Figure 5-6: Product analysis of CAN-4 **strain.**

CAN-4 was provided with 217 mM 2-KIV to induce production of BMA, in biotransformation conditions at an OD_{600} of 15. Samples were collected at 48 hours after re-suspension. RP-HPLC and GC-MS was used to determine the concentrations of ■ = substrates added to media and ■ = products of **CAN-4** at both A: 5 mM BuOH and B: 15 mM BuOH. 2-KIV = 2-ketoisovalerate, BuOH = 1-butanol, IBA = isobutyric acid, BPI = butyl propionate, BA = butyl acetate, BIB = butyl isobutyrate, BIV = butyl isovalerate.

5.2.4 Formation of IBA from CAN-5

Biotransformation using **CAN-5** without butanol should prevent AATm4 activity, thus allowing only IB-CoA production by BCKD IB-CoA will then be solely available for thioesterase activity producing IBA (Figure 5-7A). Subsequently adding butanol to the media should then introduce AATm4 activity and allow use to observe whether AATm4 activity is competitive against native thioesterase activity in *E. coli.* Therefore, we assayed **CAN-5** under the same biotransformation conditions as previously used for **CAN-4** and **HIBA-3**. Under these conditions, 69.3 mM IBA was formed when no butanol was added to the media, while 52% of the initial 30 g L^{-1} 2-KIV remained in culture. As expected, no BIB was detected in samples. Repeating this experiment with 5.00 mM butanol, thereby reintroducing AATm4 activity, resulted in a reduction of IBA concentration to 41.8 mM (Figure 5-7B). Although only 1.95 mM BIB and 1.13 mM BPI were detected, with trace concentrations of BA and BIV. The reduction in IBA concentration by 27.5 mM is not balanced by a similar scale increase in ester formation.

Assaying **CAN-5** with 15.0 mM butanol resulted in a BIB concentration of 2.80 mM, and a significant decrease in IBA concentration to 6.88 mM. As with the **CAN-4** biotransformations, a reduction in butanol concentration from 15.0 mM to 5.00 mM does not correspond to a likewise reduction in the concentrations of ester products in the extracellular samples. Indeed, 4.98 mM butanol used and 3.27 mM ester formed when 5.00 mM was added, whilst 8.28 mM butanol was consumed when 15.0 mM was added, while only 3.15 mM ester was detected.

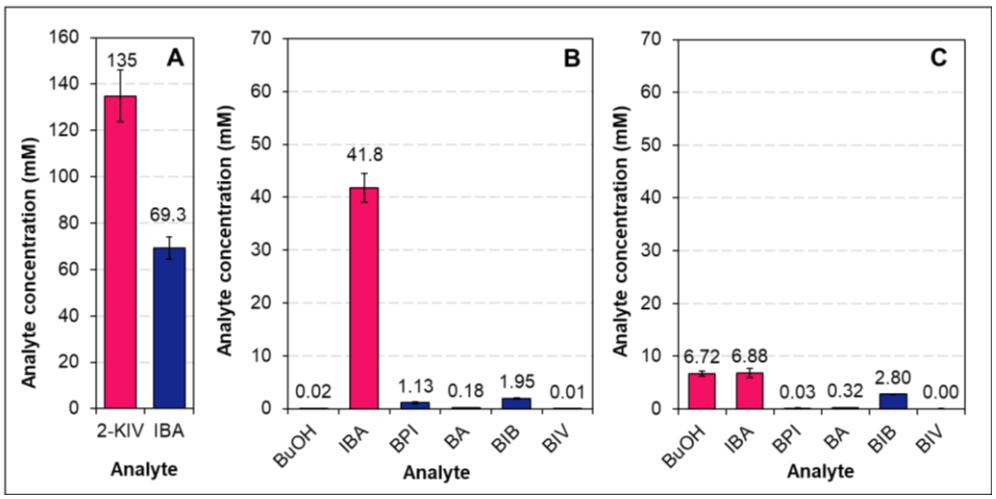


Figure 5-7: Product analysis of CAN-5 strain.

CAN-5 was provided with 217 mM 2-KIV to induce production of IBA and BIB in biotransformation conditions at an OD_{600} of 15. Samples were collected at 48 hours after re-suspension. RP-HPLC and GC-MS was used to determine the concentrations of ■ = substrates added to media and ■ = products of **CAN-5** at both A: 0 mM BuOH B: 5 mM BuOH and C: 15 mM BuOH. 2-KIV = 2-ketoisovalerate, BuOH = 1-butanol, 3-HIBA = 3-hydroxyisobutyrate, IBA = isobutyric acid.

5.2.5 Effect of expressing SoxR and AcrR mutants using CAN-4R and CAN-5R

Prior to this project, LM developed a plasmid, pSC101_AcrR_SoxR, expressing mutant AcrR and SoxR. These mutant proteins confer resistance to extracellular BMA up to 20% (v/v). pSC101_AcrR_SoxR was co-transformed into *E. coli* BW25113 Δ*ldhA* Δ*infA*:: KanR along with pCAN-4 and pCAN-5, to make the strains **CAN-4R** or **CAN-5R**, respectively.**CAN-4R** and **CAN-5R** were assayed in the biotransformation experimental set-up to assess whether introducing a plasmid conferring extracellular BMA resistance could favourably alter the formation of products in both the BMA and the BIB forming strains.

Analysis of **CAN-4R** showed a 4.3-fold increase in BMA detected as compared to **CAN-4**, at 0.026 mM (Figure 5-8A). BA was also detected in a higher proportion in **CAN-4R** compared to the non-resistant strains, although in both **CAN-4R** and **CAN-5R** BIB was the major ester product, at 1.97 mM and 3.14 mM, respectively. In **CAN-4R**, this corresponds in a change of BMA:BIB ratio from 1:28 to 1:76. There was no substantial change in net productivity from either strain compared to their non-resistant counterparts. The total ester concentration detected by GC-MS was 2.42 mM for **CAN-4R** compared to 2.66 mM for **CAN-4**, while ester concentration increased from 3.27 mM using **CAN-5** to 3.49 mM in **CAN-5R**. IBA formation remained broadly similar to the non-resistant strains at 59 mM in **CAN-4R**. In contrast, **CAN-5R** IBA formation varied widely between replicates, and was generally lower than for all other strains except **CAN-5**, at 10.4 mM (Figure 5-8B).

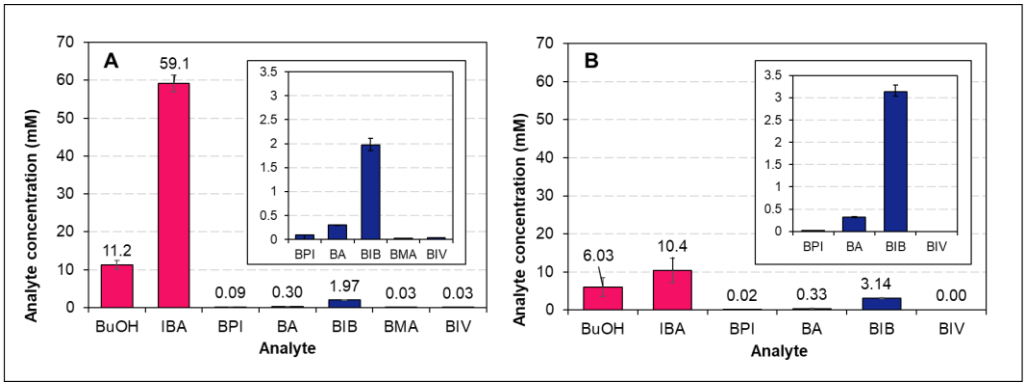


Figure 5-8: Product analysis of CAN-4R and CAN-5R strains.

CAN-4R and **CAN-5R** were provided with 217 mM 2-KIV to induce production of IBA and BIB in biotransformation conditions at an OD_{600} of 15. Samples were collected at 48 hours after re-suspension and addition of 15.0 mM butanol. RP-HPLC and GC-MS was used to determine the concentrations of ▮ = substrates added to media and ▮ = products. 2-KIV = 2-ketoisovalerate, BuOH = 1-butanol, 3-HIBA = 3-hydroxyisobutyrate, IBA = isobutyric acid.

We also carried out a growth assay using **CAN-4R** and **CAN-5R** to investigate the effect of pSC101_AcrR_SoxR on the growth of ester forming *E. coli.* **CAN-4**, **CAN-5,** **CAN-4R** and **CAN-5R** were cultured in LUND overnight at 37°C. These cultures were re-suspended to OD_{600} 0.1 in a 96-well plate in triplicate. The OD_{600} in the wells was monitored every 5 min over 60 hours from sub-culturing. **CAN-4R** exhibited little to no lag phase from OD_{600} 0.1, while **CAN-4** only exited lag phase after 24 hours (Figure

5-9A). Both strains reached similar maximum OD_{600}'s. **CAN-5** and **CAN-5R** demonstrated similar growth patterns, with a marginally longer lag phase, and slower exponential phase in **CAN-5** when compared to **CAN-5R** (Figure 5-9B).

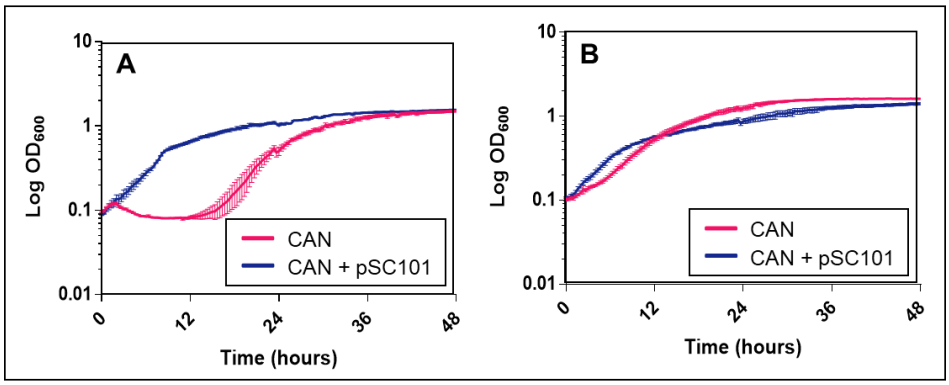
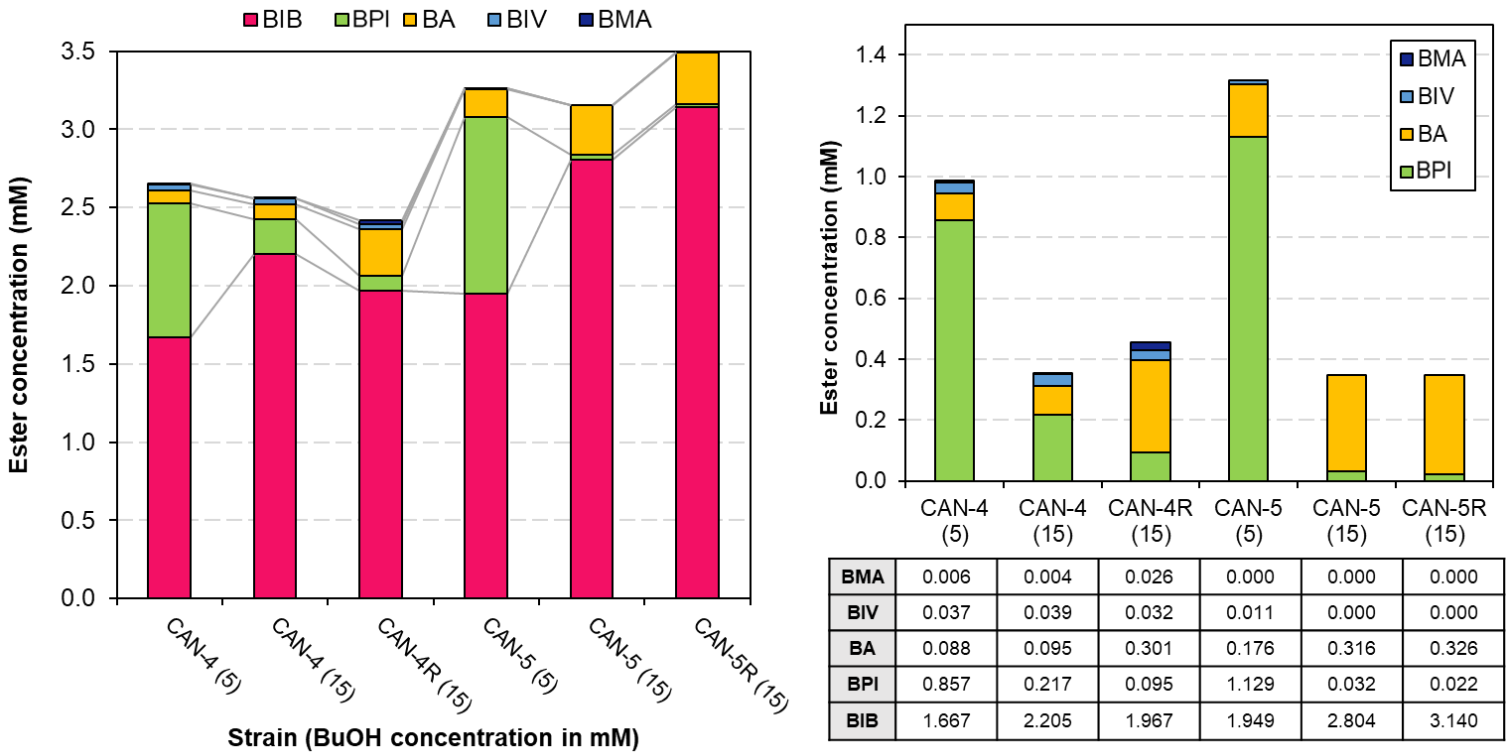


Figure 5-9: Growth curves for BMA and non-BMA resistant strains

Growth of **CAN-4**, **CAN-5**, **CAN-4R** and **CAN-5R** was monitored over a period of 60 hours from a starting point at OD_{600} 0.1. A: CAN-4, B: CAN-5. ▇ = CAN plasmid only, ▇ =.CAN plasmid and pSC101_AcrR_SoxR (LM). Readings taken using a Bioscreen C. Error bars represent the standard error calculated from a triplicate dataset. No WT measurements taken.

5.2.6 Ester specificity and formation in CAN strains

BIB was the most abundant ester product from all CAN strains utilising AAT. The concentration varied between 1.67 mM from **CAN-4** with 5 mM butanol, to 3.14 mM from **CAN-5R** with 15 mM butanol (Figure 5-10). In **CAN-4** and **CAN-5** with 5 mM butanol, BPI was the second most abundant ester, at 0.86 mM and 1.13 mM, respectively. A lower proportion of BPI was produced in strains grown with 15 mM butanol, and BPI was present in generally higher proportions in strains with pCAN-4. BA was produced by all strains expressing AATm4, with the highest concentration produced by the strains expressing mutated AcrR and SoxR; **CAN-4R** and **CAN-5R**. Approximately 0.03 mM BIV was produced by all strains expressing pCAN-4, but was detected in only one of the strains expressing pCAN-5, which was **CAN-5** grown with 5 mM butanol. In every pCAN-4 expressing strain, BMA was the least abundant detectable ester product.



	CAN-4 (5)	CAN-4 (15)	CAN-4R (15)	CAN-5 (5)	CAN-5 (15)	CAN-5R (15)
BMA	0.006	0.004	0.026	0.000	0.000	0.000
BIV	0.037	0.039	0.032	0.011	0.000	0.000
BA	0.088	0.095	0.301	0.176	0.316	0.326
BPI	0.857	0.217	0.095	1.129	0.032	0.022
BIB	1.667	2.205	1.967	1.949	2.804	3.140

Figure 5-10: Ester specificity of CAN strains.

Left: Concentrations of all ester products from CAN strains. Right: Concentrations of ester products excluding BIB from all CAN strain. Values in (brackets) on horizontal axis refer to butanol concentration in mM added to biotransformation media, where (5) = 5 mM butanol and (15) = 15 mM butanol. BIB = butyl isobutyrate, BPI = butyl propionate, BA = butyl acetate, BIV, = butyl isovalerate, BMA = butyl methacrylate.

5.2.7 Expression and purification of AtACX4 and MBP-AtACX4

The high BIB production titres compared to BMA, BA, BIV and BPI (Figure 5-10) from **CAN-4** possibly occur due to higher IB-CoA concentrations present in producer cells, as compared to M-CoA, acetyl-CoA and propionyl-CoA. Although not verified directly, this can be hypothesised because AAT enzymes have broad substrate specificity, which can be influenced by cellular substrate availability [162]. If BCKD produces sufficient quantities of IB-CoA, as must occur to produce the 159 mM 3-HIBA titres reported in 5.2.2, then IB-CoA needs to be rapidly removed to prevent significant BIB production, and diversion of AATm4 activity.

AtACX4 in combination with ECH and HCH can achieve high rate flux to 3-HIBA from **HIBA-3**. The only difference between the **HIBA-3** and **CAN-4** strains was the inclusion of AATm4 in place of ECH and HCH in the latter. The change in product composition can therefore be attributed to the introduction of AATm4. It is likely that AATm4 doesn't remove M-CoA as rapidly as ECH, as AATs generally possess higher K_M values than ECH. Therefore the hypothesis was that an AATm4 derived accumulation M-CoA accumulation resulted in AtACX4 product inhibition, thus preventing otherwise rapid removal of IB-CoA by the oxidase. To test this hypothesis, I assembled four strains to assay AtACX4 *in vitro* (Figure 5-11A). I then used these strains to calculate the K_M for IB-CoA and also the K_i of AtACX4 with M-CoA

To prepare AtACX4 and MBP-AtACX4 expression strains, I constructed the following plasmids: pOX-2_ACX4, pOX-2_ACX4-HIS, pOX-2_MBP-ACX4 and pOX-2_MBP-ACX4-HIS (Figure 5-11A). Four AtACX4 fragments were amplified from pMAE-4 by PCR. These were AtACX4 with and without a C-terminal His-tag, and MBP-AtACX4, also with and without a His-tag. Each of the four DNA fragments contained flanking *NdeI* and *XhoI* sites. We gel purified the PCR amplified AtACX4 fragments, and digested them using *NdeI* and *XhoI.* I also linearised a pET20b+ vector using *NdeI* and *XhoI*. We ligated each of the *NdeI*/*XhoI* digested PCR products into the linearised pET20b+ vector backbone. The AtACX4, AtACX4-HIS, MBP-AtACX4 and MBP-

AtACX4-HIS expressing strains are known as **OX-ACX4**, **OX-ACX4(H)**, **OX-MBP-ACX4**, and **OX-MBP-ACX4(H)** respectively.

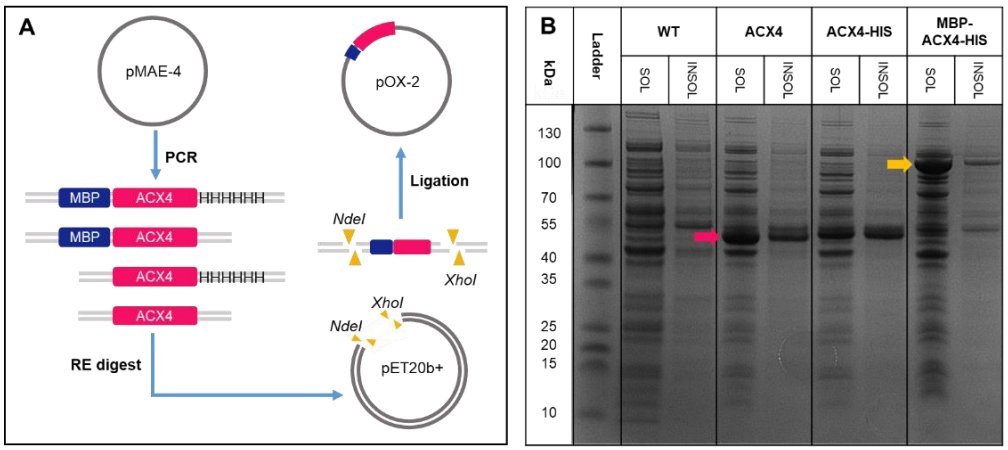


Figure 5-11: Cloning of ACX4 and MBP-ACX4 for protein expression and purification

A: Cloning of His-tagged and non-His-tagged ACX4 and MBP-ACX4 into a pET20b+ vector to produce pOX-2. B: Protein fractions from *E. coli* BL21(DE3) cells expressing either no plasmid (WT), ACX4, ACX4-HIS, MBP-ACX4 or MBP-ACX4-HIS. ▮ = AtACX4 band. ▮ = MBP-AtACX4 band. Cells were grown overnight at 37°C and 250 rpm and lysed with Bugbuster reagent. NuPAGE 4-12% Bis-Tris Gel ran at 150V for 50 min. Gel was stained using Coomassie Blue and the ladder used was PageRuler Pre-stained protein ladder.

I checked for expression of AtACX4 and MBP-AtACX4 from the newly constructed strains using SDS-PAGE to visualise both the soluble and insoluble protein fractions (Figure 5-11B). A 47 kDa band was visible in the soluble and insoluble fractions of **OX-ACX4** and **OX-ACX4(H)**, which was not present in the WT *E. coli* fractions. The solubility of AtACX4 with the addition of the His-tag did not change compared to the solubility of the protein fractions isolated from **OX-ACX4**. Approximately 40% of the total AtACX4 protein expressed from both **OX-ACX4** and **OX-ACX4(H)** was insoluble. A band corresponding to the 90 kDa MBP-ACX4 can be seen in the soluble fraction of **OX-MBP-ACX4(H)**, with a very faint band visible in the insoluble fraction. Only approximately 10% of the expressed MBP-AtACX4-HIS was insoluble. A faint protein band was also observed in both the soluble and insoluble fractions from **OX-MBP-ACX4(H)** which migrates the same distance as the AtACX4 band from **OX-ACX4** and **OX-ACX4(H)**.

His-tagged AtACX4 was expressed using *E. coli* BL21(DE3) pOX-2_ACX4-HIS. I grew **OX-ACX4(H)** to an OD_{600} 0.6-1.0 in 500 mL LB before inducing AtACX4 expression. At 15 hours after IPTG induction, the cells were lysed using sonication and then immediately purified AtACX4 using a Ni-NTA column on an AKTA FPLC. All our buffers during the purification steps contained 10 µM FAD. This prevented co-factor loss from the purified AtACX4, which would render it inactive. Fractions containing pure AtACX4 were identified (Figure 5-12, A11-D12) and pooled together. Buffer exchange was carried out by dialysis into a storage buffer containing 20 mM HEPES, 150 mM NaCl and 10 µM FAD at a pH of 7.0. Purified AtACX4 was either used directly for the kinetic assay, or stored it at -20°C until required.

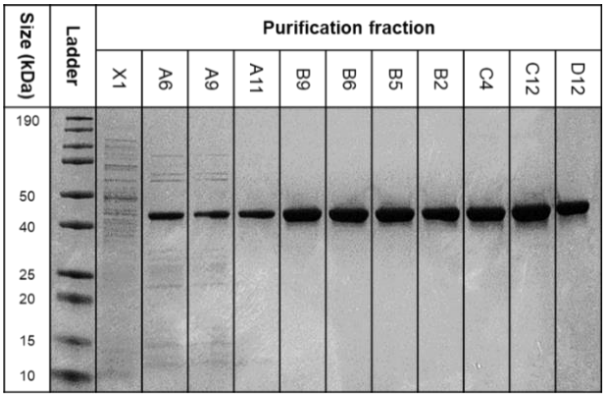


Figure 5-12: His-trap purification of AtACX4-HIS.

Fractions were collected after stepped gradient purification of AtACX4-HIS, expressed from *E. coli* BL21(DE3) pOX-2_AtACX4-HIS cells lysed by sonication. Protein visualised using a mini-PROTEAN® TGX Stain-Free™ Precast Gel, 200 V, 35 min. Stained with InstantBlue™ for 15 min. BenchMark™ Pre-stained protein ladder. Fractions A11-D12 (27 fractions) pooled and dialysed to collect purified AtACX4.

MBP-AtACX4-HIS was expressed by *E. coli* BL21(DE3) pOX-2_MBP-AtACX4-HIS cells. The cells were harvested, lysed, and the supernatant loaded onto the AKTA FPLC as for our AtACX4 purification. However, analysis of the collected fractions revealed two distinct protein bands (Figure 5-13A). These bands appeared to correspond to the approximate size of either the target MBP-AtACX4-HIS (~90 kDa) (Figure 5-13A, dark blue) or AtACX4-HIS separate from MBP (47 kDa) (Figure 5-13A, pink). Running Fraction B7 from the MBP-ACX4-HIS purification (Figure 5-13A) against a sample of purified ACX4 demonstrated that the smaller protein band in the

MBP-AtACX4-HIS purification was the same size as AtACX4 (Figure 5-13B). Repeated attempts to purify the MBP-AtACX4-HIS alone were unsuccessful, and all subsequent data refers to purified AtACX4-HIS only.

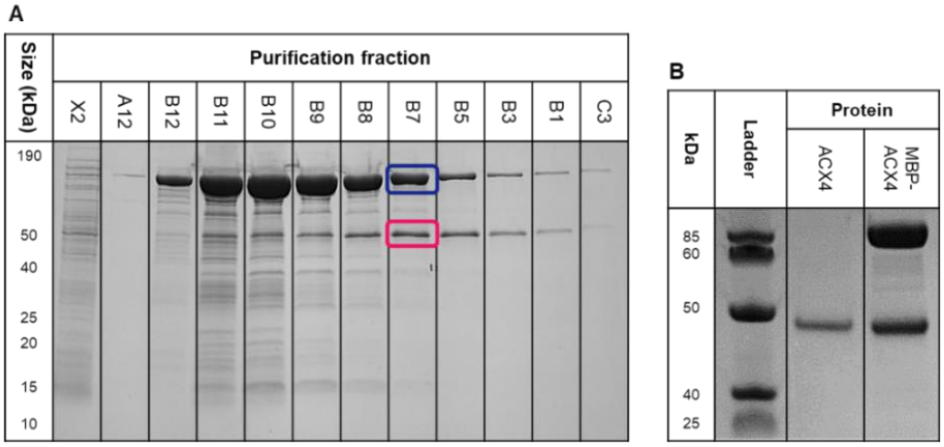


Figure 5-13: Partial purification of MBP-ACX4

A: Fractions from partial purification of MBP-ACX4 (linear gradient) from *E.coli* BL21(DE3) pOX-2_MBP-ACX4-HIS cells lysed by sonication. B: Comparison of purified fraction from MBP-AtACX4-HIS purification and AtACX4-HIS purification. Mini-PROTEAN® TGX Stain-Free™ Precast Gel, 200 V, 35 mins. Stained with InstantBlue™ for 15 minutes.

5.2.8 Kinetics of AtACX4 with isobutyryl-CoA and methacrylyl-CoA

I assayed AtACX4 activity using an S1 Clarke electrode to measure oxygen (O_2) concentration directly. When a voltage is applied to the system, the polarised Pt cathode reduces O_2 in the assay solution (Figure 5-14, dark blue). Electrons are provided *via* a KCl electrolyte solution (Figure 5-14, light blue) in contact with both the Pt cathode and an Ag anode (Figure 5-14, grey). Hydroxide ions are released as the reaction occurs, and the flow of electrons from the anode to the cathode produces a current proportional to the concentration of O_2 present in the solution. I prepared standard solutions of H_2O_2 and measured the oxygen concentration of these using the Oxygraph to ensure that H_2O_2 production by AtACX4 would not affect readings (Supplementary C.4:). For K_M determinations, the assay volume was 1 mL, containing 0.5 mg mL^{-1} AtACX4 in a pH 7.0 buffer with 20 mM HEPES, 150 mM NaCl and 10 µM FAD. Before commencing the assay, the solution was aerated to 225-230 nmol mL^{-1} O_2 and allowed to stabilize at this concentration.

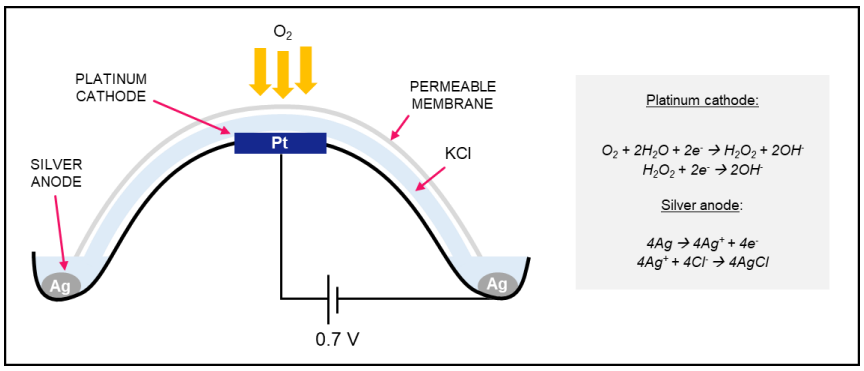
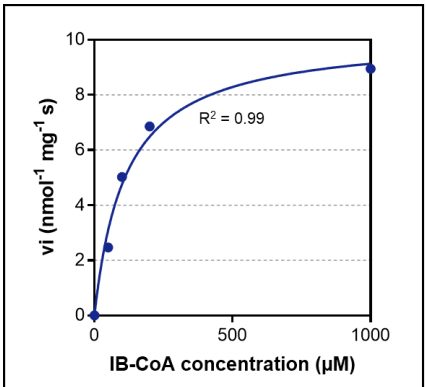


Figure 5-14: Cross-section of an Oxygraph oxygen electrode

A resin disc holds a platinum cathode (Pt) at the top and a silver anode ring (Ag) in a well around the bottom. The well is filled with KCl solution, with a paper 'wick' across the top of the disc forming a bridge. Voltage applied to the system polarises the Pt, which reduces O_2 from the assay solution.

Initially only the substrate IB-CoA was added to the assay solution. Once the substrate was added, oxygen consumption (ΔO_2) was measured over the course of 50 s, or until ΔO_2 had reached close to zero. A range of 0-1000 µM concentrations of IB-CoA was assayed with AtACX4, and the V_{max} and K_M calculated from the initial rate between 0-15 s, using GraphPad Prism 7.02 (Figure 5-15). The K_M for AtACX4 with IB-CoA was 0.14 mM, with a V_{max} of 10.2 nmol mg^{-1} s^{-1}.



Constant	Value	Units
K_M	0.14 ± 0.02	mM
V_{max}	10.2 ± 0.65	nmol mg^{-1} s^{-1}
k_{cat}	28.8	min^{-1}

Figure 5-15: K_M and V_{max} determination for His-tag purified AtACX4.

Purified AtACX4 was assayed by direct O_2 measurement using an Oxygraph oxygen electrode. 0.5 mg mL^{-1} protein at pH 7.0 and 30°C was mixed with between 0-1 mM IB-CoA. Oxygen concentration was monitored every 1 second for a period of 1 min. Data was obtained in triplicate. Michaelis-Menten curve calculated using non-linear regression on GraphPad Prism 7.

In order to calculate the K_i value for AtACX4 and M-CoA, M-CoA was synthesized and purified using RP-HPLC. Subsequent additional M-CoA was provided by MCUK. Increasing concentrations of M-CoA were added to the assay buffer, between 10-100 µM. The solution was allowed to mix and aerated as before, followed by the addition of IB-CoA to start the reaction. K_i was calculated using GraphPad Prism 7.02. Addition of increasing concentrations of M-CoA to the starting reaction resulted in noticeable decreases in the initial rates of reaction (Figure 5-16).

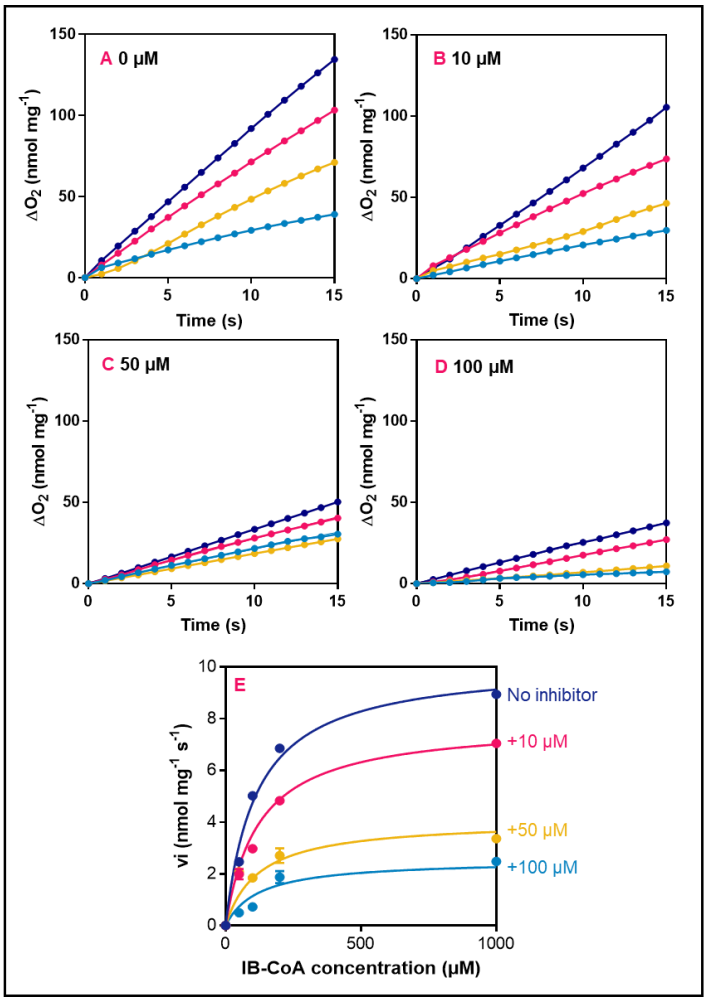


Figure 5-16: Inhibition of purified ACX4 by M-CoA.

Purified AtACX4 was assayed by direct O_2 measurement using an oxygen electrode. 0.5 mg mL^{-1} protein at pH 7.0 and 30°C was mixed with between 0-1 mM IB-CoA. Oxygen concentration was monitored every 1 second for a period of 1 min. Data was obtained in triplicate and mean values of the initial rate of reaction are plotted in A-D. E: Michaelis-Menten curve was calculated using non-linear regression on GraphPad Prism 7M-CoA concentrations shown in **bold**. ■ = 1000 µM IB-CoA, ■ = 200 µM IB-CoA, ■ = 100 µM IB-CoA, ■ = 50 µM IB-CoA.

The initial rate of reaction of AtACX4 with 1000 µM IB-CoA decreased from 8.9 ± 0.04 nmol mL^{-1} to 2.3 ± 0.01 nmol mL^{-1} with 100 µM M-CoA, a 75% reduction in activity. Between 0-100 µM M-CoA with 50 µM IB-CoA, from 2.5 ± 0.07 nmol mL^{-1} to 0.7 ± 0.01 nmol mL^{-1}, a reduction in activity of the same magnitude. Reduction in AtACX4 activity appeared to be 75% irrespective of the concentration of IB-CoA supplied.

The K_M for AtACX4 without M-CoA was determined as 0.14 ± 0.02 mM. The V_{max} was 10.2 ± 0.65 nmol mg^{-1} s^{-1} and the specific activity 9.89 ± 0.57 nmol mg^{-1} s^{-1}. This gives a turnover rate (k_{cat}) for AtACX4 of 28.8 s^{-1}. Literature values for the closest substrate match with other ACX enzymes have given K_M values of between 32-131 µM for butyryl-CoA [190, 191] and 6-92 µM for hexanoyl-CoA [190, 192, 193]. No data is published concerning the K_M of *A. thaliana* ACX4. Reported specific activities for ACX enzymes vary between 0.96 nmol mg^{-1} s^{-1} in human ACX A [194] up to 1.28 x 10^3 nmol mg^{-1} s^{-1} in *G. nicotianae* [190]. The *G. nicotianae* ACX also reportedly has a similar k_{cat} value to AtACX4, of 0.23 s^{-1} with butyryl-CoA [190].

Addition of M-CoA to the AtACX4 assay reaction mix visibly reduced the initial velocity of the reaction (Figure 5-16A-D). Almost no oxidation activity was observed when 100 µM M-CoA was added to an assay containing ≤ 100 µM IB-CoA. The initial velocity with 1 mM IB-CoA reduced by approximately half when just 50 µM M-CoA was added. Michaelis-Menten parameters were calculated for each concentration of M-CoA using non-linear regression on GraphPad 7.0 (Figure 5-16E). As M-CoA concentration increased, V_{max} decreased from 10.2 nmol mg^{-1} s^{-1} with no M-CoA to 3.0 nmol mg^{-1} s^{-1} with 100 µM M-CoA. No clear relationship could be determined between the concentration of M-CoA and the K_M. The decreasing V_{max} and non-linear K_M relationship indicated that M-CoA may inhibit AtACX4 in a mixed/non-competitive manner. Calculation of the K_i based on a non-competitive/mixed mode of inhibition gave a value of 32.8 µM M-CoA.

5.3 Discussion

5.3.1 AATm4 and AtACX4 activity during BMA biosynthesis

3-HIBA production from the biotransformations, at 159 mM, tallies well with the concentrations recorded by AY in previous experiments. This shows a 'best case' current scenario for product yield through AtACX4 to be 55% of the provided 2-KIV. In comparison BMA formed by **CAN-4R**, the highest BMA producer, reached a final titre of 0.026 mM, which reduces the product yield to 0.01%. The only change between the two strains was AATm4 replacing ECH and 3-HIBH. It is also known that BMA has a not insignificant degree of toxicity to *E. coli* although this level was not reached in the **CAN** producer cells, as previously discussed. As such, BMA toxicity is unlikely to cause low titres, particularly as cells maintain a stationary phase OD_{600} at the 48 hour time point where samples are taken. Additionally, no dramatic change in BMA formation was observed when resistance plasmid pSC101_AcrR_SoxR is added to strains **CAN-4R** and **CAN-5R**: Although a 4.3 fold increase in BMA formation was measured, this only corresponded to an actual increase of 0.02 mM BMA, and does not satisfactorily improve either the specificity or overall productivity of **CAN-4** or **CAN-4R**.

However, M-CoA toxicity can also be considered a significant factor, and the high flux observed from **HIBA-3** demonstrates that AtACX4 is capable of exhibiting sufficient activity for rapid M-CoA formation. As such, it would be expected that the BMA producing cells with a 0.01% yield of BMA would accumulate a potentially toxic intracellular M-CoA pool. In that event there would be a dramatic decrease in cell viability soon after introducing 2-KIV into the media, as AtACX4 provides a large supply of M-CoA. This effect was not observed, so either M-CoA was removed by another mechanism, or was not formed in high concentrations initially. In the first case, we would expect to see low cytoplasmic concentrations of M-CoA as it is removed by other processes. Use of metabolomics to identify any M-CoA derived

conjugates in the producer cells could in future be used to assess this hypothesis directly.

Experimental data suggests that the cytoplasmic concentration of M-CoA is 1.5 mM at the commencement of fermentation, which drops to 0.6 mM after 24 hours (Ingenza). The K_M for butanol with acetyl-CoA is 2.7 mM [138], meaning that intracellular M-CoA concentrations likely only lead to a maximum of 35% saturation of available AAT enzyme, under fermentation conditions. In this period, M-CoA may undergo Michael addition reactions with sulfhydryl groups in other metabolic products, reducing the pool available for AAT and potentially leading to a detrimental reduction in core intermediates such as oxidised glutathione [181]. In contrast, ECH has a K_M of approximately 20 µM, meaning that in **HIBA-3** a maximum of 100% of ECH active sites may be occupied by M-CoA [178, 195].

Another likely possibility resulting in low cytoplasmic M-CoA availability, AtACX4 activity limitation under certain conditions, in particular due to product inhibition. Previous data obtained by Ingenza has suggested that AtACX4 has a K_i value that lies well below the cytoplasmic M-CoA concentration found in producer cells, at around 500 µM. The data presented from this project indicates that the true K_i value is likely even lower, at 32.8 µM. If this is the case, then AtACX4 can only produce a maximum concentration of M-CoA which is 73 x less than that required to saturate half of the AAT active sites with M-CoA, at 2.7 mM. The resultant slow M-CoA removal, particularly when compared to the rapid activity of ECH, would facilitate accumulation of a higher intracellular M-CoA concentration, which in turn will limit AtACX4 activity in a negative feedback loop. In contrast, the high ECH activity can also be explained by the 32.8 µM K_i, as the ECH 20 µM K_M lies below this value. These differences between AATm4 and ECH likely account for the dramatic reduction in yield we see from 55% 3-HIBA from **HIBA-3**, to 0.01% BMA from **CAN-4R**.

Finally, the plasmid used by AY to produce high 3-HIBA titres differed from the backbone of the original pMAE-4 plasmid used in the production strain. In pMAE-4,

aatm4 and *acx4* were upstream of *bkdA1, bkdA2, bkdB*, and *ipdV* (BCKD), with all genes under the control of a single *araBAD* inducible promoter. The pHIBA-3 plasmid contains the BCKD genes upstream of *acx4*, *ech* and *hchA*, and is controlled by two constitutive Anderson series promoters: a medium strength J23104 and a strong J23119 promoter. The RBS was not varied from the original BMA pathway expression system. pCAN-4 and pCAN-5 were therefore constructed in a pHIBA-3 backbone to ascertain whether increasing pathway flux could be achieved simple by switching to a new expression system. However, despite the high 3-HIBA titres reached in the **HIBA-3** control (159 mM), a maximum of only 3.49 mM total ester was formed *via* AATm4 using **CAN-5R**, and just 2.56 mM total ester **CAN-4R**. Thus , balancing expression of the high activity BCKD step versus the AtACX4/AATm4 or AtACX4/ECH/3-HIBH using these promoters did not greatly affect BMA production. 3-HIBA titres produced by **HIBA-3** therefore appear to be the result of expressing the genes for ECH and 3-HIBH in place of AATm4, and are not due to changes introduced by the new vector backbone.

5.3.2 Competition for IB-CoA, M-CoA, and other intermediates

The capacity of AtACX4 and AATm4 to transform IB-CoA into M-CoA, and M-CoA to BMA, is a major production limitation. Diversion of intermediates into alternate metabolic pathways in *E. coli* also severely limits the selectivity and yield of BMA. This is primarily caused in two ways: endogenous enzyme activity, and broad AATm4 substrate range.

An endogenous cause of intermediate depletion occurs most clearly with IBA formation in all strains, even the 3-HIBA control. **HIBA-3** produced 5 mM IBA despite high flux through AtACX4, ECH and 3-HIBH. However, in **CAN-4** and **CAN-5** (with butanol), where the efficient removal of M-CoA and IB-CoA as in **HIBA-3** was not present, IBA concentrations ranging from 41-63 mM were formed, much closer to the IBA-only **CAN-5** (no butanol) control. This perhaps demonstrates that AATm4 activity is insufficient to remove acyl-CoA substrates before either M-CoA inhibits AtACX4

activity, cytoplasmic IB-CoA concentrations can increase, and native thioesterase activity can convert IB-CoA to IBA. In the case of the **CAN-5** strain grown both with and without butanol, we can compare the IBA produced with and without AATm4 active in the cell. **CAN-5** grown without butanol produced almost 70 mM IBA, while adding 5 mM butanol to activate AATm4 and produce BIB resulted a 2 mM reduction in IBA. Interestingly, adding 15 mM butanol to **CAN-5** resulted only 6.8 mM IBA, a change in production that cannot be attributed to increased removal of IB-CoA by AATm4, as an increase of only 1.2 mM BIB was seen compared to **CAN-5** grown on 5 mM butanol. IBA production in **CAN-4** strains, where the entire BMA pathway was active, were higher (55-63 mM) than those for **CAN-5**.

When ECH is drawing away M-CoA, AtACX4 activity appears to far outstrip thioesterase activity, which can be seen in the high ratio of 3-HIBA:IBA formed by the **HIBA-3** strain (2.78:1). Whilst AAT was utilising M-CoA the ratio of final product, in this case BMA, to IBA drops significantly ($7.27x10^{-5}$:1). In fact, minor differences were seen between the IBA formed in the **HIBA-3** strain versus **CAN-4** and **CAN-4R**.

In addition, multiple off-target esters are produced as a consequence of the broad substrate specificity of AATm4. This was observed in the first round of flask tests using pKIV and pBAD-MMA050 strains, although the ester composition was different in biotransformations. In these experiments, the major ester product from all strains at both butanol concentrations was BIB (Figure 5-10). The second most abundant product was BPI, although the proportion of this ester in samples decreased significantly with the increased butanol concentrations. Ingenza reported toxic concentrations of acetate esters in ongoing fermentations, while here I observed low titres of BA in most strains. Although there was an increase of the proportion of BA in samples taken from the **CAN-4R** strain. BIV was formed only in traces by the BIB strain in all circumstances. Introducing the BMA resistance plasmid into **CAN-5** strains appeared to increase the total extracellular ester concentrations, which was expected to some extent, as pSC101_AcrR_SoxR will affect the expression of the AcrAB-TolC

efflux pump. Most notably, BMA was no longer the major ester product. This was perhaps due to the lack of yeast extract added to BT medium, but may also have been caused by increased 2-KIV supply, constitutive expression, and a longer experimental run time.

5.3.3 Addressing product inhibition during BMA biosynthesis

The most likely explanation for low BMA production, as demonstrated through these experiments, is that it is currently caused by a combination of insufficient AATm4 activity, and the low K_i for M-CoA with AtACX4, resulting in product inhibition of the oxidase step. AATm4 activity could perhaps be improved if more M-CoA was available, but if AtACX4 is product inhibited by M-CoA then sufficient titres will not be reached. Conversely, AtACX4 activity might be improved by a more active AATm4, as efficient removal of M-CoA would prevent product inhibition becoming a limiting factor, as is observed in the 3-HIBA producing strains.

Product inhibition is not a simple challenge to approach, and is often solved by exhaustive screening of enzymes from alternative organisms until a feedback resistant enzyme is identified. In addition, evolution studies carried out by MCUK conclude that improving the AAT step often requires a trade-off between increasing activity and reducing specificity, particularly shifting product formation towards acetate esters. Two approaches to tackling low pathway flux are possible, focussing either on the oxidase or AAT catalysed steps individually, or synergistically improving flux through both terminal stages of BMA biosynthesis.

Chapter 6: Developing a semi-quantitative screen for BMA

6.1 Introduction

In order to synergistically improve BMA synthesis, I decided to modify an existing fluorescence assay for BMA, developed by Ingenza, to create a semi-quantitative screen from which a large number of BMA pathway variants could be assessed. This assay could then be used to screen a selection of novel ACX and AAT candidate enzymes, identified using bioinformatics, for improved production in a BMA producer strain.

From 5.3.1, the hypothesis was that one BMA formation bottleneck is linked to both AtACX4 and AATm4: So, enhancing AtACX4 activity should not reduce the AATm4 limitation significantly, and *vice versa*. Therefore, testing the novel ACXs and AATs in concert is more likely to provide a meaningful BMA productivity increase than assaying either of these enzymes alone. As of this point in the project, the only method on hand for assessing methacrylate formation from *E. coli* producer strains required flask biotransformations, followed by biphasic extraction of the ester products from supernatant for GC-MS analysis. This method reliably quantified the formation of BMA and the associated by-products from BMA biosynthesis, but was very low throughput. Indeed, each 'round' of testing was limited to a maximum of 5 strains, allowing for triplicates. Development of an alternative strain screening method was sought instead, which would facilitate a more expeditious assessment of a 650 strong library of BMA producer strains.

BMA is not a simple target to detect using screening. As previously discussed, BMA is a chemical with a relatively high volatility compared to those targeted in the majority of bioprocesses. This potentially makes downstream separation of BMA from fermentation broth a less complex process than if BMA remained in the liquid phase. Screening for biocatalytic formation of BMA in the liquid phase is hindered by the same property. Indeed, BMA has a solubility in water of only 800 mg L^{-1} at 25°C

(Hazardous Substances Data Bank, HSDB), making it difficult both to measure titres from liquid culture and to ensure sufficient interaction with a probe for screening. Because of this, a solid phase screening solution was considered by Ingenza prior to the commencement of this project. As part of this, a dimethyl sulfoxide (DMSO)-soluble diaryltetrazole probe was developed by MCUK. Under exposure to UV light at 302 nm, this diaryltetrazole probe binds to BMA with high specificity to produce a blue fluorescent product. This reaction, and the screening approach created by Ingenza, form the basis for our development of a BMA sensing screen with increased throughput.

6.1.1 Diaryltetrazole probes for fluorescent bioassays

When compared to other imaging techniques, such as dyes and light microscopy, fluorescent probes are a simpler solution often with much lower detection limits [196]. Although these probes have been used for decades, the development of chemo dosimeters that will react with a target analyte to produce fluorescence is more recent. Chemo dosimeters work based upon the protection-deprotection approach used frequently during organic synthesis to incorporate aldehyde, hydroxyl and amino groups in particular into a final product [197]. Variations in photophysical state when a deprotection is carried out in close proximity to a fluorescent dye may alter its emission profile, thus generating fluorescence. In the context of bioassay, these probes are altered in a bio-orthogonal, or click chemistry, reaction which can be take place under the physiological conditions of the host to produce a non-toxic product. Bio-orthogonal reactions also tend to take place with a high specificity and yield [198].

One well-known click chemistry reaction is the cycloaddition of alkenes with 2,5-diaryltetrazoles (DTZs). On exposure to UV light DTZs form a nucleophilic nitrile imine dipole. As alkenes are electron deficient, this dipole will conjugate with alkenes to produce a fluorescent pyrazoline. This occurs at a very high rate [198]. Pyrazoline is an ideal target for this type of purpose, as when quenched as a DTZ it does not fluoresce at all, and therefore any unreacted DTZ substrate need not be removed to

allow imaging. As with other biorthogonal reactions the mild conditions of the DTZ photoactivation make it a favourable choice for bioassay. However, the nitrile imine dipole formed as the reaction intermediate does encounter some problems when under physiological conditions. Namely, the dipole can react readily with the abundant thiols, amines, and water present in cells in a hydration reaction. When in an aqueous environment, such as the cytosol, the favourability of the cycloaddition reaction over hydration is much lower, and the activation time must be extended to increase pyrazoline formation. In more recent work, modifications to DTZs have been made to improve favourability of the cycloaddition, in particular by the introduction of electron withdrawing groups at the N-2-phenyl ring. MCUK developed a DTZ probe capable of binding to the methacrylate moiety on BMA (mDTZ) (Figure 6-1).

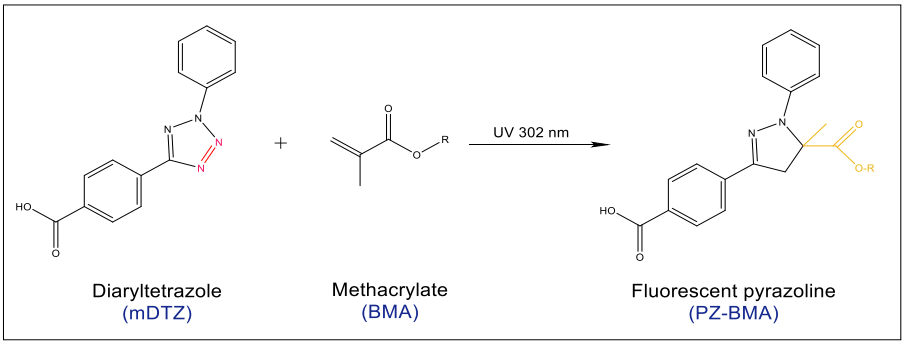


Figure 6-1: mDTZ binding to methacrylate for fluorescence emission.

Double bond shown in PINK breaks on exposure to UV 302 nm to form a nitrile imine dipole, which will readily conjugate with the nucleophilic alkene group present in methacrylate, to form the YELLOW region of the fluorescent pyrazoline product. Labels in BLUE represent components in the reaction between the MCUK diaryltetrazole probe (mDTZ) and butyl methacrylate (BMA) to form a pyrazoline-BMA conjugate (PZ-BMA).

Modifications made to this particular tetrazole probe include addition of a carboxylic acid functional group on the N-2-phenyl ring, which has been reported to boost cycloaddition efficiency. Triggering binding of mDTZ to BMA is a simple process: mDTZ and the methacrylate are mixed in a solution, and are exposed to UV light at 302 nm. This wavelength of light triggers the click 1,3 – dipolar cycloaddition reaction, to form a fluorescent pyrazoline-BMA conjugate (PZ-BMA). Fluorescent PZ-BMA emits a blue light under exposure to UV at 520 nm (Figure 6-1) and therefore the presence of BMA can be easily detected using conventional imaging equipment such

as a UV illuminator (MCUK, unpublished). In addition to the simple imaging technique, the mDTZ probe is selective for methacrylate when compared to MAA, 3-HIBA, BIB, IBA and 2-KIV, as well as LB only. MAA can be visualised instead if exposed to a different wavelength of UV light. This was demonstrated by Ingenza, who activated mDTZ in solutions mixed with the various reaction intermediates we expect to find during BMA biosynthesis (Figure 6-2).

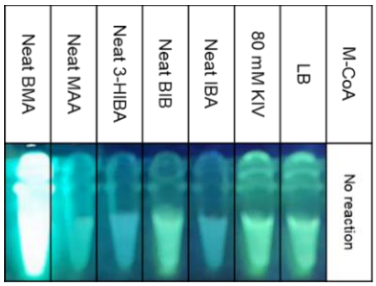


Figure 6-2: Diaryltetrazole probe with BMA reaction intermediates

Adapted from figure provided by Ingenza (Edinburgh, UK). BMA = butyl methacrylate, MAA = methacrylic acid, 3-HIBA = 3-hydroxyisobutyric acid, BIB = butyl isobutyrate, IBA = isobutyric acid, KIV = ketoisovalerate, LB = Luria-Bertani media, M-CoA = methacrylyl-CoA. Visualised using a UV illuminator at 520 nm.

At an emission wavelength of 520 nm, we can see that PZ-BMA produces significantly more fluorescent signal than any of the other tested intermediates and media components. Despite the methacrylyl moiety present on methacrylic acid little fluorescence, if any, is visible at 520 nm, as reported by MCUK. The specificity of mDTZ towards BMA, as well as its favourable reaction conditions, led to its subsequent use in colony screening experiments for BMA production, and to a lesser extent HMA biosynthesis.

6.1.2 A solid phase DTZ assay for extracellular BMA detection

Researchers at Ingenza used the mDTZ probe as the basis for a solid-phase assay measuring either BMA or HMA formation. The assay set-up they devised is fairly traditional: A liquid culture from overnight incubation of a mixture of strains is spread onto a 0.2 µm Nylon membrane on LB. This is grown for 24 hours, after which the membrane is transferred onto a solid phase assay (SPA) agarose plate containing

mDTZ, 2-KIV and 1-butanol. This is incubated for 30 min, after which the nylon membrane is removed and the mDTZ in the SPA plates activated by UV light at 302 nm. Colony 'footprints' are then visualised using a standard UV box, where fluorescence is assessed qualitatively and linked back to the colonies present on the Nylon membranes (Figure 6-3). The original colonies can be picked and assessed directly through either fermentation or biotransformation followed by GC-MS.

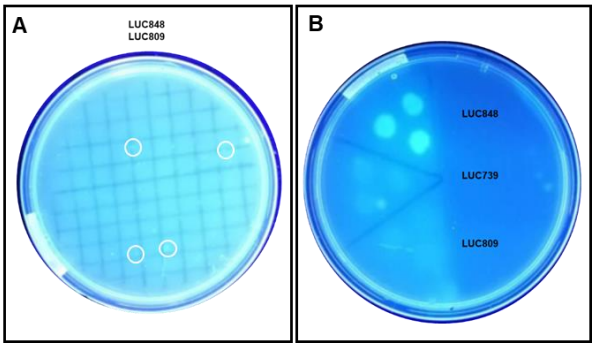


Figure 6-3: Visualisation of BMA producing strains using a UV box.

Strains used were **LUC0809**, **LUC0739**, and **LUC0848**. A: 30 µL a mixture of **LUC0848** and **LUC0809**. Brightest colonies highlighted by white circles. B: 5 µL culture was spotted onto membranes on LB in triplicate. Colonies grew at 37°C for 24 hours. Membranes incubated on SPA agarose for 30 min and mDTZ activated under UV light at 302 nm for 15 min.

One noticeable drawback with this approach is the qualitative nature of the data analysis stage. In the figure shown above (Figure 6-3, A), it can be seen that the fluorescing colonies are visible but difficult to pick out from the background interference caused produced by the SPA agarose. Although fluorescent colonies can be identified, it is time consuming and somewhat inaccurate to attempt to quantify these hits. This disadvantage can be mitigated (Figure 6-3, B) by spotting liquid culture to create larger 'footprints'. This approach however significantly decreases the throughput of the experiment, reducing the number of library entries that can be searched *per* plate. Also, although spotting the cultures creates a more concentrated signal, it still does not allow us to use this method to quantitatively compare PZ-BMA fluorescence from producer cells.

As such, I attempted to develop an improved mDTZ screening approach to quantitatively assess a large library of combinatorial BMA producer strains with a

higher throughput compared to flask biotransformations. To achieve this, I adapted the PZ-BMA fluorescence screen for use with an *in vivo* imaging system (IVIS) Spectrum, replacing the UV illuminator imaging step in the original screening method. The IVIS Spectrum is a high sensitivity *in vivo* imaging platform, primarily designed for real time fluorescent reporter tracking in living organisms [199]. In comparison to the UV illuminator, the IVIS Spectrum should display increased signal sensitivity, as well as enabling us to directly measure the radiance (in p s^{-1} cm^{-2} sr^{-1}) of individual PZ-BMA fluorescence footprints on SPA agarose plates.

Shorthand	Full name	BMA production	Author
MAE4	*E. coli* BW25113 pBAD-MMA050$_{corrected}$(mACX4)	Var	CT
CAN4	*E. coli* BW25113 Δ*infA* Δ*ldhA* :: KanR pCAN-4	High	CT
CAN5	*E. coli* BW25113 Δ*infA* Δ*ldhA* :: KanR pCAN-5	No production	CT
LUC0739	*E. coli* BW25113 Δ*tesB* Δ*yciA* pMMA050-mACX4	Low	Ingenza
LUC0809	*E. coli* BW25113 pCL1	No production	Ingenza
LUC0848	*E. coli* BW25113 Δ*tesB* Δ*yciA* pMMA050-mACX4	Low	Ingenza

Table 6-1: List of strains used in assay development chapter

6.2 Results

6.2.1 Validating DTZ screening method using UV illuminator

I received the UV illuminator screening protocol, training, and additional BMA producer control strains **LUC0739**, **LUC0809** and **LUC0848** directly from Ingenza. MCUK provided solid mDTZ which we stored at -80°C and prepared 0.12 M stock solutions in DMSO. Before beginning experiments using the IVIS Spectrum, I ran the conventional UV illuminator screen to check its functionality in our lab. For this purpose, I used the BMA producer control strains as listed in : **LUC0809**, which is *E. coli* BW25113 pCL1 and cannot produce BMA, **LUC0739**, which is *E. coli* BW25113 Δ*tesB* Δ*yciA* pMMA050 and **LUC0848**, or *E. coli* BW25113 Δ*tesB* Δ*yciA* pMMA050-mACX4, both of which produce low levels of BMA. I also included **CAN4** as the highest producer for screen development, and **CAN5** and **MAE4** from earlier shake flask cultures. **CAN5** serves as an alternative negative control to **LUC0809**, with the pCAN5 plasmid bearing more similarity to pCAN4 from **CAN4**.

Using these strains I carried out a biotransformation flask test to quantify BMA production from each, which would provide a benchmark for comparison against results from the fluorescent screen (Figure 6-4). **CAN4**, **CAN5**, **LUC0848**, **LUC0739** and **LUC0809** were grown overnight in LB, then sub-cultured to OD_{600} 0.1 in 200 mL LB supplemented with carbenicillin and 1-butanol. After the cells reached an OD_{600} >3 the cultures were harvested at 7,000 rpm for 2 min. To begin the biotransformation, the cells were re-suspended in 20 mL BT medium containing 2-KIV and 1-butanol to a final OD_{600} of 10, in triplicate. Results from this biotransformation were analysed using GC-MS, and as expected **CAN4** produced the highest concentration of BMA, followed by **LUC0848** and **LUC0739** which both produced lower BMA titres, with a marginally higher concentration detected from **LUC0848**. No BMA production was detected from either of the negative controls, **CAN5** and **LUC0809**.

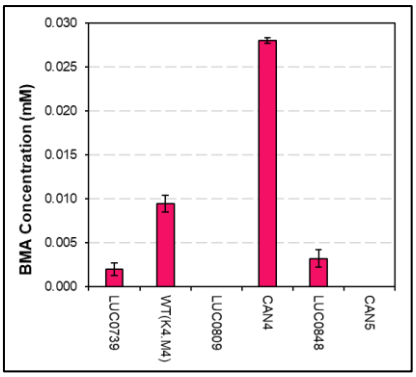


Figure 6-4: Flask determination of BMA production from screen control strains

LUC0739, **MAE4**, **LUC0809**, **CAN4**, **LUC0848** and **CAN5** were resuspended to an OD_{600} 10 in 20 mL BT media containing BuOH and Na-2-KIV. Biotransformation proceeded for 48 hours at 37°C and 250 rpm. Cells were harvested at 8,000 rpm and 10 mL supernatant extracted into EtAc for GC-MS analysis. Error bars calculated as standard error of triplicate data.

After confirming BMA formation titres from the controls, I moved on to checking the PZ-BMA fluorescence screen functionality. In the first instance, I reproduced the exact screening conditions specified by Ingenza, including using the UV illuminator visualisation method. To set this up, positive controls **LUC0739**, **LUC0848**, and **CAN4** were grown overnight in LB. These overnight cultures were diluted to 10^{-5} and I spread 30 µL of each onto nylon membranes. Under the original screening conditions, only a handful of colonies were faintly visible on the **LUC0848** and **CAN4** plates.

Nothing was visible from the **LUC0739** plate (Figure 6-5A). Although a small number of colonies were visible to the naked eye, these results could not be captured when photographs were taken. This highlights the challenges we expected in accurately matching images from the UV illuminator back to the original colonies on membranes. Nevertheless, as only a limited number of colonies could be visualised in this first experiment I repeated the process using altered conditions.

During the second PZ-BMA fluorescence assay I increased the 37°C incubation time to 48 hours on LB, before then transferring the colonies on membranes to the SPA agarose for BMA formation. As before, **LUC0739**, **LUC0848** and **CAN4** were plated as a 10^{-5} dilution, and additionally all three strains were also spotted in triplicate 10 µL spots onto membranes segmented into sixths. The membranes were incubated on SPA agarose for 30 min and activated mDTZ at 302 nm for 15 min before imaging using the UV illuminator method (Figure 6-5B).

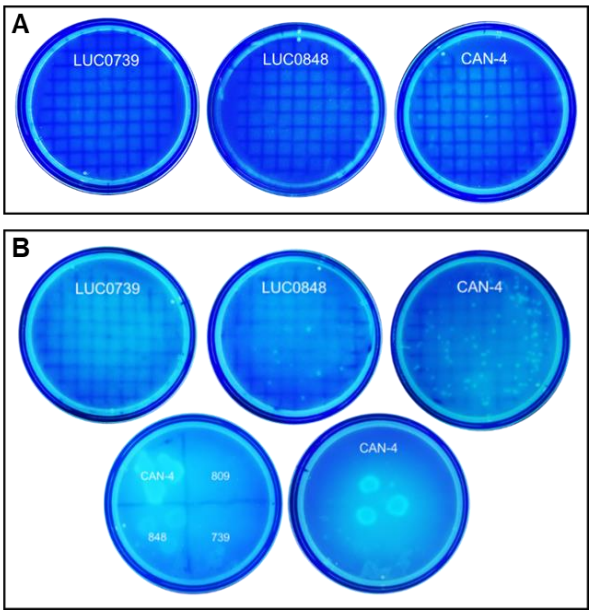


Figure 6-5: UV visualisation of BMA producer strains

Strains used were **LUC0809**, **LUC0739**, **LUC0848** and **CAN4**. A: Colonies grown for 24 hours on LB at 37°C. B: Colonies grown for 48 hours on LB at 37°C. Membranes were incubated on assay agar for 30 min, and mDTZ probe activated under UV light at 302 nm for 15 min. Visualised using a UV illuminator.

The resolution on the **CAN4** and **LUC0848** plates allowed observation of individual colonies (Figure 6-5B). Triplicate spots of **CAN4** and **LUC0848** likewise produced bright points of PZ-BMA fluorescence that could be observed using the UV illuminator. Although not quantifiable, the brightness of the **LUC0848** spots appeared similar to that of the **CAN4** spots. As might be expected, more fluorescent colonies were visible on the **CAN4** plate when compared to the **LUC0848** plate. As well as the higher BMA productivity of **CAN4**, this may be due to InfA expression in **CAN4** increasing pCAN-4 stability over the initial 48-hour incubation. In contrast, **LUC0848** and **LUC0739** are only maintained by antibiotic resistance markers. No fluorescence was observed from **LUC0739** when it was plated as a 10^{-5} dilution spread for individual colonies, but very faint spots were visible from the 10 µL spot plating (Figure 6-5B).

Following on from establishing the basic function of the PZ-BMA screen in our lab, I carried out further experiments to optimise and streamline the screening set-up. To this end we optimised several set-up conditions, these were: Length of time cells are grown on LB, length of time incubated with 2-KIV and 1-butanol on SPA agarose, distance from the UV bulb, length of time of UV exposure, and plate composition. As already explored in our initial tests, cells were grown on membranes on LB for either 24 or 48 hours. PZ-BMA fluorescence under UV was only observable from the 48 hour incubated cells. As a result, the incubation time was extended to 48 hours on LB before transfer to SPA agarose for all future screening.

I then investigated the effects of UV exposure and SPA agarose incubation time on PZ-BMA fluorescence in two ways. Firstly, **CAN4** was spotted onto membranes on LB, grew the cells for 48 hours at 37°C, and then transferred the membranes onto SPA plates. Half of the SPA plates with membranes were incubated at 37°C for 30 min, and the other half for 60 min. After incubation, the membranes were removed and the mDTZ in the plates were activated under a combination of conditions: At a long distance below the UV bulb (L) for a duration of either 15 min or 30 min, and at a short distance from the UV bulb (H) for a duration of either 15 min or 30 min (Figure

6-6A). Unfortunately, the image capture limitations make the results of this experiment difficult to observe from the images shown. Despite this it can be observed that a 60 min SPA agarose incubation time visibly increased PZ-BMA fluorescence when compared to the 30 min incubation. In contrast there was little visible difference between the fluorescence from 15 min and 30 min UV activated spots using this method. As before, there was no difference between the colonies activated close to the UV lamp (H), and colonies activated by placing in the microbiological safety cabinet (MSC) as normal (L) (Figure 6-6B).

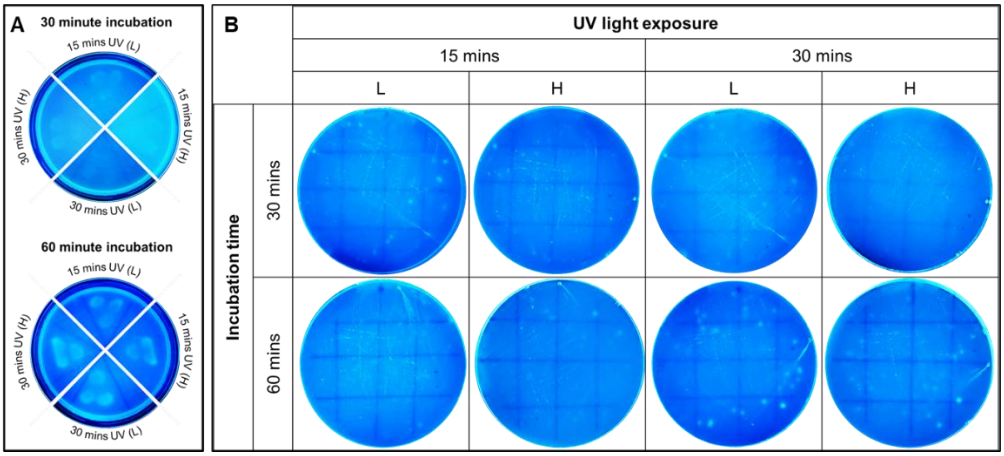


Figure 6-6: CAN4 comparison of alternate UV activation conditions

A: Spots of **CAN4** on membranes were grown for 48 hrs on LB. When membranes were transferred to SPA agarose they were incubated for either 30 or 60 min. UV activation time was varied within for colonies from each incubation, at either 15 or 30 min, as well as distance from the UV bulb. Floor = plates on floor of MSC, box = plates close to UV bulb. B: 10^{-5} diluted **CAN4** cultures were also spread on membranes on LB and tested using the same variables.

In addition to testing the assay conditions on **CAN4** spotted onto plates, I also set up gridded membranes with 10^{-5} dilutions of **CAN4**, **MAE4** and **LUC0739** in a 1:1:1 ratio mix. Incubation time, UV distance and UV activation time were varied as for the **CAN4** only experiment (Figure 6-6B). There was still no major difference in PZ-BMA fluorescence according to distance from the UV bulb. However, when inspecting the single colony plates it was possible to observe a small increase in PZ-BMA visibility where we had used a longer 60 min incubation time. I selected the 60 min SPA agarose incubation time, 30 min UV activation time, and MSC floor conditions for use in the PZ-BMA IVIS screen.

6.2.2 Optimising DTZ screening conditions for use with the *in vitro* imaging system

6.2.2.1 Imaging from BMA standard solutions

After optimising fluorescence from PZ-BMA formed from biosynthetic BMA using a UV illuminator, I also attempted to establish whether a linear relationship could be found between BMA concentration and fluorescence emission, or radiance, recorded by the IVIS Spectrum. I also sought to demonstrate that the IVIS Spectrum camera was able to detect PZ-BMA fluorescence at all. 0-10 mM BMA solutions were prepared in DMSO in 1 mL glass vials. To this we added mDTZ to a final concentration of 0.05 mM, and activated it under UV light at 302 nm for 30 min. The vials were imaged using the IVIS Spectrum, and PZ-BMA was detected from all solutions containing BMA. The radiance was measured as photons *per* second *per* the sum of the radiance from each pixel in our selected region of in the region of interest (ROI), accounting for the size of the ROI (p s^{-1} cm^{-2} sr^{-1}) (Figure 6-7).

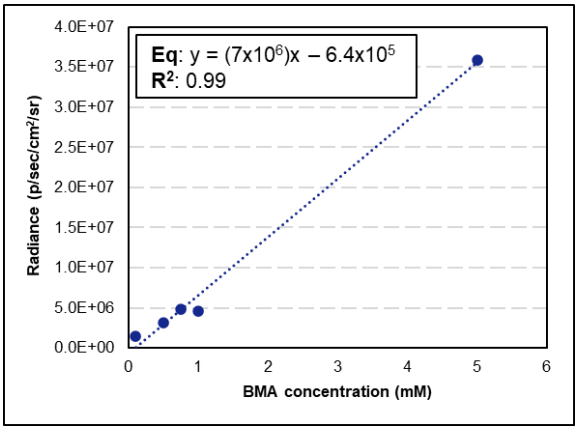


Figure 6-7: Radiance calibration with BMA in vials

0-10 mM BMA solutions were prepared in dimethyl sulfoxide in glass vials (10 mM result not shown because fluorescence was outside of IVIS range). 0.05 mM mDTZ was added and mixed, then activated under UV for 30 min to form pyrazoline-BMA. Radiance values obtained using IVIS Spectrum and Living Image Software. Version 4.3.1 (64-bit). BMA = butyl methacrylate.

The relationship between radiance and BMA concentration as measured by the IVIS Spectrum was somewhat linear, however readings were not significantly different at very low BMA concentrations (Figure 6-7). Despite this I spotted **CAN4**, **LUC0848**,

MAE4, **LUC0739** and **LUC0809** in triplicate onto membranes, carried out the PZ-BMA assay and detected the fluorescence using the IVIS Spectrum. I used the approximate BMA calibration from vials to estimate BMA production from the control strains, obtaining values of 0.4 mM BMA from **CAN4**, 0.2 mM from **LUC0848**, 0.1 mM from **MAE4**, 0.03 mM from **LUC0739**, after correcting for background fluorescence from the **LUC0809** negative control.

The ratios of BMA concentration produced by **CAN4**, **LUC0848**, **MAE4** and **LUC0739** somewhat tally with those predicted from the IVIS vial calibration. However, the mM values are much higher than the BMA titres recorded from flask experiments (Figure 5-6). The radiance of the 10 mM BMA vial was out of the range of the IVIS Spectrum and saturated readings, making it not possible to detect PZ-BMA in any of the vials of lower concentrations. I also spotted 10 µL of 0-10 mM BMA solutions onto membranes on SPA agarose to simulate the BMA production from cells. No PZ-BMA fluorescence was detectable using this method despite several attempts.

The linearity of the relationship between BMA concentration and radiance appears stronger over a larger range. Error in repeated measurements reduces the reliability of readings below 0.5 mM BMA. Variations in mM BMA concentration between repeated rounds of biotransformations causes the same issues of repeatability. Therefore, the quantitative BMA calibration did not seem powerful enough for reliable colony screening. I decided to instead gauge performance as compared to the best BMA producer, **CAN4**, and the worst, **CAN5**. In this way we can calculate the radiance of combinatorial library entries as a value relative to the fluorescence of **CAN4**.

6.2.2.2 Proof-of-principle bio-produced BMA detection using the IVIS Spectrum

The new assay protocol for the IVIS Spectrum operated in an analogous way to the original UV illuminator method (Figure 6-8). 0.2 µm Nylon membranes were gridded and sterilised by UV exposure for 20 min, on both sides. The sterile membranes were then placed on LB agar plates supplemented with carbenicillin and 1-butanol, onto

which we spotted or spread cells from overnight cultures (Figure 6-8A). Colonies grew on the membranes on LB for 48 hours at 37°C (Figure 6-8B). After incubation, the membranes were transferred to SPA agarose plates using sterile tweezers and incubated at 37°C for 60 min (Figure 6-8C). After 60 min, the membranes with cells were removed from the SPA agarose back to LB plates. Subsequently the mDTZ in the SPA agarose plates was activated under UV light at 302 nm for 30 min, catalysing formation of PZ-BMA if BMA is present (Figure 6-8D). PZ-BMA fluorescence on the SPA agarose is quantified using the IVIS Spectrum at 465/520 nm ex/em, with a 30 second exposure time, immediately after UV activation (Figure 6-8E). Once radiance data was captured using the IVIS Spectrum, we highlighted PZ-BMA emitting spots, or colonies, to obtain values for the average and maximum radiance.

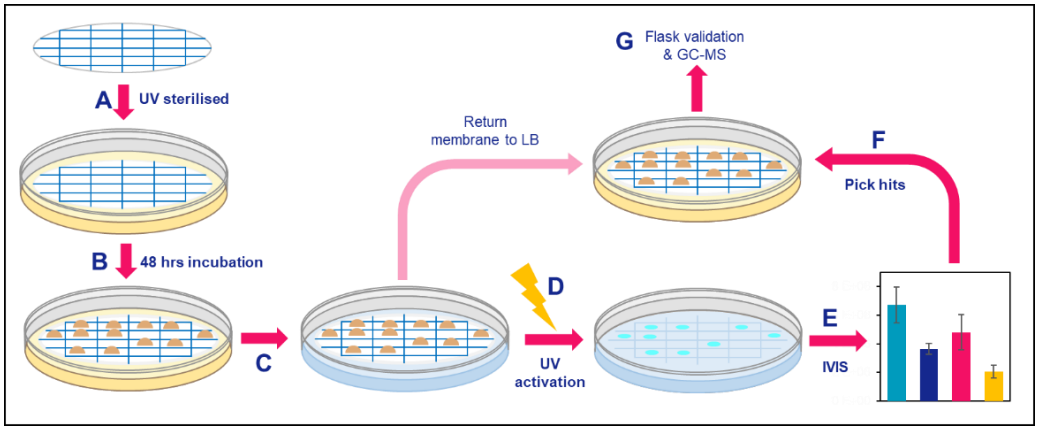


Figure 6-8: Medium-throughput screening method

A: 0.2 µm Nylon membranes are gridded and sterilised under UV light, then placed onto LB plates. B = Membranes on LB are incubated at 37°C for 48 hours. C: Membranes are transferred to SPA agarose containing mDTZ, 2-KIV and BuOH. D: After 1 hour, membrane is removed and probe is activated under UV to conjugate mDTZ with BMA. E: PZ-BMA fluorescence is quantified using the IVIS. F: Colonies with highest radiance are selected and picked from Nylon membrane.

During the ACX and AAT library screening stages, we used the radiance data to select 'hits' from the SPA agarose plates. These were colonies with a higher (%) PZ-BMA radiance signal when compared to the **CAN4** positive controls included on each plate (Figure 6-8F). Hits identified from the IVIS Spectrum data could then be traced back to the original colonies on membranes, which were stored at 4°C whilst the radiance data was collected and analysed. Picked hits from the PZ-BMA screen were

then validated in flask biotransformations, with BMA production, as well as BA and BIB concentrations, measured using GC-MS analysis (Figure 6-8G).

As an initial proof-of-principle for this approach, I spotted 3 dots of 10 µL for each of the control strains **CAN4**, **MAE4**, **LUC0739**, **LUC0809** and **LUC0848** onto sterile membranes segmented into sixths. This setup was repeated across three identical plates. After 48 hours at 37°C, membranes were transferred to SPA agarose and incubated for a further 30 min to allow BMA biosynthesis. membranes were removed and the mDTZ activated under UV light for 15 min. After 15 min the plates were immediately read using the IVIS Spectrum. Radiance was measured using an exposure time of 10 s at 465/520 ex/em. Signal was visible at all locations where **CAN4**, **MAE4 LUC0739**, **LUC809** and **LUC0848** had been plated (Figure 6-9).

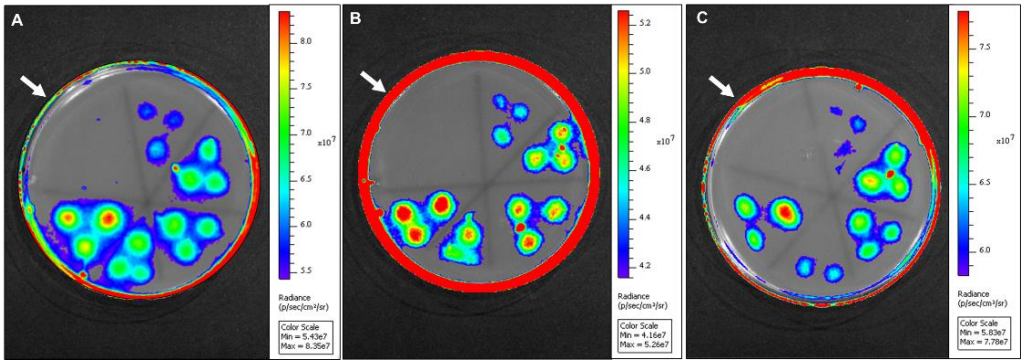


Figure 6-9: Test screen for BMA fluorescence detection using the IVIS Spectrum

Strains are as follows (clockwise from arrows): Blank, **LUC0809**, **LUC0739**, **LUC0848**, **MAE4** and **CAN4**. 10 µL culture was spotted onto membranes on LB in triplicate. Colonies grew at 37°C for 48 hours. Membranes incubated on SPA agarose for 1 hour and mDTZ activated under UV light for 30 min. Maximum radiance calculated using the IVIS and Living Image software version 4.3.1 (64-bit). Fluorescence measured at 465/520 ex/em with an exposure time of 30 seconds.

From this initial screen it was clear that the IVIS Spectrum detects a background level of signal wherever a cell footprint is present on the SPA agarose, regardless of whether that colony was from a BMA producer strain. This was perhaps due to very slight mDTZ interaction with off-target metabolites, as is visible in Figure 6-2. As PZ-BMA signal is much brighter than the off-target signal this should not affect our ability to distinguish colonies on the basis of BMA formation. As expected, **CAN4** radiance was significantly brighter than for **MAE4**, **LUC0739** and **LUC0848**. The **LUC0848** and

LUC739 radiance appeared broadly similar, with some increase from **LUC0848** on one of the replicate plates (Figure 6-9B). The radiance signal intensity, at least visually, appeared to vary between individual plates but not largely between triplicate spots on the same plate. This is expected and is the same effect seen in the BMA calibrations and flask biotransformations. **CAN4** is an exception to this, where the PZ-BMA radiance showed increased variance within the same SPA agarose plate (Figure 6-9A, C).

I quantified the radiance from each spot measured in . First, average radiance was analysed (Figure 6-10A). As indicated by the IVIS Spectrum image data, there was significant variation when the same strain was compared between replicated SPA agarose plates, but less variation between spots of the same strain on a single plate. The average radiance decreased almost twofold from plate A to plate B for **CAN4** spots from 6.1 x 10^6 p s^{-1} cm^{-2} sr^{-1} to 3.3 x 10^6 p s^{-1} cm^{-2} sr^{-1}. This was not consistent for all strains. For plate C the average radiance was higher than the equivalent value on plate B for all readings, but produced lower average radiance values for **MAE4** and **CAN4** compared to plate A, while for **LUC0848** the average radiance on plate C was lower than on plate A. When readings for all plates were averaged together, this equated to a large amount of error that read **LUC0809**, a non-producer of BMA, with no significant difference in signal from any of the BMA producer controls.

Measuring maximum radiance produced a similar effect, although the variation between plates A-C was more consistent (Figure 6-10B). Maximum radiance readings for all strains from plate C were lower than for plate A, although there was still more variation for **CAN4** and **MAE4** spots. Plate B read the lowest maximum radiance values for all strains. When data from all plates was combined, no significant difference between **CAN4**, **MAE4, LUC0809** and **LUC0848** could be observed.

I decided to move forward with measuring maximum radiance, as this may be less affected by the background fluorescence signal we appear to get from binding of mDTZ to an unknown molecule in the footprint from all *E. coli* grown on the

membranes (Figure 6-10). Inconsistent BMA formation from **MAE4** caused problems during this experiment. This unpredictability was perhaps due to leaky expression using of the genes expressed from pMAE-4, there is a lot of variation in the both the radiance and the concentration of BMA recorded from **MAE4** in flask biotransformations. Based on the reduction in error using maximum as opposed to average radiance, I decided to measure maximum radiance for all subsequent IVIS Spectrum analysis.

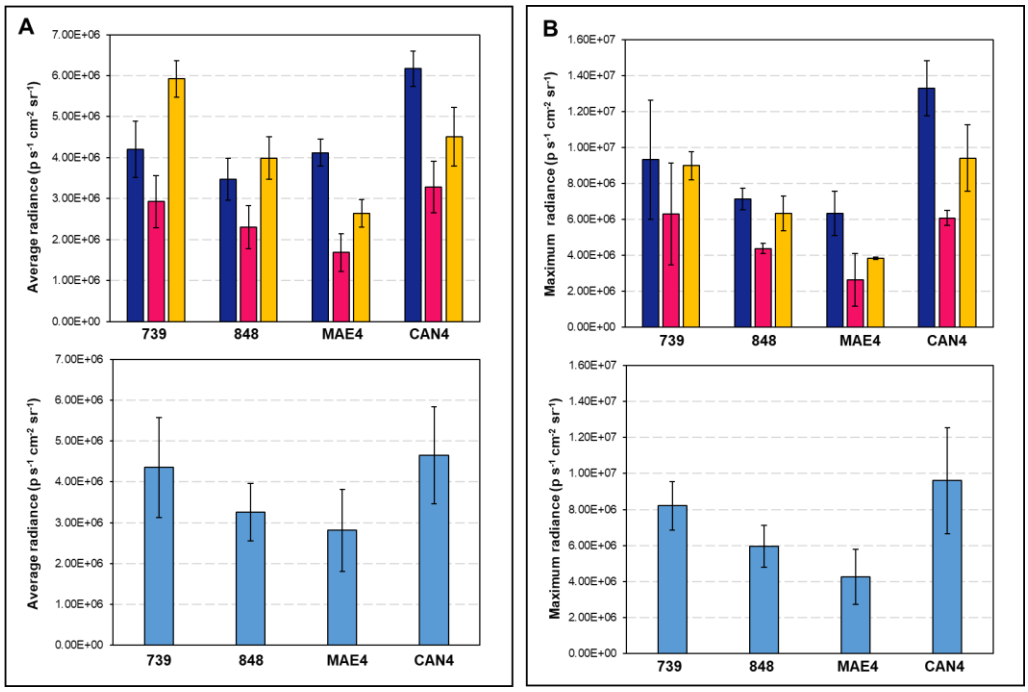


Figure 6-10: Radiance values from BMA producer controls

Using the strains **LUC0809**, **LUC0739**, **LUC0848**, **MAE4** and **CAN4** we spotted 10 µL culture onto membranes on LB in triplicate. Colonies grew at 37°C for 48 hours. Membranes incubated on SPA agarose for 1 hour and mDTZ activated under UV light for 30 min. Average and maximum radiance calculated using the IVIS and Living Image software version 4.3.1 (64-bit), with values corrected for **LUC0809** negative control grown on each plate by subtracting **LUC0809** results from test strains. Bars in top panels, ■, ■, ■, each represent the average of 3 spots on one plate. Bars in bottom panels ■ represent the average of all spots on all plates, where the largest minimum and maximum outliers are removed. Fluorescence measured at 465/520 ex/em with an exposure time of 30 seconds. Error bars calculated as standard error of the mean.

6.2.2.3 Technical developments for IVIS Spectrum data collection

A technical issue made clear from the initial IVIS Spectrum plate reads was the extreme radiance signal that appears to originate from the sides of the agar plate reflecting back towards the camera (red perimeter in Figure 6-9). Often the images

taken using the IVIS are used to detect large populations of fluorescently marked microorganisms. Therefore, the comparatively low levels of fluorescence generated by the µM BMA titres generated by our control producer strains are not sufficient to overcome any background fluorescence issues. The presence of this feedback signal may skew the IVIS Spectrum data, resulting in inaccurate comparisons between controls and experimental strains. I tried to find alternative agarose vessels that would not interfere with the already low signal emitted from the activated SPA agarose (Figure 6-11).

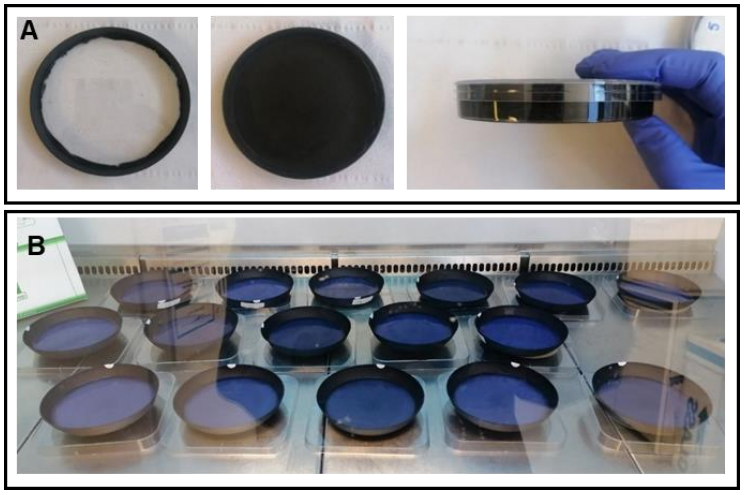


Figure 6-11: Trialling agar plate configurations to reduce feedback

Initially, I tried adding black card with a circular cut out over the top of the assay plates to block the edges of the agar plates from view of the IVIS camera. However, this did not reduce the background from the base of the agar plates. Secondly, agar plates were painted black on either just the sides or the entire plate using nail polish (Figure 6-11A). Painting only the sides of the plates did not reduce our error but painting the entire agarose plate black completely removed any interference. Although painting the agarose plates black solved the issue, the solution was high expense and low throughput. As a result, I decided to use matt black plastic plant saucers at 100 mm diameter as a reusable, higher throughput alternative that produced the same results (Figure 6-11B). Crucially, this removed the reflective feedback from readings and reduced the overall error in the results.

As a comparison between the normal and black agarose plates, I set up a series of plates in the following way: Nylon membranes were divided into six segments and plated with triplicate 10 µL spots of **LUC0809**, **LUC0739**, **LUC0848**, **MAE4** and **CAN4**. When the membranes were transferred to SPA agarose, I used either a normal agar plate, an agar plate with painted black sides, or an all black plant saucer. I carried out the screen using the conditions described above, with 1 hour incubation and 30 min UV activation on the SPA agarose. Maximum radiance was then calculated from each spot for **LUC0739**, **LUC0848**, **MAE4** and **CAN4** corrected for the **LUC809** negative control (Figure 6-12).

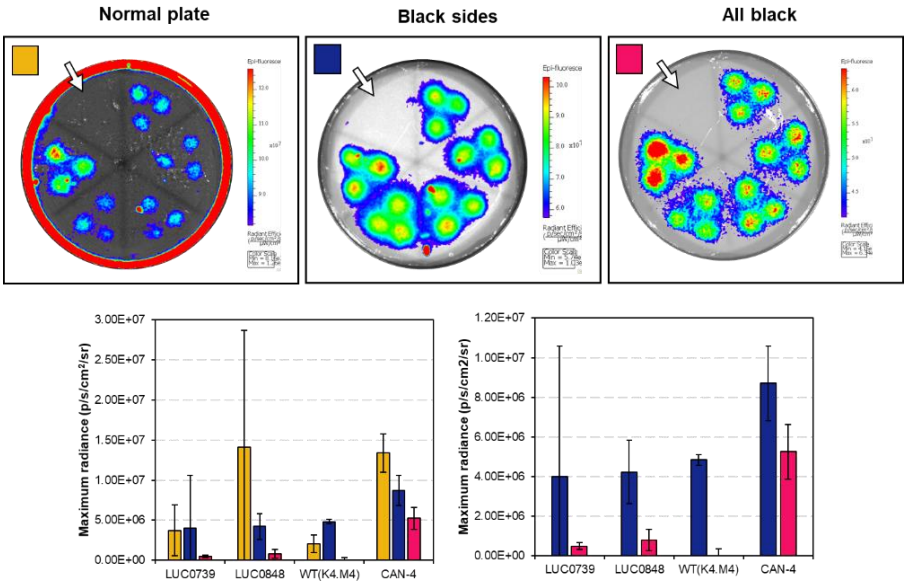


Figure 6-12: Effect of plate type on maximum radiance values

Strains used (clockwise from arrows) were **LUC0809**, **LUC0739**, **LUC0848**, **WT(K4.M4)**, and **CAN-4**. 10 µL culture was spotted onto membranes on LB in triplicate. Colonies grew at 37°C for 48 hours. Membranes incubated on SPA agarose for 1 hour and mDTZ activated under UV light for 30 min. Maximum radiance measured using the IVIS and Living Image software version 4.3.1 (64-bit), and values were corrected by subtracting the radiance of the **LUC0809** negative control. Fluorescence measured at 465/520 ex/em with an exposure time of 30 seconds.

No radiance was detectable from non-spotted areas of either the black sides or all black plates. As expected the normal agar plate had a significant fluorescent signal from the edges of the plate. Visual distinction between spots of different strains was clearest on the all black plant saucers. Maximum radiance values for the normal plate were similar for both **LUC0848** and **CAN4**, with a large amount of reading error in the

LUC0848 data that arose from one intense point of signal visible on the normal plate image. A similar problem occurred with the black sided plate data for **LUC0739**. The normal agar plate data is not consistent with flask results, and the error between replicates is too large for a robust screening method. Introducing black sides, whether only on the sides or for the entire plate, immediately improves the accuracy of the readings. In terms of selecting between black sides or an entirely black plate, as **LUC0739**, **LUC0848** and **MAE4** are similar in BMA productivity compared to **CAN4**, it is important to select for the approach that reflects this large variance in productivity. The best distinction observed was using the all black plant saucers, where **CAN4** clearly produces more fluorescence than all other strains, as well as this plate data producing the lowest error in replicates. For all subsequent screening experiments, I switched to using black plant saucers.

6.2.2.4 Culture spotting method and final screening conditions

At the outset, the aim for the final assay was to be able to spread a mixture containing the entire ACX/AAT library as a one pot mix onto membranes on LB. From which the IVIS would be used to detect variations between individual colonies containing different constructs. As showed through initial experiments, colonies grown from spots of individual cultures on membranes produce PZ-BMA signal that can be differentiated between a non-producer and a BMA-forming strain (Figure 6-13, A1). However, when **LUC0739** and **LUC0848** were assayed, which were expected to produce very similar PZ-BMA signals, measured radiance did not match the GC-MS predictions (Figure 6-13, A2). Although a small number of individual colonies were visible from the IVIS Spectrum readings, the radiance was low enough that surface reflection from the agarose obfuscated the fluorescent signal. This low radiance problem persisted even when **CAN4** was introduced to produce a stronger PZ-BMA signal (Figure 6-13, A3). This effect can be seen clearly in B (Figure 6-13, white circles), where the brightest spots on the plates are at regular positions in relation to the camera. It is possible that

the radiance produced from individual colonies was too low even for the IVIS Spectrum to reliably detect.

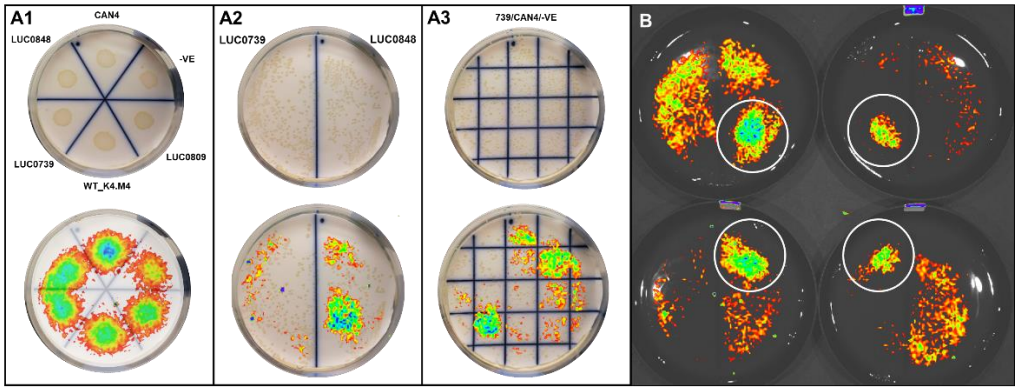


Figure 6-13: Visibility of individual colony and culture spot imprints using IVIS

Strains used are **LUC0809**, **LUC0739**, **LUC0848**, **CAN5(-VE)**, **MAE4**, and **CAN4**. A1: 10 µL culture was spotted onto membranes. A2 & A3: 30 µL of a 10^{-5} dilution was spread onto membranes on LB. Colonies grew at 37°C for 48 hours. Membranes incubated on SPA agarose for 1 hour and mDTZ activated under UV light for 30 min. B: Regions highlighted with white circles indicate areas where SPA agarose surface reflection produced a stronger reading than PZ-BMA because of the low radiance readings for individual colonies. Maximum radiance calculated using the IVIS and Living Image software version 4.3.1 (64-bit). Fluorescence measured at 465/520 ex/em with an exposure time of 30 seconds.

Due to the poor PZ-BMA signal when imaging individual colonies, I decided to return to the culture spotting method for testing our ACX/AAT libraries. However, the original colony spotting method would only allow us to compare 5 strains *per* plate, accounting for a **CAN4** comparison spot on each. Therefore, the volume of culture spotted onto membranes was reduced to decrease the size of the eventual spots measured on the IVIS Spectrum. If a smaller starting volume could be found that still permits differentiation between **CAN4**, **LUC0739**/**LUC0848** and **CAN5** then smaller end colonies and therefore test a greater number of strains and/or replicates on each plate could be produced. As a result, I tested various set up volumes and starting dilutions to find a suitable compromise. **CAN4**, **CAN5**, **LUC0739** and **LUC0848** overnight cultures were spotted onto membranes on LB in triplicate, with either 4 µL, 2 µL, or 1 µL starting volume (Figure 6-14, A-C).

As expected, spots with the highest starting volume produced images with greater overall radiance. Although the 4 µL spots were brightest, this produced **LUC0848** as

the highest producer, despite flask validation confirming otherwise. It may be that **LUC0848** over a longer period of time on plates is capable of accumulating higher BMA concentrations than **CAN4** However as the screen was designed to give results representative of what may be found in flask validations, it appeared that spotting cultures with a lower initial volume of cells improved the screen validity. It also appeared to that at lower cell concentrations the detectable radiance from the lowest producer **LUC0739** and the negative control **CAN5** became almost indistinguishable.

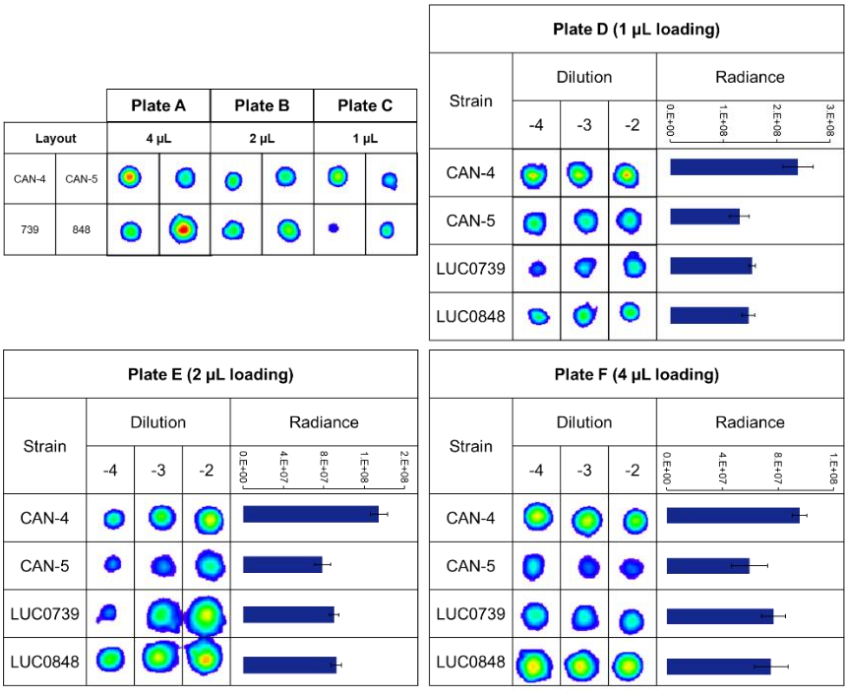


Figure 6-14: Volume and dilution comparison for IVIS screening conditions

Strains used are **LUC0739**, **LUC0848**, **CAN5**, and **CAN4**. Different culture volumes and dilutions were spotted onto membranes on LB. Colonies grew at 37°C for 48 hours. Membranes incubated on SPA agarose for 1 hour and mDTZ activated under UV light for 30 min. Bars ▮ represent maximum radiance calculated using the IVIS and Living Image software version 4.3.1 (64-bit). Fluorescence measured at em/ex 520/465 with an exposure time of 30 seconds. Error bars represent the mean of triplicate data.

I investigated this further by varying both the number of cells (by dilution) and the initial spot volume loaded onto membranes (Figure 6-14, Plates D-F). Here I photographed the colonies in addition to measuring radiance, and corrected values for the final colony size. When this was done, I found that the lowest dilution rate paired with the highest loading volume gave the most representative data for producing a distinction between test strains. At the lowest dilution, the pattern of radiance for

CAN4, **CAN5**, **LUC0739** and **LUC0848** fit the expected BMA concentrations as observed from the flask tests. As before, no real distinction could be made between **LUC0739** and **LUC0848**. As the original 10 µL starting spots produced a much better resolution than individual colonies (Figure 6-14) it is not surprising that the larger loading volumes produced stronger results. However, spots grown from 4 µL loading volumes would allow an increased number of test strains *per* plate to 30. Plating diluted starter culture appeared to have a negative impact on both error and screen resolution. I repeated the experiment, this time spotting cultures onto plates without diluting beforehand, while using three different loading volumes, 4 µL, 2 µL or 1 µL (Figure 6-15 G, H, I). Triplicate of each plate was prepared and screened under normal conditions (Figure 6-15).

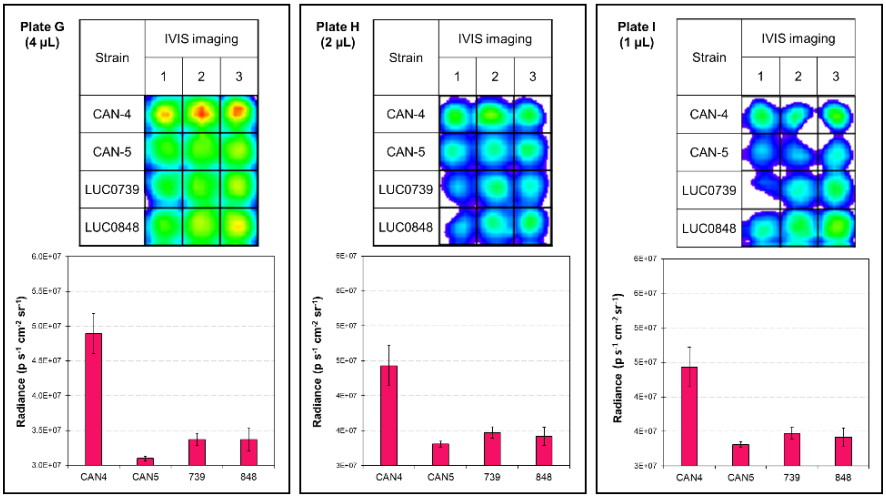


Figure 6-15: Validating screening conditions for optimal radiance measurement

Strains used are **LUC0739**, **LUC0848**, **CAN5**, and **CAN4**. Different culture volumes, undiluted, were spotted onto membranes on LB. Colonies grew at 37°C for 48 hours. Membranes incubated on SPA agarose for 1 hour and mDTZ activated under UV light for 30 min. Bars ▮ represent maximum radiance calculated using the IVIS and Living Image software version 4.3.1 (64-bit). Fluorescence measured at em/ex 520/465 with an exposure time of 30 seconds. Error bars represent the mean of triplicate spots on a plate.

From all plates, there was a high distinction between the radiance recorded from **CAN4** and the negative and low producing strains. However, on the plates with lower starting volumes (Figure 6-15 H, I) there was less difference between the radiance of **LUC0739**, **LUC0848**, and **CAN5**. Removing any dilution of the starting culture greatly improved both the resolution and the error for all loading volumes. Without this, it

appears that all the loading volumes tested would be sufficient to distinguish between **CAN4** and **CAN5**. I selected the 4 µL starting volume, as plating 30 spots *per* plate gives both sufficient numbers of reads while also being easier to handle during the set up process than using membranes with 60+ spots.

6.3 Discussion

Several changes had to be made to the original screening method to facilitate switching to our IVIS Spectrum approach. I decided to plate the test strains as spots on a gridded membrane, with a start volume of 4 µL and no dilution beforehand from the pre-culture. Incubating on LB for less than 36 hours did not produce colonies of a sufficient size to form quantifiable amounts of BMA. Therefore, an incubation time of 48 hours was used for all runs of the fluorescent bioassay. An increased time was also beneficial when incubating cells on SPA agarose and when activating mDTZ under the UV bulb, and so these were set at 1 hour and 30 min, respectively. The UV-plate distance used at the lab in Nottingham was roughly three times the distance used at Ingenza. Reducing this distance and placing plates closer to the UV source did not affect the intensity of the final PZ-BMA signal, and this metric was not altered.

Although the initial plan was to use the IVIS to compare small, single colony spots quantitatively, too much error was present in radiance readings to use this approach. In future, cells could be incubated for even longer to compensate for the low signal with a higher OD_{600}, although this may likewise reduce the productivity of *E. coli* as late stationary phase is reached. In the subsequent experiments within this project, culture spots were to allow a higher cell density without further increasing incubation time. Although this lowers the throughput of the approach, it will make the process of identifying and picking colonies far simpler once IVIS hits have been identified.

Visible light reflective surfaces formed a major challenge when trying to establish this method. As previously mentioned, the IVIS Spectrum is conventionally used to track the progression of infection in rodents. Therefore, reflection is not a usual

consideration in the function of the equipment. Replacing the traditional agar plates with matt black plates reduced the effect of this reflection, however the harder to solve problem was surface reflection from the SPA agarose. I began drying the SPA plates in the MSC before adding the Nylon membranes to reduce surface moisture. This had a negligible effect on feedback. When bright colonies such as those for **CAN4** were grown as large spots the surface reflection effect was reduced. This couldn't be accounted for in the ACX and ACX/AAT library screening, particularly as it was inevitable that some candidate enzymes would not form BMA, therefore producing no PZ-BMA signal. With further development it may be useful to try and 'mattify' the agarose surface in some way, such as preparing SPA agarose with a higher agar concentration to reduce its liquidity.

Switching to the IVIS Spectrum method allowed quantitative comparison between the relative fluorescence of PZ-BMA from the new ACX and ACX/AAT strains *versus* **CAN4**. This was an improvement from the 'by eye' approach necessitated when using the UV illuminator to image PZ-BMA. The screen as described here appears to be capable of distinguishing between a producer strain which makes approximately 0.025 mM BMA, as in the case of **CAN4**, and strains with a 5-fold lower efficiency, as found with **LUC0739**/**LUC0848**. It remains to be determined whether the dynamic range of the screen would be able to distinguish higher BMA production rates in the event of libraries in future containing a ACX or AAT with improved efficiency.

Although fluorescence-based assays are known for their high sensitivity, the small fluctuations in analyte titre observed here remain too low even for mDTZ to accurately predict productivity. By extension, part of the challenge here is that of the control 'producer' strains available, only **CAN4** produced a reliable concentration of BMA between runs. This prevented effective establishment of a radiance:BMA concentration calibration, which could then be used to more accurately predict flask titres. In future it would be useful to develop a number of pCAN4-like plasmids, with varying promoter strengths at the start of BCKD and AtACX4. This could be used to

assemble a ladder of BMA control strains from which the IVIS Spectrum screen could be calibrated. Nevertheless, it is important to use this screen as a first pass for the AAT/ACX production library, as it is able to detect any large increases in BMA formation (>1 μM) of the scale that will be of industrial interest. This detection can take place at a faster rate than is achievable using a conventional flask screening approach.

Chapter 7: Bioinformatic identification of ACX and AATs

7.1 Introduction

7.1.1 Bioinformatic selection of enzymes for increased bioprocess titres

Bioinformatics has made it increasingly possible to harness protein and nucleotide sequence 'big data' for synthetic biology applications over the past 70 years. Without the ability to rapidly align protein sequences and interrogate structures for emergent substrate promiscuity, most novel metabolic pathways used to be developed using a combination of literature searching and individual expertise [200]. The downside of this approach was both a high monetary and high time cost, as well as often not providing researchers with a good coverage of possibilities for rational pathway design [200]. Margaret Dayhoff's development of the first computational method for assembling 1° protein structure data, COMPROTEIN, kickstarted several decades of rapid improvement in our ability to collect and analyse protein sequences [201]. In the present day, multiple sequence alignment (MSA) and molecular dynamics modelling (MD) are frequently used instead to inform the selection of enzymes for *de novo* metabolic pathway design. MD in particular allows us to bypass exhaustive crystal structure determination in the initial screening for potential enzyme candidates, capitalising on our knowledge of homology to predict likely catalytic activities.

Using bioinformatics to inform metabolic pathway optimisation, as opposed to a purely synthetic biology approach, increases the biochemical flexibility of process design. Indeed, supplanting existing proteins in a novel pathway with promising "hits" identified using bioprospecting methods has the potential to circumvent many rounds of incremental improvement by working on one enzyme alone. With the increasing relevance of industrial biotechnology, which often requires the synthesis of non-native or non-natural compounds, the ability of bioinformatics to target *de novo* pathway design to an end product rather than existing metabolite is crucial [202].

In terms of BMA biosynthesis, particularly the steps catalysed by AtACX4 and AATm4, we have a head start compared to 'from scratch' pathway design. There is a wealth of data already demonstrating catalytic activity of AtACX4 and AATm4 with IB-CoA and M-CoA for flux through the BMA metabolic pathway, albeit with inefficient catalysis. A major downfall of using both AtACX4 and AATm4 in this context is that the limitation of each of these enzymes exacerbates that of the other. In practice, this means that AtACX4 is inhibited at low concentrations by its product, preventing it from facilitating significant intracellular concentrations of M-CoA. At the same time, AATm4 requires high concentrations of M-CoA in order to ensure specificity for BMA, as opposed to the more readily produced BIB. This problem is intensified by the high concentration of IB-CoA available in BMA producer cells, either as an intermediate in valine metabolism or due to BCKD activity. Therefore AtACX4 product inhibition and AATm4 inefficiencies compound, resulting in insufficient intracellular M-CoA for industrial BMA synthesis.

If we first look at AtACX4, product inhibition is not a simple problem to overcome. Determining the molecular relationship between protein sequence and inhibition can be both a time and labour-intensive process. Additionally, if inhibition is competitive or mixed, then alleviating it will frequently require active site modifications. Engineering a reduced product inhibition in this way will come at the probable cost of reducing the affinity of a protein for its intended substrate. In the case of AtACX4, a reduction in M-CoA affinity would further impair BMA formation as IB-CoA availability may increase, favouring BIB biosynthesis. There are several ways product inhibition can be dealt with. Often product-inhibited enzymes reported in literature catalyse the final step in biosynthesis of a target compound. The simplest solution in many cases, therefore, is the timely removal of the inhibitory product during the bioprocess. For example, using *in situ* product recovery. This approach is not possible in the case of BMA production of course, as the removal of our inhibitory product – M-CoA – is limited by the activity of the AATm4.

In the case of AATm4, examples of successful engineering of transferases for increased specificity are few and far between. Few predictable links have been found between AAT sequence and function, and no crystal structure data is available to aid the search. This makes rational design or targeted mutagenesis a potentially time consuming affair. Additionally, even when provided with higher substrate concentrations, the K_M of AATs for non-acetyl substrates remains insufficiently high. Despite this, the myriad specificities of acyltransferases demonstrated experimentally (Table 1-1) lends weight to the argument that there may exist unknown, favourable activity for a methacrylyl substrate in an AAT derived from a different organism.

Therefore, I attempted to use bioinformatics to identify alternative candidate enzymes to catalyse the final two reactions during BMA biosynthesis. Both AtACX4, and to some extent *M. pumila* AAT, were the first proteins to demonstrate the desired catalytic activity for this process. Investigation of related enzymes with homology or more favourable substrate kinetics that may possess more satisfactory reaction kinetics has to date been limited. Targeting a bioinformatic search for new ACX and AAT enzymes may prove to be a more expedient way to increase both the specificity and titres of BMA produced from a microbial chassis.

7.1.2 Structure, activity, and inhibition of acyl-CoA oxidase 4 from *A. thaliana*

7.1.2.1 Removing enzyme product inhibition for improved bioprocess catalysis

Enzymes are often inhibited by their products, and protein structure is linked to product inhibition *via* a range of mechanisms. By nature, a substrate often bears structural similarity to the product, leading to competitive inhibition. Secondly, the dissociation rate of the enzyme-substrate complex is a reversible process which is slowed by building product concentrations, causing non-competitive inhibition. Non-competitive inhibition is a particularly challenging problem for industrial bioprocesses, which require unnaturally high product concentrations. Lastly, product inhibition can be caused by uncompetitive binding, where the product binds to an allosteric site on the enzyme: substrate complex, limiting the forward reaction as a result [203]. As we

determined from our ACX kinetic assay (Section 5.2.8), M-CoA inhibition of AtACX4 likely occurs *via* a mixed/non-competitive mechanism. This means the concentration of product, M-CoA, will largely dictate the permitted activity of AtACX4.

Conventionally, reducing product inhibition can be achieved in three ways. Perhaps the most simple is ISPR, as mentioned above, which can meet with rapid improvements in production titres. However this is only a viable solution when product inhibition occurs at the final step in product synthesis, and thus is not applicable for our purposes of targeting AtACX4 activity.

Protein engineering, targeted towards altering active site properties, is the most common solution to reducing product inhibition [204]. For example, in the case of heterologous ß-galactosidase expression in *Pichia pastoris*, researchers working on ß-galactosidase increased its K_i with galactose from 0.76 mM to 6.46 mM, using targeted mutagenesis of the active site [205]. Similarly, in the latter series of experiments, researchers discovered that L-Serine *O*-acetyltransferase (SAT) was competitively inhibited by L-cysteine at the same location as its substrate, L-serine [206]. Comparison of L-serine *vs.* L-cysteine bound SAT revealed a single residue, Asp92, which relocated substantially only during L-cysteine binding. Site-directed mutation around Asp92 led to the identification of a substitution that renders SAT largely insensitive to L-cysteine-mediated product inhibition, with a K_i of 950 µmol mL^{-1} [206]. Targeted engineering has also successfully improved production titres in several instances, for example Atreya *et al*. increased the tolerance of the cellulose enzyme Cel7A from *Talaromyces emersonii* to its product, cellobiose [207]. Cellobiose inhibits TrCel7A *via* a mixed/competitive model. Site directed mutation of 10 previously identified sites, based on MD simulations, resulted in increased cellobiose tolerance of up to +25% compared to wild-type control. However, in all cases increasing cellobiose tolerance was concomitant with a corresponding reduction in Cel7A activity [207]. Indeed, all of the approaches outlined here come with the associated downside of either resulting in a trade-off for enzymatic activity, and

usually require a laborious process of structure determination, and random or targeted mutagenesis, with only incremental improvements [204].

The last, and often overlooked, solution is screening for alternative product-tolerant enzymes. This is likely due to the relatively recent development of the bioinformatic tools necessary to properly interrogate the available literature and published kinetic data. Although identifying alternative enzymes of the same class is an under-used approach, the few examples of its use have met with some success. For example, if we return to the problem of cellobiose-inhibited cellulase we can see an example of using previous literature to demonstrate improved cellulase activity. In this example, a group of researchers made the incidental discovery of a product-activated cellulase Cel3A from *Polyporus arculanus* (PaCel3A) [208]. On addition of 20 mM cellobiose, the activity of PaCel3A increased by more than 500%. This discovery was neglected for 10 years while other groups attempted to overcome Cel7A inhibition by site-directed mutation. However, in 2021 Zou *et al.* reported on the replacement of *Trichoderma reesi* Cel7A with PaCel3A, based on this finding [204]. Simply replacing TrCel7A with PaCel3A led to an immediate 56.4-63.0% increase in production efficiency, as cellobiose inhibition was removed from the system [204].

In the case of ACX enzymes, is it possible that we are likewise missing a simple solution for the prohibitive inhibitory effect of M-CoA on AtACX4? Several other ACX enzymes have demonstrated encouraging activity on substrates similar to IB-CoA, but have not been investigated *in vivo* for any M-CoA generating activity. It is therefore prudent to investigate alternative ACX enzymes to either identify an M-CoA independent ACX, or to identify an ACX with increased resistance to its product. Either way, this would allow us to marginally increase the size of the intracellular M-CoA pool. This in turn would favour the increased selectivity of AATm4 for M-CoA as a substrate. Unpicking the selectivity of AATs is a particular challenge, so addressing activity in this more roundabout way would be a favourable outcome.

7.1.2.2 Crystal structure and catalytic site of AtACX4

Fatty acids (FAs) are metabolised in plants by peroxisomal ß-oxidation. The first step in FA ß-oxidation is the generation of a 2-*trans*-enoyl-CoA from an acyl-CoA substrate, catalysed by an acyl-CoA oxidase (Figure 7-1) [125]. All ACX enzymes are flavoproteins, containing a flavin adenine dinucleotide (FAD) prosthetic group. During the ACX catalysed reaction, FAD is reduced to $FADH^-$, accepting an electron from the acyl-CoA substrate. $FADH^-$ is subsequently re-oxidised to FAD by molecular oxygen, forming H_2O_2 [118].

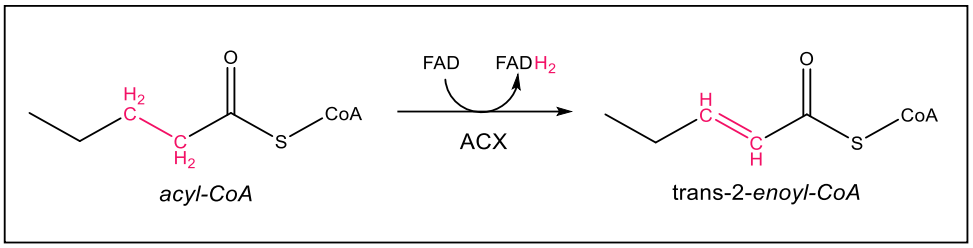


Figure 7-1: Reaction carried out by ACX enzymes

A. thaliana expresses at least 6 isozymes of ACX, each favouring a different chain length substrate, with some overlap. ACX4 is short chain specific, and is 50 kDa in size [118]. It was first purified in 1999 and exhibited activity on hexanoyl-CoA (K_M = 8.3 µM) [125, 193]. ACX4 has the least sequence identity to the other ACX isozymes in *A. thaliana*, but its reaction mechanism is identical. The arrangement of subunits in ACX4 is a dimer of dimers, with the principal interaction taking place between the C-terminal α-helix domain of one subunit with the middle ß domain of another (Figure 7-2A) [121]. It is theorised that a cysteine residue (Cys399) may form a strong disulphide bond between the two subunits *in vivo* [118]. The individual subunits of ACX4 possess a similar conformation to the dehydrogenase fold structural motif common to 20% of crystallised proteins. This consists of a 6-stranded parallel ß-sheet with a ß-α-ß-α conformation at the N-terminus, followed by a central ß-strand domain and a bundle of four C-terminal α-helices [118].

Each tetramer of ACX4 contains four FAD molecules, two at each dimer interface (Figure 7-2B). Each ACX4 subunit has an N-terminal extension towards its adjacent subunit. This extension ‘caps’ the interface where FAD is bound, occupying the site used by ETF during co-factor regeneration of ACADs [121]. Shielding of the FAD cofactor is a feature common in flavoproteins that favour oxidase over dehydrogenase activity [118].

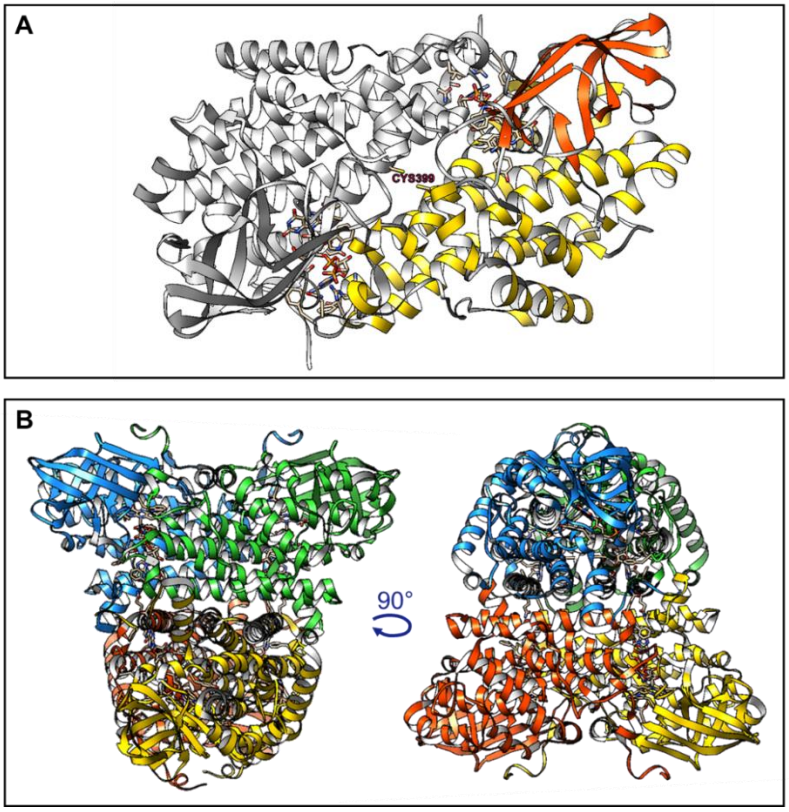


Figure 7-2: Crystal structure of AtACX4

Representation of the complete crystal structure of ACX4, with ribbon representing secondary structure. A: One dimer unit of AtACX4, and B: Final tetrameric structure of AtACX4, each subunit coloured in either blue, green, orange, or yellow. FAD cofactor, substrate analogue acetoacetyl-CoA, and active site residues are represented as sticks. Figure made using UCSF Chimera (PDB ID: 2IX6).

R-S-CoA substrates bind the ACX4 active site with the thioester bond placed between the ß-strand domain and the C-terminal four helix bundle (Figure 7-3A). The –CoA portion of the ligand then makes extensive interaction with residues within the active site (Figure 7-3B) [121]. This includes salt bridges between the diphosphate and Arg420, hydrogen bonds to Ser184, Ser181 and Glu408, as well as hydrophobic

interactions with Arg420, Leu174, Val417 and Glu408 [118]. In fact, the majority of residues within the ACX4 binding pocket are hydrophobic, and largely resemble the residue composition seen in other plant acyl-CoA oxidases. In particular, the interactions seen between Ser184, Arg420 and Lys429 with the acetoacetyl-CoA substrate are well conserved amongst other SC specific oxidases [118].

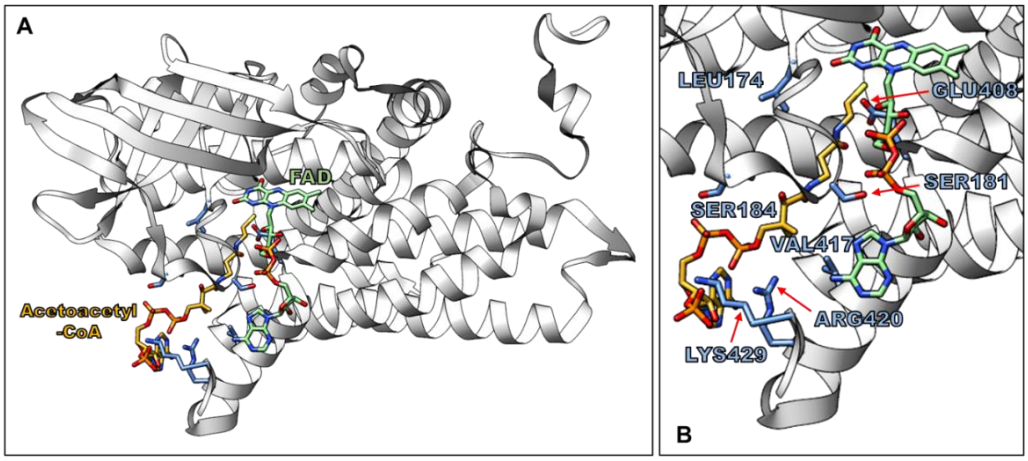


Figure 7-3: Active site residues of AtACX4 in complex with acetoacetyl-CoA

A: One subunit from the homotetramer of ACX4 from *Arabidopsis thaliana* (AtACX4). B: Amino acid residues interacting with the acetoacetyl-CoA ligand in the active site of AtACX4. FAD shown in **green**, acetoacetyl-CoA shown in **yellow**. Residues represented by ball and stick models in blue. Figure created using UCSF Chimera with the crystal structure PDB ID: 2IX5.

Indeed, maximal activity of ACX4 has been observed with butyryl-CoA, with substrates larger than octanoyl-CoA being only very weakly oxidised [125]. This is due to the small binding pocket of ACX4 not being able to physically accommodate larger molecules. This is in part why AtACX4 was initially selected for the oxidation of IB-CoA to M-CoA, and subsequently several kinetic values have been determined experimentally for the K_M of AtACX4 for IB-CoA. In 5.2.8 we determined AtACX4 K_M and V_{max} as 140 µM and 0.61 µmol min^{-1} mg^{-1}, respectively, at pH 7.5 and 30°C. Hydrogen peroxidase coupled assays carried out by Ingenza gave K_M values between 120 µM to 1 mM, depending on experimental variables such as pH, temperature, and MBP coupling. Although the K_M value varies according to the conditions of each assay, the average K_M for IB-CoA with AtACX4 is 300 µM, or 223 µM, if we exclude outliers.

Despite access to a reasonable amount of experimental data and solved crystal structures of ACX enzymes, the same logic as used in the cellulase experiments cannot be exactly applied. To date, no product resistant ACX has been described in literature. However, the bioinformatic search can be targeted towards ACX enzymes which are not able to use M-CoA or similar compounds as a substrate. Although there are no ACXs other than AtACX4 reported to utilise IB-CoA, alternative ACXs which are less active towards short-chain acyl-CoAs can be identified. For example, those utilising substrates with fewer than the four carbons present in IB-CoA, or other -CoA substrates with greater similarity towards M-CoA. These enzymes, if they can be identified, may have less affinity for M-CoA compared to AtACX4, reducing the possibility of a competitive inhibition effect of M-CoA. Previous work has also demonstrated improved activity in site-directed mutants, so it is also important to include in the final library a range of active site conformations.

Identifying a product resistant ACX, or an ACX with higher activity than AtACX4 has the potential to have a direct favourable impact on AAT activity. However, if no improvements can be made upon the IB-CoA to M-CoA step in BMA synthesis, an ISPR approach could also be taken to increase ACX activity: Under the right conditions AtACX4 can produce M-CoA at a high rate. Therefore, I targeted a second search for novel AAT enzymes that may relieve product-inhibition by drawing M-CoA more swiftly towards BMA formation.

7.1.3 AAT substrate promiscuity and structure-function relationship

Unlike ACX4 enzymes, the link between sequence and specificity of function in AATs is not so well understood. The first alcohol acetyltransferase was purified from *Cladosporium cradosporioides* in 1978, followed by *S. cerevisiae* in 1981 [209, 210]. The first plant AATs were isolated from banana (*Musa sapientum*) [210]. In fruit, AAT is one of several enzymes responsible for the cocktail of volatile compounds produced during ripening [172]. The complex combination of these compounds are what give fruit and other plants their characteristic flavour and aroma profiles [211]. The ester

portion of these profiles is most commonly produced by the transacylation from acyl-CoA to alcohol, catalysed by AAT [172]. The complex cocktail of volatile esters produced during ripening are likely caused by the promiscuity of AAT enzymes towards a diverse range of alcohol and acyl-CoA substrates.

It has long been recognised that enzymes are generally able to accept alternative substrates. The more similar the structure, the more likely it is that an enzyme will accept a new substrate [212]. This can stretch as far as one enzyme catalysing different classes of reaction; as is the case with chymotrypsin, which has amidase and phosphotriesterase activity, and myoglobin, which can catalyse O_2 binding and sulphoxidation [212]. Indeed, enzyme substrate promiscuity is thought to be more prevalent than originally theorised, with 37% of enzymes in *E. coli* able to demonstrate activity on multiple substrates [213, 214]. This flexibility improves the likelihood of finding an enzyme to produce your novel product, but by nature makes it difficult to increase substrate exclusivity. Additionally, there is currently no straightforward way to screen for potential enzymatic side activities on databases such as MetaCyc. To compound this, side activity from an enzyme may not equate to activity favoured in the context of the cell. As demonstrated through the production of BIB in this project, and occasionally BA, AATm4 still favours IB-CoA and acetyl-CoA over M-CoA in BMA producer cells. In the case of A-CoA this is most likely because of its intracellular abundance, and although less prevalent, IB-CoA is likely present in higher quantities than M-CoA.

As mentioned, the key challenge in the identification of novel AAT candidates arises from the lack of structural data, and by extension sequence-to-function data, available for the acyltransferase class of enzymes [137]. It is known that the HXXXD(G) and DF(V)GWG motifs in BAHD acyltransferase amino acid sequences are generally responsible for catalytic activity and enzyme function [215]. Yet no definitive link can be found between the amino acid sequence and AAT specificity, perhaps largely due to a lack of structural models. El-Sharkaway *et al.* compared melon and interspecies

AAT sequences to look for a relationship between sequence and function [137]. They also compared melon CmAAT1 and CmAAT2 to determine why of the two closely related AATs, only one can produce volatile esters. They concluded that AATs with low sequence identity, for example 22% between SAAT and LmAAT1, often demonstrated exceedingly similar substrate preference.

More recent research has begun to unpick individual AAT sequences in increased detail. Song *et al.* used homology modelling to map peach AAT (PpAAT1) onto an *A. thaliana* acyltransferase [216]. They used molecular docking and *in vitro* assay work to identify 9 catalytically important residues, including H165 from the HXXXD(G) motif and D381 from the DF(V)GWG motif as key for esterification and binding, respectively [216]. However all site-directed mutagenesis work reduced the k_{cat} and increased the K_M in the subsequent purified PpAAT1s. Wax synthase/diacylglycerol acyltransferases (WS/DGAT) are another group of enzymes within the BAHD acyltransferase family that bear similarity to AATs. Roulet *et al.* altered the alcohol specificity of PapA5 wax synthase from *Mycobacterium tuberculosis* and marginally improved its specificity towards ethanol, methanol and isopropanol [217] .

Despite these promising recent findings, there remains a lack of information relating to –CoA affinity and AAT structure, and what information is available has as yet failed to *improve* the activity towards targeted acyl-CoA substrates. Much information about the functionality of AATs points to the conclusion that altering availability of substrates available for AATs can often have a greater effect on their activity and selectivity than can altering the sequence [137]. This lends weight to the argument that identifying feedback resistant ACX would have the knock-on effect of increasing AAT specificity.

The AAT and ACX used were the ‘first’ to be tested in this context, and there may well be other enzymes equally, if not more, effective than the originals. Therefore, the most logical approach in this particular instance was the screening of alternative isoforms and enzymes from the wealth of kinetic and structural knowledge available. Product inhibition is usually solved using a combination of site-directed mutagenesis and

molecular dynamics. In this instance I selected alternative ACX enzymes from the kinetic data available on BRENDA (BRaunschweig ENzyme DAtabase, Technische Universität Braunschweig, Germany), as well as the significant amount of isoform labelled data on UniProt and GenBank. AAT selection provides a more complex challenge, as understanding of AAT sequence is decoupled from the specificity of this class of promiscuous enzymes. I accounted for this by using the phylogeny of a pool of AAT candidates to ensure a good coverage of sequence variation is accounted for in the alternative AAT library.

7.2 Results

7.2.1 Selecting acyl-CoA oxidase enzymes for IB-CoA activity

The first stage of identifying candidate ACX enzymes was to carry out several pBLAST (Basic Local Alignment Search Tool) searches. Initially I searched the translated sequence of AtACX4 as used in **CAN-4** against the non-redundant protein sequence database. Excluding hits from Arabidopsis, the majority of results were hypothetical or predicted sequences. On further inspection, many of these hits were also dehydrogenases, not oxidases. Removing both from the selection pool resulted in an initial list of 42 hits with an unnamed protein product from *Microthlaspi erraticum* showing the highest percentage identity of 94.51%. There was no experimental data connected with the pBLAST hits.

Running a multiple sequence alignment for the 42 pBLAST results, including the original AtACX4 sequence, allowed the exclusion of a further 13 sequences due to large insertion regions not concurrent with either ACX4, or the larger ACX1. The C-terminal SRL sequence was widely conserved between results, with the exception of ACX4 from *Populus alba, Actinidia chinensis* and *Artemisia annua*, all of which had a C-terminal SRM sequence. On checking each of the remaining 21 sequences against the UniProt database, no reviewed entries were available for any hits. Additionally, no experimental work, purifications or kinetic data had been published for any of the proteins. The majority of these sequences arise from whole genome sequencing work.

Searching for EC 1.3.3.6 on BRENDA yielded 70 entries associated with either K_M values or specificity information. Narrowing these down to hits with activity demonstrated on butyryl-CoA or hexanoyl-CoA resulted in 8 entries. No kinetic data was recorded for the interaction of ACXs with IB-CoA. The most promising entry was a *Zea mays* ACX, with a K_M value of 0.006 mM for hexanoyl-CoA. Unfortunately for several results, in particular *Glutamibacter nicotianae, Vigna radiata, Z. mays* and *Spinacia oleracea* the isoform or GenBank ID of the protein investigated was not specified within the literature. Results for *A. thaliana* were excluded, along with the mammalian ACX from *Rattus norvegicus*, and sequences for all identified ACX isoforms were identified in GenBank for the organisms where the protein was not precisely specified (Table 7-1).

Organism	Isoform	GenBank ID	Substrate	K_M (mM)
Spinacia oleracea	ACX1	XP_021856262.1	Butyryl-CoA	0.032
		KNA17362.1		
	ACXA	KNA24565.1		
	ACXB	KNA24566.1		
		KNA08139.1		
Glutamibacter nicotianae		VXC33980.1	Butyryl-CoA	0.1319
		WP_047119973.		
Vigna radiata	ACX3	XP_014524427.1	Butyryl-CoA	0.055
	ACX2	XP_014509033.1		
	ACX1	XP_014508908		
	ACX1	XP_014521199.1		
	ACX3	XP_022633883.1		
Zea mays		PWZ07906.1	Hexanoyl-CoA, C4-	0.006
		ACF78566.1		
		ACN28961.1		
	ACX1	AQL02798.1		
	ACX2	ACG45431.1		
Yarrowia lipolytica		CAA04661.1	Hexanoyl-CoA	0.067
Candida tropicalis		AAA34322.2	C4-C8	

Table 7-1: ACX enzymes with known activity on hexanoyl- or butyryl-CoA.

Data concerning K_M and substrate specificities are reported from BRENDA. GenBank IDs annotated according to all hits for ACXs from each organism on GenBank.

Of the hits investigated, several also demonstrated activity on alternative acyl-CoAs: in particular octanoyl-CoA, which could function as a substrate for *S. oleracea*, *G.*

nicotianae, V. radiata, and *Candida tropicalis* ACX [191]. The K_M for *G. nicotianae* ACX appears to decrease with chain length for the substrates tested, with a K_M of 0.1319 mM with acetyl-CoA, which decreases to 0.092 mM with hexanoyl-CoA and then 0.061 mM with octanoyl-CoA [190]. Multiple sequence alignment of the BRENDA sequences once again showed large insertions compared to *A. thaliana* ACX4. N- and C- terminal sequences were largely conserved, as seen with the results from the pBLAST search. I decided to use the sequences identified from BRENDA and BLAST, with any additional identified sequences to assemble a phylogenetic tree (Figure 7-4).

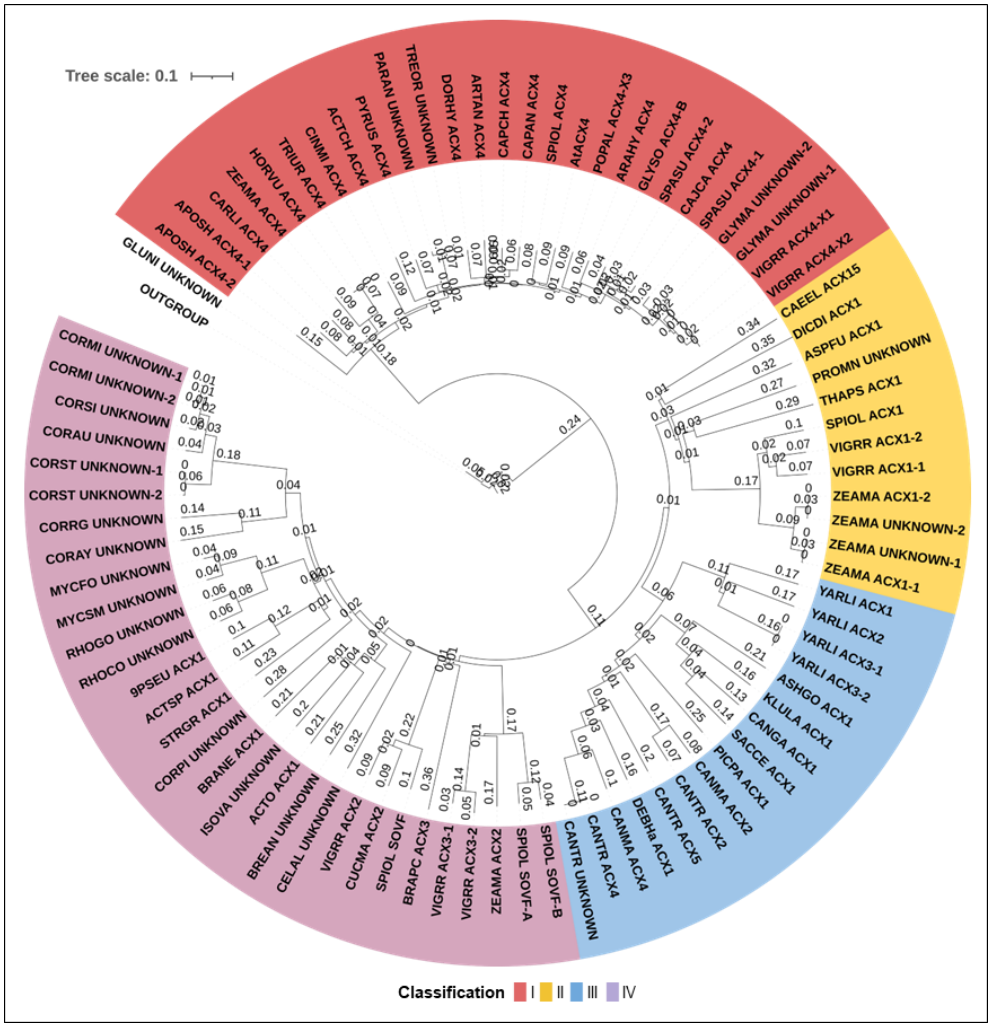


Figure 7-4: Phylogenetic tree of ACX enzymes identified through BLAST and UniProt

Full list of organisms and genes used in tree is shown in Appendix B.1: Outgroup is glutaryl-CoA dehydrogenase from *Arthrobacter* sp. (SEQ86329.1) Alignments carried out using ClustalO Multiple Sequence Alignment software (EMBL-EBI, Cambridge UK), and phylogenetic tree visualised using iTOL (interactive Tree Of Life (Biobyte Solutions, Germany) [218] Branch lengths shown on tree.

To do this, I carried out an additional search for EC 1.3.3.6 on UniProt, returning 250+ results. After removing mammalian and Arabidopsis entries 44 promising results remained. These were aligned in a multiple sequence alignment, and demonstrated much greater sequence variation than the hits identified using either BLAST or BRENDA. An MSA of all selected hits from pBLAST, BRENDA and UniProt was then carried out using ClustalO. Visualising these protein sequence alignments using a phylogeny tree, rooted to glutaryl-CoA dehydrogenase from *Arthrobacter* sp. gave four distinct clades (Figure 7-4). Unsurprisingly, many of the results from the pBLAST search reside in the Clade I with *A. thaliana* ACX, and are expected to retain some structural similarity to AtACX4 Figure 7-2, Figure 7-5A). These include the ACX4 isoforms from *Arachis hypogea, Populus alba, Apostasia shenzhenica* and unnamed proteins from *Parasponia andersonii* and *T. orientale*. ACX4 from *S. oleracea, V. radiata* and *Z. mays* are also in Clade I.

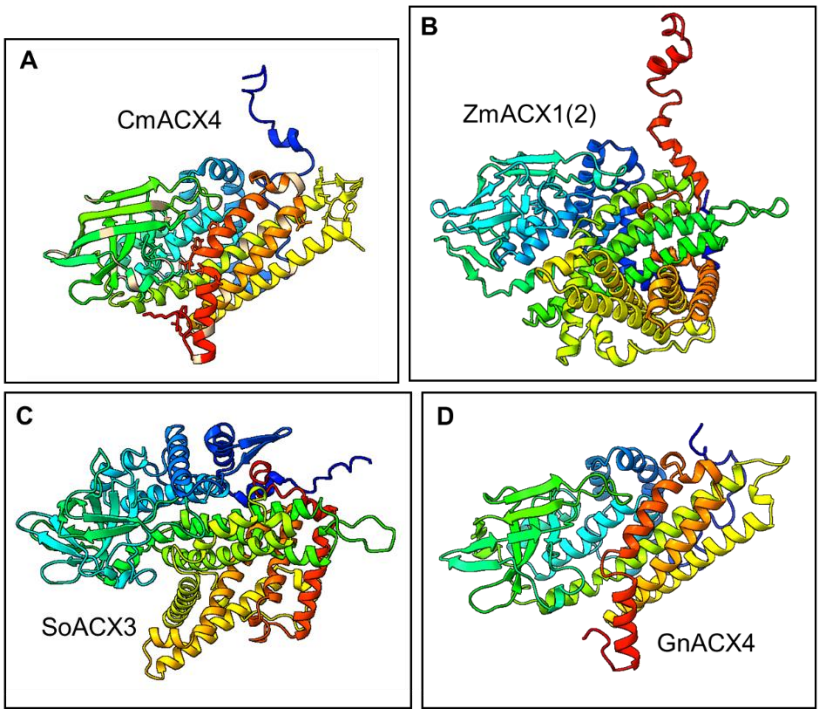


Figure 7-5: Predicted structures of AtACX4 and phylogenetically related oxidases

AlphaFold predicted ACX structures of subunit A from four enzymes identified from bioinformatic searching of candidate oxidases. A: CmACX4 = *Candida maltosa* ACX4 (CAA29901.1), B: *Zea mays* ACX1 (2) (AQL02798.1), C: SoACX3 = *Spinacea oleracea* ACX3 (KNA24566.1), and D: GnACX4 = *Glutamibacter nicotianae* ACX4-like enzyme (VXC33980.1). Structures coloured along polypeptide chain from N-terminal (blue) to C-terminal (red).

Clade II can be split into two parts, both appearing to broadly encompass the ACX1 isoforms, including those from *S. oleracea, V. radiata* and *Z. mays* in Clade IIA. In Clade IIB the enzymes are still annotated as ACX1 isoforms, but originate from various fungae, nematode and microalgae including *Neosartorya fumigata*, *C. elegans*, and *Prorocentrum minimum*. These unsurprisingly are the most divergent clade of ACX enzymes to AtACX4, due perhaps to the large C-terminal tail region present in the ACX1 subunits (Figure 7-5B).

Clade III broadly covers ACX enzymes from the yeast, or yeast-like, organisms identified in our search. This includes ACX1-3 from *Yarrowia lipolytica*, ACX2,4 and 5 from *C. tropicalis*, and ACX2 and ACX4 from *Candida maltosa.* Unsurprisingly therefore, the ACX sequences identified through BRENDA for *C. tropicalis* and *Y. lipolytica* are also within Clade III. Less common entries in Clade III were from *Pichia pastoris*, *Debaryomyces hansenii*, *Ashbya gossypii*, *Candida glabrata*, and *Kluyveromyces lactis*. It is expected that ACX1 enzymes may demonstrate more sequence divergence than the other isoforms, as ACX1 possesses a C-terminal domain that is not present in the other isoforms (Figure 7-5B). This domain is reported to prevent the formation of the homotetramers [118].

Clade IV is the largest group in the ACX phylogenetic tree. This is mostly because of the number of unique Corynebacterium sp. ACXs identified from UniProt. 9 entries in Clade IV are Corynebacterium. The remaining entries are all also gram-positive bacteria, with the exception of those from *Z. mays*, *V. radiata* and *S. oleracea*. The *S. oleracea* ACX enzymes are all closely related by sequence, as well as with the ACX3 sequences from *V. radiata* and the ACX2 from *Z. mays*. More structural similarity can be observed between SoACX3 and the ACX4 subunit structures than with ACX1 (Figure 7-5C).

Surprisingly, the ACX from *G. nicotianae* does not form part of a clade with any of the other identified enzymes. *G. nicotianae* is an Actinobacteria and therefore doesn't share a significant amount of similarity with either the plant, yeast or fungi containing

clades. However, we might have expected it to share similarity with the other gram-positive bacteria in Clade IV. If the GnACX structure is compared to the other identified isoforms of ACX, it appears to hold most similarity with ACX4 despite having slightly truncated N- and C- terminal regions in comparison (Figure 7-5D).

I selected 20 candidates from the ACX enzymes included in our phylogeny (Table 7-2). I ensured to include ACXs from each of the clades outlined above. 9 ACX4s were selected, as these have the highest likelihood of oxidising IB-CoA. I included the majority of ACXs identified from BRENDA that can accept either hexanoyl-CoA or butanoyl-CoA as substrates. This included the *G. nicotianae* ACX. Because it wasn't possible for us to identify which isoform of ACX is responsible for hexanoyl-CoA or butanoyl-CoA activity in *S. oleracea, V. radiata* or *Z. mays,* all available isoforms were included, in the most part leaving out duplicates with almost identical sequences. I also included four known ACX1 enzymes. The formation of homodimers instead of a dimer of dimers by ACX1 *vs* ACX4 will potentially result in a significant variation in enzyme activity between this group and the ACX4s.

Four of the oxidases selected are not annotated as any particular isoform of ACX. The first, ACX#2 from *Z. mays* is most likely a second ACX1. As with the other ACX1 enzymes, ACX#2 is approximately 200 amino acids longer than ACX4. In the phylogenetic tree ACX#2 also grouped into Clade II. ACX#3 from *P. andersonii* is most similar to ACX4, and indeed is closely localised to AtACX4 in Clade I, as with ACX#19 from *T. orientale*. The *G. nicotianae* ACX is harder to place, being shorter than all other selected ACXs, and with no similarity at the well-conserved C-terminus. We aligned the ACX#20 polypeptide sequence against ACX1, ACX2, ACX3 and ACX4 (WT) from *A. thaliana*. From this alignment, GnACX most closely resembled AtACX4, but the similarity was less than that for the other ACXs we annotated this way (Figure 7-6). This might be expected based upon the phylogeny shown in Figure 7-4, where ACX#20/GnACX is not in any clade.

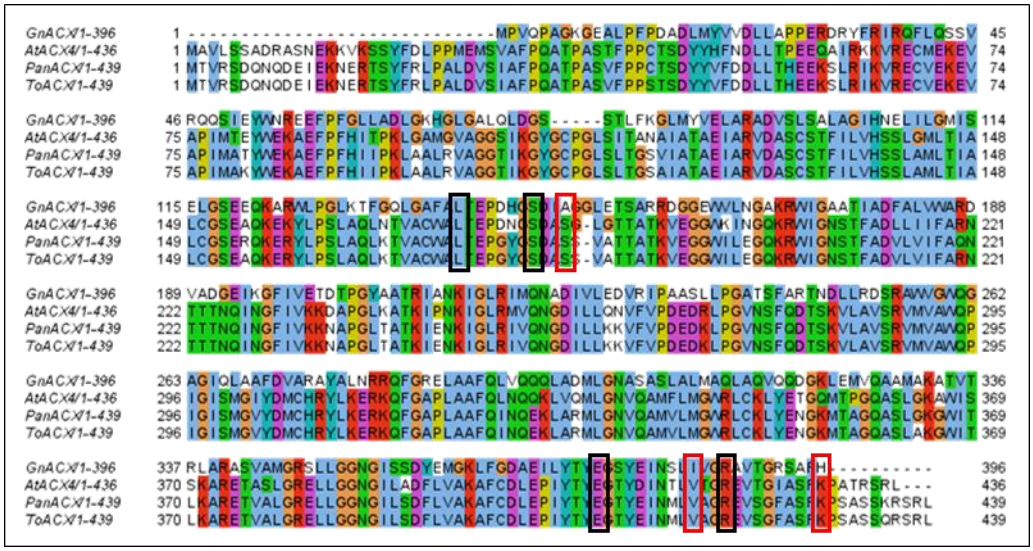


Figure 7-6: Sequence alignment of unknown ACX isoforms

Sequences aligned using ClustalO and visualised using Jalview 2.11.2.4. GnACX = *G. nicotiane* ACX (VXC33980.1), AtACX4 = *A. thaliana* ACX4 (MCC), PanACX = *P. denitrificans* ACX (PON51135.1), and ToACX = *T. orientale* ACX (PON92218.1). Key residues outlined in black if the same in all genes, and red if residue varies only in GnACX.

GenBank protein sequences from ACX enzymes in Table 7-2 were translated into nucleotide sequences and codon optimised these for expression in an *E. coli* host chassis (Supplementary A.2:). I also removed any *Eco*RI, *Afl*II, *Esp*3I and *Not*I sites from the sequences by modification of wobble bases, as these sites are required for the Golden Gate cloning approach to construct a library of BMA production strains. I designed additional flanking regions of nucleotide sequence to allow for this cloning step. We did not remove the C-terminal SRL peroxisomal signalling peptides from the ACX sequences as this was also present in the original AtACX4 sequence, as included in pMAE-4 and pCAN-4 [125].

Code	Organism	Isoform	Clade	C-term	GenBank ID
ACX#1	*Spinacia oleracea*	ACX1	II	VRTSRL	XP_021856262.1
ACX#2	*Zea mays*	ACX1	II	LKLSRL	ACF78566.1
ACX#3	*Parasponia andersonii*	ACX4	Ib	SKRSRL	PON51135.1
ACX#4	*Spinacia oleracea*	ACX3	IV	PESYCV	KNA24566.1
ACX#5	*Zea mays*	ACX1(2)	II	LKLSRL	AQL02798.1
ACX#6	*Apostasia shenzhenica*	ACX4	Ib	DKASRL	PKA62434.1
ACX#7	*Spinacia oleracea*	ACX4	Ib	AARSRL	XP_021855534.1
ACX#8	*Zea mays*	ACX4	Ib	PAKARL	ONM29903.1

Code	Organism	Isoform	Clade	C-term	GenBank ID
ACX#9	*Candida maltosa*	ACX4	III	AAILSK	CAA29901.1
ACX#10	*Vigna radiata*	ACX1(2)	II	LRNARL	XP_014521199.1
ACX#11	*Yarrowia lipolytica*	ACX3(1)	III	CELDEE	CAA04661.1
ACX#12	*Vigna radiata*	ACX3(2)	IV	SWSSQL	XP_022633883.1
ACX#13	*Candida tropicalis*	ACX4	III	AAILSK	AAA34322.2
ACX#14	*Vigna radiata*	ACX4 X2	Ib	AQRSRL	XP_014516782.1
ACX#15	*Populus alba*	ACX4	Ib	SKRSRM	TKS13357.1
ACX#16	*Vigna radiata*	ACX4 X1	Ib	AQRSRL	XP_022641792.1
ACX#17	*Arachis hypogaea*	ACX4	Ib	QKRSRL	QHO54153.1
ACX#18	*Zea mays*	ACX1(1)	II	LKLSRL	PWZ07906.1
ACX#19	*Trema orientale*	ACX4	Ib	SQRSRL	PON92218.1
ACX#20	*Glutamibacter nicotianae*	ACX4-like	Ia	GRSAFH	VXC33980.1

Table 7-2: Acyl-CoA oxidase library enzymes in this work

Isoform annotations shown in blue represent ones annotated from sequence and alignment data analysed within this thesis. GenBank IDs given for polypeptide sequences where possible.

All ACX genes were synthesised by Twist Bioscience, cloned into a pET-21(+) plasmid vector (Figure 8-2B), and were received as freeze dried samples in a 96-well microtiter plate. On arrival the plasmids were resuspended and then transformed into *E. coli* DH5α (Figure 7-7).

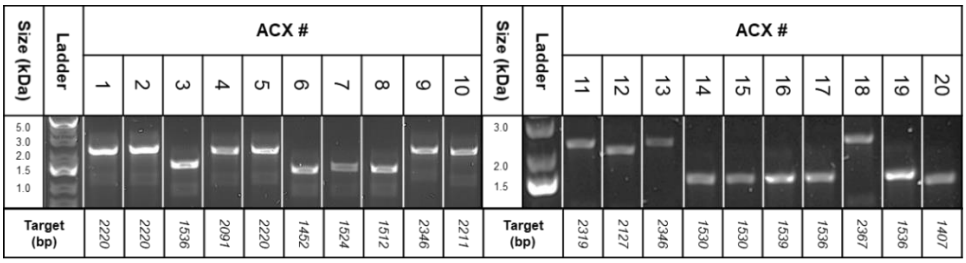


Figure 7-7: Colony PCR of strains containing ACX gene library

ACXs amplified from pOX-3 using primers T7_Fwd and T7_Rev. Colony PCR reactions were loaded onto a 1% agarose gel in TAE buffer, which ran at 80 V for 40 min. 1 kb plus DNA ladder (Thermo). Expected product sizes are shown in *italics* and correspond to those shown in Table 7-2.

7.2.2 Selecting alcohol acyltransferase enzymes for BMA production

The availability of previously obtained MCUK data screening for alternative AATs provided a baseline for identification of new transferase enzymes. In this research

they provided alcohols to the pulp of several different fruit to observe any MMA, ethyl methacrylate (EMA) or BMA synthesis (Table 7-3) (MCUK, unpublished).

Fruit	BMA yield (%)	Selectivity (MMA, EMA, BMA)
Banana	13.0	BMA
Strawberry	0.4	BMA
Kiwi	0.5	BMA
Lemon	0.0	
Apple	1.0	MMA, **EMA**, BMA
Tomato	0.0	
Melon	0.6	BMA
Pear	1.0	BMA
Lime	0.0	
Papaya	1.5	MMA, EMA, **BMA**
Avocado	1.5	MMA, EMA, **BMA**
Grape	0.0	
Blueberry	0.4	BMA

Table 7-3: Butyl methacrylate production from fruit pulp

Yield of BMA measured in fruit pulp provided with M-CoA and either methanol, ethanol, or butanol. = must include in enzyme library, = include if appears in other searches, = exclude from enzyme library. BMA = butyl methacrylate, MMA = methyl methacrylate, EMA = ethyl methacrylate. Esters shown in **bold** indicate the major product when more than one ester is produced. Table adapted from data obtained by MCUK.

The most promising result was a 13% BMA yield from banana pulp, with no EMA or MMA formation. Strawberry pulp was also specific to BMA, although only a 0.4% yield was recorded. Apple pulp produced 1% BMA, 5% EMA and 2% MMA (MCUK, unpublished). Promising results were also recorded for pear, papaya, and avocado pulp. No activity was observed while using grape, lemon, lime, or tomato pulp. The high specific BMA yield from banana pulp identifies *Musa sapientum* or *Musa acuminata* as potential sources of a superior AAT, and as a result we preferentially selected AATs from banana where these were available on the assumption that there may be many AATs present in banana, as with other fruit species that produce a cocktail of volatile esters. Apple AAT has been extensively investigated by MCUK and forms the basis for the AATm4 currently used in the BMA metabolic pathway, and so was excluded from searches.

In addition to the fruit pulp experiments, MCUK also assayed BMA production from purified AAT enzymes identified from apple, strawberry, wild strawberry, banana, and Atf1/2 from yeast (Table 7-4). In all cases the substrate preference of AAT was in formation of the acetate ester. The highest BMA forming activity was observed in the *M. pumila* (apple) AAT. Unfortunately, the banana AAT selected by MCUK was likely not the AAT responsible for the BMA formation recorded from fruit pulp, and so no BMA or BA formation was observed. SAAT and VAAT, both from strawberry, did produce BMA but with less specificity than MpAAT. No other kinetic data was recorded. Only BA formation was observed from Atf1/2, and BA was formed at a higher rate than the BA formation from any of the other purified enzymes.

	Activity (mU g^{-1})	
Enzyme	Acetate	Methacrylate
MpAAT	1410	481.5
SAAT	1770	22.8
VAAT	4990	10.2
BanAAT	0	0
Atf1	4240	0
Atf2	6750	0

Table 7-4: Butyl methacrylate forming activity of purified AATs

Ester forming activity from purified AAT enzymes provided with butanol and either acetyl-CoA or methacrylyl-CoA. MpAAT = *M. pumila* AAT, SAAT = *F. ananassa* AAT, VAAT = *F. vesca* AAT, BanAAT = banana AAT, Atf1 = *S. cerevisiae* acyltransferase 1, Atf2 = *S. cerevisiae* acyltransferase 2.

As a result of the MCUK data the AAT library search focussed on identifying any other available banana AAT sequences, as well as including the SAAT and VAAT AATs which have been shown to be active on butanol and M-CoA *in vitro* (MCUK, unpublished). Kiwi, melon, pear, and blueberry also appear to be viable candidates as all formed BMA exclusively when compared to MMA and EMA. The productivity of fruit pulp from papaya and avocado, at 1.5%, was the second highest for all tested fruits. However, the specificity of the two fruit was not as competitive, also forming MMA and EMA (MCUK, unpublished). Atf1/2, lemon, tomato, lime, and grape AATs were excluded from the selected AAT library as MCUK could not detect any methacrylate ester using extracts prepared from fruit pulp. With the factors introduced by the MCUK

data taken into account, the initial search began by BLAST searching the AATm4 protein sequence against the non-redundant protein databank (Table 7-5).

Common name	Organism	Score	Hits
-	Synthetic construct	913	1
Pear	*Pyrus ussuriensis*	868	1
	Pyrus x bretschneideri	848	8
	Pyrus ussuriensis x Pyrus communis	845	7
	Pyrus communis	839	1
Loquat	*Rhaphiolepis bibas (Eriobotrya japonica)*	836	1
Persimmon	*Diospyros kaki*	832	1
Cherry	*Prunus avium*	676	4
Peach	*Prunus persica*	676	8
Almond	*Prunus dulcis*	673	8
Plum	*Prunus mume*	672	5
Apricot	*Prunus armeniaca*	667	7
Cherry	*Prunus yedoensis var. nudiflora*	663	7
China rose	*Rosa chinensis*	643	23
Wild strawberry	*Fragaria vesca subsp. vesca*	628	4
Charcoal tree	*Trema orientale*	594	1
	Parasponia andersonii	579	1
Cork oak	*Quercus suber*	566	3
Clementine	*Citrus clementina*	563	2
Valley oak	*Quercus lobate*	562	2
Mandarin	*Citrus unshiu*	562	1
Sour gum	*Nyssa sinensis*	562	2
Sweet orange	*Citrus sinensis*	560	1
Mulberry	*Morus notabilis*	556	4
Pitcher plant	*Cephalotus follicularis*	555	1
Maple	*Acer yangbiense*	554	1
Chestnut	*Castanea mollissima*	553	1
Hornbeam	*Carpinus fangiana*	553	1
Walnut	*Juglans regia*	553	2
Date	*Ziziphus jujube*	552	5
Walnut	*Juglans microcarpa x Juglans regia*	550	1
Goldthread	*Coptis chinensis*	545	1
Buckthorn	*Rhamnella rubrinervis*	545	4
Golden kiwi	*Actinidia chinensis var. chinensis*	541	1
Petunia	*Petunia x hybrid*	541	3
Pecan	*Carya illinoinensis*	540	5

Table 7-5: AAT pBLAST search results

The top hit had a% identity of 97.6% and was labelled as a synthetic construct expressed in *C. acetobutylicum* [139]. On inspection, the AATs expressed in this experiment were the same apple AAT used as a source for AATm4 and the

strawberry AAT from *F. ananassa* known as SAAT. SAAT produced 50 mg L^{-1} butyl butyrate with an 85% selectivity when expressed in *C. acetobutylicum*. The next 17 hits all originated from pear, with *Pyrus ussuriensis* AATs most commonly appearing. Although the ester forming profile of *P. ussuriensis* has been investigated, no data could be found assaying the products of any individual AAT [219]. No hits were found for banana, melon or blueberry AATs as identified by MCUK. However, one AAT for Golden kiwifruit (*Actinidia chinensis*) was identified with a percent identity of 59.1%. The only data associated with this entry comes from an RNAseq and BLASTp annotation of the *A. chinensis* genome. A large number of hits for China rose were identified from the pBLAST search, however these were redundant annotations of AAT1 from *Rosa chinensis*. All but one of these hits, as well as any redundant hits for other organisms, were removed from the results and the remaining top 100 hits were exported as .fasta files for later evaluation.

A search for the apple AAT protein sequence *vs.* Musa sp. (banana) yielded only four results. The first, for *Musa sapientum* corresponds to the BanAAT that demonstrated no activity towards BMA when investigated by MCUK [210]. The second was a partial protein sequence from *Musa acuminata*. The third and fourth sequences were for *Musa balbisiana* and *M. acuminata*. The *M. acuminata* sequence was automatically annotated by GenBank, while the *M. balbisiana* AAT is listed only a sequence annotated during genome-wide sequencing experiments.

Searching BRENDA for EC 2.3.1.84 yielded several AATs with associated kinetic data. Seven unique AATs were returned from the search; these were AATs from *Cymbopogon martinii*, *F. ananassa*, *Hanseniaspora valbyensis*, *M. sapientum*, *Petunia hybrida*, *Rosa hybrida*, and *S. cerevisiae*. Once again the *M. sapientum* AAT is the same as the inactive BanAAT, and SAAT is the AAT identified from *F. ananassa*. The AATs from *S. cerevisiae* are Atf1/2, which are inactive on butanol and M-CoA as with the banana AAT. The only -CoA substrate for which BRENDA contains kinetic data is acetyl-CoA. The *R. hybrida* RhAAT1 was expressed in *E. coli* and

assayed for activity on several alcohol substrates, including butanol, *in vitro* [220]. Coniferyl alcohol acyltransferase (CFAT) from *P. hybrida*, as its name suggests, demonstrated highest activity utilising Coniferyl alcohol as well as other benzene-containing alcohols. In contrast CFATs activity appeared to decrease dramatically with linear alcohols < C8 in length [221]. No sequence data was available for the *H. valbyensis* and *C. martini*.

Searching for reviewed entries for EC 2.3.1.84 gave only 5 proteins. Once again two of these results were Atf1 and Atf2 from *S. cerevisiae*. Two additional transferases were identified from *S. cerevisiae*, Eht1 and Eeb1.Both enzymes are annotated as acyl-CoA:ethanol-O-acyltransferases and have been expressed in and purified from *E. coli* for *in vitro* assay [222]. Formation of ethyl butanoate, ethyl hexanoate and ethyl octanoate by Eht1 and Eeb1 were tested, with both enzymes capable of forming all products. Eht1 demonstrated a substrate preference for the shorter chain butyryl-CoA [222]. The final hit was for CFAT from *P. hybrida* once again. The promising data for Eht1 activity prompted us to search for other wax ester synthases using pBLAST and a literature search. This resulted in the identification of 7 additional wax synthase enzymes (Table 7-6).

Common name	Organism	Name	Reference
Grape	*Vitis vinifera*	WS-1	[223]
	Vitis vinifera	WS-2	[223]
	Vitis vinifera	WS-3	[223]
	Euglena gracilis	WS	[224]
	Saccharomyces cerevisiae	Eebl	[225]
	Saccharomyces cerevisiae	Ehtl	[225]
Petunia	*Petunia hybrida*	PhWS	[226]
	Thraustochytrium roseum	WSD4	[10]
	Saccharomyces cerevisiae	Atf2	[227]
	Saccharomyces cerevisiae	Atf1	[228]
Sweetcorn	*Zea mays*	Glossy2	[229]

Table 7-6: Wax synthase candidates for AAT screening

Expanding the UniProt search to include unreviewed entries gave 73 results, of which 46 were unique entries. Only 3 results from this search were not already listed or did not appear to be incorrectly annotated. These were CmAAT2 from melon, a truncated AAT from prickly pear, and a *Vasoncellea cundimarcensis* (*Vasconcellea pubescens*) AAT1. *V. pubescens* AAT1 was expressed using *E. coli* TOP10 and was assayed *in vitro* for alcohol acyltransferase activity [163]. Additionally, activity of VpAAT1 with acetyl-CoA, butanoyl-CoA and hexanoyl-CoA was determined experimentally. VpAAT1 produced butyl butanoate and ethyl butanoate with an activity of 526 pkat mg^{-1} and 83 pkat mg^{-1}, respectively. Interestingly, this productivity was also higher than VpAAT1 activity forming either butyl acetate, at 1 pkat mg^{-1} or ethyl acetate, of which only trace amounts were detected [163].

Database results for AATs with kinetic or experimental data produced limited results. Reliance purely on sequence similarity using BLAST in our enzyme selection would not be a robust strategy, as very little connection has been found between amino acid sequence and AAT specificity or activity in practice [137]. Therefore I also carried out a rigorous literature search in an attempt to identify any further enzymes with demonstrated AAT activity (Table 7-7). The plethora of AAT isoforms present in fruit and some other plants with diverse functionalities meant that a list of 41 AATs was identified. As illustrated in Table 1-1, the large majority of literature relating to this concerns assay of AATs acetyl-CoA as the acyl- donor. Additionally, many of the results from literature represent AATs that use longer chain, more complex substrates that are implicated in the production of medically relevant molecules.

Organism	Name	Ref	Organism	Name	Ref
Actinidia deliciosa	AT1	[138]	*Lavandula angustifolia*	LaAAT1	[162]
Actinidia eriantha	AT9	[138]	*Lavandula angustifolia*	LaAAT2	[162]
Cucumis melo	CmAAT1	[230]	*Papaver somniferum*	SalAAT	[231]
Cucumis melo	CmAAT4	[137]	*Rauvolfia serpentina*	VS	[232]
Cucumis melo	CmAAT3	[137]	*Solanum pimpinellifolium*	SpAAT2	[233]
Fragaria vesca	VAAT	[135]	*Clarkia brewerii*	CbBEAT	[214]
Citrus limon	LAAT	[135]	*Arabidopsis thaliana*	CER2	[234]
Mangifera indica	ManAAT	[135]	*Lupinus albus*	HMT	[235]

Solanum lycopersicum	TomAAT	[135]	*Catharanthus roseus*	DAT	[236]
Nicotiana tabacum	HQT	[237]	*Capsicum annum*	Pun1	[238]
Clarkia brewerii	CbBEBT	[239]	*Salvia splendens*	Ss5MaT2	[240]
Nicotiana tabacum	NtBEBT	[239]	*Taxus cuspidata*	DBAT	[241]
Arabidopsis thaliana	CHAT	[242]	*Taxus cuspidata*	DBBT	[241]
Arabidopsis thaliana	AtHCT	[243]	*Taxus cuspidata*	TAT	[241]
Solanum pennelli	SpAAT1	[244]	*Taxus canadensis*	DBNBT	[241]
Nicotiana tabacum	NtHCT	[245]	*Taxus cuspidata*	BAPT	[241]
Vitis labruscaXvinifera	VvAAT1	[246]	*Vitis labrusca*	AMAT	[247]
Larrea tridentata	CAAT1	[248]	*Dianthus caryophyllus*	HCBT2	[249]
Larrea tridentata	CAAT2	[248]	*Avena sativa*	AsHHT1	[250]
Petunia hybrida	PhWS	[226]	*Physalis peruviana*	PhpAAT1	[251]
Catharanthus roseus	MAT	[252]			

Table 7-7: AATs from literature search.

I constructed a phylogenetic tree for the AAT candidates (Figure 7-8) as carried out for the ACX library. This tree allowed separation of the AAT sequences into several clades, informed by the genome wide analysis for fruit AATs carried out by Liu *et al*. [253]. Due to the evidence that sequence divergence does not necessarily prevent similar substrate specificities, I aimed to include a range of AATs from all clades in the final transferase library.

Clade I is by far the largest group, comprising 37.7% of the 69 total enzymes. I have divided this into two sub-sections: Clade Ia consists of the enzymes with the highest similarity to AATm4, predictably including the *M. domestica* apple AAT. In particular two AATs from pear, *Pyrus ussuriensis* PuAAT and *Pyrus communis* PcAAT, as well as *E. japonica* EjAAT1 and *D. kaki* DkAAT1 bear the closest similarity to AATm4. Clade Ib contains all of the benzyl alcohol O-benzoyltransferase (BEBT) entries. However this Clade also contains HMT from *Lupinus albus,* which is identified as an alkaloid acyltransferase [235], CHAT from *A. thaliana* which is a cis-3-hexen-1-ol transferase [239], and the AATs from melon and tomato. Interestingly, the two hypothetical banana AAT sequences (MbAAT and MaAAT) also reside in Clade Ib instead of possessing more similarity to the BanAAT tested by MCUK, which is in Clade IIa.

Five AATs in Clade I were previously investigated by MCUK and showed no activity forming BMA from fruit pulp. These were the grape AATs VVAAT1 and AMAT, as well as all tested tomato AATs: SpAAT2, TomAAT, and SpAAT1. All of these enzymes were excluded from the final list of AAT candidates.

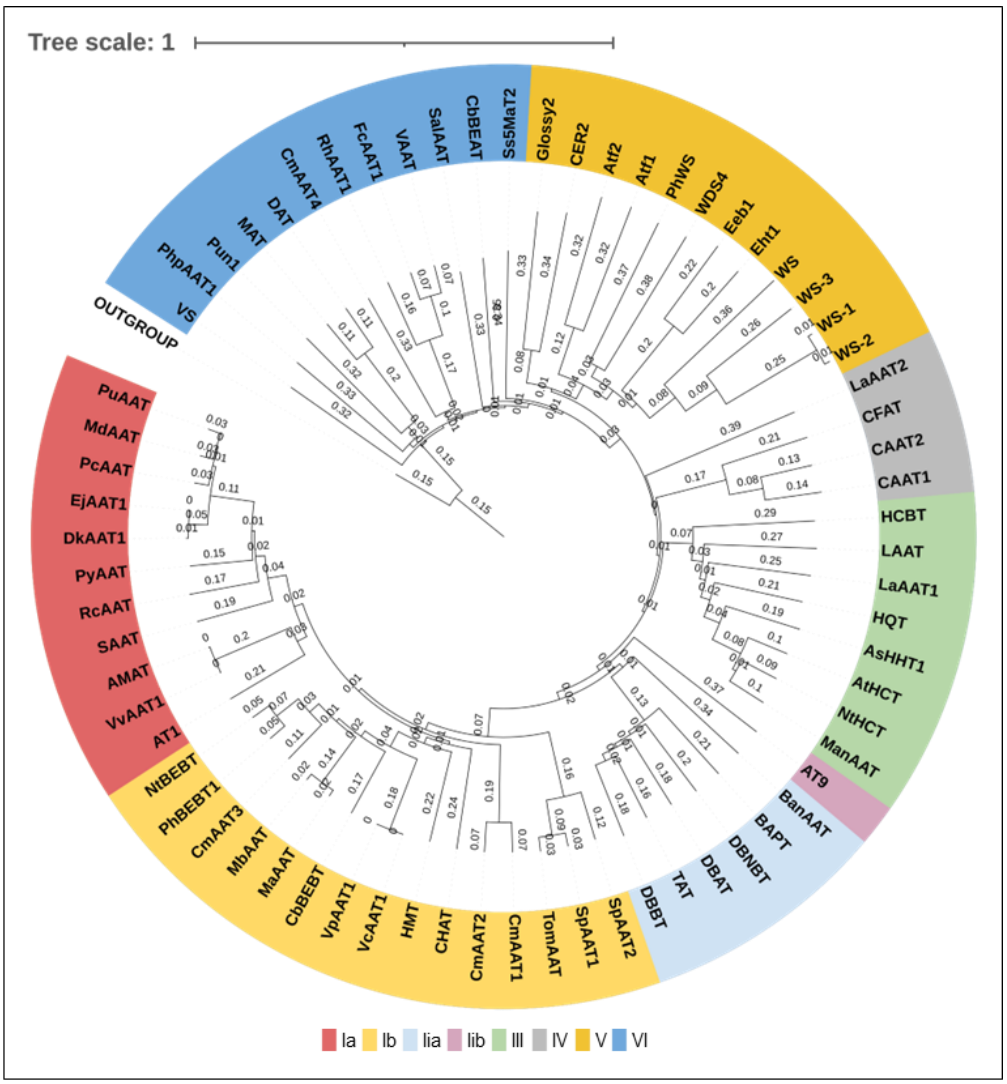


Figure 7-8: Phylogeny of AATs identified from literature and database searching.

AAT sequences represented by source organism and enzyme name, full list of sequences included in Appendix B.2: Entries divided into clades based on sequence similarity. Outgroup is a BAHD acyltransferase from *Pistacia vera*. (XP_031260545.1) Multiple sequence alignment generated using ClustalO, tree visualised using Interactive Tree of Life (iTOL). Branch lengths shown on tree.

Clade IIa contains terpenoid acyltransferase enzymes which are all from the genus Taxus, or coniferous shrub [254]. These AATs are, as indicated by the genus name, thought to be responsible for several stages in Taxol biosynthesis [255]. Therefore

their close similarity compared to the other entries is somewhat expected. TAT, DBAT and DBBT from *Taxus cuspidata* were all identified by Walker *et al.* and all act on long chain Taxol precursors [241]. However, Clade IIa also contains a group of terpenoid acyltransferases localised by both sequence and function, and BanAAT, which is classed as an alcohol acyltransferase. I randomly selected one enzyme from each Clade and aligned these against AATm4 to produce a multiple sequence alignment (Figure 7-9). In this, the selected Clade II enzyme, DBBT, diverges in similarity more than Clade Ia/b in the central 160-260 amino acid region of the polypeptide chain.

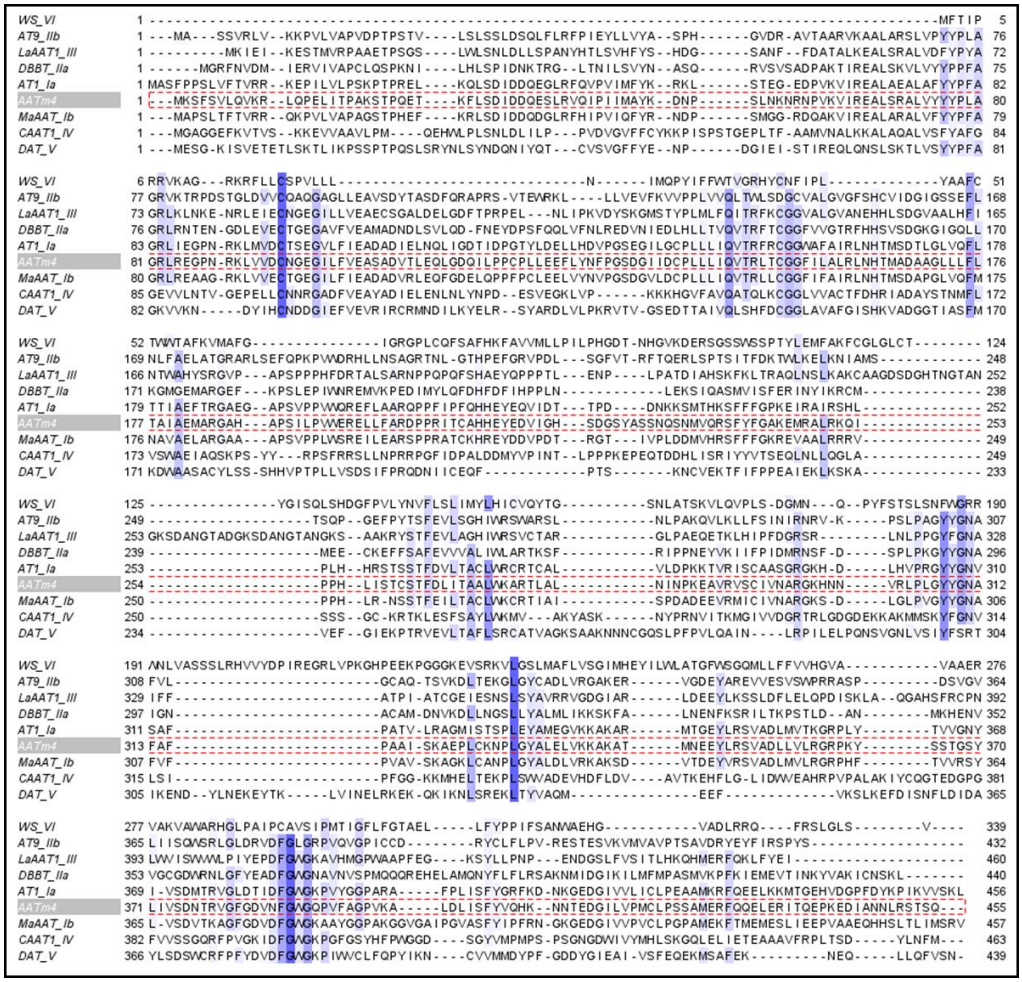


Figure 7-9: Sequence similarity of transferases from each phylogenetic clade

Sequences aligned using ClustalO and visualised using Jalview 2.11.2.4. WS_VI = *E. gracilis* WS (Clade VI), AT9_IIb = *A. eriantha* AT9 (Clade IIb), LaAAT1_III = *L. angustifolia* AAT1 (Clade III), DBBT_IIa = *T. cuspidata* DBBT (Clade IIa), AT1_Ia = *A. eriantha* AT1 (Clade Ia), MaAAT_Ib = *M. acuminata* AAT (Clade Ib), CAAT1_IV = *L. tridentata* AAT1 (Clade IV), DAT_V = *C. roseus* CAT (Clade V). AATm4 (MCC) sequence outlined in red. Residues coloured according to percentage identity.

AT9 from *Actinidia eriantha* in Clade IIb bears very little homology in comparison to the other entries, and is the only enzyme in this sub-clade. The only information regarding the substrate specificity of AT9 suggests that it can easily accept acetyl-CoA and benzyl alcohol substrates [256]. However, the sequence divergence of AT9 from all other AATs in the results makes it an interesting candidate for further analysis in the AAT library.

The AATs in Clade III appear to broadly encompass those involved in the lignin pathway in plants. For example HCT from *A. thaliana* and *Nicotiana tabacum*. HCTs accept 4-hydroxycinnamoyl-CoA as a substrate, but have also been found to utilise caffeoyl-CoA and coumaroyl-CoA [243, 245]. This grouping aligns with the classification of a clade of quinic/shikimic acid AATs as identified by Wang *et al.* [254]. As expected therefore, Clade III also contains both AATs from *Lavandula augustifolia*.

Clade IV only contains three AATs. CFAT from *P. hybrida* preferentially uses coniferyl alcohol and acetyl-CoA to form coniferyl acetate during isoeugenol biosynthesis [221]. BEBT, also from *P. hybrida*, catalyses another step in the same isoeugenol biosynthesis pathway but bears more sequence similarity to the enzymes in Clade Ib. The remaining two enzymes are CAAT1 and CAAT2, both from *Larrea tridentata*. Similar to CFAT, they are both implicated in eugenol biosynthesis, and have been shown to accept a wide variety of alcohol substrates including benzyl alcohol, cinnamoyl alcohol and coniferyl alcohol [257].

Clade V appears to contain a more diverse range of transferase functions, largely because the transferases therein have largely been examined for endogenous activity. For example, Ss5MaT2 is an anthocyanin malonyl transferase from *Salvia splendens* [240]. Vinorine synthase from *Rauvolfia serpentina* is part of the biosynthetic pathway for ajmoline, an anti-arrhythmic molecule [232]. CbBEAT from *C. brewerii* is another benzyl alcohol O-acetyltransferase [162]. Most divergent from Clade V are CER2 and Glossy2. CER2, from *A. thaliana*, is responsible for wax biosynthesis, and Glossy2 from *Z. mays* is more specifically annotated in the

biosynthesis of C32 waxes [234]. As Clade VI contains the wax synthases, the short distance between CER2 and Glossy2 and the next grouping is to be expected. In Clade V the only AATs that have been tested by MCUK are VAAT from *F. vesca* and CmAAT4 from *C. melo*. Both were able to produce BMA *in vitro*.

As mentioned, Clade VI contains the bulk of the wax synthases. From MCUK data we know that Atf1 and Atf2, as well as WS- 1, 2 and 3 from *Vitis vinifera* do not produce BMA. That leaves WS from *E. gracilis*, Eht1 and Eeb1, WDS4 from *T. roseum,* and PhWS from *P. hybrida*. WS produces medium chain wax esters, with a broad substrate specificity that tends towards myristic acid and palmitoleyl alcohol [224]. Despite its similarity to Eht1/Eeb1, PhWS does not demonstrate any DGAT activity and to date has only been shown to act as a wax synthase [224].

Perhaps reflected by the variety of substrates accepted by the transferases reported in the literature, there is very little sequence identity between the different Clade entries and AATm4 (Figure 7-9). Although characteristic features such as the DFGWG motif appear to be conserved in almost all cases. As previously mentioned, the sequences of transferases are reported to often be disconnected from substrate selectivity, at least to our current level of understanding. This sequence divergence makes it a challenging task to attempt to select AATs based on sequences alone.

Informed by this data, I selected a total of 26 AATs for our screening efforts with the aim of including the widest possible coverage of phylogeny and reported activities (Table 7-8). This includes all AATs identified as BMA forming by MCUK. I also included Eht1 and Eeb1 for their activity on butyryl-CoA. Despite the less than favourable data for some groups of AATs, for example AT9, I decided to include at least one candidate from each clade to cover the full range of AATs identified. VS was selected due to the availability of one of only a few crystal structures for the BAHD acyltransferases. I also selected several AATs closely related to MdAAT, including DkAAT1 and VpAAT1. Both new banana AAT sequences were included.

Code	Organism	Gene	Clade	HXXXD(G)	DFGWG	GenBank ID
AAT#1	*Actinidia deliciosa*	AT1	Ia	HTMSDT	DFGWG	AIC83790.1
AAT#2	*Actinidia eriantha*	AT9	IIb	HCVIDG	DFGLG	AIC83789.1
AAT#3	*Larrea tridentata*	CAAT1	IV	HRIADA	DFGWG	AHA90802.1
AAT#4	*Arabidopsis thaliana*	CER2	V	RCNDSG	EINGG	AAM64817.1
AAT#5	*Arabidopsis thaliana*	CHAT	Ib	HTMTDG	DFGPWG	AAN09797.1
AAT#6	*Cucumis melo*	CmAAT1	Ib	HTMADG	DFGWG	CAA94432.1
AAT#7	*Cucumis melo*	CmAAT4	V	HKLIDA	DFGWG	AAW51126.1
AAT#8	*Catharanthus roseus*	DAT	V	HCNDDG	DFGWG	AAC99311.1
AAT#9	*Taxus cuspidata*	DBBT	IIa	HSVSDG	DFGWG	Q9FPW3.1
AAT#10	*Diospyros kaki*	DkAAT1	Ia	HTMCDA	NFGWG	AKE98481.1
AAT#11	*Saccharomyces cerevisiae*	Eeb1	VI	-	-	Q02891.1
AAT#12	*Saccharomyces cerevisiae*	Eht1	VI	-	-	P38295.1
AAT#13	*Eriobotrya japonica*	EjAAT1	Ia	HTMCDA	NFGWG	AHC3222.2
AAT#14	*Fragaria chiloensis*	FcAAT1	V	HKLIDG	DFGWG	ACT82247.1
AAT#15	*Lavandula angustifolia*	LaAAT1	III	HHLSDG	DFGWG	DQ886904.1
AAT#16	*Musa acuminata*	MaAAT	Ib	HTMSDA	DFGWG	XP_009388282.1
AAT#17	*Mangifera indica*	ManAAT	III	HHAADG	DFGWG	CAC09378.1
AAT#18	*Musa balbisiana*	MbAAT	Ib	HTMSDA	DFGWG	THU42907.1
AAT#19	*Pyrus communis*	PcAAT	Ia	HTMCDA	NFGWG	AAS48090
AAT#20	*Petunia hybrida*	PhBEBT1	Ib	HTMSDA	DFGWG	AAU06226.1
AAT#21	*Pyrus ussuriensis*	PuAAT	Ia	HTMCDA	NFGWG	AJD18611.1
AAT#22	*Rosa hybrid cultivar*	RhAAT1	V	HKINDG	DFGWG	AAW31948.1
AAT#23	*Fragaria ananassa*	SAAT	Ia	HTICDA	DFGFG	AAG13130.1
AAT#24	*Vasconcellea pubescens*	VpAAT1	Ib	HTMSDA	DFGWG	ACT82248.1
AAT#25	*Rauvolfia serpentina*	VS	V	DCNDSG	DFGWG	CAD89104.2
AAT#26	*Euglena gracilis*	WS	VI	-	NFWG	ADI60058.1

Table 7-8: AATs selected for library screening

Complete list of alcohol acyltransferases (AATs) used in library. Gene names annotated as described on either GenBank or UniProt.

Polypeptide sequences were obtained for each of the enzymes shown in Table 7-8. These were translated to nucleotide sequences, codon optimised for *E. coli* and *Esp*3I, *Eco*RI, *Afl*II and *Spe*I sites removed as done for the ACX enzyme library (Supplementary A.3:). The AAT library was synthesised by Gene Universal and cloned into a modified pUC57-Kan plasmid, an *Esp*3I site near the origin of replication was mutated by Gene Universal to remove it from the backbone. All AAT genes were received as freeze-dried samples in tubes, on arrival these were re-suspended in dH_2O before transformation into *E. coli* DH5α (Figure 7-10).

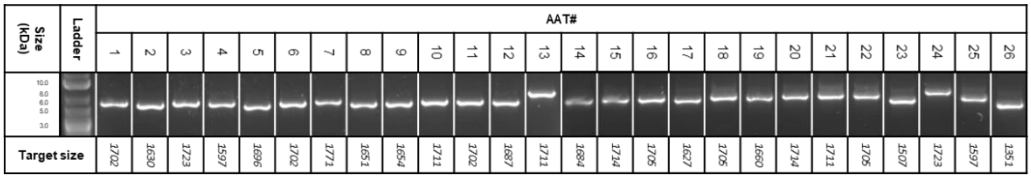


Figure 7-10: Colony PCR for synthesised AAT gene transformation into E. coli

AATs amplified from pCAN-6 using primers GG-FIX-F and GG-FIX-R. Colony PCR reactions were loaded onto a 1% agarose gel in TAE buffer, which ran at 80 V for 40 min. 1 kb plus DNA ladder (NEB). Expected product sizes are shown in *italics* and correspond to those shown in Table 7-8.

7.3 Discussion

7.3.1 ACX selection

Selections for ACXs to be used in screening were primarily based upon the limited data available about experimentally verified accepted ACX substrates. Where possible, I tried to select for ACXs with activity on substrates as similar to IB-CoA as possible. As no enzymes other than AtACX4 have been reported with IB-CoA as a substrate, we can only infer that sequence similarity and acceptance of short to medium chain -CoA substrates may translate to IB-CoA utilisation *in vivo*. The primary challenge with acyl-CoA oxidase enzymes is that while there is a wealth of data on select candidates, the number of ACXs that have been identified and studied remains relatively low. This reduced the strength of the bioinformatic approach.

It is reported that there is some overlap between substrates for the various isoforms of ACX [124]. Therefore, it seemed sensible to include a number of candidates from other entries, in particular ACX1 and ACX3. No ACX2 enzyme sequences were available with any associated kinetic data. Although literature reports ACX1 enzymes with activity on substrates similar to IB-CoA, the difference in protein quaternary structure, namely the formation of a homodimer instead of a homotetramers, may drastically alter the binding activity of ACX1s compared to ACX4. The ACX2 and ACX3 classes of oxidase diverge far less in length and structure. However, the data available about them is comparatively low. This makes it hard for us to accurately predict whether there is any chance of successful IB-CoA activity in these classes of oxidase. A benefit to assembling an ACX library as opposed to the AAT library is that

the active site of the oxidases has been well characterised. This increases the feasibility of also applying a site-directed mutagenesis approach to reducing product inhibition in future.

Introduction of ACADs, although likely tricky to engineer into *E. coli*, is another alternative option that is valuable to explore. Since the lack of endogenous co-factor regeneration in *E. coli* hinders the introduction of ACADs, other groups are currently also exploring expression of BMA biosynthesis in alternative hosts. For example, researchers at the University of Edinburgh are creating a comparison of AtACX4 against ACD1 in *S. cerevisiae*, where ACD1 is natively expressed (Florentina Winkelmann, MCUK bio-conference). Alternatively, there is evidence that some ACADs can use molecular O_2 to regenerate their FAD co-factor, for example SCAD from *Megasphaera elsdenii* [121]. This is linked to the positioning of a residue near the active site, which in *M. elsdenii* SCAD is phenylalanine. In O_2 insensitive ACADs, such as *M. elsdenii* MCAD, there is a tryptophan residue at this position, the bulky side chain of which prevents O_2 entering the active site. Engineering an IVD to introduce smaller, more permissive, amino acid residues at this position may be way to introduce an ACAD into *E. coli* without extensive exogenous protein expression.

7.3.2 AAT selection

Identifying viable AAT candidates is a more challenging task than searching for novel ACX enzymes. There is a wealth of transferase sequence data available on online databases, which usually originates from whole genome sequencing literature. However, the BAHD acyltransferases are not nearly as well characterised as oxidases. For example, it is understood that several AATs are likely present in plants, but the exact number of distinct acyltransferases in most organisms is largely unknown. This leaves a wealth of available AAT sequences, but little additional information which could allow us to select for enzymes with favourable substrate preferences or quarternary structures. To date, the only experimentally determined AAT-like crystal structures are for VS and for coffee or sorghum HCT [232, 243].

Generation of a crystal structure for the apple AAT on which AATm4 is based would perhaps inform our ability to identify positive 'hits'. Additionally, although a large number of AATs appear to be active on acetyl-CoA and overall demonstrate a broad substrate specificity, many AATs will not accept an M-CoA substrate, or similar. The challenge comes in narrowing down the broad range of candidate transferases available, with limited information as to their possible efficacy.

Here I have tried to select for a diverse range of sequences, as we cannot accurately predict specificity based on sequence alone, as well as incorporating AATs that have been previously shown to produce BMA in experiments carried out by MCUK. Precedence was given to the two AAT sequences available from banana, as this is the only organism from which fruit pulp produced more BMA than BA. MCUK previously isolated a third banana AAT, BanAAT, and found it was unable to produce BMA. If these new sequences also fail to utilise M-CoA, then a promising next step may be to attempt to isolate the AAT responsible for BMA formation direct from banana fruit pulp.

During the course of this work, MCUK have continued to improve the original apple AAT in an attempt to increase BMA selectivity and overall transferase activity. One way they have done this is by removing cysteine residues from AAT, as the occurrence of surface cysteine residues appears to correlate with increased formation of inclusion bodies, and by extension decreased solubility, in exogenously expressed AATs. As previously mentioned, they have also generated several active site mutants of AAT, including both the AATm4 used here and AATm10j. The majority of new mutants with any increased M-CoA preference came with the caveat of a large increase in acetyl-CoA selectivity, however AATm10j demonstrates a more favourable comparable selectivity for acetyl-CoA *vs.* M-CoA, and almost 3x selectivity for M-CoA *vs.* IB-CoA. Introducing this AAT into a BMA production strain may improve the removal of M-CoA, and thus allow better flux from IB-CoA through AtACX4.

Despite some improvements in AAT selectivity, the relative availability of substrates seems to largely govern acyltransferase activity. Both heterologous expression and structure-guided directed mutations are a significant challenge to successful AAT development, particularly in the formation of inactive aggregates and co-improvements in off-target -CoA activity. A more viable solution with current available knowledge is managing to increase the intracellular M-CoA pool in comparison to acetyl-CoA and IB-CoA. This would have the knock on effect of increasing the selective pressures for AAT to utilise M-CoA, as well as permitting the concentration of substrate to approach AATs K_M value.

7.3.3 Bioinformatics for the accurate prediction of heterologous enzyme properties

In using bioinformatics to select novel enzymes *in silico*, there are certain factors which cannot be predicted that will affect real-world protein expression *in vivo*. For example, pathway prediction is not yet sophisticated enough to predict the problems of context dependence. In selecting our new libraries of enzymes we cannot fully account for potential crosstalk with endogenous metabolic processes, mutations that may occur during transcription, and/or the change or total loss of function that may result from a protein from one organism being expressed under different conditions in a new host chassis [202]. For instance, many of the enzymes I have selected here are originally sourced from plants. All production experiments will be carried out at 37°C to accommodate the favoured growth conditions for *E. coli*. Considering several plants, including *Z. mays* and apple, are conventionally grown at a much lower environmental temperature, some change in function can be expected for proteins sourced from these expressed at higher temperatures.

This being said, tools are increasingly being developed to try and prioritise successful hits and predict enzyme activity from potentially thousands of candidates [213]. When a crystal structure is available, molecular dynamics can be used to simulate reactions *in silico*, while machine learning is increasingly used to predict the sequences of

enzymes that may catalyse a target reaction. For example, a Gaussian process was used by Mellor *et al.* to accurately predict a lower K_M value for heterologous N-acetyl-L-leucine production in *E. coli* [213].

Harnessing our wealth of protein sequence and structure data may be a useful way to directly solve product inhibition issues in future. Here we can hope to remove the high concentration which causes the problem in the first place. Utilising big data is a way to infer related enzymatic functions, but for many problems the most valuable data is that which has been obtained experimentally – this is limited by time and feasibility but will remain the case until *in silico* methods can improve to sufficiently predict real world kinetic values.

With two libraries, one of 20 ACXs and another containing 26 AATs, it is not practical to test each possible combination of enzyme pairs in flask culture. Therefore, I will use a Golden Gate platform to assemble an ACX library and a combinatorial ACX/AAT library. This will provide a valuable resource from which hits can be screened, picked, and then tested in biotransformations. We will test the ACX library separately using this method, as combining each new ACX with AATm4 will allow determination of whether product-inhibition has been relieved in any of the new AATs.

Chapter 8: BMA formation from ACX and AAT libraries

8.1 Section Introduction

8.1.1 A screening approach for oxidase and transferase library searching

Having optimised the plate-based screen insofar as was possible within time constraints, and having assembled a library of ACX and AAT variants for screening, the next task was to interrogate the oxidase and transferase libraries. Any hits from the solid phase IVIS screen that appeared to indicate a higher BMA concentration would be validated to determine whether there is a corresponding BMA production relationship in flask cultures. The IVIS screen's power to discriminate between different BMA production rates and concentrations was at this point untested because of all the baseline strains only **CAN4** showed measurable BMA production. The lower producer strains **LUC0739** and **LUC0848** did not differ sufficiently from **CAN5** to register any change during PZ-BMA detection.

Several of the ACX candidates had already shown activity using butyryl-CoA in the literature. This is significant because butyryl-CoA is relatively similar in structure to IB-CoA, and bears a high sequence similarity to AtACX4. These ACXs were therefore expected to have an increased likelihood of resulting in positive hits during screening. Sequence similarity also holds a greater sway over ACX activity than for AATs, and so oxidase enzymes with similar sequences to AtACX4 were also of particular focus for comparable BMA titres. The AAT screen was slightly more exploratory, as sequence similarities could not be relied upon to identify enzymes with similar substrate specificities. So promising hits were harder to identify, and were produced through a mixture of limited experimental data and a broad catchment of transferase sequences.

Therefore, of the two libraries defined in 7.3.1 and 7.3.2, I began by testing the ACX library in isolation, using only AATm4 to catalyse the transferase step. This allowed testing of the dynamic range of the screen without full combinatorial library assembly, with ACX as the single variable. This approach also makes it a simpler matter to

determine whether any BMA production changes observed during the subsequent combinatorial screen arose because of ACX, AAT or a combination of both. Testing ACXs in isolation also facilitated a more straightforward investigation of any product resistance effect, as this would be obscured by a higher AAT activity potentially present in the novel transferase enzymes, particularly as improvements in the M-CoA to BMA step would have a knock-on impact on ACX inhibition. Only slight changes were likely to be observed through looking at the ACX library, so a large change in BMA production would be a strong indicator of reduced M-CoA sensitivity in ACX. ACX constructs were designed such that if product resistance is identified, the enzymes could easily be His-tagged and assayed using the AtACX Oxygraph assay. In the event that no product resistant ACX was found, AtACX could be supplanted as the experiment baseline, or investigation into the structural differences between product resistant ACX and AtACX4 could be used to inform future engineering efforts. The screen was also used to identify oxidases that do not facilitate BMA synthesis, and therefore may not convert IB-CoA to M-CoA. These were excluded from the combinatorial library assembly. Once this streamlined list of ACXs was generated, we could the proceed with Golden Gate assembly to introduce our more complex AAT library into the combinatorial production strain library.

8.1.2 Golden Gate assembly for generating multi-module combinatorial libraries

Golden Gate assembly relies on the properties of Type IIS enzymes, which were first used carry out a one pot assembly of multiple fragments into one vector in 1996 [258]. Unlike the more widely used Type IS restriction enzymes, Type IIS's cut DNA at a location outside their recognition site. Because of this ability, users can select any DNA sequence as the overhang, sometimes referred to as a signature sequence. If designed carefully, ligating two Type IIS DNA overhangs yields a segment of DNA which does not contain the Type IIS recognition sites [259]. Because any overhang

sequence can be selected, the use of Type IIS restriction digestion can enable directional assembly of more than one DNA module at a time [260].

In 2008, the Golden Gate assembly method was developed by Engler *et al.* which demonstrated the cloning of up to 10 DNA fragments from shuffling libraries into a single Golden Gate vector plasmid [260]. The experimental set-up for this type of assembly is straightforward: Type IIS enzyme, DNA ligase, DNA modules, recipient vector, and buffer are mixed together. This mixture is then incubated at cycles of 37°C and 14°C to enable digestion and ligation, respectively. When the DNA modules correctly ligate into the recipient vector the Type IIS restriction sites are no longer present in the recombinant product, and so future cycles of digestion will not degrade the target plasmid. At the end of the assembly there is a longer 37°C incubation, which removes any re-ligated donor or recipient vector, before the enzymes are inactivated at 65°C [261]. Golden Gate assembly has been used previously to combine a number of alternative parts in a one pot mix, as was done by Rosowski *et al.* to assemble yeast surface display Fab protein libraries, with heavy/light chain variable regions [259]. Golden Gate assembly of an equimolar mixture of variable Fab proteins with *Bsa*I yielded a library of 10^8 clones in *S. cerevisiae* EBY100. More recently, Tong *et al.* used Golden Gate to combinatorially assemble 243 variants of a violacein production library in *Y. lipolytica* Polg (Leu^-), again in a one-pot mix, using *Bsm*BI [262]. Theoretically, generating the ACX only library using Golden Gate assembly using a similar method should be an uncomplicated reaction setup. However as is often the case with Golden Gate assembly, the design and construction of the components required to facilitate a one-pot reaction can pose a significant challenge.

In this chapter, I constructed a Golden Gate vector plasmid, pGGV4, into which ACX (Figure 8-1A+B, dark blue) and AAT (Figure 8-1A+B, yellow) modules can be assembled simultaneously. pGGV4 contains two Type IIS *Esp*3I recognition sites flanking a *sacB* counter selection marker operon (Figure 8-1A+B, pink). *Esp*3I cuts

outside of this region to generate two signature sequences, A and C, which correspond to the N-terminus of the ACX modules and the C-terminus of the AAT modules, respectively (Figure 8-1A+B, light blue, purple). Likewise, each of the regions neighbouring the coding sequences in the bioinformatic libraries have *Esp*3I recognition sites that cut *inside* at two flanking signature sequences. These are A and B for ACX (Figure 8-1A+B, purple, green), and B and C for AAT (Figure 8-1B, green, light blue). When assembling the ACX library, the AAT module originated from AATm4 PCR amplified from pCAN-4. The PCR product of this reaction contains the same flanking *Esp*3I, and B and C signature sequences as were added to pCAN-6 (Figure 8-1A). PCR amplified AATm4, pGGV4, and the library of pOX3-ACX#*n* plasmids were then mixed individually to carry out 20 assembly reactions. Subsequent assembly of combinations of new ACXs and new AATs was more complex.

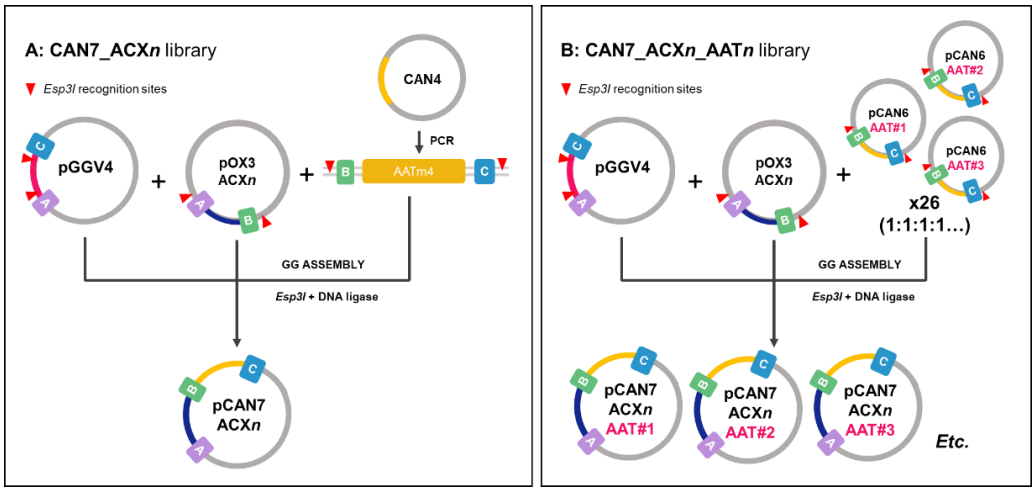


Figure 8-1: ACX and combinatorial assembly strategy using pGG vector

A: Assembly strategy for ACX library production strains. pGGV4, any single pOX3_ACX plasmid, and AATm4 PCR amplified to contain *Esp*3I sites and required signature sequences are mixed. Rounds of digestion-ligation with *Esp*3I and T4 DNA Ligase result in the formation of the target plasmid pCAN7_ACX, which contains none of the original *Esp*3I recognition sites. B: Assembly strategy for combinatorial library. pGGV4, any single pOX3_ACX plasmid, and an equimolar mix of every pCAN6_AAT plasmid, are mixed. Rounds of digestion-ligation with *Esp*3I and T4 DNA Ligase result in the formation of the target plasmid library of pCAN7_ACX_AAT (where AAT varies between each plasmid). **Red** = *Esp*3I recognition sites, **purple** = signature sequence A, **green** = signature sequence B, **light blue** = signature sequence C. **Pink** = *sacB*, **dark blue** = *acx*, **yellow** = *aat*.

In order to generate these combinatorial libraries, each with a maximum number of 26 unique clones *per* ACX enzyme to a total of 520 strains, an equimolar AAT library

master mix was prepared. Using this AAT master mix, one-pot assemblies could be carried out to combine every AAT against each IB-CoA active ACX (Figure 8-1B). Instead of isolating individual colonies from this assembly, which would require preparing hundreds of cryostocks each with a different ACX and AAT combination, I instead transformed the assembly as one AAT module mixture into *E. coli* BW25113 Δ*infA* Δ*ldhA* :: KanR. The PZ-BMA solid phase screen could then be used to search for any changes in fluorescence in resultant colonies. Sequencing of a a proportion of the colonies produced was used to check the coverage of the total library represented in the assembly products. However some AATs were expected to be preferentially integrated into pCAN-7 over others, dependent on the toxicity of any esters that may result from the constitutive expression of transferases during cloning. As with the approach to ACX screening, hits from AAT fluorescent screening would be validated using biotransformations.

8.2 Results

8.2.1 Design and assembly of a Golden Gate vector to assess candidate enzymes

The Golden Gate (pGG) plasmid was designed to facilitate rapid assembly from enzyme library plasmids directly into a constitutively expressed production plasmid (pCAN-7, Figure 8-1). The backbone of pGG was based on the sequence of pCAN4 and pHIBA3, retaining the *infA* stabilisation, BCKD genes and constitutive Anderson promoter. However, in place of the AATm4 and AtACX4 used in pCAN4, pGG contains the counter-selection marker gene *sacB*, cloned from the plasmid pEX18-Gm. *Esp*3I recognition sites flanking the new SacB gene with unique signature sequences were also included. *Esp*3I is a Type IIS enzyme, and with proper plasmid design allows digestion of pGG to create sticky ends which, once ligated together, no longer contain *Esp*3I recognition sites. Using Golden Gate the pGG cloning vector could be mixed with any combination of AAT and ACX plasmids from the enzyme library, and repeat digestion-ligation cycles facilitated one-pot assembly of production

plasmids. The enzyme libraries were designed to accommodate this, with identical *Esp*3I sites for generating signature sequences flanking genes in the pCAN6 (AAT) (Figure 8-2A) and the pOX3 (ACX) plasmids (Figure 8-2B).

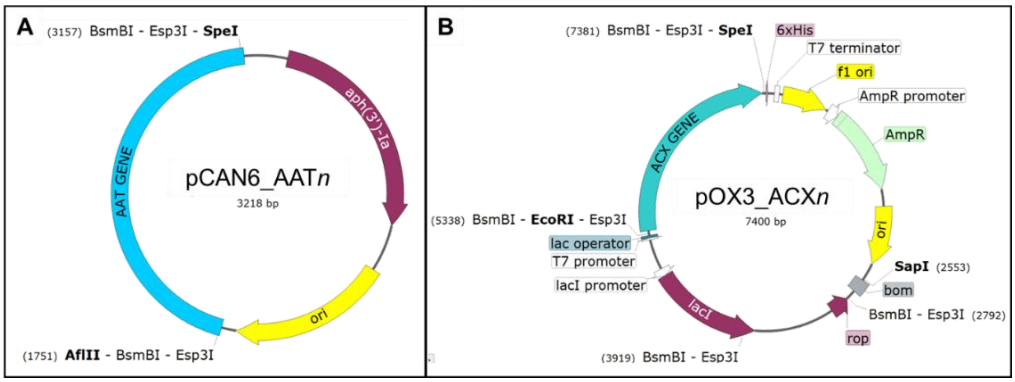


Figure 8-2: Cloning sites for AAT and ACX codon optimised gene synthesis.

A: AAT sequences were cloned into a modified pUC57-Kan vector, with *Esp*3I, *Eco*RI, *Sap*I & *Afl*II sites in the original vector removed. B: ACX sequences were cloned into pET-21(+) to allow for subsequent *in vitro* assay work.

Several methods were attempted before successful pGG assembly was eventually achieved. For the first approach (pGGV1) I endeavoured to use a 5-part HiFi assembly to construct pGGV1 directly. The five DNA segments for this assembly were as follows: pCAN4(BCKD) produced by restriction digest of pCAN4 using *Eco*RI and *Not*I (5199 bp), pEX18-Gm(SacB) from PCR amplification of *sacB* from the plasmid pEX18-Gm (1898 bp), and three backbone fragments designed to remove existing *Sap*I and *Esp*3I sites in the pCAN-4 backbone, all amplified from pCAN4. These fragments were AflII_SapI (2779 bp), SapI_Esp3I (224 bp), and Esp3I_NotI (717 bp). Re-streaking colonies from the pGGV1 assembly onto sucrose counter-selection plates resulted in colonies in all cases. Despite this result colony PCR was also used to assess the colonies, in the event that *sacB* expression for counter-selection was not working. No bands of the correct size were observed throughout several attempts of the pGGV1 method.

I hypothesised that the number of assembly fragments may reduce HiFi efficiency. Additionally, the SapI_Esp3I fragment was significantly smaller than the other fragments in the assembly, at 224 base pairs, which may also have reduced the

efficacy of HiFi assembly. Therefore, for the second strategy a pair of oligonucleotides were designed to cover the *SapI* to *NotI* region. Annealing these oligonucleotides together resulted in a 941 bp fragment, SapI_NotI. I then attempted a 3 part HiFi assembly reaction using the new SapI_NotI fragment. As with pGGV1, no plasmid was correctly amplified by colony PCR. Sequencing of products also produced negative results.

The third strategy, for pGGV3, was to return to a simpler, digestion-ligation approach. Using this method, the vector backbone was amplified from pHIBA3 using AflII_SapI_Fwd and NdeI_Rev. This was digested with *Afl*II and *Nde*I. This time a gene block (gBLOCK) fragment was used to cover the 941 bp region around the pHIBA3 backbone. The gBLOCK was blunt end ligated into pJET1.2 and then digested using *Nde*I and *Not*I. As with other approaches, BCKD was digested from pCAN4 using *Not*I and *Eco*RI. *SacB* was amplified from pEX18-Gm using pSacB2_Fwd and pSacB2_Rev before digestion using *Eco*RI and *Afl*II. This strategy produced no products with the correct amplification during colony PCR.

The final, and successful, attempt to construct pGG reverted to a four-part HiFi assembly to prepare the Golden Gate plasmid, pGGV4 (Figure 8-3A). In this approach the origin of replication, *infA* gene and promoter and Amp resistance cassette were amplified from pHIBA3 to prepare the original AflII_SacB fragment (Figure 8-3A, dark blue). This time the 941 bp fragment from pJET_gBLOCK (Figure 8-3A, yellow) was also amplified. BCKD was digested from pHIBA3 as usual using *Not*I and *Eco*RI (Figure 8-3A, light blue). *SacB* was amplified from the plasmid pEX18-Gm (Figure 8-3A, pink), which included the addition of flanking *Esp*3I sites to allow replacement of SacB with AAT and ACX in the final plasmid. After HiFi assembly, pGGV4 was purified and the plasmid was transformed into *E. coli* NEB5α. Colony PCR results confirmed the correct assembly of pGGV4. I sequenced the newly assembled joins of pGGV4 plasmid to confirm the correct construction (Supplementary A.4:).

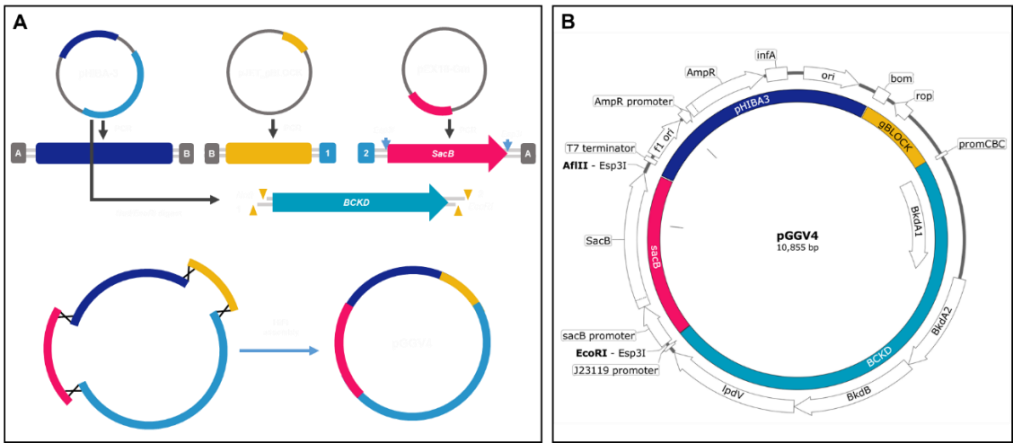


Figure 8-3: Construction of a vector for the Golden Gate assembly of an ACX/AAT library

A: A 4-part HiFi assembly strategy was employed to generate pGGV4 (a Golden Gate assembly vector). PCR amplification was used to generate 3 fragments with overlapping ends, while a fragment encoding BCKD was generated by digestion using *Not*I and *Eco*RI restriction enzymes. B: The final pGG vector is stabilised using *infA* and contains a *sacB* counter-selection marker and promotor that is removed during assembly of library strains.

Several test assemblies were carried out to check the new assembly platform: DNA fragments containing AtACX4 and AATm4 were amplified from pCAN-4 using the primer pairs OG-ACX4-F/R and OG-AAT-F-new/R, respectively. This PCR introduced the same flanking regions and signature sequences for AtACX4 and AATm4 as were included in the oxidase and transferase library constructs. 40 fmol of the donor PCR products was mixed with 20 fmol pGGV4, along with all other required components. After assembly, the plasmid products were transformed directly into *E. coli* DH5α. Colony PCR with SEQ-28 and SEQ-53 was used to check for the correct product size (Figure 8-4, pCAN-7_AtACX4-AATm4).

The success of assembling pCAN7_AATm4-ACX#2, pCAN7_AATm4-ACX#14, pCAN7_AtACX4-AAT#6, and pCAN7_AtACX4-AAT#19 into *E. coli* DH5α was also tested. I amplified the original AATm4 from pCAN4 using the primer pairs OG-AAT-NEW-F/R as for the pCAN4 re-assembly. AtACX4 was also amplified from pCAN4 using OG-ACX-F and OG-ACX-R. These oligonucleotides introduced corresponding *Esp*3I sites flanking each gene that were compatible with the Golden Gate cloning sites in pGGV4. Novel ACX and AAT genes were added to the assembly directly in the pOX3 and pCAN6 plasmids in which the genes had been synthesised. Colony

PCR was used to verify the presence of AAT/ACX genes in place of *sacB* after the Golden Gate reaction had taken place (Figure 8-4).

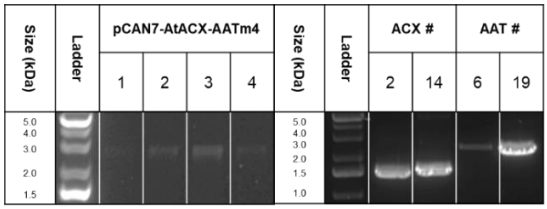


Figure 8-4: PCR verification of control Golden Gate assemblies

DNA amplified from pCAN7 Golden Gate products. PCR reactions were loaded onto a 1% agarose gel in TAE buffer, which ran at 80 V for 40 min. 1 kb plus DNA ladder. Expected band sizes: pCAN7-AtACX4-AATm4 = 2.6 kb, ACX#2 and ACX#14 = 1.5 kb, AAT#6, and AAT#19 = 2.3 kb.

8.2.2 Assembly of an oxidase library for BMA formation

Strains into which ACXs were assembled were named **CAN7-ACX#n** where "n" is the number given in the ACX code table (Table 7-2). For assembly of the ACX library strains, the Golden Gate reaction was set up as *per* the method used for the controls. After the assembly reaction, 4 µL of the Golden Gate mixture was transformed into chemically competent *E. coli* DH5α cells. Successful transformation occurred for 18 out of 20 ACXs. Despite repeated assembly attempts and screening of additional colonies, it was not possible to successfully transform ACX#11 and ACX#12 into *E coli* DH5α. DNA from the remaining 18 ACX candidates was purified and the plasmids were transformed into chemically competent *E. coli* CBC-M1 cells (provided by AY). 100% of screened colonies from this transformation contained the correct plasmid (Figure 8-5).

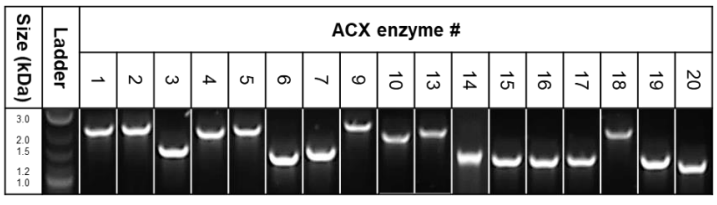


Figure 8-5: Colony PCR of ACX library transformants

Colony PCR was carried out on colonies grown from the transformation of the pCAN-7 plasmids into *E. coli* BW25113 Δ*infA* Δ*ldhA* :: KanR cells. 15 µL PCR mix was loaded directly onto a 1% agarose in TAE gel. The gel ran for 50 min at 80 V.

However, no colonies grew on **CAN-7_ACX#8** plates. It was not possible to transform the plasmid pCAN-7_ACX#8-AATm4 into *E. coli* CBC-M1 despite several attempts. Once all ACX strains were assembled, with the exception of **CAN-7_ACX#8**, **CAN-7_ACX#11**, **CAN-7_ACX#12**, colony PCR was used to verify that the correct ACXs were integrated into the production vector (Figure 8-5).

8.2.3 ACX library screening

During the ACX library screening, expression of the ACX bioinformatic library genes was not assessed. To begin the IVIS assay for ACX activity, **CAN-7_ACX** strains #1-7,9-10, and #13-20 were inoculated from cryostock into 10 mL LB (1% glycerol (v/v), 5 mM BuOH). These were incubated overnight at 37°C and 250 rpm. After 12-15 hours, overnight cultures were diluted by 1000 and spotted the solution onto pre-gridded membranes on LB agar containing BuOH and carbenicillin, in triplicate. This spotting method was replicated on two additional identical plates. After allowing time for the spots to dry, the LB plates incubated at 37°C for 48 hours. At 48 hours after spotting, membranes were transferred onto SPA agarose plates containing Na-2-KIV and 15 mM BuOH using a pair of sterilised tweezers. SPA plates were then returned to 37°C for 1 hour, before activation of the mDTZ probe for 30 min under UV light. Each plate was imaged using the IVIS Spectrum and the maximum radiance output from each spot was recorded (Figure 8-6A). Maximum radiance is reported as a percentage of the maximum radiance produced by the **CAN4** positive control (Figure 8-6B).

In many instances, the relative radiance between each ACX strain was not significantly different from the **CAN4** control. As evidenced in all three IVIS images, reflections from the agarose surface appeared to interfere with the fluorescence signal once again – as for the four central spots on all plates recorded lower signal than replicates at other positions. The relative radiance for ACX#14, ACX#19 and ACX#20 however, appeared to consistently give higher average readings between plates than for the other strains, despite the variability of replicates. One-way ANOVA for the average radiance values from each plate gave a p value of 0.000626, suggesting that

statistically not all the means are equal for each ACX enzyme, consistent across the plates.

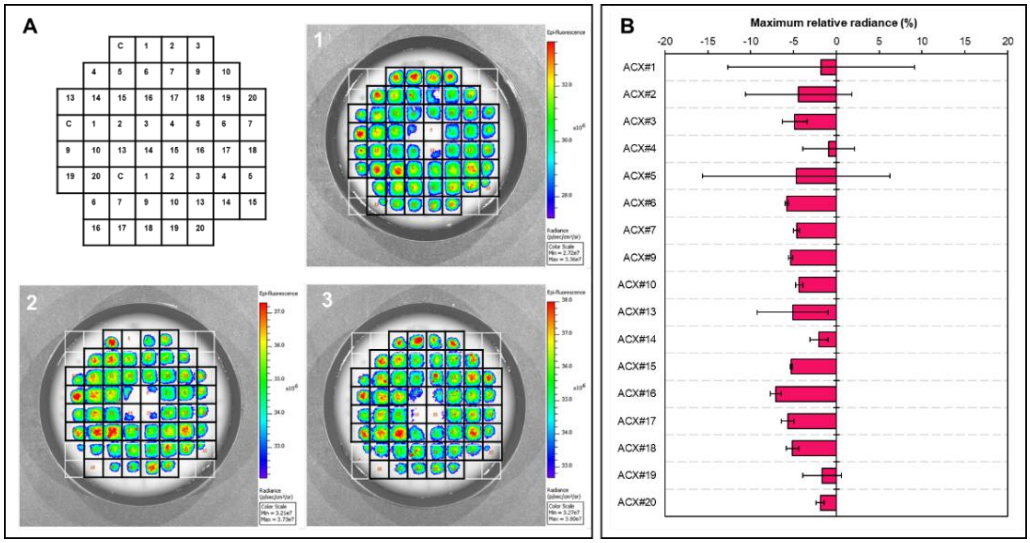


Figure 8-6: ACX library radiance with full plate spotting

Maximum radiance measured after 1 hour incubation with solid phase assay agarose, and 30 min UV activation to form PZ-BMA. Images recorded with ex/em 465/520 nm, with an exposure time of 30 seconds. A: Each plate 1, 2, and 3 represents an independent experiment each with 3 replicates. CAN4 control spotted where indicated on grid overlay (top left). B: Relative maximum radiance values calculated as percent radiance relative to CAN4 controls for each plate. Error bars represent variance calculated from one-way ANOVA. Data obtained and analysed using Living Image Software.

I repeated the experimental setup, this time with **CAN4** and **CAN5** spotted on each plate as controls. The segment of the plates with the highest reflection was left empty (Figure 8-7, yellow squares), and the controls on each plate were used to calculate the corrected radiance. Unlike the first screening attempt, no significant difference was visible between any of the means for the strains. A p value of 0.15 was determined from one-way ANOVA. However, several strains produced consistently higher radiance in both this experiment and the previous screen. These were ACX#3, ACX#4, ACX#14, ACX#19, and ACX#20. No major variation could be picked out from this data (Figure 8-7), with the highest difference from **CAN4** being from **CAN7-ACX#3** which had a 7.5% increase in fluorescence. This was followed by **CAN7-ACX#4** and **CAN7-ACX#19** which gave a 6.5% and 4.6% increase, respectively. **CAN7-ACX#7**, **CAN7-ACX#17** and **CAN7-ACX#20** all had a variance of less than 1%

radiance, 0.88%, -0.91% and 0.11%. These strains may be most similar to **CAN4** BMA productivity in flask. Particularly low relative radiance was produced by strains **CAN7-ACX#9**, **CAN7-ACX#10**, **CAN7-ACX#16** and **CAN7-ACX#18**. **CAN7-ACX#18** produced a radiance that was -5.8% of that of **CAN4**. **CAN7-ACX#3** and **CAN7-ACX#18** as 'highest and lowest' BMA producers, would be most likely of the ACX library to register distinct BMA productivities in biotransformations. However, as the fluorescence percentage changes are so small, it is unlikely that a substantial change in BMA productivity exists between **CAN7-ACX#3** and **CAN7-ACX#18**.

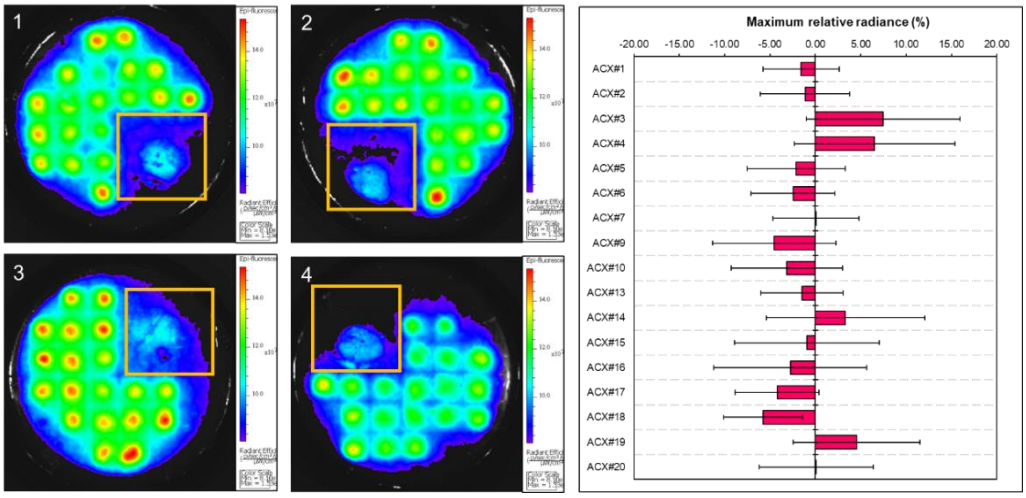


Figure 8-7: ACX library radiance with asymmetric plate spotting

Radiance values obtained from IVIS imaging for SPA agarose plates containing Na-2-KIV and BuOH, measuring the fluorescence of PZ-BMA with an ex/em 465/520 nm. Images 1-4 are biological replicates, with each strain plated once *per* plate to achieve n=3 for all strains. Maximum relative radiance calculated to **CAN4** control strains plated in triplicate on each of plates 1-4, where 0% = no change from control, positive values represent an increase, and negative values represent a decrease. Error bars represent standard deviation of 4 biological replicates.

All **CAN7-ACX** strains were assessed for BMA formation in biotransformations, using a similar experimental setup to the one employed for pathway analysis in 5.2.3. As for the plate assay colonies had to be grown for 48 hours, the biotransformation run time was adjusted to 48 hours accordingly. Na-2-KIV and BuOH supplementation were also lowered, as no titres achieved so far used more than 5% of the provided quantity. *E. coli* BW25113 *ΔldhA ΔinfA* :: KanR pCAN7-ACX was grown overnight in LB, then sub-cultured into 200 mL LUND medium with glycerol and BuOH, to an OD_{600} of 0.1. At an OD_{600}>3 the cells were harvested at 7,000 rpm for 5 min and resuspended in 20

mL BT medium supplemented with 0.1% glycerol, 5 mM BuOH and 40 mM Na-2-KIV to a final OD_{600} of 10. After 48 hours at 37°C and 250 rpm, cultures were centrifuged 8,000 rpm for 10 min. 10 mL supernatant was then mixed 1:1 with ethyl acetate to extract ester products, detecting BMA, BIB and BA using GC-MS (Table 8-1).

BMA formation	Code	Isoform	Organism	Issue		
No data	ACX#8	ACX4	*Zea mays*	Cannot clone into production strain		
	ACX#11	ACX3	*Yarrowia lipolytica*	Cannot clone into pCAN-7		
	ACX#12	ACX3	*Vigna radiata*	Cannot clone into pCAN-7		
BMA formation	**Code**	**Isoform**	**Organism**	**BMA (±)**	**BIB (±)**	**BA (±)**
No activity	ACX#1	ACX1	*Spinacea oleracea*	N.D	0.567 (0.064)	1.122 (0.063)
	ACX#2	ACX1	*Zea mays*	N.D	0.432 (0.114)	1.480 (0.235)
	ACX#5	ACX1_2	*Zea mays*	N.D	0.737 (0.016)	0.320 (0.009)
	ACX#6	ACX4	*Apostasia shenzhenica*	N.D	0.707 (0.065)	0.427 (0.012)
	ACX#9	ACX4	*Candida maltosa*	N.D	0.251 (0.022)	0.338 (0.021)
	ACX#10	ACX1	*Vigna radiata*	N.D	0.693 (0.025)	1.018 (0.074)
	ACX#13	ACX4	*Candida tropicalis*	N.D	0.500 (0.022)	0.449 (0.016)
	ACX#16	ACX4_X2	*Vigna radiata*	N.D	0.507 (0.048)	1.615 (0.142)
	ACX#18	ACX1	*Zea mays*	N.D	0.598 (0.038)	0.251 (0.022)
Low activity	ACX#3	ACX4	*Parasponia andersonii*	0.003 (0.001)	0.811 (0.115)	0.731 (0.060)
0-50%	ACX#4	ACX3	*Spinacea oleracea*	0.015 (0.001)	0.867 (0.041)	0.888 (0.040)
	ACX#17	ACX4	*Arachis hypogaea*	0.016 (0.002)	0.474 (0.073)	1.474 (0.079)
	ACX#20	ACX4	*Glutamicibacter nicotianae*	0.019 (0.001)	1.040 (0.027)	2.309 (0.040)
Medium activity 50-100%	ACX#19	ACX4	*Trema orientale*	0.053 (0.007)	0.885 (0.108)	0.904 (0.082)
High activity	ACX#7	ACX4	*Spinacea oleracea*	0.055 (0.006)	0.246 (0.032)	0.114 (0.114)
	ACX#15	ACX4_X3	*Populus alba*	0.026 (0.002)	0.483 (0.028)	0.297 (0.011)
>100%	ACX#14	ACX4_X1	*Vigna radiata*	0.096 (0.009)	0.875 (0.098)	1.935 (0.093)

Table 8-1: Summary of butyl ester formation using oxidase library strains

Butyl ester concentrations measured using GC-MS on extracts from biotransformation cultures after 48 hours. Strains sorted according to BMA production levels. All concentrations given as mM. BMA = butyl methacrylate, BIB = butyl isobutyrate, BA = butyl acetate. N.D. = none detected. BMA, BIB and BA columns coloured according to value, where highest values *per* column are coloured in **blue**. Error values (±) reported are the standard deviation of biological triplicates, and are also given in mM. **CAN4** control carried out in triplicate for each round of biotransformation and used for calculations of relative BMA production in subsequent analysis.

Nine of the 17 ACXs did not produce any BMA. This included all the strains that express ACX1 instead of ACX4 isoforms. Production of BIB and BA was retained in all the BMA negative strains, and in no strain was BMA the major ester product. Neither

of the *Candida* ACX4s produced BMA, and both produced BA and BIB. CtACX4 produced almost the same amount of both BA and BIB, at roughly 0.4 mM. CmACX4 produced 0.25 mM BIB and 0.33 mM BA. AsACX4 also didn't produce any BMA and was one of the highest producers of BIB (0.7 mM) in the no activity group of oxidases. Interestingly, VrACX4-X1 did not allow BMA formation, despite the fact that its close relative VrACX4-X2 was the highest BMA producer from our flask experiments.

Relative BMA values do not tally precisely with direct BMA concentration readings because of variation in BMA formation from **CAN4** between different rounds of biotransformation experiments. Oxidases with less than 50% relative BMA productivity were labelled as a "low activity" group. This comprised four ACX4s: PanACX4 (ACX#3), SoACX3 (ACX#4), AhACX4 (ACX#17) and GnACX4 (ACX#20). PanACX4 and AhACX4 are among the oxidases from Clade I that are most closely related to AtACX4 (Figure 7-4). PanACX4 produced roughly 7x less BMA than the other ACX4 enzymes in this group, with a low BMA titre of 0.0026 mM, and a relative BMA concentration of only 14% when compared to BMA production from the **CAN4** control for that round. The selectivity of PanACX4 was also the worst of all BMA producers, with a 1:312:281 BMA:BIB:BA ratio (Table 8-2). AhACX4 produced 0.16 mM BMA, with a relative production of 21% to **CAN4** in that round. The only ACX3 that was successfully cloned into the production strain SoACX3, is also in this grouping. SoACX3 produced a similar BMA titre to AhACX4 and GnACX4, at 0.015 mM, with the same relative BMA concentration of 21%. The PanACX4 strain preferentially formed BIB over BA and BMA, while SoACX3 produced almost a 1:1 ratio of BIB:BA. Interestingly, the GnACX4 strain **CAN7-ACX#20** formed the highest titres of BIB and BA as compared to all other strain in the oxidase library, with titres of 1.0 mM and 2.3 mM respectively.

Only one oxidase fit into the 'medium activity' group, with a BMA forming activity between 50-100% of that of **CAN4**. ACX#19, or ToACX4, produced a BMA titre of 0.053 mM, which is equivalent to the SoACX4 titre of 0.055 mM in the high activity

grouping. However, when compared to the control for that biotransformation run, the productivity was only 75% of the AtACX4 strain. ToACX4 in Clade I of the oxidase phylogeny (Figure 7-4), with high similarity to AtACX4.

The high activity group was defined based on oxidase strains with a >100% relative BMA concentration as compared to the AtACX4 strain. The lowest producer in this group was PalACX4-X3 which produced 0.026 mM BMA, but with a relative concentration of 204% compared to the control for that round of biotransformations. By far the highest producer was **CAN7-ACX#14** (VrACX4-X2), which formed 0.096 mM BMA, although this equated to a 130% increase in BMA concentration when compared to **CAN4** for that round, less than for **CAN7-ACX#15** (PalACX4-X3). The VrACX4-X2 strain also had a more favourable selectivity towards BMA, at a 1:9:20 BMA:BIB:BA ratio (Table 8-2). However, the best selectivity in both the high activity group, and from all the BMA producing **CAN7** strains, was from **CAN7-ACX#7** (SoACX4). The SoACX4 strain formed a relative BMA concentration of 315%, with a selectivity of 1:4:2 for BMA:BIB:BA.

Activity	Code	Oxidase	BMA:BIB:BA
Low	ACX#3	PanACX4	1:312:281
	ACX#4	SoACX3	1:57:58
	ACX#17	AhACX4	1:30:93
	ACX#20	GnACX4	1:55:124
Medium	ACX#19	ToACX4	1:16:17
High	ACX#7	SoACX4	1:4:2
	ACX#14	VrACX4-X2	1:9:20
	ACX#15	PalACX4-X3	1:18:11

Table 8-2: Butyl ester formation ratios from BMA producing oxidase strains

Biotransformation data for both BMA concentration and BMA relative to **CAN4** BMA in each round of experiments was compared with the relative radiance values calculated using the IVIS data (Figure 8-8). The relative radiance values do not completely correspond with the flask relative BMA values. **CAN7-ACX#14** and **CAN7-ACX#19** both produced positive relative radiance values during IVIS screening, and also produced BMA in flask experiments. Indeed, all strains with the highest relative

radiance readings were able to synthesise BMA. Similarly, strains with the lowest radiance values, **CAN7-ACX#18** and **CAN7-ACX#9** were able to form BA and BIB but not BMA during biotransformations. However, the GC-MS results for **CAN7-ACX#3**, **CAN7-ACX#4**, and **CAN7-ACX#15** do not tally with the radiance values for these strains.

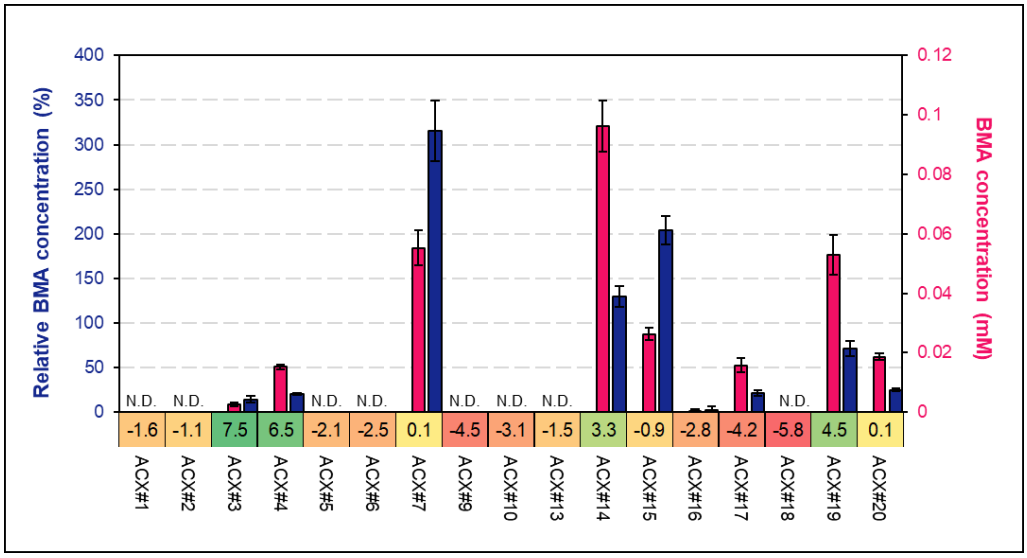


Figure 8-8: Flask and IVIS BMA production from oxidase library

BMA concentration measured using GC-MS on extracts from biotransformation cultures after 48 hours. BMA concentration in **pink**, BMA concentration relative to **CAN4** control for that round of biotransformations in blue. N.D. = none detected. Numbers shown below x-axis are% radiance from IVIS data, coloured on a scale from lowest in **red** to highest in **green**. Error bars represent standard deviation from biological triplicates.

The polypeptide sequences of the three highest performing oxidases, ACX4 from *S. oleracea*, *T. orientale* and *V. radiata*, were aligned against AtACX4 to see if there was any amino acid signature from the best performing ACXs. There was a high degree of sequence homology between all four proteins (Figure 8-9). Key catalytic residues Glu408 and Arg420 were conserved in all polypeptides, Cys399, important for the dimer-dimer interaction, was also conserved for all instances. All three of the new oxidases ended with the C terminal SRL signalling sequence.

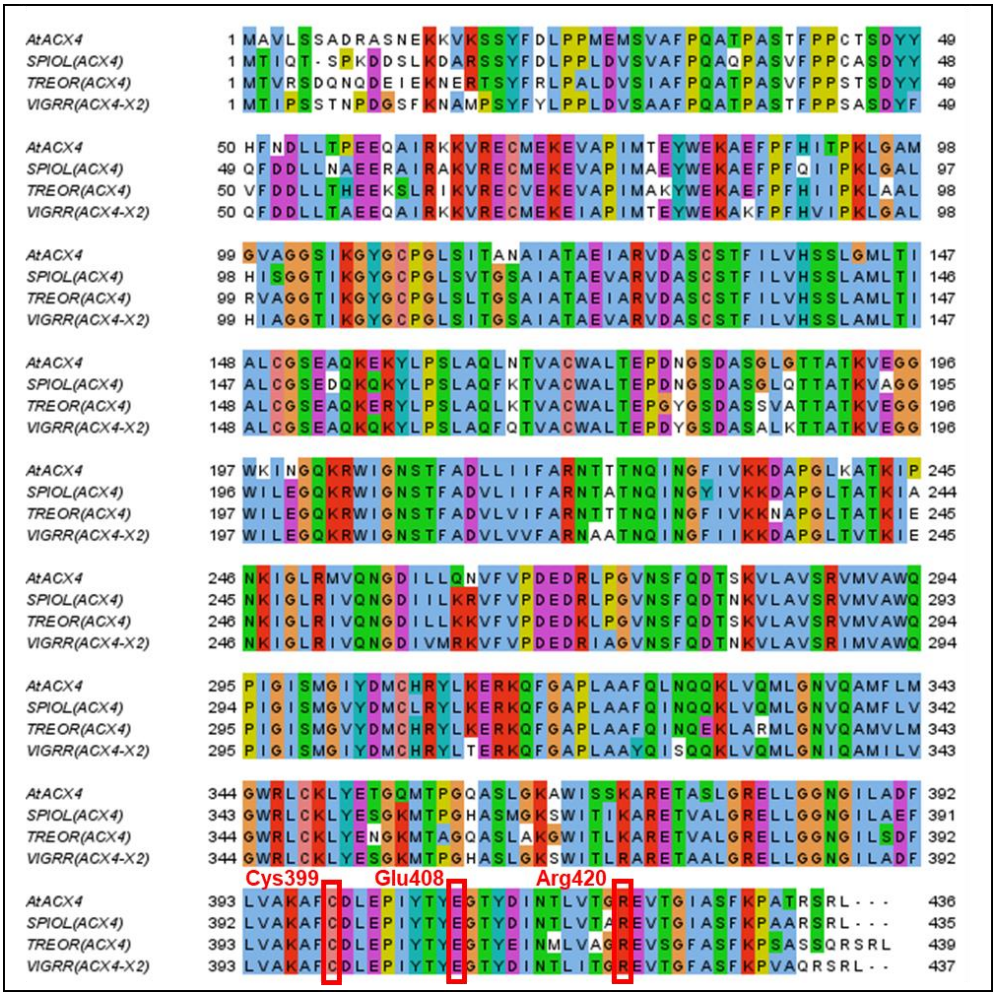


Figure 8-9: Multiple sequence alignment of AtACX4 with top hits from flask BMA screening

Sequences aligned using ClustalO, visualised in Jalview 2.11.2.1. Sequences coloured using Clustalx colours. AtACX4 = *Arabidopsis thaliana* acyl-CoA oxidase 4 (MCUK), SPIOL(ACX4) = *Spinacea oleracea* acyl-CoA oxidase 4 (XP_021855534.1), TREOR(ACX4) = *Trema orientale* acyl-CoA oxidase 4 (PON92218.1), and VIGRR(ACX4-X2) = *Vigna radiata* acyl-CoA oxidase4 isoform X2 (XP_014516782.1). Resides of interest are outlined in **red**.

Only SoACX3 differs in sequence from AtACX4 at positions directly adjoining any of the key ACX catalytic residues outlined in Figure 8-9. In this instance, there is an Alanine residue at position 419 in place of the Glycine residue present in AtACX4. However, structural analysis using the AtACX4 crystal structure demonstrates that is is unlikely this change has a large effect on ACX activity. Indeed, although Arg420 is responsible for a key hydrogen bond formation with the -CoA substrate, surface mapping showed that the adjacent glycine residue likely forms part of the outer surface of the protein (Figure 8-10A). Substitution of Alanine at this position did not

indicate any conformational changes would occur in the ACX active site (Figure 8-10B). Alanine, like Glycine, is a small and flexible amino acid, therefore a substitution of one with the other would perhaps be expected to have a minor impact on protein structure.

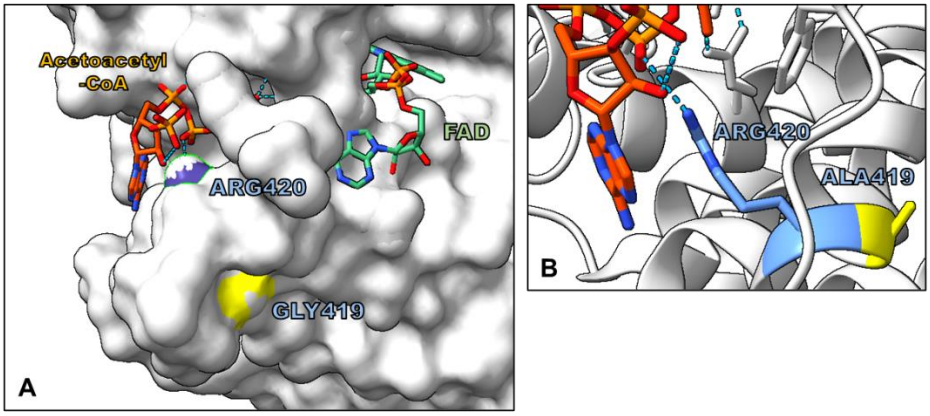


Figure 8-10: Structural differences between AtACX4 and SoACX3

A: Surface model of ACX4 from *Arabidopsis thaliana* (AtACX4). B: Arg420 and a G419A substitution in Chain A of AtACX4. FAD shown in **green**, acetoacetyl-CoA shown in **orange**. Residues represented by ball and stick models. Arg420 shown in **blue**, Ala419 shown in **yellow**. Figure created using UCSF Chimera with the crystal structure PDB ID: 2IX5. No crystal structure available for SoACX3.

8.2.4 Combinatorial Golden Gate assembly

The ACX enzymes which made up the 'no activity' group were excluded, streamlining the number of libraries to construct for the combinatorial ACX and AAT library. This was based on the assumption that the "no activity" oxidases cannot bind to and oxidise IB-CoA. BMA forming oxidases from all non-zero activity groups were included in the combinatorial assembly. This left 7 ACXs, and At-ACX4, which could be used to assemble 8 AAT variant libraries: ACX#3, ACX#4, ACX#7, ACX#14, ACX#15, ACX#17, ACX#19 and ACX#20. Before proceeding with the generation of these libraries, the two-part Golden Gate assembly was tested using AAT#19 and AtACX4, with pCAN-4 re-assembly as a control. After assembly, pCAN7-AtACX4-AATm4 and pCAN7-AtACX4-AAT#19 were transformed into *E. coli* DH5α, and colony PCR used to check five colonies from each transformation, using primers SEQ-54 and SEQ-28. 4 out of 5 colonies from the control were correctly assembled, while all the pCAN7-

AtACX4-AAT#19 transformation colonies gave the correct amplification (Figure 8-11A).

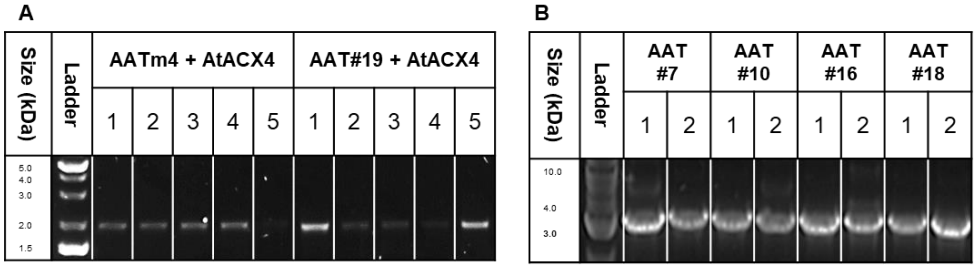


Figure 8-11: Colony PCR for assembly of combinatorial strains

A: DNA amplified from pCAN7 using the primers SEQ-54 and SEQ-58. Colony PCR reactions were loaded onto a 1% agarose gel in TAE buffer, which ran at 80 V for 40 min. 1 kb plus DNA ladder. Expected band size = 1.93 kb. B: DNA amplified from pCAN7 using primers AS-CHK-F1 and AS-CHK-R. Colony PCR reactions were loaded onto a 1% agarose gel in TAE buffer, which ran at 80 V for 40 min. 1 kb plus DNA ladder. Expected band size = 3.13 kb.

The simultaneous assembly of new ACX and AAT direct from the pOX or pCAN6 vectors was also tested. The oxidase-transferase pairings selected for this were ACX#7 and AAT#16, AAT#18, and AAT#10. As with all previous assemblies, 20 fmol of each donor vector was mixed with 40 fmol pGGV4. After assembly this mixture was transformed into *E. coli* DH5α. Colonies grew on all plates, and colony PCR used to verify the correct assembly with AS-CHK-F1 and AS-CHK-R primers (Figure 8-11B) for three colonies from each plate. On checking the remainder of the AAT library, three AATs were unable to clone into pCAN7 constructs: these were AAT#4, AAT#5, and AAT#24. Generation of the combinatorial library then commenced, and the non-clonable AATs were not used in subsequent assemblies.

An equimolar mixture of the 23 pCAN6-AAT plasmids remaining after discounting AAT#4, AAT#5, and AAT#24 was prepared. In the combinatorial assembly master mix 20 fmol of this AAT mixture was used in place of one individual AAT. 20 fmol ACX and 40 fmol pGGV4 were added as normal. Eight parallel assembly reactions were set up with this AAT library mix (ATL), one for the original ACX4, AtACX4xATL, and one for each of the BMA producing ACXs: ACX#3xATL, ACX#4xATL, ACX#7xATL, ACX#14xATL, ACX#15xATL, ACX#17xATL, ACX#19xATL and ACX#20xATL. After the assembly reaction took place, the product was transformed into *E. coli* DH5α as

usual. Colonies grew on all plates, and all colonies *per* plate were pooled together for purification and transformation into *E. coli* CBC-M1 cells (AY).

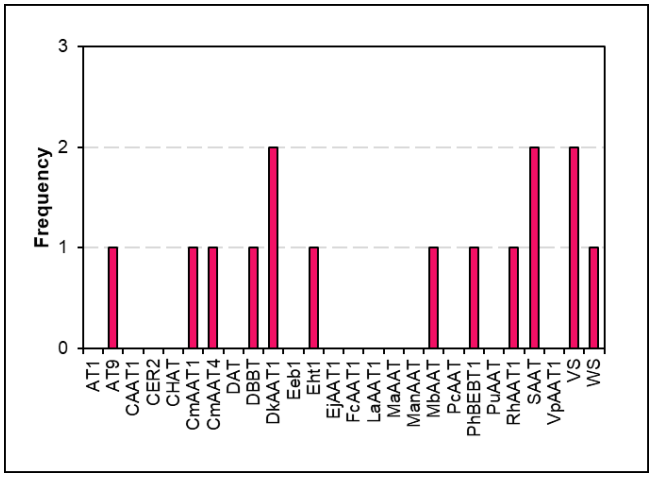


Figure 8-12: Sequencing AATs in pCAN7 library plasmids

Frequency of entries from AAT library cloned into pCAN7 simultaneously. 15 colonies sequenced with AS-CHK-R to identify AAT present in final construct. Sequencing carried out by Source Bioscience (Nottingham, UK).

To assess library coverage, 15 colonies from these transformations were randomly selected and amplified at the region around AAT using AS-CHK-F1 and AS-CHK-R primers. The PCR products from this reaction were sent for sequencing to determine whether a mixture of AATs had assembled into the final constructs. Of the 15 pCAN7 plasmids sequenced, sequencing returned the following proportion of AATs: 2x DkAAT1, 1x AT9, 1x CmAAT4, 1x WS, 1x DBBT, 2x VS, 2x SAAT, 2xRhAAT1, 1x CmAAT1, 1x EjAAT1, 1x PhBEBT1 (Figure 8-12).

8.2.5 Combinatorial library screening

The experimental setup used for combinatorial screening was almost identical to that for the ACX only screen. As each ACX was mixed with 23 AATs simultaneously, 27 colonies were picked from each library and patched onto the assay membranes. In addition three **CAN4** spots were included *per* plate as triplicate controls. The 27 colonies were inoculated from **CAN7_AtACX-ATL**, **CAN7_ACX#3-ATL**, **CAN7_ACX#4-ATL**, **CAN7_ACX#7-ATL**, **CAN7_ACX#15-ATL**, **CAN7_ACX#17-ATL**, **CAN7_ACX#19-ATL**, and **CAN7_ACX#20-ATL** each into 500 µL LB (1%

glycerol (v/v), 5 mM BuOH) in microtiter plates. The plates were incubated for 6-8 hours at 37°C and 250 rpm, to account for the decreased volume as compared to the pre-cultures used for the ACX-only screening experiments. After an OD_{600} > 1 was reached, undiluted culture was spotted onto pre-gridded nylon membranes on LB, with one plate *per* library. The plates were incubated in a static incubator at 37°C for 48 hours, then transferred the membranes transferred onto SPA agarose and incubated for a further 1 hour. The mDTZ probe was activated for 30 min under UV light at 302 nm and the maximum radiance from each colony footprint was measured using the IVIS Spectrum (Figure 8-13).

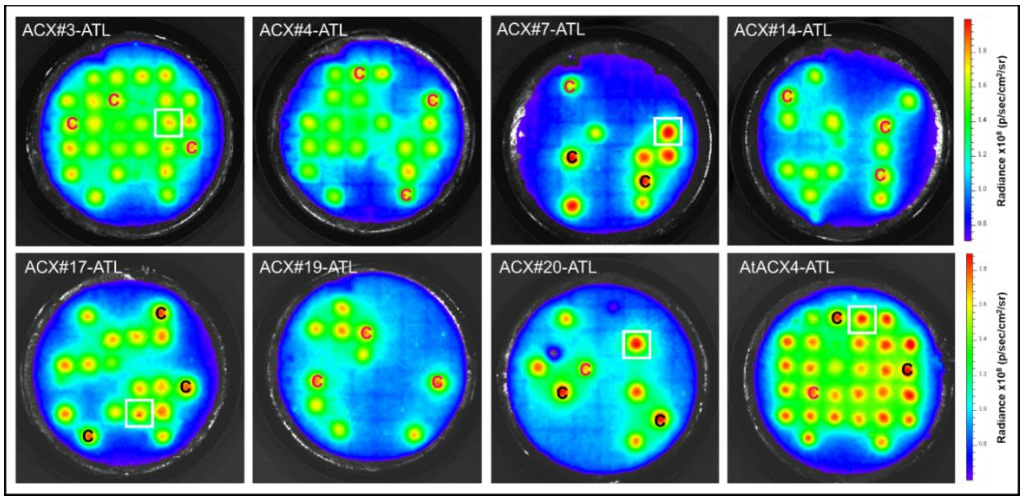


Figure 8-13: IVIS images of transferase library with individual ACX enzymes

Each plate represents colonies from a different ACX-ATL library. CAN4 controls spotted in triplicate in randomised locations where indicated with a “C”. White boxes outline colonies that were picked for flask biotransformation. Maximum radiance measured after 1 hour incubation with solid phase assay agarose, and 30 min UV activation to form PZ-BMA. Images recorded with ex/em 465/520 nm, with an exposure time of 30 seconds. Data obtained and analysed using Living Image Software.

From the images it is clear some libraries did not successfully produce BMA. In several cases only a small number of the 27 patched colonies grew. Of the AAT library produced for the highest producing ACX4, ACX#7, only 6 colonies were able to grow on LB to produce a fluorescent fingerprint on assay plates. Similar numbers of hits were observed from **CAN7_ACX#19-ATL** and **CAN7_ACX#20-ATL**. 7 colonies were picked using the IVIS Spectrum results for **CAN7_AtACX4-ATL** and **CAN7_ACX#14-ATL** using the biotransformation approach as employed for the

oxidase library tests (Table 8-3, rows 1-8). **CAN4** produced BMA of similar quantities to previous tests in all rounds of flask validation. However, no BMA was detectable from any of the selected colonies from these **CAN7_ACX-ATL** strains. On testing the remainder of the ACX-ATL libraries it was not possible to isolate any BMA producing **CAN7_ACX-ATL** strains (Table 8-3, rows 9-21). BA and BIB were also not detectable using GC-MS, although these were consistently visible as usual in the GC-MS traces for **CAN4** controls (Table 8-3).

Strain Information			Titre (mM)		
Code	**ACX**	**AAT**	**BMA**	**BIB**	**BA**
CAN4 (1)	AtACX4	AATm4	0.004	0.209	0.692
CAN7-OG (1)	AtACX4	ATL	N.D	N.D	N.D
CAN7-OG (2)	AtACX4	ATL	N.D	N.D	N.D
CAN7-OG (3)	AtACX4	ATL	N.D	N.D	N.D
CAN7-OG (4)	AtACX4	ATL	N.D	N.D	N.D
ACX14 (1)	ACX#14	ATL	N.D	N.D	N.D
ACX14 (2)	ACX#14	ATL	N.D	N.D	N.D
ACX14 (3)	ACX#14	ATL	N.D	N.D	N.D
CAN4 (2)	AtACX4	AATm4	0.002	0.540	0.541
CAN7-AX3-ATL	ACX#3	ATL	N.D	N.D	N.D
CAN7-AX4-ATL	ACX#4	ATL	N.D	N.D	N.D
CAN7-AX7-ATL	ACX#7	ATL	N.D	N.D	N.D
CAN7-AX15-ATL	ACX#15	ATL	N.D	N.D	N.D
CAN7-AX17-ATL	ACX#17	ATL	N.D	N.D	N.D
CAN7-AX19-ATL	ACX#19	ATL	N.D	N.D	N.D
CAN4 (3)	AtACX4	AATm4	0.010	0.513	0.317
CAN7-AX3-ATL	ACX#3	ATL	N.D	N.D	N.D
CAN7-AX7-ATL	ACX#7	ATL	N.D	N.D	N.D
CAN7-AX17-ATL	ACX#17	ATL	N.D	N.D	N.D
CAN7-AX20-ATL	ACX#20	ATL	N.D	N.D	N.D
CAN7-OG (5)	AtACX4	ATL	N.D	N.D	N.D
CAN7-AX7-AT16	ACX#7	AAT#16	N.D	N.D	N.D

Table 8-3: Butyl ester formation in flask from AAT library strains

Butyl ester concentrations measured using GC-MS on extracts from biotransformation cultures after 48 hours. All concentrations given as mM. BMA = butyl methacrylate, BIB = butyl isobutyrate, BA = butyl acetate, ACX = acyl-CoA oxidase, AAT = alcohol acyltransferase. N.D. = none detected. All ACX and AAT enzymes expressed from pCAN7 in *E. coli* BW25113 Δ*infA* Δ*ldhA* :: KanR. Each entry represents a strain tested in triplicate. **CAN4** control carried out in triplicate for each round of biotransformation, highlighted in blue.

Construction of a new strain into which a known AAT was cloned was also attempted. The AAT selected was AAT#16, or MaAAT. MaAAT is of particular interest due to the previous MCUK data using extract from banana pulp.Additionally, MaAAT was not

identified during sequencing of the combinatorial library assembly (Figure 8-12). AAT#16 was cloned into pCAN-7 with ACX#7, correct assembly was verified using colony PCR, and the correct plasmid was transformed into *E. coli* BW25113 Δ*infA* Δ*ldhA* :: KanR as usual. As for the ACX screen, colonies were grown from this transformation in LB (1% glycerol (v/v), 5 mM BuOH), before spotting 4 μL *per* culture onto a gridded membrane on LB. The **CAN7_ACX#7-ATL** plate incubated at 37°C for 48 hours, before membrane transfer to SPA agarose and additional incubation for 1 hour. mDTZ was activated for 30 min under UV light at 302 nm, before maximum radiance from each colony was measured on the IVIS Spectrum (Figure 8-14A).

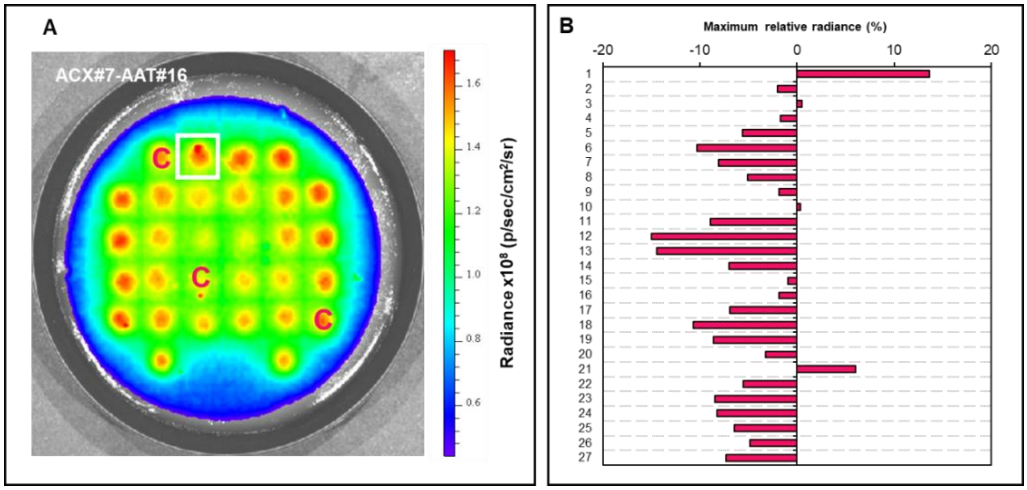


Figure 8-14: IVIS screening of CAN7_ACX#7-AAT#16

CAN4 controls spotted in triplicate in randomised locations where indicated with a "C". White box outlines the radiance signal from the colony that was picked for flask biotransformation. Maximum radiance measured after 1 hour incubation with solid phase assay agarose, and 30 min UV activation to form PZ-BMA. Images recorded with ex/em 465/520 nm, with an exposure time of 30 seconds. Data obtained and analysed using Living Image Software.

Almost all of the colonies measured from **CAN7_ACX#7-AAT#16** produced a relative radiance values of < 0% when we compared against **CAN4**. We picked colony 1 (Figure 8-14B) because its maximum radiance appeared to be greater than 10%. However, once again we found that while BMA, BIB and BA formation was consistent from **CAN4**, no ester formation could be detected from **CAN7_ACX#7-ATL** (Table 8-3). As each colony was patched once without duplicate, it is possible that the apparent hit was a false positive.

8.3 Discussion

8.3.1 Alternative oxidase enzymes with M-CoA forming activity

The oxidase screen for alternative ACXs revealed several novel ACX enzymes that could facilitate BMA formation. Perhaps unsurprisingly, most of these oxidases were the same isoform, ACX4, as the control oxidase. An interesting addition to this is the production of BMA *via* an ACX3 enzyme; SoACX3 or ACX (ACX#4) which has not previously been described. ACX3 from *A. thaliana* is classified as a medium-chain oxidase, with highest activity using C12 substrates when assayed with C8:0-C14:0 CoAs [192]. In the original list of ACXs for screening two further ACX3s were included. These were from *Y. lipolytica* and *V. radiata*, and despite many attempts I was unable to clone them into pCAN7. Considering the positive activity of SoACX3 it would therefore be interesting to assay YlACX3 and VrACX3 *in vitro*, to assess whether their poor cloning efficiencies result from any increase in oxidase activity, as the pathway enzymes are constitutively expressed from pCAN7.

All of the new oxidases with high activity however were ACX4s, and over half of the ACX4s incorporated into the production strain permitted BMA formation. Sequence similarity is well conserved between these enzymes, and as they were largely selected based on experimentally determined substrate acceptance it was expected that they may perform well. Indeed, all three strains expressing the ACX4 enzymes with reported butyryl-CoA activity (BRENDA) were capable of producing BMA. In contrast, those strains expressing ACXs selected because of hexanoyl-CoA activity or simply ‘C4’ activity did not form BMA (Table 8-4). This suggests that the range of substrates accepted by ACX4 is fairly constrained, particularly as compared to enzymes such as AAT.

No ACX1 expressing strain of *E. coli* produced any BMA. In all these strains BA and BIB were still formed, demonstrating that AATm4 was functional. It may be unlikely for ACX1 to accept IB-CoA as a substrate, purely because of the large structural divergence of ACX1 as compared to ACX4. As the active site of ACX4 is located at

the intersection of the two dimer subunits, ACX1s quaternary structure (wherein only one dimer is formed) potentially imparts a significant restriction on the size and charge of substrates that can bind to the oxidase active site. The lack of BMA production from any ACX1 strain backs up this hypothesis.

Code	ACX type	BRENDA	BMA
ACX#1	ACX1	Butyryl-CoA	-
ACX#2	ACX1	Hexanoyl-CoA	-
ACX#3	ACX4	No data	+
ACX#4	ACX3	Butyryl-CoA	+
ACX#5	ACX1	Hexanoyl-CoA	-
ACX#6	ACX4	No data	-
ACX#7	ACX4	Butyryl-CoA	++
ACX#9	ACX4	No data	-
ACX#10	ACX1	Butyryl-CoA	-
ACX#13	ACX4	C4-C6	-
ACX#14	ACX4	Butyryl-CoA	++
ACX#15	ACX4	No data	++
ACX#16	ACX4	Butyryl-CoA	-
ACX#17	ACX4	No data	+
ACX#18	ACX1	Hexanoyl-CoA	-
ACX#19	ACX4	No data	+
ACX#20	ACX4-like	Butyryl-CoA	+

Table 8-4: BRENDA and BMA production data for CAN7-ACX strains

Reported ACX activities compared against BMA formation from ACX expressing **CAN7** strains. - = no BMA, + = low BMA, ++ = BMA comparable to **CAN4**. Substrate data obtained from BRENDA K_M data for EC 1.3.3.6.

Ultimately although several IB-CoA active oxidases were identified, observed few dramatic increases in BMA formation were observed, or indeed any substantial changes to BA and BIB formation. There was a significant improvement, a 7x increase in BMA from the SoACX4, or ACX#7, strain. However, when you take into the account the scale of the increase, it was on a 0.1 µM scale. This is still far below a relevant industrial productivity, and importantly does not indicate that any of the new ACXs may be product-resistant. This is reflected in readings from the solid phase screening of the ACX library, where no significant differences from the **CAN4** control strain were observed. Because of this, further study of the top ACX candidates from this screen,

perhaps by *in vitro* assay, could demonstrate whether any of the new enzymes have the potential to actually enhance the BMA formation strains.

From the oxidase screen it is perhaps clear that the next way to tackle productivity at the IB-CoA to M-CoA step may be a more fundamental pathway modification. In particular, incorporating an acyl-CoA dehydrogenase in place of the oxidase may be the most sensible next step. If identifying successful ACXs using experimentally determined substrate specificities, as was used here, this may be a useful starting point in selecting a number of ACADs which may accept IB-CoA as a substrate. As we previously discussed in Table 1-2, IBDs are known to be active on IB-CoA, and may form a promising basis to begin for an acyl-CoA dehydrogenase (ACAD) system. This can either be expressed heterologously in *E. coli*, although this would also require the additional expression of two electron transferring proteins in order to regenerate FAD in the ACAD active site. Perhaps more easily the BMA pathway could be expressed in an organism that natively expresses isobutyryl-CoA dehydrogenase (IBD).

8.3.2 Library coverage and transferase activity from new constructs

I successfully designed and created a Golden Gate assembly system whereby any combination of ACX and AAT from the libraries could be mixed and assembled into a production plasmid in a one pot reaction. When 15 of the colonies from a mixture of individual ACXs with all of the AAT library were tested, 12 individual AAT sequences were identified. Although the library coverage may not be precisely equal, this shows that the Golden Gate setup can be used to successfully introduce a number of different enzymes from an equimolar mixture. In addition, I cloned two of the AATs that were not observed from sequencing individually with both AtACX4 and ACX#7 into pCAN7, and sequencing of these reactions demonstrated that the unobserved AATs are likely capable of integration into pCAN7. The assembly method described here sets up as a robust platform for ease of future tailoring of the final steps in BMA biosynthesis. Further enzymes can be simply introduced by either synthesis with

corresponding *Esp*3I sites, or through PCR of any gene with the corresponding overhangs and restriction sites.

Despite the successful assembly of the combinatorial library, no BMA formation from **CAN7** strains expressing the new AAT library were observed. This is despite re-sequencing all of the constructs obtained from gene synthesis, as well as sequencing of the final pCAN7 constructs which demonstrated that the ACX and AATs were integrated correctly. As successful BMA production from the ACX library was demonstrated, this lack of AAT activity suggests that transferase expression and activity may have be the problem, particularly as no BA or BIB was formed. An immediate next step must be to assess the solubility of the transferase library. This may be an underlying issue in the screen, as it has been reported previously that some AATs can be challenging to prevent forming aggregates when expressed heterologously. It has been shown however that this can be addressed by modulating the expression of AAT in the new host. If the transferases are aggregating under these new conditions, it is possible to amplify the genes with a set of different promoters, and to incorporate these into the pCAN7 construct to reduce aggregation. A more direct approach may also be to assay a selection of the AAT library *in vitro* as this will allow us to directly assess whether any of the AATs are catalytically functional. IBA production from the combinatorial strains could also be assessed as this would demonstrate whether carbon flux through to IB-CoA is performing normally. As the assembly platform is robust, there is some flexibility in future to modify the AAT library until expression and/or activity can be improved. After this the Golden Gate assembly could be attempted again before re-screening the AAT variants.

8.3.3 IVIS screen as a method for detecting BMA

No strain was identified to express superior BMA-forming activity using the ACX and AAT libraries. This made it challenging to use the fluorescent mDTZ assay to accurately identify high producers as compared to low producers. As described above, the activity of the oxidases in the ACX library is not varied on enough of a

scale to easily distinguish hits without error using the IVIS Spectrum. This means the background signal observed on the footprint where the *E. coli* has grown on the membrane remained the strongest 'output' signal. Because no strain in the combinatorial approach was producing BMA except for the control, the imaging carried out gave data that had a skewed impression over-representing background fluorescence. Much more work will need to be done to improve the production strains, therefore allowing proper determination of the dynamic range of the IVIS screen, before this can be used effectively as a viable screening approach.

The IVIS Spectrum approach may be an improvement from the UV illuminator method, but this cannot be properly determined at the µM scale required. Radiance values for PZ-BMA from our **CAN7**, and the control strains, often still sit outside the range of detection. Significant improvements are required in the BMA biosynthesis pathway, perhaps through much larger changes such as a switch in host organism or class of enzyme used, before the potential benefits of this screening method can truly be harnessed. The positive from all this is that the platform is there, and as the scope is widened for searching and using different components for BMA biosynthesis those improvements may be seen. Many avenues of research are already underway, and there are also clear moves that can be taken as a follow-on from this that will facilitate the movement towards unlocking the potential for bio-BMA production.

Chapter 9: Final conclusions and future work

9.1 Summary of outcomes, aims and objectives

During this work I sought to explore the overarching research question of what factors limit BMA biosynthesis using *E. coli*. The initial hypothesis, that BMA toxicity was a key limitation on overall titres, proved an insufficient explanation after I investigated BMA formation from both WT and product resistant *E. coli* BW25113:

- I constructed a library of production strain variants, comparing inducers, antibiotic resistance, enzyme modifications, and *E. coli* BW25113 BMA resistant mutants (Figure 4-2)
- From this library I identified several ester by-products, with butyl acetate and butyl isobutyrate as major products of BMA fermentation (Figure 4-4 A, B)
- A BMA titre of 0.175 mM over 24 hours was achieved from WT *E. coli* BW25113 pKIV4 pMAE4 (Figure 4-4B)
- I determined that the conditions required for HMA production prevented *E. coli* BW25113 from reaching a sufficient OD_{600} (Figure 4-6)
- One BMA-resistant mutant *E. coli* strain, **RNM-19_K4.M4,** had an improved BMA productivity compared to WT of 2.4 µM h^{-1} OD_{600}^{-1} (Figure 4-9)
- Product toxicity was discounted as the biggest detractor from improving BMA bioproduction in *E. coli* (Chapter 4.3.1)

In the second phase of this work, my objective was to test the hypothesis that BMA production was limited by a metabolic bottleneck, based upon the data provided by AY demonstrating g L^{-1} titres of 3-HIBA using an analogous production system. I achieved this using a library of production strains to isolate metabolism from 2-KIV to BMA, and demonstrated that a key factor was the significant product inhibition of AtACX4 by its product, M-CoA;

- I constructed a series of constitutively expressed BMA pathway variants, producing either BIB, IBA, BMA or 3-HIBA, and tested them using a biotransformation approach (Figure 5-4)
- I produced 159 mM 3-HIBA, 0.026 mM BMA, 1.95 mM BIB, and between 41-63 mM IBA as products from our bottleneck library (Figure 5-5, Figure 5-6, Figure 5-7)
- Bottleneck experiments identified AATm4 as a draw on BMA productivity, with broad activity towards BIV, BA, BPI and BIB limiting BMA specific production (Figure 5-10)
- I His-tag purified AtACX4 and designed an Oxygen electrode assay to assess its activity (Figure 5-11, Figure 5-12)
- I determined the AtACX4 K_M with IB-CoA as 0.14 mM, with a V_{max} of 10.2 nmol mg^{-1} s^{-1} (Figure 5-15)
- M-CoA was demonstrated to cause mixed/non-competitive inhibition of AtACX4 with a K_i of 32.8 µM, limiting M-CoA formation when AATm4 forms part of the pathway (Figure 5-16)

Once I had identified a source of BMA production limitation, namely the non-specific activity of AAT coupled with major product inhibition on AtACX4, the next objective was to identify a method for improving this stage of the pathway. To do this, I investigated the possibility of identifying novel enzymes to catalyse the reactions going from IB-CoA to BMA:

- I used bioinformatics to identify and design two libraries of ACX (Table 7-2) and AAT (Table 7-8) alternatives
- I designed a semi-quantitative fluorescence screen using mDTZ and an IVIS Spectrum to interrogate the ACX and AAT variant libraries with a higher throughput (Figure 6-8)

- A Golden Gate system was constructed and implemented to rapidly assemble alternative ACX and AATs into a production plasmid in *E. coli* BW25113 (Figure 8-3)

The final objective in this project was to combine the information learnt in my pathway characterisation work with the enzyme screening approach designed in the second phase of the project in order to try and find a way to 'bypass' the existing BMA production bottlenecks. To achieve this, I employed my semi-quantitative screen to successfully investigate the library of ACX enzymes I had identified from bioinformatics:

- I identified 7 new ACX enzymes able to utilise IB-CoA: PaACX4, SoACX3, AhACX4, GnACX4, ToACX4, SoACX4, and VrACX4-X2 (Table 8-1)
- I demonstrated the use of our semi-quantitative screen for identifying BMA producers from non-producers in our ACX library (Figure 8-7)
- I discovered the activity of a novel isoform of ACX, ACX3, on IB-CoA (Table 8-1)
- One ACX, SoACX4, produced BMA with a specificity of 1:4:2 BMA:BIB:BA, an improvement from the original AtACX4 (Table 8-2)
- I identified two ACXs with similar BMA titres to AtACX4, SoACX4 and VrACX4-X2 (Figure 8-8)
- I used our Golden Gate assembly system to combinatorially assemble a series of pathway variants (Figure 8-12)

The initial aim of the work was to understand whether the impact of product toxicity on *E. coli* severely impacted BMA formation. I answered this question in the first phase of the project, and was able to investigate the wider overarching question of where these production limitations were occurring during BMA production. In doing so, I provided a platform both for the rapid assembly of BMA pathway variants, as well as a screening method which, with further adaptation may significantly increase the throughput of strain validation.

9.2 Further work

9.2.1 Enhancing the current BMA synthesis platform

9.2.1.1 Metabolite toxicity and carbon flux

One thing that is clear from these BMA formation experiments is that the titre of BMA recorded from *E. coli* producer cells is not, on paper, sufficient to become lethal to *E. coli* BW25113. However, formation of certain metabolites, such as the -CoA intermediates M-CoA and IB-CoA, could possibly negatively impact cell growth and productivity. It is likely that M-CoA may inhibit *E. coli,* and several BMA tolerant mutants performed marginally better than WT *E. coli* BW25113 (Figure 4-9). For example, **RNM-19_K4.M4** had the highest specific productivity of BMA producers assayed in flask. Likewise, **RNM-18_K4.M4** was more selective towards BMA than the WT strain. Therefore, it may be useful to investigate the BMA biosynthesis intermediates in more depth, perhaps in the following ways:

- Metabolomics to assess real intracellular metabolite concentrations. This may address theories about carbon flux at the BCKD stage of production, as well as giving a true value to intracellular M-CoA concentration, which is key to both AtACX4 and AATm4 activity.
- In addition to quantifying the M-CoA pool, it may be useful to look for potential conjugates of M-CoA with intracellular metabolites. M-CoA is known to readily form conjugates with cysteine residues for example [181], which of course are present in most proteins within the cell. This introduces the possibility that M-CoA may be mopped up rapidly by this off-target activity.
- Toxicity of ester intermediates. There is some evidence that BIB causes more disruption to the *E. coli* cell membrane than BMA (MCUK Conference, unpublished). Although BMA toxicity doesn't appear to reach inhibitory levels during the current format of BMA formation, BIB is produced to several fold higher concentrations (Figure 5-7), and therefore may disrupt the *E. coli* cell membrane. Acetyl-CoA depletion by AATm4 to produce BA may also

negatively impact cell growth during BMA production. Investigating the effect of these metabolites on *E. coli* may improve our profile for whether toxicity is in fact impacting BMA formation.

9.2.1.2 Acyltransferases

What is clearly a key issue to tackle as a direct follow on from this work is the functionality of the enzymes in the novel AAT library. So far, this project has not shown evidence that the AATs are catalytically active when expressed using the current strain design and expression system, despite codon optimisation. Therefore, it must be determined whether this lack of activity is due to either: (a) Protein sequence, (b) expression in *E. coli*, or (c) experimental conditions or another section of the pathway. To do this, the following experiments need to be carried out:

- Further metabolite analysis. Although no ester production from the AAT library strains was detected, a search for other metabolites could be carried out in order to determine whether the acyltransferase step is indeed bottlenecking ester production. For example, HPLC analysis could be used to look for IBA production in the producer strains.
- Check for AAT solubility. The AAT library could be used in *E. coli* DH5α to express the proteins, lyse cells, and collect both the soluble and insoluble fractions, as done with the AtACX4 assay earlier in this thesis.
- *in vitro* AAT assays. Although a large task, direct *in vitro* assay using the AATs would reveal quickly whether the issues experienced with ester production arose from inactive enzymes or from expression of otherwise functional acyltransferases in *E. coli*.
- Golden Gate assembly verification. A number of colonies have been sequenced to assess whether multiple AATs are integrated into pCAN-7 during combinatorial assembly. However, many more colonies must be sequenced to give a truly robust picture of the likelihood of different AATs being analysed when testing colonies with the IVIS screen. Sequencing

approximately 80 colonies would provide a more thorough analysis of the Golden Gate combinatorial approach employed in this work, as well as revealing whether there are any AATs that are not integrated into pCAN-7 during assembly.

9.2.1.3 Oxidases

The expression of our novel ACX enzyme library was more fruitful, with several enzymes that permitted BMA production when used in production strains in place of AtACX4. The next steps following on from this discovery should be centred around validating these findings directly to determine whether this does correspond to the oxidases being active on M-CoA specifically.

- *in vitro* ACX assays. The new ACXs, particularly the high activity group, can easily be His-tagged and purified using the pET20b+ vectors into which the genes were synthesised. These can then be assayed using the Oxygraph setup both to compare the K_M against AtACX4, and to check for any increased K_i with M-CoA.
- M-CoA inhibition of MBP-AtACX4. Although the MBP portion of MBP-AtACX4 cleaved during attempts to purify it within this project, it may be possible to achieve this purification with more time to alter buffer composition and purification strategy. If so, then it would be useful to investigate whether the fusion of MBP to AtACX4 affects the K_i for M-CoA in any way.
- In addition to the value for K_i, it would also be useful to improve understanding of the mechanism by which M-CoA inhibits AtACX4 activity. The Oxygraph assay showed that inhibition occurs through a mixed/non-competitive model, and so determining the K_d of M-CoA with AtACX4 would be a logical next step.
- New combinatorial screen. Once issues with AAT activity have been solved in our new transferase library, the combinatorial screening can be attempted as originally planned

- Screening low activity ACX enzymes. Although when paired with AATm4, the low and medium activity ACXs may appear to provide lower flux through IB-CoA to M-CoA (Table 8-1), the possibility that may produce higher BMA titres when in combination with a different AAT enzyme cannot be discounted. Therefore, future testing could be expanded to investigate this possibility.
- ACX3 enzyme search. Due to the high activity of the SoACX3 used in this work, it may be worthwhile to expand the ACX search to focus on ACX3. This would be of particular interest as the only other similar enzyme investigated, SOVF-B from *S. oleracea*, was also capable of producing BMA in the production strain.
- Check AtACX4 mutant library for product resistance. A library of AtACX4 active site mutants has been developed by Ingenza, which can be tested using the IVIS screening setup.
- Development of a new library of ACX mutants. The sequences of several of the highest BMA producers (Figure 8-9) were compared and are well conserved. However, introducing mutations into AtACX4 which substitute amino acids present in other ACXs, particularly SoACX4 and VrACX4-X2, may improve the catalytic activity of our existing ACX.

9.2.1.4 IVIS Spectrum screening

In terms of IVIS Spectrum screen optimisations, there are several further steps that can be taken to improve the robustness, probe the dynamic range, and increase the throughput of the approach:

- Practical adjustments can be made to the physical set-up of the screen in an attempt to improve the throughput. For example, agar plates could be replaced with alternative containers, due to the reflective feedback observed during screen optimisation. Carrying out the screen on one, larger Nylon membrane in a container that will exploit the full size of the IVIS Spectrum photographic range could be investigated. This would increase the number of

spots it is possible to read at a single pass, allowing for increased replicates, or screening of higher numbers of colonies.

- Quantifying BMA method. Attempts to develop a direct calibration of BMA on membranes against BMA produced from colonies were largely abandoned because results on plates could not be easily replicated. With developments to the screening setup since these attempts, it may be worth revisiting spotting BMA onto plates to make the screen more directly quantitative.
- Understanding the dynamic range. In developing our **CAN7_ACX*n*** strains, we have obtained a series of BMA producers which give a wider spectrum of titres than our controls used in Chapter 7. We could employ these as new controls to test the robustness of the screen at a bigger range of BMA productivities. The BMA production from our alternative ACX strains is more diverse than the range given by **CAN4** and **LUC0739/848**.

As referred to in the previous section, another viable way to utilise IVIS screening could involve expanding the screen to interrogate an AtACX4 mutant library produced by Ingenza. These mutants were created by single site mutation at 9 sites of interest in WT AtACX4. The selected locations were identified by MCUK after extensive crystallographic work to investigate IB-CoA binding in AtACX4. All the mutated amino acids are within or adjacent to the binding pocket of AtACX4 (Figure 9-1). Each mutant contains variation at one base pair only. 4 sites are in the region between polypeptides 134-142, 2 at positions 407 and 408, the remaining targeted sites are at residues 172, 291, and 294 (Ingenza UK).

Based on the successes achieved through active site mutation in the cases of cellulase, β-galactosidase and SAT, using the AtACX4 mutant libraries to assay for increased product-resistance is another viable approach that could be taken to improve BMA biosynthesis. It is a simple matter to PCR amplify each of these 9 libraries to introduce *Esp*3I sites with the correct signature sequences, after which they can be assessed using the IVIS Spectrum screen and biotransformation

approach used here. Oligonucleotides MBP-RBS-ACX4-PF and MBP-RBS-ACX4-PR have been designed for this purpose.

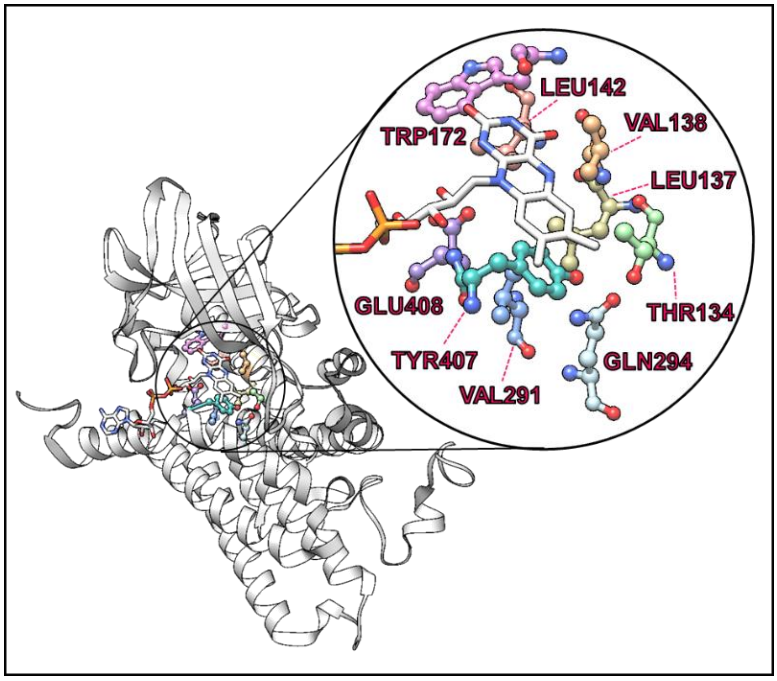


Figure 9-1: Residue selection for site-specific mutation of AtACX4.

AtACX4 subunit B is pictured. Individual residues represented as ball and stick models. FAD shown in white. (PDB ID: 2IX6). Figure made using UCSF Chimera.

9.2.2 Future directions for sustainable methacrylate synthesis

It is suspected that there may be some internal toxicity issues even at current productivity levels, as well as an almost certain toxic effect of BMA to *E. coli* if industrial titres can be achieved. Therefore turning focus to alternative hosts may be a more viable option as the BMA process develops. As well as tackling anticipated toxicity issues more simply, this may also facilitate more improvements in the BMA pathway itself. For example, the potential for replacing ACX with an ACAD for the penultimate synthesis step has been discussed above. This is difficult to achieve but could be investigated more easily in an organism that endogenously expresses the ACAD, ETF, and ETFQ. Some approaches that may facilitate this route of research would be:

- Exchanging *E. coli* for a chassis that is known to be a more robust organism. Most obviously this may include a strain such as *P. putida* KT2440 which is known for having good efflux systems and for easy evolution of resistance to toxic metabolites [263]. Additionally, organisms such as *Lactobacillus brevis* or the more commonly used *B. subtilis* possess a greater level of inherent solvent tolerance than *E. coli* and therefore may also form a good starting point [264]. An on-going study into environmental BMA resistance is underway to identify novel organisms with intrinsic BMA resistance (Thomas Hender, MCUK Conference)
- It would also be useful to consider organisms that naturally produce comparable products. This could involve using an organism that expresses AAT endogenously. Of the 52 organisms reported on BRENDA to express AAT (EC 2.3.1.84), all are from plants or fungi, particularly yeasts. The majority of the yeast AATs are orthologues of Atf1/2, which we know can't produce BMA, but one is from a non-Saccharomyces yeast, *H. valbyensis*. After re-screening the AAT library, it may also provide alternative candidates for endogenous AAT expression.
- In a similar vein, hosts that endogenously express ACXs could be explored. If plants are avoided as chassis, then *G. nicotinae* may be a useful starting point, as this project has already shown that GnACX4 can function in the BMA pathway.
- Finally, organisms that may natively express ACADs can be investigated, in particular IBD. Unfortunately, searching for IBD on either BRENDA or UniProt yields only mammalian or fungal dehydrogenases.

In a similar vein to this,it would be valuable to carry out a more in-depth investigation of the banana acyltransferase responsible for selective BMA production, as observed by MCUK. Only 3 banana AATs have ever been isolated. It may be useful to extract RNA from banana pulp and use AAT cDNA from either AATm4 or the original apple AAT to identify additional acyltransferases native to banana. As discussed, fruits

expressing AATs usually possess many different isoforms and AATs with diverse specificities. It is therefore likely that there are many as yet unidentified AATs present in banana, at least one of which may be responsible for the preferred methacrylyl- rather than acetyl-CoA substrate preference. If several novel sequences can be identified using cDNA, these proteins could be expressed and tested using our BMA production platform.

It is also possible that instead of, or in addition to, alternative organisms, non-BMA targets could once again be considered. As discussed in Chapter 5.2, high titres of IBA and 3-HIBA can be synthesised from *E. coli* expressing BCKD, or BCKD, ECH and HCH. IBA has been expressed from *E. coli* to titres as high as 1 g L^{-1} h^{-1}, but the dehydrogenation reaction to convert IBA to MAA has a conversion rate of only 40% [265]. Meanwhile 3-HIBA can reportedly be converted to MAA at a rate of >90%, making it a promising potential bioprocessing target [265].

9.3 Conclusions

PMMA has been an important industrial product for almost 100 years [65], with a range of used that only appear to expand as the scope for its use in medical technologies is further explored [68-70]. With this in mind, it is imperative that we find a more sustainable way to produce MMA on the scale required, avoiding the significant use of acetone that is presently an essential. Fortunately, there is currently a boom in emerging biotechnologies across the world, and in the UK in particular [266]. A recent report from the Royal Academy of Engineering identified over 1,800 businesses in the UK's industrial biotechnology sector, second only to the US, with recent government investment of £102 million in synthetic biology research [266]. Commercialised bioprocesses for the production of industrially relevant chemicals improve annually, growing closer to the goal of consolidated bioprocessing. In terms of bioplastics like PMMA, PHB has long been a leading success story, with a productivity of 1.6 g L^{-1} h^{-1} achieved from *C. necator* using hyacinth hydrolysate as a sole carbon source [267]. Simpler products are already produced successfully on a

large scale, such as bioethanol, the market for which is worth $58 billion *per annum*, as well as acetic acid and lactic acid [268]. The barriers to achieving success, primarily feedstock supply, process development, and policy adjustments, will increasingly step forward to allow bioprocessing of platform materials towards the target of an increasingly circular bioeconomy.

It is clear than major headway will need to be made at the metabolic engineering stage of BMA biosynthesis to move towards significantly improved production. BMA toxicity can be expected to become a key challenge further down the pipeline, but currently the main problem appears to lie with the fundamental design of BMA metabolism. The bottleneck at M-CoA prevents industrial titres of BMA being achieved, as well as introducing potentially expensive separation steps for the myriad ester by-products, in particular BIB and BA. Finding a solution to this bottleneck is a complex task, but fortunately there are many directions from which this problem can be approached. Looking at producing similar, alternative products such as BIB, or exploring BMA biosynthesis in a more robust host such as *P. pastoris* are perhaps the simplest and potentially most effective ways to rapidly make advances.

In terms of the development of a bioprocessing method for industrial BMA production, this project has provided insight into the bottlenecks in the pathway, as it currently stands. Although at the outset BMA toxicity was assumed to be the primary bottleneck, the CAN-4 and CAN-5 biotransformations demonstrate that the more pressing issue is likely metabolic flux through the pathway. Indeed, the work here describes by-products and enzymatic activities that suggest M-CoA is the focal point of BMA production bottlenecks. This provides a clear target in engineering ACX and AAT towards improved catalytic efficiency in future. This project also provided a method by which future BMA production strains may be screened in a higher throughput manner than flask cultures and fermentation. This, as well as the rapid assembly strategy designed for new CAN-7 production strains, provides a potential roadmap for continued engineering of BMA biocatalysts.

Work concerning toxic end-products in the literature continue to often taken a more conventional evolving product tolerance approach. For example, for the improvement of cadaverine production from *E. coli* by adaptive lab evolution [48]. However, increasingly other routes to address product tolerance are being explored, including engineering of specific regulation factors, or targeted membrane modifications [269]. In one recent study researchers even modified the replicative and chronological lifespan of *E. coli* to successfully improve production of lactate-co-s-hydroxybutyrate [270]. In addition to this, much research has recently focussed on fully exploring libraries of promoters, RBSs and other transcription factors in order to balance metabolite and product toxicity with optimal titres [269]. For example, the exploration of five swappable 'modules' in the violacein pathway [262], or introduction of additional tolerance related genes to reduce metabolic burden [166]. As automation and the ability to screen pathway variations at a higher throughput improves, this will become an increasingly powerful tool to explore the biosynthetic potential of novel metabolic pathways. However to date, the tried and tested method for enhancing production of toxic commodity chemicals remains the modulation of metabolic pathways towards toxic products in combination with tolerance evolution.

Currently incremental improvements to BMA formation are being made from a diverse set of angles, all of which expand our knowledge of methacrylate biosynthesis, increasing the likelihood that we can move closer to the industry 2 g L^{-1} h^{-1} target. This titre may well be achievable in the future as the result of a consortium of research, and will ideally make strides towards decoupling an important manufacturing product from the petrochemicals industry. This is of particular importance at a time where concerns about oil usage, and public opinion on sustainability, have increasingly incentivised an industry wide switch to a more sustainable economy.

References

1. Kerr, R.A., *Oil resources. The looming oil crisis could arrive uncomfortably soon.* Science, 2007. **316**(5823): p. 351.

2. Mohr, S.H., et al., *Projection of world fossil fuels by country.* Fuel, 2015. **141**: p. 120-135.

3. Delannoy, L., et al., *Peak oil and the low-carbon energy transition: A net-energy perspective.* Applied Energy, 2021. **304**.

4. Barton, N.R., et al., *An integrated biotechnology platform for developing sustainable chemical processes.* J Ind Microbiol Biotechnol, 2015. **42**(3): p. 349-60.

5. Elshahed, M., *Microbiological aspects of biofuel production: Current status and future directions.* Journal of Advanced Research, 2010. **1**: p. 103-111.

6. Chambers, G., et al., *Biotech Britain: An assessment of the impact of industrial biotechnology and bioenergy on the United Kingdom economy*. 2015, Capital Economics Limited: UK.

7. Bud, R., *History of 'biotechnology'.* Nature, 1989. **337**(6202): p. 10.

8. Bornscheuer, U.T., et al., *Engineering the third wave of biocatalysis.* Nature, 2012. **485**(7397): p. 185-94.

9. Hulse, J.H., *Biotechnologies: past history, present state and future prospects.* Trends in Food Science & Technology, 2004. **15**(1): p. 3-18.

10. Zhang, Y.H.P., J.B. Sun, and Y.H. Ma, *Biomanufacturing: history and perspective.* Journal of Industrial Microbiology & Biotechnology, 2017. **44**(4-5): p. 773-784.

11. Cavalheiro, J.M.B.T., et al., *Poly(3-hydroxybutyrate) production by Cupriavidus necator using waste glycerol.* Process Biochemistry, 2009. **44**(5): p. 509-515.

12. Gavrilescu, M. and Y. Chisti, *Biotechnology - a sustainable alternative for chemical industry.* Biotechnology Advances, 2005. **23**(7-8): p. 471-499.

13. Leon, R., et al., *Whole-cell biocatalysis in organic media.* Enzyme and Microbial Technology, 1998. **23**(7-8): p. 483-500.

14. Bornscheuer, U.T., *The fourth wave of biocatalysis is approaching.* Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, 2018. **376**(2110): p. 7.

15. Nakamura, C.E. and G.M. Whited, *Metabolic engineering for the microbial production of 1,3-propanediol.* Current Opinion in Biotechnology, 2003. **14**(5): p. 454-459.

16. Gustavsson, M. and S.Y. Lee, *Prospects of microbial cell factories developed through systems metabolic engineering.* Microbial Biotechnology, 2016. **9**(5): p. 610-617.

17. Chandel, A.K., et al., *The role of renewable chemicals and biofuels in building a bioeconomy.* Biofuels Bioproducts & Biorefining-Biofpr, 2020. **14**(4): p. 830-844.

18. Schmid, A., et al., *Industrial biocatalysis today and tomorrow.* Nature, 2001. **409**(6817): p. 258-68.

19. Yiakoumetti, A., *The production of methacrylic acid in Escherichia coli* in *Department of Chemical and Environmental Engineering*. 2016, The University of Nottingham: UK.

20. Almeida, J.R.M., L.C.L. Favaro, and B.F. Quirino, *Biodiesel biorefinery: opportunities and challenges for microbial production of fuels and chemicals from glycerol waste.* Biotechnology for Biofuels, 2012. **5**: p. 16.

21. Kitney, R. and P. Freemont, *UK Synthetic Biology Start-up Survey*. 2017, SynbiCITE: UK.

22. Clarke, L. and R. Kitney, *Developing synthetic biology for industrial*

biotechnology applications. Biochem Soc Trans, 2020. **48**(1): p. 113-122.

23. Company, E.M.F.a.M., *The New Plastics Economy - Rethinking the Future of Plastics*. 2016: World Economic Forum.

24. Turley, D., *Introduction to IBBE in the UK; Key Sectors, markets served by IBBE*. 2018, National Non-Food Crops Centre: UK.

25. Bioplastics, E., *Bioplastics market data 2017*, in *Global Production Capacities of Bioplastics 2017-2022*. 2017, Nova-Institute: Germany. p. 4.

26. Jiang, Y.J., et al., *Consolidated bioprocessing of butanol production from xylan by a thermophilic and butanologenic &ITThermoanaerobacterium&IT sp M5.* Biotechnology for Biofuels, 2018. **11**: p. 14.

27. Jenkinson, L., *Societal challenges, responsible research and innovation, and synthetic biology*. 2018, Green Biologics Ltd.: UK.

28. Van Dien, S., *From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals.* Curr Opin Biotechnol, 2013. **24**(6): p. 1061-8.

29. Golden, J.S., et al., *An Economic Impact Analysis of the U.S. Biobased Products Industry*, in *U.S. Department of Agriculture BioPreferred Program*. 2016, United States Department of Agriculture: U.S.

30. Chung, H., et al., *Bio-based production of monomers and polymers by metabolically engineered microorganisms.* Current Opinion in Biotechnology, 2015. **36**: p. 73-84.

31. Yim, H., et al., *Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol.* Nat Chem Biol, 2011. **7**(7): p. 445-52.

32. Lee, S.Y. and H.U. Kim, *Systems strategies for developing industrial microbial strains.* Nat Biotechnol, 2015. **33**(10): p. 1061-72.

33. Dunlop, M.J., *Engineering microbes for tolerance to next-generation biofuels.* Biotechnology for Biofuels, 2011. **4**: p. 9.

34. Habe, H., et al., *Microbial Production of Glyceric Acid, an Organic Acid That Can Be Mass Produced from Glycerol.* Applied and Environmental Microbiology, 2009. **75**(24): p. 7760-7766.

35. Rymowicz, W., et al., *Citric acid production from glycerol-containing waste of biodiesel industry by Yarrowia lipolytica in batch, repeated batch, and cell recycle regimes.* Applied Microbiology and Biotechnology, 2010. **87**(3): p. 971-979.

36. Yan, Q. and S.S. Fong, *Challenges and Advances for Genetic Engineering of Non-model Bacteria and Uses in Consolidated Bioprocessing.* Frontiers in Microbiology, 2017. **8**: p. 16.

37. Atsumi, S., T. Hanai, and J.C. Liao, *Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels.* Nature, 2008. **451**(7174): p. 86-9.

38. Zhang, B., et al., *Metabolic engineering of Escherichia coli for d-pantothenic acid production.* Food Chem, 2019. **294**: p. 267-275.

39. Xu, D. and L. Zhang, *Increasing Agmatine Production in Escherichia coli through Metabolic Engineering.* J Agric Food Chem, 2019. **67**(28): p. 7908-7915.

40. Li, Z., X. Wang, and H. Zhang, *Balancing the non-linear rosmarinic acid biosynthetic pathway by modular co-culture engineering.* Metab Eng, 2019. **54**: p. 1-11.

41. Jones, J.A., et al., *ePathOptimize: A Combinatorial Approach for Transcriptional Balancing of Metabolic Pathways.* Scientific Reports, 2015. **5**: p. 10.

42. Jezierska, S. and I.N.A. Van Bogaert, *Crossing boundaries: the importance of cellular membranes in industrial biotechnology.* Journal of Industrial Microbiology &

Biotechnology, 2017. **44**(4-5): p. 721-733.

43. Kruis, A.J., et al., *Microbial production of short and medium chain esters: Enzymes, pathways, and applications.* Biotechnol Adv, 2019. **37**(7): p. 107407.

44. Taylor, F., et al., *Kinetics of continuous fermentation and stripping of ethanol.* Biotechnology Letters, 1998. **20**(1): p. 67-72.

45. Qureshi, N., et al., *Genetically engineered Escherichia coli FBR5: part II. Ethanol production from xylose and simultaneous product recovery.* Biotechnol Prog, 2012. **28**(5): p. 1179-85.

46. Rodriguez, G.M., Y. Tashiro, and S. Atsumi, *Expanding ester biosynthesis in Escherichia coli.* Nat Chem Biol, 2014. **10**(4): p. 259-65.

47. Ezeji, T.C., et al., *Improving performance of a gas stripping-based recovery system to remove butanol from Clostridium beijerinckii fermentation.* Bioprocess Biosyst Eng, 2005. **27**(3): p. 207-14.

48. Wang, X., et al., *Ameliorating end-product inhibition to improve cadaverine production in engineered Escherichia coli and its application in the synthesis of bio-based diisocyanates.* Synth Syst Biotechnol, 2021. **6**(4): p. 243-253.

49. Croughan, M.S., K.B. Konstantinov, and C. Cooney, *The Future of Industrial Bioprocessing: Batch or Continuous?* Biotechnology and Bioengineering, 2015. **112**(4): p. 648-651.

50. Chen, C., H.E. Wong, and C.T. Goudar, *Upstream process intensification and continuous manufacturing.* Current Opinion in Chemical Engineering, 2018. **22**: p. 191-198.

51. Lee, S.Y., *High cell-density culture of Escherichia coli.* Trends Biotechnol, 1996. **14**(3): p. 98-105.

52. Kopp, J., et al., *The Rocky Road From Fed-Batch to Continuous Processing With E. coli.* Front Bioeng Biotechnol, 2019. **7**: p. 328.

53. Schugerl, K. and J. Hubbuch, *Integrated bioprocesses.* Current Opinion in Microbiology, 2005. **8**(3): p. 294-300.

54. Stephens, G., et al., *Presentation: Bio-based methacrylates*. 2017, University of Nottingham: UK.

55. Schugerl, K., *Integrated processing of biotechnology products.* Biotechnology Advances, 2000. **18**(7): p. 581-599.

56. Woodley, J.M., et al., *Future directions for in-situ product removal (ISPR).* Journal of Chemical Technology and Biotechnology, 2008. **83**(2): p. 121-123.

57. Kim, J.H., et al., *Isoprene production by Escherichia coli through the exogenous mevalonate pathway with reduced formation of fermentation byproducts.* Microbial Cell Factories, 2016. **15**: p. 10.

58. Burgard, A., et al., *Microorganisms and methods for producing butadiene and related compounds by formate assimilation*, W.I.P. Organisation, Editor. 2014, Genomatica, Inc.: US.

59. McKenna, R. and D.R. Nielsen, *Styrene biosynthesis from glucose by engineered E. coli.* Metabolic Engineering, 2011. **13**(5): p. 544-554.

60. Yoo, T. and S.K. Henning, *Synthesis and Characterization of Farnesene-Based Polymers.* Rubber Chemistry and Technology, 2017. **90**(2): p. 308-324.

61. Eppler, R. and D. Chan, *Compositions containing bio-based farnesene or compunds derived therefrom and their use in consumer and industrial products*, in *Patent Cooperation Treaty*, W.I.P.O.I. Bureau, Editor. 2016, Amyris, Inc.: US.

62. Eastham, G., et al., *A process for production of methacrylic acid and derivatives thereof*, W.I.P. Organisation, Editor. 2015, Lucite International UK Ltd.: UK.

63. Burg, M., et al., *Microorganisms for producing methacrylic acid and*

methacrylate esters and methods related thereto, E.P. Application, Editor. 2014, Genomatica, Inc.

64. Webb, J., *Bio-production of C5 dicarboxylic acids for chemical conversion to methyl methacrylate*, in *Department of Chemical and Environmental Engineering*. 2015, University of Nottingham: UK. p. 211.

65. Chisholm, M.S., *Artificial glass - The versatility of poly(methyl methacrylate) from its early exploitation to the new millennium.* Journal of Chemical Education, 2000. **77**(7): p. 841-845.

66. Ltd., I.C.I., *'Perspex'; acrylic materials*. 2nd ed. 1952, Welwyn Garden City: Imperial Chemical Industries.

67. Salkind, M., E.H. Riddle, and R.W. Keefer, *Acrylates and Methacrylates Ester Manufacture and Markets.* Industrial and Engineering Chemistry, 1959. **51**(11): p. 1328-1334.

68. Hosni, F., et al., *A study of the fractionation dose on the radiation response of Harwell Red-Perspex PMMA dosimeter.* Nuclear Instruments & Methods in Physics Research Section B-Beam Interactions with Materials and Atoms, 2012. **290**: p. 69-71.

69. Nottrott, M., *Acrylic bone cements: influence of time and environment on physical properties.* Acta Orthop Suppl, 2010. **81**(341): p. 1-27.

70. Barba, A., et al., *Synthesis and characterization of P(MMA-AA) copolymers for targeted oral drug delivery.* Polymer Bulletin, 2009. **62**(5): p. 679-688.

71. Crawford, J., *Improvements in or relating to the manufacture of esters of methacrylic acid*. 1934.

72. Huddle, T.A., *The continuous generation of acrylic monomers from renewable feedstocks in high temperature water*, C. University of Nottingham. Theses, Editor. 2012, University of Nottingham: UK.

73. Harris, B., *Acrylics for the future*, in *Ingenia*. 2010, Ingenia. p. 3.

74. Webb, J.P., et al., *Efficient bio-production of citramalate using an engineered Escherichia coli strain.* 2018.

75. Rohwerder, T. and R.H. Muller, *Biosynthesis of 2-hydroxyisobutyric acid (2-HIBA) from renewable carbon.* Microbial Cell Factories, 2010. **9**: p. 10.

76. Lang, K., K. Buehler, and A. Schmid, *Multistep Synthesis of (S)-3-Hydroxyisobutyric Acid from Glucose using Pseudomonas taiwanensis VLB120 B83 T7 Catalytic Biofilms.* Advanced Synthesis & Catalysis, 2015. **357**(8): p. 1919-1927.

77. Disley, Z., *Towards the bioproduction of methyl methacrylate: solving the problem of product toxicity*, in *Department of Chemical and Environmental Engineering*. 2016, The University of Nottingham: UK.

78. Teo, W.S., et al., *Metabolic engineering of Saccharomyces cerevisiae for production of fatty acid short- and branched-chain alkyl esters biodiesel.* Biotechnol Biofuels, 2015. **8**(177): p. 177.

79. Kalscheuer, R., et al., *Neutral lipid biosynthesis in engineered Escherichia coli: Jojoba oil-like wax esters and fatty acid butyl esters.* Applied and Environmental Microbiology, 2006. **72**(2): p. 1373-1379.

80. Kim, S., et al., *Enhancement of carbon dioxide reduction and methane production by an obligate anaerobe and gas dissolution device.* New Biotechnology, 2016. **33**(1): p. 196-205.

81. Martins, L., *Understanding butyl methacrylate tolerance in Escherichia coli mutant strains*. 2017, The University of Nottingham: UK.

82. Nicolaou, S.A., S.M. Gaida, and E.T. Papoutsakis, *A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to*

biocatalysis and bioremediation. Metabolic Engineering, 2010. **12**(4): p. 307-331.

83. Laane, C., et al., *Rules for Optimization of Biocatalysis in Organic Solvents.* Biotechnology and Bioengineering, 2009. **102**(1): p. 2-8.

84. Fujisawa, S. and E. Masuhara, *Determination of Partition-Coefficients of Acrylates, Methacrylates, and Vinyl Monomers Using High-Performance Liquid-Chromatography (Hplc).* Journal of Biomedical Materials Research, 1981. **15**(6): p. 789-793.

85. Yeh, V., et al., *Membrane Stability in the Presence of Methacrylate Esters.* Langmuir, 2020. **36**(33): p. 9649-9657.

86. Takatsuka, Y., C. Chen, and H. Nikaido, *Mechanism of recognition of compounds of diverse structures by the multidrug efflux pump AcrB of Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(15): p. 6559-6565.

87. Ankarloo, J., S. Wikman, and I.A. Nicholls, *Escherichia coli mar and acrAB Mutants Display No Tolerance to Simple Alcohols.* International Journal of Molecular Sciences, 2010. **11**(4): p. 1403-1412.

88. Jain, K. and S. Saini, *MarRA, SoxSR, and Rob encode a signal dependent regulatory network in Escherichia coli.* Molecular Biosystems, 2016. **12**(6): p. 1901-1912.

89. Dietrich, L.E. and P.J. Kiley, *A shared mechanism of SoxR activation by redox-cycling compounds.* Mol Microbiol, 2011. **79**(5): p. 1119-22.

90. Menchavez, R., *Host cell engineering for the production of methacrylate esters*. 2018, The University of Nottingham: UK.

91. Vinue, L., L.M. McMurry, and S.B. Levy, *The 216-bp marB gene of the marRAB operon in Escherichia coli encodes a periplasmic protein which reduces the transcription rate of marA.* Fems Microbiology Letters, 2013. **345**(1): p. 49-55.

92. Wang, Z., et al., *An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump.* Elife, 2017. **6**: p. 19.

93. Park, J.H., et al., *Fed-Batch Culture of Escherichia coli for L-Valine Production Based on In Silico Flux Response Analysis.* Biotechnology and Bioengineering, 2011. **108**(4): p. 934-946.

94. Lejona, S., et al., *PhoP can activate its target genes in a PhoQ-independent manner.* Journal of Bacteriology, 2004. **186**(8): p. 2476-2480.

95. Aono, R., *Improvement of organic solvent tolerance level of Escherichia coli by overexpression of stress-responsive genes.* Extremophiles, 1998. **2**(3): p. 239-48.

96. Dunlop, M.J., *Engineering microbes for tolerance to next-generation biofuels.* Biotechnol Biofuels, 2011. **4**: p. 32.

97. Park, Y.C., C.E. Shaffer, and G.N. Bennett, *Microbial formation of esters.* Appl Microbiol Biotechnol, 2009. **85**(1): p. 13-25.

98. Fischer, E. and A. Speier, *Darstellung der Ester.* Berichte der deutschen chemischen Gesellschaft, 1895. **28**(3): p. 3252-3258.

99. Menendez-Bravo, S., et al., *Metabolic engineering of microorganisms for the production of structurally diverse esters.* Appl Microbiol Biotechnol, 2017. **101**(8): p. 3043-3053.

100. Lamare, S., et al., *Production of Natural Esters at the Pre-Industrial Scale by Solid/Gas Biocatalysis.* Biocatalysis and Biotransformation, 2009. **19**(5-6): p. 361-377.

101. Guo, D., H. Pan, and X. Li, *Metabolic engineering of Escherichia coli for production of biodiesel from fatty alcohols and acetyl-CoA.* Appl Microbiol Biotechnol, 2015. **99**(18): p. 7805-12.

102. Tucci, S., et al., *Variability of wax ester fermentation in natural and bleached Euglena gracilis Strains in response to oxygen and the elongase inhibitor flufenacet.* J Eukaryot Microbiol, 2010. **57**(1): p. 63-9.

103. Hari Krishna, S. and N.G. Karanth, *Lipases and Lipase-Catalyzed Esterification Reactions in Nonaqueous Media.* Catalysis Reviews, 2002. **44**(4): p. 499-591.

104. Onwueme, K.C., et al., *The dimycocerosate ester polyketide virulence factors of mycobacteria.* Prog Lipid Res, 2005. **44**(5): p. 259-302.

105. Lortie, R., *Enzyme catalyzed esterification.* Biotechnology Advances, 1997. **15**(1): p. 1-15.

106. Kusano, M., et al., *A novel hemiacetal dehydrogenase activity involved in ethyl acetate synthesis in Candida utilis.* J Biosci Bioeng, 1999. **87**(5): p. 690-2.

107. Iwaki, H., et al., *Cloning and characterization of a gene cluster involved in cyclopentanol metabolism in Comamonas sp. strain NCIMB 9872 and biotransformations effected by Escherichia coli-expressed cyclopentanone 1,2-monooxygenase.* Appl Environ Microbiol, 2002. **68**(11): p. 5671-84.

108. Kotani, T., et al., *Novel acetone metabolism in a propane-utilizing bacterium, Gordonia sp. strain TY-5.* J Bacteriol, 2007. **189**(3): p. 886-93.

109. Olias, R., A.G. Perez, and C. Sanz, *Catalytic properties of alcohol acyltransferase in different strawberry species and cultivars.* J Agric Food Chem, 2002. **50**(14): p. 4031-6.

110. Christiaens, J.F., et al., *The fungal aroma gene ATF1 promotes dispersal of yeast cells through insect vectors.* Cell Rep, 2014. **9**(2): p. 425-32.

111. Yamauchi, H., et al., *Cell-free Synthesis of Ethyl Hexanoate by Extract fromNeurosporasp., Containing a Novel Acyl Coenzyme A: Alcohol Acyltransferase.* Agricultural and Biological Chemistry, 2014. **53**(3): p. 821-825.

112. Marshall, V.D. and J.R. Sokatch, *Regulation of valine catabolism in Pseudomonas putida.* Journal of Bacteriology, 1972. **110**(3): p. 1073-&.

113. Tai, Y.S., M. Xiong, and K. Zhang, *Engineered biosynthesis of medium-chain esters in Escherichia coli.* Metab Eng, 2015. **27**: p. 20-28.

114. Yu, P., et al., *Engineering artificial metabolic pathways for biosynthesis.* Current Opinion in Chemical Engineering, 2012. **1**(4): p. 373-379.

115. Zhang, K.C., et al., *A Synthetic Metabolic Pathway for Production of the Platform Chemical Isobutyric Acid.* Chemsuschem, 2011. **4**(8): p. 1068-1070.

116. Bi, H.P., et al., *Engineered short branched-chain acyl-CoA synthesis in E-coli and acylation of chloramphenicol to branched-chain derivatives.* Applied Microbiology and Biotechnology, 2013. **97**(24): p. 10339-10348.

117. Mooney, B.P., J.A. Miernyk, and D.D. Randall, *The complex fate of alpha-ketoacids.* Annu Rev Plant Biol, 2002. **53**: p. 357-75.

118. Mackenzie, J., et al., *Controlling electron transfer in Acyl-CoA oxidases and dehydrogenases: a structural view.* J Biol Chem, 2006. **281**(41): p. 31012-20.

119. Cardoso, A.R., P. Kakimoto, and A.J. Kowaltowski, *Diet-Sensitive Sources of Reactive Oxygen Species in Liver Mitochondria: Role of Very Long Chain Acyl-CoA Dehydrogenases.* Plos One, 2013. **8**(10): p. 13.

120. Bremer, J. and A.B. Wojtczak, *Factors controlling the rate of fatty acid -oxidation in rat liver mitochondria.* Biochim Biophys Acta, 1972. **280**(4): p. 515-30.

121. Kim, J.J. and R. Miura, *Acyl-CoA dehydrogenases and acyl-CoA oxidases. Structural basis for mechanistic similarities and*

differences. Eur J Biochem, 2004. **271**(3): p. 483-93.

122. Battaile, K.P., et al., *Structures of isobutyryl-CoA dehydrogenase and enzyme-product complex: comparison with isovaleryl- and short-chain acyl-CoA dehydrogenases.* J Biol Chem, 2004. **279**(16): p. 16526-34.

123. Pedersen, L. and A. Henriksen, *Expression, purification and crystallization of two peroxisomal acyl-CoA oxidases from Arabidopsis thaliana.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 6): p. 1125-8.

124. Arent, S., V.E. Pye, and A. Henriksen, *Structure and function of plant acyl-CoA oxidases.* Plant Physiol Biochem, 2008. **46**(3): p. 292-301.

125. Hayashi, H., et al., *A novel acyl-CoA oxidase that can oxidize short-chain acyl-CoA in plant peroxisomes.* J Biol Chem, 1999. **274**(18): p. 12715-21.

126. Camoes, F., et al., *New insights into the peroxisomal protein inventory: Acyl-CoA oxidases and -dehydrogenases are an ancient feature of peroxisomes.* Biochim Biophys Acta, 2015. **1853**(1): p. 111-25.

127. Gerbling, H. and B. Gerhardt, *Oxidative decarboxylation of branched-chain 2-oxo Fatty acids by higher plant peroxisomes.* Plant Physiol, 1988. **88**(1): p. 13-5.

128. Ghisla, S. and C. Thorpe, *Acyl-CoA dehydrogenases. A mechanistic overview.* Eur J Biochem, 2004. **271**(3): p. 494-508.

129. Zhang, Y.X., et al., *Genes encoding acyl-CoA dehydrogenase (AcdH) homologues from Streptomyces coelicolor and Streptomyces avermitilis provide insights into the metabolism of small branched-chain fatty acids and macrolide antibiotic production.* Microbiology, 1999. **145 (Pt 9)**: p. 2323-34.

130. Maggio-Hall, L.A. and N.P. Keller, *Mitochondrial beta-oxidation in Aspergillus nidulans.* Mol Microbiol, 2004. **54**(5): p. 1173-85.

131. Korotkova, N., et al., *Glyoxylate regeneration pathway in the methylotroph Methylobactetaum extorquens AM1.* Journal of Bacteriology, 2002. **184**(6): p. 1750-1758.

132. Defilippi, B.G., A.A. Kader, and A.M. Dandekar, *Apple aroma: alcohol acyltransferase, a rate limiting step for ester biosynthesis, is regulated by ethylene.* Plant Science, 2005. **168**(5): p. 1199-1210.

133. Layton, D.S. and C.T. Trinh, *Expanding the modular ester fermentative pathways for combinatorial biosynthesis of esters from volatile organic acids.* Biotechnol Bioeng, 2016. **113**(8): p. 1764-76.

134. Rottig, A. and A. Steinbuchel, *Acyltransferases in bacteria.* Microbiol Mol Biol Rev, 2013. **77**(2): p. 277-321.

135. Beekwilder, J., et al., *Functional characterization of enzymes forming volatile esters from strawberry and banana.* Plant Physiol, 2004. **135**(4): p. 1865-78.

136. Tashiro, Y., S.H. Desai, and S. Atsumi, *Two-dimensional isobutyl acetate production pathways to improve carbon yield.* Nat Commun, 2015. **6**: p. 7488.

137. El-Sharkawy, I., et al., *Functional characterization of a melon alcohol acyl-transferase gene family involved in the biosynthesis of ester volatiles. Identification of the crucial role of a threonine residue for enzyme activity*.* Plant Mol Biol, 2005. **59**(2): p. 345-62.

138. Souleyre, E.J., et al., *An alcohol acyl transferase from apple (cv. Royal Gala), MpAAT1, produces esters involved in apple fruit flavor.* FEBS J, 2005. **272**(12): p. 3132-44.

139. Noh, H.J., et al., *Metabolic engineering of Clostridium acetobutylicum for the production of butyl butyrate.* Appl Microbiol Biotechnol, 2018. **102**(19): p. 8319-8327.

140. Mills, T.Y., N.R. Sandoval, and R.T. Gill, *Cellulosic hydrolysate toxicity and tolerance mechanisms in Escherichia coli.* Biotechnology for Biofuels, 2009. **2**: p. 11.

141. Rau, M.H., et al., *Genome-wide Escherichia coli stress response and improved tolerance towards industrially relevant chemicals.* Microbial Cell Factories, 2016. **15**: p. 18.

142. Dahl, R.H., et al., *Engineering dynamic pathway regulation using stress-response promoters.* Nature Biotechnology, 2013. **31**(11): p. 1039-+.

143. Bertani, G., *Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli.* J Bacteriol, 1951. **62**(3): p. 293-300.

144. Hanahan, D., *Studies on transformation of Escherichia coli with plasmids.* Journal of Molecular Biology, 1983. **166**(4): p. 557-580.

145. Woo, K.S., et al., *Characteristics of the Thermal Degradation of Glucose and Maltose Solutions.* Prev Nutr Food Sci, 2015. **20**(2): p. 102-9.

146. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6640-5.

147. Ltd, I., *Development of a scalable and cost-competitive bioprocess for manufacture of methacrylate esters*, in *MCC Quarterly Report*. 2017, Roslin BioCentre: Scotland.

148. Flentie, K., et al., *A bioluminescent transposon reporter-trap identifies tumor-specific microenvironment-induced promoters in Salmonella for conditional bacterial-based tumor therapy.* Cancer Discov, 2012. **2**(7): p. 624-37.

149. Weber, H., et al., *Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity.* J Bacteriol, 2005. **187**(5): p. 1591-603.

150. Oktyabr'skii, O.N., et al., *The role of thiol redox systems in the response of Escherichia coli to far-UV irradiation.* Microbiology, 2009. **78**(3): p. 290-295.

151. Nakata, K., et al., *All genomic mutations in the antimicrobial surfactant-resistant mutant, Escherichia coli OW66, are involved in cell resistance to surfactant.* Appl Microbiol Biotechnol, 2010. **87**(5): p. 1895-905.

152. Park, K., et al., *Novel cF-dependent genes of Escherichia coli found using a specified promoter consensus.* FEMS Microbiology Letters, 2001. **202**: p. 243-250.

153. Xu, H., et al., *Role of the lpxM lipid A biosynthesis pathway gene in pathogenicity of avian pathogenic Escherichia coli strain E058 in a chicken infection model.* Vet Microbiol, 2013. **166**(3-4): p. 516-26.

154. Sahara, T., et al., *Cloning, sequencing, and expression of a gene encoding the monomeric isocitrate dehydrogenase of the nitrogen-fixing bacterium, Azotobacter vinelandii.* Biosci Biotechnol Biochem, 2002. **66**(3): p. 489-500.

155. Kornberg, H., *Routes for fructose utilization by Escherichia coli.* Journal of Molecular Microbiology and Biotechnology, 2001. **3**(3): p. 335-339.

156. Riley, A.T., et al., *Semi-enzymatic synthesis of pseudouridine.* Bioorg Med Chem Lett, 2021. **44**: p. 128105.

157. Rau, M.H., et al., *Genome-wide Escherichia coli stress response and improved tolerance towards industrially relevant chemicals.* Microb Cell Fact, 2016. **15**(1): p. 176.

158. El Hadi, M.A., et al., *Advances in fruit aroma volatile research.* Molecules, 2013. **18**(7): p. 8200-29.

159. Dixon, J. and E. Hewett, *Factors affecting apple aroma/flavour volatile concentration: A Review.* New Zealand Journal of Crop and

Horticultural Science, 2010. **28**(3): p. 155-173.

160. Takamura, Y. and G. Nomura, *Changes in the intracellular concentration of acetyl-CoA and malonyl-CoA in relation to the carbon and energy metabolism of Escherichia coli K12.* J Gen Microbiol, 1988. **134**(8): p. 2249-53.

161. Bastian, S., et al., *Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli.* Metab Eng, 2011. **13**(3): p. 345-52.

162. Landmann, C., et al., *Substrate promiscuity of a rosmarinic acid synthase from lavender (Lavandula angustifolia L.).* Planta, 2011. **234**(2): p. 305-20.

163. Balbontin, C., et al., *VpAAT1, a gene encoding an alcohol acyltransferase, is involved in ester biosynthesis during ripening of mountain papaya fruit.* J Agric Food Chem, 2010. **58**(8): p. 5114-21.

164. Ansari, S., et al., *A newly identified prophage-encoded gene, ymfM, causes SOS-inducible filamentation in Escherichia coli.* J Bacteriol, 2021.

165. Tropp, B.E., *Cardiolipin synthase from Escherichia coli.* Biochim Biophys Acta, 1997. **1348**(1-2): p. 192-200.

166. Jung, H.R., et al., *Increased Tolerance to Furfural by Introduction of Polyhydroxybutyrate Synthetic Genes to Escherichia coli.* J Microbiol Biotechnol, 2019. **29**(5): p. 776-784.

167. Suppini, J.P., et al., *Complementation of an Escherichia coli DnaK defect by Hsc70-DnaK chimeric proteins.* J Bacteriol, 2004. **186**(18): p. 6248-53.

168. Neumann, P., et al., *Structural basis for membrane binding and catalytic activation of the peripheral membrane enzyme pyruvate oxidase from Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(45): p. 17390-17395.

169. Steinman, H.M. and R.L. Hill, *Bovine Liver Crotonase (enoyl coenzyme A hydratase).* Methods in Enzymology, 1975. **35**: p. 136-151.

170. Wanders, R.J., M. Duran, and F.J. Loupatty, *Enzymology of the branched-chain amino acid oxidation disorders: the valine pathway.* J Inherit Metab Dis, 2012. **35**(1): p. 5-12.

171. Li, D., et al., *Molecular cloning and expression of a gene encoding alcohol acyltransferase (MdAAT2) from apple (cv. Golden Delicious).* Phytochemistry, 2006. **67**(7): p. 658-67.

172. Aharoni, A., et al., *Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays.* Plant Cell, 2000. **12**(5): p. 647-62.

173. Stern, J.R. and A. del Campillo, *Enzymes of Fatty Acid Metabolism: Properties of Crystalline Crotonase.* Journal of Biological Chemistry, 1956. **218**(2): p. 985-1002.

174. Holland, P.C., A.E. Senior, and H.S. Sherratt, *Biochemical effects of the hypoglycaemic compound pent-4-enoic acid and related non-hypoglycaemic fatty acids. Effects of their coenzyme A esters on enzymes of fatty acid oxidation.* Biochem J, 1973. **136**(1): p. 173-84.

175. Waterson, R.M., et al., *Purification and Characterization of Crotonase from Clostridium acetobutylicum.* Journal of Biological Chemistry, 1972. **247**(16): p. 5266-5271.

176. Fong, J.C.S., H., *Short-chain and long-chain enoyl-CoA hydratases from pig heart muscle.* Methods in Enzymology, 1981. **71**(Lipids Part C): p. 390-398.

177. Stern, J.R., A. del Campillo, and I. Raw, *Enzymes of Fatty Acid Metabolism: General Introduction; Crystalline Crotonase.* Journal of Biological Chemistry, 1956. **218**(2): p. 971-983.

178. Fujita, Y., T. Shimakata, and T. Kusaka, *Purification of two forms of enoyl-CoA hydratase from*

Mycobacterium smegmatis. J Biochem, 1980. **88**(4): p. 1045-50.

179. Binstock, J.F.S., H., *Fatty Acid Oxidation Complex from Escherichia coli.* Methods in Enzymology, 1981. **71**: p. 403-411.

180. Wilbanks, B. and C.T. Trinh, *Comprehensive characterization of toxicity of fermentative metabolites on microbial growth.* Biotechnol Biofuels, 2017. **10**: p. 262.

181. Brown, G.K., et al., *beta-hydroxyisobutyryl coenzyme A deacylase deficiency: a defect in valine metabolism associated with physical malformations.* Pediatrics, 1982. **70**(4): p. 532-8.

182. Speir, T.W. and E.A. Barnsley, *The conjugation of glutathione with unsaturated acyl thiol esters and the metabolic formation ofS-carboxyalkylcysteines.* Biochemical Journal, 1971. **125**(1): p. 267-273.

183. Li, F., et al., *Identification of the wax ester synthase/acyl-coenzyme A: diacylglycerol acyltransferase WSD1 required for stem wax ester biosynthesis in Arabidopsis.* Plant Physiol, 2008. **148**(1): p. 97-107.

184. Liu, C.L., et al., *Multi-modular engineering for renewable production of isoprene via mevalonate pathway in Escherichia coli.* J Appl Microbiol, 2019. **126**(4): p. 1128-1139.

185. Avalos, J.L., G.R. Fink, and G. Stephanopoulos, *Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols.* Nature Biotechnology, 2013. **31**(4): p. 335-+.

186. Brat, D., et al., *Cytosolic re-localization and optimization of valine synthesis and catabolism enables increased isobutanol production with the yeast Saccharomyces cerevisiae.* Biotechnology for Biofuels, 2012. **5**: p. 16.

187. Lucchetta, L., et al., *Biochemical and catalytic properties of three recombinant alcohol acyltransferases of melon. sulfur-containing ester formation, regulatory role of CoA-SH in activity, and sequence elements conferring substrate preference.* J Agric Food Chem, 2007. **55**(13): p. 5213-20.

188. Anderson, J. *Promoters/Catalog/Anderson*. Registry of Standard Biological Parts 2006 [cited 2022 04/09/2022].

189. Olins, P. and S. Rangwala, *A novel sequence element derived from bacteriophage-T7 messenger-RNA acts as an enhancer of translation of the lacZ gene in Escherichia coli. .* Journal of Biological Chemistry, 1989. **264**(29): p. 16973-16976.

190. Sztajer, H., G. Wagner, and R.D. Schmid, *Bacterial Short-Chain Acyl-Coa Oxidase - Production, Purification and Characterization.* Applied Microbiology and Biotechnology, 1993. **39**(6): p. 708-713.

191. Gerhardt, B., *Substrate-Specificity of Peroxisomal Acyl-Coa Oxidase.* Phytochemistry, 1985. **24**(2): p. 351-352.

192. Froman, B.E., et al., *ACX3, a novel medium-chain acyl-coenzyme A oxidase from Arabidopsis.* Plant Physiol, 2000. **123**(2): p. 733-42.

193. Hooks, M.A., K. Bode, and I. Couee, *Higher-plant medium- and short-chain acyl-CoA oxidases: identification, purification and characterization of two novel enzymes of eukaryotic peroxisomal beta-oxidation.* Biochem J, 1996. **320 (Pt 2)**: p. 607-14.

194. Ferdinandusse, S., et al., *Adult peroxisomal acyl-coenzyme A oxidase deficiency with cerebellar and brainstem atrophy.* J Neurol Neurosurg Psychiatry, 2010. **81**(3): p. 310-2.

195. Agnihotri, G. and H.W. Liu, *Enoyl-CoA hydratase. reaction, mechanism, and inhibition.* Bioorg Med Chem, 2003. **11**(1): p. 9-20.

196. Pak, Y.L., K.M. Swamy, and J. Yoon, *Recent Progress in Fluorescent Imaging Probes.*

Sensors (Basel), 2015. **15**(9): p. 24374-96.

197. Tang, Y., et al., *Development of fluorescent probes based on protection-deprotection of the key functional groups for biological imaging.* Chem Soc Rev, 2015. **44**(15): p. 5003-15.

198. Benassi, A., et al., *The Quest for the Right Trade-Off for an Efficient Photoclick Monitoring Reaction.* ChemPhotoChem, 2021. **6**(3).

199. Corporation, X., *IVIS® Spectrum Hardware Manual*. 2007.

200. Chen, T., et al., *Efficient Biofilm-Based Fermentation Strategies for L-Threonine Production by Escherichia coli.* Front Microbiol, 2019. **10**: p. 1773.

201. Gauthier, J., et al., *A brief history of bioinformatics.* Brief Bioinform, 2019. **20**(6): p. 1981-1996.

202. Alterovitz, G., T. Muso, and M.F. Ramoni, *The challenges of informatics in synthetic biology: from biomolecular networks to artificial organisms.* Brief Bioinform, 2010. **11**(1): p. 80-95.

203. Duggleby, R.G., *Product inhibition of reversible enzyme-catalysed reactions.* Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1994. **1209**(2): p. 238-240.

204. Zou, G., et al., *Alleviating product inhibition of Trichoderma reesei cellulase complex with a product-activated mushroom endoglucanase.* Bioresour Technol, 2021. **319**: p. 124119.

205. Hu, X., et al., *Engineering of a fungal β-galactosidase to remove product inhibition by galactose.* Applied Microbiology and Biotechnology, 2010. **87**(5): p. 1773-1782.

206. Kai, Y., et al., *Engineering of Escherichia colil-serine O-acetyltransferase on the basis of crystal structure: desensitization to feedback inhibition by l-cysteine.* Protein Engineering, Design and Selection, 2006. **19**(4): p. 163-167.

207. Atreya, M.E., K.L. Strobel, and D.S. Clark, *Alleviating product inhibition in cellulase enzyme Cel7A.* Biotechnol Bioeng, 2016. **113**(2): p. 330-8.

208. Ishihara, H., et al., *Enhancement of the viscometric endocellulase activity of Polyporus arcularius CMCase IIIa by cellobiose and cellooligosaccharides.* Mycoscience, 2005. **46**(3): p. 148-153.

209. Yoshioka, K. and N. Hashimoto, *Ester Formation by Alcohol Acetyltransferase from Brewers-Yeast.* Agricultural and Biological Chemistry, 1981. **45**(10): p. 2183-2190.

210. Harada, M., Y. Ueda, and T. Iwata, *Purification and Some Properties of Alcohol Acetyltransferase from Banana Fruit.* Plant and Cell Physiology, 1985. **26**(6): p. 1067-1074.

211. Malcorps, P. and J.P. Dufour, *Short-chain and medium-chain aliphatic-ester synthesis in Saccharomyces cerevisiae.* Eur J Biochem, 1992. **210**(3): p. 1015-22.

212. O'Brien, P.J. and D. Herschlag, *Catalytic promiscuity and the evolution of new enzymatic activities.* Chemistry & Biology, 1999. **6**(4): p. R91-R105.

213. Mellor, J., et al., *Semisupervised Gaussian Process for Automated Enzyme Search.* ACS Synth Biol, 2016. **5**(6): p. 518-28.

214. Nam, K.H., N. Dudareva, and E. Pichersky, *Characterization of benzylalcohol acetyltransferases in scented and non-scented Clarkia species.* Plant Cell Physiol, 1999. **40**(9): p. 916-23.

215. Bayer, A., X. Ma, and J. Stockigt, *Acetyltransfer in natural product biosynthesis--functional cloning and molecular analysis of vinorine synthase.* Bioorg Med Chem, 2004. **12**(10): p. 2787-95.

216. Song, Z.Z., et al., *Site-directed mutagenesis identified the key active site residues of alcohol acyltransferase PpAAT1 responsible for aroma biosynthesis in peach fruits.* Hortic Res, 2021. **8**(1): p. 32.

217. Roulet, J., et al., *Modification of PapA5 acyltransferase substrate selectivity for optimization of short-chain alcohol-derived multimethyl-branched ester production in Escherichia coli.* Applied Microbiology and Biotechnology, 2020. **104**(20): p. 8705-8718.

218. Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation.* Nucleic Acids Research, 2021.

219. Zhou, X., et al., *Low temperature conditioning prevents loss of aroma-related esters from 'Nanguo' pears during ripening at room temperature.* Postharvest Biology and Technology, 2015. **100**: p. 23-32.

220. Shalit, M., et al., *Volatile ester formation in roses. Identification of an acetyl-coenzyme A. Geraniol/Citronellol acetyltransferase in developing rose petals.* Plant Physiol, 2003. **131**(4): p. 1868-76.

221. Dexter, R., et al., *Characterization of a petunia acetyltransferase involved in the biosynthesis of the floral volatile isoeugenol.* Plant J, 2007. **49**(2): p. 265-75.

222. Saerens, S.M., et al., *The Saccharomyces cerevisiae EHT1 and EEB1 genes encode novel enzymes with medium-chain fatty acid ethyl ester synthesis and hydrolysis capacity.* J Biol Chem, 2006. **281**(7): p. 4446-56.

223. Lefort, F., et al., *Characterization of new microsatellite loci from Vitis vinifera and their conservation in some Vitis species and hybrids.* Molecular Ecology Notes, 2002. **2**(1): p. 20-21.

224. Teerawanichpan, P. and X. Qiu, *Fatty acyl-CoA reductase and wax synthase from Euglena gracilis in the biosynthesis of medium-chain wax esters.* Lipids, 2010. **45**(3): p. 263-73.

225. Knight, M.J., I.D. Bull, and P. Curnow, *The yeast enzyme Eht1 is an octanoyl-CoA:ethanol acyltransferase that also functions as a thioesterase.* Yeast, 2014. **31**(12): p. 463-74.

226. King, A., et al., *Cuticular wax biosynthesis in petunia petals: cloning and characterization of an alcohol-acyltransferase that synthesizes wax-esters.* Planta, 2007. **226**(2): p. 381-94.

227. Verstrepen, K.J., et al., *Expression levels of the yeast alcohol acetyltransferase genes ATF1, Lg-ATF1, and ATF2 control the formation of a broad range of volatile esters.* Appl Environ Microbiol, 2003. **69**(9): p. 5228-37.

228. Fujii, T., et al., *Molecular cloning, sequence analysis, and expression of the yeast alcohol acetyltransferase gene.* Appl Environ Microbiol, 1994. **60**(8): p. 2786-92.

229. Tacke, E., et al., *Transposon tagging of the maize Glossy2 locus with the transposable element En/Spm.* Plant J, 1995. **8**(6): p. 907-17.

230. Aggelis, A., et al., *Characterization of two cDNA clones for mRNAs expressed during ripening of melon (Cucumis melo L.) fruits.* Plant Mol Biol, 1997. **33**(2): p. 313-22.

231. Lenz, R. and M.H. Zenk, *Acetyl coenzyme A:salutaridinol-7-O-acetyltransferase from papaver somniferum plant cell cultures. The enzyme catalyzing the formation of thebaine in morphine biosynthesis.* J Biol Chem, 1995. **270**(52): p. 31091-6.

232. Ma, X., et al., *Crystal structure of vinorine synthase, the first representative of the BAHD superfamily.* J Biol Chem, 2005. **280**(14): p. 13576-83.

233. Matsuba, Y., et al., *Evolution of a complex locus for terpene biosynthesis in solanum.* Plant Cell, 2013. **25**(6): p. 2022-36.

234. Negruk, V., et al., *Molecular cloning and characterization of the CER2 gene of Arabidopsis thaliana.* Plant J, 1996. **9**(2): p. 137-45.

235. Okada, T., et al., *Molecular characterization of a novel quinolizidine alkaloid O-tigloyltransferase: cDNA cloning, catalytic activity of recombinant protein and expression analysis in*

Lupinus plants. Plant Cell Physiol, 2005. **46**(1): p. 233-44.

236. Power, R., W.G. Kurz, and V. De Luca, *Purification and characterization of acetylcoenzyme A: deacetylvindoline 4-O-acetyltransferase from Catharanthus roseus.* Arch Biochem Biophys, 1990. **279**(2): p. 370-6.

237. Chang, J., J. Luo, and G. He, *Regulation of polyphenols accumulation by combined overexpression/silencing key enzymes of phenylpropanoid pathway.* Acta Biochim Biophys Sin (Shanghai), 2009. **41**(2): p. 123-30.

238. Stewart, C., Jr., et al., *The Pun1 gene for pungency in pepper encodes a putative acyltransferase.* Plant J, 2005. **42**(5): p. 675-88.

239. D'Auria, J.C., F. Chen, and E. Pichersky, *Characterization of an acyltransferase capable of synthesizing benzylbenzoate and other volatile esters in flowers and damaged leaves of Clarkia breweri.* Plant Physiol, 2002. **130**(1): p. 466-76.

240. Suzuki, H., et al., *Identification and characterization of a novel anthocyanin malonyltransferase from scarlet sage (Salvia splendens) flowers: an enzyme that is phylogenetically separated from other anthocyanin acyltransferases.* Plant J, 2004. **38**(6): p. 994-1003.

241. Walker, K. and R. Croteau, *Taxol biosynthesis: molecular cloning of a benzoyl-CoA:taxane 2alpha-O-benzoyltransferase cDNA from taxus and functional expression in Escherichia coli.* Proc Natl Acad Sci U S A, 2000. **97**(25): p. 13591-6.

242. D'Auria, J.C., *Acyltransferases in plants: a good time to be BAHD.* Curr Opin Plant Biol, 2006. **9**(3): p. 331-40.

243. Eudes, A., et al., *Exploiting the Substrate Promiscuity of Hydroxycinnamoyl-CoA:Shikimate Hydroxycinnamoyl Transferase to Reduce Lignin.* Plant Cell Physiol, 2016. **57**(3): p. 568-79.

244. Goulet, C., et al., *Divergence in the Enzymatic Activities of a Tomato and Solanum pennellii Alcohol Acyltransferase Impacts Fruit Volatile Ester Composition.* Mol Plant, 2014.

245. Hoffmann, L., et al., *Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism.* J Biol Chem, 2003. **278**(1): p. 95-103.

246. Ji, X.-h., et al., *Differences of aroma development and metabolic pathway gene expression between Kyoho and 87-1 grapes.* Journal of Integrative Agriculture, 2021. **20**(6): p. 1525-1539.

247. Wang, J. and V. De Luca, *The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including 'foxy' methylanthranilate.* Plant J, 2005. **44**(4): p. 606-19.

248. Kim, S.J., et al., *Allyl/propenyl phenol synthases from the creosote bush and engineering production of specialty/commodity chemicals, eugenol/isoeugenol, in Escherichia coli.* Arch Biochem Biophys, 2014. **541**: p. 37-46.

249. Yang, Q., et al., *Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, Dianthus caryophyllus L.* Plant Mol Biol, 1997. **35**(6): p. 777-89.

250. Yang, Q., et al., *Analysis of the involvement of hydroxyanthranilate hydroxycinnamoyltransferase and caffeoyl-CoA 3-O-methyltransferase in phytoalexin biosynthesis in oat.* Mol Plant Microbe Interact, 2004. **17**(1): p. 81-9.

251. Vega-Paulino, R.J. and M.E. Zúniga-Hansen, *Potential application of commercial enzyme preparations for industrial production of short-chain fructooligosaccharides.* Journal of Molecular Catalysis B: Enzymatic, 2012. **76**: p. 44-51.

252. Laflamme, P., B. St-Pierre, and V. De Luca, *Molecular and biochemical*

analysis of a Madagascar periwinkle root-specific minovincinine-19-hydroxy-O-acetyltransferase. Plant Physiol, 2001. **125**(1): p. 189-98.

253. Liu, C., et al., *Genome-wide comparative analysis of the BAHD superfamily in seven Rosaceae species and expression analysis in pear (Pyrus bretschneideri).* BMC Plant Biol, 2020. **20**(1): p. 14.

254. Wang, L., et al., *Catalytic function, mechanism, and application of plant acyltransferases.* Crit Rev Biotechnol, 2021: p. 1-20.

255. Walker, K., R. Long, and R. Croteau, *The final acylation step in taxol biosynthesis: cloning of the taxoid C13-side-chain N-benzoyltransferase from Taxus.* Proc Natl Acad Sci U S A, 2002. **99**(14): p. 9166-71.

256. Gunther, C.S., et al., *Characterisation of two alcohol acyltransferases from kiwifruit (Actinidia spp.) reveals distinct substrate preferences.* Phytochemistry, 2011. **72**(8): p. 700-10.

257. Dhar, N., et al., *Characterization of a sweet basil acyltransferase involved in eugenol biosynthesis.* J Exp Bot, 2020. **71**(12): p. 3638-3652.

258. Padgett, K.A. and J.A. Sorge, *Creating seamless junctions independent of restriction sites in PCR cloning.* Gene, 1996. **168**(1): p. 31-35.

259. Rosowski, S., et al., *A novel one-step approach for the construction of yeast surface display Fab antibody libraries.* Microb Cell Fact, 2018. **17**(1): p. 3.

260. Engler, C., R. Kandzia, and S. Marillonnet, *A one pot, one step, precision cloning method with high throughput capability.* PLoS One, 2008. **3**(11): p. e3647.

261. Engler, C. and S. Marillonnet, *Combinatorial DNA assembly using Golden Gate cloning.* Methods Mol Biol, 2013. **1073**: p. 141-56.

262. Tong, Y., et al., *A Golden-Gate Based Cloning Toolkit to Build Violacein Pathway Libraries in Yarrowia lipolytica.* ACS Synth Biol, 2021. **10**(1): p. 115-124.

263. Kim, J., et al., *Properties of alternative microbial hosts used in synthetic biology: towards the design of a modular chassis.* Essays Biochem, 2016. **60**(4): p. 303-313.

264. Nawab, S., et al., *Genetic engineering of non-native hosts for 1-butanol production and its challenges: a review.* Microb Cell Fact, 2020. **19**(1): p. 79.

265. Lebeau, J., J.P. Efromson, and M.D. Lynch, *A Review of the Biotechnological Production of Methacrylic Acid.* Front Bioeng Biotechnol, 2020. **8**: p. 207.

266. Shott, I., et al., *Engineering biology: A priority for growth*. 2019, Royal Academy of Engineering. p. 20.

267. Radhika, D. and A.G. Murugesan, *Bioproduction, statistical optimization and characterization of microbial plastic (poly 3-hydroxy butyrate) employing various hydrolysates of water hyacinth (Eichhornia crassipes) as sole carbon source.* Bioresour Technol, 2012. **121**: p. 83-92.

268. Lee, J.-Y., S.-E. Lee, and D.-W. Lee, *Current status and future prospects of biological routes to bio-based products using raw materials, wastes, and residues as renewable resources.* Critical Reviews in Environmental Science and Technology, 2021. **52**(14): p. 2453-2509.

269. Tao, Y., et al., *Strategies to improve the stress resistance of Escherichia coli in industrial biotechnology.* Biofuels, Bioproducts and Biorefining, 2022. **16**(4): p. 1130-1141.

270. Guo, L., et al., *Engineering Escherichia coli lifespan for enhancing chemical production.* Nature Catalysis, 2020. **3**(3): p. 307-318.

Supplementary A: Sequence information

A.1: pGGV4

All sequences are given in FASTA format, written 5'-3'.

>pGGV4

TCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTT
CTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAG
CGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCA
GAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACT
CTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCG
ATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGG
GCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGAT
ACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATC
CGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGT
ATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGT
CAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT
GCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTA
CCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGA
GCGAGGAAGCGGAAGAGGGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCAC
ACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATAC
ACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGAC
GCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCTGCATGTGT
CAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGG
TCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGA
AGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTC
ACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAG
AGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGT
AAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGC
TTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGG
AACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAG
ACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCG
CGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAAC
GACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGTGATCTCGATCCCG
CGAAATGCGCGGCCGCTTGACAGCTAGCTCAGTCCTAGGTATTGTGCTAGCACTAGTGAGACC
ACAACGGTTTCCCTCTAGAAATAATTTTGTGCTAGCGTTTGGCCCGGACCGGGCCGGAGGCCT
GTCATGAGTGATTACGAGCCGTTGCGTCTGCATGTCCCGGAGCCCACCGGGCGTCCTGGCTGC
AAGACCGACTTTTCCTATCTGCACCTGTCCCCCGCCGGCGAGGTACGCAAGCCGCCGGTGGAT
GTCGAGCCCGCCGAGACCAGCGACCTGGCCTACAGCCTGGTACGTGTGCTCGACGACGACGGC
CACGCCGTCGGTCCCTGGAATCCGCAGCTCAGCAACGAACAACTGCTGCGCGGCATGCGGGCG
ATGCTCAAGACCCGCCTGTTCGACGCGCGCATGCTCACCGCGCAACGGCAGAAAAAGCTTTCC
TTCTATATGCAATGCCTCGGCGAGGAAGCCATCGCCACCGCCCACACCCTGGCCCTGCGCGAC
GGCGACATGTGCTTTCCGACCTATCGCCAGCAAGGCATCCTGATCACCCGCGAATACCCGCTG
GTGGACATGATCTGCCAGCTTCTCTCCAACGAGGCCGACCCGCTCAAGGGCCGCCAGCTGCCG
ATCATGTACTCGAGCAAGGAGGCAGGTTTCTTCTCCATCTCCGGCAACCTCGCCACCCAGTTC
ATCCAGGCGGTCGGCTGGGGCATGGCCTCGGCGATCAAGGGCGACACGCGCATCGCCTCGGCC
TGGATCGGCGACGGCGCCACCGCCGAGTCGGACTTCCACACCGCCCTCACCTTCGCCCATGTC
TACCGCGCGCCGGTAATCCTCAACGTGGTCAACAACCAGTGGGCGATCTCCACCTTCCAGGCC
ATCGCCGGCGGCGAAGGCACCACCTTCGCCAACCGTGGCGTGGGCTGCGGGATCGCCTCGCTG
CGGGTCGACGGCAATGACTTCCTGGCGGTCTACGCCGCCTCCGAGTGGGCCGCCGAGCGCGCC
CGGCGCAACCTCGGGCCGAGCCTGATCGAATGGGTCACCTACCGCGCCGGCCCGCACTCGACT
TCGGACGACCCGTCCAAGTACCGCCCCGCCGACGACTGGACCAACTTCCCGCTGGGCGACCCG
ATCGCCCGCCTGAAGCGGCACATGATCGGCCTCGGCATCTGGTCGGAGGAACAGCACGAAGCC
ACCCACAAGGCCCTCGAAGCCGAAGTACTGGCCGCGCAGAAACAGGCGGAGAGCCATGGCACC

CTGATCGACGGCCGGGTGCCGAGCGCCGCCAGCATGTTCGAGGACGTCTATGCAGAACTGCCG
GAGCACCTGCGCCGGCAACGCCAGGAGCTCGGGGTATGAATGCCATGAACCCGCAACACGAGA
ACGCCCAGACGGTCACCAGCATGACCATGATCCAGGCGCTGCGCTCGGCGATGGACATCATGC
TCGAGCGCGACGACGACGTGGTGGTATTCGGCCAGGACGTCGGCTACTTCGGCGGCGTGTTCC
GCTGCACCGAAGGCCTGCAGAAGAAATACGGCACCTCGCGGGTGTTCGATGCGCCGATCTCCG
AGAGCGGCATCATCGGCGCCGCGGTCGGCATGGGTGCCTACGGCCTGCGCCCGGTGGTGGAGA
TCCAGTTCGCCGACTACGTCTACCCGGCCTCCGACCAGTTGATCTCCGAGGCGGCGCGCCTGC
GCTATCGCTCGGCCGGCGACTTCATCGTGCCGATGACCGTACGCATGCCCTGTGGCGGCGGCA
TCTACGGCGGGCAAACGCACAGCCAGAGCCCGGAGGCGATGTTCACCCAGGTCTGCGGCCTGC
GCACGGTGATGCCGTCCAACCCCTACGACGCCAAGGGCCTGCTGATCGCCTGCATCGAGAACG
ACGACCCGGTGATCTTCCTCGAGCCCAAGCGCCTCTACAACGGCCCGTTCGATGGCCACCACG
ACCGCCCGGTGACGCCCTGGTCCAAGCATCCGGCCAGCCAGGTGCCGGACGGCTACTACAAGG
TGCCGCTGGACAAGGCGGCGATCGTCCGCCCCGGCGCGGCGCTGACCGTGCTGACCTACGGCA
CCATGGTCTACGTGGCCCAGGCCGCGGCCGACGAGACCGGCCTGGACGCCGAGATCATCGACC
TGCGCAGCCTCTGGCCGCTGGACCTGGAAACCATCGTCGCCTCGGTGAAGAAGACCGGCCGCT
GCGTCATCGCCCACGAGGCGACCCGCACCTGCGGGTTCGGCGCCGAGCTGATGTCGCTGGTGC
AGGAGCACTGCTTCCACCACCTGGAGGCGCCGATCGAGCGCGTCACCGGTTGGGACACCCCCT
ACCCGCATGCCCAGGAGTGGGCGTATTTCCCCGGCCCCGCGCGCGTCGGCGCGGCATTCAAGC
GTGTGATGGAGGTCTGAATGGGTACCCATGTGATCAAGATGCCGGACATCGGGGAAGGCATCG
CCGAGGTCGAACTGGTGGAGTGGCATGTCCAGGTCGGCGACTCGGTCAATGAAGACCAGGTCC
TCGCCGAGGTGATGACCGACAAGGCCACGGTGGAGATTCCCTCGCCGGTGGCCGGACGCATCC
TCGCCCTCGGCGGCCAGCCGGGCCAGGTGATGGCGGTGGGCGGCGAACTGATCCGCCTGGAGG
TGGAAGGCGCCGGCAACCTCGCCGAGAGTCCGGCCGCGGCGACGCCGGCCGCGCCCGTCGCCG
CCACCCCGGAGAAACCGAAGGAAGCCCCGGTCGCGGCGCCGAAAGCCGCCGCCGAAGCGCCGC
GCGCCTTGCGCGACAGCGAGGCGCCACGGCAGCGGCGCCAGCCCGGCGAACGCCCGCTGGCCT
CCCCCGCGGTGCGCCAGCGCGCCCGCGACCTGGGCATCGAGTTGCAGTTCGTGCAGGGCAGCG
GTCCCGCCGGACGCGTCCTCCACGAGGACCTCGATGCCTACCTGACCCAGGATGGCAGCGTCG
CGCGCAGCGGCGGCGCCGCGCAGGGGTATGCCGAGCGACACGACGAACAGGCGGTGCCGGTGA
TCGGCCTGCGTCGCAAGATCGCCCAGAAGATGCAGGACGCCAAGCGACGCATCCCGCATTTCA
GCTATGTCGAGGAAATCGACGTCACCGATCTGGAAGCCCTGCGCGCCCATCTCAACCAGAAAT
GGGGTGGCCAGCGCGGCAAGCTGACCCTGCTGCCGTTCCTGGTCCGCGCCATGGTCGTGGCGC
TGCGCGACTTCCCGCAGTTGAACGCGCGCTACGACGACGAGGCCGAGGTGGTCACCCGCTACG
GCGCGGTGCACGTCGGCATCGCCACCCAGAGCGACAACGGCCTGATGGTGCCGGTGCTGCGCC
ACGCCGAATCGCGCGACCTCTGGGGCAACGCCAGCGAAGTGGCGCGCCTGGCCGAAGCCGCAC
GCAGCGGCAAGGCGCAACGCCAGGAGCTGTCCGGCTCGACCATCACCCTGAGCAGCCTCGGCG
TGCTCGGCGGGATCGTCAGCACACCGGTGATCAACCATCCGGAGGTGGCCATCGTCGGCGTCA
ACCGCATCGTCGAGCGACCGATGGTGGTCGGCGGCAACATCGTCGTGCGCAAGATGATGAACC
TCTCCTCCTCCTTCGACCACCGGGTGGTCGACGGGATGGACGCGGCGGCCTTCATCCAGGCCG
TGCGCGGCCTGCTCGAACATCCCGCCACCCTGTTCCTGGAGTAAGCGATGAGCCAGATCCTGA
AGACTTCCCTGCTGATCGTCGGCGGCGGTCCCGGCGGCTACGTCGCGGCGATCCGTGCCGGGC
AACTGGGCATTCCCACCGTACTGGTGGAGGGCGCCGCCCTCGGCGGCACCTGCCTGAACGTCG
GCTGCATCCCGTCGAAGGCGCTGATCCACGCCGCCGAGGAATACCTCAAGGCCCGCCACTATG
CCAGCCGGTCGGCGCTGGGCATCCAGGTACAGGCGCCGAGCATCGACATCGCCCGCACCGTGG
AATGGAAGGACGCCATCGTCGACCGCCTCACCAGCGGCGTCGCCGCGCTGCTGAAGAAACACG
GGGTCGATGTCGTCCAGGGCTGGGCGAGGATCCTCGACGGCAAAAGCGTGGCGGTCGAACTCG
CCGGCGGCGGCAGCCAGCGCATCGAGTGCGAGCACCTGCTGCTGGCCGCCGGCTCGCAGAGCG
TCGAGCTACCGATCCTGCCGCTGGGCGGCAAGGTGATCTCCTCCACCGAGGCGCTGGCGCCCG
GCAGCCTGCCCAAGCGCCTGGTGGTGGTCGGCGGCGGCTACATCGGCCTGGAGCTGGGTACCG
CCTACCGCAAGCTCGGCGTCGAGGTGGCGGTGGTGGAAGCGCAACCACGCATCCTGCCGGGCT
ACGACGAAGAACTGACCAAGCCGGTGGCCCAGGCCTTGCGCAGGCTGGGCGTCGAGCTGTACC
TCGGGCACAGCCTGCTGGGCCCGAGCGAGAACGGCGTGCGGGTCCGCGACGGCGCCGGCGAGG
AGCGCGAGATCGCCGCCGACCAGGTACTGGTGGCGGTCGGCCGCAAGCCGCGCAGCGAAGGCT
GGAACCTGGAAAGCCTGGGCCTGGACATGAACGGCCGGGCGGTGAAGGTCGACGACCAGTGCC
GCACCTCGATGCGCAATGTCTGGGCCATAGGCGATCTCGCCGGCGAGCCGATGCTCGCGCACC
GGGCCATGGCCCAGGGCGAGATGGTCGCCGAGCTGATCGCCGGCAAGCGTCGCCAGTTCGCCC
CGGTGGCGATCCCCGCGGTGTGCTTCACCGATCCGGAAGTGGTGGTCGCCGGGTTGTCCCCGG
AGCAGGCGAAGGATGCCGGCCTGGACTGCCTGGTGGCGAGCTTCCCGTTCGCCGCCAACGGTC
GCGCCATGACCCTGGAGGCCAACGAAGGCTTCGTCCGCGTGGTGGCGCGTCGCGACAACCACC
TGGTCGTCGGCTGGCAGGCGGTGGGCAAGGCGGTCTCGGAACTGTCCACGGCCTTCGCCCAGT

CGCTGGAGATGGGCGCCCGCCTGGAAGACATCGCCGGCACCATCCACGCCCATCCGACCCTCG
GCGAAGCGGTCCAGGAAGCCGCCCTGCGCGCGCTGGGACACGCCCTGCACATCTGATCTAGAG
TCGACCTGCAGGTATTATCTCGAGCGTCAATTGTCTGATTCGTTACCAAAGATCTCTGACAGC
TAGCTCAGTCCTAGGTATAATGCTAGCACCCGTTTTTTTGGGAATTCGAGACGCACATATACC
TGCCGTTCACTATTATTTAGTGAAATGAGATATTATGATATTTTCTGAATTGTGATTAAAAAG
GCAACTTTATGCCCATGCAACAGAAACTATAAAAAATACAGAGAATGAAAAGAAACAGATAGA
TTTTTTAGTTCTTTAGGCCCGTAGTCTGCAAATCCTTTTATGATTTTCTATCAAACAAAAGAG
GAAAATAGACCAGTTGCAATCCAAACGAGAGTCTAATAGAATGAGGTCGAAAAGTAAATCGCG
CGGGTTTGTTACTGATAAAGCAGGCAAGACCTAAAATGTGTAAAGGGCAAAGTGTATACTTTG
GCGTCACCCCTTACATATTTTAGGTCTTTTTTTATTGTGCGTAACTAACTTGCCATCTTCAAA
CAGGAGGGCTGGAAGAAGCAGACCGCTAACACAGTACATAAAAAAGGAGACATGAACGATGAA
CATCAAAAAGTTTGCAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTGGCAGGAGG
CGCAACTCAAGCGTTTGCGAAAGAAACGAACCAAAAGCCATATAAGGAAACATACGGCATTTC
CCATATTACACGCCATGATATGCTGCAAATCCCTGAACAGCAAAAAAATGAAAAATATCAAGT
TCCTGAGTTCGATTCGTCCACAATTAAAAATATCTCTTCTGCAAAAGGCCTGGACGTTTGGGA
CAGCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAACTATCACGGCTACCACATCGTCTT
TGCATTAGCCGGAGATCCTAAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAGT
CGGCGAAACTTCTATTGACAGCTGGAAAAACGCTGGCCGCGTCTTTAAAGACAGCGACAAATT
CGATGCAAATGATTCTATCCTAAAAGACCAAACACAAGAATGGTCAGGTTCAGCCACATTTAC
ATCTGACGGAAAAATCCGTTTATTCTACACTGATTTCTCCGGTAAACATTACGGCAAACAAAC
ACTGACAACTGCACAAGTTAACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGTAGA
GGATTATAAATCAATCTTTGACGGTGACGGAAAAACGTATCAAAATGTACAGCAGTTCATCGA
TGAAGGCAACTACAGCTCAGGCGACAACCATACGCTGAGAGATCCTCACTACGTAGAAGATAA
AGGCCACAAATACTTAGTATTTGAAGCAAACACTGGAACTGAAGATGGCTACCAAGGCGAAGA
ATCTTTATTTAACAAAGCATACTATGGCAAAAGCACATCATTCTTCCGTCAAGAAAGTCAAAA
ACTTCTGCAAAGCGATAAAAAACGCACGGCTGAGTTAGCAAACGGCGCTCTCGGTATGATTGA
GCTAAACGATGATTACACACTGAAAAAAGTGATGAAACCGCTGATTGCATCTAACACAGTAAC
AGATGAAATTGAACGCGCGAACGTCTTTAAAATGAACGGCAAATGGTATCTGTTCACTGACTC
CCGCGGATCAAAAATGACGATTGACGGCATTACGTCTAACGATATTTACATGCTTGGTTATGT
TTCTAATTCTTTAACTGGCCCATACAAGCCGCTGAACAAAACTGGCCTTGTGTTAAAAATGGA
TCTTGATCCTAACGATGTAACCTTTACTTACTCACACTTCGCTGTACCTCAAGCGAAAGGAAA
CAATGTCGTGATTACAAGCTATATGACAAACAGAGGATTCTACGCAGACAAACAATCAACGTT
TGCGCCTAGCTTCCTGCTGAACATCAAAGGCAAGAAAACATCTGTTGTCAAAGACAGCATCCT
TGAACAAGGACAATTAACAGTTAACAAATAAAAACGCAAAAGAAAATGCCGATTATGGTGCAC
TCTCGTCTCCTTAAGATACACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAG
TTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTG
AGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCG
GCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCC
TAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTC
AAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCA
AAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCC
CTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCA
ACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAA
AAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTT
CAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATT
CAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGA
AGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTC
CTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCAC
GAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAG
AACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTG
ACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACT
CACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCA
TAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGC
TAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGC
TGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGT
TGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGA
TGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTG
CTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATG
GTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAA

ATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACGGTAGTGATACCCCA
GAGGATTAGATGGCCAAAGAAGACAATATTGAAATGCAAGGTACCGTTCTTGAAACGTTGCCT
AATACCATGTTCCGCGTAGAGTTAGAAAACGGTCACGTGGTTACTGCACACATCTCCGGTAAA
ATGCGCAAAAACTACATCCGCATCCTGACGGGCGACAAAGTGACTGTTGAACTGACCCCGTAC
GACCTGAGCAAAGGCCGCATTGTCTTCCGTAGTCGCTAACTGTCAGACCAAGTTTACTCATAT
ATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTT
GATAATCTCATGACCAAAA

A.2: ACX sequences

Regions in red are the sequences added to incorporate Esp3I, EcoRI and SpeI cloning sites for Golden Gate assembly.

>SPIOL_ACX1
cgtctcgaattcttaactttaagaaggagatataccATGGAAGATCTGGAAGGCGGCGTGGAT
TATCTGGCGGATGAACGCAAAAAAGCGCAGTTTGATGTGAACGCGATGAAAATTGTGTGGGCG
GGCAGCCAGCATAAACTGCAGCTGAGCGATCGCATGGCGCGCCTGGTGGCGAGCGATCCGGAA
TTTAAAAAAGATGATCGCGTGCGCCTGGATCGCAAAGAACTGTTTAAAAACACCCTGCGCAAA
GCGAGCCATGCGTGGAAACGCATTATTGAACTGCGCCTGACCGAAGAAGAAGCGAGCCAGCTG
CGCTTTTTTGTGGATCAGCCGGCGTATACCGATCTGCATTGGGGCATGTTTGTGCCGGCGATT
AAAGGCCAGGGCACCGAAGAACAGCAGGCGAAATGGCTGCCGCTGGCGTATCGCATGCAGATT
ATTGGCTGCTATGCGCAGACCGAACTGGGCCATGGCAGCAACGTGCAGGGCCTGGAAACCACC
GCGACCTTTGATCCGGAAACCGATGAATTTGTGATTCATAGCCCGACCCTGACCAGCAGCAAA
TGGTGGCCGGGCGGCCTGGGCAAAATTAGCACCCATGCGATTGTGTATGCGCGCCTGATTACC
GATGGCAAAGAACATGGCGTGCATGGCTTTATTGTGCAGCTGCGCAGCCTGGATGATCATCTG
GCGCTGCCGGGCATTACCGTGGGCGATATTGGCATGAAATTTGGCAACGCGGCGTATAACACC
ATGGATAACGGCGTGCTGCGCTTTGATCATGTGCGCATTCCGCGCGATCAGATGCTGATGCGC
GTGCTGCAGGTGACCCGCGAAGGCAACGTGGTGAAAAGCAACATTCCGCGCCAGCTGATTTAT
GGCACCATGGTGTATGTGCGCCAGACCATTGTGTATGATGCGAGCCGCGCGCTGAGCAAAGCG
GTGTGCATTGCGACCCGCTATAGCGCGGTGCGCCGCCAGTTTGGCAGCGAAAACGGCCGCGAA
ACCCAGGTGATTGATTATAAAACCCAGCAGAACCGCCTGTTTCCGCTGCTGGCGAGCGCGTAT
GCGTTTCGCATTGTGGGCGATTGGCTGAGCTGGCTGTATACCGATGTGACCAAACGCCTcCAG
GCGAACGATTTTAGCACCCTGCCGGAAGCGCATGCGTGCACCGCGGGCCTGAAAAGCCTGACC
ACCAGCGTGACCGCGGATGCGATTGAAGAATGCCGCAAACTGTGCGGCGGCCATGGCTATCTG
TGCAGCAGCGGCCTGCCGGAACTGTTTGCGGTGTATGTGCCGGCGTGCACCTATGAAGGCGAT
AACGTGGTGCTGCTGCTGCAGGTGGCGCGCATTCTGATGAAAACCGTGAGCCAGCTGGGCAGC
GGCAAACAGCCGGTGGGCACCATGGCGTATATGGGCCGCGCGGAACATCTGCTGCAGAGCCAG
TGCAGCGTGCAGAAAGCGGAAGATTGGCTGAACCCGGATATTGTGCTGGAAGCGTTTGAAGCG
CGCAGCGTGCGCATGGCGGTGGCGTGCGCGAAAAACCTGAGCGAATTTAGCAACCCGGAAGAA
GGCTTTCTGGAACTGAGCGCGGATCTGGTGGGCGCGAGCACCGCGCATTGCCAGCTGATTGTG
GTGAGCAAATTTATTGAAAAACTGCAGGAAGATATTCCGGGCGAAGGCGTGAAAGAACAGCTG
CAGGCGCTGTGCGGCATTTATGCGCTGAGCATTATGGGCAAACATCTGGGCGATTTTCTGAGC
ACCGGCTGCATTAAACCGAAACAGGCGAGCCTGGCGAACGATCTGCTGCGCAGCCTGTTTAGC
AAAGTGCGCCCGAACGCGATTGCGCTGGTGGATGCGTTTAACTATACCGATCATTTTCTGGGC
AGCGTGCTGGGCCGCTATGATGGCAACGTGTATCCGAAACTGTATGAAGAAGCGTGGAAAGAT
CCGTTTAACGATACCGCGGTGCCGGATGGCTATCATGAATATGTGCGCCCGCTGCTGAAACGC
CAGGTGCGCACCAGCCGCCTGtaactcgagtaaactagtgagacgcctgcagg

>SPIOL_ACXb
cgtctcgaattcttaactttaagaaggagatataccATGGCGAGCCGCGAAGAAATTATTATT
TTTGATACCAAAAAACTGCGCGCGCTGCTGGATATGCATAACATTGAAGATCGCGATAGCCTG
TATAACCTGATTATTCAGAGCAGCGTGTTTAACCCGCGCAAAATTGGCAGCCGCGTGTTTGTG
AGCCCGGATTATAACGAACCGATGGAACTGCAGCGCGAAGTGACCATGAAACGCCTGCGCTAT
ATGGTGGATCATGGCGCGTTTAAAGGCTGGATTACCGAAGGCGGCGCGGCGGTGGAACTGCGC
CGCATGGCGCTGAACGAAGTGGTGCAGATGTATGATCATAGCCTGGCGGTGATGCTGGGCGCG
CATTATTTTCTGTGGGGCGGCGCGATTCAGTTTTTTGGCACCAAACGCCATCATGATAAATGG

CTGAAAATTACCGAAGATTATGGCATTAAAGGCTGCTTTGCGATGACCGAACTGGGCCATGGC
AGCAACGTGCGCGGCATTGAAACCATTACCACCTATGATCCGAACGCGGGCGAATTTATTATT
AACACCCCGTGCGAAAGCGCGCAGAAATATTGGATTGGCGGCGCGGCGAAACATGCGACCCAT
ACCGTGGTGTTTAGCCAGCTGCATATTAACGGCAAAGATGAAGGCGTGCATGCGTTTATTGTG
CAGATTCGCGATGAAGAAGGCAACGTGAGCCCGAACATTCGCATTGCGGATTGCGGCCATAAA
ATTGGCCTGAACGGCGTGGATAACGGCCGCATTTGGTTTGATAACGTGCGCATTCCGCGCGAA
AACCTGCTGAACAGCGTGGCGGATGTGAGCCCGGATGGCCAGTATCTGAGCGCGATTAAAGAT
CCGGTGCAGCGCTTTGCGGCGTTTCTGGCGCCGCTGACCAGCGGCCGCGTGAACATTGCGGTG
AGCAGCATTTATATTAGCAAAGTGGGCCTGGCGACCGCGCTGCGCTATAGCCTGAGCCGCCGC
GCGTTTAGCCTGGCGGCGAACGAACCGGAAGTGCTGCTGCTGGATTATCCGAGCCATCAGCGC
CGCCTGCTGCCGCTGCTGGCGAAAACCTATGCGATGAGCTTTGCGGGCAACTATCTGAAAAGC
ATTTATGTGAAACGCAGCCCGGAAACCAACAAAAACATTCATGTGGTGAGCAGCGCGCTGAAA
GCGACCCTGACCTGGCATAACATGCGCACCCTcCAGGAATGCCGCGAAGCGTGCGGCGGCCAG
GGCCTGAAAACCGAAAACCGCGTGGGCCATCTGAAAGGCGAATATGATGTGCAGAGCACCTTT
GAAGGCGATAACAACGTGCTGATGCAGCAGGTGAGCAAAGCGCTGTTTGGCGAATATATTGCG
ACCCAGAAACAGAACCGCCCGTTTAAAGGCCTGTGGCTGGAACATATGAACGGCCCGAGCCCG
GTGATTCCGGCGCAGCTGACCAGCAGCATTCTGCGCAGCAGCCAGATTCAGACCGATATTTTT
TATCTGCGCGAACGCGATCTGCTGAACCGCTTTGTGGCGGAAGTGAGCGCGCATCAGAAACAG
GGCCGCAACAAAGAATATGCGTTTAGCCTGAGCTATCAGCTGGCGGAgGATCTGGGCCGCGCG
TTTGCGGATAAAGCGATTCTGCTGACCAGCATTGAAGCGGAAGCGAACGTGCCGGCGGGCCCG
CTGAAAGATGTGCTGGCGCTGCTGCGCACCCTGTATGCGCTGGTGATTCTGGAAGAAGATGCG
AGCTTTCTGCGCTATGGCTATCTGAGCGTGGAAAACAGCGCGGTGGTGCGCCAGGAAGTGATT
AAACTGTGCAGCGCGAGCAGCCATCCGTGCATTGGCCAGTTTTTTCGCCATACCCGCTGCGTG
CCGGAAAGCTATTGCGTGtaactcgagtaaactagtgagacgcctgcagg

>SPIOL_ACX4
cgtctcgaattcttaactttaagaaggagatataccATGACCATTCAGACCAGCCCGAAAGAT
GATAGCCTGAAAGATGCGCGCAGCAGCTATTTTGATCTGCCGCCGCTGGATGTGAGCGTGGCG
TTTCCGCAGGCGCAGCCGGCGAGCGTGTTTCCGCCGTGCGCGAGCGATTATTATCAGTTTGAT
GATCTGCTGAACGCGGAAGAACGCGCGATTCGCGCGAAAGTGCGCGAATGCATGGAAAAAGAA
GTGGCGCCGATTATGGCGGAATATTGGGAAAAAGCGGAATTTCCGTTTCAGATTATTCCGAAA
CTGGGCGCGCTGCATATTAGCGGCGGCACCATTAAAGGCTATGGCTGCCCGGGCCTGAGCGTG
ACCGGCAGCGCGATTGCGACCGCGGAAGTGGCGCGCGTGGATGCGAGCTGCAGCACCTTTATT
CTGGTGCATAGCAGCCTGGCGATGCTGACCATTGCGCTGTGCGGCAGCGAAGATCAGAAACAG
AAATATCTGCCGAGCCTGGCGCAGTTTAAAACCGTGGCGTGCTGGGCGCTGACCGAACCGGAT
AACGGCAGCGATGCGAGCGGCCTGCAGACCACCGCGACCAAAGTGGCGGGCGGCTGGATTCTG
GAAGGCCAGAAACGCTGGATTGGCAACAGCACCTTTGCGGATGTGCTGATTATTTTTGCGCGC
AACACCGCGACCAACCAGATTAACGGCTATATTGTGAAAAAAGATGCGCCGGGCCTGACCGCG
ACCAAAATTGCGAACAAAATTGGCCTGCGCATTGTGCAGAACGGCGATATTATTCTGAAACGC
GTGTTTGTGCCGGATGAAGATCGCCTGCCGGGCGTGAACAGCTTTCAGGATACCAACAAAGTG
CTGGCGGTGAGCCGCGTGATGGTGGCGTGGCAGCCGATTGGCATTAGCATGGGCGTGTATGAT
ATGTGCCTGCGCTATCTGAAAGAACGCAAACAGTTTGGCGCGCCGCTGGCGGCGTTTCAGATT
AACCAGCAGAAACTGGTGCAGATGCTGGGCAACGTGCAGGCGATGTTTCTGGTGGGCTGGCGC
CTGTGCAAACTGTATGAAAGCGGCAAAATGACCCCGGGCCATGCGAGCATGGGCAAAAGCTGG
ATTACCATTAAAGCGCGCGAAACCGTGGCGCTGGGCCGCGAACTGCTGGGCGGCAACGGCATT
CTGGCGGAATTTCTGGTGGCGAAAGCGTTTTGCGATCTGGAACCGATTTATACCTATGAAGGC
ACCTATGATATTAACACCCTGGTGACCGCGCGCGAAGTGACCGGCATTGCGAGCTTTAAACCG
GCGGCGCGCAGCCGCCTGtaactcgagtaaactagtgagacgcctgcagg

>VIGRR_ACX1
cgtctcgaattcttaactttaagaaggagatataccATGGAAGGCGTGGATCATCTGGCGTTT
GAACGCAACAAAGCGCAGTTTGATGTGGATGAAATGAAAATTGTGTGGGCGGGCAGCCGCGAA
GATTTTGAACTGAGCGATCGCATTAGCCGCCTGGTGGCGAGCGATCCGGTGTTTCGCAAAGAT
GATCGCGCGGCGCTGGGCCGCAAAGAACTGTTTAAAAACACCCTGCGCAAAGCGGCGCATGCG
TGGAAACGCGTGATTGAACTGCGCCTGAGCGAACCGGAAGCGTATAAACTGCGCGCGTTTGTG
GATCAGCCGGCGTTTACCGATCTGCATTGGGGCATGTTTGTGCCGGCGATTCAGGGCCAGGGC
ACCGATGAACAGCAGCAGAAATGGCTGCCGCTGGCGTATCGCATGCAGATTATTGGCTGCTAT
GCGCAGACCGAACTGGGCCATGGCAGCAACGTGCAGGGCCTGGAAACCACCGCGACCTTTGAT
CCGAAAACCGATGAATTTGTGATTCATAGCCCGACCCTGACCAGCAGCAAATGGTGGCCGGGC

GGCCTGGGCAAAGTGAGCACCCATGCGGTGGTGTATGCGCGCCTGATTATTGATGGCCAGGAT
TATGGCGTGCATGGCTTTATTGTGCAGCTGCGCAGCCTGGATGATCATCTGCCGCTGCCGGGC
ATTACCGTGGGCGATATTGGCATGAAATTTGGCAACGCGGCGTATAACACCATGGATAACGGC
GTGCTGACCTTTGATCATGTGCGCATTCCGCGCAACCAGATGCTGATGCGCGTGAGCCAGGTG
ACCCGCGAAGGCAAATATGTGCATAGCAACGTGCCGCGCCAGCTGGTGTATGGCACCATGGTG
AACGTGCGCCAGAAAATTGTGGCGGATGCGAGCATTGCGCTGAGCCGCGCGGTGTGCATTGCG
ACCCGCTATAGCGCGGTGCGCCGCCAGTTTGGCAGCCATAACGGCGGCCTGGAAACCCAGGTG
ATTGATTATAAAACCCAGCAGAGCCGCCTGTTTCCGCTGCTGGCGAGCGCGTATGCGTTTCGC
TTTGTGGGCGAATGGCTGCTGTGGCTGTATACCGATGTGACCGAACGCCTGAAAGCGAACGAT
TTTAGCACCCTGCCGGAAGCGCATGCGTGCACCGCGGGCCTGAAAAGCCTGACCACCACCGCG
ACCGCGGATGGCATTGAAGAATGCCGCAAACTGTGCGGCGGCCATGGCTATCTGTGCAGCAGC
GGCCTGCCGGAACTGTTTGCGGTGTATGTGCCGGCGTGCACCTATGAAGGCGATAACATTGTG
CTGCTGCTGCAGGTGGCGCGCCATCTGATGAAAACCGTGAGCCAGCTGAGCACCGGCAACAAA
CCGGTGGGCACCACCGCGTATATGGCGCGCCTGGAACAGCTGCTGCAGTATCGCAGCGATGTG
AAAAAAGCGGAAGATTGGCTGAAACCGAACGTGGTGCTGGAAGCGTTTGAAGCGCGCGCGGCG
CGCATGAGCGTGGCGTGCGCGCAGAACCTGAGCAAATTTAGCAACCCGGAAGAAGGCTTTCAG
GAACTGGCGGCGGATCTGGTGGAAGCGGCGGTGGCGCATTGCCAGCTGATTGTGGTGAGCAAA
TTTATTGATAAACTGCGCCAGGATATTCCGGGCAAAGGCGTGAAACATATTCTGGAAGTGCTG
TGCAGCGTGTATGCGCTGAGCCTGCTGCATAAACATCTGGGCGATTTTATTAGCACCGGCTGC
ATTGATCCGAAACAGGGCAGCCTGGCGAACGAACAGCTGCGCAACCTGTATAGCCAGGTGCGC
CCGAACGCGGTGGCGCTGGTGGATGCGTATAACTATACCGATCATTATCTGGGCAGCATTCTG
GGCCGCTATGATGGCAACGTGTATCCGAAACTGTATGAAGAAGCGTGGAAAGATCCGCTGAAC
GATAGCATTGTGCCGGATGGCTTTAAAGAATATATTCAGCCGCTGCTGAAACAGCAGCTGCGC
AACGCGCGCCTGtaactcgagtaaactagtgagacgcctgcagg

>VIGRR_ACX3
cgtctcgaattcttaactttaagaaggagatataccATGGATCATCAGGTGAGCCGCCGCACC
GAAATTCTGACCAACCATCTGCTGCTGCGCGCGCCGCCGCCGAGCAGCGTGCTGCAGCCGCAT
CGCTGCCTGAGCTATAGCCCGCCGGAACTGAGCAACGAATTTGCGTTTGATCTGCGCGAAATG
CGCCGCCTGATGGATGGCCATAACCTGGAAGATCGCGATTGGCTGTTTAGCGTGATTGTGCAG
AGCGCGCTGTTTAACCGCCGCGAACGCGCGGGCCGCATTTTTGTGTGCCCGGATTATAACCAG
AGCATGGAACAGCAGCGCGAAGCGACCATGCGCCGCATTGAATATCTGGTGGAACGCGGCGTG
TTTCGCGGCTGGCTGACCGGCGAAGGCCCGGCGGAAGAACTGCGCAAACTGGCGCTGCATGAA
GTGATGGGCATGTATGATCATAGCCTGAGCGTGAAACTGGGCGTGCATTATTTTCTGTGGGGC
GGCGCGGTGAAATTTCTGGGCACCAAACATCATCATGATAAATGGCTGCGCGCGACCGAAAAC
TATGATATGAAAGGCTGCTTTGCGCTGAGCGAACTGGGCCATGGCAGCAACGTGCGCGGCATT
GAAACCATTACCACCTATGATAGCAACACCGGCGAATTTGTGGTGAACACCCCGTGCGAAAGC
GGCCAGAAATATTGGATTGGCGGCGCGGCGAACCATGCGACCCATACCATTGTGTTTAGCCAG
CTGTATATTAACGGCAGCAACCAGGGCGTGCATGCGTTTATTGCGCAGATTCGCGATAGCGAT
GGCAACATTTGCCCGAACATTCGCATTGCGGATTGCGGCCATAAAATTGGCGTGAACGGCGTG
GATAACGGCCGCATTTGGTTTGATAACGTGCGCATTCCGCGCGAAAACCTGCTGAACAGCGTG
GCGGATGTGAGCCCGACCGGCGAATATCTGAGCGCGATTAAAAACGTGGATCAGCGCTTTGCG
GCGTTTCTGGCGCCGCTGACCAGCGGCCGCTTTAGCGGCCATTTTCCGATTTGGATGTTTATT
AGCCTGCTGCTGTGGTTTCGCTATGCGATGAGCTTTAGCGCGAACGATCTGAAAATTATGTAT
GTGAAACGCACCCCGAAAAGCAACAAAGCGATTCATATTATTAGCAGCGCGTATAAAGCGACC
TTTACCTGGAACAACATTCGCACCCTGCTGGAATGCCGCGATGCGTGCGGCGGCCAGGGCGTG
AAAAGCGAAAACCGCGTGGGCCTGTTTATGGGCGAATTTGAAATGCATAGCACCTTTGAAGGC
GATAACAAAGTGCTGATGCAGCAGATTAGCAAAGCGCTGTTTGCGGAATATGTGGCGTGCCAT
AAACAGAACAAACCGTTTAGCGGCCTGGGCCTGGAACATATGAACAAACCGCTGCTGGTGATT
CCGAGCCATCTGACCACCCCGACCATTCGCAGCAGCGAATTTCAGATTGATCTGTTTCATCTG
CGCGAACGCGATCTGCTGCGCCGCTTTGTGGAAGAAGTGAGCGAATATCAGAGCCGCGGCGAA
AGCAAAGAAAGCGCGTTTATTCTGAGCTATCAGCTGGCGGAgGATCTGAGCCGCGCGTTTAGC
GAACGCGCGATTCTGAAAACCTTTATGGATGCGGAAAGCACCCTGGCGGCGGGCAGCCTGAAA
AACGTGCTGGGCCTGCTGCGCAGCCTGTATGCGATGATTAGCGTGGATGAAGATGCGAGCTTT
CTGCGCTATGGCTATCTGAGCAAAGAAAACGCGAGCGCGGTGCGCAAAGAAGTGCCGAAACTG
TGCGCGGAACTGCGCCCGCATGCGCTGGCGCTGGTGAGCAGCTTTGGCATTCCGGATGCGTTT
CTGAGCCCGATTGCGTATAACTGGCTGGATAGCAACAGCTGGAGCAGCCAGCTGtaactcgag
taaactagtgagacgcctgcagg

>VIGRR_ACX4_X2
cgtctcgaattcttaactttaagaaggagatataccATGACCATTCCGAGCAGCACCAACCCG
GATGGCAGCTTTAAAAACGCGATGCCGAGCTATTTTTATCTGCCGCCGCTGGATGTGAGCGCG
GCGTTTCCGCAGGCGACCCCGGCGAGCACCTTTCCGCCGAGCGCGAGCGATTATTTTCAGTTT
GATGATCTGCTGACCGCGGAAGAACAGGCGATTCGCAAAAAAGTGCGCGAATGCATGGAAAAA
GAAATTGCGCCGATTATGACCGAATATTGGGAAAAAGCGAAATTTCCGTTTCATGTGATTCCG
AAACTGGGCGCGCTGCATATTGCGGGCGGCACCATTAAAGGCTATGGCTGCCCGGGCCTGAGC
ATTACCGGCAGCGCGATTGCGACCGCGGAAGTGGCGCGCGTGGATGCGAGCTGCAGCACCTTT
ATTCTGGTGCATAGCAGCCTGGCGATGCTGACCATTGCGCTGTGCGGCAGCGAAGCGCAGAAA
CAGAAATATCTGCCGAGCCTGGCGCAGTTTCAGACCGTGGCGTGCTGGGCGCTGACCGAACCG
GATTATGGCAGCGATGCGAGCGCGCTGAAAACCACCGCGACCAAAGTGGAAGGCGGCTGGATT
CTGGAAGGCCAGAAACGCTGGATTGGCAACAGCACCTTTGCGGATGTGCTGGTGGTGTTTGCG
CGCAACGCGGCGACCAACCAGATTAACGGCTTTATTATTAAAAAAGATGCGCCGGGCCTGACC
GTGACCAAAATTGAAAACAAAATTGGCCTGCGCATTGTGCAGAACGGCGATATTGTGATGCGC
AAAGTGTTTGTGCCGGATGAAGATCGCATTGCGGGCGTGAACAGCTTTCAGGATACCAACAAA
GTGCTGGCGGTGAGCCGCATTATGGTGGCGTGGCAGCCGATTGGCATTAGCATGGGCATTTAT
GATATGTGCCATCGCTATCTGACCGAACGCAAACAGTTTGGCGCGCCGCTGGCGGCGTATCAG
ATTAGCCAGCAGAAACTGGTGCAGATGCTGGGCAACATTCAGGCGATGATTCTGGTGGGCTGG
CGCCTGTGCAAACTGTATGAAAGCGGCAAAATGACCCCGGGCCATGCGAGCCTGGGCAAAAGC
TGGATTACCCTGCGCGCGCGCGAAACCGCGGCGCTGGGCCGCGAACTGCTGGGCGGCAACGGC
ATTCTGGCGGATTTTCTGGTGGCGAAAGCGTTTTGCGATCTGGAACCGATTTATACCTATGAA
GGCACCTATGATATTAACACCCTGATTACCGGCCGCGAAGTGACCGGCTTTGCGAGCTTTAAA
CCGGTGGCGCAGCGCAGCCGCCTGtaactcgagtaaactagtgagacgcctgcagg

>VIGRR_ACX4_X1
cgtctcgaattcttaactttaagaaggagatataccATGACCATTCCGAGCAGCACCAACCCG
GTGTGCGCGGATGGCAGCTTTAAAAACGCGATGCCGAGCTATTTTTATCTGCCGCCGCTGGAT
GTGAGCGCGGCGTTTCCGCAGGCGACCCCGGCGAGCACCTTTCCGCCGAGCGCGAGCGATTAT
TTTCAGTTTGATGATCTGCTGACCGCGGAAGAACAGGCGATTCGCAAAAAAGTGCGCGAATGC
ATGGAAAAAGAAATTGCGCCGATTATGACCGAATATTGGGAAAAAGCGAAATTTCCGTTTCAT
GTGATTCCGAAACTGGGCGCGCTGCATATTGCGGGCGGCACCATTAAAGGCTATGGCTGCCCG
GGCCTGAGCATTACCGGCAGCGCGATTGCGACCGCGGAAGTGGCGCGCGTGGATGCGAGCTGC
AGCACCTTTATTCTGGTGCATAGCAGCCTGGCGATGCTGACCATTGCGCTGTGCGGCAGCGAA
GCGCAGAAACAGAAATATCTGCCGAGCCTGGCGCAGTTTCAGACCGTGGCGTGCTGGGCGCTG
ACCGAACCGGATTATGGCAGCGATGCGAGCGCGCTGAAAACCACCGCGACCAAAGTGGAAGGC
GGCTGGATTCTGGAAGGCCAGAAACGCTGGATTGGCAACAGCACCTTTGCGGATGTGCTGGTG
GTGTTTGCGCGCAACGCGGCGACCAACCAGATTAACGGCTTTATTATTAAAAAAGATGCGCCG
GGCCTGACCGTGACCAAAATTGAAAACAAAATTGGCCTGCGCATTGTGCAGAACGGCGATATT
GTGATGCGCAAAGTGTTTGTGCCGGATGAAGATCGCATTGCGGGCGTGAACAGCTTTCAGGAT
ACCAACAAAGTGCTGGCGGTGAGCCGCATTATGGTGGCGTGGCAGCCGATTGGCATTAGCATG
GGCATTTATGATATGTGCCATCGCTATCTGACCGAACGCAAACAGTTTGGCGCGCCGCTGGCG
GCGTATCAGATTAGCCAGCAGAAACTGGTGCAGATGCTGGGCAACATTCAGGCGATGATTCTG
GTGGGCTGGCGCCTGTGCAAACTGTATGAAAGCGGCAAAATGACCCCGGGCCATGCGAGCCTG
GGCAAAAGCTGGATTACCCTGCGCGCGCGCGAAACCGCGGCGCTGGGCCGCGAACTGCTGGGC
GGCAACGGCATTCTGGCGGATTTTCTGGTGGCGAAAGCGTTTTGCGATCTGGAACCGATTTAT
ACCTATGAAGGCACCTATGATATTAACACCCTGATTACCGGCCGCGAAGTGACCGGCTTTGCG
AGCTTTAAACCGGTGGCGCAGCGCAGCCGCCTGtaactcgagtaaactagtgagacgcctgca
gg

>ZEAMA_ACX1
cgtctcgaattcttaactttaagaaggagatataccATGTTTAGCCGCAACCTGACCAGCAGC
CGCCCGCGCCTGCCGCGCTGCGTGAGCCGCCCGCCGATTTATATTTATCATAAAAAAACCGCG
AAAAGCAGCGAAGCGACCCGCGCGAGCAGCAGCTGGCGCGTGGGCGCGGAAATTGCGATGGAT
GCGAGCGCGGAAGTGGATCATCTGGCGGCGGAACGCAGCGCGGCGCGCTTTGATGTGGAAGCG
ATGAAAGTGGCGTGGGCGGGCAGCCGCCATGCGGTGGAAGTGGGCGATCGCATGGCGCGCCTG
GTGGCGAGCGATCCGGTGTTTCGCAAAGATAACCGCACCATGCTGAGCCGCAAgGATCTGTTT
AAAGATACCCTGCGCAAAGCGGCGCATGCGTGGAAACGCATTGTGGAACTGCGCCTGACCGAA
GAAGAAGCGGGCATGCTGCGCCTGTATGTGGATCAGCCGGGCTATGTGGATCTGCATTGGGGC
ATGTTTGTGCCGGCGATTAAAGGCCAGGGCACCGAAGAACAGCAGAAAAAATGGCTGCCGATG

GCGTATAAATTTCAGATTATTGGCTGCTATGCGCAGACCGAACTGGGCCATGGCAGCAACGTG
CAGGGCCTGGAAACCACCGCGACCTTTGATCCGAAAACCGATGAATTTGTGATTCATAGCCCG
ACCCTGACCAGCAGCAAATGGTGGCCGGGCGGCCTGGGCAAAGCGAGCACCCATGCGGTGGTG
TATGCGCGCCTGATTACCGAAGGCAAAGATTATGGCATTCATGGCTTTATTGTGCAGCTGCGC
AGCCTGGAAGATCATAGCCCGCTGCCGGGCATTACCCTGGGCGATATTGGCGGCAAATTTGGC
AGCGGCGCGTATAACAGCATGGATAACGGCGTGCTGCGCTTTGATCATGTGCGCATTCCGCGC
GATCAGATGCTGATGCGCCTGAGCCAGGTGACCCGCGAAGGCAAATATGTGCATAGCGATGTG
CCGAAACAGCTGCTGTATGGCACCATGGTGTTTGTGCGCCAGACCATTGTGGCGGATGCGAGC
AAAGCGCTGAGCCGCGCGGTGTGCATTGCGGTGCGCTATAGCGCGATTCGCAAACAGTTTGGC
AGCCAGGATGGCGGCCCGGAAACCAAAGTGCTGGATTATAAAACCCAGCAGAGCCGCCTGTTT
CCGCTGCTGGCGAGCGCGTATGCGTTTCGCTTTGTGGGCGATTGGCTGAAATGGCTGTATATG
GATGTGACCCAGAAACTGGAAGCGAAAGATTATAGCACCCTcCAGGAAGCGCATGCGTGCACC
GCGGGCCTGAAAGCGGTGACCACCAGCGCGACCGCGGATGCGATTGAAGAATGCCGCAAACTG
TGCGGCGGCCATGGCTATCTGAACAGCAGCGGCCTGCCGGAACTGTTTGCGGTGTATGTGCCG
GCGTGCACCTATGAAGGCGATAACATTGTGCTGCTGCTGCAGGTGGCGCGCATTCTGATGAAA
ACCCTGAGCCAGCTGACCAGCGGCAAACAGCCGGTGGGCACCATGGCGTATATGGGCAACGTG
CAGTATCTGATGCAGTGCAAATGCGCGGTGAACACCGCGGAAGATTGGCTGAACCCGGTGGCG
ATTCAGGAAGCGTTTGAAGCGCGCGCGCTGCGCATGGCGGTGAACTGCGCGCAGAACATTGGC
CAGGCGGCGAACCAGGAAGAAGGCTTTTATGAACGCAGCCCGGATCTGCTGGAAGCGGCGGTG
GCGCATATTCAGCTGGTGATTGTGACCAAATTTATTAGCAAAGTGCAGCAGGATATTCCGGGC
CCGGGCGTGAAAGAACAGCTGCAGAACCTGTGCAACGTGTATGCGCTGTATATTCTGCATAAA
CATCTGGGCGATTTTCTGGCGACCGGCTGCATTACCCCGAAACAGGGCGCGCTGGCGAACGAA
CAGCTGGGCAAACTGTATGCGCAGGTGCGCCCGAACGCGGTGGCGCTGGTGGATGCGTTTAAC
TATACCGATCATTATCTGGGCAGCGTGCTGGGCCGCTATGATGGCAACGTGTATCCGGCGCTG
TATGAAGAAGCGTGGAAAGATCCGCTGAACGAAACCGTGGTGCCGGAAGGCTATCATGAATAT
CTGCGCCCGCTGCTGAAACAGCAGCTGAAACTGAGCCGCCTGtaactcgagtaaactagtgag
acgcctgcagg

>ZEAMA_ACX
cgtctcgaattcttaactttaagaaggagatataccATGGATGCGAGCGCGGAAGTGGATCAT
CTGGCGGCGGAACGCAGCGCGGCGCGCTTTGATGTGGAAGCGATGAAAGTGGCGTGGGCGGGC
AGCCGCCATGCGGTGGAAGTGGGCGATCGCATGGCGCGCCTGGTGGCGAGCGATCCGGTGTTT
CGCAAAGATAACCGCACCATGCTGAGCCGCAAAGAcCTGTTTAAAGATACCCTGCGCAAAGCG
GCGCATGCGTGGAAACGCATTGTGGAACTGCGCCTGACCGAAGAAGAAGCGGGCATGCTGCGC
CTGTATGTGGATCAGCCGGGCTATGTGGATCTGCATTGGGGCATGTTTGTGCCGGCGATTAAA
GGCCAGGGCACCGAAGAACAGCAGAAAAAATGGCTGCCGATGGCGTATAAATTTCAGATTATT
GGCTGCTATGCGCAGACCGAACTGGGCCATGGCAGCAACGTGCAGGGCCTGGAAACCACCGCG
ACCTTTGATCCGAAAACCGATGAATTTGTGATTCATAGCCCGACCCTGACCAGCAGCAAATGG
TGGCCGGGCGGCCTGGGCAAAGCGAGCACCCATGCGGTGGTGTATGCGCGCCTGATTACCGAA
GGCAAAGATTATGGCATTCATGGCTTTATTGTGCAGCTGCGCAGCCTGGAAGATCATAGCCCG
CTGCCGGGCATTACCCTGGGCGATATTGGCGGCAAATTTGGCAGCGGCGCGTATAACAGCATG
GATAACGGCGTGCTGCGCTTTGATCATGTGCGCATTCCGCGCGATCAGATGCTGATGCGCCTG
AGCCAGGTGACCCGCGAAGGCAAATATGTGCATAGCGATGTGCCGAAACAGCTGCTGTATGGC
ACCATGGTGTTTGTGCGCCAGACCATTGTGGCGGATGCGAGCAAAGCGCTGAGCCGCGCGGTG
TGCATTGCGGTGCGCTATAGCGCGATTCGCAAACAGTTTGGCAGCCAGGATGGCGGCCCGGAA
ACCAAAGTGCTGGATTATAAAACCCAGCAGAGCCGCCTGTTTCCGCTGCTGGCGAGCGCGTAT
GCGTTTCGCTTTGTGGGCGATTGGCTGAAATGGCTGTATATGGATGTGACCCAGAAACTGGAA
GCGAAAGATTATAGCACCCTcCAGGAAGCGCATGCGTGCACCGCGGGCCTGAAAGCGGTGACC
ACCAGCGCGACCGCGGATGCGATTGAAGAATGCCGCAAACTGTGCGGCGGCCATGGCTATCTG
AACAGCAGCGGCCTGCCGGAACTGTTTGCGGTGTATGTGCCGGCGTGCACCTATGAAGGCGAT
AACATTGTGCTGCTGCTGCAGGTGGCGCGCATTCTGATGAAAACCCTGAGCCAGCTGACCAGC
GGCAAACAGCCGGTGGGCACCATGGCGTATATGGGCAACGTGCAGTATCTGATGCAGTGCAAA
TGCGCGGTGAACACCGCGGAAGATTGGCTGAACCCGGTGGCGATTCAGGAAGCGTTTGAAGCG
CGCGCGCTGCGCATGGCGGTGAACTGCGCGCAGAACATTGGCCAGGCGGCGAACCAGGAAGAA
GGCTTTTATGAACGCAGCCCGGATCTGCTGGAAGCGGCGGTGGCGCATATTCAGCTGGTGATT
GTGACCAAATTTATTAGCAAAGTGCAGCAGGATATTCCGGGCCCGGGCGTGAAAGAACAGCTG
CAGAACCTGTGCAACGTGTATGCGCTGTATATTCTGCATAAACATCTGGGCGATTTTCTGGCG
ACCGGCTGCATTACCCCGAAACAGGGCGCGCTGGCGAACGAACAGCTGGGCAAACTGTATGCG
CAGGTGCGCCCGAACGCGGTGGCGCTGGTGGATGCGTTTAACTATACCGATCATTATCTGGGC

AGCGTGCTGGGCCGCTATGATGGCAACGTGTATCCGGCGCTGTATGAAGAAGCGTGGAAAGAT
CCGCTGAACGAAACCGTGGTGCCGGAAGGCTATCATGAATATCTGCGCCCGCTGCTGAAACAG
CAGCTGAAACTGAGCCGCCTGtaactcgagtaaactagtgagacgcctgcagg

>ZEAMA_ACX1_2
cgtctcgaattcttaactttaagaaggagatataccATGGCGGAAGTGGATCATCTGGCGGCG
GAACGCGCGACCGCGCGCTTTGATGTGGAAGAAATGAAAGTGGCGTGGGCGGGCAGCCGCCAT
GCGGTGCATGTGGCGGATCGCATGGCGCGCCTGGTGGCGAGCGATCCGGTGTTTCGCAAAGAT
AACCGCACCATGCTGAGCCGCAAAGAACTGTTTAAAGATACCCTGCGCAAAGCGGCGCATGCG
TGGAAACGCATTGTGGAACTGCGCCTGACCGAAGAAGAAGCGGATCTGCTGCGCCAGTATGTG
GATCAGCCGGGCTATGTGGATCTGCATTGGGGCATGTTTGTGCCGGCGATTAAAGGCCAGGGC
ACCGAAGAACAGCAGCATAAATGGCTGCCGCTGGCGTATAAATTTCAGATTATTGGCTGCTAT
GCGCAGACCGAACTGGGCCATGGCAGCAACGTGCAGGGCCTGGAAACCACCGCGACCTTTGAT
CCGAAAACCGATGAATTTGTGATTCATAGCCCGACCCTGACCAGCAGCAAATGGTGGCCGGGC
GGCCTGGGCAAAGCGAGCACCCATGCGGTGGTGTATGCGCGCCTGATTACCGAAGGCAAAGAT
TATGGCATTCATGGCTTTATTGTGCAGCTGCGCAGCCTGGAAGATCATAGCCCGCTGCCGGGC
GTGACCCTGGGCGATATTGGCGGCAAATTTGGCAGCGGCGCGTATAACAGCATGGATAACGGC
GTGCTGCGCTTTGATCATGTGCGCATTCCGCGCGATCAGATGCTGATGCGCCTGAGCCAGGTG
ACCCGCGAAGGCAAATATGTGCATAGCAACGTGCCGAAACAGCTGCTGTATGGCACCATGGTG
TATGTGCGCCAGAGCATTGTGGCGGATGCGAGCAAAGCGCTGAGCCGCGCGGTGTGCATTGCG
GTGCGCTATAGCGCGGTGCGCAAACAGTTTGGCAGCCAGGATGGCGGCCCGGAAACCCAGGTG
CTGAACTATAAAACCCAGCAGAGCCGCCTGTTTCCGCTGCTGGCGAGCGCGTATGCGTTTCGC
TTTGTGGGCGATTGGCTGAACTGGCTGTATACCGATGTGACCCAGAAACTGGAAGCGAAAGAT
TTTAGCACCCTcCAGGAAGCGCATGCGTGCACCGCGGGCCTGAAAGCGGTGACCACCAGCGTG
ACCGCGGATGCGATTGAAGAATGCCGCAAACTGTGCGGCGGCCATGGCTATCTGAACAGCAGC
GGCCTGCCGGAACTGTTTGCGGTGTATGTGCCGGCGTGCACCTATGAAGGCGATAACGTGGTG
CTGCTGCTGCAGGTGGCGCGCATTCTGATGAAAACCGTGAGCCAGCTGGCGAGCGGCAAACAG
CCGGTGGGCACCATGGCGTATATGGGCAAAGTGCAGTATCTGATGCAGTGCAAAAGCGCGGTG
AGCAGCGCGGAAGATTGGCTGAACCCGGATGCGATTCAGGAAGCGTTTGAAGCGCGCGCGCTG
CGCATGGCGGTGAACTGCGCGCAGAACATTAGCCAGGCGGCGAGCCAGGAAGAAGCGGCGGGC
TTTTATGAACGCAGCCCGGATCTGCTGGAAGCGGCGGTGGCGCATATTCAGCTGATTATTGTG
ACCAAATTTATTGAAAAAGTGCATCAGGAAATTCCGGGCCATGGCGTGAAAGAACAGCTGCAG
GCGCTGTGCAACGTGTATGCGCTGTATATTCTGCATAAACATCTGGGCGATTTTCTGGCGACC
GGCTGCATTACCCCGCGCCAGGGCGCGCTGGCGAACGAACAGCTGGGCAAACTGTATGCGCAG
GTGCGCCCGAACGCGGTGGCGCTGGTGGATGCGTTTAACTATACCGATCATTATCTGGGCAGC
GTGCTGGGCCGCTATGATGGCGATGTGTATCCGGCGCTGTATGAAGAAGCGTGGAAAGATCCG
CTGAACGAAACCGTGGTGCCGGAAGGCTATCATGAATATCTGCGCCCGCTGATTAAACAGCAG
CAGCTGAAACTGAGCCGCCTGtaactcgagtaaactagtgagacgcctgcagg

>ZEAMA_ACX4
cgtctcgaattcttaactttaagaaggagatataccATGGCGGGCAAACGCGTGACCGGCCGC
GATGGCCAGGATGAAGGCGGCGCGAAAGCGGGCCTGCCGGCGATGGATATTAGCCTGGCGTTT
CCGCAGGCGACCCCGGCGAGCATTTTTCCGCCGAGCGCGAGCGATTATTATCAGAGCGATGAT
CTGCTGACCAGCGAAGAACGCAGCATTCGCATGAAAGTGCGCGGCATTATGGAAAAAGAAATT
GCGCCGATTATGAGCGCGTATTGGGAAAAAGCGGAATTTCCGTTTAACGCGATTCCGAAACTG
GCGAGCCTGGGCGTGGCGGGCGGCACCATTAAAGGCTATGGCTGCCCGGGCCTGAGCATTACC
GCGAGCGCGGTGACCATGAGCGAAATTGCGCGCGTGGATGCGAGCTGCAGCACCTTTATTCTG
GTGCATAGCAGCCTGGCGATGGTGACCATTGCGCTGTGCGGCAGCGAAGCGCAGAAACAGAAA
TATCTGCCGAGCCTGGCGCAGCTGACCACCGTGGGCTGCTGGGCGCTGACCGAACCGAACTAT
GGCAGCGATGCGAGCAGCCTGCGCACCATTGCGACCAAAGCGCCGGGCGGCTGGCATATTGAT
GGCCAGAAACGCTGGATTGGCAACAGCACCTTTGCGGATGTGCTGGTGGTGCTGGCGCGCAAC
GCGGATACCCAGCAGCTGAACGGCTTTATTGTGCGCAAAGGCGCGCCGGGCCTGAAAGCGACC
AAAATTGAAAACAAAATTGGCCTGCGCATGGTGCAGAACGGCGATATTGTGCTGAACAAAGTG
TTTGTGCCGGATGAAGATCGCCTGCCGGGCATTAACAGCTTTCAGGATATTAGCAAAGTGCTG
GCGATTAGCCGCATTATGGTGGCGTGGCAGCCGGTGGGCATTAGCATGGGCGTGTTTGATATG
TGCCATCGCTATCTGAAAGAACGCAAACAGTTTGGCGTGCCGCTGGCGGCGTTTCAGCTGAAC
CAGGAAAAACTGGCGCGCATGCTGGGCAACGTGCAGAGCATGGTGCTGGTGGGCTGGCGCCTG
TGCAAACTGTATGAAAGCGGCAAAATGACCCCGGGCCATGCGAGCCTGGGCAAAGCGTGGAAC
AGCCGCACCGCGCGCGAAGTGGTGAGCCTGGGCCGCGAACTGCTGGGCGGCAACGGCATTCTG

GCGGATTTTCTGGTGGCGAAAGCGTTTTGCGATCTGGAACCGATTTATAGCTATGAAGGCACC
TATGATATTAACAGCCTGGTGACCGGCCGCGAAGTGACCGGCATTGCGAGCTTTCGCCCGGCG
GCGCCGGCGAAAGCGCGCCTGtaactcgagtaaactagtgagacgcctgcagg

>YARLI_ACX3
cgtctcgaattcttaactttaagaaggagatataccATGATTAGCCCGAACCTGACCGCGAAC
GTGGAAATTGATGGCAAACAGTATAACACCTTTACCGAACCGCCGAAAGCGCTGGCGGGCGAA
CGCGCGAAAGTGAAATTTCCGATTAAAGATATGACCGAATTTCTGCATGGCGGCGAAGAAAAC
GTGACCATGATTGAACGCCTGATGACCGAACTGGAACGCGATCCGGTGCTGAACGTGAGCGGC
GATTATGATATGCCGAAAGAACAGCTGCGCGAAACCGCGGTGGCGCGCATTGCGGCGCTGAGC
GGCCATTGGAAAAAAGATACCGAAAAAGAAGCGCTGCTGCGCAGCCAGCTGCATGGCATTGTG
GATATGGGCACCCGCATTCGCCTGGGCGTGCATACCGGCCTGTTTATGGGCGCGATTCGCGGC
AGCGGCACCAAAGAACAGTATGATTATTGGGTGCGCAAAGGCGCGGCGGATGTGAAAGGCTTT
TATGGCTGCTTTGCGATGACCGAACTGGGCCATGGCAGCAACGTGGCGGGCCTGGAAACCACC
GCGACCTATATTCAGGATACCGATGAATTTATTATTAACACCCCGAACACCGGCGCGACCAAA
TGGTGGATTGGCGGCGCGGCGCATAGCGCGACCCATACCGCGTGCTTTGCGCGCCTGCTGGTG
GATGGCAAAGATTATGGCGTGAAAATTTTTGTGGTGCAGCTGCGCGATGTGAGCAGCCATAGC
CTGATGCCGGGCATTGCGCTGGGCGATATTGGCAAAAAAATGGGCCGCGATGCGATTGATAAC
GGCTGGATTCAGTTTACCAACGTGCGCATTCCGCGCCAGAACATGCTGATGAAATATGCGAAA
GTGAGCAGCACCGGCAAAGTGAGCCAGCCGCCGCTGGCGCAGCTGACCTATGGCGCGCTGATT
GGCGGCCGCGTGACCATGATTGCGGATAGCTTTTTTGTGAGCCAGCGCTTTATTACCATTGCG
CTGCGCTATGCGTGCGTGCGCCGCCAGTTTGGCACCACCCCGGGCCAGCCGGAAACCAAAATT
ATTGATTATCCGTATCATCAGCGCCGCCTGCTGCCGCTGCTGGCGTTTACCTATGCGATGAAA
ATGGCGGCGGATCAGAGCCAGATTCAGTATGATCAGACCACCGATCTGCTGCAGACCATTGAT
CCGAAAGATAAAGGCGCGCTGGGCAAAGCGATTGTGGATCTGAAAGAACTGTTTGCGAGCAGC
GCGGGCCTGAAAGCGTTTACCACCTGGACCTGCGCGAACATTATTGATCAGTGCCGCCAGGCG
TGCGGCGGCCATGGCTATAGCGGCTATAACGGCTTTGGCCAGGCGTATGCGGATTGGGTGGTG
CAGTGCACCTGGGAAGGCGATAACAACGTGCTGTGCCTGAGCATGGGCCGCGGCCTGATTCAG
AGCTGCCTGGGCCATCGCAAAGGCAAACCGCTGGGCAGCAGCGTGGGCTATCTGGCGAACAAA
GGCCTGGAACAGGCGACCCTGAGCGGCCGCGATCTGAAAGATCCGAAAGTGCTGATTGAAGCG
TGGGAAAAAGTGGCGAACGGCGCGATTCAGCGCGCGACCGATAAATTTGTGGAACTGACCAAA
GGCGGCCTGAGCCCGGATCAGGCGTTTGAAGAACTGAGCCAGCAGCGCTTTCAGTGCGCGAAA
ATTCATACCCGCAAACATCTGGTGACCGCGTTTTATGAACGCATTAACGCGAGCGCGAAAGCG
GATGTGAAACCGTATCTGATTAACCTGGCGAACCTGTTTACCCTGTGGAGCATTGAAGAAGAT
AGCGGCCTGTTTCTGCGCGAAGGCTTTCTGCAGCCGAAAGATATTGATCAGGTGACCGAACTG
GTGAACCATTATTGCAAAGAAGTGCGCGATCAGGTGGCGGGCTATACCGATGCGTTTGGCCTG
AGCGATTGGTTTATTAACGCGCCGATTGGCAACTATGATGGCGATGTGTATAAACATTATTTT
GCGAAAGTGAACCAGCAGAACCCGGCGCAGAACCCGCGCCCGCCGTATTATGAAAGCACCCTG
CGCCCGTTTCTGTTTCGCGAAGATGAAGATGATGATATTTGCGAACTGGATGAAGAAtaactc
gagtaaactagtgagacgcctgcagg

>CANTR_ACX
cgtctcgaattcttaactttaagaaggagatataccATGACCTTTACCAAAAAAAACGTGAGC
GTGAGCCAGGGCCCGGATCCGCGCAGCAGCATTCAGAAAGAACGCGATAGCAGCAAATGGAAC
CCGCAGCAGATGAACTATTTTCTGGAAGGCAGCGTGGAACGCAGCGAACTGATGAAAGCGCTG
GCGCAGCAGATGGAACGCGATCCGATTCTGTTTACCGATGGCAGCTATTATGATCTGACCAAA
GATCAGCAGCGCGAACTGACCGCGGTGAAAATTAACCGCATTGCGCGCTATCGCGAACAGGAA
AGCATTGATACCTTTAACAAACGCCTGAGCCTGATTGGCATTTTTGATCCGCAGGTGGGCACC
CGCATTGGCGTGAACCTGGGCCTGTTTCTGAGCTGCATTCGCGGCAACGGCACCACCAGCCAG
CTGAACTATTGGGCGAACGAAAAAGAAACCGCGGATGTGAAAGGCATTTATGGCTGCTTTGGC
ATGACCGAACTGGCGCATGGCAGCAACGTGGCGGGCCTGGAAACCACCGCGACCTTTGATAAA
GAAAGCGATGAATTTGTGATTAACACCCCGCATATTGGCGCGACCAAATGGTGGATTGGCGGC
GCGGCGCATAGCGCGACCCATTGCAGCGTGTATGCGCGCCTGATTGTGGATGGCCAGGATTAT
GGCGTGAAAACCTTTGTGGTGCCGCTGCGCGATAGCAACCATGATCTGATGCCGGGCGTGACC
GTGGGCGATATTGGCCCGAAAATGGGCCGCGATGGCATTGATAACGGCTGGATTCAGTTTAGC
AACGTGCGCATTCCGCGCTTTTTTATGCTGCAGAAATTTTGCAAAGTGAGCGCGGAAGGCGAA
GTGACCCTGCCGCCGCTGGAACAGCTGAGCTATAGCGCGCTGCTGGGCGGCCGCGTGATGATG
GTGCTGGATAGCTATCGCATGCTGGCGCGCATGAGCACCATTGCGCTGCGCTATGCGATTGGC
CGCCGCCAGTTTAAAGGCGATAACGTGGATCCGAACGATCCGAACGCGCTGGAAACCCAGCTG

ATTGATTATCCGCTGCATCAGAAACGCCTGTTTCCGTATTTTGTGCCGCCGATGAGCAGCCCG
AGCGTGCCGAGCCGCCTGAACACCCCGAGCCGCCCGCCGTGGAGCAACTGGACCAGCCCGCTG
AAACGCACCACCCCGCGCCTGATTTTTAAAAGCATTGATGATATGAAAAGCCTGTTTGTGGAT
AGCGGCAGCCTGAAAAGCACCGCGACCTGGCTGGGCGCGGAAGCGATTGATCAGTGCCGCCAG
GCGTGCGGCGGCCATGGCCATAGCAGCTATAACGGCTTTGGCAAAGCGTATAACGATTGGGTG
GTGCAGTGCACCTGGGAAGGCGATAACAACGTGCTGGGCATGAGCGTGGGCAAACCGATTGTG
AAACAGGTGATTAGCATTGAAGATGCGGGCAAAACCGTGCGCGGCAGCACCGCGTTTCTGAAC
CAGCTGAAAGAATATACCGGCAGCAACAGCAGCAAAGTGGTGCTGAACACCGTGGCGGATCTG
GATGATATTAAAACCGTGATTAAAGCGATTGAAGTGGCGATTATTCGCCTGAGCCAGGAAGCG
GCGAGCATTGTGAAAAAAGAAAGCTTTGATTATGTGGGCGCGGAACTGGTGCAGCTGAGCAAA
CTGAAAGCGCATCATTATCTGCTGACCGAATATATTCGCCGCATTGATACCTTTGATCAGAAA
GAACTGGCGCCGTATCTGATTACCCTGGGCAAACTGTATGCGGCGACCATTGTGCTGGATCGC
TTTGCGGGCGTGTTTCTGACCTTTAACGTGGCGAGCACCGAAGCGATTACCGCGCTGGCGAGC
GTGCAGATTCCGAAACTGTGCGCGGAAGTGCGCCCGAACGTGGTGGCGTATACCGATAGCTTT
CAGCAGAGCGATATGATTGTGAACAGCGCGATTGGCCGCTATGATGGCGATATTTATGAAAAC
TATTTTGATCTGGTGAAACTGCAGAACCCGCCGAGCAAAACCAAAGCGCCGTATAGCGATGCG
CTGGAAGCGATGCTGAACCGCCCGACCCTGGATGAACGCGAACGCTTTCAGAAAAGCGATGAA
ACCGCGGCGATTCTGAGCAAA

>POPAL_ACX4_X3
ATGAGCATTCCGAAAAACAAAGATGAA
GTGGAACAGAACCCGCGCAGCAGCTATTTTAACCTGCCGCCGCTGGATGTGAGCGTGGCGTTT
CCGCAGGCGACCCCGGTGAGCACCTTTCCGCCGAGCGTGAGCGATTATTATCAGTTTAGCGAT
CTGCTGACCGTGGAAGATCAGGCGCTGCGCATGCGCGTGCGCGAATGCATGGAAAAAGAAATT
GCGCCGATTATGACCGAATATTGGGAAAAAGCGAAATTTCCGTTTCATGCGATTCCGAAACTG
GGCAGCCTGGGCGTGGCGGGCGGCACCATTAAAGGCTATGGCTGCCCGGGCCTGAGCATTACC
GCGAGCGCGATTGCGATGGCGGAAGTGGCGCGCGTGGATGCGAGCTGCAGCACCTTTATTCTG
GTGCATAGCAGCCTGGCGATGCTGACCATTGCGTTTTGCGGCAGCGAAGAACAGAAACTGAAA
TATCTGCCGAGCCTGGCGCAGTTTAGCACCGTGGCGTGCTGGGCGCTGACCGAACCGGATTAT
GGCAGCGATGCGAGCGGCCTGAAAACCACCGCGATTAAAGTGGAAGGCGGCTGGATTCTGGAA
GGCAAAAAACGCTGGATTGGCAACAGCACCTTTGCGGATCTGCTGGTGGTGTTTGCGCGCAAC
ACCACCACCAACCAGATTAACGGCTTTATTGTGAAAAAAGATGCGCCGGGCCTGACCGTGACC
AAAATTGAAAACAAAATTGGCCTGCGCATTGTGCAGAACGGCGATATTGTGATGAAACGCGTG
TTTGTGCCGGATGAAGATCGCCTGCCGGGCGTGAACAGCTTTCAGGATACCAACAAAGTGCTG
GCGGTGAGCCGCGTGATGGTGGCGTGGCAGCCGATTGGCATTAGCATGGGCGTGTATGATATG
TGCCATCGCTATCTGAAAGAACGCAAACAGTTTGGCGCGCCGCTGGCGGCGTTTCAGATTAAC
CAGCAGAAACTGGTGCATATGCTGGGCAACGTGCAGGCGATGGTGCTGGTGGGCTGGCGCCTG
TGCAAACTGTATGAAAAAGAAAAAATGACCCCGGGCCAGGCGAGCCTGGCGAAAGCGTGGATT
AGCCTGAAAGCGCGCGAAACCGCGAGCATTGGCCGCGAACTGCTGGGCGGCAACGGCATTCTG
GCGGATTTTCTGGTGGCGAAAGCGCTGTGCGATCTGGAACCGATTTATACCTATGAAGGCACC
TATGATATTAACACCCTGATTACCGGCCGCGAAATTACCGGCCTGGCGAGCTTTAAACCGGCG
ATGCTGAGCAAACGCAGCCGCATG

>ARAHY_ACX4
ATGGCGATTCATAGCAGCAGCAACCAG
GATGATAGCAACAAACGCGCGATGCCGAGCTATTTTAACCTGCCGCCGCTGGATGTGAGCAAC
GCGTTTCCGCAGGCGACCCCGGCGAGCGTGTTTCCGCCGTGCGCGAGCGATTATTTTCAGTTT
GATGATCTGCTGACCCCGGAAGAACAGGCGGTGCGCAAAAAAGTGCGCGAATGCATGGAAAAA
GAAATTGCGCCGGTGATGACCGAATATTGGGAAAAAGCGAAATTTCCGTTTCAGTGCATTCCG
AAACTGGCGGCGCTGCGCCTGGCGGGCGGCACCATTAAAGGCTATGGCTGCCCGGGCCTGAGC
ATTCTGGGCAGCGCGGTGGGCACCGCGGAAGTGGCGCGCGTGGATGCGAGCTGCAGCACCTTT
ATTCTGGTGCATAGCAGCCTGGCGATGCTGACCATTGCGCTGTGCGGCAGCGAAGCGCAGAAA
CAGAAATATCTGCCGAGCCTGGCGCAGTTTCAGACCATTGGCTGCTGGGGCCTGACCGAACCG
GATTATGGCAGCGATGCGAGCGCGCTGAAAACCACCGCGACCAAAGTGGAAGGCGGCTGGATT
CTGGATGGCCAGAAACGCTGGATTGGCAACAGCACCTTTGCGGATATTCTGGTGATTTTTGCG
CGCAACACCAGCACCAACCAGATTAACGGCTATATTGTGAAAAAAGATGCGCCGGGCCTGACC
GTGACCAAAATTGAAAACAAAATTGGCCTGCGCATTGTGCAGAACGGCGATATTGTGATGAAA
AAAGTGTTTATTCCGGATGAAGATCGCCTGACCGGCGTGAACAGCTTTCAGGATACCAACAAA
GTGCTGGCGGTGAGCCGCGTGATGGTGGCGTGGCAGCCGATTGGCATTAGCATGGGCATTTAT

GATATTTGCCATCGCTATCTGAAAGAACGCAAACAGTTTGGCGCGCCGCTGGCGGCGTTTCAG
ATTAACCAGCAGAAACTGGTGCAGATGCTGAGCAACGTGCAGGCGATGTTTCTGGTGGGCTGG
CGCCTGTGCAAACTGTATGAAAGCGGCAAAATGACCCCGGGCCATGCGAGCCTGGCGAAAAGC
TGGATTACCCTGCGCGCGCGCGAAACCGCGGCGCTGGGCCGCGAACTGCTGGGCGGCAACGGC
ATTCTGGCGGATTTTCTGGTGGCGAAAGCGTTTTGCGATATTGAACCGATTTATACCTATGAA
GGCACCTATGATATTAACACCCTGGTGACCGGCCGCGAAGTGACCGGCTTTGCGAGCTTTAAA
CCGGTGATTGCGCAGAAACGCAGCCGCCTG**taactcgagtaaactagtgagacgcctgcagg**

>TREOR_ACX
cgtctcgaattcttaactttaagaaggagatataccATGACCGTGCGCAGCGATCAGAACCAG
GATGAAATTGAAAAAAACGAACGCACCAGCTATTTTCGCCTGCCGGCGCTGGATGTGAGCATT
GCGTTTCCGCAGGCGACCCCGGCGAGCGTGTTTCCGCCGAGCACCAGCGATTATTATGTGTTT
GATGATCTGCTGACCCATGAAGAAAAAAGCCTGCGCATTAAAGTGCGCGAATGCGTGGAAAAA
GAAGTGGCGCCGATTATGGCGAAATATTGGGAAAAAGCGGAATTTCCGTTTCATATTATTCCG
AAACTGGCGGCGCTGCGCGTGGCGGGCGGCACCATTAAAGGCTATGGCTGCCCGGGCCTGAGC
CTGACCGGCAGCGCGATTGCGACCGCGGAAATTGCGCGCGTGGATGCGAGCTGCAGCACCTTT
ATTCTGGTGCATAGCAGCCTGGCGATGCTGACCATTGCGCTGTGCGGCAGCGAAGCGCAGAAA
GAACGCTATCTGCCGAGCCTGGCGCAGCTGAAAACCGTGGCGTGCTGGGCGCTGACCGAACCG
GGCTATGGCAGCGATGCGAGCAGCGTGGCGACCACCGCGACCAAAGTGGAAGGCGGCTGGATT
CTGGAAGGCCAGAAACGCTGGATTGGCAACAGCACCTTTGCGGATGTGCTGGTGATTTTTGCG
CGCAACACCACCACCAACCAGATTAACGGCTTTATTGTGAAAAAAAACGCGCCGGGCCTGACC
GCGACCAAAATTGAAAACAAAATTGGCCTGCGCATTGTGCAGAACGGCGATATTCTGCTGAAA
AAAGTGTTTGTGCCGGATGAAGATAAACTGCCGGGCGTGAACAGCTTTCAGGATACCAGCAAA
GTGCTGGCGGTGAGCCGCGTGATGGTGGCGTGGCAGCCGATTGGCATTAGCATGGGCGTGTAT
GATATGTGCCATCGCTATCTGAAAGAACGCAAACAGTTTGGCGCGCCGCTGGCGGCGTTTCAG
ATTAACCAGGAAAAACTGGCGCGCATGCTGGGCAACGTGCAGGCGATGGTGCTGATGGGCTGG
CGCCTGTGCAAACTGTATGAAAACGGCAAAATGACCGCGGGCCAGGCGAGCCTGGCGAAAGGC
TGGATTACCCTGAAAGCGCGCGAAACCGTGGCGCTGGGCCGCGAACTGCTGGGCGGCAACGGC
ATTCTGAGCGATTTTCTGGTGGCGAAAGCGTTTTGCGATCTGGAACCGATTTATACCTATGAA
GGCACCTATGAAATTAACATGCTGGTGGCGGGCCGCGAAGTGAGCGGCTTTGCGAGCTTTAAA
CCGAGCGCGAGCAGCCAGCGCAGCCGCCTG**taactcgagtaaactagtgagacgcctgcagg**

>PARAN_ACX
cgtctcgaattcttaactttaagaaggagatataccATGACCGTGCGCAGCGATCAGAACCAG
GATGAAATTGAAAAAAACGAACGCACCAGCTATTTTCGCCTGCCGGCGCTGGATGTGAGCATT
GCGTTTCCGCAGGCGACCCCGGCGAGCGTGTTTCCGCCGTGCACCAGCGATTATTATGTGTTT
GATGATCTGCTGACCCATGAAGAAAAAAGCCTGCGCATTAAAGTGCGCGAATGCGTGGAAAAA
GAAGTGGCGCCGATTATGGCGACCTATTGGGAAAAAGCGGAATTTCCGTTTCATATTATTCCG
AAACTGGCGGCGCTGCGCGTGGCGGGCGGCACCATTAAAGGCTATGGCTGCCCGGGCCTGAGC
CTGACCGGCAGCGTGATTGCGACCGCGGAAATTGCGCGCGTGGATGCGAGCTGCAGCACCTTT
ATTCTGGTGCATAGCAGCCTGGCGATGCTGACCATTGCGCTGTGCGGCAGCGAACGCCAGAAA
GAACGCTATCTGCCGAGCCTGGCGCAGCTGAAAACCGTGGCGTGCTGGGCGCTGACCGAACCG
GGCTATGGCAGCGATGCGAGCAGCGTGGCGACCACCGCGACCAAAGTGGAAGGCGGCTGGATT
CTGGAAGGCCAGAAACGCTGGATTGGCAACAGCACCTTTGCGGATGTGCTGGTGATTTTTGCG
CAGAACACCACCACCAACCAGATTAACGGCTTTATTGTGAAAAAAAACGCGCCGGGCCTGACC
GCGACCAAAATTGAAAACAAAATTGGCCTGCGCATTGTGCAGAACGGCGATATTCTGCTGAAA
AAAGTGTTTGTGCCGGATGAAGATAAACTGCCGGGCGTGAACAGCTTTCAGGATACCAGCAAA
GTGCTGGCGGTGAGCCGCGTGATGGTGGCGTGGCAGCCGATTGGCATTAGCATGGGCGTGTAT
GATATGTGCCATCGCTATCTGAAAGAACGCAAACAGTTTGGCGCGCCGCTGGCGGCGTTTCAG
ATTAACCAGGAAAAACTGGCGCGCATGCTGGGCAACGTGCAGGCGATGGTGCTGATGGGCTGG
CGCCTGTGCAAACTGTATGAAAACGGCAAAATGACCGCGGGCCAGGCGAGCCTGGGCAAAGGC
TGGATTACCCTGAAAGCGCGCGAAACCGTGGCGCTGGGCCGCGAACTGCTGGGCGGCAACGGC
ATTCTGAGCGATTTTCTGGTGGCGAAAGCGTTTTGCGATCTGGAACCGATTTATACCTATGAA
GGCACCTATGAAATTAACATGCTGGTGGCGGGCCGCGAAGTGAGCGGCTTTGCGAGCTTTAAA
CCGAGCGCGAGCAGCAAACGCAGCCGCCTG**taactcgagtaaactagtgagacgcctgcagg**

>APOSH_ACX4
cgtctcgaattcttaactttaagaaggagatataccATGGGCCAGATGCTGCGCGGCTGCAAC
ATGAGCGGCATTGAACCGCGCGCGAGCGATTATTATATTCTGGATGATCTGCTGAACAGCGAA

GAACGCAGCCTGCGCAAAATTGTGCGCGAAGTGATGGAAAAAGAAGCGGCGCCGATTATGGTG
GAATATTGGGAAAAAGCGGAATTTCCGTTTCATATTATTCCGAAACTGAGCAGCCTGCATGTG
GTGGGCGGCATTATTAAAGGCTATGGCTGCCCGGGCCTGAGCATGACCGCGACCGCGATTAGC
GTGATTGAAATGGCGCGCGTGGATGCGAGCCTGAGCAGCTTTATTGTGGCGCATAGCCCGCTG
GCGATGCTGACCATTGCGAAATTTGGCAGCGAAACCCAGAAACAGAAATATCTGCCGAGCCTG
GCGCAGCTGGATACCATTGCGTGCTGGGCGCTGACCGAACCGGGCTATGGCAACGATGCGAGC
AGCCTGAAAGCGACCGCGACCAAAGTGAACGGCGGCTGGATTCTGAACGGCCAGAAACGCTGG
GTGGGCAACAGCACCTTTGCGGATGTGTTTGTGCTGTTTGCGCGCAACACCAGCACCAGCCAG
ATTAACTGCTTTATTATTAAAAAAGGCGCGCCGGGCCTGCGCGCGACCAAAATTGAAAACAAA
GTGGCGCTGCGCATTGTGCAGAACGGCGATATTCTGCTGAAAAACGTGTTTGTGCCGGATGAA
GATTGGATGCCGGGCGTGAACAGCTTTCAGGAAACCAACGAAATTCTGGCGCATGCGCGCGTG
ATGGTGGCGAGCCAGGCGGTGGGCCTGAGCATGGGCGTGTTTGATGCGTGCAACCGCTATCTG
AAAGAACGCAAACAGTTTGGCGCGCCGCTGGCGGCGTTTCAGATTAACCAGGAAAAACTGGCG
CGCATGCTGGGCAACATTCAGGGCATGATTCTGGTGGGCTGGCGCCTGTGCAAACTGTATGAA
AGCGGCAAAATGACCCCGGGCCAGGCGAGCCTGGGCAAAGCGTGGACCAGCAAAATTGCGCGC
GAAACCGTGAGCCTGGGCCGCGAACTGCTGGGCGGCAACGGCATTCTGACCGATTTTCTGGTG
GGCAAAGCGTTTTGCGATCTGGAAAGCATTTTTAGCTATGAAGGCAGCTATGATGTGAACAGC
CTGATTACCGGCCGCGAAATTACCGGCATTGCGAGCTTTAAACCGCCGCTGTTTGATAAAGCG
AGCCGCCTGtaactcgagtaaactagtgagacgcctgcagg

>CANMA_ACX
cgtctcgaattcttaactttaagaaggagatataccATGACCTTTACCAAAAAAAACGTGAGC
GTGAGCCAGGGCCCGGATCCGCGCACCAGCATTCAGACCGAACGCGCGAACAGCAAATTTGAT
CCGGTGACCATGAACTATTTTCTGGAAGGCAGCAAAGAACGCAGCGAACTGATGAAAAGCCTG
GCGCAGCAGATTGAACGCGATCCGATTCTGTTTACCGATGGCAGCTATTATGATCTGACCAAA
GATCAGCAGCGCGAACTGACCGTGCTGAAAATTAACCGCCTGAGCCGCTATCGCGAAGGCGAT
AGCGTGGATACCTTTAACAAACGCCTGAGCATTATGGGCGTGGTGGATCCGCAGGTGGCGACC
CGCATTGGCGTGAACCTGGGCCTGTTTCTGAGCTGCATTAGCGGCAACGGCACCGCGGAACAG
TTTAAATATTGGGCGATTGATAAAGGCACCCATAACATTCAGGGCCTGTATGGCTGCTTTGGC
ATGACCGAACTGGGCCATGGCAGCAACGTGGCGGGCGTGGAAACCACCGCGACCTTTGATAAA
GAAACCGATGAATTTGTGATTAACACCCCGCATATTGGCGCGACCAAATGGTGGATTGGCGGC
GCGGCGCATAGCGCGACCCATTGCAGCGTGTATGCGCGCCTGGTGGTGGATGGCAAAGATTAT
GGCGTGAAAACCTTTGTGGTGCCGCTGCGCGATAGCAACCATGATCTGATGCCGGGCGTGACC
GTGGGCGATATTGGCGCGAAAATGGGCCGCGATGGCATTGATAACGGCTGGATTCAGTTTAGC
AACGTGCGCATTCCGCGCTTTTTTATGCTGCAGAAATTTTGCAAAGTGAGCGCGGAAGGCGAA
GTGGTGCTGCCGCCGCTGGAACAGCTGAGCTATAGCGCGCTGCTGGGCGGCCGCGTGATGATG
GTGCTGGATAGCTATCGCATGCTGGCGCGCGTGAGCACCATTGCGCTGCGCTATGCGATTGGC
CGCCGCCAGTTTAAAGGCGATAACGTGGATCAGAACGATCCGAACGCGCTGGAAACCCAGCTG
ATTGATTATCCGCTGCATCAGAAACGCCTGTTTCCGTATCTGGCGGCGGCGTATGTGGTGAGC
ACCGGCGCGCTGAAAGTGGAACATACCATTCAGAGCACCCTGGCGACCCTGGATGCGGCGGTG
GAAAACAACGATACCACCGCGATTTTTAAAAGCATTGATGATATGAAAAGCCTGTTTATTGAT
AGCGGCAGCCTGAAAGCGACCACCACCTGGCTGGCGGCGGAAGCGATTGATCAGTGCCGCCAG
GCGTGCGGCGGCCATGGCTATAGCAGCTATAACGGCTTTGCGAAAGCGTTTAACGATTGGGTG
GTGCAGTGCACCTGGGAAGGCGATAACAACGTGCTGAGCCTGAGCGTGGGCAAACCGATTATT
AAACAGATTATTGGCATTGAAGATAACGGCAAAACCGTGCGCGGCAGCACCGCGTTTCTGAAC
CAGGTGAAAGATTTTACCGGCAGCAACGCGAGCAAAGTGGTGCTGAACAACACCAGCGATCTG
AACGATATTAACAAAGTGATTAAAAGCATTGAAGTGGCGATTATTCGCCTGGCGCATGAAGCG
GCGATTAGCGTGCGCAAAGAAAGCCTGGATTTTGCGGGCGCGGAACTGGTGCAGATTAGCAAA
CTGAAAGCGCATCATTATCTGCTGACCGAATTTGTGAAACGCGTGGGCGAATTTGAACATAAA
GAACTGGTGCCGTTTCTGAACACCATTGGCCGCCTGTATAGCGCGACCGTGGTGCTGGATAAA
TTTGCGGGCGTGTTTCTGACCTTTAACGTGGCGAGCCCGCAGGCGATTACCGATCTGGCGAGC
ACCCAGATTCCGAAACTGTGCGCGGAAGTGCGCCCGAACGTGGTGGCGTATACCGATAGCTTT
CAGCAGAGCGATATGGTGATTAACAGCGCGATTGGCAAATATGATGGCGATGTGTATGAAAAC
TATTTTGATCTGGTGAAACAGCTGAACCCGCCGAAAAACACCAAAGCGCCGTATACCGCGGCG
CTGGAAGGCATGCTGAACCGCCCGAGCCTGGAAGCGCGCGAACGCTATGAAAAAAGCGATGAA
ACCGCGGCGATTCTGAGCAAAtaactcgagtaaactagtgagacgcctgcagg

>GLUNI_ACX

cgtctcgaattcttaactttaagaaggagatataccATGCCGGTGCAGCCGGCGGGCAAAGGC
GAAGCGCTGCCGTTTCCGGATGCGGATCTGATGTATGTGGTGGATCTGCTGGCGCCGCCGGAA
CGCGATCGCTATTTTCGCATTCGCCAGTTTCTGCAGAGCAGCGTGCGCCAGCAGAGCATTGAA
TATTGGAACCGCGAAGAATTTCCGTTTGGCCTGCTGGCGGATCTGGGCAAACATGGCCTGGGC
GCGCTGCAGCTGGATGGCAGCAGCACCCTGTTTAAAGGCCTGATGTATGTGGAACTGGCGCGC
GCGGATGTGAGCCTGAGCGCGCTGGCGGGCATTCATAACGAACTGATTCTGGGCATGATTAGC
GAACTGGGCAGCGAAGAACAGAAAGCGCGCTGGCTGCCGGGCCTGAAAACCTTTGGCCAGCTG
GGCGCGTTTGCGCTGACCGAACCGGATCATGGCAGCGATATTGCGGGCGGCCTGGAAACCAGC
GCGCGCCGCGATGGCGGCGAATGGGTGCTGAACGGCGCGAAACGCTGGATTGGCGCGGCGACC
ATTGCGGATTTTGCGCTGGTGTGGGCGCGCGATGTGGCGGATGGCGAAATTAAAGGCTTTATT
GTGGAAACCGATACCCCGGGCTATGCGGCGACCCGCATTGCGAACAAAATTGGCCTGCGCATT
ATGCAGAACGCGGATATTGTGCTGGAAGATGTGCGCATTCCGGCGGCGAGCCTGCTGCCGGGC
GCGACCAGCTTTGCGCGCACCAACGATCTGCTGCGCGATAGCCGCGCGTGGGTGGGCTGGCAG
GGCGCGGGCATTCAGCTGGCGGCGTTTGATGTGGCGCGCGCGTATGCGCTGAACCGCCGCCAG
TTTGGCCGCGAACTGGCGGCGTTTCAGCTGGTGCAGCAGCAGCTGGCGGATATGCTGGGCAAC
GCGAGCGCGAGCCTGGCGCTGATGGCGCAGCTGGCGCAGGTGCAGCAGGATGGCAAACTGGAA
ATGGTGCAGGCGGCGATGGCGAAAGCGACCGTGACCCGCCTGGCGCGCGCGAGCGTGGCGATG
GGCCGCAGCCTGCTGGGCGGCAACGGCATTAGCAGCGATTATGAAATGGGCAAACTGTTTGGC
GATGCGGAAATTCTGTATACCTATGAAGGCAGCTATGAAATTAACAGCCTGATTGTGGGCCGC
GCGGTGACCGGCCGCAGCGCGTTTCATtaactcgagtaaactagtgagacgcctgcagg

A.3: AAT sequences

Regions in red are the sequences added to incorporate Esp3I, EcoRI and SpeI cloning sites for Golden Gate assembly.

>AT1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGCATCTTTCCCCCCCAGCC
TTGTGTTCACGGTTCGTCGTAAGGAGCCGATCCTGGTGCTTCCATCCAAACCGACACCTCGTG
AGTTAAAGCAGTTATCCGACATCGACGATCAAGAAGGATTGCGTTTTCAAGTACCAGTAATTA
TGTTCTACAAACGTAAGCTTAGTACGGAAGGCGAAGATCCCGTAAAAGTTATCCGTGAAGCCC
TTGCAGAAGCTTTAGCCTTCTACTACCCTTTTGCAGGACGCTTGATTGAGGGACCTAATCGTA
AACTGATGGTAGATTGTACCTCCGAAGGGGTACTTTTCATCGAAGCCGACGCCGATATTGAGC
TTAACCAATTGATCGGTGATACCATCGATCCTGGCACGTACTTGGACGAGCTGTTGCACGACG
TGCCTGGCTCCGAGGGGATTTTGGGCTGCCCCTTGCTGTTGATCCAAGTGACGCGCTTCCGTT
GTGGAGGTTGGGCCTTTGCAATCCGTTTAAATCACACAATGAGTGACACATTAGGACTTGTCC
AGTTTTTAACTACGATTGCAGAATTTACTCGTGGGGCGGAAGGAGCACCTAGCGTCCCCCCCG
TCTGGCAGCGCGAATTTTTAGCTGCCCGCCAACCCCCTTTCATCCCGTTTCAGCATCATGAAT
ACGAGCAGGTCATCGATACAACACCAGATGACAATAAAAAGAGCATGACGCATAAATCGTTCT
TTTTCGGGCCCAAGGAGATTCGTGCCATTCGTTCTCACCTTCCTTTACACCACCGTTCAACAT
CTAGCACTTTCGACGTATTGACAGCATGTCTTTGGCGTTGCCGCACCTGTGCTTTGGTATTAG
ATCCGAAAAAAACAGTACGTATCTCCTGTGCAGCATCTGGACGTGGTAAGCATGACTTACACG
TCCCGCGTGGGTATTACGGTAACGTAAGTGCTTTCCCAGCAACAGTTCTTCGTGCGGGAATGA
TCTCAACGAGCCCCTTGGAATACGCAATGGAGGGTGTCAAGAAGGCCAAAGCACGCATGACTG
GGGAATACCTGCGCAGCGTCGCTGATCTGATGGTAACGAAGGGACGCCCCCTGTACACAGTTG
TGGGAAACTATATTGTGAGCGACATGACCCGTGTGGGACTTGATACAATTGACTTCGGGTGGG
GAAAGCCGGTCTATGGAGGCCCGGCCCGCGCCTTCCCGCTTATTTCATTCTATGGCCGTTTCA
AGGACAATAAGGGGGAAGATGGTATCGTCGTTCTTATCTGTCTTCCCGAGGCTGCCATGAAGC
GCTTCCAAGAGGAGCTGAAGAAGATGACTGGAGAACACGTGGACGGACCGTTTGACTACAAGC
CAATCAAGGTCGTTAGCAAGTTATAACTTAAGGAGACG

>AT9

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGCGTCGTCTGTACGCCTGG
TAAAAAAACCAGTCCTTGTTGCTCCCGTTGATCCGACACCCAGTACGGTACTGTCCTTATCGT

CTTTGGACAGTCAGCTTTTTTTGCGTTTTCCAATCGAATACTTACTTGTGTATGCGTCGCCAC
ACGGCGTTGATCGCGCGGTAACAGCCGCACGCGTCAAGGCGGCTTTAGCTCGCAGCCTTGTTC
CCTATTATCCGTTGGCTGGCCGCGTAAAGACTCGCCCAGACAGTACGGGTCTTGACGTTGTAT
GCCAAGCACAGGGCGCAGGATTACTGGAAGCGGTTTCGGACTATACAGCGAGTGATTTCCAGC
GTGCGCCTCGCTCGGTGACGGAGTGGCGCAAGTTATTGTTGGTTGAGGTATTCAAAGTCGTGC
CCCCATTGGTAGTTCAATTAACGTGGTTATCGGACGGTTGTGTCGCGCTTGGGGTTGGTTTTT
CACACTGCGTGATCGATGGAATTGGTTCTTCCGAATTTCTTAACCTTTTCGCTGAACTTGCCA
CTGGGCGCGCACGCTTGTCTGAGTTCCAACCCAAACCAGTTTGGGATCGTCACTTGTTGAACT
CTGCAGGTCGCACCAACTTAGGTACCCACCCGGAGTTCGGGCGTGTGCCTGACTTGAGCGGTT
TTGTAACCCGCTTCACGCAGGAGCGTCTGAGTCCGACCTCTATTACATTTGATAAGACGTGGC
TTAAAGAGTTGAAGAATATTGCGATGTCGACAAGTCAGCCTGGGGAATTTCCTTACACCAGTT
TTGAGGTCTTATCCGGCCACATTTGGCGCTCGTGGGCGCGTTCCTTAAACCTTCCAGCGAAGC
AAGTATTGAAACTGTTATTTTCTATTAATATCCGCAATCGTGTGAAACCTTCCTTGCCAGCTG
GTTACTACGGGAATGCTTTCGTACTTGGATGTGCTCAAACTTCCGTAAAGGACTTGACCGAAA
AGGGGTTAGGTTATTGTGCTGATCTGGTACGCGGTGCTAAAGAACGCGTTGGCGATGAGTACG
CGCGTGAGGTTGTAGAGTCGGTGTCATGGCCCCGCCGTGCCTCTCCTGACTCTGTAGGCGTTT
TGATCATCAGTCAGTGGTCTCGTTTAGGATTAGACCGCGTCGATTTCGGGTTAGGCCGTCCGG
TACAAGTGGGGCCTATCTGCTGTGACCGCTATTGCTTGTTCCTGCCCGTACGTGAGAGTACGG
AGTCAGTTAAGGTGATGGTAGCAGTCCCTACAAGTGCGGTAGATCGTTACGAGTATTTTATTC
GTTCACCGTATTCCTAACTTAAGGAGACG

>CAAT1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCGCGGGAGGAGAGTTTA
AAGTAACTGTATCGAAGAAGGAGGTAGTGGCCGCTGTACTGCCGATGCAGGAACACTGGCTTC
CCTTGTCAAATCTTGACTTGATTCTGCCGCCGGTCGACGTTGGGGTCTTCTTCTGTTATAAGA
AACCAATTAGCCCCAGCACCGGAGAGCCTCTTACATTTGCAGCCATGGTGAACGCGCTGAAGA
AGGCGTTAGCGCAAGCTCTGGTAAGTTTCTACGCATTTGGGGGGGAAGTTGTATTGAACACTG
TGGGAGAACCCGAATTGTTGTGCAACAACCGTGGGGCTGATTTTGTGGAGGCTTATGCTGACA
TTGAGCTTGAAAACCTTAACCTGTACAATCCTGACGAAAGTGTCGAAGGTAAACTGGTCCCTA
AAAAAAAACACGGCGTTTTCGCTGTCCAAGCCACTCAACTTAAATGCGGGGGCCTTGTGGTGG
CATGTACGTTTGACCACCGTATCGCAGATGCGTATTCAACCAATATGTTTCTGGTTTCCTGGG
CGGAGATCGCGCAGTCCAAGCCGTCATATTATCGTCCGAGTTTCCGTCGTTCACTTCTTAATC
CCCGCCGCCCAGGTTTCATTGATCCTGCGTTAGATGACATGTATGTCCCGATTAATACACTTC
CGCCGCCCAAGGAACCAGAGCAGACAGACGACCATTTAATTAGCCGTATCTACTATGTAACAA
GCGAACAATTAAACTTATTGCAGGGTCTTGCGTCCAGCTCTGGCTGTAAGCGCACAAAACTGG
AGAGTTTCAGCGCCTATCTTTGGAAGATGGTTGCTAAATATGCGTCCAAAAATTATCCGCGTA
ACGTCATTACCAAAATGGGGATCGTCGTCGATGGACGTACTCGCCTTGGTGACGGCGACGAAA
AGAAAGCAAAAATGATGAGCAAGTACTTTGGAAACGTGCTGTCCATTCCCTTTGGAGGCAAAA
AGATGCATGAATTAACTGAGAAGCCACTGAGCTGGGTAGCCGATGAGGTGCATGACTTCCTGG
ACGTTGCTGTTACTAAAGAGCACTTCCTTGGATTGATCGACTGGGTTGAGGCACATCGCCCCG
TGCCGGCCCTTGCCAAGATCTACTGTCAAGGGACGGAGGATGGTCCTGGCTTCGTGGTGTCGA
GCGGTCAACGTTTCCCAGTAGGTAAAATCGACTTTGGTTGGGGTAAACCTGGTTTCGGGTCCT
ACCATTTCCCTTGGGGAGGCGATAGTGGGTATGTTATGCCGATGCCCTCGCCTTCGGGGAACG
GTGACTGGATTGTCTACATGCACTTGAGTAAGGGCCAGTTAGAATTGATTGAGACAGAGGCGG
CTGCTGTCTTCCGTCCATTAACGAGCGACTATCTTAATTTTATGTAACTTAAGGAGACG

>CER2

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGAGGGTTCGCCTGTCACCT
CTGTGCGTCTGTCGTCGGTAGTGCCTGCATCGGTTGTGGGTGAGAATAAACCTCGTCAGTTGA
CGCCTATGGATTTGGCCATGAAACTTCACTACGTCCGCGCGGTTTACTTTTTTAAGGGCGCAC
GCGATTTCACAGTGGCGGATGTGAAGAATACGATGTTCACGTTGCAGTCCTTGCTTCAATCCT
ATCACCATGTCTCCGGTCGTATTCGCATGTCAGATAATGATAACGACACGAGTGCCGCTGCGA
TCCCCTATATTCGCTGTAATGACTCAGGAATCCGCGTCGTAGAAGCCAACGTGGAGGAGTTTA
CAGTCGAGAAATGGTTAGAATTGGATGACCGCTCCATCGATCACCGTTTCTTAGTATACGATC
ACGTATTGGGGCCGGATCTTACCTTTAGTCCATTAGTATTTCTGCAAATCACACAGTTTAAAT
GTGGAGGTTTATGTATCGGATTATCATGGGCTCACATCCTGGGGGATGTCTTTTCAGCATCAA
CGTTTATGAAGACACTGGGACAGCTTGTCAGTGGACATGCTCCGACCAAACCTGTCTATCCCA

AGACACCCGAGCTGACCTCACATGCTCGCAATGACGGGGAGGCCATCAGTATTGAAAAGATCG
ACTCCGTGGGGGAGTATTGGCTGCTGACCAATAAATGTAAGATGGGGCGCCACATTTTTAACT
TTTCGCTTAACCATATTGACAGTCTTATGGCGAAATACACCACCCGTGACCAGCCATTCTCCG
AGGTGGACATTCTGTATGCTCTTATCTGGAAAAGTTTGTTAAATATTCGTGGTGAAACAAATA
CGAACGTCATCACAATTTGCGATCGCAAAAAGTCTTCGACGTGCTGGAATGAAGACCTTGTCA
TTAGTGTGGTAGAAAAGAACGATGAGATGGTTGGAATTTCGGAATTAGCCGCCCTTATCGCAG
GTGAAAAGCGTGAAGAAAATGGTGCCATTAAACGCATGATTGAACAAGATAAGGGTAGCTCCG
ACTTTTTTACCTACGGCGCTAATCTGACTTTCGTTAACCTTGATGAAATCGACATGTACGAGT
TAGAGATCAATGGTGGCAAGCCAGATTTCGTTAATTACACAATTCATGGAGTTGGGGACAAGG
GCGTAGTTCTTGTATTTCCGAAACAGAATTTCGCCCGTATTGTCTCGGTCGTCATGCCGGAAG
AGGACTTGGCTAAGCTGAAAGAGGAAGTTACAAATATGATTATCTAACTTAAGGAGACG

>CHAT

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGACCATCAGGTTAGTTTGC
CCCAGTCCACCACTACTGGGCTTAGTTTCAAAGTGCACCGCCAGCAACGTGAATTAGTAACCC
CGGCGAAGCCAACCCCTCGTGAGCTGAAACCATTATCGGACATTGATGACCAACAGGGCCTTC
GTTTTCAGATCCCTGTAATCTTTTTCTATCGCCCCAATTTGTCGTCAGATTTGGATCCCGTTC
AGGTTATCAAGAAGGCTTTAGCAGATGCCCTGGTATATTATTATCCGTTCGCAGGTCGTTTGC
GTGAATTAAGTAATCGTAAGTTGGCCGTGGACTGCACTGGTGAGGGAGTGTTGTTTATTGAGG
CAGAGGCGGACGTTGCGCTGGCGGAGCTTGAGGAAGCTGACGCCTTATTACCGCCCTTCCCCT
TCCTGGAAGAGTTGCTGTTCGACGTAGAAGGTAGTTCCGATGTGTTAAATACCCCATTGTTGT
TGGTTCAGGTGACACGTCTGAAATGTTGCGGCTTCATCTTCGCTCTGCGCTTCAACCACACTA
TGACAGATGGTGCTGGATTATCCTTGTTTCTGAAGTCGCTTTGTGAATTGGCTTGTGGTTTAC
ATGCACCTTCTGTGCCTCCTGTCTGGAACCGTCATCTTCTTACGGTAAGTGCGTCCGAGGCCC
GCGTTACACACACGCACCGTGAGTACGATGATCAGGTTGGGATTGATGTCGTCGCCACCGGTC
ATCCATTAGTCTCACGCTCATTTTTTTTCCGTGCAGAGGAGATTAGTGCAATCCGCAAATTGC
TTCCTCCTGATTTGCATAACACTTCCTTTGAGGCTCTTTCCTCTTTTTTATGGCGTTGTCGTA
CGATTGCATTAAACCCCGACCCTAATACAGAGATGCGTTTAACGTGCATCATCAACAGCCGTA
GCAAACTGCGTAATCCGCCTCTGGAACCCGGATACTATGGTAATGTCTTCGTTATTCCTGCCG
CTATCGCAACCGCCCGCGATCTGATCGAAAAGCCACTTGAATTTGCATTGCGCCTTATCCAGG
AGACCAAATCCTCAGTAACGGAGGACTATGTACGTTCGGTGACAGCCTTGATGGCAACCCGTG
GGCGTCCCATGTTTGTGGCGTCCGGCAACTATATCATCTCAGACTTACGCCACTTCGACTTAG
GAAAGATCGACTTCGGCCCGTGGGGGAAGCCTGTTTATGGAGGTACGGCAAAAGCTGGTATCG
CTTTATTCCCCGGTGTATCGTTTTACGTCCCGTTTAAAAACAAGAAGGGTGAGACCGGCACAG
TGGTTGCTATTTCATTGCCGGTACGCGCAATGGAAACTTTCGTTGCCGAGCTGAACGGGGTAT
TAAATGTTTCGAAGGGATAACTTAAGGAGACG

>CmAAT1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGATTTCAGCTTTCATGTTC
GTAAATGTCAGCCGGAACTGATTGCACCGGCAAATCCGACACCGTATGAATTTAAACAGCTGA
GTGATGTTGATGATCAGCAGAGCCTGCGTCTGCAGCTGCCGTTTGTTAACATTTATCCGCATA
ATCCGAGCCTGGAAGGTCGTGATCCGGTTAAAGTTATTAAAGAAGCAATTGGTAAAGCCCTGG
TGTTCTATTATCCGCTGGCAGGTCGTCTGCGTGAAGGTCCGGGTCGTAAACTGTTTGTTGAAT
GTACCGGTGAAGGCATCCTGTTTATTGAAGCAGATGCAGATGTTTCCCTGGAAGAATTTTGGG
ATACCCTGCCGTATAGCCTGAGCAGCATGCAGAATAACATTATTCATAATGCCCTGAACTCCG
ATGAAGTTCTGAATAGTCCGCTGCTGCTGATTCAGGTTACCCGTCTGAAATGTGGTGGTTTTA
TCTTTGGTCTGTGCTTTAATCATACCATGGCAGATGGTTTTGGCATCGTGCAGTTTATGAAAG
CAACCGCAGAAATTGCACGTGGTGCATTTGCACCGAGCATTCTGCCGGTTTGGCAGCGTGCCC
TGCTGACCGCACGTGATCCGCCTCGTATTACCTTTCGTCACTATGAATATGACCAGGTGGTTG
ATATGAAAAGCGGTCTGATTCCGGTGAACAGCAAAATTGATCAGCTGTTCTTTTTTAGCCAGC
TGCAAATTAGCACCCTGCGTCAGACCCTGCCTGCGCATCTGCATGATTGTCCGAGCTTTGAAG
TGCTGACAGCATATGTTTGGCGTCTGCGTACCATTGCACTGCAGTTTAAACCGGAAGAAGAGG
TTCGTTTTCTTTGCGTTATGAATCTGCGTAGCAAAATCGATATTCCGCTGGGTTATTATGGTA
ATGCAGTTGTTGTTCCGGCAGTTATTACCACCGCAGCAAAACTGTGTGGTAATCCGCTGGGCT
ATGCAGTTGATCTGATTCGTAAAGCAAAAGCCAAAGCAACCATGGAATATATCAAAAGCACCG
TGGATCTGATGGTGATTAAAGGTCGTCCGTATTTTACCGTTGTTGGTAGCTTCATGATGAGCG
ATCTGACCCGTATTGGTGTTGAAAATGTTGATTTTGGTTGGGGCAAAGCCATTTTTGGCGGTC

CGACCACCACCGGTGCACGTATTACCCGTGGTCTGGTTAGCTTTTGTGTTCCGTTTATGAATC
GCAATGGTGAAAAAGGCACCGCACTGAGCCTGTGCCTGCCTCCGCCTGCAATGGAACGTTTTC
GTGCAAATGTTCATGCAAGCCTGCAGGTTAAACAGGTTGTTGATGCCGTTGATAGCCATATGC
AGACCATTCAGAGCGCAAGCAAATAACTTAAGGAGACG

>CmAAT4

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGAGGTGAAAGTACTGAGCA
AGGAGACTATTATTCCGAGTTCCCCAACGCCGCCCCACTTACAGCCACTGAATCTTTCACTGT
TAGACCAACTTTCTCCAATGCTGTATATCCCTCTGCTGCTGTTCTACCCTATGAAAAAGAGCT
ACCAGCATCAAGACCATAATAAAGCGATTGCTACACTTAAAACCTCTCTGTCAAAGACACTTT
CGCGCTTCTACTTGCTGGCAGGGCGTATTATCGGGAAGTCTATTCACTGTAATGATAAAGGGG
CCGTTTTTATGGAAGCTACGATTAACTCGAACATGTTTGATATCTTAAAAGAACCCAATAATG
AAGTGCTTACAAAGTTGCTTCCCTGCTCACTGTTATGTAACACCAAACCGATTGAAGAATACC
CCCAGATCGTGGTCCAGGCAAATATCTTCGAGTGCGGTGGCATTGCCATCAGCCTGTGTTTAC
TTCACAAATTAATTGATGCTGCTACCTTCTGCTGTTTCTTACGTTCGTGGGCCACGACCAACC
GTGAGTTGCTTTCTCTTGACCATTCTTCGCCTAACAACAATATGGTCTGTGTTGATTACAAGA
GTTTCTCGTCTTTATTCCCACAGACTAATCTTTTACCGTTTCACCAGTCATTGATTAACAATG
ATAAGGCGGTTGTCCCACCTTCTAGCATCTTTAACCGCAAGCGCCGCTTTCAGCGCTTTGTAT
TTCGCAGTGAGGCGATTTTGGATTTAAAGGCAAAGGCGAAATCTTGCGATATCCCTAACCCTA
CGTGCGTTGAGACATTAACGTGCTTTATTTGGAAATACTTAATGAAAGTTGCCGATGACGGAG
ACAGCCAACGTCCTTCCACTTTATCTCATGTAGTGAACATCCGCAAAATGCTGGAGCCGAGCT
TAGGTGAAGTCAGCCTTGGTAATATTATGTGGGGTACAGTCGCTCACCATTTCAGTACAACAC
GCAACGAGGAATTTGAGGGGTTAGAACTGTCAAAGTTAGTTAGTTTGTTACGCCAGAGCTTCA
AGAAGATCAATAAAGACTATATTAAGGAGCTGATTATGGGGGGTGACAAGGAACGCCGCAACG
GGGTTATGAAGTTAGTGGGTGAAATCAACAAGTGGCCGATCTCGAATTATTATTTCTTCACAT
CGTGGAAAAACCTGAAGCTGAATGAGGTCGACTTCGGTTGGGGAAAACCCCTTTGGTCCGCGA
TTGCGGGAGATCCGAACGAAATGATGGGTAACATCATTGTCTTGGTTGATAACGTCCTGGATG
ACGGAAGTACCGAGGCGTGGATCTTACTGGATGAGAAGGAGATGCAACTTCTGGAACAGATTC
CACAGTTCCTTGAATTTGCTTTGTTAAATCCCTCTATCAATTTGCCGCATAACCAGAAAACAG
CTGATGAAATTTTTTCTAACAAGTTGATTTAACTTAAGGAGACG

>DAT

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGAATCTGGGAAAATTTCAG
TAGAGACTGAAACTTTGTCAAAAACGTTGATCAAGCCTTCCTCGCCCACCCCACAGTCCCTTT
CACGTTACAATCTGAGCTATAACGACCAAAACATTTATCAGACTTGCGTGTCGGTCGGGTTCT
TTTACGAAAACCCAGATGGTATTGAGATTTCGACTATCCGCGAACAATTGCAAAACAGCTTGT
CAAAGACATTAGTTTCATACTATCCATTTGCCGGCAAAGTGGTCAAAAATGATTACATTCATT
GCAATGACGACGGTATCGAATTTGTGGAGGTGCGTATCCGTTGCCGTATGAACGACATCCTTA
AATACGAGTTACGCTCATACGCACGTGATTTGGTATTGCCGAAACGTGTCACAGTTGGCTCCG
AAGACACCACCGCCATTGTCCAACTGAGCCACTTTGATTGCGGGGGGTTAGCGGTCGCATTCG
GTATTAGTCACAAAGTCGCGGACGGCGGAACGATCGCTAGTTTCATGAAAGACTGGGCGGCGA
GTGCGTGTTACCTGTCCAGCTCCCATCACGTACCAACCCCGTTATTGGTGTCGGACTCAATCT
TTCCACGTCAGGACAATATTATTTGTGAACAGTTCCCAACTAGCAAGAACTGCGTCGAGAAGA
CATTCATTTTTCCACCAGAAGCGATTGAGAAGTTGAAGTCAAAGGCGGTTGAGTTTGGCATCG
AAAAGCCAACACGTGTTGAAGTCCTGACCGCTTTTCTTTCGCGTTGCGCCACAGTGGCAGGCA
AGTCAGCAGCTAAGAACAACAATTGTGGACAGTCCCTGCCATTTCCCGTTCTTCAGGCGATTA
ACTTGCGTCCGATCCTGGAGCTTCCGCAAAACAGTGTCGGTAACCTTGTATCTATTTACTTCT
CGCGTACGATCAAGGAGAATGATTATCTGAACGAAAAGGAGTACACAAAGTTAGTGATCAATG
AGCTGCGCAAAGAGAAACAAAAGATTAAAAATTTAAGTCGTGAAAAGTTGACCTATGTGGCAC
AGATGGAAGAATTTGTCAAATCGCTGAAAGAATTTGACATCAGTAATTTTCTGGACATCGATG
CATACCTGAGCGATTCCTGGTGCCGCTTTCCATTTTATGACGTTGATTTTGGGTGGGGTAAGC
CGATCTGGGTATGTCTGTTCCAACCTTACATCAAAAACTGCGTCGTAATGATGGACTACCCCT
TTGGCGACGACTACGGGATTGAGGCGATTGTGTCATTCGAGCAAGAAAAAATGTCTGCGTTCG
AGAAAAATGAGCAATTACTTCAATTTGTGTCTAATTAACTTAAGGAGACG

>DBBT

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCCGCTTTAACGTGGATA
TGATTGAGCGCGTGATTGTGGCACCGTGCCTGCAATCACCCAAAAATATCTTACATTTGTCAC
CCATCGACAATAAAACACGCGGTTTGACTAATATCTTGTCTGTGTATAACGCCTCCCAACGTG
TCAGCGTTTCTGCAGACCCTGCGAAAACAATCCGTGAGGCTCTGTCTAAAGTATTAGTCTATT
ACCCCCCGTTTGCAGGCCGCTTACGTAATACGGAGAATGGTGACTTGGAAGTCGAATGCACCG
GCGAGGGCGCTGTATTCGTAGAGGCTATGGCAGATAATGATTTATCGGTTCTGCAAGATTTCA
ATGAGTACGACCCGTCATTCCAGCAGCTGGTTTTCAATTTACGCGAGGACGTAAACATTGAGG
ACCTGCACCTTCTGACGGTTCAGGTAACTCGCTTCACGTGCGGAGGATTTGTCGTTGGTACTC
GTTTCCATCATAGTGTCTCGGATGGGAAAGGTATCGGTCAGCTTTTGAAAGGCATGGGTGAGA
TGGCACGTGGAGAGTTCAAACCTTCATTAGAGCCCATCTGGAATCGTGAAATGGTCAAACCGG
AAGACATCATGTATCTGCAATTCGATCACTTTGATTTCATCCACCCTCCACTTAATTTGGAGA
AATCTATCCAAGCCAGTATGGTGATCAGTTTCGAACGTATTAACTACATCAAGCGTTGTATGA
TGGAGGAGTGTAAAGAGTTCTTTAGTGCATTCGAAGTGGTCGTAGCCTTAATCTGGCTTGCAC
GCACCAAGTCCTTTCGTATCCCTCCGAACGAATATGTCAAAATCATTTTCCCTATCGATATGC
GCAATTCTTTCGATTCACCTTTGCCTAAGGGCTACTACGGGAATGCAATTGGAAACGCTTGTG
CCATGGACAACGTCAAGGACTTATTAAACGGCTCATTGTTATATGCGCTGATGCTGATCAAGA
AGTCAAAATTTGCACTGAACGAAAATTTCAAATCACGCATCTTGACCAAACCCTCTACGTTAG
ACGCCAATATGAAACACGAAAATGTAGTAGGCTGCGGGGACTGGCGCAATCTTGGATTCTACG
AGGCTGACTTCGGCTGGGGTAATGCAGTGAACGTGAGTCCCATGCAGCAGCAACGTGAGCACG
AGCTTGCGATGCAAAATTACTTCTTGTTCTTACGTTCAGCCAAAAACATGATCGACGGCATCA
AAATCTTGATGTTTATGCCAGCGTCCATGGTAAAGCCGTTCAAAATCGAGATGGAGGTCACCA
TTAACAAATATGTGGCAAAGATTTGTAATTCCAAGCTTTAACTTAAGGAGACG

>DkAAT1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGATGCCACTTTCAGTGCTGC
AGGTCAAACGTCTGCAGCCGGAGTTCATTACACCGGCGAAACCTATGCCCCAAGAGACAAAGT
TTTTATCAGATATCGACGATCAAGAGGGGTTACGTTTTCACGTCCCAGTTATTATGCGTTATA
AGGATAACCCCTCTCTGAATGAGAATCGCAATCCGGTGAAAGTTATCCGCGAAGCCGTATCGC
GCGCCTTAGTGTACTACTACCCGTTAGCCGGCCGCTTACGCGAATGGCCTAACCGTAAGCTTG
TGGTCGACTGTAATTCGGAAGGAATCCTTTTTGTAGAAGCTTTCGCTAATGTAACTCTGGAAC
AGCTTGGTGACAAGATTCTGCCACCCTGCCCCCTGTTAGAAGGATTCTTATTCAACGTCCCAG
GAAGTGACGGAATTATTGGCTGCCCCTTACTTTTGGTCCAAGTAACGTGCCTGACCTGCGGCG
GGTTTATTTTGGCTCTGCGCTTAAACCACACGATGTGTGACGCCCCAGGCTTGGTGCTGTTTC
TGACCGCTATCGCAGAAATGGCGAGCGGGGCGCACGCCCCAAGCATTCTTCCGGTGTGGGAAC
GCGAATTGCTGTTCGCGCGTGACCCCCCACGCATCACATGTGCCCACCACGAATACGAGGATG
TGATTGATCACTCAGATGGCAGCTATGCCTCCAGTAATCAAAGTAACATGGTGCAGCGCAGCT
TTTATTTCGGGGCAAAGGAGATGCGTGTATTACGCAAACAGATTCCTCCTCATCTTATCTCTA
CTTGCAGCACGTTCGATCTTATTCCTGCGTGTTTGTGGAAATGTCATACCCTTGCATTAAAAA
TTAACCCAAAACAAGTTGTGCAAGTGTCTTTTACTGTTAATGCGCGTGGTAAGCATCACAATG
TTCGTTTGCCACTTGGCTACTATGGTAATGCCTTTGCATTTCCAGCAGCTGTGTCTAAGGCCG
AACCTTTATGTAAGAACCCATTAGGATATGCGTTGGAGCTTGTCAAGAAGGCTAAAGCCACAA
TGAACGAGGAATACTTACGTTCAGTCGCAGACCTGTTAGTGTTACGCGGACGCCCCCAGTACC
TTTATACCGGATCCTATTTCATCGTAAGCGACAACACGCGCGCCGGATTTGGGGATGTAAACT
TCGGATGGGGTCAGCCAGTTTTCGCCGGGCCTGCGAAAGCTGTCGACCTTTTCCGCTTCTACG
TACAACATAAGAATAATACAGAGGATGGTATCCTGGTTCCCATGTGTCTGCCAAGTAGTGCTA
TGGAACGCTTTCAGCAAGAGCTTGAGCGTATCACGCAGGAGCCGAAAGAGGATATCTGCAATA
ACTTACGTTCCACACGTATCATGTCCATGATGTAACTTAAGGAGACG

>Eeb1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGTTTCGTTCTGGTTATTACC
CAACAGTAACACCTTCGCACTGGGGATACAACGGTACGGTTAAGCACGTACTGGGTGAGAAAG
GCACCAAGTCGTTGGCCTTCCGTGACTCCAAGCGCCAGATCCCGTTGCATGAGTTTGTGACCA
AACACGTTCCAACATTGAAGGACGGCGCTAACTTTCGTTTGAATAGCTTGCTTTTCACGGGTT
ATTTGCAGACCCTGTACTTGAGCGCAGGTGATTTCTCCAAGAAGTTTCAAGTATTCTACGGTC
GTGAAATCATTAAGTTTAGTGATGGAGGGGTTTGTACAGCGGATTGGGTCATGCCGGAATGGG
AACAGACATACTCTTTGAACGCAGAAAAAGCGAGTTTCAATGAAAAGCAATTCTCAAATGACG
AAAAAGCTACACATCCCAAAGGTTGGCCGCGTTTACACCCGCGTACACGCTACTTGAGTTCTG

AGGAGCTGGAAAAGTGTCACTCAAAGGGCTACTCGTATCCGCTGGTGGTCGTCCTTCACGGTC
TGGCTGGAGGCAGCCATGAGCCCTTAATCCGTGCCTTGTCAGAAGACTTGTCGAAAGTAGGCG
ATGGCAAATTTCAAGTGGTCGTACTTAACGCTCGCGGATGCTCCCGTTCGAAGGTGACAACGC
GCCGCATCTTCACGGCTCTGCATACCGGTGACGTCCGCGAGTTCTTGAATCATCAGAAAGCGT
TGTTCCCGCAGCGCAAGATTTATGCCGTCGGCACCAGTTTCGGGGCCGCCATGCTGACAAACT
ACCTTGGTGAAGAAGGTGATAACTGCCCTCTTAATGCCGCAGTCGCATTAAGCAATCCTTGGG
ACTTCGTTCACACTTGGGATAAATTGGCTCATGATTGGTGGAGTAATCATATCTTTTCGCGCA
CCTTAACGCAGTTTTTAACCCGTACAGTTAAAGTTAACATGAACGAGCTTCAAGTTCCCGAGA
ATTTTGAGGTGAGTCACAAACCCACGGTTGAAAAACCGGTCTTTTATACCTATACACGCGAGA
ATTTGGAGAAAGCCGAGAAATTCACCGATATTTTAGAATTTGACAATCTGTTCACTGCGCCAT
CAATGGGGTTACCGGATGGGTTGACATATTACCGTAAGGCTTCTTCCATTAACCGTTTGCCGA
ACATTAAAATCCCAACCTTGATCATCAATGCGACCGATGATCCAGTAACTGGCGAGAATGTGA
TTCCATACAAGCAGGCCCGCGAGAATCCTTGCGTTTTGCTGTGTGAGACTGACCTGGGCGGCC
ACTTGGCCTACCTTGACAACGAAAGTAACAGCTGGCTTACAAAGCAAGCGGCGGAGTTCCTTG
GGTCGTTCGACGAGTTAGTACTGTAACTTAAGGAGACG

>Eht1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGTCCGAGGTCTCTAAATGGC
CAGCCATCAATCCGTTTCACTGGGGATATAACGGTACGGTATCGCATATCGTAGGAGAGAATG
GTTCGATTAAACTTCATCTGAAAGATAATAAAGAACAAGTTGACTTCGATGAATTTGCAAACA
AATACGTACCAACTTTGAAAAATGGAGCACAGTTCAAGTTATCGCCATACCTGTTTACAGGGA
TTTTGCAGACACTTTATTTGGGCGCTGCTGATTTTTCAAAAAAATTTCCTGTTTTCTACGGTC
GTGAAATTGTAAAGTTTTCAGATGGTGGGGTTTGTACGGCTGATTGGCTTATCGATAGTTGGA
AGAAGGACTACGAATTTGACCAATCGACAACGTCCTTTGACAAAAAAAAGTTCGATAAAGATG
AAAAGGCTACCCATCCGGAAGGTTGGCCACGCTTACAACCCCGCACCCGCTACCTTAAAGACA
ACGAACTGGAAGAGTTACGTGAGGTAGACCTTCCATTGGTCGTAATTTTACATGGCCTTGCGG
GAGGTTCCCACGAGCCAATTATCCGTAGCTTGGCCGAAAACCTTTCTCGCAGCGGACGCTTCC
AAGTCGTCGTACTTAACACGCGCGGGTGTGCACGTTCGAAAATCACCACTCGCAATCTTTTCA
CCGCTTATCATACCATGGATATTCGCGAATTTTTGCAACGTGAGAAACAACGCCATCCCGACC
GTAAGCTGTACGCGGTAGGGTGTAGTTTCGGGGCCACTATGTTGGCCAATTACCTGGGCGAGG
AGGGTGACAAAAGTCCACTGTCCGCTGCTGCCACCCTGTGCAACCCGTGGGATTTACTGTTGA
GTGCGATCCGCATGAGCCAGGACTGGTGGTCACGTACCTTATTTAGCAAGAATATTGCTCAAT
TCCTGACTCGCACGGTCCAGGTGAATATGGGAGAACTTGGGGTGCCTAATGGAAGTCTGCCGG
ACCATCCCCCTACCGTAAAAAACCCATCATTTTATATGTTCACCCCGGAAAATTTAATCAAGG
CCAAGTCTTTCAAATCAACCCGTGAGTTCGACGAGGTCTACACTGCTCCCGCGTTGGGCTTTC
CCAACGCAATGGAGTATTACAAAGCAGCTTCGTCGATTAATCGTGTGGACACCATTCGCGTTC
CAACTTTAGTCATTAACAGTCGCGATGATCCAGTAGTCGGTCCAGATCAACCGTATAGCATCG
TTGAGAAAAACCCGCGCATCTTGTACTGCCGCACAGACTTAGGAGGCCACTTAGCGTATCTGG
ATAAGGACAATAATAGCTGGGCGACAAAAGCGATTGCAGAATTTTTTACGAAATTCGACGAGT
TAGTTGTTTAACTTAAGGAGACG

>EjAAT1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGATGCCGTTATCGGTCTTGC
AAGTCAAGCGTTTGCAGCCAGAATTTATCACGCCAGCAAAGCCCATGCCCCAAGAGACTAAAT
TCCTTTCCGATATTGATGACCAAGAAGGGCTGCGCTTCCATGTTCCTGTCATCATGTGCTACA
AAGACAACCCCTCTTTGAATGAGAACCGTAATCCCGTGAAAGTAATTCGCGAAGCTGTATCAC
GCGCTCTGGTTTACTATTACCCGCTGGCTGGACGTCTTCGCGAATGGCCGAATCGTAAACTGG
TAGTAGATTGCAATAGCGAGGGTATTTTATTTGTTGAAGCCTTCGCTAATGTCACCCTGGAGC
AGCTTGGCGATAAAATCTTGCCTCCGTGCCCTCTTTTAGAAGAGTTCCTTTTTAACGTGCCCG
GGTCAGACGGCATCATTGGTTGTCCACTTTTGCTTGTTCAAGTAACCTGTTTAACTTGTGGGG
GCTTTATCTTGGCACTGCGCCTGAATCATACTATGTGTGATGCCCCAGGCTTAGTCTTATTCT
TGACGGCAATCGCGGAAATGGCATCCGGCGCTCATGCGCCCTCAATCTTGCCTGTATGGGAGC
GTGAGCTTCTTTTTGCACGCGACCCACCCCGTATCACGTGCGCACATCACGAATACGAAGATG
TAATTGACCATTCCGACGGCTCCTACGCGTCTAGCAATCAATCTAATATGGTACAACGCTCGT
TTTATTTCGGAGCAAAAGAGATGCGTGTACTGCGCAAACAAATCCCGCCCCACCTTATTAGTA
CCTGTAGCACTTTTGACCTTATCACTGCATGTTTGTGGAAGTGTCACACGCTGGCGTTAAAAA
TCAACCCAAAGCAGGTAGTGCAAGTGTCGTTCACCGTCAATGCCCGTGGCAAACACCACAACG

TGCGCCTGCCATTAGGCTACTATGGTAACGCTTTTGCCTTCCCAGCCGCAGTAAGTAAGGCTG
AGCCTCTGTGTAAGAATCCTCTGGGTTACGCCTTGGAATTAGTAAAGAAAGCGAAGGCCACGA
TGAATGAAGAATATTTACGTTCAGTTGCCGATCTTCTGGTCCTTCGCGGCCGTCCTCAGTATC
TTTACACGGGTAGCTATTTCATCGTGTCCGACAATACTCGTGCCGGGTTTGGTGATGTTAATT
TTGGATGGGGACAGCCCGTATTCGCGGGTCCTGCTAAAGCGGTTGACTTGTTTCGCTTTTACG
TACAGCATAAAAACAACACTGAAGACGGTATTCTTGTTCCAATGTGCCTTCCGTCGAGTGCCA
TGGAACGCTTCCAGCAAGAATTAGAGCGCATCACACAGGAACCGAAGGAAGACATCTGCAACA
ATCTGCGCAGCACACGCATTATGTCTATGATGTAACTTAAGGAGACG

>FcAAT1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGAGAAAATTGAGGTTTCTA
TTATTTCTAAGTATACAATTAAACCTTCATCTAGTTTATTGCAGCCTTACAAATTAAGCTTGT
TAGACCAACTGACCCCACCGGCGTATGTACCGATGGTTTTCTTTTACCCTATCACAGAGCATG
TATTTAACTTACCCCAAACACTGGCCGACTTACGCCAGAGCCTTTCCGAGACCCTGGCGCTGT
ACTACCCCTTGTCCGGGCGCGTCAAAAATAACCTGTATATTGATAACTTCGAAGAGGGTGTAC
CGTACTTGGAGGCCCAAGTGAATTGTGACATGACAGACTTTTTGCGTTTGGGGAAGATCGAGT
GCTTAAACGAGTTCGTCAGTATTAAGCCCTTTTCTATGGAGGCTATTAGTGACGAACGTTACC
CACTGCTTGGCGTTCAGGTTAATGTGTTTGATTCAGGTATTGCAATTGGGGTTAGCTTGTCAC
ACAAGTTAATTGACGGCCGTACGGCGTATTGTTTCTTGAAATCGTGGGGGGCCGTGTTTCGTG
GATGCCGCGAAGACGTCATTCACCCATCGCTGTCGGAAGCAGCCTTACTGTTCCCTCCACGCG
ATGATTTGCCAGAGAAGTATGCGGATCAGATGGAAGGGTTATGGTTTGCTGGGAAAAAGGTTG
CAACCCGTCGTTTTGTATTCGGGGCAAAGGCTATCAGCAGTATTCAAGACGAAGCGAAGTCTG
AATCGGTGCCCAAGCCTTCCCGCGTGCAAGCTGTAACAGGTTTTCTGTGGAAGCACTTAATTG
CAGCGTCGCGTGCTCTTACCAGTGGCACTACGTCCACTCGCCTTTCAATCGCAGCACAGGCAG
TTAATCTGCGTACCCGCATGAATATGGAAACGGTCTTGGACAATGCTACCGGGAACCTTATCT
GGTGGGCCCAAGCCATTTTAGAGCTGAGCCATACGACGCCGGAAATTTCCGACTTGAAACTTT
GCGATCTGGTTAACCTGCTTAATGGATCTGTGAAACAGTGCAATGGAGACTACTTCGAGACCT
TCAAAGGGAAAGAAGGGTACGGACGCATGTGTGAGTACTTGGACTTTCAACGCACCATGTCCT
CTATGGAACCGGCCCCAGATATCTACCTGTTCTCGTCCTGGACCAACTTTTTCAACCCGCTTG
ATTTCGGCTGGGGTCGTACCAGCTGGATTGGAGTGGCGGGGAAGATCGAATCCGCGTCGTGCA
AATTCATCATTTTGGTCCCAACCCAGTGCGGATCAGGCATTGAGGCGTGGGTAAATTTAGAAG
AGGAGAAAATGGCTATGCTGGAGCAAGACCCGCACTTCCTTGCTTTAGCGAGCCCGAAAACAC
TTATTTAACTTAAGGAGACG

>LaAAT1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGAAAATCGAAATTAAGGAGT
CGACTATGGTACGCCCAGCGGCAGAAACGCCAAGTGGCAGTTTATGGCTGAGTAACTTGGACT
TACTTTCCCCGGCCAACTACCATACGCTTAGCGTGCATTTTTACTCCCATGATGGCTCCGCCA
ACTTTTTTGATGCAACGGCTTTGAAGGAGGCACTGTCCCGTGCACTGGTCGACTTCTATCCGT
ATGCAGGTCGCCTTAAACTTAATAAGGAAAACCGTTTGGAAATCGAGTGTAATGGGGAGGGAA
TCTTATTGGTAGAAGCCGAGTGTTCAGGAGCATTGGATGAATTGGGAGATTTCACCCCGCGTC
CAGAACTGAATTTAATCCCCAAAGTAGACTACTCTAAAGGCATGTCCACGTATCCCCTTATGT
TATTCCAAATTACTCGCTTCAAATGTGGCGGGGTCGCTTTAGGGGTTGCCAACGAGCACCACC
TGAGTGACGGGGTCGCGGCATTGCACTTTATTAACACCTGGGCTCATTACTCTCGCGGCGTAC
CTGCTCCCAGCCCACCACCTCACTTCGATCGCACAGCGCTTTCCGCTCGCAACCCGCCTCAAC
CCCAATTTTCGCACGCGGAGTACCAACCTCCGCCCACACTGGAGAATCCCCTTCCTGCGACCG
ACATCGCGCATAGCAAATTCAAGTTAACTCGCGCACAGTTGAACAGTCTTAAAGCAAAATGTG
CGGCGGGGGATTCGGACGGTCATACGAACGGGACCGCCAATGGGAAATCCGACGCAAACGGTA
CGGCAGACGGCAAATCAGACGCCAATGGGACTGCTAATGGAAAGTCAGCGGCTAAGCGTTATA
GCACGTTTGAGGTATTGGCGGGTCACATTTGGCGCTCGGTATGCACAGCTCGTGGTTTGCCAG
CAGAACAGGAAACTAAACTGCACATCCCCTTCGATGGGCGCTCGCGTTTGAACCTTCCACCGG
GGTATTTTGGTAACGCTATTTTCTTTGCGACGCCGATCGCCACCTGCGGTGAAATTGAGAGTA
ACAGTTTAAGCTATGCAGTGCGCCGCGTTGGAGATGGAATCGCGCGTCTTGATGAAGAGTATT
TAAAGTCATCGCTGGACTTTTTGGAACTGCAGCCAGACATCTCTAAATTAGCGCAGGGCGCCC
ACTCTTTCCGTTGCCCCAACCTGTGGGTGATTAGCTGGGTATGGCTTCCAATTTACGAACCAG
ATTTTGGCTGGGGCAAGGCCGTACATATGGGCCCATGGGCAGCGCCATTTGAGGGGAAGTCAT

ACCTGTTACCAAACCCCGAGAACGACGGCAGTCTGTTCGTGTCTATCACTCTGCACAAGCAGC
ATATGGAGCGTTTTCAGAAGCTGTTTTATGAGATTTAACTTAAGGAGACG

>MaAAT

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGCTCCATCACTTACTTTCA
CGGTTCGTCGCCAGAAGCCTGTACTGGTGGCACCTGCCGGATCCACTCCGCACGAATTtAAAC
GTTTGAGTGATATTGATGATCAAGACGGTTTACGCTTCCACATCCCGGTAATTCAGTTCTACC
GTAACGATCCTTCCATGGGGGGCCGTGACCAGGCTAAGGTCATCCGCGAGGCTCTTGCTCGTG
CCTTGGTTTTCTACTATCCGTTCGCGGGGCGCCTGCGCGAGGCCGCCGGGCGTAAATTGGTAG
TGGAGTGTACTGGCGAAGGGATCCTTTTTATTGAGGCTGATGCCGACGTTCGCTTGGAGCAGT
TTGGAGATGAGCTGCAACCACCATTCCCTTGTCTGGAAGAACTTGTTTATAATGTGCCGGGAT
CAGACGGAGTTTTGGACTGCCCATTGTTATTGATCCAGGTAACCCGCCTTCTTTGCGGTGGCT
TCATCTTCGCAATCCGCTTGAACCACACTATGTCTGACGCGCCCGGATTAGTCCAGTTTATGA
ACGCTGTAGCCGAGTTAGCACGTGGTGCTGCAGCACCATCAGTGCCCCCACTTTGGTCACGCG
AGATCCTGGAAGCACGTTCGCCCCCACGCGCCACATGCAAGCACCGTGAATACGACGACGTTC
CAGACACACGCGGCACTATCGTCCCATTGGACGATATGGTCCATCGTTCATTTTTCTTCGGCA
AGCGTGAAGTCGCAGCTTTACGCCGTCGCGTTCCACCCCATCTGCGCAACAGTTCGACGTTTG
AAATCTTGACAGCCTGTCTTTGGAAGTGTCGCACGATTGCTATTAGTCCTGACGCAGACGAGG
AAGTCCGCATGATCTGTATCGTTAACGCCCGCGGCAAAAGTGACTTGGGGCTGCCGGTTGGGT
ACTATGGGAACGCATTTGTATTCCCAGTGGCAGTATCAAAAGCGGGCAAATTGTGCGCTAATC
CTTTAGGTTATGCGTTGGATCTTGTCCGCAAAGCGAAGTCGGATGTCACGGATGAGTATGTCC
GTTCAGTCGCCGACTTAATGGTTTTACGTGGGCGTCCTCATTTCACTGTAGTTCGTAGTTACC
TGGTCTCGGACGTCACAAAGGCAGGCTTTGGTGACGTAGACTTCGGATGGGGAAAAGCCGCCT
ATGGAGGTCCAGCCAAGGGAGGCGTTGGAGCTATTCCCGGAGTAGCCAGTTTTTATATTCCCT
TTCGCAACGGAAAGGGAGAGGACGGCATCGTCGTGCCTGTGTGTCTTCCGGGCCCGGCGATGG
AGAAGTTCACGATGGAAATGGAATCGTTGATCGAAGAACCAGTCGCGGCTGAGCAGCACCATT
CGTTGACGTTAATCATGTCCCGCGTTTtaaCTTAAGGAGACG

>ManAAT

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGATTATCACGGTTAAAGAAT
CCACAATGGTCCCTCCCAGCGCCGAAACACCTCGCATTTCCCTGTGGAACTCTAACGCAGATT
TGGTCGTGCCTCGCTTCCACACACCTTCTGTATATTTCTATCGCCCTACAGGGGCCATCAATT
TCTTCGACGGAAAGCTTCTGAAGGAAGCACTGGGAAAAGCGCTGGTTCCTTTTTATCCTATGG
CAGGGCGTCTGAAACGTGATGAAGATGGGCGTATCGAAATCGATTGCAATGCCGAGGGAGTGC
TTTTTGTAGAGGCAGAAACACCCTCTGTTATTGATGATTTTGGAGATTTTGCACCTACACTTG
AACTTAAACAACTTATTCCGACGGTCGACTACAGTGGTGGAATCTCCACTTACCCACTGCTGG
CCTTACAAGTCACGCACTTTAAGTGCGGGGGTGTATCGTTGGGTGTCGGGATGCAACACCACG
CGGCCGATGGGTTCTCAGGGTTACACTTCGTCAATACTTGGTCAGATATTGCCCGTGGGTTAG
ATGTGAACATCACCCTGTTCATTGACCGTACATTGTTGCGCGCTCAGGACCCCCCTCAGCCTA
CGTTCCCACATACTTGGAATACCCGTCCACCCCCAAGCCTTAAAACACCGCCTCCTGCGGTGT
CGGAACCCACCGCTGTCTCAATCTTTAAGTTAACGCGCGACCAGTTGAATATTTTAAAGGCGA
AGGCGAAGGAGGACGGTAATACTATCAATTATTCGTCGTATGAAATGCTTGCTGGCCATGTGT
GGCGCAGTGCCTGTAAGGCTCGCGGTCTTTCCGATGATCAGGAAACAAAATTATATATCGCCA
CAGACGGTCGCGCTCGTCTTATCCCCCCACTGCCGCCCGGCTACTTTGGAAATGTCATCTTTA
CGGCAACCCCCATGGCTGTGGCGGGAGACCTGCAATCTAAGCCGATTTGGTATGCCGCGGGTC
AGATTCACGATGCGCTGGTTCGCATGGACAATGATTATTTACGCTCGGCCCTTGACTACCTTG
AACTGCAACCAGATTTATCCGCGTTAGTCCGCGGAGCGCACACGTTTCGTTGTCCAAATTTGG
GGATCACTAGcTGGGTTCGTCTGCCCATTCACGATGCAGATTTTGGGTGGGGCCCACCTACGT
TTATGGGTCCAGGCGGTATCGCTTACGAGGGCCTGAGCTTTGTGTTACCCTCACCGACCAACG
ACGGTTCTTTAAGTGTAGCCATTTCCCTTCAATCGGAACACATGAAACTGTTCCAAAAGTTTT
TTTATGATATTtaaCTTAAGGAGACG

>MbAAT

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGCTCCGAGTCTTACATTTA
CCGTACGTCGCCAGAAGCCTGTCCTTGTTGCTCCATCTGGCCCCACGCCGCATGAGTTCAAGC
GCCTGAGTGATATTGATGACCAAGATGGCCTTCGTTTTCACATCCCCGTAATTCAATTTTATC

GTAACGATCCCTCGATGGGAGGACGTGATCAGGCCAAAGTGATCCGCGAGGCGTTGGCACGCG
CACTGGTGTTCTACTACCCGTTCGCCGGACGCTTACGTGAAGCAGCAGGTCGCAAGCTGGTGG
TTGAGTGTACTGGAGAGGGAATCCTTTTTATCGAGGCTGATGCGGATGTGCGTCTTGAACAAT
TCGGGGACGAGCTTCAACCGCCATTTGCGTGTTTGGAGGAATTAGTTTATAACGTGCCTGGCT
CAGACGGTGTATTGGATTGTCCTTTATTACTTATCCAGGTAACACGTTTGTTGTGTGGAGGGT
TTATCCTGGCTATCCGTTTGAATCACACGATGTCCGACGCGCCGGGACTTGTTCAGTTCATGA
ATGCAGTAGCGGAGCTGGCACGTGGAGCCGCCGCTCCTAGCGTCCCTCCGGTCTGGGCTCGTG
AGATTCTGGAGGCCCGTTCTCATCCACGCGCCACCTGTAAGCATCGCGAGTATGACGATGTGC
CGGATACTCGCGGAACCATCATTCCCTTGGACGACATGGTTCACCGTTCATTTTTTTCGGTA
CACGCGAGGTAGCTGCCTTACGCCGCCGCGTCCCCCCACATCTTCGTAATTCTTCTACCTTCG
AaATCTTAACGGCTTGCTTGTGGAAATGTCGCACGATCGCCATCAGTCCTGACGCAGATGAAG
AGGTTCGCATGATCTGTATCGTGAATGCACGTGGAAAAAGCGATTTGGGTCTTCCGGTAGGTT
ATTACGGTAACGCTTTTGTATTCCCTGTGGCGGTAAGTAAAGCCGGCAAACTTTGTGCTAACC
CTTTAGGGTACGCCTTAGACCTGGTACGTAAAGCCAAGTCAGACGTAACTGATGAGTACGTGC
GTAGTGTTGCAGACCTTATGGTTCTGCGTGGACGTCCGCACTTTACTGTCGTTCGTTCGTACT
TGGTGTCCGACGTAACAAAGGCCGGGTTCGGGGATGTTGACTTCGGTTGGGGTAAGGCAGCAT
ACGGCGGACCTGCTAAGGGAGGTGTGGGAGCTATTCCTGGGGTTGCCAGCTTTTATATTCCGT
TTCGCAATCGTAAGGGGAAAGATGGTATTGTTGTCCCGGTCTGTCTGCCGGGCCCCGCGATGG
AGAAATTCACTATGGAGATGGAATCGTTAATTGAGGAGCCAGTGGCTGCCGAACAGCATCATA
GTGTCACTCTTGTGATGTCAAAATTCtaaCTTAAGGAGACG

>PcAAT

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGATGAGCCTGAGCGTTTTAC
AGGTCAAACGCTTGCAGCCAGAGTTAATCACACCGGCCAAGCCGACCCCGCAGGAAACTAAGT
TCCTGAGCGATATTGACGACCAAGAAGGGCTTCGCTTCCAATTACCTGTAATCATGTGTTACA
AAGACAACCCCTCTTTGAATAAAAACCGCAATCCAATCAAAGTCATTAAGGAAGCCTTGTCTC
GCGCTCTTGTCTACTACTACCCCCTGGCCGGTCGTCTGCGTGAAGGGCCAAACCGCAAATTAA
TGGTAAACTGTAACGGTGAGGGCATCTTATTCGTGGAGGCTAGCGCGGATGTGACTCTGGAGC
AATTGGGTGATAAaATCTTGCCTCCTTGTCCTTTGTTAGAAGAATTTCTTTTTAATTTTCCAG
GCAGCGACGGTATTATCGGGTGTCCATTATTACTTGTCCAGGTGACATGTTTGACTTGCGGAG
GATTTATTTTGGCCTTGCGCTTGAACCACACCATGTGCGACGCCACAGGATTACTGATGTTCT
TGACTGCGATCACAGAAATGGGACGCGGAGCTGACGCTCCAAGTATTCTTCCAGTGTGGGAAC
GTGAACTGCTTTTCGCACGTGATCCGCCTCGTATTACCTGCGCTCATTACGAGTACGAAGACG
TCATTGATCATAGCGATGGGTCTTACGCATTTTCCAACCAGAGCAATATGGTTCAGCGCAGTT
TTTATTTTGGTGCTAAAGAGATGCGCGTGTTACGCAAACAAATTCCCCCCCATTTAATCTCAA
CGTGTAGTACTTTCGACCTTATCACCGCCTGCCTGTGGAAATGTCGTACTTTAGTCCTGAAGA
TCAATCCAAAACAGGCCGTCCGCGTATCCTGCATTGTGAACGCACGTGGGAAACACAATAACG
TACACATTCCTCTGGGTTACTACGGCAACGCATTCGCCTTCCCGGCTGCGGTCAGTAAAGCCG
AACCTCTGTGTAAAAACCCGCTGGGGTATGCGTTAGAGTTGGTGAAGAAAGCGAAGGCGACCA
TGAATGAAGAGTACTTACGTAGCGTCGCCGATTTGTTGGTCCTGCGTGGTCGTCCGCAGTACA
GCTCTACTGGCTCTTATCTGATTGTGAGTGACAATACGCGCGCAGGGTTTGGTGATGTCAATT
TTGGGTGGGGGCAGCCCGTATTCGCAGGCCCCGCTAAAGCACTGGACTTAATCTCGTTCTATG
TCCAACATAAGAACAACATCGAGGATGGTATTCTTGTGCCCATGTGTTTACCCTCGTCGGCGA
TGGAACGTTTTCAGCAAGAACTGGAGCGTATCACAACAGGAACCtaaCTTAAGGAGACG

>PhBEBT1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGATTCGAAGCAATCCTCTG
AACTGGTATTTACAGTCCGCCGCCAAGAGCCTGAGCTTATTGCACCCGCAAAACCAACACCTC
GTGAGACTAAGTTTCTGAGTGACATCGATGATCAAGAAGGCCTTCGTTTTCAAATCCCAGTCA
TTAACTTTTACCGCAAGGATAGTAGTATGGGCGGCAAAGACCCAGTTGAAGTCATCAAGAAAG
CGATTGCGGAAACTCTGGTATTTTATTACCCGTTTGCAGGACGCCTGCGCGAAGGAAACGATC
GCAAGTTAATGGTAGACTGCACTGGTGAAGGGGTAATGTTTGTAGAAGCAAACGCCGATGTTA
CGTTGGAGGAATTTGGGGATGAACTGCAGCCGCCTTTTCCCTGTCTTGAGGAGTTATTGTATG
ACGTACCTGGCTCGGCCGGAGTCTTACACTGCCCCCTTTTGCTGATTCAAGTTACTCGCTTAC
GCTGTGGGGGATTTATTTTTGCACTGCGTCTGAATCACACGATGAGTGACGCCCCAGGATTAG
TCCAATTTATGACCGCGGTCGGTGAGATGGCTCGCGGCGCAACCGCACCATCTACATTACCGG
TTTGGTGTCGCGAGTTACTGAATGCACGCAATCCACCCCAAGTAACTTGCACCCATCACGAAT

ATGAAGAAGTTCCCGACACAAAAGGTACTCTTATTCCTCTGGACGATATGGTTCACCGCTCTT
TTTTTTTCGGGCCGACAGAAGTATCAGCACTTCGTCGTTTTGTTCCACCCCACTTGCATAATT
GCAGTACCTTTGAGGTTCTTACTGCTGCTCTGTGGCGCTGTCGCACGATCTCTATCAAACCTG
ACCCGGAGGAGGAAGTACGTGTTTTGTGTATTGTGAACGCTCGTTCTCGTTTCAATCCTCAGT
TGCCAAGTGGTTATTATGGCAACGCCTTTGCGTTCCCTGTCGCAGTGACTACCGCAGAAAAAT
TATGCAAAAACCCTCTGGGATACGCTCTGGAACTTGTGAAAAAAACGAAAAGCGACGTTACAG
AAGAATATATGAAGTCGGTGGCCGATTTGATGGTGATCAAAGGTCGTCCGCATTTTACGGTCG
TTCGCACATACCTGGTTTCGGACGTGACACGTGCTGGTTTTGGGGAAGTGGACTTCGGATGGG
GGAAGGCGGTATATGGGGGGCCCGCCAAAGGTGGGGTTGGGGCCATTCCTGGAGTAGCTAGTT
TCTACATTCCTTTTCGCAATAAAAAAGGTGAAAATGGTATTGTCGTCCCGATCTGCTTACCGG
GGTTCGCAATGGAAAAATTCGTAAAGGAGCTGGACAGTATGCTTAAAGGAGATGCTCAGTTAG
ATAATAAGAAATACGCCTTCATCACGCCCGCCTTAtaaCTTAAGGAGACG

>PuAAT

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGATGCCTCTGTCCGTACTTC
AGGTTAAGCGCTTGCAACCGGAGCTTATTACACCTGCAAAAAGCACGCCGCAAGAGACTAAAT
TTTTATCCGACATTGATGATCAGGAATTTTTACGCTTCCATGTACCCGTGATCATGTGCTATA
AGGACAACCCTTCGCTTAACAAAAACCGCAATCCAGTCAAAGTAATCCGCGAAGCACTGTCAC
GTGCCTTGGTGTACTATTATCCTCTTGCTGGTCGCCTGCGCGAGGGTCCAAACCGTAAGCTGG
TGGTTAACTGTAACGGCGAGGGCATCTTATTTGTGGAGGCTTCAGCTAACGTCACGCTTGAGC
AGTTAGGAGACAAAATCCTGCCTCCCTGCCCCTTATTAGAGGAGTTTTTGTTTAACTTTCCGG
GCTCTGATGGGATTATCGGTTGTCCTTTGCTGCTTGTTCAAGTGACTTGCCTGACGTGTGGGG
GGTTCATTCTTGCATTACGTTTAAATCATACGATGTGTGATGCAACTGGACTTCTGCTTTTCT
TAACAGCAGTTGCTGAGATGGCTCGTGGCGCTCATGCCCCTTCTATCCTGCCCGTGTGGGAAC
GTGAACTTCTGTTCGCCCGCGATCCTCCTCGCATCACCTGTGCTCACCACGAGTACGAGGACG
TGATCGATCATTCGGACGACAGTTACTCCTTCTCGAACCAAAGTAATATGGTGCAGCGCTCTT
TTTACTTTGGTGCGAAGGAGATGCGTGTTTTGCGTAAGCAAATTCCTCCACACTTGATTTCAA
CATGCTCAACTTTCGATTTAATCACGGCATGCTTATGGAAATGCCGCACACTGGCGCTGAAAA
TCAATCCTAAACAGGTAGTGCGCGTGTCGTGTGTGGTGAACGCCCGTGGGAAACACCATAATG
TTCGTTTACCACTGGGATATTATGGAAATGCCTTTGCCTTCCCCGCGGCGGTTAGCAAAGCCG
AGCCTTTATGCAAGAATCCCTTAGGCTACGCATTGGAGTTGGTAAAGAAGGCCAAAGCGACCA
TGAATGAGGAGTACTTACGCTCTGTAGCTGACCTGTTAGTCTTACGCGGTCGCCCCCAGTATA
GTTCCACTGGATCTTATCTTATTGTTAGCGACAACACCCGTGTCGGATTCGGAGATGTTAATT
TTGGGTGGGGACAGCCGGTCTTCGCAGGGCCTGCAAAAGCATTGGACCTGATTAGCTTCTACG
TTCAACACAAAAACAACACTGAAGACGGGATTCTGGTTCCGATGTGTCTGCCATCGTCAGCAA
TGAAACGTTTCCAGCAGGAGCTTGCACGCATTACACAGGAGCCGAACGAGGACATTTGTAACA
ATCTGCGTAGTACTCGCATCATGTCCATGATGtaaCTTAAGGAGACG

>RhAAT1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGAGAAAATCGAGGTAAGTA
TTATTTCACGTGATACAATTAAGCCCTCGGCTGCTAGTTCCTCGCTTCATCCTTACAAACTGT
CGATTATTGATCAATTTACACCAACAACATATTTTCCGGTAATTTTCTTTTATCCCATCACTG
ATCGTGTCTTTAACTTGCCACAAACCTTGACAGACTTGAAGAATACCGTATCACAAGCGCTGA
CATTATACCACCCTCTTTCTGGCCGCATTAAGAACAATCTGTACATCGACGATTTTGAAGCAG
GGATTCCGTATTTAGAAGCTCGTGTGAATTTTCACATGATTGATTTCTTGCGTTTGCCCAAGA
TCGAATGGTTGAACGAGTTCGTTCCGATGGCCCCGTATCGCAAAGAAACGATCTCCGAATTTC
TGCCTTTGTTAGGCATCCAAGTCAACATTTTCGATTCCGGCATTGCTATCGGTGTGAGCTTTA
GCCACAAGATCAACGATGGCCAGACGGCTTCTTGTTTTTTGAAGTCTTGGGTAGCGATTTTCC
GTGGGTACCGTAACAAGATTATCCACCCAAATCTGAGCCAGGCTGCATTGCTGTTACCATCAC
GTGATGACTTGCCTGAGAAATATGTTGCAATGATGGAGCGCATGTGGTTTGGAGAAAAAAAGG
TAGTTACACGTCGCTTCGTGTTTGACGCTAAGGCAATTTCGGCCTTGCAGGATGAAGGGAAAT
CGGAATACGTGCCAAAACCCTCGCGTGTCCAGGCCCTTACAGGGTTTTTGTGGAAACATCAGC
TTGCAGCCTCACGCGCATTATCTTCCGGAACCTCGACCCGTTTTTCGGTCGCTTCACAGACAG
TGAACTTACGCTCAAAAATGAATATGAAGACGACTTTAGATAACGCCATCGGCAACATCTTTT
TGTGGGCGTCGGCACGCTTAGACCTGAACGACACCGCCCCAGGTTCCTCGGACTTAAAGTTAT
GTGACTTGGTCAATTTACTGAATGAAAGTATTAAGGAATTTAACTCAGACTACTTAGAAATCT
TAAAAGGAAAGGAGGGGTACGGTGGTATGTGCGACCTGCTTGATTTTATGGAGGAGGGTTCTT

TCGTAGAGCCAGCGCCGGAATTTTATAGCTTCTCTAGTTGGACGCGCTTCTTCGATCAGGTCG
ACTTTGGGTGGGGCCGTCCCTCCTGGGTAGGTTTCTCCGGTCGCGTGGAGACACGTAATTTCA
CCATTTTTGTAGAGACTCAATGCGACGACGGTATTGATGCCTGGGTCACTGTAGACGAAAAGC
AGATGGCAATGTTGGAGCAAGACCCCCAGTTCTTAGCTTTCGCTTCCCCGAATCCCCGTATCA
GTATCGCTTCATCTGTGGGGATGGATtaaCTTAAGGAGACG

>SAAT

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGTCATATAAGAATAATCACT
CTATCCTTTCTAAACCTAATGATCCCGTTGAGGTAATCCGCGACGCACTGAGTAAAGCGTTGC
AATTTTACTACCCGCTTGCCGGACGTTTGCGCGAAGGACCAAACAAAAAGTTGATGGTAGACT
GTACCGGCGAGGGTATCTTATTCGTGGAAGCAAATGCGGAAGTCACCCTTGACGAGTTAGGGG
ATGCTATTTTGCCGCCATGTCCGTTTCTTGACGGCTTCTTGTTCAATGTGCCCGGTTCTGACG
GGATCTTGGGATCACCTTTATGCCTGATCCAAGTCACACGTTTAAGTTGCGGTGGGTTCATCT
TCGCGCTTCGCTTAAATCACACGATCTGCGATGCACTGGGTTTAGTCCAATTCTTAAACGCAG
TGGGTGAGATCGCTCAAGGAAAATACGCCCCATCCATCACACCTGTCTGGGAACGTGAGCTGT
TGTCAGCACGTGACCCACCCCGTATTTCTTGCACTCATGAGGAGTTTGACGACTCCATTGATC
ACTCTTACCCAAACTATGGGGCGACTGTCCAACAATGTTACTGTTTCGGACCGAAAGAGATTA
AGAGTCTGCGCGAGCATTTGCCTCCACATTTGTCAACATGTTCTTCCACGTTCGAACTGATCA
CTGCGTGTGTGTGGAAATGCCGCACCATCTCTTTGGACATGGATCCAGAACAGATTGTTCGTC
TGTCGTGTGTGGTCACTGCGCTGGGTAAACATAATAACGTGTGTTTACCGTTAGGGTACTACG
GAAATACATTCACCTATCCAGCGGTAGTTTCTACTGCGGAGCGTTTGTGTAATAGTCCCTTAG
GATATGCAGTGGAATTGGTCAAAAAATCTAAGGCTAAGATGTCGGAGGAATACTTGCGCTCCG
CCATCGATTTCGTTGAGGTCCGTGGCCGTCCGCCATTCGCTCTTGAAGGGATGTCAGATTTCT
TGGTCTCGGACAACACCCGTACTGGATTAGGCGAAATTGATTTTGGTTTTGGCAAACCCGTCT
ATGCGGGTGTTGCGAAGAGTACTGACCTTATTTCATTCTACGTTCGCAGCACCAACAAGGAGG
AGCGCGAAATTCTGGTTCCAGTATGTTTGCCGATTTTAAGTATGGAGATCTTTCAACAGGAAC
TGAAAAAGATGATTGGAtaaCTTAAGGAGACG

>VpAAT1

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGCAGAGAAGGCCTCAAGTT
TAATGTTCAATGTCCGTCGTCATGAGCCGGAATTGATCACCCCTGCTAAGCCAACGCCCCGCG
AAATCAAATTATTATCCGATATCGACGACCAGGACGGGCTTCGTTTCCAGGTGCCGATCATTC
AATTTTACAAGAATAATAGTTCTATGCAGGGGAAGAATCCCGCGAAGATTATTAAGAGTGCGC
TGGCCGAAACACTGGTTCATTATTATCCTCTGGCGGGGCGTTTACGCGAGGGGTTTGGGCGCA
AGTTAATGGTGGAATGTACAGGAGAGGGCATTCTGTTTATCGAGGCAGACGCAGATGTGACGC
TTCATGAGTTTGGGGATGACCTTCCTCCTCCATTCCCCTGTCTTGTGGAGCTTTTATACGACG
TTCCTGGGAGCAGTGGCATTATTGACACCCCGTTGTTGTTGATTCAAGTCACGCGTCTGAAGT
GTGGGGGGTTTATTTTTGCGTTGCGCTTGAATCACACAATGTCTGACGCCTCGGGGTTAGTTC
AGTTTATGACCGCGGTAGGTGAAATGGCACGTGGTCAACGTAGTCTGAGTATTCAGCCGGTAT
GGGAGCGCCATTTGCTGAATGCTCGCGATCCCCCCCGTGTGACTCATATTCATCATGAGTATG
ACGATCTGGAAGATACTAAAGGAACAATCATCCCGTTAGATGACATGGTACATCGTAGTTTTT
TCTTTGGGCCAAGCGAGATGGCAGCCATTCGTCGTCTTGTGCCAGCCCATTTTCATCGTTCGA
CGACGAGCGAGGTGTTGACCGCCTATTTGTGGCGCTGTTACACTATTGCCTTACAACCTGATC
CTGAGGAGGAGATGCGCGTAATCTGTGTGGTGAACAGCCGCACCAAGCTTAATCCGCCACTGC
CCACTGGATTTTATGGCAATGGCATCGCGTTCCCGGCGGCGATCTCACAGGCAAAAAAAATCT
GTGAGAACCCCTTCGGATATACTTTACAGCTTGTCAAGCAGACAAAAGTGGATGTGACAGAAG
AGTACATGCGTAGCGCCGCCGATCTTATGGCTATGAAAGGTCGCCCTCACTTCACTGTCGTCC
GCCGTTACATGGTGAGTGACGTGACCCGCGCGGGCTTTGGATTGGTCGACTTCGGCTGGGGTC
GTCCTGAGCCGGTTTATGGAGGGCCAGCTAAGGGTGGGGTTGGTCCCATTCCCGGGGTTACAA
GTTTCTTCGTTCCGTTCAAGAATCGTAAAGGTGAAAAAGGGATTGTAGTTCCCACTTGCTTAC
CAACGCCTGCTATGGAACGCTTCGCAAAGTTAATGAACGAGATTCTGCAAAACCAGCTGCTGG
TGTCTGCGGAAGAGAATAAAAGCGTATTCATTGTGAGTGCGATCtaaCTTAAGGAGACG

>VS

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGGCACCCCAGATGGAGAAAG
TCTCGGAAGAATTGATCTTACCCAGCTCGCCGACCCCACAATCTTTAAAGTGTTATAAGATTA

GTCACCTGGACCAGCTTCTTTTAACGTGCCACATTCCCTTTATCCTGTTTTATCCGAATCCAT
TAGACAGTAATTTGGACCCGGCGCAAACTTCTCAACACCTGAAGCAGTCACTGTCCAAAGTTC
TTACACATTTCTACCCACTTGCCGGACGCATTAATGTCAACTCGTCAGTGGACTGCAATGATT
CAGGAGTGCCTTTTGTTGAAGCTCGTGTGCAGGCCCAGTTATCTCAGGCGATCCAGAACGTAG
TGGAATTGGAGAAATTGGATCAATATTTACCATCTGCCGCCTATCCGGGTGGGAAAATCGAGG
TGAACGAAGACGTCCCCTTAGCTGTAAAAATTTCATTTTTTGAATGTGGTGGGACTGCTATCG
GCGTAAACTTATCCCATAAAATTGCTGACGTGTTGTCACTGGCGACGTTTTTGAATGCTTGGA
CTGCGACTTGCCGTGGCGAAACTGAGATCGTCTTGCCAAATTTCGACCTGGCAGCTCGCCACT
TTCCTCCGGTCGACAACACCCCGAGCCCTGAGCTGGTGCCAGACGAGAACGTAGTGATGAAGC
GTTTTGTTTTTGACAAAGAGAAGATTGGGGCGTTGCGCGCCCAAGCATCCAGCGCTTCCGAAG
AGAAGAATTTCAGTCGCGTACAACTGGTAGTCGCTTACATTTGGAAGCACGTGATCGACGTTA
CACGCGCAAAATACGGCGCTAAAAACAAGTTTGTTGTCGTGCAGGCAGTCAACTTACGTTCCC
GCATGAACCCCCCTTTACCTCATTATGCTATGGGTAATATCGCGACCCTGTTGTTCGCTGCAG
TGGATGCCGAGTGGGACAAGGATTTCCCGGACCTTATTGGGCCTCTGCGCACGTCCTTAGAAA
AGACAGAAGACGACCATAACCATGAATTATTAAAGGGCATGACATGCCTGTATGAATTGGAGC
CTCAAGAGTTGTTGAGCTTCACCTCTTGGTGTCGTTTGGGTTTTTATGACTTGGACTTTGGAT
GGGGAAAACCGCTGTCTGCATGCACCACCACTTTCCCGAAACGCAATGCCGCTTTATTGATGG
ACACCCGCAGCGGAGATGGGGTAGAGGCCTGGCTGCCTATGGCCGAAGATGAAATGGCGATGT
TACCGGTTGAGTTACTTAGCTTAGTTGACTCGGACTTTAGCAAAtaaCTTAAGGAGACG

>WS

CGTCTCACTAGTTTTGTTTAACTTTAAGAAGGAGATATACCATGTTCACCATTCCCCGCCGTG
TGAAGGCGGGGCGTAAGCGTTTCTTGCTTTGTAGTCCTGTGCTGCTGCTTAACATTATGCAGC
CTTACATTTTCTTCTGGACGGTCGGTCGCCATTACTGTAATTTCATTCCTTTGTATGCCGCGT
TTTGTACATGGTGGACGGCCTTTAAAGTGATGGCGTTTGGGATCGGACGCGGCCCTCTGTGCC
AGTTCTCGGCTTTCCATAAATTTGCCGTCGTGATGTTACTGCCCATCTTGCCTCATGGCGATA
CCAACCACGGGGTCAAAGACGAACGCAGTGGGAGCTCTTGGTCGAGTCCGACATACCTGGAAA
TGTTTGCAAAATTCTGTGGCCTGGGTCTTTGCACCTATGGCATTTCACAGCTGTCGCATGACG
GCTTCCCGGTATTGTATAATGTGTTTTTATCGCTTATCATGTACTTACACATCTGCGTCCAGT
ACACAGGTTCCAACCTTGCTACCTCGAAGGTCCTTCAAGTACCATTATCGGATGGAATGAACC
AACCATATTTTTCTACCAGTCTGTCGAACTTCTGGGGACGCCGCTGGAATCTGGTGGCATCCT
CGTCGTTGCGCCATGTTGTGTACGATCCAATCCGCGAAGGCCGCTTAGTCCCAAAGGGTCATC
CCGAGGAGAAACCCGGTGGTGGAAAAGAAGTTAGCCGCAAGGTGTTAGGCTCTCTTATGGCGT
TCTTGGTGAGTGGAATTATGCATGAATACATTCTGTGGTTAGCCACAGGGTTTTGGTCTGGTC
AGATGTTGCTTTTCTTCGTGGTACATGGGGTGGCGGTAGCGGCGGAGCGTGTTGCCAAGGTGG
CGTGGGCGCGTCATGGTTTGCCTGCCATCCCCTGTGCAGTCTCGATTCCAATGACTATCGGTT
TCTTGTTTGGCACAGCGGAATTATTATTCTATCCGCCGATTTTCTCAGCGAATTGGGCCGAGC
ATGGTGTGGCCGATTTGCGTCGCCAATTCCGTTCCCTTGGTCTTTCTGTCtaaCTTAAGGAGA
CG

A.4: pGGV4 sequencing results

Sequencing results for joins in new pGGV4 plasmid shown below (5'→3'):

> TC5_pGGV4_SEQ1

ACCCCAAGAAATTTCTCTTCTGCAAAGGCCTGGACGTTTGGGACAGCTGGCCATTAC
AAAACGCTGACGGCACTGTCGCAAACTATCACGGCTACCACATCGTCTTTGCATTAG
CCGGAGATCCTAAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAGTCG
GCGAAACTTCTATTGACAGCTGGAAAAACGCTGGCCGCGTCTTTAAAGACAGCGACA
AATTCGATGCAAATGATTCTATCCTAAAAGACCAAACACAAGAATGGTCAGGTTCAG
CCACATTTACATCTGACGGAAAAATCCGTTTATTCTACACTGATTTCTCCGGTAAAC
ATTACGGCAAACAAACACTGACAACTGCACAAGTTAACGTATCAGCATCAGACAGCT
CTTTGAACATCAACGGTGTAGAGGATTATAAATCAATCTTTGACGGTGACGGAAAAA

CGTATCAAAATGTACAGCAGTTCATCGATGAAGGCAACTACAGCTCAGGCGACAACC
ATACGCTGAGAGATCCTCACTACGTAGAAGATAAAGGCCACAAATACTTAGTATTTG
AAGCAAACACTGGAACTGAAGATGGCTACCAAGGCGAAGAATCTTTATTTAACAAAG
CATACTATGGCAAAAGCACATCATTCTTCCGTCAAGAAAGTCAAAAACCTCTGCAAA
GCGATAAAAAACGCACGGCTGAGTTAGCAAACGGCGCTCTCGGTATGATTGAGCTAA
ACGATGATTACACACTGAAAAAAGTGATGAAACCGCTGATTGCATCTAACACAGTAA
CAGATGAAATTGAACGCGCGAACGTCTTTAAAATGAACGGCAAATGGTATCTGTTCA
CTGACTCCCGCGGATCAAAAATGACGATTGACGGCATTACGTCTAACGATATTTACA
TGCTTGGTTATGTTTCTAATTCTTTAACTGGCCCATACAAGCCGCTGAACAAAACTG
GCCTTGTGTTAAAAATGGATCTTGATCCTAACGATGTAACCTTTACTTACTCACACT
TCGCTGTACCTCAAGCGAAAGGAAACAATGTCGTGATTACAAGCTATATGACAAACA
GAGGATTCTACGCAGACAAACAATCAACGTTTGCGCCTAGCTTCCTGCTGAACATCA
AAGGCAAGAAAACATCTGTTGTCAAGACAGCATCCTTGAACAAGGACATTTACCGTT
TACAAATAAAAACGCAAAGAAAATGCCGATTAAGGTGCATCCCGTCCCCTTAAAATA
CACCTGAAATCCGGTGGTAACAAACCCAAAAGAAACTAAATTGGTTCGGCCCCGCTG
ACAAAAACTTGCTA

> TC5_pGGV4_SEQ2

CCCGCATTGGTGAATTGAGCTAACGATGATTACACACTGAAAAAAGTGATGAAACCG
CTGATTGCATCTAACACAGTAACAGATGAAATTGAACGCGCGAACGTCTTTAAAATG
AACGGCAAATGGTATCTGTTCACTGACTCCCGCGGATCAAAAATGACGATTGACGGC
ATTACGTCTAACGATATTTACATGCTTGGTTATGTTTCTAATTCTTTAACTGGCCCA
TACAAGCCGCTGAACAAAACTGGCCTTGTGTTAAAAATGGATCTTGATCCTAACGAT
GTAACCTTTACTTACTCACACTTCGCTGTACCTCAAGCGAAAGGAAACAATGTCGTG
ATTACAAGCTATATGACAAACAGAGGATTCTACGCAGACAAACAATCAACGTTTGCG
CCTAGCTTCCTGCTGAACATCAAAGGCAAGAAAACATCTGTTGTCAAAGACAGCATC
CTTGAACAAGGACAATTAACAGTTAACAAATAAAAACGCAAAAGAAAATGCCGATTA
TGGTGCACTCTCGTCTCCTTAAGATACACCACTGAGATCCGGCTGCTAACAAAGCCC
GAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTG
GGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT
TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTAC
GCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTT
CCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCT
CCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTA
GGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGAC
GTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAA
CCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTG
GTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAAC
GCTTACAATTTAGGTGGACTTTTCGGGGAAATGGGCGCGGAACCCCTATTTGTTTAT
TTTCTAAATAATTCAAATATGTATCCGCCTCTGAAAAAATAACCCTGAAAAAGGGTT
CCATAAATTGAAAAGGGAGAAATTGAATATTCAAATTTCCGGGGCGCCTTATTCCTT
TTTTGGGGATTTTTCCTTCCGTTTTGCCCCCCCCCAAACCCGGGGGGGGAA

> TC5_pGGV4_SEQ3

TCCCGCAAGGAGCGCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCG
CAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCC
TTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCC
TTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGG
TGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTT
GGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCC

TATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTT
AAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCT
TACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTT
TTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTT
CAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATT
CCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAA
GTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTC
AACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGC
ACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAG
CAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTC
ACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATA
ACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAG
GAGCTAACCGCTTTTTTGCACAACATGGGGGGATCATGTAACTCGCCTTGATCGTTG
GGAACCGGAACTGAATGAAGCCCTACCAAACGACGAGCGTGACCCCCGAATGCCTGC
AGCAAGGCCACAACCGTTGCCCAACTATTAACTGGCGAATTAATTTATCTAACTTTC
CGGGACCAATTAAATAACTGGGATGGAGGCGGGAAAAGTTGGGAGGACTCCTTTTGG
GCTTGGGGCCTTCCGGGAGGGGGGGGTTTATGTGGTGAAATATCGTGGGAGGGCGGG
GGAAGGTGAGGGCCCCCCCGTTGCATTTTTGTCCACTGGGGGGCCCAG

> TC5_pGGV4_SEQ4

CCCCCCCCTTACGAAGTAGGTGTTCCCAGGGTAGCCAGCAGCATCCTGCGATGCAGA
TCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACAC
GGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGT
CGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCC
GCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGG
ACCCAACGCTGCCCGTGATCTCGATCCCGCGAAATGCGCGGCCGCTTGACAGCTAGC
TCAGTCCTAGGTATTGTGCTAGCACTAGTGAGACCACAACGGTTTCCCTCTAGAAAT
AATTTTGTGCTAGCGTTTGGCCCGGACCGGGCCGGAGGCCTGTCATGAGTGATTACG
AGCCGTTGCGTCTGCATGTCCCGGAGCCCACCGGGCGTCCTGGCTGCAAGACCGACT
TTTCCTATCTGCACCTGTCCCCCGCCGGCGAGGTACGCAAGCCGCCGGTGGATGTCG
AGCCCGCCGAGACCAGCGACCTGGCCTACAGCCTGGTACGTGTGCTCGACGACGACG
GCCACGCCGTCGGTCCCTGGAATCCGCAGCTCAGCAACGAACAACTGCTGCGCGGCA
TGCGGGCGATGCTCAAGACCCGCCTGTTCGACGCGCGCATGCTCACCGCGCAACGGC
AGAAAAAGCTTTCCTTCTATATGCAATGCCTCGGCGAGGAAGCCATCGCCACCGCCC
ACACCCTGGCCCTGCGCGACGGCGACATGTGCTTTCCGACCTATCGCCAGCAAGGCA
TCCTGATCACCCGCGAATACCCGCTGGTGGACATGATCTGCCAGCTTCTCTCCAACG
AGGCCGACCCGCTCAAGGGCCGCCAGCTGCCGATCATGTACTCGAGCAAGGAGGCAG
GTTTCTTCTCCATCTCCGGCAACCTCGCCACCCAGTTCATCCAGGCGGTCGGCTGGG
GCATGGCCTCGGCGATCAAGGGCGACCCGCGCATCGCCTCGGCCTGGATCGGCGACG
GCGCCCCGCCGAGTCGGACTTCCAACCGCCCTCACTTTCCCCATGTCTACCGCGCGC
CGGTAATCCTCCACGGGGTCAACAACCATTGGGCGATCTCAACTTTCAGGGCATCGC
CGGCGGGGAAGGCCCCACTTCGCCAACCGGGGCTGGGGCTGCGGAATCCCCCCCTTG
GGGGTCGAACGGCAAATAATTTCTGGGGGGGCTTACCCCTCCCCCAAAGGGGCCCCC
AAAAAGGGCCCGGGGAAAACTCGGGG

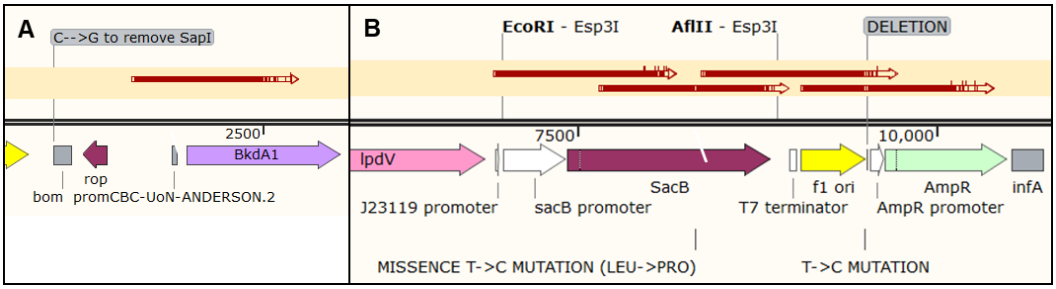


Figure S 1: Sequencing of pGGV4

A: SapI removal region, B: ACX/AAT insertion region. Sequencing carried out by Eurofins; results aligned using Snapgene.

Supplementary B: Bioinformatics

B.1: Sequences used for ACX phylogeny

Label	*Organism* annotation	GenBank ID
AtACX4	*Arabidopsis thaliana* ACX4	N.A. (mutant)
SPIOL(ACX4)	*Spinacea oleracea* ACX4	XP_021855534.1
ZEAMA(ACX4)	*Zea mays* ACX4	ONM29903.1
VIGRR(ACX4-X2)	*Vigna radiata* ACX4 isoform X2	XP_014516782.1
VIGRR(ACX4-X1)	*Vigna radiata* ACX4 isoform X1	XP_022641792.1
GLUNI_UNKNOWN	*Glutamicibacter nicotianae* unknown isoform ACX	VXC33980.1
SPIOL(ACX1)	*Spinacea oleracea* ACX1	XP_021856262.1
SPIOL(SOVF-A)	*Spinacea oleracea* ACXA	KNA24565.1
SPIOL(SOVF-B)	*Spinacea oleracea* ACXB	KNA24566.1
SPIOL(SOVF)	*Spinacea oleracea* unknown isoform ACX	KNA08139.1
VIGRR(ACX3-1)	*Vigna radiata* ACX3	XP_014524427.1
VIGRR(ACX2)	*Vigna radiata* ACX2	XP_014509033.1
VIGRR(ACX1-1)	*Vigna radiata* ACX1	XP_014508908.1
VIGRR(ACX1-2)	*Vigna radiata* ACX1	XP_014521199.1
VIGRR(ACX4-2)	*Vigna radiata* ACX3	XP_022633883.1
ZEAMA(ACX1-1)	*Zea mays* ACX1	PWZ07906.1
ZEAMA(UNKNOWN-1)	*Zea mays* unknown isoform ACX	ACF78566.1
ZEAMA(UNKNOWN-2)	*Zea mays* unknown isoform ACX	CAN28961.1
ZEAMA(ACX1-2)	*Zea mays* ACX1	AQL02798.1
ZEAMA(ACX2)	*Zea mays* ACX2	ACG45431.1
YARLI(ACX3-1)	*Yarrowia lipolytica* ACX3	CAA04661.1
CANTR(UNKNOWN)	*Candida tropicalis* unknown isoform ACX	AAA34322.2
SPASU(ACX4-1)	*Spatholobus suberectus* ACX4	TKY63513.1
ARTAN(ACX4)	*Artemisia annua* ACX4	PWA67585.1
POPAL(ACX4-X3)	*Populus alba* ACX4 X3	TKS13357.1
GLYMA(UNKNOWN-1)	*Glycine max* unknown isoform ACX	NP_001236991.2
ARAHY(ACX4)	*Arachis hypogaea* ACX4	QHO54153.1
TREOR(UNKNOWN)	*Trema orientale* unknown isoform ACX	PON92218.1
ACTCH(ACX4)	*Actinidia chinensis* ACX4	PSS23831.1
PARAN(UNKNOWN)	*Parasponia andersonii* unknown isoform ACX	PON51135.1
DORHY(ACX4)	*Dorcoceras hygrometricum* ACX4	KZV15167.1
PYRUS(ACX4)	*Pyrus ussuriensis* ACX4	KAB2621925.1
CAPAN(ACX4)	*Capsicum annuum* ACX4	PHT70509.1
CARLI(ACX4)	*Carex littledalei* ACX4	KAF3333609.1
CAJCA(ACX4)	*Cajanus cajan* ACX4	KYP46695.1
CINMI(ACX4)	*Cinnamomum micranthum* ACX4	RWR86286.1
GLYMA(UNKNOWN-2)	*Glycine max* unknown isoform ACX	BAG09369.1
APOSH(ACX4-1)	*Apostasia shenzhenica* ACX4	PKA67096.1
SPASU(ACX4-2)	*Spatholobus suberectus* ACX4	TKY65814.1
GLYSO(ACX4-B)	*Glycine soja* ACX4b	RZB52878.1
CAPCH(ACX4)	*Capsicum chinense* ACX4	PHU05248.1
TRIUR(ACX4)	*Triticum urartu* ACX4	EMS57934.1
HORVU(ACX4)	*Hordeum vulgare* ACX4	KAE8784658.1

YARLI(ACX3-2)	*Yarrowia lipolytica* ACX3	O74936
SACCE(ACX1)	*Saccharomyces cerevisiae* ACX1	P13711
CANTR(ACX4)	*Candida tropicalis* ACX4	P06598
YARLI(ACX2)	*Yarrowia lipolytica* ACX2	O74935
CUCMA(ACX2)	*Cucurbita maxima* ACX2	O64894
YARLI(ACX1)	*Yarrowia lipolytica* ACX1	O74934
DICDI(ACX1)	*Dictyostelium discoideum* ACX1	Q54GQ6
CAEEL(ACX15)	*Caenorhabditis elegans* ACX1.5	P34355
CANTR(ACX5)	*Candida tropicalis* ACX5	P08790
CANTR(ACX2)	*Candida tropicalis* ACX2	P11356
ASHGO(ACX1)	*Ashbya gossypii* ACX1	Q756A9
CANMA(ACX4)	*Candida maltosa* ACX4	P05335
CANMA(ACX2)	*Candida maltosa* ACX2	Q00468
PICPA(ACX1)	*Pichia pastoris* ACX1	Q9Y7B1
CANGA(ACX1)	*Candida glabrata* ACX1	Q6FY63
DEBHA(ACX1)	*Debaryomyces hansenii*	Q6BRD5
KLULA(ACX1)	*Kluyveromyces lactis* ACX1	Q6CKK7
PROMN(UNKNOWN)	*Prorocentrum minimum* unknown isoform ACX	A0A2K8DRF5
THAPS(ACX1)	*Thalassiosira pseudonana* ACX1	B8CA48
CORMI(UNKNOWN-1)	*Corynebacterium minutissiumum* unknown isoform ACX	A0A2X4R989
CORMI(UNKNOWN-2)	*Corynebacterium minutissimum* unknown isoform ACX	A0A376D3B9
BRANE(ACX1)	*Brachybacterium nesterenkovii* ACX1	A0A1X6X9H6
CORST(UNKNOWN-1)	*Corynebacterium striatum* unknown isoform ACX	A0A376GPP0
CORST(UNKNOWN-2)	*Corynebacterium striatum* unknown isoform ACX	A0A449HG44
CORAU(UNKNOWN)	*Corynebacterium aurimucosum* unknown isoform ACX	C3PIV8
ACTO(ACX1)	*Actinomycetales* bacterium ACX1	VA0A1R4EPF4
BRAPC(ACX3)	*Brachionus plicatilis* ACX3	A0A3M7Q8K5
ASPFU(ACX1)	*Neosartorya fumigata* ACX1	Q4WGS8
CORRG(UNKNOWN)	*Corynebacterium resistens* unknown isoform ACX	F8E372
9PSEU(ACX1)	*Alloactinosynnema* sp. ACX1	A0A0H5D1M9
CORSI(UNKNOWN)	*Corynebacterium singular* unknown isoform ACX	A0A0B6F3H7
APOSH(ACX4-2)	*Apostasia shenzhenica* ACX4	A0A2I0B3Q8
CORPI(UNKNOWN)	*Corynebacterium pilosum* unknown isoform ACX	A0A376CLC1
CORAY(UNKNOWN)	*Corynebacterium amycolatum* unknown isoform ACX	A0A376C6R8
ISOVA(UNKNOWN)	*Isoptericola variabilis* unknown isoform ACX	F6FSZ4
CELAL(UNKNOWN)	*Cellulophaga algicola* unknown isoform ACX	E6X4N8
ACTSP(ACX1)	*Actinokineospora* ACX1	W7IT94
MYCFO(UNKNOWN)	*Mycolicibacterium fortuitum* unknown isoform ACX	A0A0N9XLM7
MYCSM(UNKNOWN)	*Mycolicibacterium smegmatis* unknown isoform ACX	A0A0D6IPE5
RHOGO(UNKNOWN)	*Rhodococcus gordoniae* unknown isoform ACX	A0A379M002
RHOCO(UNKNOWN)	*Rhodococcus coprophilus* unknown isoform ACX	A0A2X4XC01
BREAN(UNKNOWN)	*Brevibacterium antiquum* unknown isoform ACX	A0A2H1HS78
STRGR(ACX1)	*Streptomyces griseus* ACX1	A0A380N8I8

B.2: Sequences used for AAT phylogeny

Label	*Organism* annotation	GenBank ID
PuAAT	*Pyrus ussuriensis* AAT	AJD18611.1
EjAAT1	*Eriobotrya japonica* AAT1	AHC3222.2
PcAAT	*Pyrus communis* AAT	AAS48090
DkAAT1	*Diospyros kaki* DkAAT1	AKE98481.1
MbAAT	*Musa balbisiana* AAT	THU42907.1
MaAAT	*Musa acuminata* AAT	XP_009388282.1
MdAAT	*Malus domestica* AAT	AAS79797.1
SAAT	*Fragaria ananassa* AAT	AAG13130.1
BanAAT	*Musa sapientum* AAT	CAC09063.1
VAAT	*Fragaria vesca* AAT	CAC09062.1
ManAAT	*Mangifera indica* AAT	CAC09378.1
TomAAT	*Solanum lycopersicum* AAT	AAS48091.1
LAAT	*Citrus limon* AAT	CAC09049.1
WS	*Euglena gracilis* wax synthase	ADI60058.1
WDS4	*Tricothecium roseum* acyl-coenzyme A: diacylglycerol acyltransferase 4	ASA49417.1
WS-2	*Vitis vinifera* wax synthase 2	AAO18665.1
WS-3	*Vitis vinifera* wax synthase 3	AAO18666.1
WS-1	*Vitis vinifera* wax synthase 1	AAO18664.1
PyAAT	*Prunus yeodensis* AAT	PQQ18992.1
RcAAT	*Rosa chinensis* AAT	PRQ19177.1
CER2	*Arabidopsis thaliana* eceriferum 2	AAM64817.1
Glossy2	*Zea mays* Glossy2 locus ORF 2	CAA61258.1
SalAAT	*Papaver somniferum* AAT	AAK73661.1
Pun1	*Capsicum annum* acyltransferase Pun1	AAV66311.1
DAT	*Catharanthus roseus* deacetyl vindoline O-acetyltransferase	AAC99311.1
MAT	*Catharanthus roseus* minovincinine 19-hydroxy-O-acetyltransferase	AAO13736.1
VS	*Rauvolfia serpentina* vinorine synthase	CAD89104.2
Ss5MaT2	*Salvia splendens* pelargonidin 3-O-(6-caffeoylglucoside) 5-O-(6-O-malonylglucoside) 4-malonyltransferase	AAR26385.1
CbBEAT	*Clarkia brewerii* acetyl-CoA-benzyl alcohol acetyltransferase	AAC18062.1
CmAAT4	*Cucumis melo* AAT4	AAW51126.1
RhAAT1	*Rosa hybrid cultivar* AAT1	AAW31948.1
CHAT	*Arabidopsis thaliana* (Z)-3-hexen-1-ol acetyltransferase	AAN09797.1
HMT	*Lupinus albus* 13-hydroxylupanine O-tigloyltransferase	BAD89275.1
AMAT	*Vitis labrusca* methanol O-anthraniloyltransferase	AAW22989.1
CbBEBT	*Clarkia brewerii* benzyl alcohol O-benzoyltransferase	AAN09796.1
CmAAT3	*Cucumis melo* AAT3	AAW51125.1
NtBEBT	*Nicotiana tabacum* benzyl alcohol O-benzoyltransferase	AAN09798.1
PhBEBT1	*Petunia hybrida* benzyl alcohol O-benzoyltransferase 1	AAU06226.1
CmAAT1	*Cucumis melo* AAT1	CAA94432.1
CmAAT2	*Cucumis melo* AAT2	AAL77060.1
DBNBT	*Taxus canadensis* 3'-N-debenzoyl-2'-deoxytaxol N-benzoyltransferase	AAM75818.1

BAPT	*Taxus cuspidata* phenylpropanoyltransferase	AAL92459.1
DBAT	*Taxus cuspidata* 10-deacetylbaccatin III 10-O-acetyltransferase	AAF27621.1
DBBT	*Taxus cuspidata* 2-α-hydroxytaxane 2-O-benzoyltransferase	Q9FPW3.1
TAT	*Taxus cuspidata* taxadien-5-α-ol O-acetyltransferase	AAF34254.1
HCBT	*Dianthus caryophyllus* anthranilate N-benzoyltransferase protein 2	CAB06430.1
HQT	*Nicotiana tabacum* hydroxycinnamoyl CoA quinate transferase	CAE46932.1
AsHHT1	*Avena sativa* hydroxycinnamoyl-CoA:5-hydroxyanthranilate N-hydroxycinnamoyl transferase 1	BAC78633.1
NtHCT	*Nicotiana tabacum* shikimate O-hydroxycinnamoyl transferase	CAD47830.1
AtHCT	*Arabidopsis thaliana* shikimate O-hydroxycinnamoyl transferase	NP_199704.1
Eht1	*S. cerevisiae* medium-chain fatty acid ethyl ester synthase/esterase 2	P38295.1
Eeb1	*S. cerevisiae* medium-chain fatty acid ethyl ester synthase/esterase 1	Q02891.1
Atf1	*Saccharomyces cerevisiae* alcohol O-acetyltransferase 1	CAA99708.1
Atf2	*Saccharomyces cerevisiae* alcohol O-acetyltransferase 2	CAA97203.1
CFAT	*Petunia hybrida* coniferyl alcohol acyltransferase	ABG75942.1
VvAAT1	*Vitis vinifera* AAT1	ART85743.1
LaAAT1	*Lavandula angustifolia* AAT1	DQ886904.1
SpAAT1	*Solanum pennelli* AAT1	AIW04708.1
LaAAT2	*Lavandula angustifolia* AAT2	ABI48361.1
PhpAAT1	*Physalis peruviana* AAT1	AFW03968.1
FcAAT1	*Fragaria chiloensis* AAT1	ACT82247.1
AT1	*Actinidia deliciosa* AAT1	AIC83790.1
AT9	*Actinidia eriantha* AAT9	AIC83789.1
CAAT1	*Larrea tridentata* AAT1	AHA90802.1
CAAT2	*Larrea tridentata* AAT2	KF543261.1
PhWS	*Petunia hybrida* wax synthase	AAZ08051.1
SpAAT2	*Solanum pimpinellifolium* AAT2	AGK82816.1
VpAAT1	*Vasconcellea pubescens/Vasconcellea cundinamarcensis* AAT1	ACT82248.1

Supplementary C: Analytical figures

C.1: GC-MS

Calibration curves for butyl acetate (BA), butyl propionate (BPI), butyl isobutyrate (BIB), butyl methacrylate (BMA) and butyl isovalerate (BIV) for concentrations from 0.031 mM to 2.5 mM:

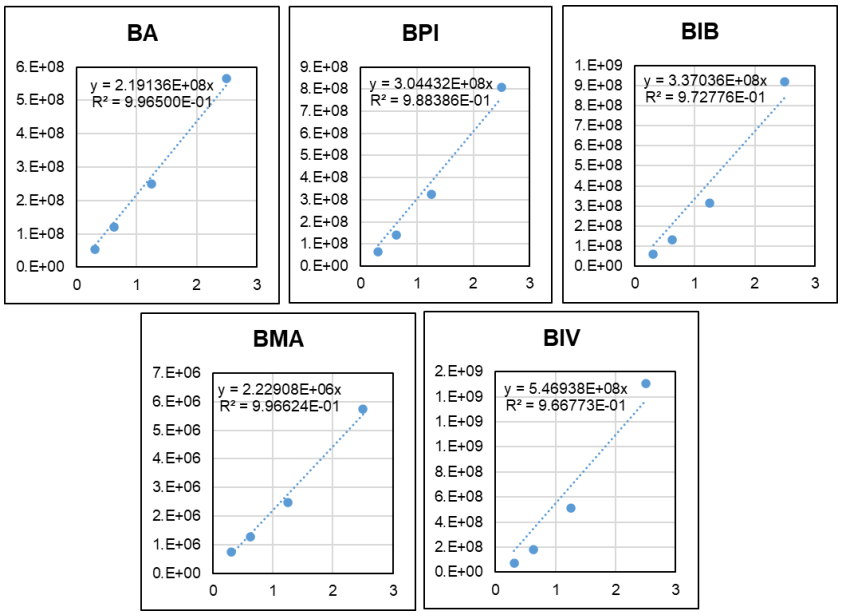


Figure S 2

Example GC-MS trace from a calibration using different concentrations of each of the esters of interest. Butyl acetate (BA) at 4.9 min, butyl propionate (BPI) at 7.1 min, butyl isobutyrate (BIB) at 7.8 min, butyl methacrylate (BMA) at 8.1 min, and butyl isovalerate (BIV) at 8.9 min:

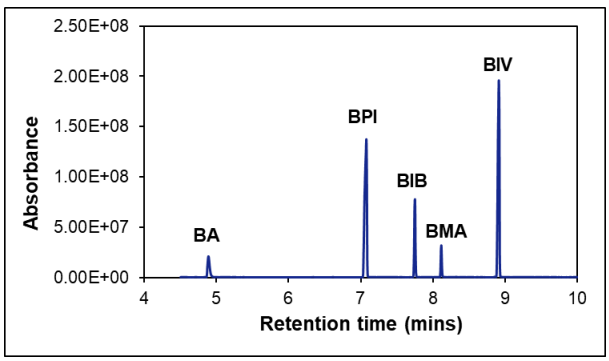
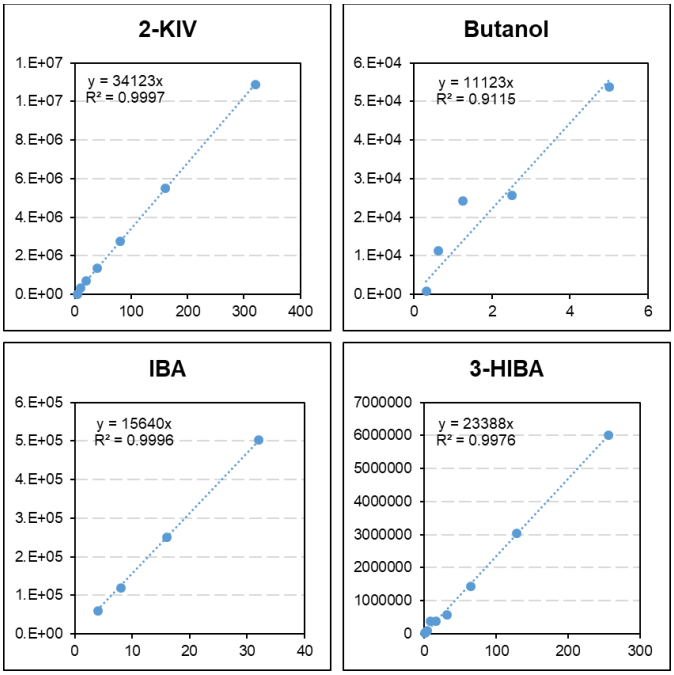


Figure S 3

C.2: HPLC

Calibration curves for 2-KIV, BuOH, IBA, and 3-HIBA using HPLC. All concentrations in mM.



C.3: BSA assay calibration curve

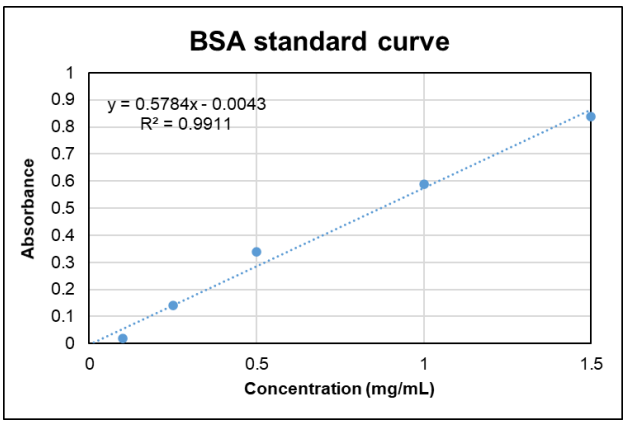


Figure S 4

C.4: Oxygen electrode assay

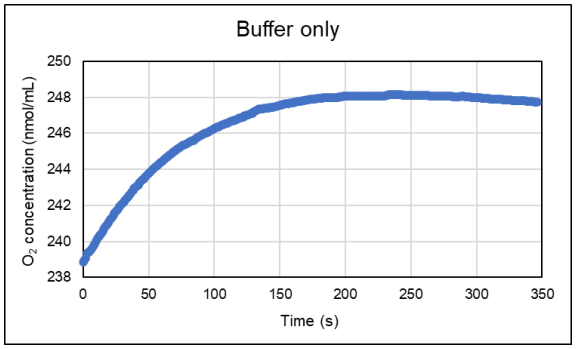


Figure S 5

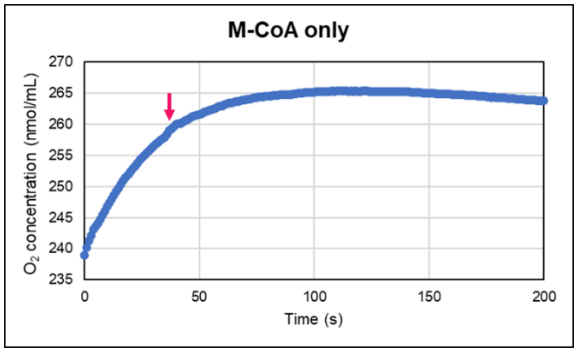


Figure S 6