The University Of Sheffield.

Lignin it, to win it: Transformation, Toxicity and Transport in the Microbial Utilisation of Lignin.

A thesis submitted in part fulfilment for the degree of

Doctor of Philosophy

Department of Molecular Biology and Biotechnology,

University of Sheffield

Calum A Pattrick

MBiolSci. Biochemistry and Microbiology, University of Sheffield

September 2017

Summary

Microbial utilisation of lignocellulosic biomass is now a key area of biotechnology, with the focus on production of biofuels as a result of the shift to renewable energy. However, industrial application of such processes is frequently not yet cost-effective. Therefore, increased valorisation of biomass through utilisation of the often untapped lignin component can facilitate the future profitability of this industry. Essential to enhancing biotransformation processes is overcoming problems with the toxicity of substrates and products, and the transport of these compounds across biological membranes; this study has explored the microbial physiology of these processes.

Rhodopseudomonas palustris is a model organism for the anaerobic degradation of lignin-derived aromatic compounds, and has great potential for exploitation for biotechnological purposes. This study investigated harnessing *R. palustris* genes for enhanced microbial biotransformation of ferulate to the commercially valuable flavouring compound, vanillin. Engineering a recombinant *E. coli* strain with the ability to transport ferulate and convert it into vanillin was only partially succesful, but the data obtained revealed some unexpected novel findings on aromatic metabolism in *R. palustris*.

Microbial production of vanillin has been achieved in a range of species, but the toxic effects of this aromatic aldehyde are inhibitory to yield and also poorly understood. A high-resolution global-proteomics approach was used to elucidate the mechanism of toxicity in *E. coli*. Combined with physiological follow-up experiments, a severe oxidative stress response and potential targets for enhancing vanillin tolerance were identified. Furthermore, vanillin-tolerant strains of *E. coli* were evolved and their genomes sequenced, which provided several novel insights into the toxicity of vanillin. These findings can inform toxicity mitigation for other similar industrially-relevant compounds.

Key to the biotechnological potential of *R. palustris* is its versatile metabolism and extensive transportome. The secondary TRAP transport systems are relatively under-studied, and this work addressed the functions of the high-affinity solute binding proteins of several *R. palustris* TRAP systems. This resulted in the identification of a C4-dicarboxylate DctPQM homolog, and a medium-chain dicarboxylate transporter with high binding affinity for pimelate, RPA1975. Furthermore, functional redundancy — a common feature in the *R. palustris* genome — was investigated through comparison with ABC transport binding proteins that also showing binding with medium- and long-chain dicarboxylates.

Acknowledgements

I would like first thank Dave Kelly for giving me this opportunity, and for his continual support, guidance and mentoring from my initial time in F1 as an undergrad and throughout for the course of the four years of this PhD. While he's a major factor in making life and research in F1 enjoyable/bearable, I must also thank all the other people I've had the pleasure of working alongside and have often offered their assistance: Aidan, Rob, Mike, Kai, Wei, Leo, Nits, Hannah, Tom, Pete, Abrar, Reem, Smarty, Majeed, and anyone else I've forgotten.

There are several people outside the lab that have helped without whom I would not have been able to complete this work: Richard Benniston and Mark Collins for their help with the MS and proteomic data analysis; Andrea Hounslow for her help with all the NMR; Roy Chaudhuri for help with genome assembly and bioinformatics; John Rafferty for help with x-ray crystallography; Craig MacGregor-Chatwin for his help with the EM; Adam Brooks for help with the fluorescence microscopy; and Andy Hitchcock for a large amount of help and knowledge on too many things to mention.

To all my friends in Sheffield and beyond, many who I have already mentioned by name, thank you for providing me with distractions and a life outside of the lab, whether it be organising and enjoying hitchhikes across Europe or winning the IM Thursday 5pm 6-a-side football league with Plasmiddlesbrough FC. I must also thank Macclesfield Town FC who have continuously provided me with experiences of disappointment and misery for many years, without which I may not have been able to cope with what my PhD had to offer.

My family have always been there to offer support and encouragement, so thank you Mum, Dad, Sally and Alastair; special thanks must got to my Dad who's experience of academic life has proved invaluable in terms of guidance and reading thousands of words (many of which he doesn't believe are real), despite his preference for rocks. And lastly, but definitely not leastly, I must thank Lucy who has suffered through the majority of my PhD with me while offering unwaivering support, keeping me sane, and encouraging me both to carry on, and (equally importantly) to stop working, when I needed to hear it.

CP

2017

Publications and Presentations

PROPOSED PUBLICATION:

Pattrick, C., Collins, M., Chaudhuri, R. and Kelly, D. J. (TBC) Comprehensive analysis of vanillin toxicity in *E. coli* identifies means of engineering a robust microbial platform for biotransformation.

N.B. Chapter 4 contains the proposed manuscript for this publication that, as of the time of submission of this thesis, is set for imminent submission.

ORAL PRESENTATIONS:

Pattrick, C., Salmon, R., Bisson, C. and Kelly, D. J. (2016) Exploring the potential TRAP and ABC solute uptake transporters from the phototroph *Rhodopseudomonas palustris* for use in biotechnology. *The Microbiology Society Annual Conference 2016*, Liverpool, UK. 22 March 2016.

Pattrick, C., Salmon, R., Kelly, D. J. (2016) Bioconversion of ferulate to vanillin: genomics and proteomics approaches to understanding and mitigating vanillin toxicity. *17th Annual European Congress on Biotechnology 2016*, Krakow, Poland.

Abbreviations

ABC	ATP-binding cassette
Amp	Ampicillin
ADP	Adenosine diphosphate
ALE	Adaptive laboratory evolution
AMP	Adenosine monophosphate
APS	Ammonioum persulfate
ATP	Adenosine triphosphate
BCS	Bathocuproinedisulfonic acid disodium salt
cAMP	Cyclic adenosine monophosphate
CCD	Charge-coupled device
CFE	Cell-free extract
Cm	Chloramphenicol
CoA	Coenzyme A
COG	Cluster of orthologous groups
CRP	cAMP-receptor protein
dATP	Deoxyadenosine triphosphate
DCFDA	2',7'-dichlorodihydrofluorescein diacetate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
DH	Dehydrogenase
dH_2O	Distilled water
DHAP	Dihydroacetone phosphate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dTDP	Deoxythymidine diphosphate
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
ECF	Energy-coupling factor
EDTA	Ethylenediamine tetra-acetic acid
EET	External electron transfer
EK	Enterokinase
EM	Electron microscopy
EVT	Evolvd vanillin tolerance
FCCP	Carbonyl cyanide-4-phenylhydrazone
FDR	False-discovery rate
$FMNH_2$	Reduced flavin mononucleotide
FTMS	Fourier-transform mass-spectrometry
gDNA	Genomic DNA
GMP	Guanoside monophosphate
GSH	Glutathione
GSSG	Glutathione disulfide
H_2O	Water
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMPHP	4-hydroxy-3-methoxyphenyl-β-hydroxypropionic acid
HMPKP	4-hydroxy-3-methoxyphenyl-β-ketopropionic acid
HSQC	Heteronuclear single-quantum correlation spectroscopy
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin

K_d	Dissociation constant
K_m	Michaelis constant
KO	Knock-out
LB	Lysogeny broth
LFQ	Label-free quantification
LTQ	Linear trap quadrapole
MBP	Maltose-binding protein
MCS	Multiple-cloning site
MES	2-(N-morpholino)ethanesulfonic acid
MFS	Major facilitator superfamily
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	Messenger RNA
MRP	Multidrug resistance protein
NAD	Nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NBD	Nucleotide-binding domain
NMR	Nuclear magnetic resonance
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBP	Periplasmic binding protein
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEP	Phosphoenolpyruvate
PMF	Proton motive force
PPP	Pentose phosphate pathway
PTS	Phosphotransferase system
PYE	Peptone yeast extract
qPCR	Quantitative PCR
RARE	Reduced aromatic aldehyde reduction
RBS	Ribosomal binding site
RCV	*Rhodobacter capsulatus* vitamins
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPA	*Rhodopseudomonas palustris*
RT-PCR	Real-time PCR
SBP	Substrate binding protein
SDS	Sodium dodecyl sulfate
SNP	Single nucleotide polymorphism
TAE	Tris base, acetic acid, EDTA
TAXI	TRAP-associated extracytoplasmic immunogenic
TCA	Tricarboxylic acid
TE	Tris base, EDTA
T_m	Melting temperature
TMD	Transmembrane domain
TOCSY	Total correlation spectroscopy
TRAP	Tripartite ATP-independent periplasmic transporter
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
TSP	Trimethylsilylpropanoic acid

TTT	Tripartite tricarboxylic acid transporter
UDP	Uridine diphosphate
UV	Ultraviolet
v/v	Volume/volume (concentration)
WT	Wildtype
w/v	Weight/volume (concentration)

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1. Lignin

1.1.1. Structure, composition and function

Lignocellulosic biomass is one of the most abundant reserves of fixed carbon on the planet and as such is a natural resource of high-importance in biotechnology. There are three components of lignocellulose: cellulose, a polysaccharide of β(1-4)-linked glucose molecules; hemicellulose, a mixture of 5- and 6-carbon sugars such as arabinose, galactose and xylose forming amorphous structures; and lignin, a complex polymer of aromatic phenylpropanoids joined by α- and β- ether bonds (Saha, 2003). These components are themselves interweaved by cross-linkages to form a matrix that makes up the majority of plant matter.

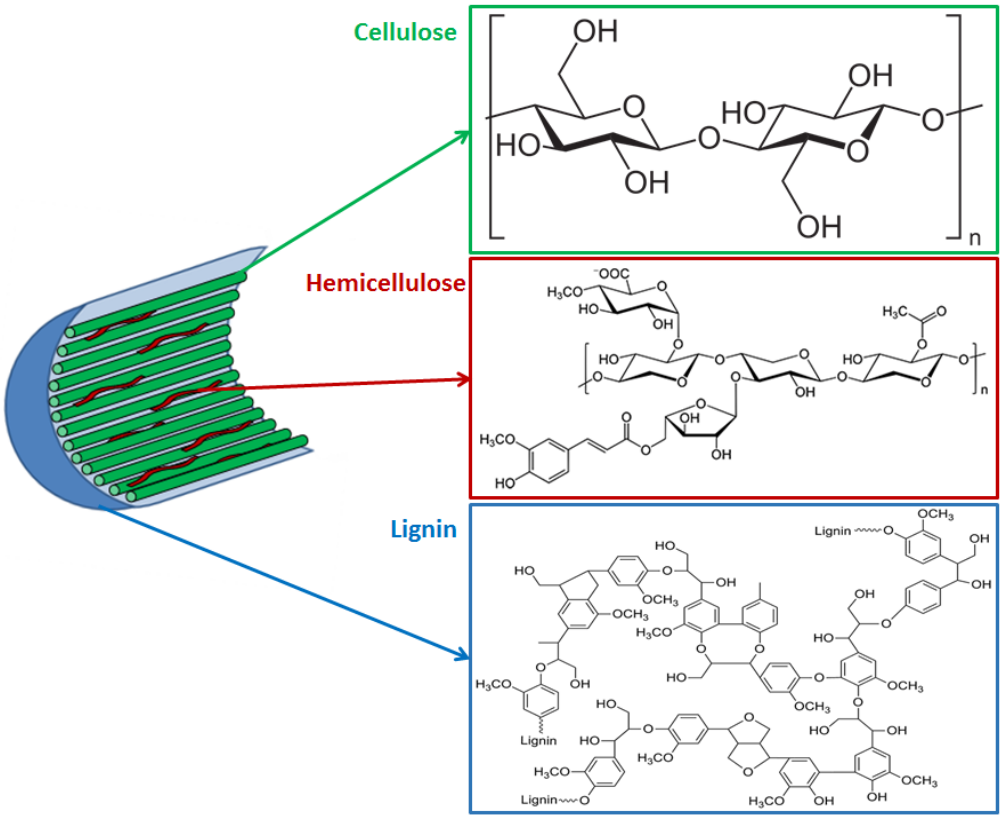


Figure 1: Outline of the organisation of lignocellulose and its three major components. The chemical structures displayed highlight the organised, regular sugar structure of cellulose; the branched sugar structure of hemicellulose; and the complex, varied aromatic structure of lignin.

Lignin's function in plants is primarily structural, providing rigidity and strength to cell walls and the plant as a whole (Valadares, 2014). This is due to its highly branched and varied 3D macromolecular structure comprised of monomers with a variety of chemical bonds formed between their functional groups, including robust ether-linkages, resulting in a highly recalcitrant arrangement. These properties also contribute to the defence of the plant against mechanical and chemical attack, with antimicrobial defence additionally exhibited (Cruz et al., 2001; Wainhouse et al., 1990). Lignin is also vital in water and solute transport; unlike the hydrophilic, sugar-based cellulose and hemicellulose, lignin is hydrophobic and as such prevents water absorption in cell walls and allows water transport in plants' vascular structures (Boyce et al., 2004).

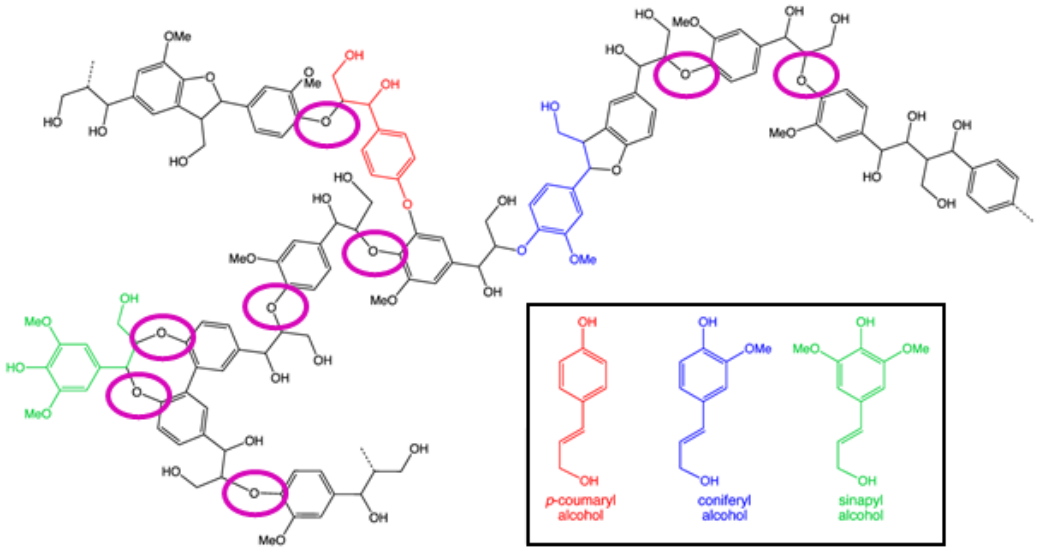


Figure 2: Chemical structure of a lignin polymer. The three monolignol-based monomers are highlighted in colour: *p*-coumaryl alcohol (H) in red, coniferyl alcohol (G) in blue, and sinapyl alcohol (S) in green. The β-aryl ether linkages are circled in pink (Adapted from https://commons.wikimedia.org/wiki/File:Lignin.png).

Lignin is a highly heterologous polymer with its composition variable. While it has no defined constitution, lignin is predominantly made up of phenylpropane subunits on a monomeric level. Three monolignol compounds – coniferyl alcohol, sinapyl alcohol and *p*-coumaryl alcohol – are assimilated in lignin biosynthesis in varying amounts dependent on plant species, the chemical structures of these are shown in Figure 2 (Agrawal et al., 2014; Vanholme et al., 2010). Gymnosperms, such as conifers, have high amounts of coniferyl alcohol (or guaiacyl, G); angiosperms, which encompass most flowering plants, have amounts of sinapyl alcohol (or syringyl, S) equal to that of guaiacyl; while lignin in grasses

contains a mixture of all three including higher than average levels of *p*-coumaryl alcohol (or *p*-hydroxyphenyl, H) (Boerjan et al., 2003; Li and Chapple, 2010). As well as these primary monolignols, alternative subunits, with substituted or additional side groups on the aromatic ring, can appear in trace amounts such as dihydroconiferyl alcohol (Ralph et al., 2004; Steeves et al., 2001).

1.1.2. Lignin on an ecological scale

As the second-most abundant source of fixed carbon, after cellulose, lignin must be degraded in order to complete recycling as part of the carbon cycle. Therefore, its recalcitrant nature must be overcome. The combustion of wood – woody plants have higher than average lignin content – achieves this to some degree and makes the most of the large amount of energy stored in lignin's chemical bonds. However, some microorganisms are capable of degrading lignin naturally and allowing the complete utilisation of lignocellulosic biomass along with the more accessible cellulose and hemicellulose components (Hara et al., 2000).

It is the breakage of the strong β-aryl ether linkages that is the major obstacle for these organisms to overcome in order to incorporate lignin into their metabolism. White-rot fungi, including *Phanerochaete chrysosporium*, *Coprinus cinereus* and *Dichomitus squalens*, secrete extracellular enzymes that possess the capability to break these bonds aerobically and release lignin-derived monomers for sequestration. The enzymes from these *Basidiomycota* species include laccases, and lignin- and manganese-peroxidases (Hatakka, 1994). These enzymes, which use oxygen and hydrogen peroxide as electron accepting co-substrates, efficiently degrade lignin when supplemented with other complementary enzymes such as H_2O_2-producing glyoxal oxidases and aryl-alcohol oxidases (Dashtban et al., 2010). Some bacteria have also been identified that produce ligninolytic enzymes that act in a similar fashion (de Gonzalo et al., 2016; Plácido and Capareda, 2015). *Sphingomonas paucimobilis* (formerly *Pseudomonas paucimobilis*) was shown to be capable of breaking the β-aryl ether linkage with identification and characterisation of enzymes responsible for this in the form of a beta-etherase, LigE, and an extradiol dioxygenase, LigZ (Masai et al., 1999; Peng et al., 1998). *Rhodococcus jostii* produces a lignin/manganese peroxide, DypB, in an encapsulin nanocompartment capable of breaking the aryl-linkage in lignin (Ahmad et al., 2011; Rahmanpour and Bugg, 2013). More recently a plethora of soil bacteria have been identified as result of screening for lignolytic activity, highlighting the importance of bacteria in ecological lignin degradation (Huang et al., 2013; Strachan et al.,

2014; Wang et al., 2016). It appears while bacteria are incapable of producing the more complex, multi-ion-containing, glycosylated peroxidases that fungi use, they possess a repertoire of coordinating and complementary enzymes capable of releasing phenylpropanoid monomers from lignin polymers (de Gonzalo et al., 2016).

1.1.3. Biotechnological uses of lignin

Lignin's primary role in biotechnology is as a feedstock with lignocellulosic biomass sourced from agricultural, paper pulping and household waste, and from specific crops such as *Miscanthus* sp. and *Populus* sp. grown on semi-arable land. While it is often the cellulosic components of this biomass that is predominantly sought as a source of sugars for fermentation into biofuels, the lignin constitutes a sizeable proportion of it and therefore offers a relatively untapped pre-existing resource in biotechnology. To overcome the recalcitrant nature of lignin, a ligninolytic pre-treatment step is required. This can be a thermochemical biomass step using acids and solvents, or a biological step using whole microbes or ligninolytic enzymes for the depolymerisation of lignin; *Sphingobium* sp. SYK-6 has proven capable of achieving bacterial pre-treatment (Linger et al., 2014; Peralta-Yahya et al., 2012). Valorisation of lignin is achievable through use as a carbon source for aromatic-degrading bacteria, and its aromatic monomers as precursors of high value phenolic-based products such as the solvents benzene and toluene and the flavourings vanillin and benzaldehyde. Furthermore, bacteria have been identified for the conversion of lignin into polyhydroxyalkanoates for bioplastic production (Numata and Morisaki, 2015). Ultimately, with the cost-effectiveness of lignocellulosic biomass-based biotechnological production constantly under scrutiny, especially with the rise in cheaper fossil fuel-based energy sources such as shale gas, maximum valorisation of all its components is essential. Utilising the aromatic building blocks of lignin in addition to the sugars of cellulose and hemicellulose is vital to achieving this.

1.2. Bacterial Solute Transport

The ability of bacteria to transport molecules across their membranes in both directions, both with and against gradients, is of the upmost importance. The bacterial cell envelope must act as a barrier from a potentially hostile external environment, while allowing selective and regulated movement of substances in and out of the cell as required. Gram-positive and Gram-negative bacteria differ in the ultrastructure of their cell envelopes; while both possess a phospholipid bilayer-based inner membrane, the latter have an additional, outer membrane comprised of a phospholipid interior leaflet and a lipopolysaccharide exterior leaflet. The presence of this Gram-negative-specific organelle creates the periplasm between the two membranes. The long established belief is that certain molecules – such as gases and small molecules – are capable of moving across these hydrophobic membranes independently; however, there is a recent paradigm shift away from this with a growing theory that all transmembrane movement of molecules is via proteinaceous bodies (Kell et al., 2015). Work pioneered by Douglas Kell has led to the argument that most xenobiotics don't passively diffuse across membranes driven by their concentration gradient and that protein transport systems are necessary and thus highly important (Dobson and Kell, 2008; Kell et al., 2011; Kell and Oliver, 2014). Furthermore, biological transport systems for an array of small, uncharged molecules previously assumed to diffuse such as carbon dioxide, dioxygen, water, urea, ammonia, hydrogen peroxide, glycerol and fatty acids, have been identified (Bienert and Chaumont, 2014; Day et al., 2014; Ishibashi et al., 2011; Itel et al., 2012; Kai and Kaldenhoff, 2014; Strugatsky et al., 2013; van den Berg, 2005; van den Berg et al., 2004; Wang et al., 2013a; Wang and Tajkhorshid, 2010). Ultimately, while the degree to which transmembrane solute movement is mediated by proteins is debated, the importance of proteins in the translocation of solutes is unquestionable.

1.2.1. Transport across the outer membrane

The outer membrane of Gram-negative bacteria is the first barrier for the uptake of solutes. As the inner membrane acts as the major permeability barrier, solutes cross the outer membrane mostly through largely nonspecific channels. These aqueous channels are formed by the "pore-like" structure of β-barrels, mostly consisting of between 8 and 24 β-strands, and are classified as porins (Wiener and Horanyi, 2011). Solutes diffuse through the channels lined with hydrophilic residues, with selectivity and the rate of diffusion derived from the physiochemical property of the molecule and the size of the porin's

aperture (Koebnik et al., 2000; Nikaido, 1993). Though there are recently characterised exceptions, most bacterial porins have the same trimeric structure and are divided into 16-stranded, nonspecific porins and 18-stranded, specific porins, as shown in Figure 3; the former involved in the uptake of mostly small, inorganic molecules and the latter the uptake of larger molecules such as sugars (Zachariae et al., 2006; Zeth and Thein, 2010). The *E. coli* porins OmpG and FadL are exceptions, with a 14-strand monomeric structures (van den Berg et al., 2004; Yildiz et al., 2006). The former appears to function as a nonspecific channel for oligosaccharides, and the latter uses a high affinity ligand binding for specific diffusion of fatty acids, and potentially phenylpropanoids (Lepore et al., 2011; Zhou et al., 2015). It is estimated that there are up to 100,000 copies of porins per bacterial cell wall, with one form of porin dominating in abundance e.g. OmpF in *E. coli* (Zeth and Thein, 2010). However, other outer membrane transport proteins have been identified, such as the polysaccharide secreting Wza, and the TonB classification of energy-driven, substrate binding outer membrane proteins such as FepA (Collins and Derrick, 2007; Ye and van den Berg, 2004).

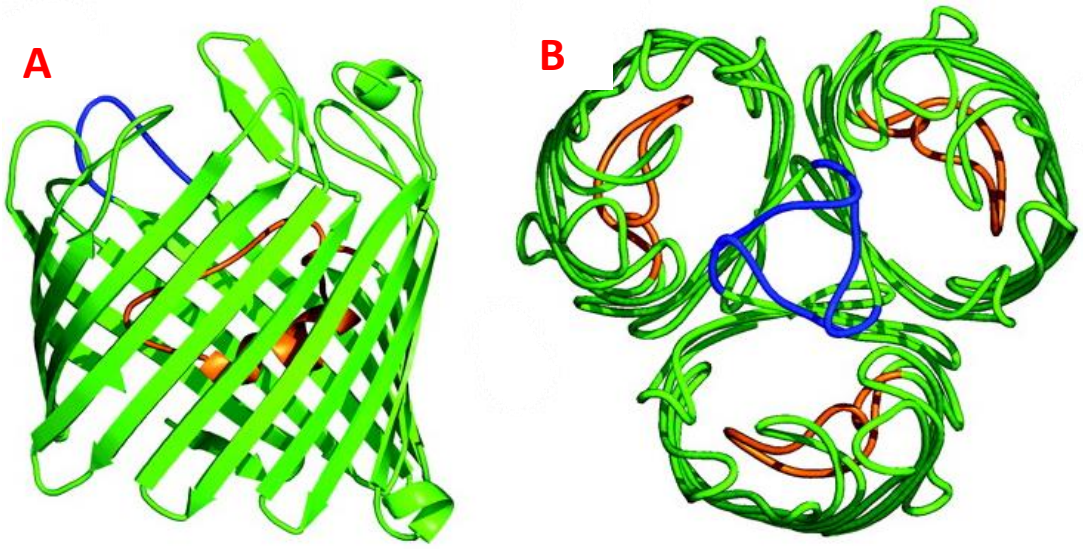


Figure 3: Structure of the *E. coli* outer membrane porin, OmpF. A) The side-on view of an OmpF monomer displaying the 16-sheet β-barrel structure. B) The top-down view with the organisation of these barrels in the trimeric structure. The loop (Loop L2) highlighted in blue is key to the interactions between monomeric subunits. The loop (Loop L3) shown in orange is responsible for constricting the opening (Taken from Nikaido et al, 2003).

1.2.2. Active transport across the inner membrane

It is vital that bacteria are capable of maintaining a selective internal environment and in many environments vital nutrients are at a lower concentration outside the bacterial cell than in. Therefore protein transporters have evolved for the required active transport of solutes against the concentration gradient with the inner membrane and its protein transport systems the major selectivity barrier. Active transport is necessary for the uptake of metabolic substrates and inorganic ions but also the efflux of unwanted and toxic substances such as antibiotics. Active transport systems are broadly classified as either primary or secondary transporters dependent on the means of energy coupling; primary transporters are driven directly by the free energy of chemical hydrolysis, while secondary transports are driven by energy derived from electrochemical gradients across the membrane such as the proton motive force (PMF) or cation gradients (Figure 4) (Busch and Saier, 2003). A third group of active transporters known as group translocators do exist in bacteria but are entirely restricted to proteins involved in the phosphoenolpyruvate-carbohydrate phosphotransferase system (PTS) for sugar uptake; a multicomponent system where uptake is coupled to enzymatic sugar phosphorylation and the energy (and phosphoryl group) is derived from phosphoenolpyruvate (Deutscher et al., 2014).

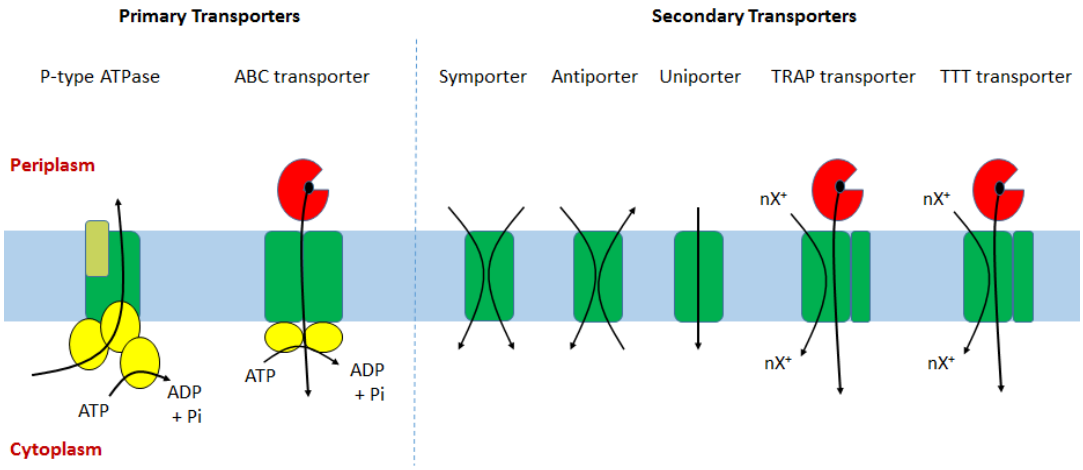


Figure 4: Schematic highlight key differences in inner membrane active transporters mentioned in Section 1.2. Primary transporters are coupled directly to ATP hydrolysis that occurs at cytoplasmic domains. Secondary transporters are couple to the proton motive force or the movement of ions down their gradient across the membrane. The substrate binding proteins are shown in red, and the nucleotide-binding domains responsible for ATP-hydrolysis are shown in yellow.

1.2.3. Primary transport: ATP-dependent and independent primary transport systems

Transport systems involved in primary transport can be further classified into two major groups based on whether they utilise the hydrolysis of ATP or not. While the vast majority are the former, bacteria do possess ATP-independent primary transporters. Examples of these are the light-driven pumps bacteriorhodopsin and halorhodopsin that are especially prominent in the Halobacteriacea class, and transporters that couple the decarboxylation of intracellular substances - such as methylmalonyl-CoA, malonate and oxaloacetate - to ion translocation (Ashwini et al., 2017; Balsera et al., 2011; Berg et al., 1997; Huder and Dimroth, 1993; Kandori, 2015).

ATP-dependent primary transporters include ATP-binding cassette (ABC) transporters, F-type ATPases, P-type ATPases and V-type ATPases. F-type ATPases (or ATP synthases) are vital in bacteria due to their function producing ATP from ADP and inorganic phosphate (Muller and Gruber, 2003; Steed and Fillingame, 2014). This occurs in the peripheral (cytoplasmic in bacteria) F_1 domain and is energetically unfavourable, requiring passive H^+ movement across the membrane through the ATPase to drive it. P-type ATPases are an extensive and ubiquitous family of ATP-hydrolysis-powered pumps (Kuhlbrandt, 2004). These ATPases are sub-classified dependent on their substrates with most involved in the transfer of ions across the membrane; the most well-characterised examples are the Na^+/K^+-, Ca^+- and H^+-pumps, with all P-type ATPases sharing sequence homology and a conserved aspartate residue that is phosphorylated, causing the conformational change required for substrate translocation (Ekberg et al., 2013). A range of transition metal ions (e.g. Cu^+, Zn^{2+}, Pb^{2+}, Ag^+) are also substrates for certain classes of P-type ATPases that are key to metal homeostasis and efflux of accumulating toxic ions (Raimunda et al., 2012; Rensing et al., 2000; Sharma et al., 2000; Wijekoon et al., 2017). P-type ATPases also include lipid flippases; ATP-driven transporters responsible for translocating phospholipids between the leaflets of bilayered membranes (Lopez-Marques et al., 2014). V-type ATPases have a similar structure and mechanism to F-type ATPases and function as irreversible H^+ and Na^+ (sometimes Li^+) translocators using ATP hydrolysis (Lolkema et al., 2003; Murata et al., 2008).

ABC transporters represent one of the largest families of membrane transporters and have structure and function distinctive to the ATPases. With diverse uptake and efflux capabilities ABC transporters are involved a wide-range of bacterial physiological processes. Unlike eukaryotic ABC transporters that are considered one polypeptide, ABC transport

systems found in prokaryotes share an ultrastructure that typically consists of two transmembrane domains (TMD) and two cytoplasmic ATP-hydrolysing domains (termed nucleotide-binding domains; NBDs) (Davidson et al., 2008). This canonical organisation is slightly varied between the four subfamilies: Exporters, Type I importers, Type II importers and Type III (energy-coupling factor, ECF) importers (Figure 5), while further variation originates from potential fusions between the TMDs and NBDs (Holland and Blight, 1999; Rice et al., 2014).

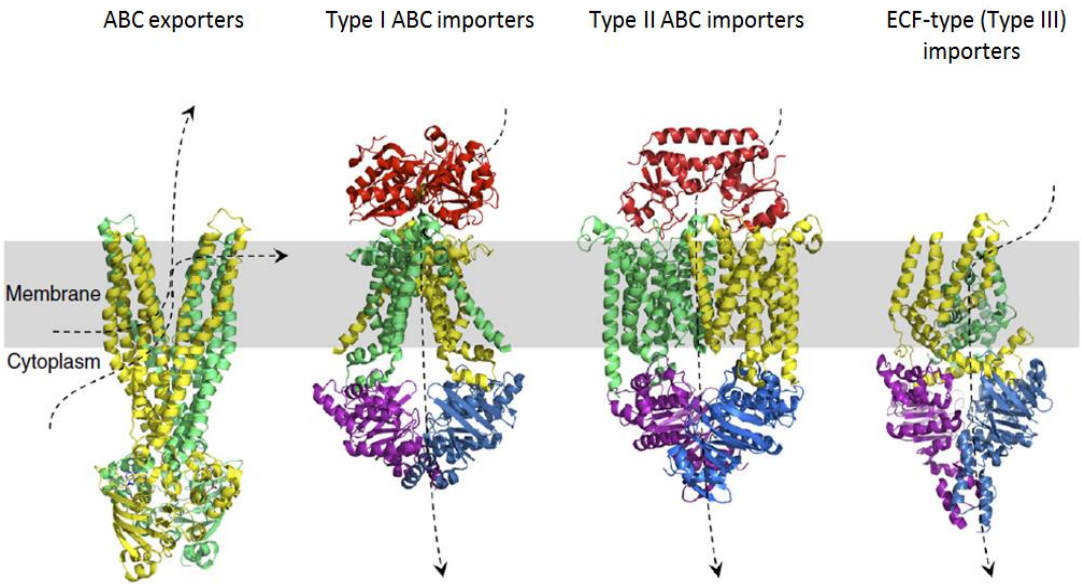


Figure 5: Schematic diagrams showing the mechanisms for ABC exporters, Type I importers, Type II importers and ECF-type (Type III) importers. The SBP is shown in red in the Type I and II importers. The TMDs are shown in green and yellow, and the NBDs in blue and purple; in exporters the NBDs are fused to the TMDs (Taken from Locher et al, 2016).

The NBDs that are highly conserved across all phylogenetic groups are the characteristic feature of ABC transport systems (Wilkens, 2015). Each NBD has an overall L-shaped fold consisting of two subdomains linked by the Q-loop and the Pro-loop (Schmitt et al., 2003). The larger subdomain is a RecA-like domain consisting of the Walker A and Walker B motifs, and the smaller is an α-helical domain that contains a conserved LSGGQ motif (Zaitseva et al., 2005). The primary structures of TMDs are much more varied, allowing for the accommodation of a wide range of substrates. This variation extends to the number of transmembrane helices in the domains with the number mostly consistent within a subfamily (as displayed in Figure 5): for example, Type I importer TMDs are mostly two homodimers (or are at least structurally similar) of six helices each; while with Type II importers the TMDs consist of ten helices each, and exhibit two-fold symmetry (ter Beek et

al., 2014). Nevertheless, the overall topology is a typical transmembrane cavity capable of alternating between which side of the membrane it is accessible from.

These pairs of domains work cooperatively to provide a mechanism for translocation of substrates from one side of the membrane to the other. The substrate binds between the TMDs in the open face, before a conformational change is induced by the interaction of the NBDs with ATP that results in the cavity being open to the opposite side of the membrane. There are two theories as to the nature of the ATP-induced "power stroke" conformational change with evidence that it is either hydrolysis of two molecules of ATP by the NBDs, or ATP-binding to the NBDs that causes them to dimerise, with this motion transferred to the TMDs (Higgins and Linton, 2004; Khare et al., 2009; Orelle et al., 2010).

Most ABC transporters involved in uptake possess an extracytoplasmic solute binding protein (SBP, or periplasmic binding protein, PBP) located in the periplasm of Gram-negative bacteria, and often anchored to the external surface of the membrane in Gram-positive bacteria via an N-terminal lipid moiety (Scheffel et al., 2004; Sutcliffe and Russell, 1995). These SBPs all have a similar structure of two-domains connected by a hinge linker that forms a binding pocket for the substrate. This binding site confers a high affinity (i.e. even sub nanomolar K_d) and specificity for the relevant substrate, allowing the SBP to act as an extra cytoplasmic scavenger and the transporter to translocate substrates present at extremely low concentrations, and additionally form a bound-substrate reservoir in the periplasm ready for uptake (Eitinger et al., 2011; Erkens and Slotboom, 2010). Binding induces a conformational change and the two domains close around the substrate in what is described as a "venus flytrap" mechanism (Berntsson et al., 2010).

These soluble proteins offer a relatively easy target for characterisation and so a huge number of structures have been deposited in the Protein Data Bank, with five topologically different clusters described for ABC SBPs (Figure 6) (Berntsson et al., 2010). The distinguishing feature of each of these clusters is, typically, the hinge between the two domains. For example the SBP of the *E. coli* vitamin B_{12} ABC importer, BtuF, belongs to cluster A and possess an α-helical hinge that confers a rigid overall structure (Hvorup et al., 2007). In comparison, the maltose-binding protein, MBP, belongs to cluster D, with a much smaller hinge of two 4-5 amino acid long sequences, and a small additional subdomain (Bertz and Rief, 2009). Figure 6 clearly highlights these differences in the hinges that define each cluster.

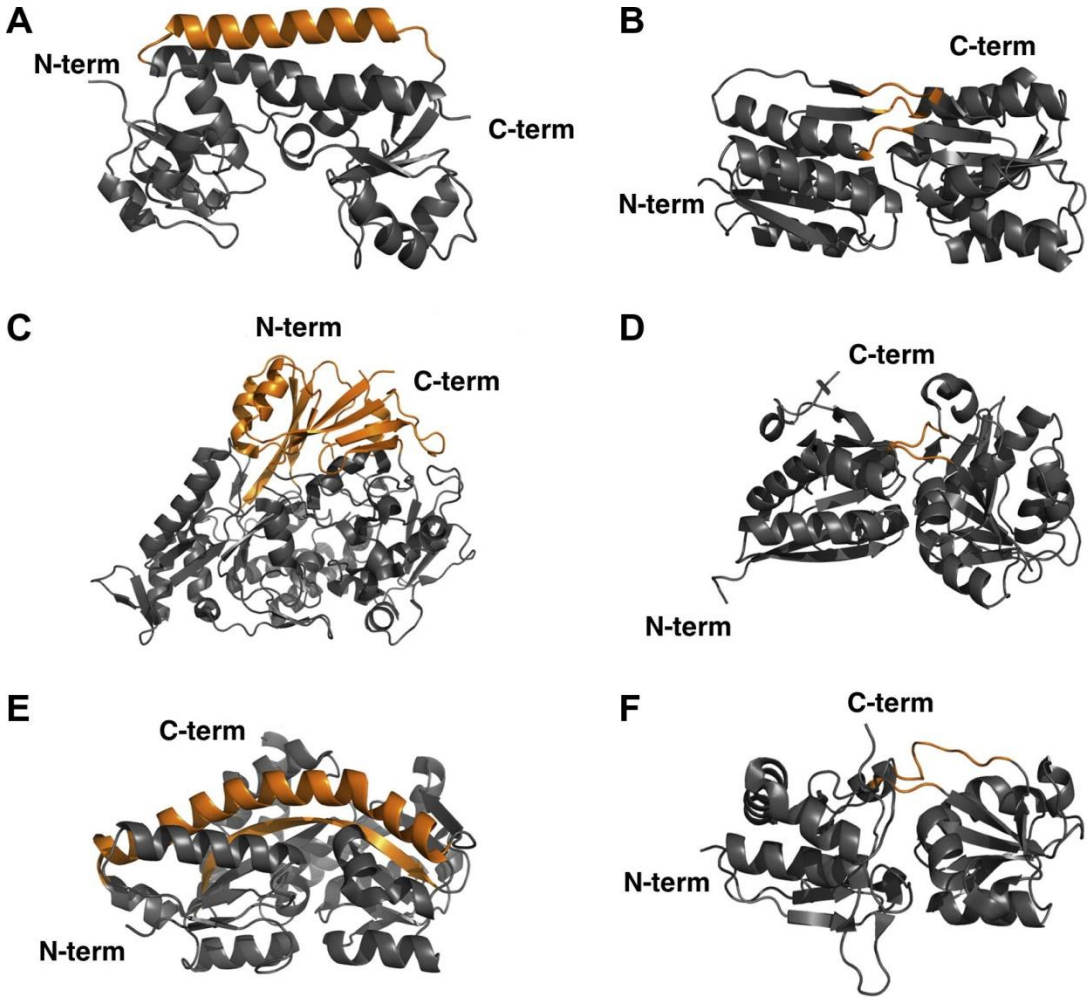


Figure 6: The clusters of SBPs classified based on structure as defined in Berntsson et al, 2010 (Berntsson et al., 2010). A-D and F are SBPs from ABC transporters, and E is from TRAP transporters. The linkages are the key differential between clusters and are highlighted in colour.

Further to the high affinity they provide to the transport system, SBPs play an integral role in the mechanism for uptake. Studies on the *E. coli* maltose import system MalFGK$_2$-E, a Type I importer, have shown that maltose binding enables the SBP MalE to interact with a periplasmic loop of TMD MalF and cause a conformational change necessary for NBD dimerisation, in addition to the cytoplasmic ATP binding (Jacso et al., 2009). This subsequently alters the TMD in the externally open conformation and reduces the affinity for maltose of MalE. Maltose dissociates and binds in the membrane, causing ATP hydrolysis and a final conformational change that exposes maltose to the cytoplasm (Chen, 2013; Oldham and Chen, 2011). A similar mechanism is shown with the aforementioned BtuF that relays binding of vitamin B_{12} to NBD dimerisation via coupling helices in the TMDs.

SBPs are not always necessary to uptake mechanisms, as is the case with the ECF class of ABC transporters that lack periplasmic SBPs entirely (Slotboom, 2014).

ABC transporters play important roles in bacterial physiology defined by the substrates they translocate. They are required for the uptake of carbohydrates and cofactors for metabolism, as is the case for the maltose and vitamin B_{12} uptake systems, with the high affinity scavenging of the SBPs allowing bacteria to utilise scarce compounds as carbon sources. They can also be involved in iron homeostasis with the Ybt and Irt systems of the pathogens *Yersinia pestis* and *Mycobacterium tuberculosis*, respectively, necessary for growth in iron-limiting conditions due to their function as siderophore importers (Bobrov et al., 2014; Rodriguez and Smith, 2006). ABC importers also have a role in response to stress, with upregulation of the Ehu and Ous systems to increase the uptake of the compatible solutes ecotine and glycine betaine in *Sinorhizobium meliloti* and *Erwinia chrysanthemi*, respectively, when faced with osmotic stress (Choquet et al., 2005; Jebbar et al., 2005).

ABC exporters play a vital role in antimicrobial resistance and have evolved to proficiently efflux toxic compounds. The acridine resistance complex of *E. coli*, AcrAB, is a model example of this with homologous examples in most Gram-negative bacteria (Du et al., 2014). AcrAB combines with the outer membrane channel TolC to transport a wide range of antimicrobial compounds in addition to acridines, such as tetracycline, ampicillin and chloramphenicol (More detail in 1.5.4) (Baucheron et al., 2004; de Cristobal et al., 2006). Overexpression of such efflux systems are common features of multidrug resistant strains of bacteria.

1.2.4. Secondary transport: Symporters, antiporters and uniporters

Secondary transporters are generally simpler than ATP-dependent transporters, often encoded by just one gene; the exceptions being the tripartite families discussed in 1.2.5 and 1.2.6. These single component systems are principally members of the major facilitator superfamily (MFS), and typically have a lower affinity but higher V_{max} than their SBP-utilising counterparts (Chen and Beattie, 2008; Law et al., 2008). They share structural homology, mostly consisting of two near-symmetrical domains of six transmembrane helices with some larger proteins consisting of four of these domains (Boudker and Verdon, 2010). The organisation of these structural repeats suggests gene fusion or duplication was responsible for the evolution of MFS transporters from the earliest, more basic transporters.

These transporters can be divided into three groups based on their mechanism of action, though all use energy from a pre-existing gradient across the membrane. Symporters transport solutes against their concentration gradient, simultaneously transporting an ion in the same direction with its electrochemical gradient as an obligatory coupled source of energy. A prominent and well-studied example is the *E. coli* lactose permease LacY that acts as a proton/galactoside symporter and exhibits the rocker-switch mechanism for translocation (Chaptal et al., 2011; Kaback et al., 2011). This entails a periplasmic H^+ binding the protein in its resting state where the hydrophilic cavity is open towards the cytoplasm; the residues vital to protonation of the protein are mostly located in the C-terminal domain. The movement of the H^+ through the protein releases free energy that induces a conformational change whereby the hydrophilic cavity is opened to the periplasm, allowing the substrate, i.e. lactose, to bind. Substrate binding subsequently induces a second conformational change back to the original orientation, allowing release into the cytoplasm. Interestingly movement of the galactoside substrate down its gradient can result in uphill H^+ translocation (Guan and Kaback, 2006). More recently, an alternative alternating access mechanism has been discovered (Reyes et al., 2009); the elevator-type mechanism involves one domain exhibiting a vertical translocation perpendicular to the membrane and has been shown to be utilised by Na^+/succinate and Na^+/citrate transporters in *Vibrio cholerae* and a Na^+/citrate transporter in *Klebsiella pneumoniae* (Kim et al., 2017; Mulligan et al., 2016).

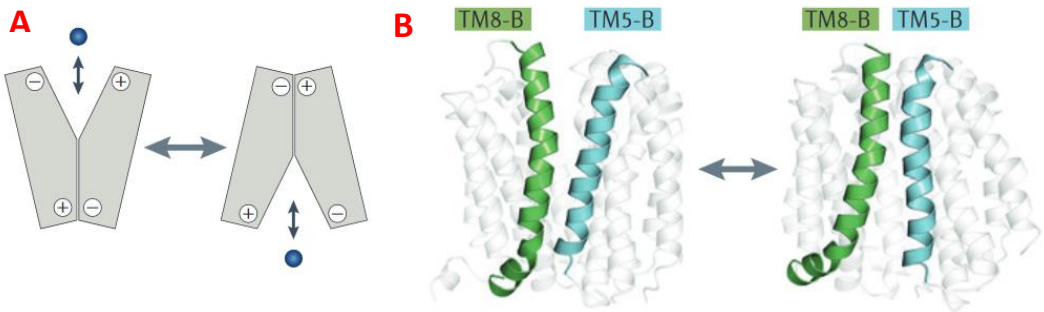


Figure 7: The alternating access mechanism of MFS family transporters (Quistgaard et al., 2016). A) The basic conformational shift in the transporter on substrate binding with ionic interactions between each subunit. B) The change in position of two key helices in the lactose permease LacY between the outward and inward open states.

Antiporters work in a similar fashion employing the alternative access mechanism, but with movement of the solute and the coupled ion (or secondary substrate) occurring in opposite directions (Lee et al., 2013). Ligand binding is essential for translocation with the charge of

the substrate interacting with residues of the protein to cause the required conformational change (Gao et al., 2009; Hunte et al., 2005). Their functions vary, including protection against electrophiles as with the K^+/H^+ antiporter KefB that effluxes potassium ions in order to acidify the cytoplasm, and protection against acidic environments as with the arginine/agmatine antiporter AdiC that responds to pH changes and works to produce a net efflux of protons (Iyer et al., 2003; Roosild et al., 2010).

The physiological role of antiporters can also be the simultaneous import of a metabolic substrate and export of the unwanted product resulting from utilisation of the substrate, thus allowing maintenance of a desirable ratio of substrate:product to expedite cytoplasmic conversion. An example of this in *E. coli* is the DcuB antiporter that imports fumarate for use as a terminal electron acceptor during anaerobic growth and exports the product of fumarate reduction, succinate; both dicarboxylic acids have a charge of $^{2-}$ and so the process is electroneutral (Kleefeld et al., 2008). The electrogenic transport of histidine/histamine, malate/lactate and oxalate/formate antiporters coupled to proton-consuming enzymatic conversions is used to generate a PMF for ATP production (Uzdavinys et al., 2017; Wolken et al., 2006).

Uniporters are responsible for the transport of a single solute across the membrane and as such are the most simplistic secondary transporters. Typically transport is driven by the membrane potential with the negatively charged interior driving movement of positively-charged ions into the cytoplasm against their concentration gradient. Examples in bacteria include Ca^{2+} and K^+ ion transporters and the *E. coli* protein AmtB that exhibits ammonia-specific transmembrane movement (Callaway et al., 1999; Kusaka and Matsushita, 1987; Nygaard et al., 2006); however, overall there have been scarce recent characterisations of uniporters.

1.2.5. Tripartite ATP-independent periplasmic (TRAP) transporters

A distinct family of homologous transporters has been identified that shares traits with both ABC and secondary transporters. These systems consist of three subunits: a periplasmic SBP, a larger membrane-protein subunit and a smaller membrane-protein subunit; in some cases these two integral membrane subunits are fused as one (Fischer et al., 2010). Like ABC systems they utilise the substrate-specific binding of the SBP in the periplasmic as a means of selectivity, however they completely lack ATP-binding domains and uptake has been shown to be driven by Na^+ ions (Kelly and Thomas, 2001; Mulligan et al., 2009. Their uneven-sized membrane proteins (as well as the SBP) distinguish them from

other secondary transporters, with the larger of the two domains composed of twelve transmembrane helices, and the smaller four (Wyborn et al., 2001). While the larger domain acts as the transmembrane solute carrier or permease, the function of the latter is unknown.

The first TRAP transporter was identified in the photosynthetic purple bacterium *Rhodobacter capsulatus* and TRAP transporters are found in many bacteria and archaea but are absent in eukaryotes (Forward et al., 1997; Mulligan et al., 2011). The SBP of this transport system, DctP, was characterised first by showing binding of the dicarboxylic acids malate, succinate and fumarate; these substrates were indicative of the role of TRAP transporters as they are almost ubiquitously involved in the uptake of organic acids. Following identification of the SBP it was noted there were no genes encoding NBDs in the neighbouring genomic region as one would expect with an ABC system, only genes for two disparately sized membrane proteins designated DctQ (smaller) and DctM (larger). With no apparent coupling to ATP hydrolysis, studies were conducted using protonophores such as 2,4-dinitrophenol and carbonyl cyanide-4-(trifluoromethoxy) phenylhydrazone (FCCP) that showed collapsing the proton motive force inhibited TRAP-mediated uptake (Chae and Zylstra, 2006; Forward et al., 1997).

Since identification of the initial DctPQM system, many more TRAP transporters have been identified in bacteria and archaea covering transport of a range of substrates including four-carbon dicarboxylic acids, monocarboxylic acids, neuraminic acids, amino acids and compatible solutes like ectoine, with the only known substrate of a TRAP transporter that doesn't possess a carboxylic acid moiety being taurine (Denger et al., 2006; Lecher et al., 2009; Mulligan et al., 2007). Many of these species have been isolated from marine, and other "nutrient-poor", environments where organic compounds are apparently in low concentrations; this also explains the use of Na^+ as a coupling ion.

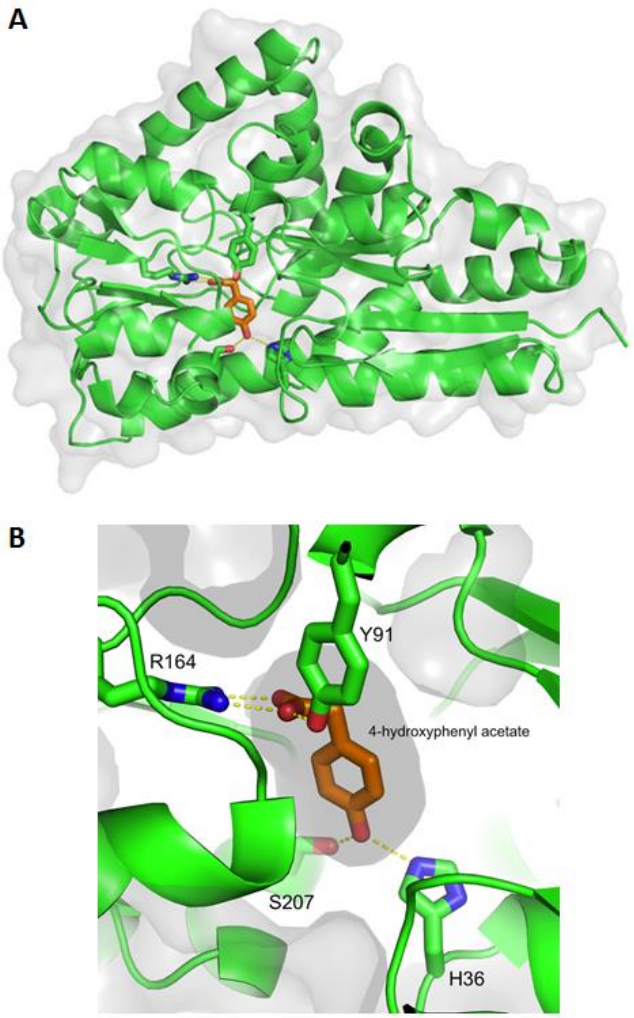


Figure 8: Structure of a TRAP SBP determined by x-ray crystallography. A) Overall structure of the SBP from the *Rhodopseudomonas palustris* TRAP transporter TarPQM showing the ligand, hydroxyphenylacetate, in the binding pocket. B) The positioning of the ligand in the binding pocket highlighting the salt being formed between the carboxylic acid moiety and the conserved arginine residue (R164) (Figures from personal correspondence with D. J. Kelly).

As with ABC transporters, substrate specificity is conferred by high-affinity ligand-binding by the SBP. While there is little sequence homology between the SBPs of TRAP and ABC systems (and in fact amongst TRAP SBPs themselves) the structural fold is similar, with TRAP SBPs also employing the "venus flytrap" mechanism described previously (Section 1.2.3). The expansive characterisation of TRAP systems has recognised two distinct families of SBP: the DctP-type family consisting of orthologues of the initial DctPQM system, and the TRAP-associated extracytoplasmic immunogenic (TAXI) family (Kelly and Thomas, 2001; Mulligan et al., 2011). All TAXI SBPs are associated with a fused DctQM, and all TRAP

systems in archaea utilise TAXI-family SBPs. Both families are similar sizes (300-400 residues), employ N-terminal signal sequences and employ a conserved arginine residue shown to form salt bridge with the carboxylic acid moiety of the ligand on binding (Fischer et al., 2010).

Recently electron paramagnetic resonance spectroscopy has revealed that TRAP SBPs only transition between two states, open-unliganded and closed-liganded (Glaenzer et al., 2017). Both forms have been shown to interact with the TMDs and initiate the conformation changes for translocation, with the unliganded state interacting to trigger efflux; thus displaying the bidirectional potential of TRAP systems, though only in presence of an excess of unliganded SBP (Mulligan et al., 2009). In two studies coordinating ions have been identified in the substrate-binding pocket of DctP-family SBPs; a sodium ion with TakP and calcium ion with LacP, both of which operate as dimers. However, these cations were shown to be unessential for substrate binding (Akiyama et al., 2009; Gonin et al., 2007).

1.2.6. Tripartite tricarboxylic transporter (TTT) family

A third family of SBP-dependent transports has been identified in bacteria that, like TRAP transporters exhibit secondary transport. The tripartite tricarboxylic transporters (TTT) were first characterised in *Salmonella typhimurium* in the form of the TctABC system (previously shown to bind the tricarboxylic acids citrate and isocitrate), and have a similar structural organisation to TRAP transporters – two transmembrane proteins and a periplasmic SBP – but share no sequence homology (Sweet et al., 1984; Winnen et al., 2003). TTT proteins are relatively understudied compared to other transport classes with few characterised, though they were found to be vastly overrepresented in the Gram-negative pathogen *Bordetella pertussis* (Antoine et al., 2003). In this bacterium two pairs of genes encoding TTT membrane proteins were identified based on sequence, and 90 encoding TTT SBPs. The occurrence of SBPs with no associated membrane proteins – termed ‘orphans’ – is a common feature of the TTT family, and could contribute to the necessary uptake of a diverse range of solutes or be involved in chemotaxis and signalling (Piepenbreier et al., 2017). The citrate-binding SBP of the BctCAB system in *B. pertussis* was shown to be involved in both citrate uptake and direct regulation of citrate operon (Antoine et al., 2005).

1.2.7. Exploitation of membrane transporters

The study of bacterial transport systems garners much interest due to the potential applications in the biotechnological and pharmacological industries. The use of bacteria as whole-cell biocatalysts is expanding rapidly with the majority of efforts to increase yield focused on intracellular metabolic processes with the potential involvement of transporters historically somewhat neglected. However, it is clear that discovering novel transporters for relevant substrates and hijacking them to increase the import of precursors and export of the commercially viable products will help drive metabolic flux in a desirable direction (Kell et al., 2015). Bacterial gene products are relatively poorly characterised functionally – around 25% (Genee et al., 2016); therefore research into bacterial transporters offer ways to enhance industrial applications. Identification of uptake systems for polluting hydrocarbons demonstrate the potential for bacteria to be used as biocatalysts in bioremediation and recombinant expression of a *P. putida* transporter in *E. coli* improved whole-cell biotransformation capabilities by facilitating uptake of the hydrophobic fatty acid methyl esters (Hua et al., 2013; Julsing et al., 2012). Efflux systems are vital to increasing the yield of the products, such as with protein secretion in *Streptomyces* sp. and microbial biofuel production with the directed evolution of native AcrB in *E. coli* for the excretion of biofuels such as octane and pinene (Anne et al., 2014; Foo and Leong, 2013). Accumulation of many commercially viable products in bacteria often gives rise to toxicity – especially in recombinant strains not acquitted to certain metabolites – with efflux systems vital to alleviating this while simultaneously improving accessibility of the product (Fisher et al., 2014).

1.3. Aromatic Metabolism

1.3.1. Aerobic aromatic metabolism in bacteria

Bacteria encounter a range of aromatic compounds in the environments they survive in, and some have evolved the ability to utilise them in their metabolism with both aerobic and anaerobic pathways identified. The major energetic "stumbling block" in aromatic degradation is the cleavage of the carbon ring, and aerobically bacteria can employ molecular oxygen for this step with oxygenases the key enzymes in the aerobic pathway. These enzymes, which include flavoprotein monooxygenases, Rieske iron dioxygenases and di-iron monooyxgenases, break a carbon-carbon bond in the ring via hydroxylation (Pérez-Pantoja et al., 2010).

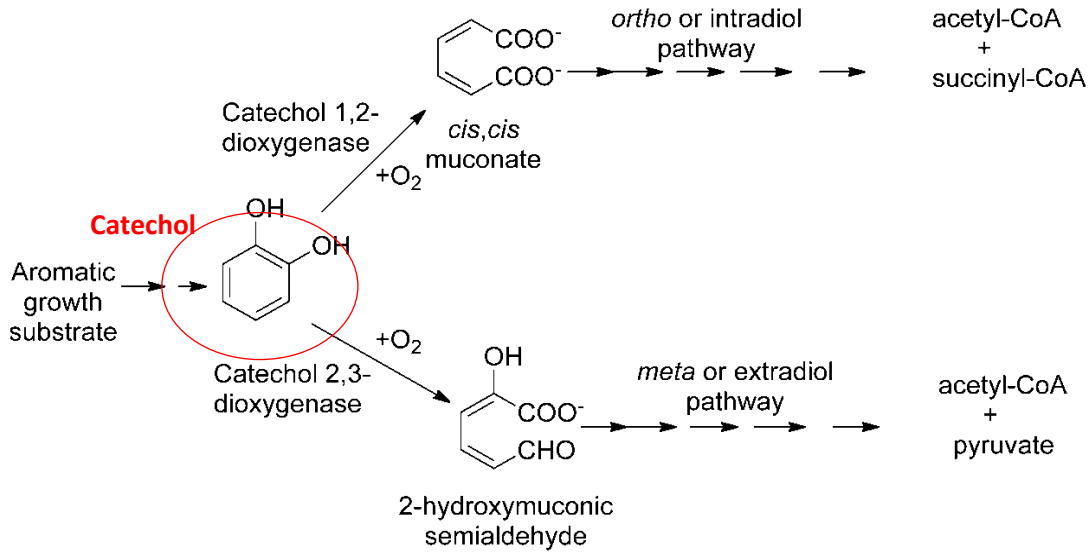


Figure 9: Basic process of aerobic aromatic metabolism by bacteria. Peripheral pathways convert a multitude of aromatic growth substrates to the central metabolite, catechol. Molecular oxygen and a non-haem Fe^{3+} complex are utilised by the key enzymes, catechol dioxygenases, to cleave the aromatic ring. An aliphatic dicarboxylic product such as *cis,cis*-muconate can then be readily metabolised through a pathway such as the β-ketoadipate pathway (Taken from Singh et al, 2015).

Due to the diversity of aromatic compounds encountered, bacteria utilise an array of peripheral pathways to "funnel" these compounds into central intermediates for catabolism. In Figure 9 catechol is this intermediate with catechol and substituted catechols, such as protocatechuate, commonly used by bacteria and the substrate for this ring-cleavage by oxygenases. Further metabolic pathways can then utilise the ring-cleavage products to produce acetyl-CoA to feed into the tricarboxylic acid cycle and other

carbohydrates for biosynthesis. Coastal marine bacteria of the *Roseobacter* genus have several gene clusters containing the *pca* genes that encode enzymes for the complete metabolism of protocatechuate to succinyl-CoA and acetyl-CoA via the β-ketoadipate pathway (the *ortho* or intradiol branch in Figure 9); *pcaGH* encode the non-haem dioxygenase for ring cleavage (Alejandro-Marin et al., 2014).

1.3.2. Anaerobic aromatic metabolism in bacteria

Bacterial anaerobic aromatic metabolism adopts a similar overall schematic, with peripheral pathways feeding into a central metabolite that undergoes ring cleavage (Fuchs, 2008). This is most commonly benzoyl-CoA, with the formation of CoA derivatives of the aromatic substrate required and performed by ATP-dependent CoA ligases (Porter and Young, 2014). Reduction of this central aromatic intermediate is the key and unique step of the anaerobic pathway. The rings of benzoyl-CoA derivatives are mechanistically difficult to reduce, unlike alternative intermediates such as resorcinol, which possess keto/enol moieties that weaken the ring structure. For benzoyl-CoA a combination of an electron donor, ferredoxin, and a benzoyl-CoA reductase are required, with energy input in the form of ATP; the presence of the CoA thioester facilitates this reduction by significantly lowering the redox potential of electron transfer (Carmona et al., 2009).

Genome sequencing and functional assignment has identified several proteobacteria that use benzoyl-CoA degradation pathways, with the genes encoding the enzymes forming distinct clusters. These include strains of α-proteobacteria, such as *Rhodopseudomonas palustris* and *Magnetospirillum magneticum*; β-proteobacteria, such as *Thauera aromatica* and *Aromatoleum aromaticum*; and δ-proteobacteria, such as *Geobacter metallireducens* and *Syntrophus aciditrophicus* (Carmona et al., 2009). The outlines of the benzoyl-CoA pathways used by two of these are shown in Figure 10A, with both initiated with a benzoyl-CoA ligase and reduction by a four-subunit reductase facilitated by a reduced ferredoxin; for *R. palustris* these enzymes are BadDEFG and BadB, respectively. Both pathways then undergo a second, subsequent reduction of the aromatic ring, but this step is performed differently and the pathways diverge. The alicyclic compounds are then hydrated, and the ring subsequently cleaved; the enzyme responsible for this in *R. palustris* is the 2-ketocyclohexanecarboxyl-CoA hydrolase, BadI (Pelletier and Harwood, 2000). Figure 10B displays the gene cluster found in *R. palustris* containing the complete repertoire of *bad* genes responsible for the enzymes of this pathway, the MarR-family transcriptional repressor, BadR, that controls expression of the CHC operon, and the Rrf2-family

transcription factor, BadM, that controls expression of the Bad operon (Hirakawa et al., 2015). Through this process the six-carbon ring of an aromatic compound results in the CoA-derivative of the seven-carbon dicarboxylic acid, pimelyl-CoA, which can feed into a downstream pathway for conversion to acetyl-CoA; this is described in more detail for *R. palustris* in Section 1.6.

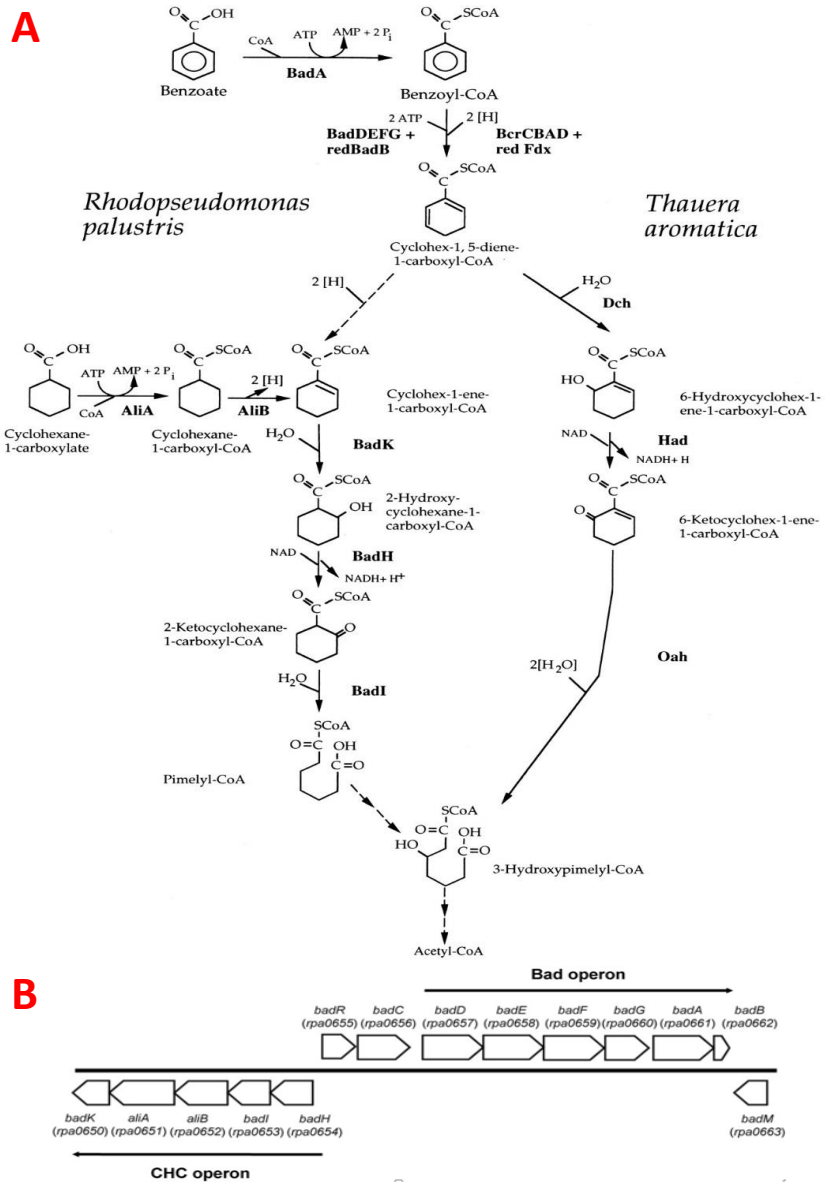


Figure 10: Bacterial benzoyl-CoA reduction and ring cleavage pathways. A) The pathways for ring cleavage of model aromatic compound benzoate via benzoyl-CoA in *R. palustris* and *T. aromatica.* B) The gene cluster from *R. palustris* encoding the complete set of enzymes required. (Figure taken from Pelletier and Harwood, 2000; and Hirakawa et al, 2015)

.

1.4. Microbial Production of Vanillin

1.4.1. The commercial potential of vanillin

Vanillin (IUPAC name: 4-hydoxy-3-methoxybenzaldehyde) is an aromatic organic compound that is the principal organoleptic component of vanilla, a flavouring with a global demand and significant commercial value that is naturally sourced from the curing of the beans (or pods) of the *Vanilla planifolia* and *Vanilla tahitensis* plants. Vanilla extract itself is a complex mix of over 200 chemicals with a composition of mostly phenols and alcohols. While the phenolic aldehyde vanillin provides the distinctive sweet vanilla aroma, other chemicals such as vanillic acid, vanilyl alcohol, hydroxybenzaldehyde and lactones produce various nuances in the overall flavour (Baqueiro-Peña and Guerrero-Beltrán, 2016; Brunschwig et al., 2016). However, it is purified vanillin that is sought for its use as an additive in numerous industries. It's wide use in food, beverages, cosmetics and pharmaceuticals results in a huge demand across the world with annual global sales of vanillin estimated to be over 20,000 tonnes in 2015, with 2% of food products and 3% of beverages launched in the 12-months up to September 2016 advertised as vanilla-flavoured, thus making it a notable contributor to a global flavours and fragrances market expected to exceed 40 billion dollars by 2023 (Fache et al., 2016; Global Market Insights, 2016). The extensive labour required for harvesting and extraction of vanillin from beans and pods make it the second most expensive spice by weight in the world after saffron (Odoux and Grisoni, 2010; Walton et al., 2003). The direct consequence of this is that only 0.2% of vanillin is currently sourced naturally by isolation from vanilla pods, which are mostly cultivated in Madagascar (accounting for ~80% of natural vanillin production). The rest of the world's vanillin market is fulfilled by chemical synthesis.

Vanillin can be synthesised chemically through various mechanisms. Vanillin has been synthesised from the phenylpropanoid eugenol, which is found at relatively high concentrations in the essential (aromatic) oils of plants such as clove, cinnamon and nutmeg, since the 19th century (Bullerman et al., 1977; Cortés-Rojas et al., 2014; Lampman et al., 1977). This requires a two-step process of alkaline isomerisation (by potassium hydroxide) and subsequent oxidation (by nitrobenzene). Another aromatic compound can be used as a precursor; vanillin can be synthesised from 4-hydroxybenzaldehyde by electrophilic bromination and copper-catalysed methoxylation. Guaiacol produced either from lignin or from hydrolysis of catechol grew in popularity as a precursor in the 20th century (Li and Rosazza, 2000). Synthesis from guaiacol by condensation with glyoxylic acid,

oxidation and decarboxylation is estimated to account for 80 – 90% of synthesised vanillin. Vanillin can also be produced directly from kraft lignin by alkaline oxidation, allowing for production from industrial waste (Fargues et al., 1996). However, the huge amount of pollutants produced in the process makes it somewhat undesirable e.g. 160 tons of caustic waste per ton of vanillin produced.

These environmental concerns, combined with the continually growing consumer demand for "naturally sourced" products, have resulted in a shift towards naturally produced vanillin across the market (Ma and Daugulis, 2014). Major companies in the sector such as Nestlé, Kraft and a range of high-street restaurant brands have recently moved to be completely free of artificial flavours, a move that encompasses a necessity for natural vanillin. This demand has seen prices rocket to over $1500/kg for natural vanillin compared to $10-20/kg for the chemically synthesised product (Evolva, 2017). Furthermore, this demand cannot be met by sourcing from vanilla pods alone, especially so with recent poor harvests and profiteering hoarders in Madagascar. Legal regulations in both Europe (EG 1334/2008) and the USA (Code of Federal Regulation, Title 21) define the term natural in regards to foods, flavourings and cosmetics in a similar fashion, specifically *compounds extracted by physical, enzymatic or microbiological processes* (Krings and Berger, 1998; Muheim and Lerch, 1999). Therefore, the biotechnological exploitation of the microbial conversion of wide range of substrates into biovanillin is a key area of interest (Gallage et al., 2014; Kaur and Chakraborty, 2013; Ni et al., 2015).

1.4.2. Microbial species used as platforms for the biotransformation of vanillin

Vanillin is a naturally occurring secondary metabolite in plants and a catabolic intermediate in various aromatic degradation pathways in bacteria, therefore numerous existing enzymatic pathways exist to be utilised (Gallage et al., 2014; Hua et al., 2007a; Masai et al., 2002; Overhage et al., 2003). These pathways use a variety of precursors including aromatic compounds such as eugenol, isoeugenol, hydroxybenzaldehyde, ferulate and tyrosine, and simple carbon sources such as glucose, xylose and glycerol (Ni et al., 2015). A variety of microbial species are also being exploited, including both species that harbour these pathways endogenously and engineered recombinant strains.

The first patent in this area was a process using Enterobacter, *Klebsiella*, and *Serratia* sp. for conversion of eugenol/isoeugenol to vanillin was filed in 1990, and the 21st century has seen several commercial applications of biotransformed vanillin reach the market (Leffingwell and Leffingwell, 2015; Rabenhorst and Hopp, 1991):

- *Solvay* produce Rhovanil® from ferulate using a fermentative approach;
- *De Monchy Aromatics* produce vanillin from curcumin sourced from turmeric;
- *Mane* produce Sense Capture™ Vanillin from eugenol;
- *Evolva* produce vanillin using an extensive pathway from glucose;
- *Conagen* and *Givaudan* have a joint venture with a patent for producing vanillin from *p*-coumarate, via ferulate;
- *BASF* produce vanillin from ferulate using *Pseudomonas* strains; and
- *Shanghai Apple Flavor and Fragrance* produce vanillin from ferulate using *Streptomyces* strains.

1.4.3. Microbial production of vanillin from ferulate

The majority of biotransformation processes use ferulate as a precursor, even though simple carbon sources (e.g. glucose) offer much cheaper alternatives. This is mostly due to the relatively simple pathway required for the conversion of ferulate into vanillin due to the structural similarities, and ferulate's high abundance in the cell wall of plants as a derivative of both lignin and hemicelluloses such as the arabinoxylan of *Miscanthus*. As mentioned previously various soil microbes metabolise ferulate as part of aromatic degradation, providing numerous opportunities for exploitation (Section 1.3). Such is the interest that Zamzuri et al., 2013, have developed a method for rapidly screening bacterial strains for their ability to produce vanillin from ferulate (Zamzuri et al., 2014).

Multiple pathways for conversion have been identified in microbes differing by the initial step. The majority of ferulate metabolising organisms use one (or in some cases more than one as exhibited by *Pseudomonas fluorescens*) one of five potential pathways: CoA-dependent retro-aldol reduction, CoA-independent retro-aldol reduction, CoA-dependent β-oxidation, non-oxidative decarboxylation, and reductive pathway via vanillic acid (Gallage and Moller, 2015). Both the CoA-dependent and independent retro-aldol reduction pathways are particular desirable for biotechnological exploitation as they are relatively simple processes requiring two and one enzymes respectively (Figure 11A and B) (Gallage et al., 2014). The cofactor can be used to activate ferulate in a reaction catalysed by a 4-hydroxycinnamate-CoA ligase (also labelled feruloyl-CoA synthetase, Fcs) (Plaggenborg et al., 2003; Yang et al., 2013). The resultant CoA-thioester feruloyl-CoA is then converted to vanillin via a transient intermediate, 4-hydroxy-3-methoxyphenyl-β-hydroxypropionyl-CoA, in a two-step reaction by a single enzyme, an enoyl-CoA hydratase/lyase (Ech) (Plaggenborg et al., 2003; Plaggenborg et al., 2001). The CoA-independent pathway proceeds via a

hydration step and the transient intermediate 4-hydroxy-3-methoxyphenyl-β-hydroxypropionic acid (Batista, 2014).

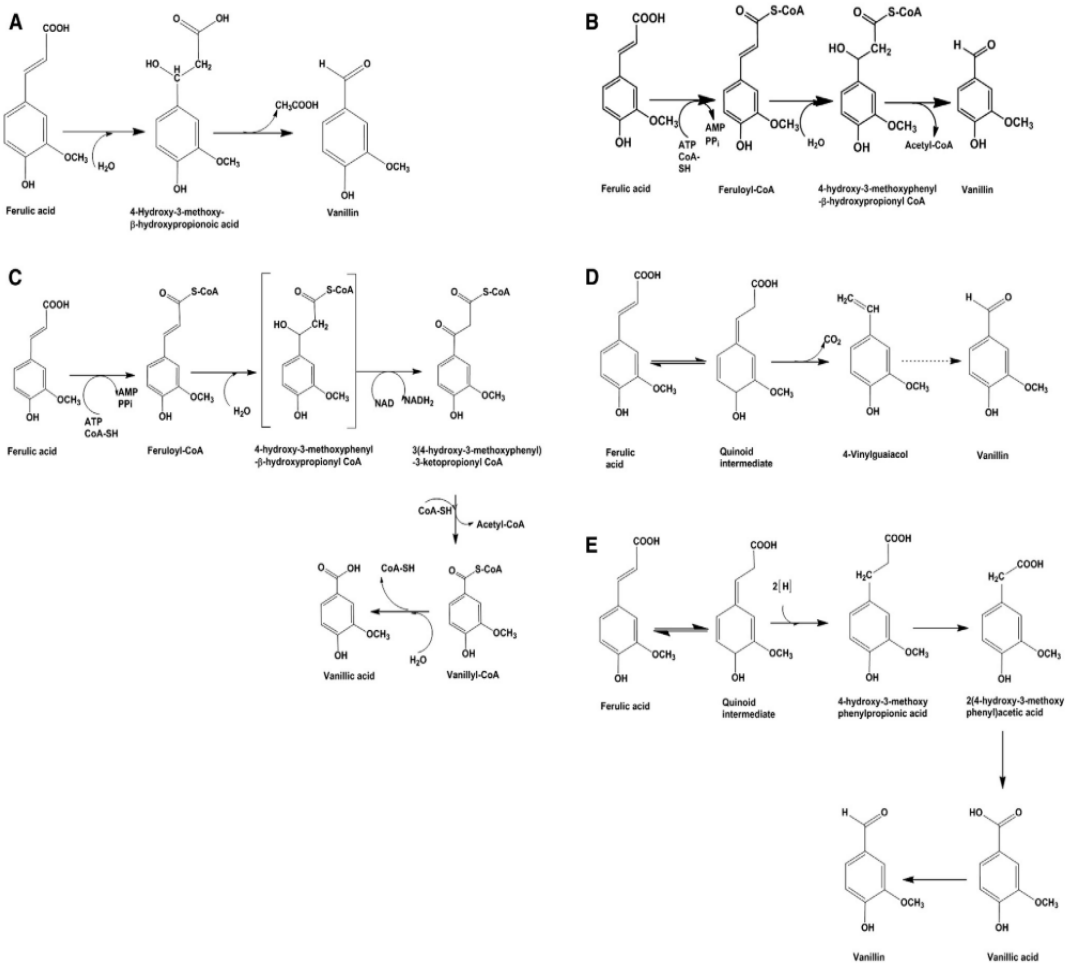


Figure 11: Five possible routes for the enzymatic bioconversion of ferulate to vanillin. A) CoA-independent retro-aldol reduction; B) CoA-dependent retroaldol reduction; C) CoA-dependent β-oxidation; D) non-oxidative decarboxylation; and E) reductive pathway (Gallage and Moller, 2015).

Organism	Study
Bacteria	
Pseudomonas putida	(Labuda et al., 1992).
Amycolatopsis sp	(Achterholt et al., 2000).
Bacillus subtilis	(Gurujeyalakshmi and Mahadevan, 1987).
Pseudomonas mira	(Jurková and Wurst, 1993).
Bacillus coagulans	(Karmakar et al., 2000).
Streptomyces setonii	(Muheim and Lerch, 1999; Sutherland et al., 1983).
Pseudomonas fluorescens	(Narbad and Gasson, 1998)
Escherichia coli	(Yoon et al., 2005) (Torres et al., 2009).
Pseudomonas sp	(Overhage et al., 1999).
Streptomyces halstedii	(Brunati et al., 2004).
Oenococcus oeni	(Bloem et al., 2007)
Lactobacillus sp	(Bloem et al., 2007)
Enterobacter sp. Px6-4	(Li et al., 2008).
Corynebacterium glutamicum	(Labuda et al., 1992)
Pseudomonas acidovorans	(Toms and Wood, 1970).
Fungi	
Paecilomyces variotti	(Rahouti et al., 1989).
Aspergillus niger + Pycnoporus cinnabarinus	(Lesage-Meessen et al., 1999).
Brettanomyces anomalus	(Edlin et al., 1995).
Pycnoporus cinnabarinus	(Falconnier et al., 1994).
Polyporus versicolor	(Ishikawa et al., 1963).
Fomes fomentarius	(Ishikawa et al., 1963).
Debaryomyces hansenii	(Mathew and Abraham, 2004).
Fusarium solani	(Nazareth and Mavinkurve, 1986).

Table 1: List of organisms utilised for microbial production of vanillin from ferulate as of 2014 (Batista, 2014).

Genes encoding the enzymes required for the conversion of ferulate to vanillin have been identified in a variety of microbial species, some of which are listed in Table 1. Studies have been conducted into the direct utilization and optimization of these native vanillin producers with varying results. Genetic manipulation of *P. fluorescens* via concurrent amplification of the *fcs* gene for the synthesis of feruloyl-CoA result in yields of 1.28g l^{-1} of vanillin from ferulate (Di Gioia et al., 2011); *Rhodococcus jostii* has produced yields of 0.095g l^{-1} of vanillin when grown on minimal medium with wheat straw lignocellulose as the substrate source (Sainsbury et al., 2013); and 19.2g l^{-1} was achieved with use of

Streptomyces and adsorbent resins allowing a high concentration of ferulate to be fed into the culture continuously (Hua et al., 2007b). One issue that arises from the use of natural vanillin producers as a platform is the unwanted subsequent catabolism of the accumulating intermediate vanillin to vanillic acid, via vanillin dehydrogenase, (EC 1.2.1.67). In order to prevent this unwanted oxidation of vanillin, a precise *vdh* deletion mutant of the vanillin-producing actinomycete *Amycolatopsis* sp. strain ATCC 39116 was created (Fleige et al., 2013). The *Amycolatopsis* sp. ATCC 39116 Δ*vdh*::Km(r) mutant was shown to produce vanillin at a concentration 2.3-fold larger than the original strain. While the highest yields of vanillin production have been seen through use of actinomycetes (such as *Amycoloatopsis* and *Streptomyces*), filamentous growth of these leads to untenably viscous cultures, unfavourable pellet formation and uncontrollable lysis of the mycelium that could cause problems with downstream processing when applied on an industrial scale (Barghini et al., 2007). Therefore the use of unicellular organisms should not encounter similar problems with the scaling up of the process.

1.4.4. E. coli as a platform for vanillin production

An alternative approach is to use *E. coli* as the platform for vanillin production via heterologous expression of the required enzymes, thus eliminating problems arising from downstream metabolism of accumulated vanillin through endogenous pathways. Furthermore, the unparalleled genetic, physiological and metabolic knowledge of *E. coli* means it is an extensively used biotechnological platform (Bokinsky et al., 2011; Woodruff et al., 2013). Production of recombinant strains of *E. coli* for vanillin production has been attempted: Yoon et al. (2005) cloned genes from *Amycolatopsis* sp. Strain HR104 *and Delftia acidovorans* into *E. coli* using the P_{BAD} expression system (Yoon et al., 2005). The yields of 580 mg l^{-1} from 1 g l^{-1} ferulate highlighted the feasibility of such as approach and multiple further efforts have been made to use *E. coli* for vanillin production: Fcs and Ech from *P. fluorescens* were expressed using a low-copy number plasmid with yields 2.52 g l^{-1} (Barghini et al., 2007); corn cob hydrolysate was used a ferulate feedstock produced 0.239 g l^{-1} (Torres et al., 2009); increased acetyl-CoA consumption and forcing *E. coli* to utilise the glyoxalate shunt increased production up to 5.14 g l^{-1} (Lee et al., 2009); the knockout of six endogenous *E. coli* aldehyde reducing enzymes increased vanillin yield 55-fold (Kunjapur et al., 2014); and a coenzyme-independent decarboxylase/oxygenase cascade engineered into *E. coli* produced 1.2 g l^{-1} (Furuya et al., 2015). While the use of *E. coli* alleviates one issue of vanillin yield being reduced by inherent degradation pathways of aromatic metabolising microbes, an area improved further by the use of the RARE strain by

Kunjapur et al. (2014), the problem of toxicity still persists as limiting overall production (Kunjapur et al., 2014).

1.5. Bacterial Toxicity and Stress Responses in Biotechnology

The use of microbial cell factories for the production of high-value chemicals and renewable fuels is rising sharply due to increased demand, a greater understanding of bacteria's metabolic capabilities through genomic characterisation, and developments in genetic engineering and protein expression procedures. Whether inherent producers of the compound of interest are manipulated to accumulate the product, or a recombinant strain is imbued with the ability to produce something it would not normally encounter, toxicity of the substances involved impairs growth, reduces yield, and has to be addressed if industrial scale production is to be achieved.

The stress responses seen in bacteria in the industrial biotechnology field have a lot in common with those studied in regards to antimicrobial research. Bacteria have evolved stress responses to cope with the presence of antimicrobials that are intrinsically involved in the development of antimicrobial resistance, with protective responses often including biofilm formation, membrane alterations, upregulation of efflux systems and growth cessation (de Bruijn, 2016). Unlike pharmacological research, in industrial biotechnology an understanding of the mechanisms and regulation of such adaptations is sought with the aim to promote them, with the exception perhaps of growth cessation and dormancy. With bacteria such as *E. coli* used as a model organism for both clinical and industrial research, outcomes in one can often be applied to the other.

A huge range of compounds that exhibit toxicity are encountered in biotechnology, with varying levels of understanding as to their effects at a physiological level. The toxicity of solvents such as aliphatic alcohols and alkanes has been linked to membrane damage due the correlation between the compounds toxicity and their hydrophobicity based on their partition coefficient (logP) (Rutherford et al., 2010). However, the response to alcohols is not always consistent between bacteria, and differs depending on chain length (Kabelitz et al., 2003). The toxicity of aldehydes is significantly less well-characterised despite their endogenous occurrence, though their high reactivity has been linked to multiple chaotropic effects (See Section 1.5.5). Volatile- and semi-volatile organic compounds such as polycyclic aromatic hydrocarbons (PAHs) have been shown to be mutagenic and carcinogenic in animals, but bacteria are utilised in their bioremediation (Yung et al., 2016a). Despite being capable of metabolising these compounds, bacteria do produce stress responses to alleviate toxicity. Many of the precursors and products of biotechnology produce common effects, such as oxidative stress and respiratory distress, and elicit common responses, such

as increased efflux, direct detoxification to a less toxic variant and alteration of membrane composition. Approaches to understanding stress can be broad, whole-cell based or focused on specific targets.

1.5.1. Omics-based analysis of tolerance

Developments in genomics, transcriptomics, proteomics and metabolomics methods has increased their effectiveness; thus, they have been widely used to elucidate the stress responses of bacteria to industrial compounds to form a basis for engineering tolerance.

Production of 3-hydropropionic acid by bacteria is beneficial for its use as a precursor for the synthesis of its biodegradable polymer polyester. A transcriptomics study in *E. coli* identified genes involved in oxidative stress, DNA protection and arginine tolerance mechanisms due to acidification of the cytoplasm (Yung et al., 2016b). There were also indications that the 3HP was reduced to its aldehyde form with genes encoding the aldehyde reductase YqhD and formaldehyde detoxification enzymes upregulated. A genomic study into n-butanol stress in *E. coli* discovered it caused perturbation of respiration, indicating an increased requirement for energy through increased long-term transcript levels of the *nuo* and *cyo* operons (Rutherford et al., 2010). This study also discovered n-butanol caused oxidative, heat shock and envelope stress responses, and an increase in transcription of the periplasmic maltose-binding protein gene *malE*. Another *E. coli* sugar transport system, the mannose transporter ManXYZ that is part of the PTS system, has been identified through transcriptional analysis to be involved in organic solvent tolerance; furthermore overexpression of this system improved hexane tolerance by decreasing its intracellular level (Okochi et al., 2007). This showed that these studies can prove useful starting points. A genomic library screen identified 270 genes in *E. coli* that showed a significant change in regulation in the presence of n-butanol (Reyes et al., 2011). Based on this, genetic targets were identified and overexpressing the iron transport and metabolism genes *entC* and *feoA*, and deleting *astE*, which encodes an enzyme involved in arginine degradation, all increased tolerance. A comparative quantitative proteomics approach also identified 108 butanol-responsive proteins in *B. subtilis*, which exhibits higher than average butanol tolerance, and highlighted that the bacteria accumulated proline as a compatible solute as its major defence mechanism (Mahipant et al., 2017).

1.5.2. Adaptive laboratory evolution (ALE)

An “irrational” approach to engineering tolerance in bacteria is adaptive laboratory evolution (ALE), where a species of bacteria is grown and sub-cultured in a defined environment for an extended period of time (Dragosits and Mattanovich, 2013). For biotechnology this can involve growing the bacteria in increasing concentrations of a relevant, toxic substance and positively selecting cells that survive due to having evolved adaptations. Improvements in the speed and quality of whole-genome sequencing allows for identification of the changes to the bacterium’s genome that have imparted this tolerance, such as single nucleotide polymorphisms, indels and large-scale amplifications or deletions (Conrad et al., 2011; Liu et al., 2012). A famous example of ALE is the ongoing long-term evolution experiment of *E. coli* by the Lenski group that has gone through over 66,000 generations since its instigation in 1988 (Cooper and Lenski, 2000; Tenaillon et al., 2016).

This method has been applied for industrial biotechnology purposes. *P. putida* was exposed to increasing *n*-butanol concentrations resulting in isolation of a strain that could survive in 6% butanol, where normal tolerated levels are <2%, with a four-fold increase in redox cofactor regeneration levels identified and linked to this tolerance (Ruhl et al., 2009). A similar degree of tolerance to ethanol was developed in a *Thermoanaerobacter* species, which offers the potential to conduct biotransformations at temperatures up to 65°C due to its thermophilic nature, with growth at 6% ethanol shown (Lin et al., 2013). However, a major contributing factor to this was overexpression of an alcohol dehydrogenase which would negatively impact yield. Several lineages of *E. coli* with evolved *n*-butanol tolerance showed different genetic mechanisms, including increased fatty acid biosynthesis and both increased and decreased Fur activity that affected iron homeostasis (Reyes et al., 2012). Genome sequencing of evolved isopropanol tolerant *E. coli* showed mutations in five genes – *relA*, *marC*, *proQ*, *yfgO*, and *rraA*, with transcriptomic analysis of these strains indicating increased iron uptake, amino acid biosynthesis (except arginine biosynthesis that experienced the opposite) and oxidative stress defence, and decreased respiration and central carbon metabolism (Horinouchi et al., 2017). As well as product tolerance, feedstock tolerance can be evolved as shown by ALE of *Klebsiella pneumoniae* tolerance to cotton stalk hydrosylate, which improved 2,3-butanediol production through achievement of higher biomass levels (Li et al., 2016).

While mutations happen spontaneously at a rate of approximately 10^{-10} per base pair per replication, efforts have been made to synthetically increase the rate of mutagenesis by using a strain deficient in *dnaQ*, a subunit of *E. coli* DNA polymerase III (Drake, 1991; Luan et al., 2013). This strain exhibited a 317-fold increased mutation rate and proof of principle was achieved through rapid evolution of *n*-butanol and acetate tolerance. This concept was advanced further with development of an inducible synthetic module that controlled expression of six DNA replication genes and increased the mutation rate almost 3000-fold (Zhu et al., 2015). Mutagenesis can also be promoted by chemical mutagens or radiation (Winkler and Kao, 2014).

Some cases of ALE that have successfully increased tolerance to the substance of interest have shown decreased tolerance to a different stress; ethanol tolerance in *E. coli* showed a decrease in acid tolerance, and an *n*-butanol tolerant strain showed decreased oxidative stress tolerance (Dragosits et al., 2013; Goodarzi et al., 2010). Overall, ALE in combination with omics-level analysis is a useful method of working towards engineered tolerance, though with some caveats such as this decrease tolerance to other stresses that may be problematic in an industrial environment, while the developed tolerance may be attributed to a mechanism that actually reduces the yield through transformation to a less toxic substance.

1.5.3. Oxidative Stress

A common stress response exhibited by microbes during biotransformation processes is the oxidative stress response as a result of an increase in reactive oxygen species (ROS) such as superoxide, peroxides and hydroxyl radicals (Figure 12). While ROS are produced in aerobic conditions endogenously, a significant rise in these levels results in chaotropic enzyme and DNA damage resulting in growth defects and increased mutagenesis (Imlay, 2013). Bacteria have inherent mechanisms to cope with ROS, expression of which is increased during oxidative stress. The major regulatory process for this is the SoxRS regulon, which is triggered by oxidation of the 2Fe-2S cluster in SoxR by superoxides resulting in activation of SoxS transcription (Blanchard et al., 2007; Liochev et al., 1999). There is also an extensive SoxRS-independent response involved in oxidative stress, mediated by the regulators OxyR, MarA, CysB, IscR, BirA and Fur; OxyR exists in two redox states and exposure to peroxides results in its oxidation, with $OxyR^{ox}$ activating transcription. These regulatory networks are interconnected and result in expression of ROS scavengers including superoxide dismutases (e.g. the iron-containing SodA, and

mangangese-containing SodB in *E. coli*) catalases (KatE and KatG in *E. coli*) and peroxidases (Ahp in *E. coli*). Genes transcriptionally regulated by SoxRS and OxyR in *E. coli* can be seen in Figure 13.

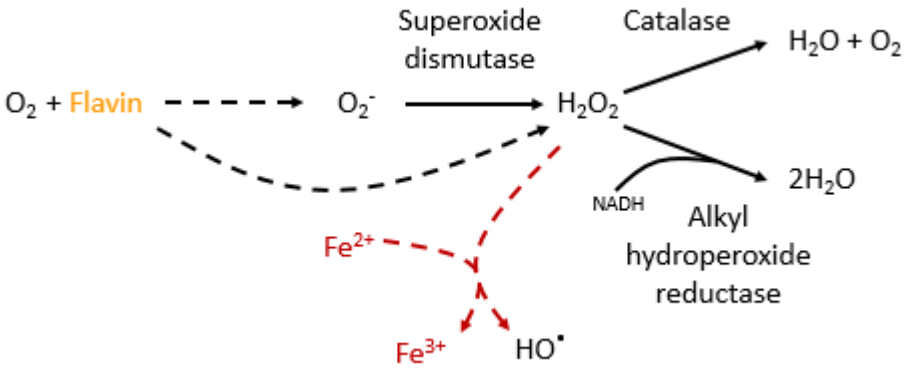


Figure 12: Biological formation of reactive oxygen species. Simple schematic showing ROS generation and the action of scavenging enzymes (solid lines) to alleviate oxidative stress. A basic portrayal of Fenton chemistry is shown in red-brown, displaying the generation of the highly destructive hydroxyl radical.

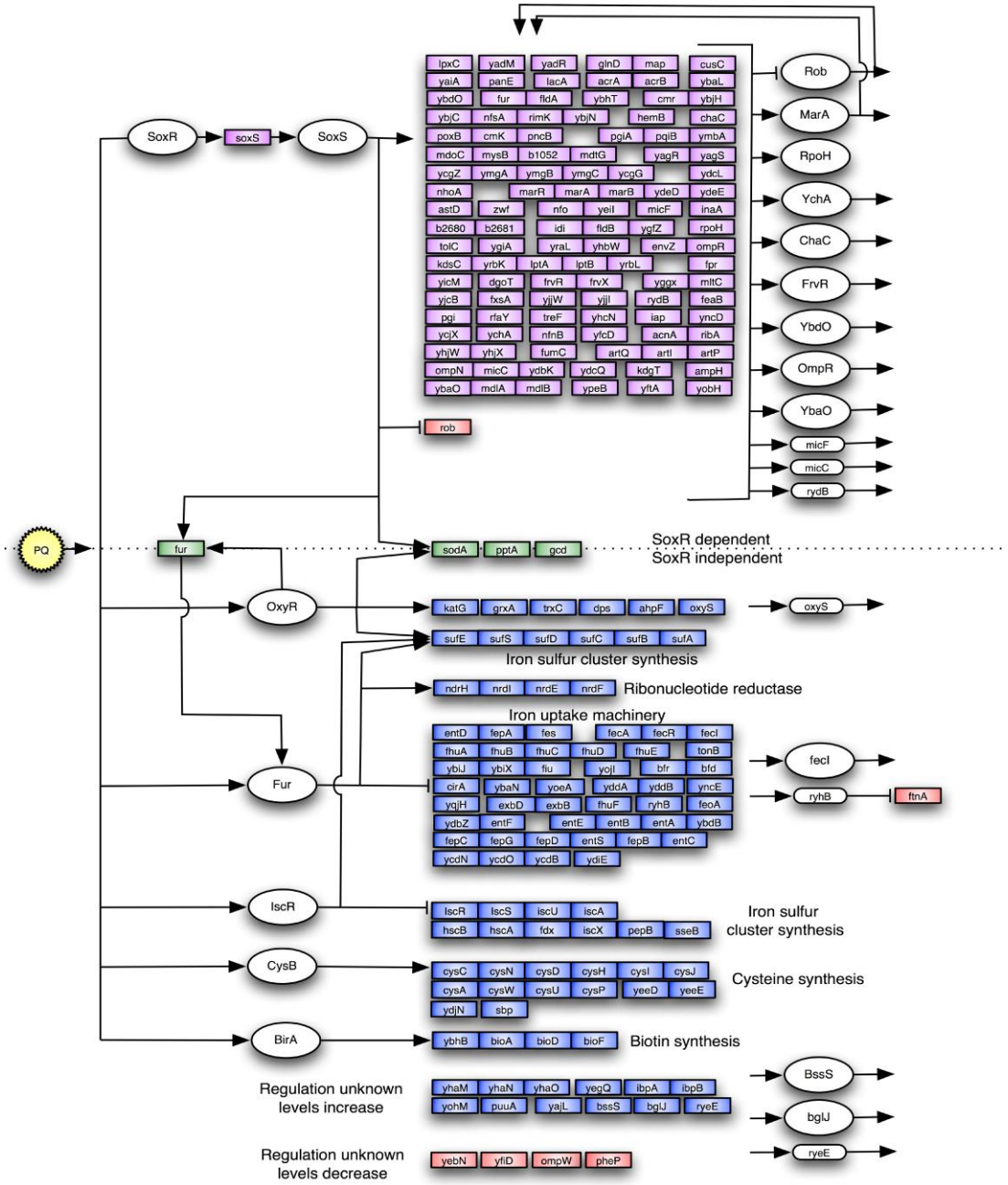


Figure 13: The model for the superoxide transcriptional response in *E. coli* by Blanchard et al, 2007 (Blanchard et al., 2007). This figure was produced through an amalgamation of microarray data, bioinformatics, and primary literature based on paraquat (PQ) treatment, and is a non-extensive overview. mRNAs are shown in boxes, regulatory proteins in ovals and sRNAs in rounded boxes. Connected boxes indicate chromosomal neighbours.

The primary source of superoxide and hydrogen peroxide are intracellular flavoproteins; molecular oxygen inadvertently collides with the dihydroflavin resulting in a single electron transfer event producing O_2^-, and a flavosemiquinone species (Imlay, 2013). This O_2^- can

either diffuse away from the protein immediately, or undergo a second electron transfer producing H_2O_2 (Messner and Imlay, 2002). Different flavoproteins result in different rates of this autooxidation event, and thus ROS production is directly dependent on the abundance of such proteins (and oxygen concentration); e.g. fumarate reductases are a predominant source of superoxide (Korshunov and Imlay, 2010).

Iron is an essential nutrient required as a cofactor in a large array of bacterial enzymes, but unincorporated intracellular iron generates the extremely disruptive hydroxyl radical by the Fenton reaction with peroxides; hydroxyl radicals cause cell-wide damage to biological components. Therefore, iron levels must be strictly controlled. The major regulator of iron homeostasis in *E. coli* is the ferric uptake regulator, Fur. This transcription factor represses iron uptake machinery when iron is bound, i.e. when intracellular iron levels are sufficient. Oxidative stress defence is coupled to iron metabolism by SoxS and OxyR promoting *fur* expression, thus coordinating regulation and reducing iron uptake when ROS levels rise.

Fenton chemistry:

$$Fe^{2+} + H_20_2 \rightarrow Fe^{3+} + OH^- + OH^{\cdot} \quad (1)$$

$$OH^{\cdot} + H_2O_2 \rightarrow HO_2^{\cdot} + H_2O \quad (2)$$

$$Fe^{3+} + HO_2^{\cdot} \rightarrow Fe^{2+} + H^+ + O_2 \quad (3)$$

$$Fe^{2+} + HO_2^{\cdot} \rightarrow Fe^{3+} + HO_2^- \quad (4)$$

$$Fe^{2+} + OH^{\cdot} \rightarrow Fe^{3+} + OH^- \quad (5)$$

Therefore, during oxidative stress key downstream effects are reduced synthesis of iron importers and siderophores, and increased synthesis of ferritin-class iron scavengers (Imlay, 2013). Perturbation of iron homeostasis, in addition to oxidative stress, has been identified during studies into bacterial tolerance in biotechnology (Sections 1.5.1 and 1.5.2).

As mentioned previously, exogenous *n*-butanol and 3-hydroxypropionic acid elicited an oxidative stress response in *E. coli* (Rutherford et al., 2010; Yung et al., 2016b). Bacteria can encounter heavy metals at toxic levels during bioremediation processes that inflict oxidative stress; cadmium caused lipid peroxidation in cyanobacteria found in contaminated water, and an increase in catalase activity is required to achieve tolerance (Haghighi et al., 2017).

Increasing the oxidative stress response has proven successful in a biotechnology setting, with overexpression of the antioxidant enzymes ferredoxin-NADP(+) reductase, Fpr; superoxide dismutases, SodA and SodB; and alkyl hydroperoxide reductase, AhpC, enhancing a strain of *Pseudomonas*'s ability to degrade naphthalene (Kang et al., 2007). In the cyanobacterium *Synechocystis* sp. Strain PCC 6803, overexpression of SodB improved *n*-butanol tolerance, though overexpression of a peroxiredoxin system known to reduce H_2O_2 levels had no effect (Anfelt et al., 2013).

1.5.4. Efflux systems as a specific means of tolerance

Active efflux is of vital importance to bacteria in their ability to tolerate toxicity as a result of antibiotics, detergents and environmental compounds. The multicomponent acridine resistance efflux system (AcrAB-TolC) of *E. coli* is perhaps the most-studied efflux pump in bacteria, and is the model for multicomponent systems (Du et al., 2014). AcrB is an inner-membrane-located secondary transporter while AcrA acts as a periplasmic bridge connecting AcrB with TolC, an outer-membrane channel. AcrAB-TolC's broad specificity is conveyed through AcrB's binding pocket; substrates have been shown to bind to both a narrow groove and a wide cave in this area of the structure (Takatsuka et al., 2010).

Regulation of AcrAB-TolC is complicated and heavily entwined with multiple stress responses, with regulation mediated by MarA, SoxS, Rob and AcrR, with overlapping and crosstalk between these regulatory systems (Chubiz et al., 2012). The MarRAB operon is vital to antibiotic resistance and controls transcription of regulon in response to xenobiotic stress, with the operon itself controlled by MarR (Alekshun and Levy, 2007; Martin and Rosner, 2004). Transcriptional repression of the operon by MarR has been shown to be reversed by binding of the aromatic acid salicylate and other phenolic antibiotics; and oxidation by Cu^{2+} (Duval et al., 2013; Hao et al., 2014). This leads to MarA production, a transcriptional activator that acts on the antibiotic resistance regulon. The genes controlled overlap with those of the SoxRS regulon (Blanchard et al., 2007; Liochev et al., 1999). With oxidative stress a common effect of accumulation of biotechnological-relevant compounds, upregulation of AcrAB-TolC is central to vital stress responses.

Overexpression of AcrAB-TolC in *E. coli* and *Salmonella typhimurium*, and its homolog MexAB–OprM in *Pseudomonas aeruginosa*, is a common feature in multidrug resistant strains (Llanes et al., 2004; Sato et al., 2013). In biotechnology, the potential for engineering AcrAB-TolC to improve tolerance has also been identified (Mukhopadhyay, 2015). The efflux pump has been shown to be required for tolerance to aliphatic

hydrocarbons such as hexene and octanol, and aromatic compounds such as styrene (Foo and Leong, 2013; Mingardon et al., 2015). A successful application of this has been shown by overexpression of AcrAB-TolC increasing the tolerance of terpenes such as amorphadiene and kaurene, which are used as precursors in the production of medicines, cosmetics and biofuels (Wang et al., 2013b). However, AcrAB-TolC's capabilities do have limitations as it has proven incapable of efflux of more polar, short-chain alcohols. Furthermore, though it has shown to efflux toxic levels of fatty acids such as dodecanoic acid, overexpression did not increase tolerance or yield (Lennen et al., 2013).

Interestingly higher tolerance to isobutanol was actually achieved by knockout of the *acrB* gene (Ankarloo et al., 2010). This may be a result of an indirect effect on expression of other genes caused by deletion of *acrB*, or *acrA*, as shown by investigations of the effect of knocking out *acrB* in *Salmonella* (Blair et al., 2015; Webber et al., 2009). This includes upregulation of transport proteins, i.e. AcrD and AcrF; upregulation of the regulator RamA, a member of the AraC/XylS family of transcription factors that includes MarA, SoxS and Rob; and downregulation of the anaerobic metabolism *nar* and *nir* operons.

Other bacterial efflux systems have been evaluated for their industrial potential. The AcrAB-TolC homologs TtgABC, TtgDEF, TtgGHI in *P. putida* have been identified as toluene efflux systems and are key to its tolerance, though expression of *ttgABC* in *E. coli* did not confer tolerance to toluene (Dunlop et al., 2011; Rojas et al., 2001). An alternative multidrug resistance efflux complex in *E. coli* is the EmrAB system that also complexes with TolC similar to AcrAB. Another *P. putida* strain shows upregulation of an EmrAB-TolC homolog in response to production of the aromatic compound p-hydroxybenzoic acid from glucose, and deletion of this efflux system reduces yield and cell growth (Verhoef et al., 2010). However, overexpression, again, did not affect tolerance. PAHs are other industrially-relevant aromatic compounds for which a bacterial efflux system has been characterised; the EmhABC system in *Pseudomonas fluorescens* was shown to export phenanthrene, anthracene and fluroanthene (Hearn et al., 2003). The industrial implications of this contrast with that of other transporters, as expression of EmhABC reduces the bioremediation capabilities of *P. fluorescens* (Adebusuyi et al., 2012). That this system also effluxes the antibiotics chloramphenicol and nalidixic acid highlights the relationship between clinical and industrial tolerance in bacteria at a physiological level.

To fully harness controlled expression of efflux systems to increase tolerance in industrial biotechnology, reducing the abundance of other transporters not involved in efflux of the

compound of interest, as shown inherently by *Pseudomonas* sp., may help, while co-expression of the SecYEG translocation machinery may be required to facilitate correct expression of the efflux system (Jones et al., 2015; Mulder et al., 2013).

1.5.5. Aldehydes in biotechnology

The production of aldehydes in biotechnology has great value due to this class of compounds' importance in the flavouring and fragrance industries, and also as precursors for the synthesis of pharmaceuticals and other high-value chemicals. Aldehydes are organoleptic due to their binding of G-protein-coupled receptors involved in mammalian olfaction and gustation; aliphatic aldehydes such hexanal, octanal and decanal produce fruity flavourings such as apple and citrus, while aromatic aldehydes such as vanillin, cinnamaldehyde and benzaldehyde are the key chemical components of the natural flavourings vanilla, cinnamon and almond, respectively (Kunjapur and Prather, 2015). Benzaldehyde is required to produce phenylacetylcarbinol, which itself is necessary for synthesis of the pharmaceuticals ephedrine and pseudoephedrine, by fermentation in combination with dextrose. Overall, purified aldehydes have an extremely high market value. The biotransformation of aldehydes using bacteria is therefore of great interest (See Section 1.4), though the aldehydes' reactivity makes them typically highly toxic. However, unlike other compounds such as ethanol and butanol, the mechanism of toxicity of aldehydes is less comprehensively understood.

An investigation into the inhibitory effects of several aldehydes on an ethanologenic *E. coli* strain reported several key findings (Zaldivar et al., 1999). Most importantly, a relationship between the hydrophobicity of the aldehyde compounds and the level of toxicity was discovered. This would suggest that it was their ability to interact with membranes that was the determining factor to their toxic effect, however they did not detect significant membrane damage as a result of treatment. Direct DNA damage is another potential mechanism of aldehyde toxicity, as shown by a study into acetaldehyde highlighting genotoxicity by causing single- and double- stranded DNA breaks (Singh and Khan, 1995). Aldehydes have also been shown to elicit and oxidative stress response in yeast (Nguyen et al., 2014).

Bacteria have methods of coping with the accumulation of aldehydes. Several enzymes have been identified in *E. coli* responsible for the reduction of aldehydes to their typically less toxic alcohol forms, including keto-aldo reductases and aldehyde dehydrogenases (Kunjapur et al., 2014). While over expression of such enzymes would theoretically improve

the tolerance of bacteria to aldehydes, it would do so by negatively impacting on the yield of the product.

In addition to enzymes capable of converting aldehydes to potentially less toxic compounds, *E. coli* has other established mechanisms for coping with aldehyde toxicity due to the endogenous occurrence of certain aldehydes. Highly reactive unsaturated aldehydes such as 2-nonenal and malondialdehyde are generated from lipid peroxidation, while methylglyoxal is a toxic aldehyde produced as a glycolytic intermediate that accumulates under certain metabolic conditions (Ferguson et al., 1998; Grimsrud et al., 2008). Methylglyoxal's mechanisms of toxicity emanate from its highly electrophilic nature, causing it to interact and modify DNA and proteins forming unwanted adducts (LoPachin and Gavin, 2014). Glycation of nucleotides, amino acids and lipids cause wide-scale damage to the cell, while methylglyoxal treatment of bacteria has been linked to fimbriae and flagella damage (Hayashi et al., 2014; Murata-Kamiya and Kamiya, 2001; Rabie et al., 2016). Therefore, it is essential bacteria have a process of detoxifying methylglyoxal with the glutathione (GSH)-dependent glyoxalase pathway – and subsequent acidification of the cytoplasm though KefBC potassium efflux – both central to this (Ferguson et al., 1998). GSH is also vital to *E. coli*'s detoxification of formaldehyde, an aldehyde generated through multiple endogenous processes (Denby et al 2016). Formaldehyde and GSH react spontaneously to form the adduct *S*-hydroxymethylglutathione that is then processed by the formaldehyde dehydrogenase FrmA, and hydrolases FrmB and YeiG (Gonzalez et al., 2006). Confirmation that an industrially relevant aldehyde product was stimulating a response akin to one of these processes would help elucidate its mechanism of toxicity, and perhaps provide a means of reducing aldehyde toxicity without reducing yield.

1.6. *Rhodopseudomonas palustris*

1.6.1. Background

Rhodopseudomonas palustris is a purple, non-sulfur, photosynthetic *Alphaproteobacterium*. Typically, cells are rod-shaped with flagellated poles, with reproduction occurring through budding. The *R. palustris* genome has been sequenced; it consists of a singular circular chromosome 5.46 Mbp in length holding 4,826 genes with an additional plasmid 8,427 bp in length (Larimer et al., 2004). *R. palustris* is known and well-studied for its extremely versatile metabolism allowing it to grow in a wide range of environments such as soils, pond water, marine sediments, agricultural and industrial waste and effluents, marshes and alkaline lakes (Madigan and Gest, 1988; Oda et al., 2003). An integral part of this versatility is the ability of *R. palustris* to uptake a plethora of solutes for metabolic purposes, and it is well equipped to do this with an extensive array of transport systems; ~15% of the genome encodes transport proteins, which is a larger proportion than most bacteria (typically 5-6%). Furthermore, unlike most purple photosynthetic bacteria that are predominantly solely photoheterotrophic, *R. palustris* is capable of switching between chemoautotrophic, chemoheterotrophic, photoautotrophic and photoheterotrophic metabolic systems; the former two as aerobic growth and the latter two as anaerobic growth (Figure 14).

Ultimately, its broad and extensive metabolic capabilities has resulted in *R. palustris* being used as a model organism for bacterial photophosphorylation, nitrogen fixation and anaerobic degradation of aromatic compounds (Pechter et al., 2015).

1.6.2. Metabolic traits

As stated above, *R. palustris*'s metabolism is extremely flexible. Aerobically, the bacterium uses oxygen as the terminal electron acceptor in a membrane-located electron transport chain to create a proton-motive force that is harnessed for the regeneration of ATP by phosphorylation of ADP by ATP synthases. *R. palustris* either utilises a wide-range of compounds that it actively transports into the cell via its many transport systems as carbon sources, or fixes carbon from reduction of CO_2. The latter involves uptake and utilisation - via oxidation - of inorganic molecules such as molecular hydrogen, thiosulfate and ferrous ions as electron donors. This chemosynthetic process allows formation of carbohydrates, and subsequently other organic "building blocks" required to produce cell material, from CO_2.

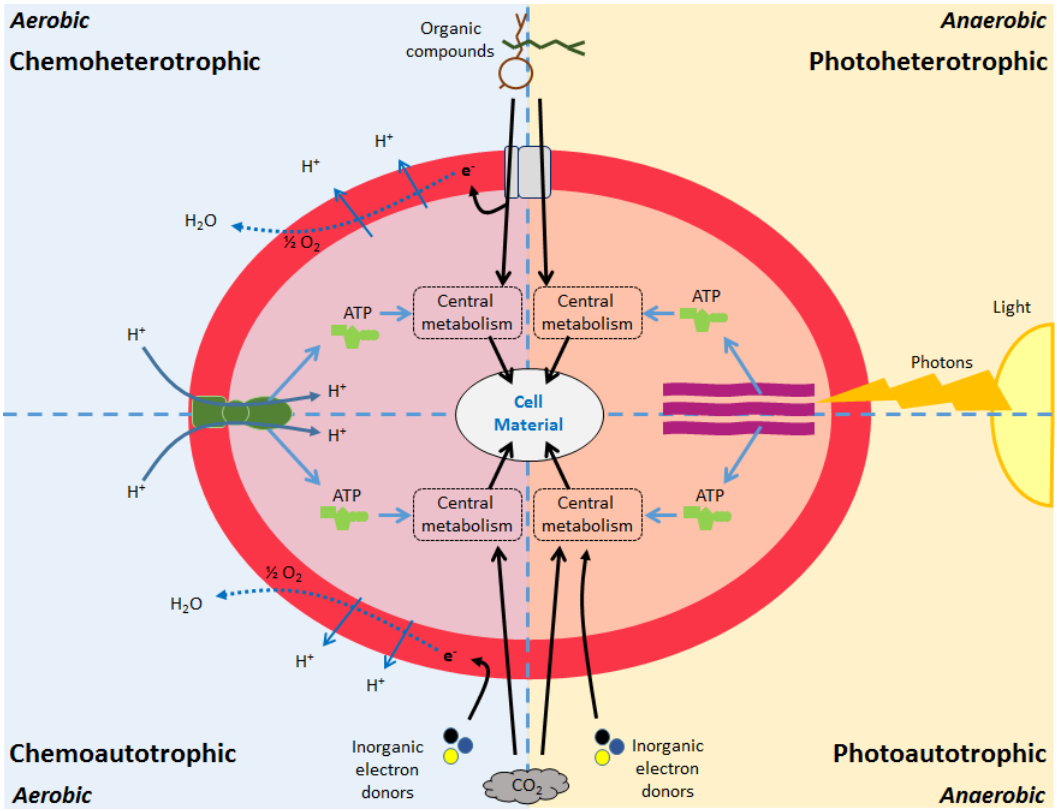


Figure 14: The four modes of metabolism that *R. palustris* is capable of switching between. (Adapted from Figure 1, Larimer et al, 2004)

Anaerobically, energy captured from photons by light-harvesting complexes is instead used for the creation of a proton-motive force; ATP synthases are still used for ATP production. *R. palustris* produces multiple forms of light-harvesting complex 2 (LH2) that absorb different wavelengths of light, allowing it harness light over a broad spectrum (Gall and Robert, 1999). The range of phytochromes synthesised by *R. palustris* allow it to respond to changes in its light environment (Giraud et al., 2005). As is the case in other photosynthetic bacteria, energy is transfered from these peripheral LH2 complexes to the linked LH1 complexes and reaction centres (Cogdell et al., 1999; Scheuring et al., 2006). *R. palustris* can also respond to the level of light intensity; the abundance of photosynthetic apparatus is regulated and a circadian regulatory system is employed to conserve energy (Luer et al., 2012; Ma et al., 2016).

Carbon can be sourced solely from CO_2 autotrophically using photosynthesis or additionally by breakdown of organic molecules using ATP. For photosynthesis *R. palustris* produces two different forms of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), the enzyme integral to the Calvin cycle. The encoding of genes for both a two-subunit Form I, *cbbLS*, and a single-subunit Form II RuBisCO, *cbbM*, is somewhat unusual but a trait *R.*

palustris shares with other non-sulfur photosynthetic bacteria such as *Rhodobacter capsulatus* and *Rhodobacter sphaeroides* (Joshi et al., 2013; Romagnoli and Tabita, 2006).

R. palustris has also been shown to acquire electrons externally by extracellular electron transfer (EET) using ferrous iron, Fe^{2+}, or even electrodes as electron donors (Bose et al., 2014). This process involves a set of proteins produced from the *pioABC* operon essential for iron oxidation, and is upregulated during phototrophic growth (Bose and Newman, 2011; Hernandez and Newman, 2001; Jiao and Newman, 2007).

Due to the large number of genes involved in transport in *R. palustris* it is unsurprising that its transportome exhibits a certain level of redundancy. In the CouR-regulated operon that encodes the enzymes for the initial steps of the metabolism of hydroxycinnamic acids, genes were identified for both an ABC and a TRAP transport system. Studies based on this have since confirmed that the SBPs of both systems have similar binding affinities with a range of hydroxycinnamic substrates, implying that *R. palustris* has two distinct methods for the uptake of these compounds as a carbon source making each one non-essential (Salmon et al., 2013). As each transport system is driven by a different energy coupling method this could allow for the highly efficient uptake of the substrates regardless of which energy source is more favourable. Redundancy is also shown in other areas of the *R. palustris* genome including the degradation pathways of substrates transported by these ABC and TRAP systems (1.6.3; Figure 15); by *R. palustris* possessing two paralogous NADH-dehydrogenases for the transfer of electrons from NADH to quinone, though only one is essential (Pechter et al., 2015); and, by the three nitrogenases, one iron-cofactor-, one molybdenum-cofactor- and one vanadium-cofactor-utilising nitrogenases (Oda et al., 2005).

1.6.3. Anaerobic aromatic degradation in R. palustris

As described in Section 1.3, anaerobic aromatic degradation is “more difficult” and less common than the oxygenase-dependent approach to ring cleavage, with *R. palustris* being a model organism for the metabolism of aromatic compounds without molecular oxygen. *R. palustris*’s extensive metabolism includes the uptake of a wide range of aromatic compounds as growth substrates, and the required peripheral pathways to funnel these into the central anaerobic degradation pathway, the benzoyl-CoA pathway described in detail in Section 1.3.2. One of these peripheral pathways involves the conversion of hydroxycinnamates that are representative of phenylpropanoid monomers derived from lignin, such as cinnamate, coumarate, caffeate and ferulate, to benzoyl-CoA. The pathway has been studied as its central to *R. palustris*’s ability to grow on lignin-derived aromatics.

The pathway can follow two routes, the β-oxidation and non-β-oxidation branches, as shown in Figure 15. While a transcriptomics and proteomics study suggests the non-β-oxidation route is preferable, based on upregulation of the genes encoding the first two enzymes when the bacterium was grown with coumarate as the sole carbon source. *R. palustris* possesses gene clusters for both routes and this is another example of functional redundancy in the *R. palustris* genome (Hirakawa et al., 2012; Pan et al., 2008). Both pathways are encoded by genes in distinct operons: for the β-oxidation route a TetR-family-regulator-regulated operon, genes *RPA1703*-RPA1710; and for the non-β-oxidation route the CouR-regulated operon, which also includes the aforementioned CouPSTU and TarPQM transporters, genes *RPA1782-RPA1795*. Both routes start with a CoA-ligation reaction, and the two respective enzymes for this step have been characterised with the same range of hydroxycinnamate substrates (Hirakawa et al., 2012)(*Unpublished work in PhD the thesis of R. Salmon*).

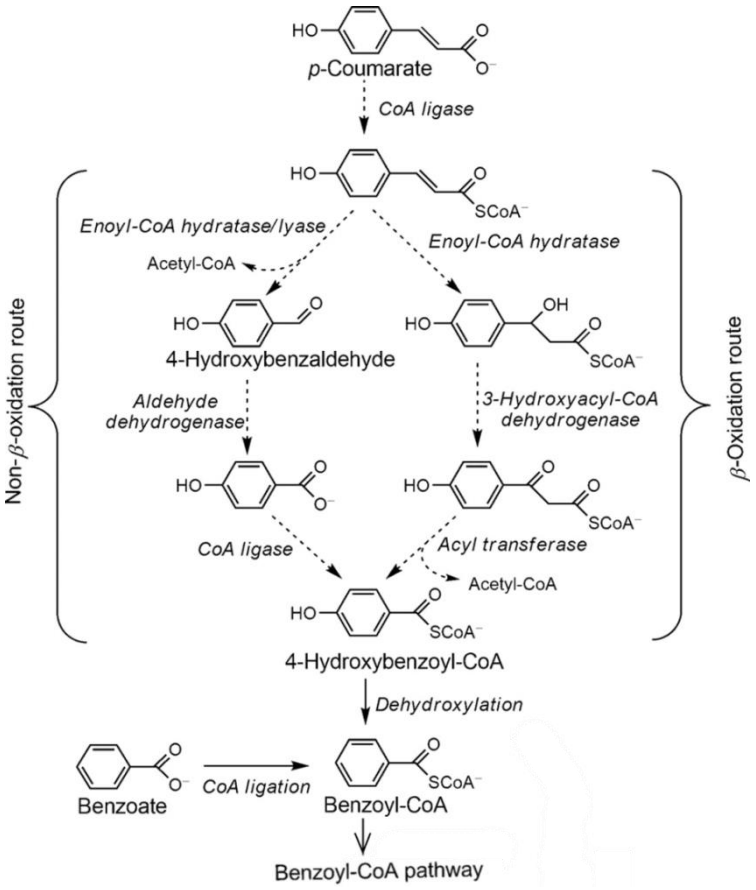


Figure 15: Peripheral degradation pathway for lignin-derived aromatics in *R. palustris*. Coumarate is shown as the substrate in this diagram, though other related hydroxycinnamates, such as cinnamate, caffeate and ferulate are expected to follow the same steps (Adapted from Pan et al, 2008).

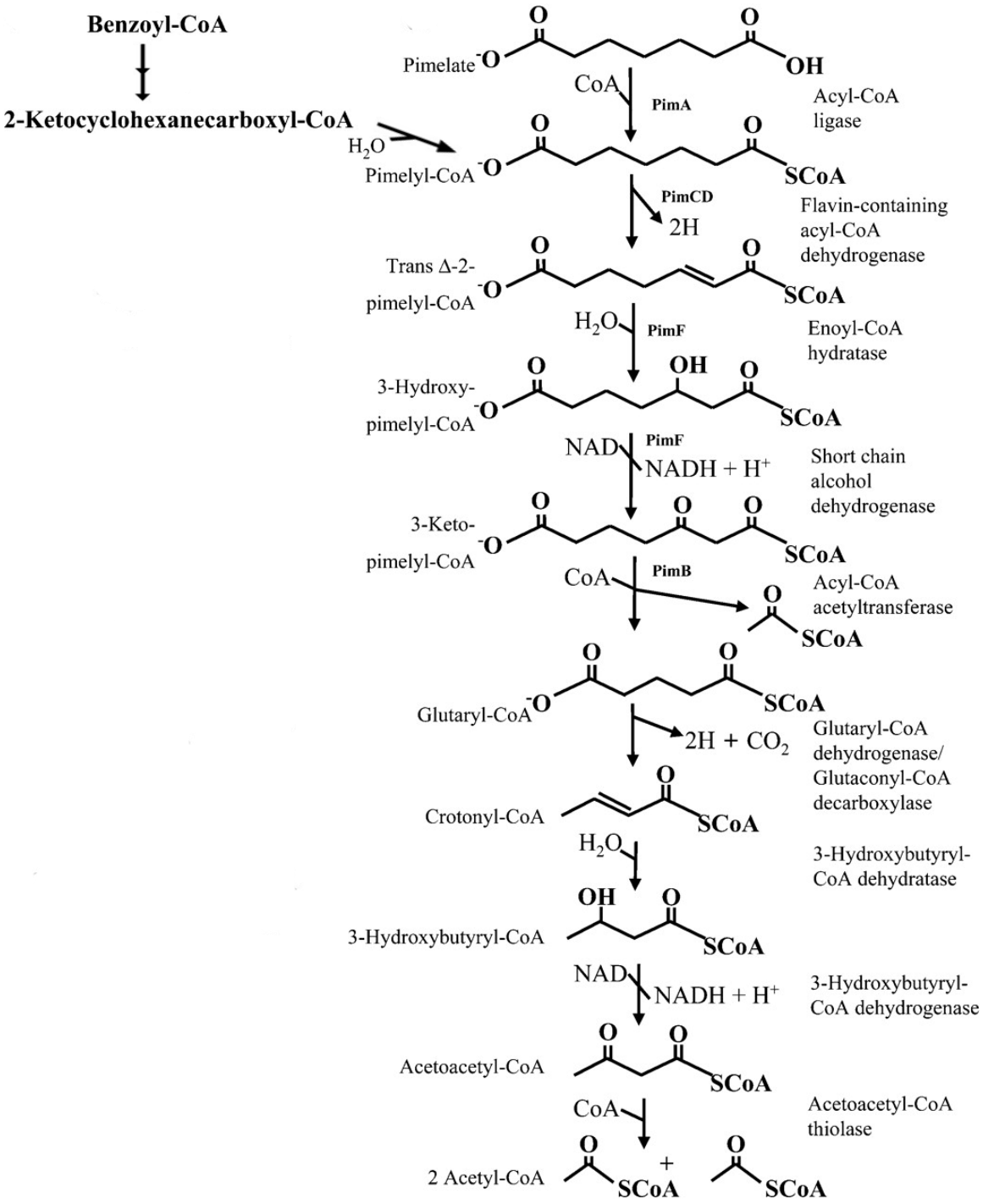


Figure 16: The β-oxidation pathway identified in *R. palustris* for degradation of aromatic compounds to acetyl-CoA via 3-hydroxypimelyl-CoA. The *R. palustris* benzoyl-CoA pathway in Figure 10 showed cleavage of the carbon ring by the 2-ketocyclohexanecarboxyl-CoA hydrolase, BadI, to produce pimelyl-CoA; this product can then feed directly into the β-oxidation for the metabolism of dicarboxylates. The result of this pathway is ultimately three molecules of acetyl-CoA that can feed into the tricarboxylic acid cycle (Harrison and Harwood, 2005).

1.6.4. Industrial biotechnological exploitation

In addition to being studied as a model organism, *R. palustris* garners interest for its potential utilisation in biotechnology due to the possibilities provided by its wide-ranging, flexible metabolism. The ability of *R. palustris* to degrade aromatic compounds is key to its commercial exploitation. Firstly, this allows the use of low cost agricultural and industrial waste as a feedstock, which combined with its photosynthetic capabilities, allow for cheap biomass production and maintenance making this species a useful biotechnological platform. Secondly, *R. palustris*'s ability to degrade a diverse selection of aromatic compounds allows it to be used in bioremediation. Aromatic structures are common in the majority of environmentally damaging pollutants such as industrial waste, sewage and waste waters, petrofuels, plastics, and pesticides (Diaz, 2004). The soluble, aromatic compound benzene is prevalent in many waste products due to its widespread use an industrial solvent, and it is responsible for damaging ecosystems being toxic to animals including humans (Yardley-Jones et al., 1991). *R. palustris*'s aerobic and anaerobic aromatic degradation pathways allow it to metabolise benzene (Egland et al., 1997; Harwood and Gibson, 1988). Furthermore, *R. palustris* is capable of degrading halogenated aromatics such as 3-chlorobenzoate via a reductive dehalogenation step, with some strains able to use it as a sole carbon source (Egland et al., 2001; McGrath and Harfoot, 1997; Oda et al., 2001). *R. palustris* has also been shown to be involved in the degradation of other polluting compounds such as phenanthrene, 2,4,6-Trichlorphenol and coloured dyes (Mehrabi et al., 2001; Wang et al., 2008; Zhao et al., 2011). As well as aromatic compounds, *R. palustris*'s bioremediation capabilities extend to inorganic substances. This includes compounds containing metalloids and metals such as arsenic, chromium and cobalt (Batool et al., 2017; Gao et al., 2017; Gogada et al., 2015; Mehrabi et al., 2001; Merugu et al., 2013; Zhao et al., 2015).

The most extensive studies of the potential of *R. palustris* in the production of various substances is the production of clean-burning hydrogen gas as an alternative, renewable fuel. Microbial biohydrogen production has grown in popularity as a result of the effort to shift away from fossil fuel reliance, and *R. palustris* was highlighted as a potential candidate due to its three different nitrogenases, including a vanadium-containing nitrogenase that produces three times as much hydrogen as the molybdenum-containing counterpart (Chandrasekhar et al., 2015; Larimer et al., 2004; Rey et al., 2007). Various groups have worked on engineering *R. palustris* to enhance its role in biohydrogen production including fermentative H_2 production (Oh et al., 2002; Oh, 2004); using inorganic electron donors

such as thiosulfate (Huang et al., 2010); co-culturing with *Clostridium cellulovorans* and cyanobacteria (Lu et al., 2016; Pandey et al., 2006); using crude glycerol from waste for H_2 production (Pott et al., 2014); capnophilic lactic fermentation (Dipasquale et al., 2015); adjust central carbon metabolism (McKinlay et al., 2014); and utilisation of carbon monoxide for the process (Pakpour et al., 2014).

Another area of interest is use of *R. palustris* as a microbial factory for bioplastic production, as the bacteria can synthesise polyhydroxyalkanoates and polyhydroxybutyrate (PHB) for use as carbon and energy stores (Higuchi-Takeuchi et al., 2016; Larimer et al., 2004). Levels of accumulation are variable and can be enhanced by altering the growth conditions and carbon source, and by nutrient limitation (Carlozzi and Sacchi, 2001; Wu et al., 2012). *R. palustris* can also be used for the production of the antioxidant squalene, 5-aminolevulinic acid for clinical use, and carotenoids (Liu et al., 2016; Xu et al., 2016; Zhang et al., 2013). The bacterium has also been investigated for use as a microbial fuel cell, and a plant-growth enhancer (Lee et al., 2016; Xing et al., 2008). Ultimately, a greater understanding and characterisation of *R. palustris*'s physiology is key to furthering its role in biotechnology.

1.7. Aims and Objectives

The overall aim of this study was to investigate the physiology of relevant bacteria in regards to enhancing their capabilities as a useful tool in biotechnology.

Therefore, the initial planned approaches were:

- Assess the potential of utilising *R. palustris* genes as a basis for engineering a recombinant strain of *E. coli* for the biotransformation of ferulate to vanillin.
- Determine the benefit of additional engineering of *E. coli* to produce a ferulate uptake system, towards the biotransformation process, and compare two distinct *R. palustris* transport systems capable of this uptake.
- Investigate vanillin toxicity in *E. coli* through a comprehensive analysis of the bacterium's response to treatment, with a view to enhancing *E. coli*'s tolerance and thus ability to accumulate it.
- Elucidate the roles of *R. palustris*'s uncharacterised TRAP transporters through biochemical and physiological studies of their substrate binding proteins.

2. Materials and Methods

2.1. Organisms and Growth Media

2.1.1. Organisms used in this study

The full list of bacterial strains used in this study is included in Table 2.

Table 2: List of organisms used in this study.

Strain	Description	Source
Rhodopseudomonas palustris		
CGA009	Genome-sequenced strain used as wildtype	Prof. DK Newman, California Institute of Technology
ΔRPA1206	CGA009 with unmarked deletion of *rpa1206* gene	This study
ΔRPA1975	CGA009 with unmarked deletion of *rpa1975* gene	Rob Salmon, University of Sheffield
ΔRPA3723-3725	CGA009 with unmarked deletions of *rpa3723*, *rpa3724* and *rpa3725* genes	This study
ΔRPA1975 ΔRPA3723-3725	CGA009 with unmarked deletions of *rpa1975, rpa3723, rpa3724* and *rpa3725* genes	This study
Escherichia coli		
BW25113	F^-, DE(*araD-araB*)567, *lacZ*4787(del)::*rrnB*-3, LAM^-, rph-1, DE(*rhaD-rhaB*)568, *hsdR514*	(Datsenko and Wanner, 2000)
DH5α	F^- *endA*1 *glnV*44 *thi*-1 *recA1 relA1 gyrA96 deoR nupG* Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–	(Hanahan, 1983)

Strain	Description	Source
BL21(DE3)	F^- *ompT hsdS*$_B$($r_B^-m_B^-$) *gal dcm* (DE3)	Invitrogen
Origami B(DE3)	F^- *ompT hsd*S_B(r_B^- m_B^-) *gal dcm lacY1 ahpC* (DE3) *gor522*:: Tn*10 trxB*	Novagen
TOP10	F- *mcrA* Δ(*mrr-hsd*RMS-*mc*rBC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*)7697 *gal*U *gal*K *rps*L (StrR) *endA1 nupG*	Invitrogen
S17-1	*recA pro hsdR* RP4-2-Tc::Mu-Km::Tn7 integrated into the chromosome	ATCC#47055

2.1.2. Rhodopseudomonas palustris growth media and conditions

Medium was typically prepared in distilled water (dH_2O) and sterilised by autoclaving, and stored at room temperature. To prepare solid medium 1.5% (w/v) agar (Melford) was added prior to sterilisation. Required supplements were added after cooling of medium and solid medium plates were stored at 4°C.

R. palustris was grown in peptone yeast extract (PYE) medium, as rich medium; *Rhodobacter capsulatus* vitamins (RCV) medium (Weaver et al., 1975), as minimal medium; and M22 medium, as defined medium.

- PYE medium consisted of 5 g l^{-1} peptone (Becton/Dickinson), 5 g l^{-1} yeast extract and 5 g l^{-1} succinate (Oxoid).
- RCV medium consisted of following reagents with final concentrations noted: 30 mM succinate; 7.5 mM $(NH_4)_2SO_4$; 1 mM $MgSO_4.7H_2O$; 0.5 mM $CaCl_2.2H_2O$; 50 µM ethylene-diaminetetraacetic acid (EDTA); 1 mM NaCl; 40 µM $FeSO_4.7H_2O$; 10 µM $MnSO_4$; 3.3 mM $NaMoO_4.2H_2O$; 0.9 mM $ZnSO_4.7H_2O$; 0.2 mM $Cu(NO_3)_2.2H_2O$; 6 µM thiamine HCl; 0.2 µM biotin; and 8 mM nicotinic acid. The following reagents were added after sterilisation $NaHCO_3$, 10 mM, was added as a source of CO_2; 17 µM p-aminobenzoic acid; 10 nM cobalamin; 4 µM calcium pantothenate; and 16 mM potassium phosphate buffer, pH 7. Succinate was substituted for alternative carbon

sources at a range of concentrations sterilised by 0.2 µM filter sterilisation for the appropriate growth experiments.

- M22 media consisted of the following reagents with final concentrations noted: 22 mM sodium lactate; 4 mM ammonium sulfate; 8.5 mM NaCl; 25 mM sodium succinate; 1.6 mM sodium glutamate; 0.3 mM aspartic acid; 1 mM nitrilotriacetic acid; 5 mM $MgCl_2$; 0.6 mM $CaCl_2$; 40 µM $ZnCl_2$; 40 µM $FeCl_2$; 14 µM $MnCl_2$; 1.2 µM $CuCl_2$; 1 µM EDTA; 0.16 µM $(NH_4)_2MoO_4$; 1.4 µM $Co(NO_3)_2$; and 1.8 µM orthoboric acid, made up in 20 mM potassium phosphate buffer, pH 6.8. Casein hydrosylate (acid) was added to a final concentration of 0.5 g l^{-1}.

R. palustris strains were stored in PYE with 25% (v/v) glycerol at -80°C. Bacteria were streaked out from glycerol stocks onto solid PYE media plates, with 20 µg ml^{-1} chloramphenicol; and grown in the dark at 30°C; plates were stored in the dark at room temperature. Single colonies were resuspended in 40 µL PYE prior to initial growth in PYE media. All liquid growth was conducted anaerobically and photoheterotrophically at 30°C in glass on a magnetic stirrer, or in plastic on a tube rotator. Light was provided by a low-heat, 9-watt LED lamp. Alternative growth conditions were required for mutagenesis and are described in detail in Section 2.7. *R. palustris* culture turbidity was measured at OD_{660} in a Jenway 6305 190(UV/Vis) spectrophotometer.

2.1.3. Escherichia coli growth media and conditions

Medium was typically prepared in distilled water (dH_2O) and sterilised by autoclaving, and stored at room temperature. To prepare solid media 1.5–2% (w/v) agar (Melford) was added prior to sterilisation. Required supplements were added after cooling of media and solid media plates were stored at 4°C.

E. coli was typically grown in lysogeny broth (LB) media (Melford). For growth experiments where minimal media was required, M9 media was used:

- M9 media consisted of 42 mM Na_2HPO_4; 22 mM KH_2PO_4; 18 mM NH_4Cl; and 8.5 mM NaCl in dH_2O. Following autoclaving 140 µM $CaCl_2$; 1mM $MgSO_4.7H_2O$; 1.8 µM thiamine; and 0.2 % (w/v) glucose were added.

Liquid cultures were typically grown aerobically at 37°C, shaking at 250 rpm in LB. *E. coli* strains were stored in LB with 25% (v/v) glycerol at -80°C. *E. coli* culture turbidity was measured at OD_{600} in a Jenway 6305 190(UV/Vis) spectrophotometer.

2.1.4. Antibiotics

Stock solutions of antibiotics were made in AnalaR NORMAPUR® water or 70% ethanol and filter sterilised. Final concentrations of the following antibiotics were used: carbenicillin, 50–100 μg ml^{-1}; kanamycin, 30–100 μg ml^{-1}; chloramphenicol 20–40 μg ml^{-1} (Sigma Aldrich).

2.2. DNA Manipulation

2.2.1. Isolation of DNA

R. palustris and *E. coli* genomic DNA was isolated using the Wizard® Genomic DNA Purification Kit (Promega). Plasmid DNA was isolated using the Wizard® *Plus* Midipreps DNA Purification System (Promega). Both kits were used as to the manufacturer's instructions.

2.2.2. Agarose gel electrophoresis

In this study all DNA samples were analysed using DNA gel electrophoresis. Agarose gels (0.7 % standard) were prepared by adding 0.35 g agarose to 50 ml 1% TAE buffer (40 mM Tris-acetate, 1 mM EDTA at pH 8.0) and heating until dissolved. Once cooled ethidium bromide was added to a final concentration of 50 μg ml^{-1}, and the mixture was poured into a cast and a comb was placed in to set. The solid gel was placed in a gel electrophoresis dock and covered in 1% TAE buffer. Samples of DNA were mixed with 5x loading buffer (0.05% (w/v) bromophenol blue, 40% (w/v) glycerol in dH_2O) and loaded alongside Hyperladder 1 (Bioline) as a marker for estimating DNA sample size. A standard gel was exposed to 110 V for 45 min, or until the samples had migrated sufficiently, and then visualised under UV light.

Table 3: List of plasmids used in this study.

Plasmid	Description	Resistance Marker	Source
pET21a(+)	Vector for cytoplasmic overexpression of proteins with C-terminal 6xHis-tag under control of IPTG-induced T7 promoter.	Amp	Novagen
pET22b(+)	Vector for periplasmic overexpression of proteins with C-terminal 6xHis-tag under control of IPTG-induced T7 promoter, with N-terminal *pelB* leader sequence.	Amp	Novagen
pBAD/HisB	Vector for cytoplasmic overexpression of proteins with N-terminal 6xHis-tag under control of the arabinose-induced P_{BAD} promoter.	Amp	Thermo Fisher Scientific
pCOLADuet™-1	Vector for cytoplasmic co-expression of multiple genes using two multiple cloning sites under control of IPTG-induced T7 promoter.	Kan	Novagen
pACYCDuet™-1	Vector for cytoplasmic co-expression of multiple genes using two multiple cloning sites under control of IPTG-induced T7 promoter.	Cm	Novagen
pMAL-p2x™	Vector for overexpression of protein with an N-terminal maltose-binding protein (MBP) fusion, with an enterokinase cut	Amp	New England Biolabs

Plasmid	Description	Resistance Marker	Source
	site in the linker region.		
pK18mob*sacB*	Suicide vector for mutagenesis in *R. palustris*. Contains kanamycin (*kanR*) resistance and sucrose sensitivity (*sacB*) markers.	Kan, sucrose sensitivity	Prof. Neil Hunter, University of Sheffield
pBBRBB-pBAD	Vector for cytoplasmic overexpression in *R. palustris.* Engineered from the pBBRBB-BioBrick™ with the P_{puf} promoter replaced with a P_{BAD} promoter (Tikh et al., 2014)	Kan	Dr. Andrew Hitchcock, University of Sheffield
pET21a:*yqhD*	pET21a(+) containing the *E. coli yqhD* gene.	Amp	This study
pET21a:*aldB*	pET21a(+) containing the *E. coli aldB* gene.	Amp	This study
pET21a:*yhbW*	pET21a(+) containing the *E. coli yhbW* gene.	Amp	This study
pET21a:*1975*	pET21a(+) containing the *RPA1975* gene.	Amp	This study
pET21a:*2047*	pET21a(+) containing the *RPA2047* gene.	Amp	This study
pET21a:*4510*	pET21a(+) containing the *RPA4510* gene.	Amp	This study
pET21a:*3723*	pET21a(+) containing the *RPA3723* gene.	Amp	This study
pET21a:*3724*	pET21a(+) containing the *RPA3724* gene.	Amp	This study
pET21a:*3725*	pET21a(+) containing the *RPA3725* gene.	Amp	This study
pET21a:*4421*	pET21a(+) containing the *RPA4421* gene.	Amp	This study

Plasmid	Description	Resistance Marker	Source
pET21b:*1975*	pET21b(+) containing the *RPA1975* gene.	Amp	This study
pET21b:*3458*	pET21b(+) containing the *RPA3458* gene.	Amp	This study
pET21b:3723	pET21b(+) containing the *RPA3723* gene.	Amp	This study
pET21b:3724	pET21b(+) containing the *RPA3724* gene.	Amp	This study
pET21b:3725	pET21b(+) containing the *RPA3725*gene.	Amp	This study
pET21b:*4556*	pET21b(+) containing the *RPA2047* gene.	Amp	This study
pBAD:*couAB*	pBAD/HisB containing the *couAB* genes complete with intergenic region.	Amp	This study
pBAD:*tarPQM*	pBAD/HisB containing the *couAB* genes complete with intergenic regions.	Amp	This study
pCOLA:*couAcouB*	pCOLADuet™-1 containing the *couA* and *couB* genes in separate MCS sites.	Kan	This study
pACYC:*tarPtarQM*	pACYCDuet™-1 containing the *tarP* and *tarQM* genes in separate MCS sites.	Cm	This study
pMAL-p2x:*1975*	pMAL-p2x with *RPA1975* cloned in to form the MBP-fusion.	Amp	This study
pK18mob*sacB*:1206	pK18mob*sacB* containing flanking regions of the *rpa1206* gene cloned into *EcoR1.*	Kan, sucrose sensitivity	This study
pK18mob*sacB*:3723-25	pK18mob*sacB* containing flanking regions of the *rpa3723 and rpa3725* genes cloned into *EcoR1.*	Kan, sucrose sensitivity	This study

Plasmid	Description	Resistance Marker	Source
pBBRBB-pBAD: *couAB*	pBBRBB-pBAD containing genes for *couAB* for overexpression in *R. palustris* inducible by arabinose.	Kan	This study

2.2.3. Polymerase chain reaction (PCR) amplification

Amplification primers were obtained from Sigma Aldrich, resuspended in AnalaR NORMAPUR® water and stored at a concentration of 10 µM at -20°C. PCR were conducted using Accuzyme DNA polymerase (Bioline), Phusion Flash High-fidelity PCR Master Mix (Thermo Fischer) and MyTaq™ Red Mix (Bioline) according to manufacturer's instructions with adjustments, in a Techne Techgene Thermal cycler (Techne). Primers were used at a final concentration of 0.25 µM, with 0.1–0.5 µg of template DNA per reaction. Amplification of high-GC content *R. palustris* genes used Accuzyme with 2xPolyMate additive (Bioline) and 1.25 mM $MgCl_2$, with an initial 96°C, 5 min denaturation for a 'hot-start' protocol. MyTaq Red Mix was used for colony PCR to determine the presence of inserts using whole cells picked with a toothpick as a DNA template.

Table 4: List of primers used in this study.

Name	5' → 3' sequence
1206_cPCR_F	AAATTCGTCGGAGCCCGATCATAT
1206_cPCR_R	AATCGTGCGCTGAACCGCCG
1206_KO_F1	CACACAGGAAACAGCTATGACATGATTACGTCGTGGCGTTCCTGAAATT
1206_KO_F2	CGATCACTATCAGCACACCAA
1206_KO_R1	AGCAGGTTCTTGGTGTGCTGATAGTGATCGATGAAGTTGCCATAGCGGGC
1206_KO_R2	TAGAGGATCCCCGGGTACCGAGCTCGAATTTCTTCGACGTTCTTGATCGT
1206_KOscreen_F	AAATTCGTCGGAGCCGATCATAT
1206_KOscreen_R	AATCGTGCGCTGAACCGCCG
1975_21a_F	ATTATCATATGCAATCGTCGCCCGAGGTG
1975_21ab_R	ATTATGCGGCCGCGCCGCGGGTGCGCGAGCG

Name	5’ → 3’ sequence
1975_21b_F	ATTATCCATGGAACAATCGTCGCCCGAGGTG
1975_pBAD_F	ATATCAGATCTCAATCGTCGCCCGAGGTGAA
1975_pBAD_R	ATTAAGCTTGCCGCGGGTGCGCGAGC
1975_pMalp2e_F	AATATGAATCGCTCAATCGTCGCCCGAGGT
1975_pMalp2e_R	ATATAAGCTTTCAGCCGCGGGTGCGCGAGC
1975_RTPCR_F	ACGTCGTGACTTTCTGAA
1975_RTPCR_R	GTAGTACGACACGGTATGG
2047_21a_F	CGTATCATATGCAGTCGCCGATCGTCATC
2047_21a_R	ATTATAAGCTTCTTGGTCGGGCCGCCGCC
3458_21a_F	ATTATCATATGGACGACATCAAGCTACCA
3458_22b_F	ATTATCCATGATGACGACATCAAGCTACCA
3458_R	ATTATCTCGAGCTGCTCGAACCCGTTCGGCAG
3723/4/5_KOcPCR_F	GACAGCTACACCATCCAGCTC
3723/4/5_KOcPCR_R	TGTCGAGGAACAGGTCGTTGA
3723–5_KO_F1	CACACAGGAAACAGCTATGACATGATTACGATCGCGCTGTTCCTCGG
3723–5_KO_F2	CAGCTGATGCGGTTCAA
3723–5_KO_R1	CTTCTCGCCCTTGAACCGCATCAGCTGCAGGGACGGCGGCTGACAGT
3723–5_KO_R2	TAGAGGATCCCCGGGTACCGAGCTCGAATTGTGTTCGAACTGCTCGAGAA
3724_KO_F1	CACACAGGAAACAGCTATGACATGATTACGATCCACAGTGGCGCGACGAT
3724_KO_F2	CGAGAAATCTGGTAGTTCTACTGA
3724_KO_R1	ATTAGGTCAGTAGAACTACCAGATTTCTGCGCCGCTAAACGCGGCGAAAA
3724_KO_R2	TAGAGGATCCCCGGGTACCGAGCTCGAATTAGCGAATTGAAGATCAGCA
3724_KOcPCR_F	TATGCCAAGTACATCCTGG
3724_KOcPCR_R	TCTTCGAAGCCTTTAAGATA
3724_RTPCR_F	AGAAGAAATACGACACCG
3724_RTPCR_R	ATAGCTGATGAAGTTGATCTT
4421_21a_F	ATCGTACATATGGAGACCGAAATGCCCGACTTC
4421_21a_F	ATCGTACATATGGAGACCGAAATGCCCGACTTC
4421_21a_R	TATTAAGGCTTCGACGCCGCCTTCATGCC

Name	5’ → 3’ sequence
4421_21a_R	TATTGCGGCCGCCGACGCCGCCTTCATGCC
4510_21a_F	ATTATCATATGCAGCAGTTCGTCAACGTG
4510_21a_R	ATTATAAGCTT CTTCAGCACGCCCTTTTCCTT
4556_21a_F	ATTATCATATGCTCGAACTCAAGGTTGCC
4556_22b_F	ATTATCCATGGATCTCGAACTCAAGGTTGCC
4556_R	ATTATCTCGAGCTTGCTCGCACCAGCGGCAAG
aaeA_RTPCR_F	GATGGTGAACAGTATCTGTCCT
aaeA_RTPCR_R	GTCGTATTAGTCATTCTGGCC
aaeB_RTPCR_F	GCAATACAGCCAATAAATGTG
aaeB_RTPCR_R	TGCTAACCAACATATTCGCT
acrA_RTPCR_F	TAGTAACAGTCAAAACTGAACC
acrA_RTPCR_R	TTTCGCACTGTCGTATGT
acrB_RTPCR_F	TAATTTCTTTATCGATCGCC
acrB_RTPCR_R	GTCACTGTTAGAGGACATGTA
acrD_RTCPR_F	TCTTTATTGATCGCCCCATT
acrD_RTPCR_R	TATCGAGGCCGGTCATATTT
acrF_RTPCR_F	TATTCGACGACCGATATTTGC
acrF_RTPCR_R	TGTTCGATAACCTGCGTCAC
aldB_21a_F	ATTATCATATGACCAATAATCCCCCTTCAGCA
aldB_21a_R	ATTATAAGCTTGAACAGCCCCAACGGTTTATC
copA_RTPCR_F1	AAACTATCGACCTGACCCT
copA_RTPCR_R1	TTTGGGTGGCTTACAGAT
couA_cola1_F	ATTATCCATGGGACTGACCGGCAACGCT
couA_cola1_R	GTTCTGGATCCTTAAGACGGACGAACTTT
couAB_pBAD_F	ATATTCCATGGGGCTCACAGGAAACGCTCAG
couAB_pBAD_R	ATCGTGAATTCTCACTCCAGCGCAATCACATG
couB_cola2_F	ATTATCATATGATGGACGCTATGACCGAC
couB_cola2_R	ATTATCTCGAGTTATTCCAGAGCGATAACGTG
couP_RTPCR_F	ACCAAGTTCAAGCTATCC
couP_RTPCR_R	ATTTCCTTAGCCACGAAT
marA_RTPCR_F	CAATACTGACGCTATTACCA
marA_RTPCR_R	GGTTTCTTTTTTAAACATCC

Name	5’ → 3’ sequence
marR_21a_F	ATTATGCTAGCAAAAGTACCAGCGATCTG
marR_21a_R	ATTATCTCGAGCGGCAGGACTTTCTTAAG
marR_RTPCR_F	GTTCAATGAAATTATTCCATT
marR_RTPCR_R	CAGTTCAACCGGAGTAATAC
oxyS_RTPCR_F	TATTCGTGATCTTGAGTACCT
oxyS_RTPCR_R	CTTTAAGGACTTTCACCTCA
pK18mobsacB_seq_F	CGACTGGAAAGCGGGCAGTGA
pK18mobsacB_seq_R	TCGCCTTGCAGCACATCCCCCTT
rrsA_RTPCR_F	ACCAGGGTATCTAATCCTGTT
rrsA_RTPCR_R	GTTAATACCTTTGCTCATTGA
soxS_RTPCR_F	GCATATTGACCAGCCGCTTAA
soxS_RTPCR_R	TTACAGGCGGTGGCGATAATCGCT
tarP_acyc2_F	ATTATCATATGCGTAAAACCCTGCTGGCT
tarp_acyc2_R	ATTATCTCGAGTTACAGACCAGCGTCGTA
tarP_RTPCR_F	ACAAAACTGTCAACTGGAA
tarP_RTPCR_R	GATTGACGTAGGTGACATC
tarPQM_pBAD_F	ATATTCCATGGGGAGGAAGACTCTGCTCGCGC
tarPQM_pBAD_R	TAGAATCTCGAGCTACCCGATCATCCGCATCGG
tarQM_acyc1_F	ATTATCCATGGTTGACTCTGGTCTGGCTCTG
tarQM_acyc1_R	ATTATGGATCCCTACCCGATCATCCGCAT
yhbW_21a_F	ATTATGCTAGCACTGATAAAACCATTGCGTTTT
yhbW_21a_R	ATTATCTCGAGTCCCAACAACTCTTCCTTAAC
yqhD_21a_F	ATATTAGCTAGCAACAACTTTAATCTGCACACCCCA
yqhD_21a_R	ATATTACTCGAGGCGGGCGGCTTCGTATAT

2.2.4. DNA purification

PCR product purification was conducted using the QIAquick® PCR Purification Kit (Qiagen). If required, DNA was isolated from agarose gels using the QIAquick® Gel Extraction Kit (Qiagen).

2.2.5. Restriction digestion of DNA

Restriction digests of DNA were carried out using restriction enzymes (NEB) in 20 µL reactions according to manufacturer's instructions. For double digests the optimal buffer was selected. Reactions were typically performed at 37°C for 1.5 h. Reactions where digestion products were inserts required for subsequent cloning were heat inactivated at 65°C for 20 min and purified using the QIAquick PCR purification kit.

2.2.6. Phosphatase treatment

Digested vectors were 5'-phosphatase treated using Antarctic Phosphatase (NEB) at 37°C for 20 minutes, with subsequent heat inactivation at 65°C for 10 min, prior to purification using the QIAquick PCR purification kit.

2.2.7. DNA ligation

DNA ligation reactions were performed using T4 DNA ligase (NEB) according to manufacturer's instructions. Typically 10–50 ng of vector DNA was used per 10 µL reaction, with a 1:1 to 3:1 insert: vector molar ratio. An insert-free reaction was used as a negative control to determine level of undigested/re-ligated vector. Reactions were performed for a minimum of 10 min at room temperature. Ligation products were stored at 4°C until transformation.

2.2.8. Isothermal assembly

Isothermal assembly was used for ligation of three or more DNA fragments. Plasmids were linearised by the appropriate restriction enzyme and phosphatase treated. Insert fragments were amplified by PCR and purified. DNA fragments were added at equimolar concentrations with a total DNA volume of 5 µL, in a total reaction volume of 20 µL. Reactions consisted of the following reagents with final concentrations noted: 100 mM Tris-HCl, pH 7.5; 5 mM $MgCl_2$; 0.2 mM dGTP; 0.2 mM dCTP; 0.2 mM dTTP; 0.2 mM dATP; 10 mM dithiothreitol (DTT); 1 mM Nicotinamide adenine dinucleotide (NAD); 5% (w/v) polyethylene glycol- (PEG-)8000). Enzymes were added in following amounts per 20 µL reaction: T5 exonuclese (NEB), 0.1 µL of 8-fold diluted enzyme; Phusion High-Fidelity DNA Polymerase (NEB), 0.25 µL; and, Taq DNA Ligase (NEB), 2 µL. Reaction mixes were incubated at 50°C for 1 h.

2.2.9. Quantification of DNA

DNA concentrations were determined using a Jenway™ Genova Nano spectrophotometer.

2.3. Transformation of Competent Bacteria

2.3.1. Preparation of chemically competent bacteria

E. coli cells were made competent as per the Hanahan method (Hanahan, 1983). 50 ml *E. coli* cultures were grown in LB to mid-exponential phase, incubated on ice for 20 min and then pelleted by centrifugation (6000 g, 20 min, 4°C). Pellets were resuspended in 50 ml filter-sterilised RF1 solution (100 mM KCl; 50 mM $MnCl_24H_2O$; 30 mM potassium acetate; 10 mM $CaCl_2.2H_2O$; 15% (w/v) glycerol; adjusted to pH 5.8). Cell suspensions were incubated on ice for 15 min, centrifuged as before and resuspended in 8 ml filter-sterilised RF2 solution (10 mM MOPS; 10 mM KCl; 75 mM $CaCl_2.2H_2O$; 15% (w/v); adjusted to pH 6.8). Cell suspensions were incubated on ice for 20 min, aliquoted and stored at -80°C until required for transformation.

2.3.2. Transformation of competent bacteria

An aliquot of competent bacteria was thawed on ice. 2 µL of ligation reaction mixture or 1 µL of plasmid was mixed with 150 µL of competent cells and incubated on ice for 30 min. Cells were heat-shocked at 42°C for 1 min, then incubated on ice for 2 min. 1 ml of LB was added and cells grown at 37°C for 1 h. Following recovery growth, cells were pelleted at max speed, resuspended in fresh LB and spread on a solid LB plate with the appropriate antibiotic for selection. Insert-free ligation reaction mixtures were also transformed as a negative control to assess levels of uncut/religated vector. Following overnight growth colonies were screened by PCR.

2.4. Recombinant Protein Over-production, Preparation and Purification

This study entailed over-production of a range of proteins using several expression vectors in relevant protein production strains.

2.4.1. Over-production of recombinant proteins in E. coli

Genes encoding proteins of interest from *E. coli* and *R. palustris* were cloned into pET21a(+), for cytoplasmic production, and pET22b(+), for periplasmic production via incorporation of a N-terminal *pelB* leader sequence, vectors (Novagen). These vectors were

utilised for producing proteins with an incorporated C-terminal 6xHis-tag under control of the isopropyl β-D-1-thiogalactopyranoside- (IPTG-) induced T7 promoter. pET21a(+) was used primarily, with pET22b(+) used for SBPs that contained multiple disulfide bonds and experienced low solubility when produced in the cytoplasm. Native stop codons and signal sequences from genes were excluded for cloning, to allow for 6xHis-tag expression and correctly localised production respectively. Constructs were initially used for the transformation of competent *E. coli* DH5α for high-yield isolation by QIAprep Spin Miniprep Kit (Qiagen). Constructs were subsequently sequenced by GATC Biotech to confirm accurate cloning, before being transformed into BL21(DE3) for overexpression. For proteins that exhibited low solubility, pet21a(+) constructs were used for the transformation of *E. coli* Origami B, which harbours mutations in genes for thioredoxin reductase and glutathione reductase, for enhanced disulfide-bond formation in the cytoplasm.

Overnight cultures of overexpression strains were used to inoculate larger cultures to an OD_{600} of ~0.05 for protein production. These cultures were grown to a mid-log phase OD_{600} of 0.4–0.6 and expression induced by IPTG. Protein production proceeded for several hours and cells were harvested by centrifugation (16,000 x g, 10 min, 4°C) and pellets frozen at -80°C. Production was optimised by varying IPTG concentration (0.1–0.4 mM), growth temperature (18°C–37°C), and production time (1–24 hours); protein yield was assessed by SDS-PAGE (Section 2.4.3) of whole-cell samples, and protein solubility by SDS-PAGE of cell-free extracts. Once optimal conditions were identified, production was upscaled for purification.

RPA1975 proved to be largely insoluble when overproduced in *E. coli*. The *rpa1975* gene was cloned into the pMAL-p2x vector for periplasmic production of the protein with an N-terminal maltose-binding protein (MBP) fusion. Protein production was performed as with pET vectors but with the addition of 0.2% (w/v) glucose to growth medium.

The pBAD/HisB system was also used for protein production. Genes were inserted into the pBAD/HisB vector for arabinose-inducible expression, using the araBAD promoter, with an N-terminal 6xHis-tag. In contrast to the pET system, pBAD constructs were transformed into *E. coli* TOP10; protein production was performed as above but with induction by arabinose (0.001%–0.1% [w/v]).

2.4.2. Overproduction of protein in R. palustris

Overproduction of protein in *R. palustris* was achieved by insertion of genes into the pBBRBB-pBAD vector supplied by A. Hitchcock (University of Sheffield). This construct was cloned into *E. coli* S17-1 for conjugation into *R. palustris* strains and allowed arabinose-inducible overproduction of proteins.

2.4.3. One-dimensional SDS-polyacrlyamide gel electrophoresis (PAGE)

SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) gels were formed from a 12% resolving gel and a 6% stacking gel. The resolving gel was composed as follows: 2.5 ml of 30% acrylamide (w/v), 2.35 ml 1M Tris HCl (pH8.8), 1.25 ml MQ H_2O, 62.25 ml 10% (w/v) SDS, 6.25 µl tetramethylethylenediamine (TEMED) and 62.5 µl ammonium persulfate (APS). Once gently mixed by pipetting, the mixture was added to a gel cast and a thin layer of 100% H_2O-saturated butanol or ethanol was poured on top. Once the resolving gel has set, the alcohol was blotted off using blot paper and the stacking gel was added which was composed of the following: 750 µl of 30% acrylamide, 470 µl 1 M Tris HCl (pH 8.8), 2.46 ml MQ H_2O, 37.5 ml 10% SDS, 3.75 µl TEMED and 37.5 µl 10% (w/v) APS. Once poured a gel comb was added and left to set. Once set the gel comb was removed and the gel placed in the "Protein 3 electrophoresis apparatus" (Biorad). The tank was filled to the appropriate level with 1 x SDS-PAGE running buffer (0.1 M Tris base, 0.38 M glycine, 0.1% [w/v] SDS).

Protein samples for loading were prepared by mixing 20 µL of the cell-free extract with 20 µl of protein loading buffer (25 µl β-mercaptoethanol added to 475 µl of 2x protein loading buffer: 125 mM Tris-HCl pH 6.8, 4 % [w/v] SDS, 0.05 % [w/v] bromophenol blue, 20 % [w/v] glycerol). Samples were then boiled for 5 min at 100^{o}C and then centrifuged at 13,000rpm for 5 min to remove cell debris. Samples were then added in appropriate volumes to the SDS-PAGE gel with 5 µl PageRuler Pre-stained Protein Ladder Plus (Fermentas) as a molecular weight marker. The gel was run at 180 V until samples had migrated sufficiently. Once completed, gels were stained with coomassie blue staining solution (0.02% brilliant/coomassie blue, 50% methanol, 40% H_2O, 10% acetic acid) for a minimum of 1 h, shaking gently. The stain was poured off and a destaining solution added (50% methanol, 40% H_2O 10% acetic acid). The gel was then shaken gently until clear, distinct bands were visible.

2.4.4. Cell-free extract (CFE) creation by mechanical cell lysis

Harvested cell pellets were typically resuspended in 20 mM sodium phosphate buffer buffer, pH 7.4. If required for protein purification, 0.5 M NaCl and 40 mM imidazole were added to the buffer. Cell suspensions were kept ice cold at all times. For small-scale CFE creation (<5 ml), cell lysis was achieved by sonication: suspensions were subjected to four 15-20 s pulses of 10-15 microns amplitude on ice, in a Soniprep150 Ultrasonic Disintegrator (SANYO). For larger-scale CFE creation, cell lysis was achieved by French press: components of the French press equipment were pre-chilled for 30 min at 4°C and suspensions lysed at 10,000 psi. Following lysis, suspensions were centrifuged at 16,000 rpm for 10–20 min at 4°C, and the supernatant retained on ice (pellets discarded).

2.4.5. Protein purification via Nickel-Affinity Chromatography

Purification of 6xHis-tagged recombinant proteins was performed using 1 ml/5 ml HisTrap™ HP columns (GE Healthcare) according to manufacturer's instructions. Columns were stored in 20% ethanol and equilibrated with 20 mM sodium phosphate, 0.5 M NaCl, 40 mM imidazole, pH 7.4, before injection of CFE. Proteins were eluted over a 40 ml, 40 – 500 mM imidazole gradient into 1 ml samples. Purity of samples was assessed by SDS-PAGE. All buffers were degassed thoroughly and sterilised using a 0.22 µm filter.

2.4.6. Protein purification by Maltose-Affinity Chromtography

Purification of MBP-fused recombinant proteins was performed using 5 ml MBPTrap™ HP columns (GE Healthcare) packed with dextrin sepharose, according to manufacturer's instructions. Columns were stored in 20% ethanol and equilibrated 20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, pH 7.4, before injected of CFE. Proteins were eluted over a 40 mL, 0–10 mM maltose gradient into 1 ml samples.

2.4.7. Buffer exchange, desalting and concentration of purified protein

Vivaspin centrifugal protein concentrators (GE Healthcare, Sartorius) with an appropriate MW cut-off were used according to manufacturer's instructions. Columns were centrifuged at 4–25°C depending on stability of the protein of interest.

2.4.8. Refolding of purified proteins

Endogenously bound ligands were removed from purified periplasmic binding proteins prior to the proteins utilisation in characterisation or crystallisation experiments, by dialysis. Purified protein samples were dialysed against 1 l of 6 M urea for 12 – 20 h at 4°C to achieve unfolding. Unfolded protein was subsequently dialysed against 2 l of 50 mM Tris-HCl, 100 mM NaCl, pH 7.4 for 4 x 2 h, followed by final dialysis for 16 hrs. Precipitated protein was removed by centrifugation (13,000 g, 10 min, 4°C).

2.4.9. Gel filtration chromatography

For molecular size-based separation of proteins, in particular purification of MBP-fused RPA1975, high resolution gel filtration chromatography was conducted using a total volume of 5 ml of concentration protein solution post maltose-affinity chromatography. Protein solution was loaded onto a pre-chilled HiLoad Superdex 200 Gel Filtration Column (GE Healthcare) and eluted into the desired buffer according to manufacturer's instructions.

2.4.10. Protein concentration determination

Protein concentration of a purified protein of known amino acid sequence was determined using the Beer-Lambert Law, and the sample's OD_{280} - OD_{320} and an extinction coefficient calculated by the ProtParam tool (ExPASy). Protein concentration of CFE and soluble protein mixtures were determined using the Bio-rad Protein Assay Reagent kit (Bio-rad) as per manufacturer's instructions.

2.5. Protein Structure Determination by X-ray Crystallography

2.5.1. Preliminary crystallisation trials

Purified proteins were concentrated to an approximate concentration of 10 mg ml^{-1} in 10 mM Tris-HCl. Trials were conducted with both *apo*-protein and protein mixed with a potential ligand or substrate. Proteins were screened in the commercial 96-well, sitting-drop PACT, JCSG+ and Proplex plates for initial crystallisation trials by vapour diffusion. Liquid handling was performed by a Matrix_Hydra II Plus One crystallisation robot. Crystal growth was monitored and used to identify optimal conditions for crystallisation of each protein.

2.5.2. Follow-up crystallisation optimisation

Optimisation crystallisation experiments were based on conditions identified as favourable for crystal growth in the screening, and conducted in 24-well plates using the hanging-drop methodology. *Apo*-protein was screened with and without potential ligands, and 1 µL mixed with 1 µL of reservoir solution to form the drop; reservoir solution conditions were based on previous trials with PEG concentration, salt concentration and pH variants trialled in each plate.

2.5.3. Data collection and structural determination

Successfully grown crystals were extracted and placed in a cryoprotectant solution consisting of 25% (v/v) ethylene glycol in the appropriate reservoir solution. All data were collected from crystals at the Diamond Light Source (Harwell, UK). Data collection and processing was performed by Dr John Rafferty (University of Sheffield).

2.6. Biochemical Assays

2.6.1. CoA-ligase activity assay

The activity of the CoA-ligase enzyme, CouB, was assayed to confirm recombinant production of functional CouA and CouB protein in *E. coli*. Activity was measured at 30°C spectrophotometrically using a Shimadzu UV-240-recording spectrophotometer in a quartz cuvette (Hellma). The total reaction volume was 1 mL, and was composed of 50 mM Tris-HCl pH 8, 2 mM $MgCl_2$, 2 mM ATP, 400 µM CoA, and 0.01 – 2 mM hydroxycinnamate substrate, with 50 µl of *E. coli* CFE. All reagents excluding the CoA were mixed in the cuvette, the absorbance blanked, and the CoA added to initiate the reaction. Formation of the enoyl-CoA thioester was measured at the specific wavelength for maximal absorbance of each hydroxycinnamoyl-CoA product: *p*-coumaroyl-CoA = 333 nm, cinnamoyl-CoA = 311 nm, caffeoyl-CoA and feruloyl-CoA = 346 nm.

2.6.2. Thermal shift assay

Purified periplasmic binding proteins were screened against a range of potential ligands using thermal shift assays (ThermoFluor) in MicroAmp Optical 96-well Reaction plates (Thermo Scientific). Each 50 µL reaction was typically composed of the SYPRO dye reagent (Invitrogen) at working concentration, 10 µM protein, and 60 µM ligand in 50 mM Tris-HCl, 0.1 M NaCl, pH 7.4. A ligand-free control was performed for each protein, and a 1 mg ml^{-1} lysozyme control performed for each plate. All reactions were conducted in triplicate.

Protein concentrations were varied for optimisation due to variable protein stability. Reaction mixes were made up in low-light with the dye added immediately before use to minimalise photodenaturation.

Thermal denaturation of proteins was measured via fluorescence emission using a Stratagene MX3005p thermo-cycler (Agilent), and curves plotted using MxPRO qPCR software (Agilent). Data were analysed using Microsoft Excel to determine a T_M for each protein with a ligand for comparison to the ligand-free control.

2.6.3. Tryptophan fluorescence spectroscopy assays

Ligand binding of periplasmic binding proteins was characterised by measurement of the quench (or, rarely, enhancement) of intrinsic fluorescence emission of proteins on binding. All measurements were performed using a Cary Eclipse Fluorescent Spectrophotometer (Varian) in 3 ml volumes in a quartz cuvette with magnetic stirring. Proteins were assayed at 0.2 µM in 10 mM Tris-HCl, pH 7.4, with an excitation wavelength of 280 nm (5 mM slit width) and an emission wavelength of 300–400 nm (20 mM slit width). Initially potential ligands were added in excess to protein solution and fluorescence emission recorded over this range, with a quench or enhancement of fluorescence indicative of binding. Ligand titrations were performed with an emission wavelength that corresponded to the peak fluorescence emission for a given protein (typically ~340 nm), with ligands titrated in from 0.01 µM to 1 µM. The % of change in fluorescence was plotted as a function of ligand concentration and binding characteristics calculated by iterative fitting to the quadratic equation for tight binding using the Levenberg-Marquadt non-linear “least squares” method in Microsoft Excel. Titrations and calculations were conducted in triplicate.

2.7. *R. palustris* mutagenesis

The suicide vector pK18mobsacB was used for marker-free gene deletion in *R. palustris*. 400–500 bp DNA fragments homologous to the sequences upstream and downstream of the target gene, or genes, were amplified by PCR, and ligated into pK18mobsacB by isothermal assembly. This construct was used for the transformation of competent *E. coli* S17-1.

2.7.1. Conjugation of R. palustris

Freshly grown CGA009 *R. palustris* was used to inoculate 8 ml of PYE and grown photo-heterotrophically for 3-4 days. This starter culture was then used to inoculate a 100 ml

culture and grown for 48 h, and 40 ml harvested by centrifugation (6000 x g, 15 min, 30°C), the supernatant removed and the pellet resuspended in 100 µL fresh PYE. *E. coli* S17-1 harbouring the pK18mobsacB construct was grown overnight, and 20 µL added to the *R. palustris* suspension and thoroughly mixed by pipetting. 50 µL of this mixture was carefully pipetted onto a well-dried PYE plate (no antibiotic) and incubated for 12–16 h at 30°C. Following incubation each mixed-bacteria spot was scraped off the plate, streaked out onto a kanamycin 80 µg ml^{-1} + chloramphenicol 20 µg ml^{-1} M22 plate, and grown at 30°C until single colonies appear (1–2 weeks).

2.7.2. Recombination during kanamycin/sucrose selection

Colonies were streaked out on kanamycin 80 µg ml^{-1} + chloramphenicol 20 µg ml^{-1} M22 plates once more, and grown as before. Single colonies were then used to inoculate 3 ml kanamycin 80 µg ml^{-1} + chloramphenicol 20 µg ml^{-1} liquid PYE and grown for 3–5 days aerobically, in the dark, until the culture had turned red. Subsequently 300 µL of this culture was used to inoculate 80 ml kanamycin 80 µg ml^{-1} + chloramphenicol 20 µg ml^{-1} M22 and grown for 3–5 days as before. Once grown, a range of dilutions 1:10 to 1:10,000,000 were spread on solid chloramphenicol 20 µg ml^{-1} M22 + 10% sucrose plates. Plates were grown for 1–2 weeks until single colonies appear.

These single colonies were picked and stabbed onto replica solid chloramphenicol 20 µg ml^{-1} M22 + 10% sucrose, and then kanamycin 80 µg ml^{-1} chloramphenicol 20 µg ml^{-1} M22 + 10% sucrose plates, and grown for 1–2 weeks. Colonies which were sucrose resistant and kanamycin sensitive were screen by colony PCR using primers ~100 bp upstream and downstream of the fragments in the original construct to ascertain gene deletion. CGA009 was used as a control. Colonies where gene deletion was successfully identified were streaked out on chloramphenicol 20 µg ml^{-1} PYE plates, and subsequently grown for the purpose of gDNA extraction for final confirmation of deletion.

2.8. Quantitative RT-PCR

2.8.1. Bacterial RNA extraction

E. coli RNA was extracted using the SV Total RNA Isolation System kit (Promega) per manufacturer's instructions with modifications. Overnight cultures were used to inoculate at and OD_{600} of 0.05 or 0.1 grown to mid-log phase for a maximum for 2 hrs, centrifuged at max speed and the pellets resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8) with 1 mg ml^{-1} lysozyme (Sigma Aldrich) added, and incubated for 5 min. Following RNA

extraction by the kit, a second DNase-treatment step was performed using the Turbo DNA-free kit (Ambion) and purified RNA stored at -80 °C until required.

R. palustris RNA was extracted using the TRI reagent method with some modifications. Freshly growing *R. palustris* in RCV with 10 mM succinate was used to inoculate RCV media containing 5 mM succinate at an OD_{660} of 0.25 and grown for 16 h. A sterile solution of the carbon source being investigated was then added to triplicate cultures and grown for 4 h. 10 ml of culture was added to a pre-chilled vessel containing 1.2 ml 100% ethanol and 62.5 µL phenol, and centrifuged at 12,000 g for 10 min at 4°C. The supernatant was completely removed and pellets stored at -80°C until required for RNA extraction.

Frozen pellets were thawed on ice and resuspended in 1 ml TRI reagent (Sigma). GlycoBlue™ Coprecipitant was added to a final concentration of 50 µg ml^{-1} and samples incubated at room temperature for 5 min. 200 µL chloroform was added before vigorous mixing and incubation on ice for 10 min. Samples were centrifuged at 12,000 g for 15 min at 4°C, and the clear aqueous layer removed to a fresh, sterile tube. 500 µL ice-cold isopropanol and 90 µL ammonium acetate were added, samples mixed and incubated at 4°C for 1 h, and then centrifuged as before. The supernatant was completely removed and the pellet resuspended in 1 ml ice-cold 75% (v/v) ethanol to wash. Samples were centrifuged at 7,500 g for 5 min at 4°C, the ethanol removed and the pellets air-dried until residual ethanol had completely evaporated. Purified pellets were resuspended in nuclease-free dH_2O and DNase treated using Turbo DNA-free kit (Ambion). Purified RNA was stored at -80 °C until required.

All RNA concentrations were determined using a Jenway™ Genova Nano spectrophotometer.

2.8.2. qRT-PCR

One-Step quantitative RT-PCR was performed using the SensiFAST SYBR Lo-ROX One-step Kit (Bioline) in MicroAmp Optical 96-well Reaction plates (Thermo Scientific) with each biological replicate performed in triplicate. Primers of genes of interest were designed to amplify ~200 bp at the N-terminus, with a T_m of ~60°C. Each reaction consisted of 10 µL of 2xSensimix SYBR-Lo-ROX, 0.2 µL reverse transcriptase, 0.4 µL RNAse inhibitor, 0.25 µM primers and 20 ng template RNA, made up to a total volume of 20 µL with nuclease-free dH_2O. For RT-PCR of *E. coli* genes, *rrsA* expression was used as a control, and for *R. palustris* genes, *rpoD* was used. A standard curve for each pair of primers was produced using

serially diluted (10^4 –10^6) gDNA with reactions performed in duplicate. Template-free control reactions were performed for each pair of primers. PCR amplification was performed using a Stratagene MX3005p thermo-cycler (Agilent). The cycling protocol was as followed: 10 min at 45°C; 2 min at 95°C; 40 cycles of 20 s at 95°C, 30 s at 55°C and 20 s at 72°C. Threshold cycle (C_T) values were determined by MxPRO qPCR software (Agilent) and relative gene expression was calculated using the comparative method as described in User Bulletin #2: ABI PRISM 7700 Sequence Detection System (https://www.core-facility.uni-freiburg.de/qpcrddpcr/lc480obj/ub2).

2.9. Proton Nuclear Magnetic Resonance (^{1}H-NMR)

2.9.1. Sample preparation

Samples were taken from the growth cultures of interest and centrifuged at 13,000 rpm for 5 min. The supernatant was removed and frozen at -20°C until required (< 1 week). On thawing 450 µl of each sample was mixed with 50 µl deuterated water (D_2O). 1 µl trimethylsilyl propionate (TSP) was added to a final concentration of 200 µM for use as a calibration standard. Samples were mixed thoroughly and placed into ultra-precision, 5 mm rounded-bottom NMR tubes for use.

2.9.2. ^{1}H-NMR and ^{13}C-NMR analysis

Data were collected using an 800 MHz Bruker Avance NMR spectrometer at 25 °C with a z gradient coil. Chemical shifts were determined in parts per million (ppm), and referenced to TSP as in internal standard of 0 ppm. One-dimension ^{1}H-NMR analysis was performed with water suppression via presaturation or excitation sculpting to reveal peaks hidden by the water resonance (~4.8 ppm). The recycle delay was 10 s to enable accurate integration of the peaks. Two-dimensional ^{1}H-^{1}H total correlation spectroscopy (TOCSY) analysis was performed with a mixing time of 120 ms to determine connectivities within components of the supernatant. Two-dimensional ^{1}H-^{13}C heteronuclear single-quantum correlation spectroscopy (HSQC) analysis was performed to establish the ^{13}C chemical shifts of the components of the supernatant to confirm chemical structure.

3. Harnessing the aromatic metabolism of *Rhodopseudomonas palustris* for the production of vanillin

3.1. Background

Rhodopseudomonas palustris's aromatic degradation capabilities are well-established with several peripheral pathways characterised for the funnelling of a variety of aromatic compounds into the central benzoyl-CoA ring-cleavage pathway. This includes pathways for the metabolism of the lignin-derived phenylpropanoic acids, or hydroxycinnamic acids, such as cinnamate, coumarate, caffeate and ferulate. Initially, two alternative pathways were proposed for the conversion of the hydroxycinnamates into benzoyl-CoA: the β-oxidation pathway and the non-β-oxidation pathway. Both proceed via the same first step that sees the substrate activated by a CoA-ligase producing CoA derivatives of the carboxylic acid group (Figure 15). The β-oxidation pathway then proceeds through shortening of the aliphatic side chain with the CoA remaining attached. In contrast, the non-β-oxidation pathway cleaves acetyl-CoA to shorten the chain producing an aldehyde that is subsequently oxidised before undergoing a second CoA-ligation step. The initial two-steps of this resemble the CoA-independent retro-aldol reduction pathway mentioned in section 1.4.3, and where ferulate is the starting substrate, the aldehyde intermediate of the pathway is vanillin.

A transcriptomics- and proteomics-based study identified a cluster of genes significantly upregulated in the presence of coumarate that contained two putative non-β-oxidation genes (Figure 17)(Pan et al., 2008). The *p*-coumarate CoA-ligase (CouB) and enoyl-CoA hydratase/lyase (CouA) have been fully characterised highlighting their potential for use in biotransformation of ferulate to vanillin (Hirakawa et al., 2012). Furthermore, the gene cluster contains two distinct, complete transport systems: an ABC system, CouPSTU; and a TRAP system, TarPQM. Both of these systems have been characterised through binding studies into their respective solute binding proteins, showing their role in the uptake of hydroxycinnamates, including ferulate, in *R. palustris.* This has involved combinations of fluorescence spectroscopy and isothermal titration calorimetry showing their high affinity for the substrates including ferulate (K_d values determined by fluorescence titrations: CouP, 15 ± 6 nM; TarP, 15 ± 3 nM) (Salmon et al., 2013).

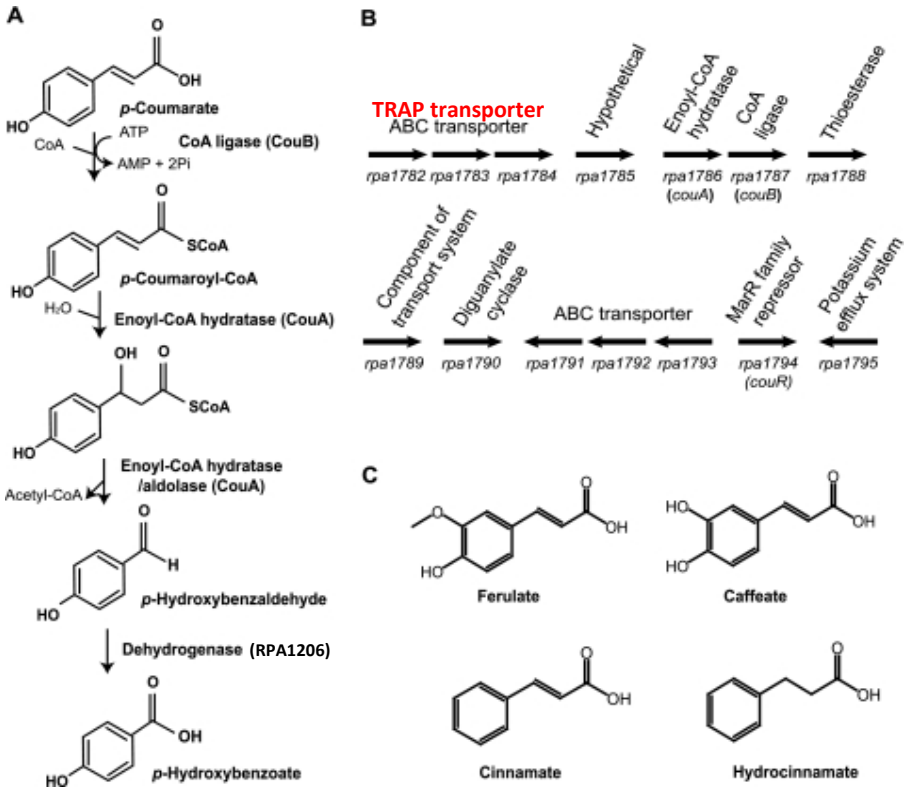


Figure 17: The gene cluster from *R. palustris* discussed in this chapter (Hirakawa et al., 2012). A) The non-β-pathway for the degradation of *p*-coumarate to *p*-hydroxybenzoate via an aldehyde intermediate. B) The RPA1782-1795 gene cluster including enzymes and two transport systems; in Hirakawa et al (2012) paper RPA1782-1784 is incorrectly annotated as an ABC transporter. C) Hydroxycinnamate alternatives to coumarate as substrate of the pathway.

The use of a recombinant strain of *E. coli* possessing genes encoding the enzymes for the biotransformation of ferulate to vanillin has been widely explored (Section 1.4.4); however, no approach considers the uptake of ferulate – a substrate not inherently metabolised by *E. coli*. The success of *E. coli* as a platform for vanillin production suggests that there is a means for exogenously provided ferulate to enter the cytoplasm for conversion. To improve vanillin production in *E. coli* we hypothesise that expression of confirmed, high-affinity ferulate uptake systems, such as the CouPSTU and TarPQM systems, in addition to the required enzymes would result in an increase in ferulate uptake and thereby facilitate the biotransformation process. Therefore this chapter reports attempts to engineer a strain of *E. coli* with the ability to convert exogenous ferulate into vanillin through the recombinant production of the *R. palustris* CouA and CouB enzymes, and uptake systems.

The potential use of the gene product RPA4421, which is the only other gene annotated as a putative coumarate CoA-ligase in the *R. palustris* genome, as an alternative enzyme will be assessed through recombinant production, purification and characterisation.

Additionally, the potential of *R. palustris* itself as a biocatalyst for vanillin production is explored. The advantage of this is the bacterium already possesses the enzymes and inherently expresses the transport systems, with expression promoted in response to the CoA derivative being produced. As mentioned previously (1.4.2), the problem with using microbial species that inherently metabolise aromatic compounds for vanillin production is the downstream degradation of the product for its use as a carbon source. The "omics" study of Pan et al. (2008) showed that the gene *rpa1206*, encoding an aldehyde dehydrogenase, was significantly upregulated in the presence of *p*-coumarate. As the only aldehyde dehydrogenase responding in this way, it is proposed that the gene product RPA1206 is responsible for the subsequent step of the non-β-oxidation pathway downstream of CouAB. Therefore, enhanced vanillin accumulation in *R. palustris* via knock-out of the *rpa1206* gene was planned. Additionally, plasmid-based overexpression of the CouAB enzymes, to enhance the conversion of ferulate to vanillin, was investigated.

3.2. Results

3.2.1. Recombinant production of the CouA and CouB enzymes in E. coli

The first step to assessing the potential for using *R. palustris* proteins in the biotransformation of ferulate to vanillin was the cloning of the two-gene *couAB* enzyme operon for their co-expression in *E. coli*. The full 2706 bp sequence, which includes both genes and the 8 bp sequence between them, was amplified by PCR from *R. palustris* CGA009 gDNA. The primers used were couAB_pBAD_F and couAB_pBAD_R, which added an *Nco*I and *Eco*RI restriction enzyme sites for ligation into the pBAD/HisB vector. The insert was inserted into the plasmid excluding the N-terminal polyhistidine tag and the enterokinase recognition site, as these proteins were not to be purified. The pBAD:*couAB* construct was transformed into *E. coli* TOP10 for arabinose-induced expression; the titratable nature of the P_{BAD} promoter allowed for more sensitive control of expression via the addition of arabinose compared to the IPTG-induced expression achieved through use of the lac repressor. Another advantage of controlling transcription via the P_{BAD} promoter is that arabinose occurs in lignocellulosic biomass — e.g. as a branched-sugar monomer in hemicelluloses — and thus could be a source of auto-induction.

An extensive series of protein production trials were conducted, using SDS-PAGE to assess production of the two enzymes with whole-cell samples and the soluble fractions. Cultures were induced with a range of arabinose concentrations: 0.002, 0.02 and 0.2 % (w/v); incubated at different temperatures: 25 °C and 37 °C; induced at different cell densities: OD_{600} of 0.4 – 0.8; and samples taken at different time points over a 24 h induction/protein production timeframe: 0 h (negative control), 1 h, 3 h, 5 h and 24 h. Initially, the whole-cells were used for SDS-PAGE analysis to determine whether production was occurring. Figure 18A shows an SDS-PAGE analysis where protein expression was detected; cells induced with the arabinose concentrations shown and incubated at 25 °C displayed a band at ~28.5 kDa from 5 h onwards indicating production of CouA, though no clear band was visible at ~68 kDa that would have indicated simultaneous production of CouB. The level of CouA appeared to increase up to a maximum at 24 h with no sign of degradation affecting yield. The soluble fractions of these samples, produced by careful sonication of cell suspensions and centrifugation, were also analysed by SDS-PAGE. Figure 18B shows the soluble fractions of the 24 h samples and indicates the presence of a reasonable yield of CouA and a faint band that suggested CouB was being produced, but at a much lower yield.

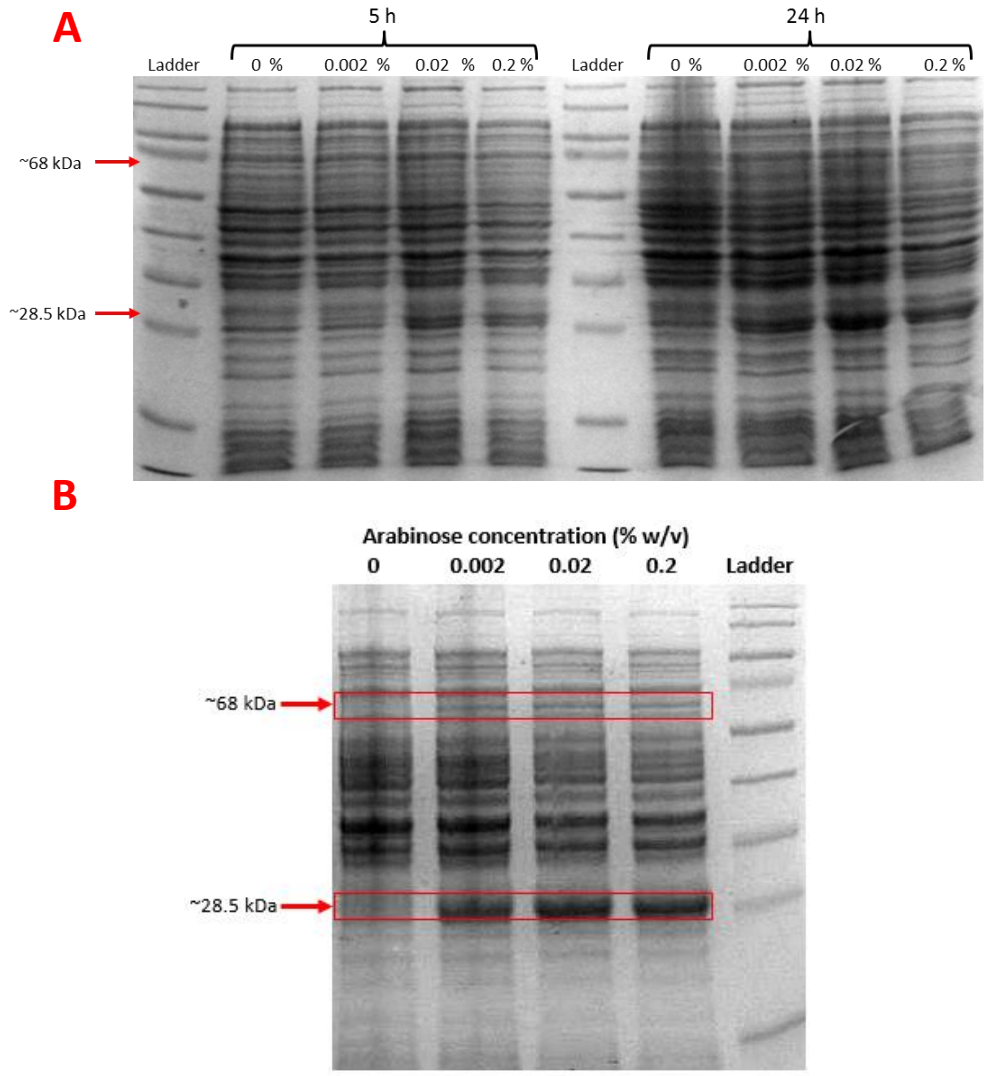


Figure 18: Production of recombinant CouA and CouB in *E. coli* TOP10. SDS-PAGE showing results of the production trials of CouA and CouB using the arabinose-inducible pBAD vector. A) The whole-cell samples taken after 5 h and 24 h of induction at 25 °C, by the arabinose concentrations labelled. A band is visible at ~28.5 kDa, but not at ~68 kDa. The ladder used can be seen on page 257. B) The soluble-fraction samples with protein production at 25 °C for 24 hours; the arabinose concentrations used for induction are shown. The band at 28.5 kDa shows a good yield of CouA; and smaller band was visualised at 68 kDa suggesting production of CouB was successful but with a much lower yield. The ladder used can be seen on page 255.

3.2.2. Spectrophotometric assays to assess CouA and CouB activity in E. coli

These results had indicated that the recombinant enzymes were being produced in *E. coli* to a certain degree, and the next step was to determine whether the enzymes exhibited catalytic activity. Assays were conducted *in vitro* using the cell-free extract, created by

sonication and centrifugation, of *E. coli* TOP10 that had produced the proteins. A spectrophotometric assay has been used extensively to characterise CoA ligases (including CouB) based on the measured increase in absorbance at the wavelength that corresponds to maximal absorbance of the enoyl-CoA thioester product of the ligation reaction, e.g. feruloyl-CoA formation measured at 346 nm. No assays exist for characterisation of the enoyl-CoA hydratase/lyase activity that CouA harbours, due to the enoyl-CoA thioester substrate being impossible to obtain in large, pure quantities.

Figure 19 highlights that these assays produced unexpected results that suggested the recombinant enzymes were not exhibiting the expected activity. On addition of CoA to reaction mixtures containing ferulate as the substrate (and with two other hydroxycinnamates, coumarate and caffeate) the absorbance decreased, as shown in Figure 19A. Only when cinnamate was the substrate was there an increase in absorbance at the wavelength relevant for the enoyl-CoA product. The rates of decrease observed were substrate-concentration dependent, though higher concentrations were required than expected based on the kinetic constants for CouB previously published, i.e. Figure 19A shows the rate of decrease with substrates at 0.5 mM, though K_m values were reported at ~20 µM. Furthermore, Figure 19C shows that after an extended period of time the rate tails off and the reaction reaches completion presumably due to depletion of the ferulate substrate, indicating there is enzymatic activity. Figure 19D shows that with cinnamate as the substrate the increase in absorbance appeared to plateau before then decreasing. Throughout these experiments the rates of the change in absorbance exhibited large variability. A purified sample of CouB provided by Abrar Akbar (University of Sheffield) was used as a positive control to confirm the assay conditions were correct, with this enzyme sample showing activity indicated by an increase in absorbance with all four substrates.

The hydroxycinnamate substrates, such as ferulate, do absorb at the respective wavelengths of their enoyl-CoA derivatives though at a much lower level. As the absorbance is blanked with the substrate in the reaction mixture prior to reaction initiation, the decrease exhibited likely corresponded to a decrease in concentration of the substrate, but not resulting in the enoyl-CoA thioester accumulating. Based on the protein production yields shown in Figure 18, the high ratio of abundance of CouA to CouB in the CFE used in the assays could result in any enoyl-CoA thioester produced being rapidly converted to the aldehyde form resulting in no detectable accumulation of the enoyl-CoA thioester intermediate. However, vanillin also absorbs significantly at the wavelength measured for feruloyl-CoA, indicating that no vanillin was being formed (Figure 19B).

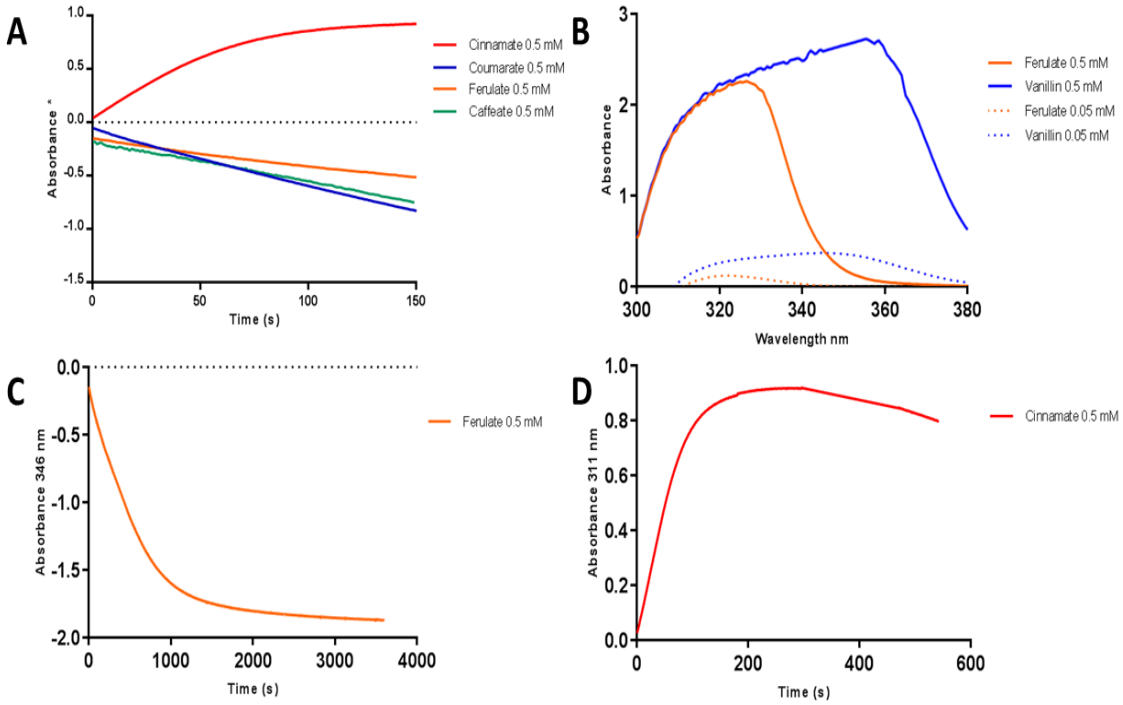


Figure 19: Spectrophotometric assessment of CouAB activity. A) Reactions were initiated by addition of CoA to a mix containing the cell-free extract of CouAB-producing *E. coli* and the hydroxycinnamate substrates shown. Absorbance was measured at the wavelength of maximal absorbance for the expected enoyl-CoA product: *p*-coumaroyl-CoA = 333 nm, cinnamoyl-CoA = 311 nm, caffeoyl-CoA and feruloyl-CoA = 346 nm. B) Absorbance spectra of ferulate and vanillin at pH 7.4 highlighting vanillin production should be observed by measurement at 346 nm. C) The reaction with ferulate as a substrate reaches completion after approximately 1 h. D) The reaction with cinnamate showed an initial increase in absorbance at 311 nm indicating cinnamoyl-CoA formation; on reaching a maximal absorbance this was followed by a slow decrease.

Based on these results it was apparent that this attempt to express the *couA* and *couB* genes in *E. coli* would not form the basis for biotransformation of ferulate to vanillin in *E. coli*. The low level of recombinant protein production of CouB in particular was a likely contributing factor. This could have been due to the cloning strategy; there is an 8 bp sequence in between the *couA* and *couB* genes with the ribosomal binding site (RBS) expected to occur at the 3' region of the *couA* gene. This RBS site may not be as effective in *E. coli* as in *R. palustris* and thus may have resulted in poor translation efficiency (while the CouA was translated efficiently due to the optimised RBS on the vector). The *R. palustris* genome is GC-rich (>60%) compared to *E. coli*, and therefore sub-optimal codon usage may also have reduced protein production. As CouA's enoyl-CoA hydratase/lyase activity cannot

be characterised it is impossible to determine the catalytic turnover rates of the two enzymes; the difference in these may actually mean optimal metabolism is achieved in *R. palustris* with a high CouA to CouB ratio, however.

Ultimately, an alternative cloning and recombinant protein production strategy was required in order to achieve the expected biotransformation of ferulate to vanillin by CouA and CouB.

3.2.3. Assessment of RPA4421 as an alternative CoA-ligase

In the *R. palustris* genome *rpa4421* is the only gene, other than *couB* and *rpa1707*, annotated as a coumaroyl-CoA ligase, and therefore potentially could catalyse the first step of hydroxycinnamate degradation. As no attempt to biochemically determine its specificity and activity had been published, characterisation of the enzyme's CoA-ligase activity was performed to assess whether overproduction of this enzyme would be preferable to that of CouB in the biotransformation of ferulate to vanillin.

The *rpa4421* gene was amplified by PCR using the 4421_21a_F and 4421_21a_R primers for insertion into the pET21a(+) vector using the *Nde*I and *Not*I restriction sites; the stop codon was not included in the PCR product to allow for incorporation of the C-terminal polyhistidine tag. The construct was transformed into *E. coli* BL21(DE3) for overproduction, with a high yield and solubility of the protein achieved through induction using 0.4 mM IPTG and incubation at 37 °C. Figure 20A shows the optimal production time was 3 h, with apparent degradation of the recombinant protein decreasing yield after that time point. These conditions were used to upscale production for purification by nickel-affinity chromatography, with purification via imidazole gradient achieving a yield of >35 mg from 1 l of *E. coli* culture (Figure 20B).

The spectrophotometric CoA-ligase assay described above was used to assess purified RPA4421's activity with ferulate, however no change in absorbance at 346 nm was recorded suggesting no feruloyl-CoA ligase activity. Therefore, it was clear that recombinant product of RPA4421, as an alternative to CouB, would not be viable for biotransformation of ferulate to vanillin.

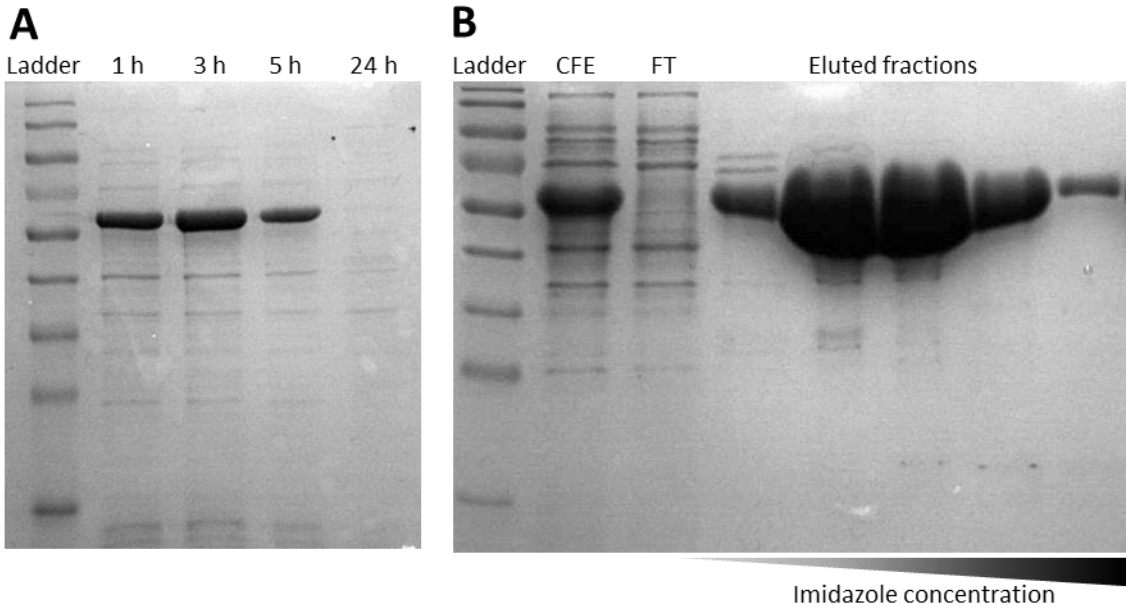


Figure 20: Production and purification of RPA4421. A) SDS-PAGE of soluble fractions taken from RPA4421 production trials where cells were induced with 0.4 mM IPTG and incubated at 37 °C. The band ~56 kDa indicates a high yield of soluble RPA4421; degradation of protein or the high yield causing the protein to become insoluble results in no recombinant protein seen in the 24 h sample. B) SDS-PAGE showing the steps of purification of RPA4421 using nickel-affinity chromatography. The band at ~56 kDa in the cell free extract (CFE) shows soluble RPA4421 presence, which is absent in the flow through (FT) of unbound protein. An imidazole gradient then results in the bound RPA4421 eluting from the nickel column, with these eluted fractions showing an extremely high yield of relatively pure RPA4421 was achieved. The ladder used are shown on page 257.

3.2.4. Recombinant production of the TarPQM and CouPSTU transport systems in E. coli

The SBPs of both the TRAP and ABC transport systems had been produced in *E. coli* and purified for biochemical characterisation by Salmon et al. (2013), but the membrane protein subunits had never been engineered to produce a fully functioning transport system (Salmon et al., 2013). Therefore, the first step was to confirm successful production of all the recombinant proteins in *E. coli*.

The genes encoding TarPQM are arranged in an "orthodox" fashion (as shown in Figure 17), and therefore the full 3,064 bp sequence, including the 85 bp sequence between the *tarP* and *tarQ* genes, was amplified from *R. palustris* gDNA using the tarPQM_pBAD_F and tarPQM_pBAD_R primers. *NcoI* and *XhoI* restriction sites were used for insertion into the

pBAD vector; due to an *NcoI* cut site within the *tarQ* gene partial restriction digests with gel extraction of the DNA fragment of the correct size required to produce the pBAD:*tarPQM* construct. This construct was transformed into *E. coli* TOP10 as before.

As with the production trials for the enzymes, conditions were varied in an attempt to find the optimal conditions for production of TarP, TarQ and TarM, using SDS-PAGE to assess yield. Despite a large number of attempts with every condition varied no production of the transport system was ever confirmed. Therefore, production of the CouPSTU transport system in *E. coli* – the cloning of which is more complicated due to the arrangement of the genes (Figure 17) – was not pursued until a new strategy for engineering the transport systems into *E. coli* was devised.

3.2.5. Use of Duet™ vectors

Due to the problems that arose from the previous attempts to produce both the enzymes and the transporters, an alternative approach to plasmid-based expression of the genes from *R. palustris* in *E. coli* was attempted using the Duet™ vectors (Novagen). This series of vectors are designed for co-expression of multiple target genes in *E. coli* through the use of two multiple-cloning sites (MCS) per vector, each controlled by *T7lac* promoters and with individual RBSs. The vectors each carry replicons and antibiotic resistance markers compatible with each other allowing for co-expression of multiple constructs in the same cell. For co-expression of the *R. palsutris* enzymes and transport systems, pCOLADuet™-1 and pACYCDuet™-1 were used; pCOLADuet-1 uses the COLA replicon and a kanamycin resistance cassette, while pACYCDuet-1 uses the P15A replicon and a chloramphenicol resistance cassette.

In order to alleviate the potential problem of the 8-bp intragenic sequence and the *R. palustris* RBS, the *couA* and *couB* genes were amplified by PCR separately. Primers couA_cola1_F and couA_cola1_R were used to produce a PCR product for insertion of *couA* into pCOLADuet™-1 between the *NcoI* and *NotI* sites of MCS1, and primers couB_cola2_R and couB_cola2_R were used to produce a PCR product for insertion of *couB* between the *NdeI* and *XhoI* sites of MCS2.

A similar approach was used for cloning of the genes for the TarPQM transporter. There is an 85 bp sequence between *tarP* and *tarQ* in the *R. palsutris* genome that may impede protein production in *E. coli*; therefore, that *tarP* gene was amplified separately from the *tarQM* genes, for cloning. Primers tarP_acyc2_F and tarP_acyc2_R were used to produce a

PCR product for insertion of *tarP* into pACYCDuet™-1 between the *NdeI* and *XhoI* sites of MCS2, and primers tarQM_acyc1_F and tarQM_acyc1_R were used to produce a PCR product for insertion of the *tarQM* sequence between the *NcoI* and *BamHI* sites of MCS1.

Figure 21 shows the results of using these two Duet vector expression systems in tandem for the recombinant production of CouA, CouB, TarP, TarQ and TarM. *E. coli* BL21(DE3) cells cotransformed with both constructs were induced with 0.4 mM IPTG and incubated at 25 °C, with samples taken over a 24 h timecourse for analysis by SDS-PAGE. The bands at ~68 kDa indicated that the yield of CouB produced was much higher than achieved through use of the pBAD construct; however, CouA production appeared to be much lower with no clear band visible ~28.5 kDa. TarP produced a high yield as displayed by the very large band at ~35 kDa from 1 h of production onwards, though only a small proportion of the protein proved soluble. No significant bands indicated production of the membrane proteins TarQ and TarM, up until 24 h when a prominent band at ~25 kDa indicated a high yield of the smaller membrane protein TarQ, which was also clear in the soluble cell fraction.

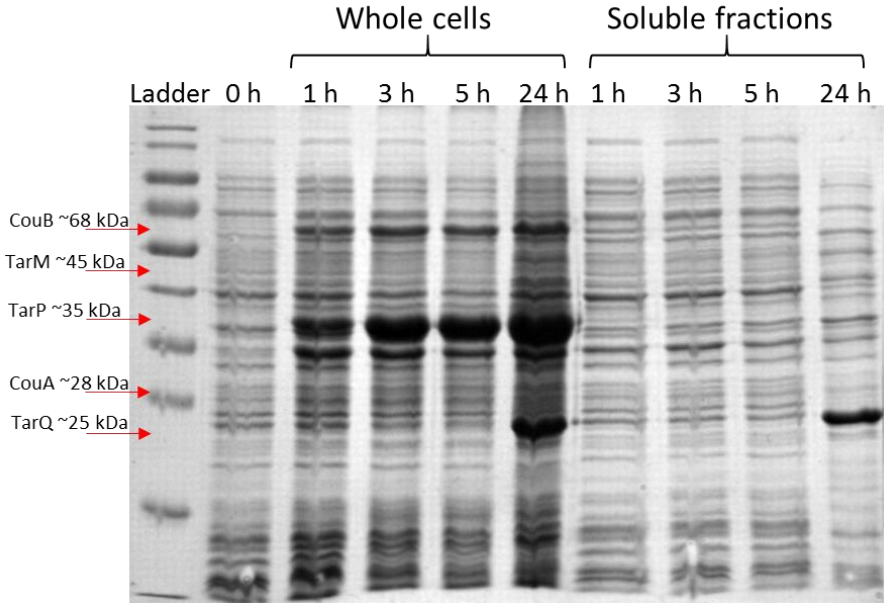


Figure 21: SDS-PAGE assessing recombinant production of *R. palustris* proteins using Duet vectors. The protein production of the two *R. palustris* enzymes and the TRAP transport system in *E. coli* through use of the pCOLADuet:*couAcouB* and pACYCDuet:*tarPtarQM* constructs in tandem, was assessed at 25 °C with induction by 0.4 mM IPTG. SDS-PAGE shows the proteins present in whole cells and soluble fractions of the samples of each time point shown; the expected sizes of each of the five proteins are marked.

3.2.6. Overexpression of the couAB genes in R. palustris

As *R. palustris* inherently expresses the genes for the enzymes and transporters with transcription shown to be upregulated by enoyl-CoA intermediates, and is capable of correct production and assembly of the transport systems, its potential as a biocatalyst for vanillin production was assessed in comparison with other endogenous ferulate metabolising bacteria.

A relatively simple approach to potentially improving the rate of biotransformation would be to overexpress the genes encoding the enzymes using a plasmid overexpression vector. A plasmid for arabinose-inducible overexpression in purple bacteria, pBBRBB had been created and was provided by Andrew Hitchcock (University of Sheffield). In this work, the genetic region encoding CouA and CouB with the intergenic region intact, was amplified by PCR from *R. palustris* gDNA with addition of *Nde*I and *Not*I restriction sites for insertion into this vector; the stop codon was not removed as expression of the C-terminal 6xHis-tag was not required. This pBBRBB-pBAD:*couAB* construct was transformed into *E. coli* S17-1 cells for conjugative transfer into *R. palustris*, with kanamycin selection and colony PCR screening used for confirmation.

R. palustris cells harbouring this construct were grown photoheterotrophically in rich medium to mid-log phase and overexpression was induced with the addition of 0.1% (w/v) arabinose. Samples were taken at time-points over a 48 h post-induction period and analysed with SDS-PAGE to confirm overproduction of the enzymes; the results are shown in Figure 22. This shows that CouA was overproduced to a high level from 4 h onwards; overproduction of CouB appeared less successful with a less prominent band from 24 h onwards.

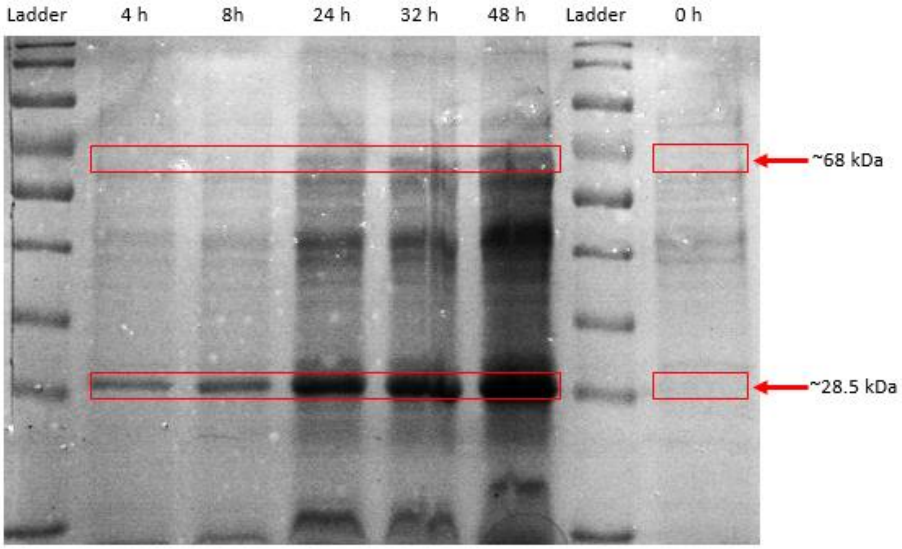


Figure 22: SDS-PAGE showing overexpression of CouA and CouB in *R. palustris.* *R. palustris* cells transformed with the pBBRBB-pBAD:*couAB* construct were grown overnight to a mid-log phase and induced with 0.1 % (w/v) arabinose, t = 0 h. SDS-PAGE shows increasing levels of protein bands at ~28.5 kDa and ~68 kDa, corresponding to CouA and CouB, respectively. The ladder used can be seen on page 255.

3.2.7. Creation of a Δ1206 mutant strain of R. palustris

The main disadvantage of using *R. palustris* for vanillin accumulation is that vanillin is predicted to be an intermediate in a metabolic pathway for catabolism of phenylpropanoids, with the gene *rpa1206* encoding the putative aldehyde dehydrogenase responsible for oxidation of vanillin to vanillic acid, based on its regulation. Therefore, an unmarked single-gene *Δ1206* knockout mutant of *R. palustris* was created. The ~500 bp regions flanking the *rpa1206* were amplified by PCR and inserted into the pK18mobsacB suicide vector linearised by *Eco*RI, using isothermal assembly. The pK18mobsacB:*Δ1206* construct was transformed into *E. coli* S17-1 and conjugated into *R. palustris*. The vector contains a kanamycin resistance cassette, which was used for confirmation of conjugation. These positive colonies were picked and further grown in minimal media in the presence of kanamycin that was used to select for successful completion of the first homologous recombination event. The vector also contains the *Bacillus subtilis sacB* gene encoding the levansucrase enzyme that produces levan from sucrose; levan accumulates in the periplasm and is toxic to gram-negative bacteria. This allows for counter-selection, and the addition of sucrose to minimal media provided selective pressure for the *R. palustris* cells to undergo a second recombination event whereby the plasmid was looped back out of the chromosome. This produced the desired, markerless chromosomal deletions of *rpa1206* and undesired "back-mutant" wildtype cells. Therefore colony PCR using *rpa1206* primers was used to distinguish between the two; and colonies that showed deletion of the gene were cultured for gDNA extraction. The gDNA of these potential mutants was used as the template for PCR amplification of *rpa1206* with the results shown in Figure 23. The KO gDNA produced a PCR product of size ~800 bp, and the WT a gene product of size ~2000 bp, indicating the knock out ~1200 bp from the *rpa1206* gene.

The growth of the *Δ1206* mutant was assessed with succinate, coumarate and ferulate as the carbon source (Figure 24). The mutant was capable of growth with all three of these carbon sources, and growth was comparable to that exhibited by wildtype *R. palustris*; this was not unexpected as a molecule of acetyl-CoA is produced from ferulate metabolism prior to the expected step catalysed by RPA1206.

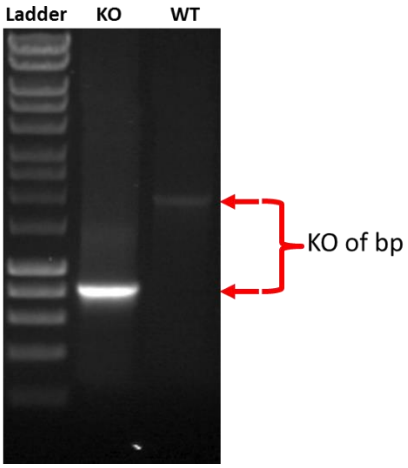


Figure 23: Confirmation of the knockout of the RPA1206 gene. gDNA extracted from a potential knockout mutant (KO) and wildtype (WT) was used as the template for PCR with primers outside of the *rpa1206* gene. The KO gDNA resulted in a ~800 bp product, and the WT gDNA resulted in a ~2000 bp product.

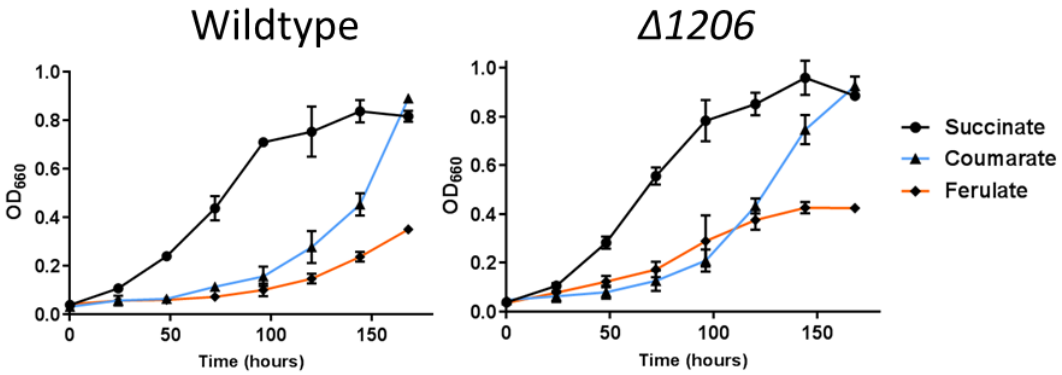


Figure 24: Growth of Δ1206 *R. Palustris mutant.* Wildtype and the *Δ1206* knock-out mutant strain were grown photoheterotrophically in minimal media with succinate, coumarate and ferulate as the sole carbon source. Growth was measured by measurement of the OD_{660} of the cultures. Cultures were grown in triplicate and the means plotted; the standard deviation is shown as error bars.

3.2.8. Biotransformation of ferulate to vanillin in E. coli cultures

Despite the problems identified with production of the *R. palustris* proteins in *E. coli*, experiments were conducted to determine if production of vanillin from ferulate was possible with the levels of protein production obtained. Overnight cultures of *E. coli* BL21(DE3) containing no plasmid, containing just the pCOLA:*couAcouB* construct, and containing both the pCOLA:*couAcouB* and pACYC:*tarPtarQM* constructs were used to

inoculate 100 ml of fresh medium at an OD_{600} of 0.05. M9 minimal media — supplemented with magnesium sulfate, thiamine and glucose as the carbon source — was used to facilitate nuclear magnetic resonance (NMR) analysis of the supernatant composition. Cells were grown to an OD_{600} of 0.4 representing mid-log phase growth, and induced with 0.4 mM IPTG, with ferulate added to a final concentration of 0.5 g l^{-1}. Cultures were moved to 25 °C and grown for 24 hours with 1 ml samples taken after 1 h, 3 h, 5 h and 24 h. Samples were centrifuged and the supernatant removed and frozen until analysis. All cultures were grown, and samples taken, in triplicate (Figure 25). Growth was slow due to the use of minimal media and incubation at 25 °C; growth was retarded further in the cultures containing expression constructs likely due to the antibiotics added (kanamycin, 25 µg ml^{-1}; chloramphenicol 20 µg ml^{-1}) and the added demand of overproduction of recombinant proteins.

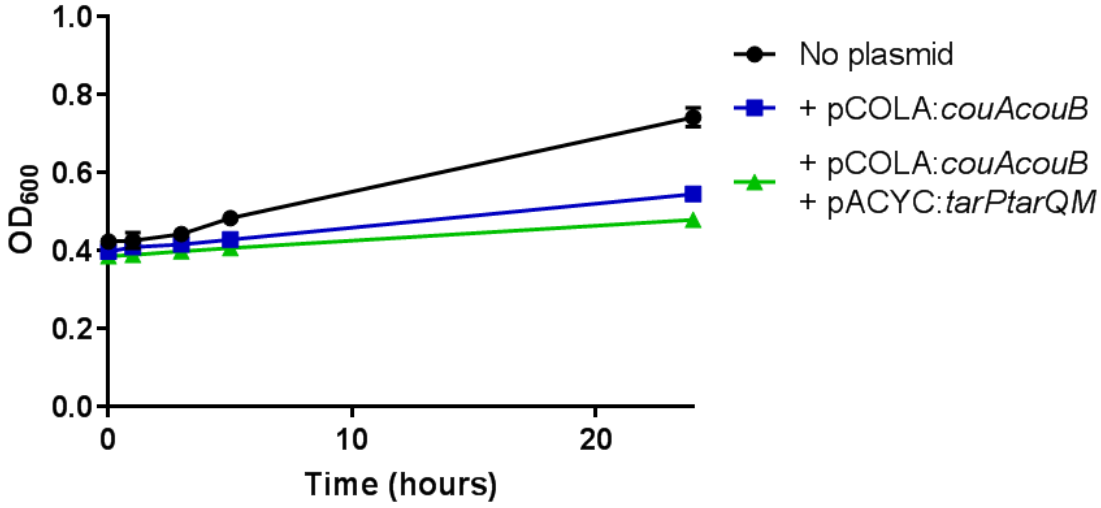


Figure 25: Growth of *E. coli* cultures during biotransformation. Each of the three cultures were inoculated at t = 0 h and the OD_{600} measured at each supernatant sampling time point. Values shown are the mean of the three cultures with error bars representing standard deviation.

3.2.9. Biotransformation of ferulate to vanillin in R. palustris cultures

A similar approach was used to determine the ability of *R. palustris* to convert ferulate into vanillin, and to assess the effect of overexpressing the *couA* and *couB* genes and the knockout of the *rpa1206* gene on this process. Therefore, four *R. palustris* cultures CGA009 (WT), CGA009 harbouring pBBRBB-pBAD:*couAB*, *Δ1206* knockout mutant, and *Δ1206* knockout mutant harbouring pBBRBB-pBAD:*couAB* were used for assessment. Starter cultures were grown for 5 days in minimal medium with succinate as the carbon source and

used to inoculate small batch cultures of 160 ml. These were grown photoheterotrophically for 24 hours to an optical density representative of mid-logarithmic phase in minimal medium with an excess of succinate as the sole carbon source. Ferulate, at a final concentration of 0.5g l^{-1}, and arabinose for induction of the overexpression of the enzymes, were added simultaneously at a time designated t = 0 h. Arabinose was also added to the cultures that did not harbour the overexpression plasmid for consistency during NMR analysis. 1 ml samples of the culture were taken over an extended time course of 96 h after 16 h, 24 h, 48 h, 72 h and 96 h. These were centrifuged and the supernatants removed and frozen until analysis (Figure 26). Growth was slightly reduced in cultures harbouring the overexpression construct, likely due to a combination of the presence of 20 µg l^{-1} kanamycin and the added demand due to overproduction of the enzymes.

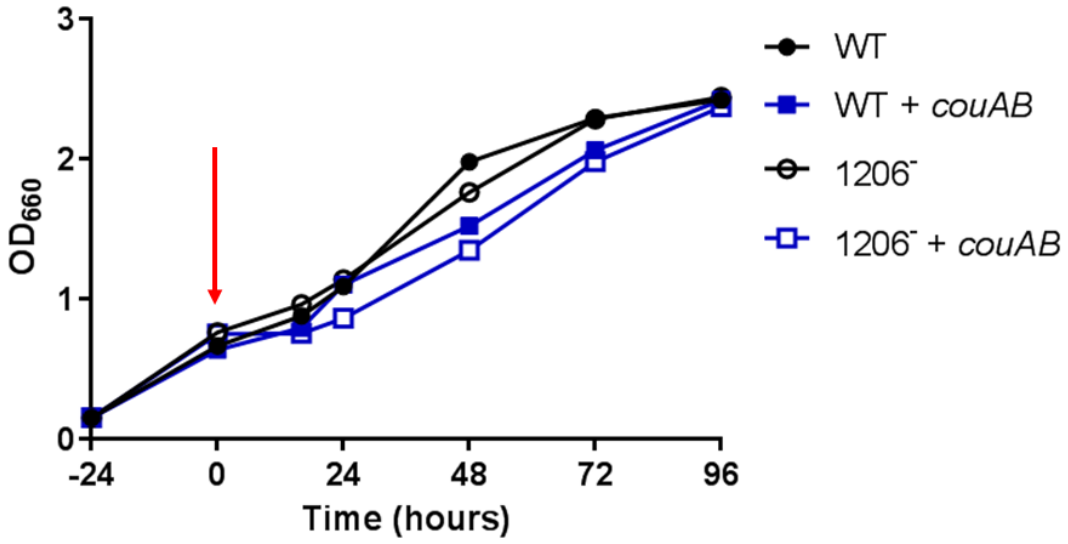


Figure 26: Growth of *R. palustris* cultures during biotransformation. Each of four *R. palustris* cultures were initially inoculated at t = -24 (OD_{660} of 0.15). After 24 h ferulate and arabinose was added to each culture (shown by the red arrow). Values shown are the mean of three cultures with error bars representing standard deviation.

3.2.10. NMR analysis of E. coli biotransformation experiment

Due to difficulties quantifying ferulate and vanillin using spectrophotometric methods, proton nuclear magnetic resonance (^{1}H-NMR) was used to determine changes in the extracellular concentration of ferulate, vanillin and other metabolites of the *E. coli* cultures during the biotransformation experiments. The *E. coli* supernatant samples were mixed with D_2O and trimethylsilyl propionate (TSP) added as an internal standard for 0 ppm. Samples were analysed by one dimensional ^{1}H-NMR and the spectra are shown in Figure 27, Figure 28, and Figure 29. Peaks visible in the spectra were assigned to compounds, and quantitation of compounds was performed by integration of the peaks; this allowed for calculation of the abundance of the compounds relative to the initial ferulate concentration (Figure 30). Reference spectra of compounds discussed can be found on page 258.

The spectra for the *E. coli* lacking expression of the *R. palustris* genes showed no change in peaks identified as corresponding to ferulate, highlighting that *E. coli* is inherently unable to metabolise ferulate and so its concentration remained unchanged (Figure 27). The only changes in the spectra were the decrease in the peaks corresponding to glucose over time and an increase in a peak corresponding to acetate, which accumulates during metabolism of glucose.

In contrast, the cells that were overexpressing the genes encoding CouA and CouB showed a very different outcome (Figure 28). After 24 h almost all the ferulate appeared to have been transformed with very little remaining. However, it was very clear that no vanillin had accumulated during the time course as shown by anabsence of a peak at ~9 – 10 ppm that would correspond to the aldehyde group. An originally unidentified aromatic compound appeared to be accumulating instead. The peaks at 6.9 ppm and 7.2 ppm in a 2:1 ratio indicated an aromatic ring, while a methoxy group was still present but the peak had shifted slightly to 3.7 ppm (from 3.9 in ferulate). Adjusting the water suppression identified a peak at ~4.8, and a new peak had also appeared ~2.7 ppm. Two-dimensional ^{1}H-^{1}H total correlation spectroscopy (TOCSY) analysis and two-dimensional ^{1}H-^{13}C heteronuclear single-quantum correlation spectroscopy (HSQC) analysis were utilised to determine whether these peaks were related and confirmed they were all part of the same chemical structure. These peaks did not correspond to the likely vanillin detoxification products vanillyl alcohol or vanillic acid.

Searching the literature regarding ferulate metabolism and its degradation products analysed by using NMR identified a study investigating ferulate degradation in

Agrobacterium fabrum (Campillo et al., 2014). As with the *R. palustris* enzymes utilised here, they had predicted *A. fabrum* would convert ferulate to vanillin as an intermediate of its degradation pathway, but actually discovered accumulation of another compound, identified as 4-hydroxy-3-methoxyphenyl-β-hydroxypropionic acid (HMPHP). They used a combination of mass spectrometry and NMR to identify the HMPHP, with their NMR data matching the compound the *E. coli* expressing *couA* and *couB* was producing. Therefore, it was clear that the recombinant CouA and CouB enzymes produced in *E. coli* were catalysing the accumulation of HMPHP. This compound is the free acid form of HMPHP-CoA, the predicted intermediate that the two-step action of CouA proceeds via with conversion of feruloyl-CoA to vanillin.

The only other difference in the spectra between the *E. coli* with no plasmid (Figure 27), and the *E. coli* producing CouA and CouB (Figure 28) was an increase in a peak ~1.4 ppm corresponding to an accumulation of lactate. The additional expression of the *tarP, tarQ and tarM* genes showed similar results but did not appear to have resulted in an increase in rate of ferulate uptake; HMPHP still appeared to be the product of ferulate metabolism, but this process actually occurred slower, rather than faster (Figure 29). This could be due to the lower cell density as shown in Figure 25. There are several explanations for induction of this expression cassette not influencing ferulate metabolism:

a) as Figure 21 highlighted, production of the membrane proteins was not confirmed, though the levels may be too low to detect using SDS-PAGE;
b) TarP should be produced in abundance, but its low solubility when overproduced in *E. coli* may prevent it functioning;
c) the native *R. palustris* signal sequence for TarP may not result in periplasmic targeting of the SBP;
d) or, the membrane proteins may be produced but complete TarPQM transport systems may not be structurally arranged correctly in the membrane.

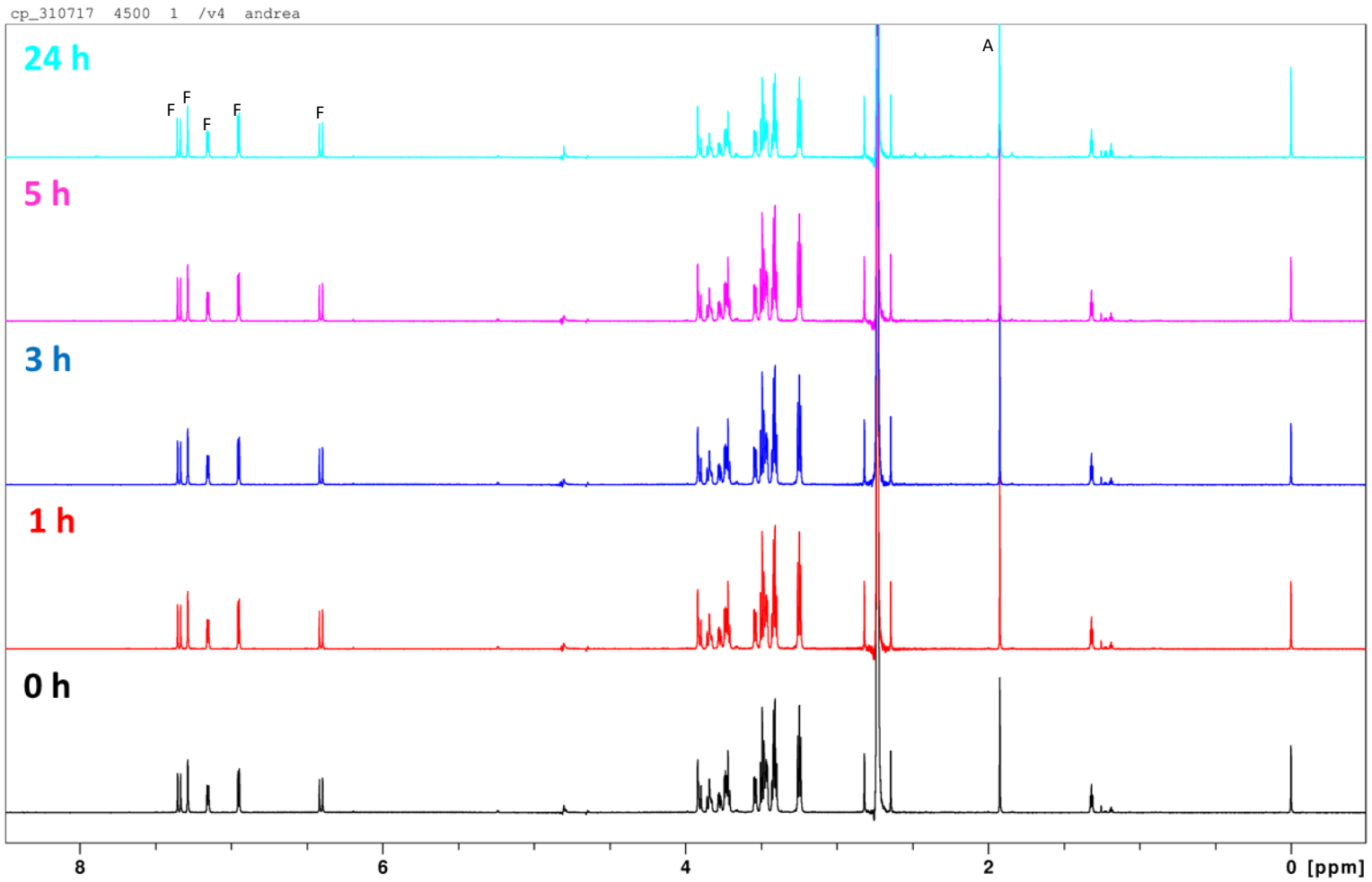


Figure 27: ^{1}H-NMR spectra of *E. coli* (control) supernatant samples. The peaks between 6.4 and 7.5 ppm represent the aromatic ring of ferulate (F), and are unchanged over 24 h; the peak at 3.9 ppm is the methoxy side group. The only change is the peak at ~1.9 ppm is the acetate (A) forming from metabolism of glucose (peaks around 3.5 ppm) as the carbon source. Spectra were normalised to the TSP peak at 0 ppm, with H_2O suppression applied.

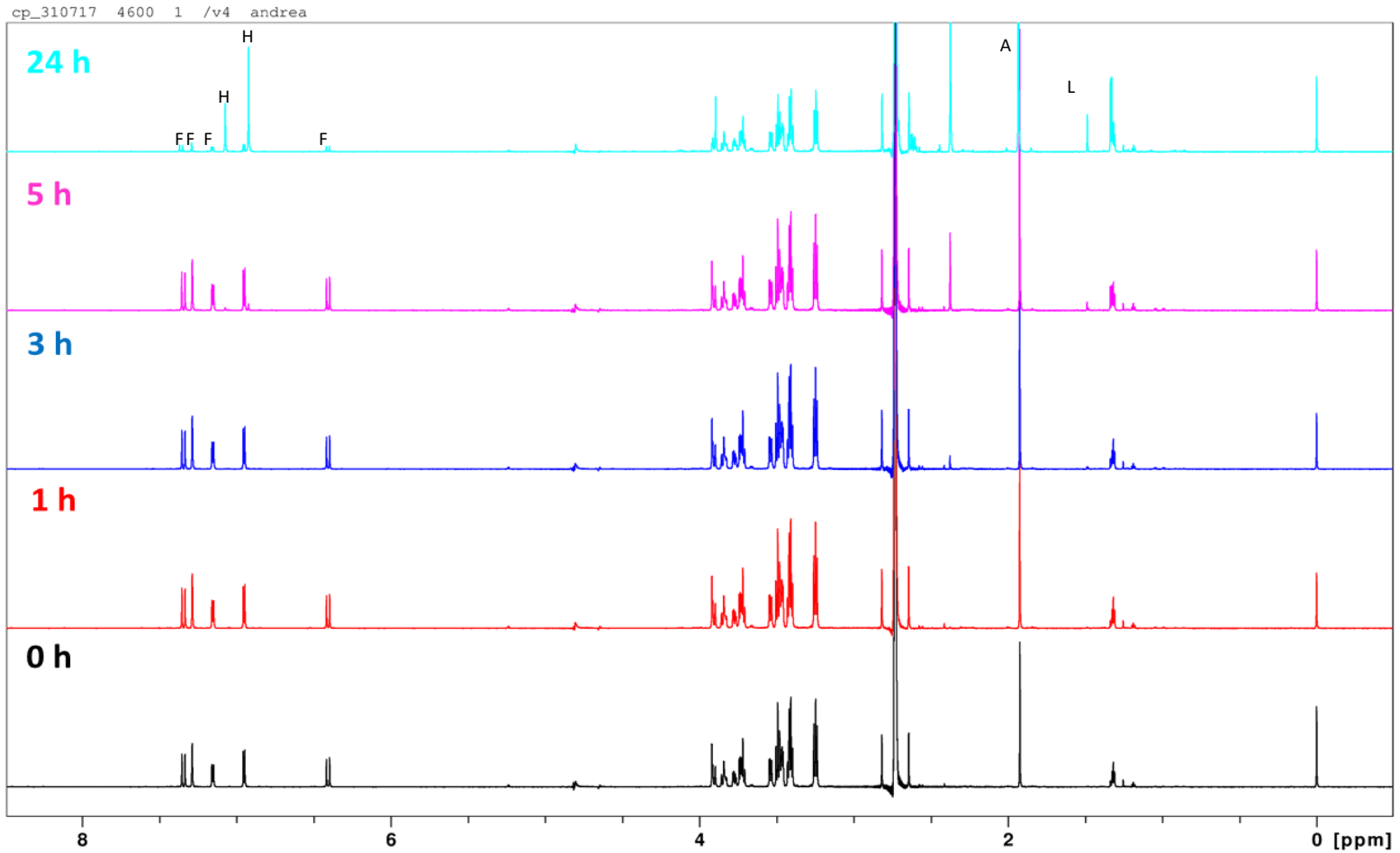


Figure 28: ^{1}H-NMR spectra of *E. coli* (pCOLA:*couAcouB*) supernatant samples. The plasmid-based expression of the *couA* and *couB* genes resulted in the ferulate (F) peaks decreasing to almost nothing by 24 h, with two new peaks in the aromatic region appearing at ~6.9 and ~7.1 ppm in a 2:1 ratio (H). Additionally the methoxy peak shifts from 3.9 to 3.8 ppm and a peak at 2.7 ppm appears, which were shown to be structurally linked by 2D-NMR. The peak growing at ~1.3 is expected to represent lactate (L). Spectra were normalised to the TSP peak at 0 ppm, with H_2O suppression applied. (A) = acetate.

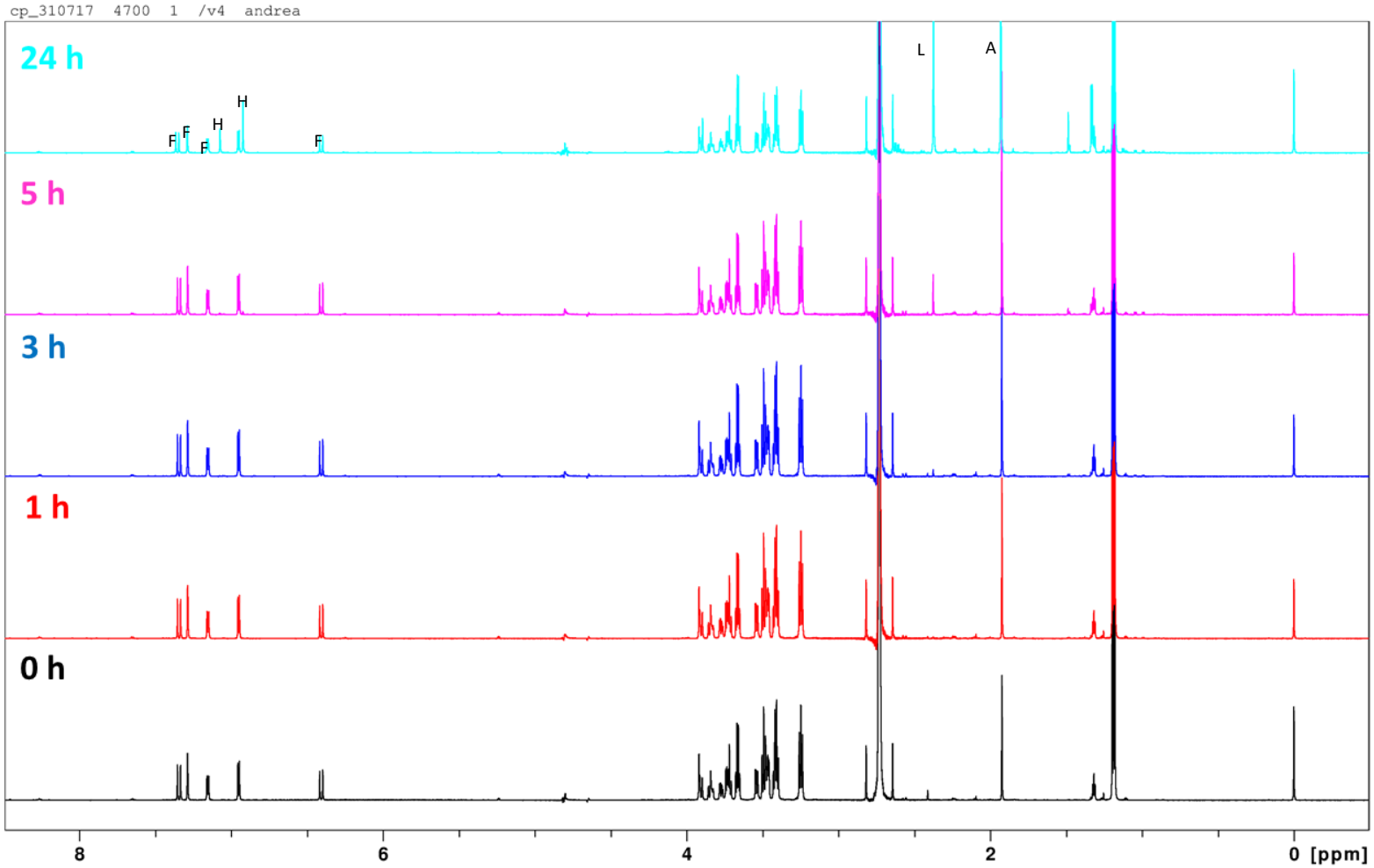


Figure 29: ^{1}H-NMR spectra of *E. coli* (pCOLA:*couAcouB* pACYC:*tarPtarQM*) supernatant samples. Additional expression of the *tarP* and *tarQM* genes showed similar spectra (H) to those in Figure 28. The only difference is the consistent peak at 1.4 ppm, which is the ethanol from the addition of chloramphenicol. Spectra were normalised to the TSP peak at 0 ppm, with H_2O suppression applied. (A) = acetate; (L) = lactate; (F) = ferulate.

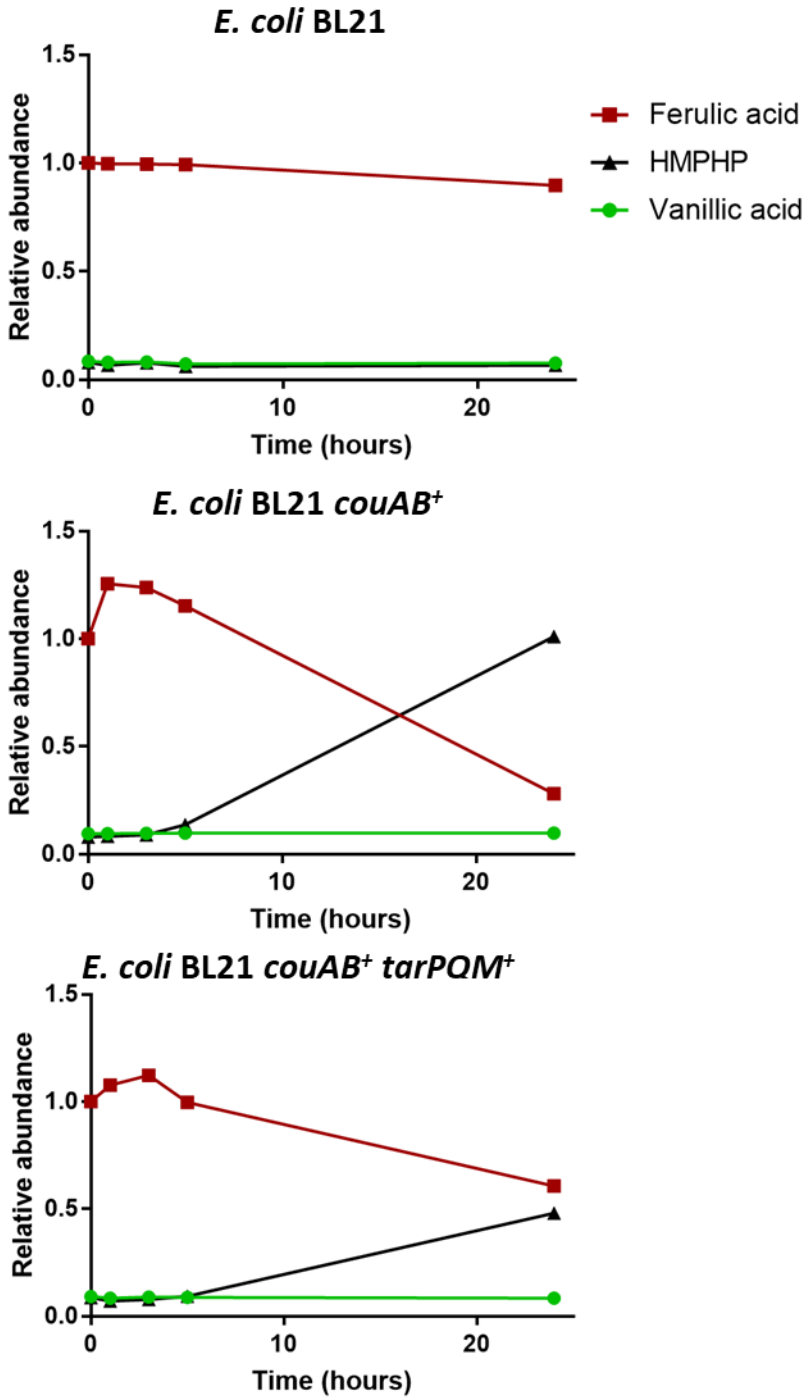


Figure 30: Changes in relative abundance of compounds from supernatants of *E. coli* cultures. The relative abundance of ferulate, HMPHP and vanillic acid were calculated by integration of each compound's specific peaks: 7.497 and 7.573 ppm for vanillic acid; 7.185, 7.309 and 7.38 ppm for ferulate; and 7.105 ppm for HMPHP. Values at each time point are scaled relative to the TSP peak at 0 ppm. Relative abundances shown here are scaled by the integral of the ferulate peak at t = 0.

3.2.11. NMR analysis of R. palustris biotransformation experiment

The supernatant samples from the *R. palustris* biotransformation cultures were analysed by NMR using the same methodology as the *E. coli* samples; Figure 31, Figure 32, Figure 33 and Figure 34 show the ^{1}H-spectra for each sample and the change in concentrations of compounds identified are compared in Figure 35. The wildtype culture showed that by 16 h the majority of the ferulate had been degraded, with two aromatic compounds accumulating (Figure 31). One of these was HMPHP, while the other was identified as vanillic acid; as with the *E. coli* no vanillin was detected (at any time point). By 24 h the ferulate had been entirely depleted, the level of HMPHP had decreased slightly and the level of vanillic acid had increased. By 48 h the HMPHP had been depleted too, with vanillic acid representing 100% of the aromatic compounds detected. The spectra also showed degradation of succinate over the 96 h.

The NMR data for the pBBRBB-pBAD:*couAB* culture showed that the additional, arabinose-induced overexpression of the genes resulted in the same products forming but the process occurring more rapidly; by 16 h the ferulate had been depleted with higher levels of HMPHP and vanillic acid (Figure 32). HMPHP levels then decreased and vanillic acid levels increased, though this occurred slower than in the plasmid-free culture. This is likely due to the lower cell density and growth rate shown by the cells harbouring the plasmid and overproducing the protein (Figure 26). The *Δ1206* knockout and *Δ1206* knockout pBBRBB-pBAD:*couAB* cultures showed nearly identical results to their equivalent wildtype strains, and with no vanillin accumulating (Figure 33 and Figure 34).

Overall, these data suggest that *R. palustris* metabolises ferulate to vanillic acid via HMPHP-CoA, with the enzymes CouA and CouB involved but no evidence that vanillin is an intermediate in this pathway. The aldehyde dehydrogenase RPA1206 does not have a significant role in this process.

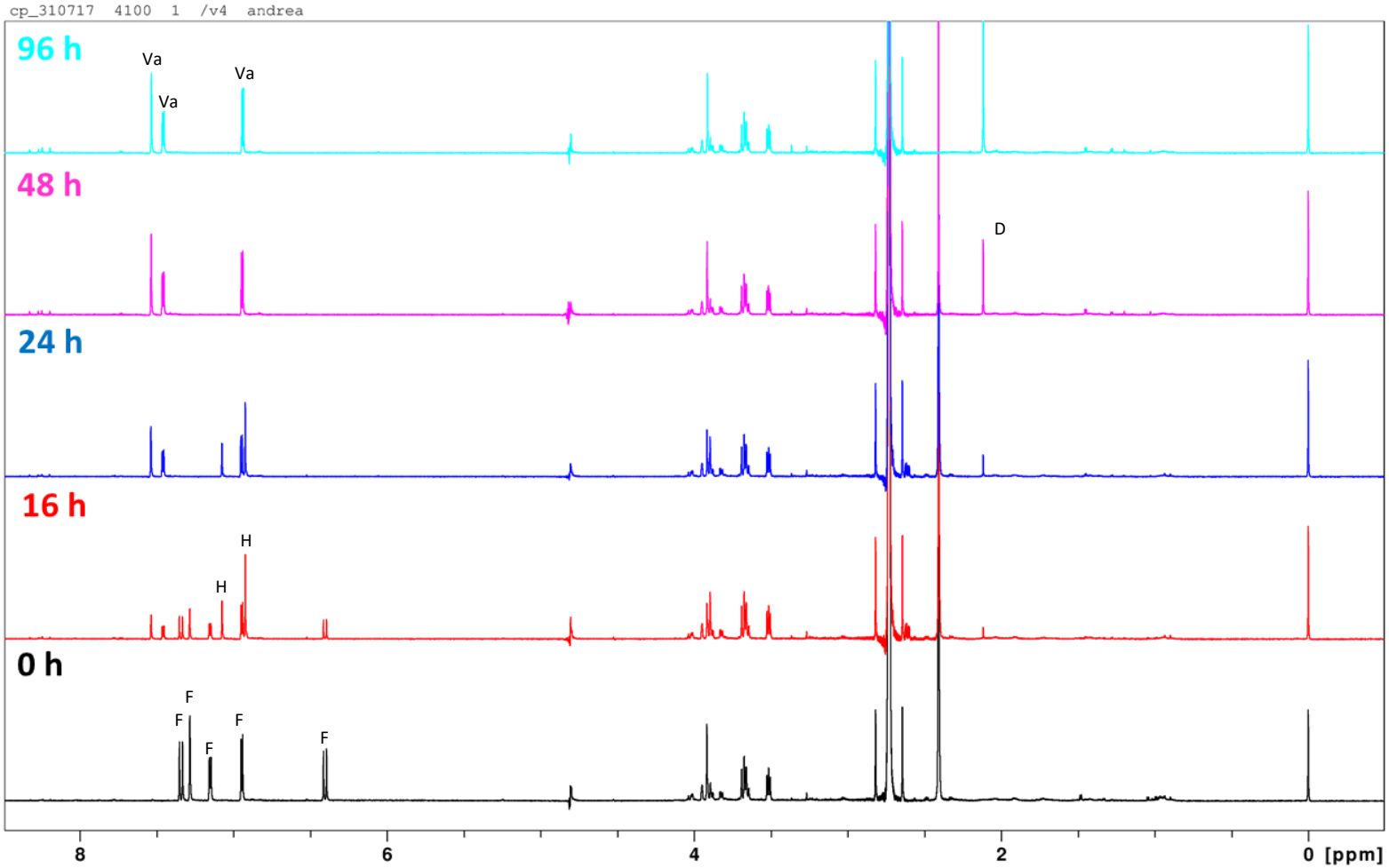


Figure 31: ^{1}H-NMR spectra of WT *R. palustris* supernatant samples. The ferulate (F) peaks decrease by 16 h and are gone by 24 h. Peaks at 6.9 and 7.1 ppm represent intermediary produced HMPHP (H) , and the unique peaks at ~7.5 and 7.6 ppm represent vanillic acid (Va). The peaks in the sugar region at 3.5 and 3.7 ppm show arabinose levels unchanging. The peak at ~2.1 ppm indicates probable accumulation of a degradation product (D) of succinate metabolism. Spectra were normalised to the TSP peak at 0 ppm, with H_2O suppression applied.

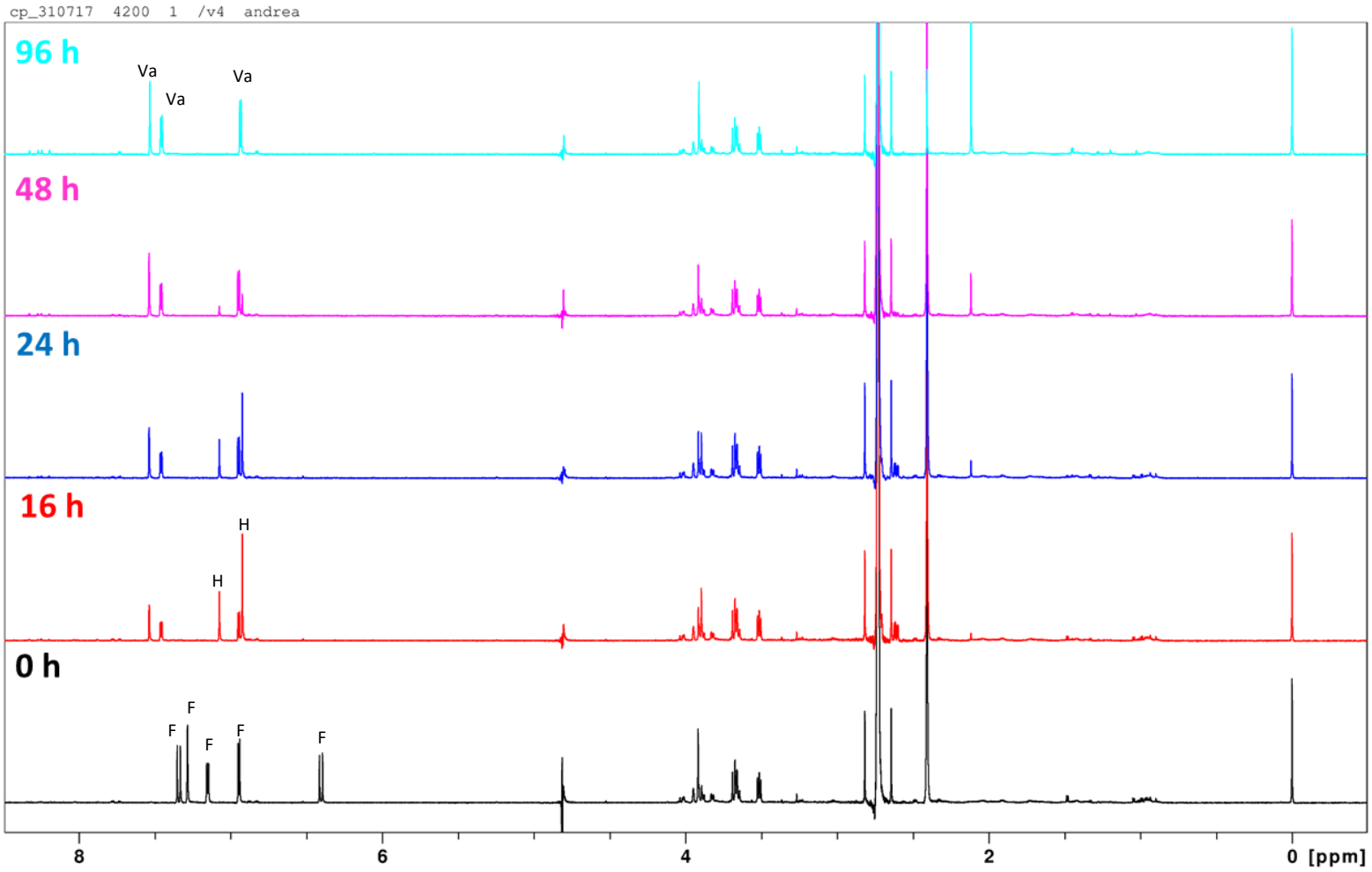


Figure 32: ^{1}H-NMR spectra of WT *R. palustris* (pBBRBB:*couAB*) supernatant samples. Plasmid-based overexpression of *couAB* resulted in a faster decrease of the ferulate (F) peaks, with no unique ferulate peaks visible in the 16 h sample (unlike in Figure 31). Spectra were normalised to the TSP peak at 0 ppm, with H_2O suppression applied. (H) = HMPHP; (Va) = vanillic acid.

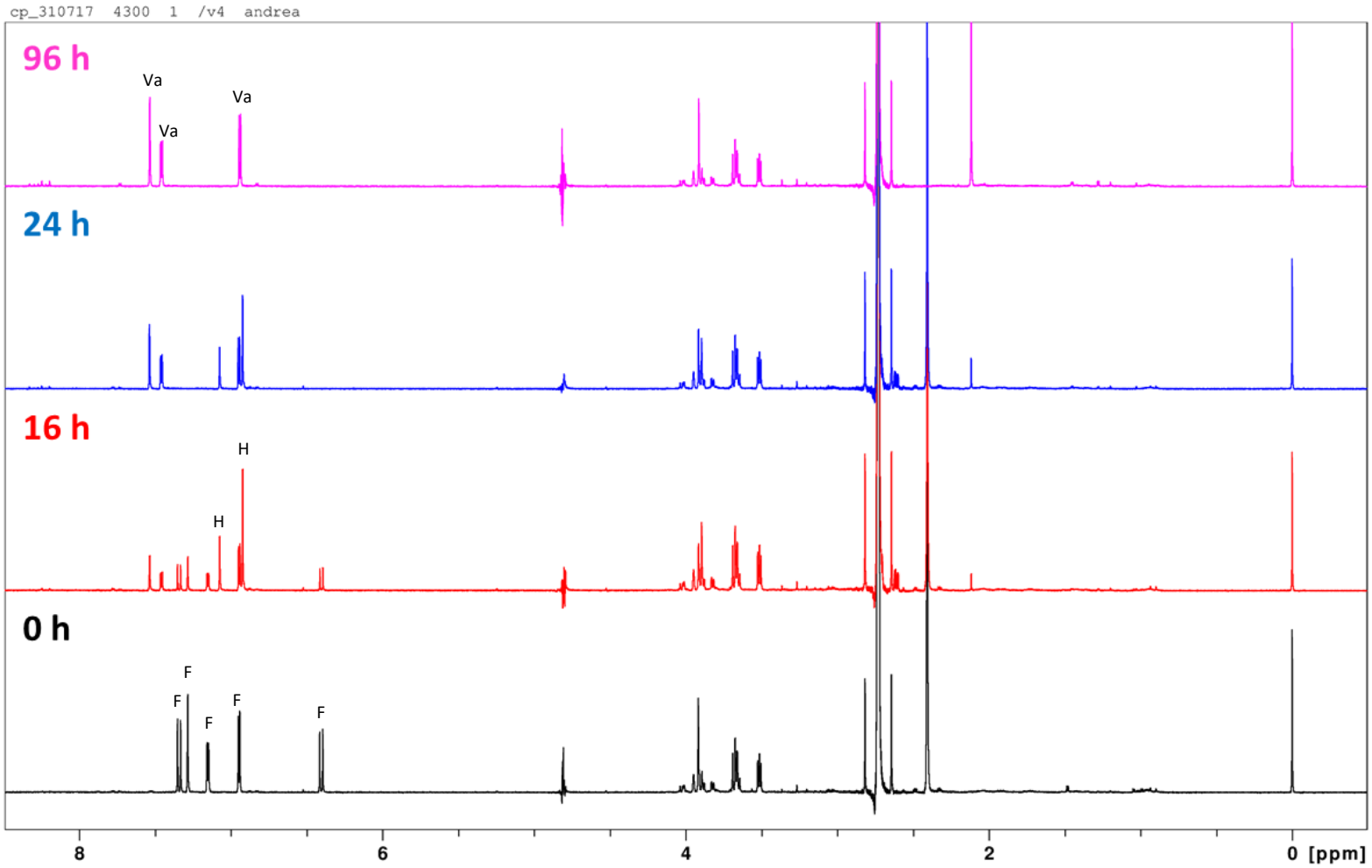


Figure 33: ^{1}H-NMR spectra of *R. palustris Δ1206* supernatant samples. The spectra shown are almost identical to those in Figure 31, implying no effect of the *rpa1206* knockout. Spectra were normalised to the TSP peak at 0 ppm, with H_2O suppression applied. (F) = ferulate; (H) = HMPHP; (Va) = vanillic acid.

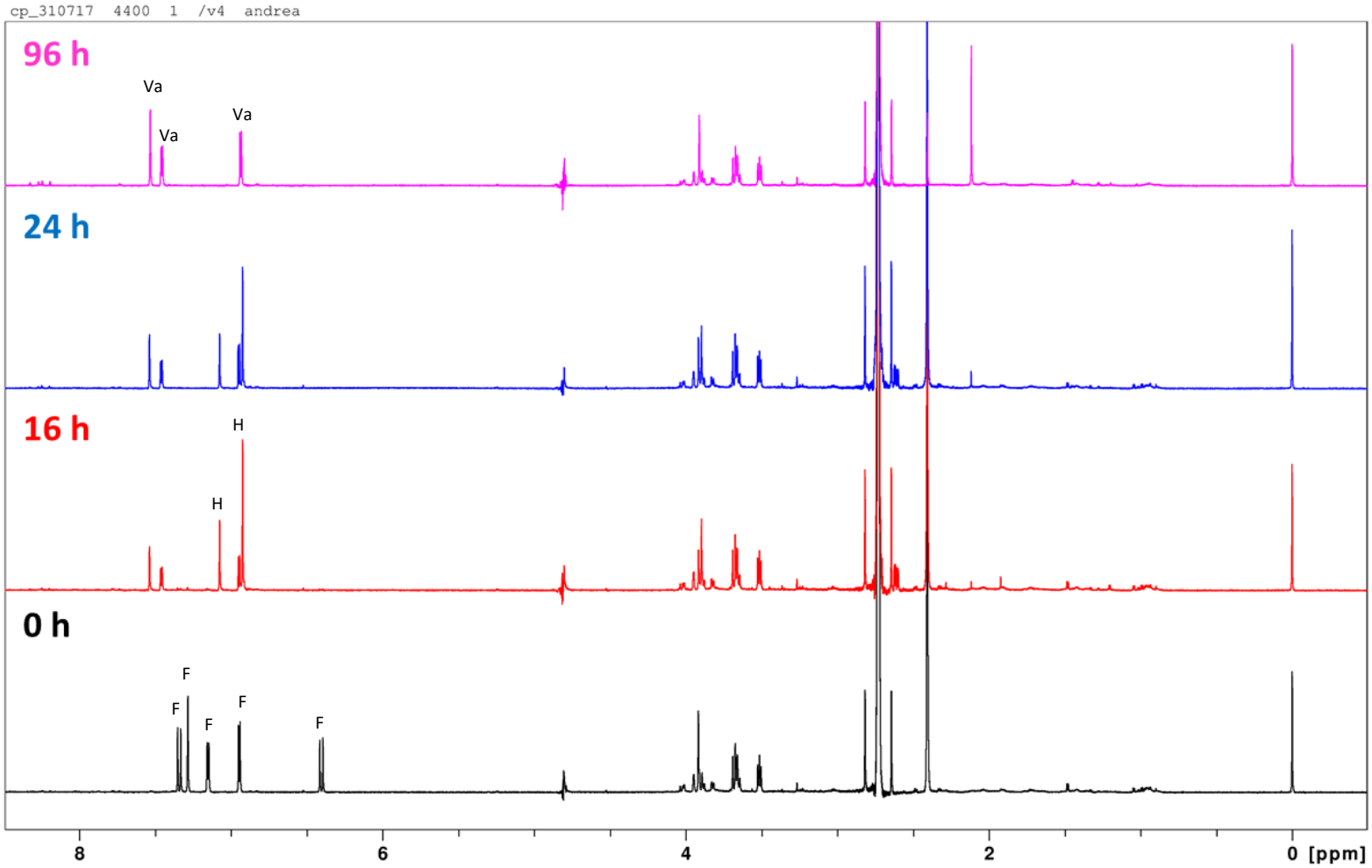


Figure 34: ^{1}H-NMR spectra of *R. palustris Δ1206* (pBBRBB:*couAB*) supernatant samples. The spectra shown are almost identical to those in Figure 32, implying no effect of the *rpa1206* knockout. Spectra were normalised to the TSP peak at 0 ppm, with H_2O suppression applied. (F) = ferulate; (H) = HMPHP; (Va) = vanillic acid.

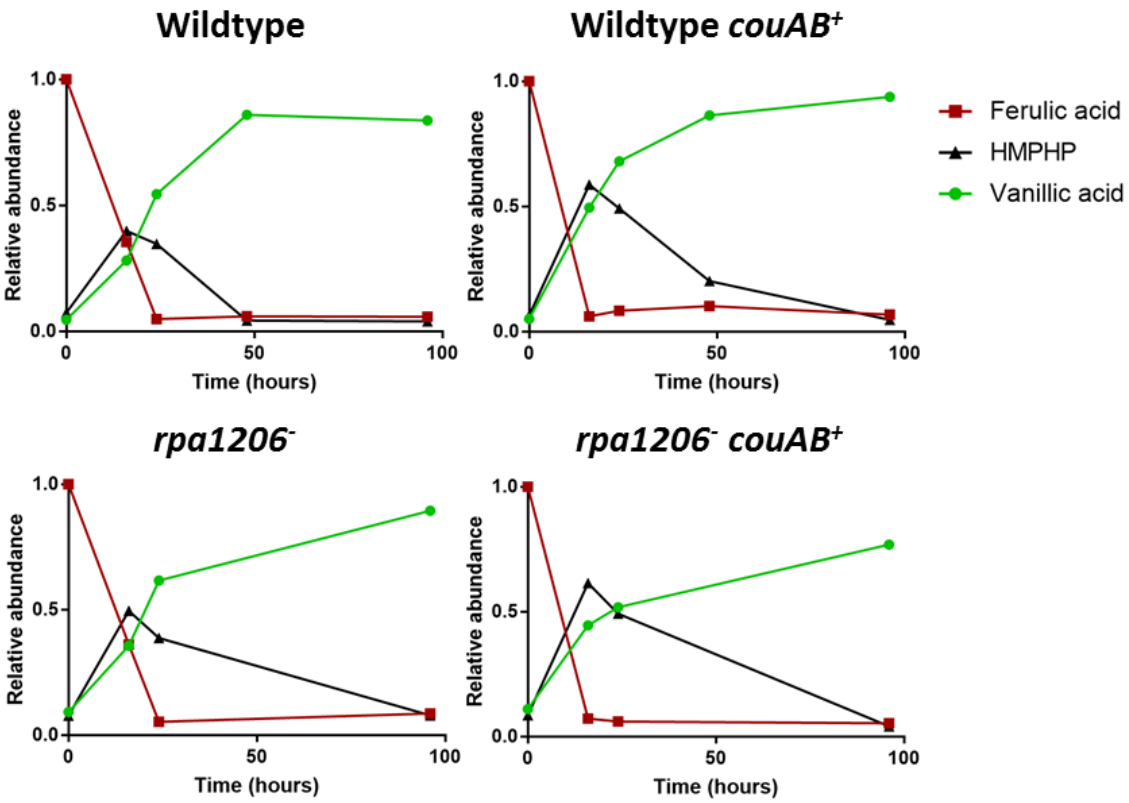


Figure 35: Changes in relative abundance of compounds from supernatants of *R. palustris* cultures. The relative abundance of ferulate, HMPHP and vanillic acid were calculated by integration of each compounds specific peaks: 7.497 and 7.573 ppm for vanillic acid; 7.185, 7.309 and 7.38 ppm for ferulate; and 7.105 ppm for HMPHP. Values at each time point are scaled relative to the TSP peak at 0 ppm. Relative abundances shown here are scaled by the integral of the ferulate peak at t = 0.

3.2.12. NMR analysis assessing E. coli's processing of exogenous vanillin

In order to assess *E. coli*'s ability to transform vanillin, a wildtype *E. coli* culture was grown in the presence of exogenously added vanillin. As before, samples of the culture were taken at time points over 24 h and the supernatants processed for ^{1}H-NMR analysis. The results are displayed in Figure 36 and show that *E. coli* must uptake extracellular vanillin, as with ferulate, and convert it into another compound identified as the less toxic vanillyl alcohol. Comparing this to the data shown in Figure 28 reinforces the conclusion that no vanillin was formed from ferulate by the action of CouA and CouB, as no vanillyl alcohol was detected as accumulating in the biotransformation experiments.

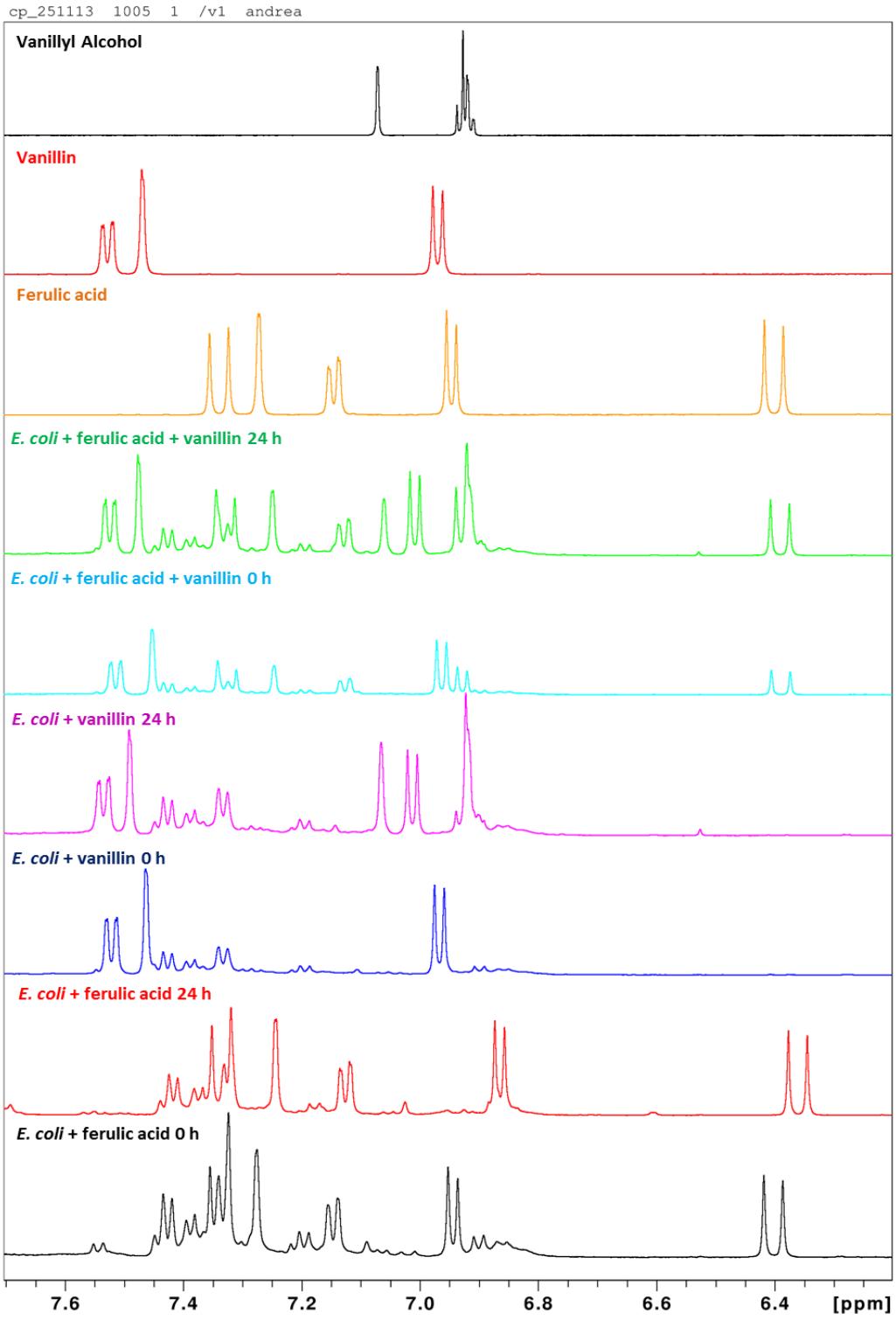


Figure 36: ^{1}H-NMR spectra showing *E. coli*'s detoxification of vanillin. Wildtype *E. coli* cells were grown in the presence of ferulate and/or vanillin, with the spectra here showing no change in the ferulate peaks while the presence of vanillin resulted in peaks representing vanillyl alcohol forming. Reference spectra for ferulate, vanillin and vanillyl alcohol are shown for comparison.

3.3. Discussion

The initial aim of the work of this chapter was to assess the utilisation of *R. palustris* genes expressed in *E. coli* for the biotransformation of ferulate into vanillin with the further aim of facilitating this process through increasing the uptake of the substrate. In addition the use of *R. palustris* as a microbial whole-cell cell catalyst for the process was also to be investigated. Initial efforts to overproduce the recombinant enzymes produced mixed results, while implementing overproduction of the transport systems proved highly problematic. Multiple other strategies could have been pursued to improve recombinant protein production such as using codon optimised genes, integration of the genes into the *E. coli* chromosome for more stable expression, or use of alternative *E. coli* strains and/or plasmids. Other areas of research that were not investigated based on these results include using the recombinant *E. coli* strain in upscaled batch and continuous cultures to produce vanillin with a view to optimising growth conditions to improve yield; and engineering the *E. coli* biocatalyst strain further, such as overexpression of citrate synthase and influencing the tricarboxylic acid (TCA) cycle to improve intracellular CoA levels. While planned, these approaches were not pursued as the results of this chapter rendered them redundant.

Ultimately, despite the issues with recombinant production levels of the CouA and CouB enzymes in *E. coli* experienced, the NMR analysis of the biotransformation experiment indicates that both were actually produced in an active form. Comparing the NMR data of the *E. coli* and *E. coli* pCOLADuet:*couAcouB* experiments highlights that *E. coli* lacks a degradation pathway for hydroxycinnamates and is unable to metabolise ferulate, with recombinant production of CouA and CouB the only possible factor that results in ferulate being catabolised and HMPHP accumulating (Figure 30). Enzyme activity was imbued upon the *E. coli*, but the expected CoA-ligase and enoyl-CoA hydratase/lyase activity was clearly not exhibited. The lack of CouA lyase activity could have been explained in its recombinant production in *E. coli* as a result of incomplete folding or erroneous disulphide bond formation; however, this is ruled out by *R. palustris* also showing HMPHP accumulation.

The detected accumulation of HMPHP and vanillic acid, but no vanillin (Figure 31), in *R. palustris* matched the results shown by an investigation into the degradation of ferulate in the soil bacterium *Agrobacterium fabrum* (Campillo et al., 2014). The expectation was, based on the annotation of genes in the gene cluster shown to be essential for growth on ferulate, that *A. fabrum* employed a non-β-oxidation pathway for metabolism of ferulate via vanillin. NMR and MS were used to characterise the intermediate compounds of

ferulate metabolism, with HMPHP and vanillic acid identified. MS also confirmed the presence of another intermediate identified as 4-hydroxy-3-methoxyphenyl-β-ketopropionic acid (HMPKP), which is highly unstable and was not detected by NMR. Based on these data and their use of single-gene knock out mutants Campillo *et al* (2014) proposed a β-oxidation CoA-dependent deacetylation pathway for the degradation of ferulate, and identified the enzymes responsible for each step (Figure 37).

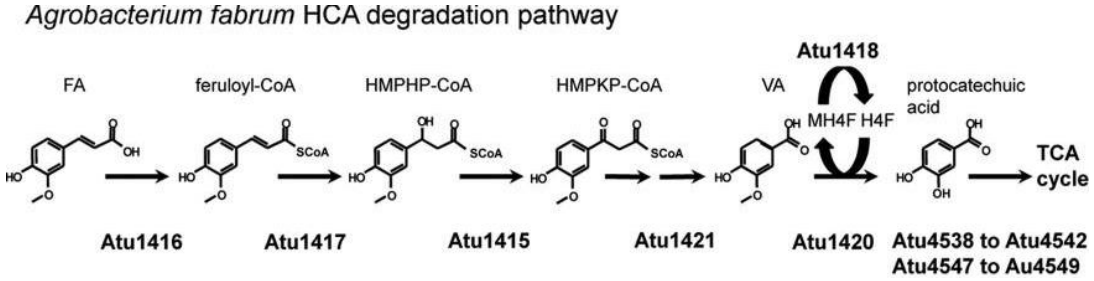


Figure 37: The β-oxidation CoA-dependent deacetylation pathway identified in *A. fabrum.* Based on their findings, Campillo *et al* (2014) proposed this pathway for the complete metabolism of ferulate (FA) for use in the tricarboxylic acid (TCA) cycle (Campillo et al., 2014). VA = vanillic acid; HMPHP = ; 4-hydroxy-3-methoxyphenyl-β-hydroxypropionic acid HMPKP = 4-hydroxy-3-methoxyphenyl-β-ketopropionic acid; MH4F = methylene-tetrahydrofolate; H4F = methyl-tetrahydrofolate.

The NMR data shown in this study align with the results of the Campillo et al (2014) paper to a large degree. The enzymes Atu1416 and Atu1417 show 45% and 49% sequence identity to CouB and CouA respectively. Combined with the data showing that overproduction of recombinant CouB and CouA in *E. coli* resulted in the bacterium metabolising ferulate and HMPHP accumulating, and that *R palustris* metabolising ferulate resulted in HMPHP and vanillic acid accumulating – the rate of which was increased significantly by the induced overproduction of CouB and CouA, this indicates that *R. palustris* metabolises ferulate through a similar pathway, with CouB and CouA catalysing the first two steps. Central to this is the observed function of CouA.

During the characterisation of CouA and CouB by Hirakawa et al (2012) the activity of CouA appeared to show conversion of coumaroyl-CoA to 4-hydroxy benzaldehyde, detected by analysis of the product by mass spectrometry (Hirakawa et al., 2012). The results of the NMR shown in this current study suggest that when feruloyl-CoA is the substrate, CouA only exhibits hydratase activity, with the lack of lyase activity resulting in the product being HMPHP-CoA rather than vanillin. Based on amino acid sequence, CouA is closely related to

Atu1417, with a very low BLAST expectation (E) value of $2x10^{-82}$. In comparison, other bacterial enoyl-CoA hydratase/lyases that have been proven to convert ferulate to vanillin show much less similarity (Figure 38). A structural analysis of the enoyl-CoA hydratase/lyase from *Pseuduomonas fluorescens* AN103 characterised the active site and identified its key residues including Glu143, Met70, Tyr75 and Tyr239 (Figure 38C) (Bennett et al., 2008). Site-directed mutagenesis revealed that Glu143 was essential for activity, and Tyr239 was involved in substrate specificity, with the two tyrosines and the methionine forming the part of the substrate binding pocket where the side groups of aromatic rings orientate. While CouA and Atu1417 do possess the conserved Glu143 residue, they both lack the methionine and tyrosine residues that the three enoyl-CoA hydratase/lyases shown to exhibit lyase activity with feruloyl-CoA have (Figure 38B). These substituted key residues could prevent lyase activity for substrates with the 4-methoxy side group, while not affecting activity with coumaroyl-CoA.

Ultimately, the results of the current study suggest that there are two subfamilies of enoyl-CoA hydratase/lyase based on their specificity for the second reaction step they catalyse, determined by residues forming their active sites (Figure 38). While both are involved in the metabolism of ferulate, only one subfamily can be harnessed for the production of vanillin.

While HMPHP-CoA is the product of this reaction, HMPHP was the compound detected by NMR in the supernatant. The presence of the CoA group would make HMPHP-CoA impossible to be exported out of the cytoplasm, while its removal to produce the propionic acid form could be spontaneous or catalysed by thioesterases, which both *E. coli* and *R. palustris* possess. In fact, RPA1788 is a thioesterase found in the CouR operon, and has been shown to be upregulated by coumaroyl-CoA via the action of CouR, so could be also be upregulated as a result the increasing feruloyl-CoA level (Phattarasukol et al., 2012).

The evidence that *R. palustris* metabolises ferulate through this alternative pathway extends to the other reaction steps. The initial CoA-ligation step can be performed by CouB as clearly shown by the *E. coli* NMR results, though the *R. palustris* genome harbours an extensive range of other CoA-ligases, including RPA1707 that is part of the *rpa1703-1710* β-oxidation pathway gene cluster, and is annotated as a feruloyl-CoA synthetase (Elder et al., 1992). The gene products of this operon could catalyse the pathway for degradation of

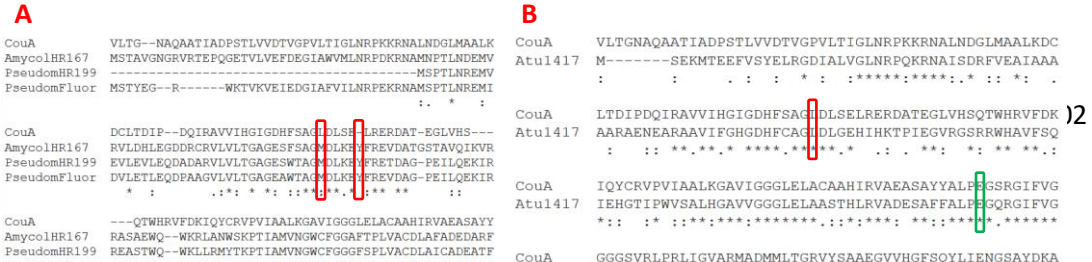


Figure 38: Comparison of CouA and alternative enoyl-CoA hydratase amino acid sequences. A) Sequence alignment of CouA and three enoyl-CoA hydratases (Ech) that have been shown to produce vanillin from feruloyl-CoA (Achterholt et al., 2000; Di Gioia et al., 2011). B) Sequence alignment of CouA and the enoyl-CoA Atu1417 from *Agrobacterium fabrum* that produces 4-hydroxy-3-methoxyphenyl-β-hydroxypropionic acid (HMPHP) from feruloyl-CoA. The coloured boxes highlight key residues of the active site shown by Bennett et al, 2008 (Bennett et al., 2008). C) The active site of the Ech from *Pseudomonas fluorescens* with vanillin and the cleaved acetyl-CoA (AcCoA), showing the location of key residues. D) The similarities in amino acid sequence of enoyl-CoA hydratases compared with CouA as shown by BLASTp searches.

hydroxycinnamatess to benzoyl-CoA, but the omics-based study of Pan et al 2008, indicated that the CouAB, non-β-oxidation pathway was the major metabolic pathway due to its significant upregulation during growth on *p*-coumarate (Pan et al., 2008).

The reaction subsequent to the formation of HMPHP-CoA produces a keto intermediate through dehydrogenation of the hydroxyl on the propionyl carbon chain. In *A. fabrum* this is catalysed by Atu1415, to which two *R. palustris* proteins show 55% sequence identity – RPA1705 and RPA2203. The former is located in the aforementioned β-oxidation operon, while the latter is a gene cluster that contains similarly annotated genes but has not been linked to metabolism of hydroxycinnamates. The product of this reaction is the intermediate HMPKP-CoA that was not detected as accumulating via NMR in this current study, probably due to its instability.

The *A. fabrum* study identified that the 4-hydroxyphenyl-β-ketoacyl-CoA hydrolase, Atu1421, catalyses the next step that resulted in the production of vanillic acid (Campillo et al., 2014). The author suggests that the product of the reaction is vanilloyl-CoA, which is spontaneously converted to vanillic acid. A BLAST search revealed that the closest homolog in *R. palustris* is RPA4198 with 73% identity and an E value of $5x10^{-156}$, which is annotated as an amidohydrolase. However, transcriptomics and proteomics showed that RPA4198 and its neighbour RPA4199 were upregulated during growth on coumarate, relative to succinate and benzoate (Pan et al., 2008). Furthermore, expression of *rpa4198* and *rpa4199* was shown to be strongly repressed by CouR and thus regulated via the same mechanism as CouA and CouB; they are the only genes upregulated by coumaroyl-CoA as part of the CouR regulon that aren't part of the *rpa1782-1793* gene cluster but were not discussed further, probably due to their annotation (Phattarasukol et al., 2012). Therefore it is proposed here that RPA4198 is a 4-hydroxyphenyl-β-ketoacyl-CoA thiolase that catalyses the conversion of HMPKP-CoA to vanillic acid, via vanilloyl-CoA. This step produces an acetyl-CoA molecule that allows *R. palustris* to utilise ferulate as a carbon source. RPA1703 is annotated as a thiolase and is the gene product of a gene in the β-oxidation operon and therefore could catalyse this reaction, but showed no homology to Atu1421. As before, the removal of the CoA could be catalysed by a thioesterase, such as the CouR-regulated RPA1788.

RPA4199 is annotated as a halide hydrolase (EC 3.8.1.2), and is predicted to be involved in the chloroalkane/chloroalkene degradation pathway based on this. Blast searching using the amino acid sequence of RPA4199 identified no homologs in *A. fabrum*. If its function is

as a halide hydrolase this could suggest that *R. palustris* could use part of this phenylpropanoid pathway to degrade chlorinated aromatics. Ether hydrolases, such as isochorimatase, are capable of demethylation of methoxy groups. If RPA4199 exhibited ether hydrolase activity it could demethylate the 4-methoxy side group of vanillic acid to produce protocatechuic acid; however, as this compound was not detected by NMR this activity is unlikely.

The Campillo et al (2014) study showed that interrupting the ferulate degradation pathway in *A. fabrum* after vanillic acid (ΔAtu1420) reduced growth on ferulate but did not prevent it; highlighting that the one molecule of acetyl-CoA produced up to this point can feed into the TCA cycle, but conversion into protocatechuic acid is required for complete metabolism (Campillo et al., 2014). This *A. fabrum* mutant did not have a phenotype when coumarate was used as a carbon source as a 4-hydroxybenzoate 3-monooxygenase (Atu4544) can perform this step; *R. palustris* possesses an enzyme that has amino acid sequence 68% identity to Atu4544: RPA1781 (PhbH). RPA1781 is encoded next to the the CouR operon, and was shown to be upregulated during growth on coumarate (Pan et al., 2008). It is likely that RPA1781 is used in aerobic metabolism of aromatics. Aerobically *R. palustris* can metabolise ferulate fully, via vanillic acid and the protocatechuic acid pathway (Figure 9), but anaerobically the VanAB system essential to this cannot demethylate vanillic acid and this product accumulates as shown by the NMR data here (Figure 31).

Ultimately, this suggests that ferulate is metabolised through a pathway distinct to that of the previously proposed non-β-oxidation pathway for coumarate metabolism. This proposed alternative pathway is shown in Figure 39, and offers alterations to the existing model of the two-branched peripheral pathway for the metabolism of lignin-derived hydroxycinnamate monomers in *R. palustris* shown in Figure 15. Central to this proposal is the bifunctional role of CouA, and anticipated function of RPA4198. This theory proposes that the initial, peripheral portion of the pathway for metabolism of lignin-derived phenylpropanoids in *R. palustris* is the same under aerobic and anaerobic conditions.

Figure 39 shows that *R. palustris* is capable of fully metabolising coumarate or ferulate aerobically, but anaerobically will only produce one acetyl-CoA molecule per molecule of ferulate, while vanillic acid will accumulate, as shown by the NMR data. This conversion of ferulate to vanillic acid does not go via vanillin as an intermediate in *R. palustris* as expected, and based on these data neither *R. palustris* nor recombinant production of its enzymes CouA and CouB can be utilised for the biotransformation of vanillin directly.

Vanillic acid does have commercial value as a flavouring agent and as a pharmaceutical precursor. Furthermore, it can be used as a precursor for vanillin production; aldehyde oxidoreductase enzymes such as the carboxylic acid reductase from the Gram-positive bacterial genus *Nocardia* have been shown to reduce vanillic acid. Recombinant production of such enzymes in *R. palustris* could feasibly enable it to be used as a microbial biocatalyst for vanillin production.

The role of RPA1206 was briefly investigated during this chapter. It was identified as the only enzyme with this activity in *R. palustris* that was more abundant at a proteomic level when the bacterium was grown on *p*-coumarate and was predicted to be the aldehyde dehydrogenase for the non-β-oxidation pathway, and thus capable of oxidising vanillin to vanillic acid (Pan et al., 2008). There is currently no published characterisation of RPA1206 with 4-hydroxybenzaldehyde to prove its role in coumarate metabolism, nor with vanillin. Regardless, the findings of this study largely render the knockout of the *rpa1206* gene redundant due to the lack of vanillin being produced. Therefore, the similarity of NMR data between the wildtype *R. palustris* and the *Δ1206* mutant is unsurprising. While it was originally thought that *E. coli*'s lack of an aromatic metabolism pathway would aid vanillin accumulation, the amino acid sequence of RPA1206 shows strong homology with the well-characterised, broad-substrate *E. coli* aldehyde dehydrogenase, AldB (Figure 40). However, Figure 36 shows that *E. coli* detoxifies vanillin by reduction to vanillyl alcohol rather than oxidation to vanillic acid, diminishing AldB's potential role. Within the *R. palustris* genome the protein that showed the most similarity was HpcC (42% identity), involved in catabolism of aromatic compounds post-ring cleavage. Based on this it is hard to predict whether RPA1206 and HpcC's acitvities may overlap as another example of redundancy in the metabolism of *R. palustris.*

The evidence for this pathway is mostly circumstantial; there are several key follow-up experiments that should be conducted to investigate this further. *In vitro* experiments with purified CouAB should be conducted with ferulate, and the final product analysed by NMR to confirm HMPHP is produced rather than vanillin. The aldehyde dehydrogenase RPA1206 should also be purified and its activity with the aldehyde derivatives of the hydroxycinnamates, e.g. vanillin, 4-hydroxybenzaldehyde, determined. Single knock-out mutants in RPA1705, RPA4198 and RPA1703, and a double knock-out mutant in RPA4198 and RPA1703 should be created. A ΔRPA1705 strain should be unable to grow on ferulate, or show an impaired-growth phenotype if there is an alternative phenylhydroxypropionyl-CoA dehydrogenase. A transcriptomics/proteomics-based approach to determine *R.*

palustris's response to ferulate would provide evidence on the enzymes and pathways involved, and would be useful for comparing with the equivalent coumarate data published. A proteomics study conducted in this lab by Abrar Akbar identified that CouA, CouB, RPA4198, RPA4199, RPA1706 were all more abundant in *R. palustris* grown on an alternative phenylpropanoid, 5-phenylvalerate, compared to growth on benzoate. This suggests the proposed pathways shown in Figure 39 may be key to the peripheral metabolism of other phenylpropanoids distinct from hydroxycinnamates, and further work should not be restrictive in the substrates investigated.

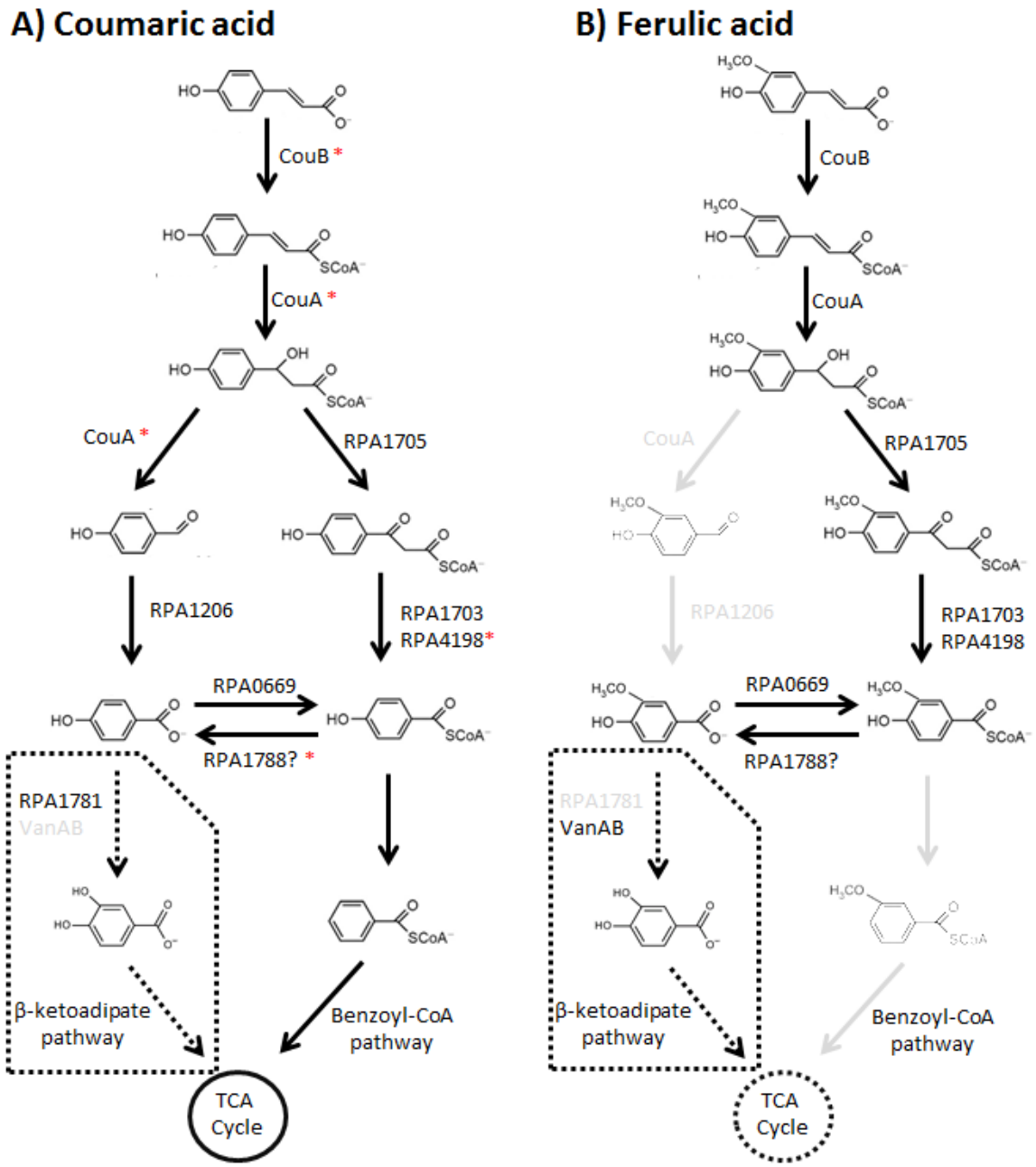


Figure 39: Proposed updated peripheral pathway for metabolism of lignin-derived hydroxy-cinnamates in *R. palustris*. A) The proposed means by which *R. palustris* metabolised coumarate, with the pre-existing paradigm of CouA's dual-activity allowing progression through the non-β-oxidation pathway. Red asterisks highlight proteins for which transcription of their genes has been shown to be controlled by CouR. B) The proposed pathway for metabolism of ferulate due to the lack of HMPHP-CoA-lyase activity exhibited by CouA. The faded steps cannot be followed for ferulate. The dotted lines depict pathways that can only occur under aerobic conditions.

Figure 40: Multiple sequence alignment of aldehyde dehydrogenases. BLASTing of the amino acid sequence of RPA1206 exclusively against *E. coli* and *R. palustris* revealed, respectively, that RPA1206 shows 70% sequence identity with AldB in *E. coli*, while the closest similarity shown by another *R. palustris* aldehyde dehydrogenase is the 5-carboxy-2-hydroxymuconate semialdehyde dehydrogenase involved in the catabolism of 4-hydroxyphenylacetate.

4. Comprehensive analysis of vanillin toxicity in *E. coli* identifies means of engineering a robust microbial platform for biotransformation.

Calum Pattrick, Mark Collins, Roy Chaudhuri, David J Kelly.

Preface

The following chapter has been prepared in manuscript format and is set for imminent submission. The vast majority of the experimental design, work and writing of the manusctipt has been carried out by myself; Mark Collins performed the mass spectrometry and proteomic data analysis and has written the methods for these sections, Roy Chaudhuri performed the genome assembly and analysis; Dave Kelly was involved in the planning and supervision of the work, and editing of the manuscript. The proposed author list reflects these contributions and is unlikely to change significantly.

4.1. Abstract

Vanillin is a high value compound due to its use as a flavouring compound and its production by biotransformation using bacteria is well-established. One of the key issues affecting production of vanillin, and other biotechnologically-relevant aromatic aldehydes, is the toxicity of the product, the mechanism of which is relatively poorly understood. In this study, the effect of vanillin treatment on the physiology of *E. coli* is elucidated, by assessing the global proteomic response supported by specific physiological experiments. The results reveal that vanillin elicits a significant oxidative stress response and related perturbations in metal homeostasis and central metabolism. Additionally, the role of, and interactions with, copper were revealed to be key to vanillin's toxicity. *E. coli* attempts to adapt to vanillin stress through upregulation of aldehyde detoxification enzymes, while these data also identify AcrD and AaeAB as potential vanillin efflux systems that could hold great potential in simultaneously increasing tolerance and the yield of intracellularly produced vanillin. Furthermore, four vanillin-tolerant strains of *E. coli* have been isolated through adaptive laboratory evolution, with each strain having a distinct non-synonymous SNP in *gltA.* Other detected mutations in *cpdA*, *rob* and the *marRAB* region also appear to contribute to conferred tolerance.

A range of gene targets for engineering a vanillin-tolerant strain of *E. coli* have been identified, including knock-out of *eamA*, *acrA,* and *acrB*; and overexpression of oxidative stress defences, copper-stress defences, NADPH-producing enzymes and efflux systems. Combined, this could greatly facilitate microbial production of vanillin and other aromatic and aldehyde compounds.

4.2. Introduction

Vanillin (4-hydoxy-3-methoxybenzaldehyde) is the principal organoleptic component of vanilla flavouring; a flavour highly sought after globally that is naturally sourced from the curing of the beans of the *Vanilla planifolia* plant. It's wide use in food, beverages, cosmetics and pharmaceuticals results in a huge demand, while the extensive labour required for harvesting and extraction make it the second most expensive spice in the world after saffron (Odoux and Grisoni, 2010; Walton et al., 2003). The consequence of this is that less than 1% is now sourced naturally, with the rest produced through chemical synthesis. With a greater consumer demand for 'natural flavours', and the pollution created from the chemical synthesis, there is a growing desire to produce vanillin by biotransformation from natural substrates (Ma and Daugulis, 2014). Therefore, the biotechnological exploitation of the microbial conversion of substrates such as eugenol, isoeugenol and ferulic acid into biovanillin is a key area of interest and commercial value (Kaur and Chakraborty, 2013).

Current approaches to biovanillin synthesis involve either manipulating organisms that have the inherent ability to metabolise vanillin, such as *Rhodococcus* sp., *Pseudomonas* sp., *Amycolatopsis* sp. and *Streptomyces* sp., or imbuing an alternative species with the capacity to produce it via recombinant expression of the enzymes required (Di Gioia et al., 2011; Ghosh et al., 2007; Plaggenborg et al., 2006). While the former have the advantage of already producing the necessary enzymes to convert the phenylpropanoid precursors, which originate in abundance in plant biomass, to vanillin, these enzymes are only steps in extended metabolic pathways that utilise aromatic compounds as carbon sources. Therefore, vanillin yield is limited by downstream conversion to further intermediates by enzymes such as vanillin dehydrogenases, with the toxicity of vanillin driving a need for this process to be rapid (Ding et al., 2015). The targeted knockout of such enzymes has been investigated, with mixed results (Fleige et al., 2013; Martínez-Cuesta et al., 2005). However, vanillin is not an intermediate in such metabolic pathways in a model organism such as *E. coli*, which is used as a platform for vanillin production through heterologous expression of the required enzymes (Barghini et al., 2007; Furuya et al., 2015). Nevertheless, *E. coli* still harbours endogenous enzymes capable of the conversion of aromatic aldehydes into alcohols such as aldo-keto reductases (DkgA, DkgB and YeaE) and aldehyde dehydrogenases (YqhdD, YahK and YjgB) (Koma et al., 2012; Pugh et al., 2015). Deletion of

genes encoding these enzymes has resulted in increased accumulation of aromatic aldehydes in *E. coli* (Kunjapur et al., 2014a).

One of the issues that arises with utilisation of any microbes in biotechnology are the toxic effects of substrates and products involved. Lignocellulosic feedstock used in many biotechnological processes contains a range of inhibitory compounds, such as furan derivatives, phenylpropanoids and weak acids; tolerance to which is much sought after (Forsberg et al., 2015; Jonsson et al., 2013). Toxicity of biofuels and solvents microbially produced in biotechnology, such as ethanol, butanol and toluene, arises from membrane damage and disruption of intracellular physiology; the potential benefit of reducing the levels of toxicity in organisms producing these products garners much interest and investment (Alsaker et al., 2010; Chong et al., 2013; Dunlop, 2011; Minty et al., 2011b; Sardessai and Bhosle, 2002; Zhang et al., 2015).

Toxicity of aldehydes is an issue commonly faced in biotechnology, with the strong, distinctive odours of aldehydes such as hexanal, octanal, nonanal, decanal, benzaldehyde, cinnamaldehyde – in addition to vanillin – utilised in the perfume and flavouring industries. Furthermore, furfural is an aldehyde found in lignin hydrolysates used extensively as feedstocks that inhibits growth of bacteria and yeast, and shown to be responsible for reducing microbial production of ethanol (Almeida et al 2009). The antimicrobial effect of aldehydes, such as acetaldehyde, formaldehyde and aromatic aldehydes, has been well documented (Esterbauer et al., 1991; Grimsrud et al., 2008; Singh and Khan, 1995). While the electrophilic nature of an aldehyde functional group is clearly a source of chaotropic behaviour, a comprehensive mechanism of aldehyde toxicity at a physiological level is yet to be established. Furthermore, direct DNA damage and a detrimental NADPH drain have been identified as effects of aldehydes on bacteria, while signs of oxidative stress were shown in yeast (Jarboe, 2011; Kunjapur et al., 2014a; Miller et al., 2009b; Nguyen et al., 2014; Singh and Khan, 1995).

A previous study into vanillin toxicity in bacteria identified it as having a bacteriostatic effect as a result of being membrane active (Fitzgerald et al., 2004). The hydrophobic nature of vanillin, as a phenolic compound, was highlighted as key to its detrimental interference with the integrity of the membrane in bacteria, and subsequent ion efflux. Aldehyde toxicity, including vanillin, in an ethanologenic *E. coli* strain was investigated and identified a correlation between the compounds' hydrophobicity and the degree of toxicity (Zaldivar et al., 1999). Plants accumulate phenolics such as vanillin, which are intermediates

in phenylpropanoid pathways, as a defence against phytopathogens (Maddox et al., 2010). Various plant-derived phenolics have been investigated for their antibacterial effect against *E. coli* and other foodborne pathogens (Cetin-Karaca and Newman, 2015; Rodriguez-Perez et al., 2016; Sanhueza et al., 2017).

A proteomics-based study analysed the response to growth on vanillin of *Pseudomonas putida* KT2240, an engineered biocatalyst with inherent vanillin-metabolising pathways (Simon et al., 2014). Unlike *E. coli*, *P. putida*'s broad metabolism encompasses a degradation pathway where vanillin is an intermediate (Graf and Altenbuchner, 2014). In addition to the expected changes in carbon metabolism shown, which are inapplicable to *E. coli*, their proteomics data also showed an increase in oxidative stress defences (catalases), chaperones (GroEL, GrpE, DnaK), low MW compound synthesis (trehalose, proline and betaine accumulation), and a wide range of efflux pumps that have been shown to be involved in solvent tolerance (Simon et al., 2014).

In this study we use quantitative label-free shotgun proteomics as a starting point for investigating the mechanism of toxicity of vanillin in *E. coli* BW25113 by assessing the bacterium's response to vanillin at the proteomic level. *E. coli* BW25113 is a well-characterised strain of *E. coli* from which the Keio collection of in-frame, single-gene knockouts was created (Baba et al., 2006), and has been engineered for production of chemicals previously (Marzan et al., 2017; Shen and Liao, 2008). The results of the proteomics are used to identify and investigate potential protein and pathway targets that could be manipulated to improve microbial vanillin tolerance. Furthermore, four strains have been irrationally evolved from BW25113 by repeated sub-culturing in the presence of increasing concentrations of vanillin. Their genomes have been sequenced to provide further insight into the physiological basis of vanillin toxicity and tolerance.

4.3. Methods

4.3.1. Bacterial strains and growth

E. coli BW25113, a derivative of the F^-, λ^-, *E. coli* K-12 strain BD792, was used as the standard wildtype strain. Single-gene knockout strains were obtained from the Keio collection (Baba et al., 2006). *E. coli* cultures were typically grown aerobically in LB medium at 37 °C, with flasks shaken at 200–250 rpm. For minimal media growth *E. coli* was grown in M9 medium supplemented with 1.8 µM thiamine and 0.2 % (W/v) carbon source. Antibiotics were used at the following concentrations: carbenicillin, 100 µg/ml; kanamycin, 50 µg/ml. Vanillin sensitivity was assessed by comparing growth with and without 10 mM vanillin in 50 ml cultures.

4.3.2. Peptide sample preparation for mass spectrometry

E. coli BW25113 cultures were grown overnight in 50 ml LB medium at 37 °C and used to inoculate 1 l flasks of LB with starting optical density (OD_{600}) 0.1. Cultures were grown either in the absence of vanillin or in the presence of 10 mM vanillin to an OD_{600} 0.5 and harvested by centrifugation at 10,000g for 10 min at 4 °C. Cell pellets were washed twice with 20 mM sodium phosphate buffer (pH 7.4) and stored at -80 °C until use. A cell-free extract was prepared by sonication and the protein concentration measured by Bradford assay. A total of 50 µg of protein was added to each well of a 12.5% SDS-PAGE gel and run at 160 V for 15 minutes. The gel was stained with Colloidal Coomassie (Invitrogen) for 1 h, de-stained with 25% methanol for 1 h and each lane was divided into 6 gel fractions and cut into 2 mm cubes. Gel slices were reduced with 10 mM DTT at 56 °C for 1 h and alkylated with 55 mM iodoacetamide in the dark, at room temperature for 30 min. Proteins were digested with 120ng Pierce™ trypsin protease for 1 hour at 37°C with gentle agitation followed by overnight incubation at 25°C. Peptides were extracted from gel fractions by treatment with 100% acetonitrile followed by treatment with 0.5% formic acid after which the supernatant was transferred to a peptide collection tube. This process was repeated three times before a final treatment with 100% acetonitrile. Finally, the samples were dried using a SpeedVac. For mass spectrometric analysis samples were re-suspended in 0.5% formic acid.

4.3.3. Mass spectrometry analysis

Extracted peptides were analysed by nano-liquid chromatography tandem mass spectrometry (LC-MS/MS) on a LTQ Orbitrap Elite (Thermo Fisher) hybrid mass

spectrometer equipped with a nanospray source, coupled with an Ultimate RSLCnano LC System (Dionex). The system was controlled by Xcalibur 2.1 (Thermo Fisher) and DCMSLink 2.08 (Dionex). Peptides were desalted on-line using a micro-Precolumn cartridge (C18 Pepmap 100, LC Packings) and then separated using a 60 min reversed phase gradient (4-32% acetonitrile/0.1% formic acid) on an EASY-Spray column, 15 cm x 50 µm ID, PepMap C18, 2 µm particles, 100 Å pore size (Thermo). The LTQ-Orbitrap Elite was operated with a cycle of one MS (in the Orbitrap) acquired at a resolution of 60,000 at m/z 400, with the top 20 most abundant multiply-charged (2^+ and higher) ions in a given chromatographic window subjected to MS/MS fragmentation in the linear ion trap. A Fourier-transform mass-spectrometry (FTMS) target value of 1E6 and an ion trap MSn target value of 1E4 was used and with the lock mass (445.120025) enabled. Maximum FTMS scan accumulation time of 500 ms and maximum ion trap MSn scan accumulation time of 100 ms were used. Dynamic exclusion was enabled with a repeat duration of 45 s with an exclusion list of 500 and exclusion duration of 30 s.

4.3.4. Proteomics data analysis

MS data was analysed data using MaxQuant version 1.5.2.8 (Cox and Mann, 2008). Data was searched against an *E. coli* UniProt sequence database using the following search parameters: trypsin with a maximum of 2 missed cleavages, 7 ppm for MS mass tolerance, 0.5 Da for MS/MS mass tolerance, with acetyl (Protein N-term) and oxidation (M) set as variable modifications and carbamidomethylation (C) set as a fixed modification. A protein FDR of 0.01 and a peptide FDR of 0.01 were used for identification level cut offs. Label free quantification was performed using MaxQuant calculated protein intensities with matching between runs (with a 2-minute retention time window) enabled (Cox et al., 2014). Data was filtered so that at least 3 valid LFQ intensity values were present in either group. Missing values were imputed using Perseus (1.4.1.3) and two sample t-testing was performed with a permutation based FDR calculation in Perseus (Tyanova et al., 2016).

4.3.5. RNA extraction and qRT-PCR

E. coli cultures were grown to OD_{600} 0.5, centrifuged at 10,000 g and pellets resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8) with 1 mg ml^{-1} lysozyme (Sigma Aldrich) added. Following incubation for 5 minutes, RNA extraction was performed using the SV Total RNA Isolation System kit (Promega). Purified RNA underwent a second DNase-treatment step using the Turbo DNA-free kit (Ambion) and stored at -80 °C until required. qRT-PCR were performed using the SensiFAST SYBR Lo-ROX One-step Kit (Bioline) in 96-well

optical reaction plates with each reaction performed in triplicate. Reactions used 20 ng of template RNA in a volume of 20 µl. A standard curve for each pair of primers was produced using *E.* coli genomic DNA, and gene expression was calculated relative to the reference gene, *rrsA.*

4.3.6. Detection of reactive oxygen species

2', 7'-dihydrodichlorofluorescein diacetate (H_2DCFDA) was used to detect endogenous ROS production in cell suspensions. *E. coli* cultures were grown to an OD_{600} of 1.0 in LB, then washed and resuspended in phosphate-buffered saline (pH 7.4) to a final OD_{600} of 0.5. Cell suspensions were incubated at 37 °C with 0.5 mM vanillin, vanillic acid or vanillyl alcohol, or 0.1% DMSO in the presence of 10 µM H_2DCFDA. Fluorescence was measured in a Cary Eclipse fluorimeter with excitation at 485 nm and emission at 538 nm. Cell-free and dye-free controls were assessed for artefact effects.

4.3.7. H_2DCFDA Fluorescence microscopy

E. coli BW25113 cells were washed and resuspended in PBS as before and incubated with 0.5 mM vanillin or 0.1% DMSO, and 10 µM H_2DCFDA for 80 min. Cells were briefly washed and resuspended once more before being dropped onto slides for visualisation using a Nikon Eclipse Ti-E with a PlanApo 60× NA = 1.40 oil-immersed objective. Images were produced using wide-field illumination and epi-fluorescence microscopy with illumination at 488 nm, with Nikon NIS-Elements software.

4.3.8. Inductively coupled plasma mass spectrometry

E. coli cultures were grown to OD_{600} 0.5 and centrifuged at max speed at 10 °C for 20 minutes. Pellets were resuspended in wash buffer (10 mM HEPES, 0.5 M sorbitol, 100 µM EDTA, pH 7.5) and centrifuged again, twice; and then finally once in wash buffer without EDTA. Washed pellets were resuspended in 1 mL 65% nitric acid and left for 48 h. The abundance of metal was determined by inductively coupled plasma mass spectrometry (ICP-MS).

4.3.9. DNA manipulation

Standard cloning and transformation methods were used. Genomic DNA was extracted using GenElute Bacterial Genome DNA Kit (Sigma Aldrich). PCR was performed using Phusion Flash High-Fidelity PCR Master Mix (ThermoFisher Scientific) with products purified

by QIAquick PCR purification kit (Qiagen). Restriction endonucleases, phosphatase and ligases were purchased from New England Biolabs.

4.3.10. Protein over-production and purification

Overexpression constructs were assembled for *yqhD*, *aldB*, and *yhbW* by cloning the amplicons derived from PCR with the primers shown S. Fig 13 into pET-21a, and these plasmids were then transformed in *E. coli* BL21(DE3). Overnight cultures of transformed *E. coli* were used to inoculate 1 l volumes of LB, grown to an OD_{600} of ~0.4 at 37 °C and expression was induced with 0.1 – 0.4 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Protein production typically lasted 3 h before cells were centrifuged (10,000 g at 4 °C, 10 min) and resuspended in buffer (20 mM sodium phosphate, 0.5 M NaCl, 40 mM imidazole, pH 7.4). Cell-free extracts were created by French press and subsequent centrifugation (20, 000 g at 4 °C, 20 min). Purification was achieved via nickel affinity chromatography using a HisTrap HP column (GE Healthcare). Concentration, desalting and buffer exchange (where necessary) was conducted using Vivaspin columns (GE Healthcare) of appropriate MW cut-off according to manufacturer's instructions. Purified protein concentrations were measured using the specific extinction coefficient calculated from the amino acid sequence and absorbance at 280 nm.

4.3.11. Fumarase activity assay

Total fumarase activity of cell-free extracts was determined by measuring the rate of conversion of fumarate to malate *in vitro*. Overnight cultures of *E. coli* were used to inoculate fresh LB to an OD_{600} of 0.05. Cells were grown until OD_{600} of 0.5, harvested by centrifugation at maximum speed and pellets resuspended in 50 mM Tris-HCl. Cell-free extracts were created by sonication on ice and subsequent centrifugation. 0.4 mM fumarate was added to 50 µl of cell-free extract diluted in 50 mM Tris-HCl to a total reaction volume of 1 ml and absorbance measured at 240 nm.

4.3.12. Transmission electron microscopy (TEM)

An overnight culture of *E. coli* BW25113 was diluted 100-fold and grown to an OD_{600} of 0.5 in LB media. Cells were then treated with vanillin at a final concentration of 5 mM and 10 mM, and incubated at 37 °C for 6 h. 1 ml samples were taken from the culture prior to treatment, after 1 h, 3 h and 6 h, and washed and resuspended in 20 mM pH 7.4 sodium phosphate buffer. 20 µl was placed onto a glow-discharged carbon-coated copper grid, stained with 0.75% uranyl formate and subsequently rinsed with dH_2O before air-drying.

TEM images were collected using a Philips CM100 microscope at 100kV equipped with a Gatan Ultrascan 667 CCD camera.

4.3.13. Whole-genome sequencing and analysis of vanillin-tolerant E. coli

E. coli BW25113 cells were initially grown in 10 mM vanillin overnight, before repeated sub-culturing into fresh media containing vanillin over a 12 week period; the vanillin concentration used was gradually increased over time up to concentrations in excess of 20 mM. Four independent lines of sub-culturing in vanillin were maintained, while *E. coli* cells were also sub-cultured in the absence of vanillin to provide a reference strain. Single colonies of each of these five strains were isolated, re-plated and samples sent for whole-genome sequencing conducted by MicrobesNG. Single nucleotide polymorphisms (SNPs) and small insertions/deletions (indels) were identified using the Snippy pipeline version 3.0 (Seemann, 2015). The trimmed reads obtained from MicrobesNG were mapped to the genome of E.coli strain BW25113 (GenBank accession number CP009273) using BWA mem version 0.7.12 (Li, 2013), and SNPs were called using FreeBayes version 1.1.0 (Garrison and Marth, 2012). The default Snippy settings were used, which exclude SNPs identified in regions with less than 10x sequence coverage and SNPs supported by less than 90% of the overlapping reads. Larger structural variants (SVs) such as insertions, deletions, inversions and other rearrangements were identified using LUMPY version 0.2.13 (Layer et al., 2014). SVs with fewer than 10 supporting pieces of evidence (split reads or discordant read pairs) were excluded from further analysis. The effects of the SNPs and SVs on any overlapping genes were determined using SnpEff (Cingolani et al., 2012). SNPs and SVs were displayed using Dalliance (Down et al., 2011).

4.3.14. Abiotic metal interaction assays

Metal interaction assays were performed based on methods described previously (Karlíčková et al., 2015) Copper reduction was assayed *in vitro* at 37 °C by addition of a compound to a reaction solution containing 0.25 mM $CuCl_2$ and 0.3 mM bathocuproinedisulfonic acid disodium salt in HEPES pH 7.4 buffer. Absorbance was measured at 484 nm. Controls were performed using DMSO as a baseline level, and with cuprous ions replacing $CuCl_2$. Iron reduction assays were performed similarly, with 0.25 mM $FeCl_3$ and 0.3 mM ferrozine; absorbance was measured at 562 nm.

4.3.15. Protein concentration determination

The concentration of a purified protein with defined amino acid sequence was determined by measurement of the absorbance at 280 nm of a solution for calculation according to the Beer-Lambert Law, with the extinction coefficient of the protein calculated from the sequence using the ExPASy ProtParam tool (http://web.expasy.org/protparam/). The total protein concentration of cell-free extracts was determined using the Bio-Rad Protein Assay Dye Reagent according to manufacturer's instructions.

4.4. Results

4.4.1. Effects of vanillin on global protein expression

Four independent cultures were grown in the absence or presence of a concentration of vanillin (10 mM) that slowed growth but was not completely inhibitory (typical growth curves are shown in S. Fig 3). Cell-free extracts were prepared from cells that had reached mid-exponential phase in each case and processed as described in Methods. A combination of 1-D SDS-PAGE and LC-MS/MS analysis was used in a label-free quantitative shotgun proteomics approach to determine changes in protein abundance in response to treatment with vanillin. A 1% false discovery rate (FDR) cut-off at the peptide and protein level was used for protein identification using MaxQuant resulting in the identification of 1,885 proteins, amounting to 46.0% coverage of the *E. coli* BW25113 proteome (S. Fig 1.) (Cox and Mann, 2008). Gene Ontology (GO) representation analysis performed using Panther (http://pantherdb.org/) showed that, while both cytoplasmic and membrane proteins were identified, the latter were under-represented with a fold enrichment of 0.76 ($p = 2.95x10^{-8}$) relative to the total *E. coli* proteome (Mi et al., 2013).

For label-free quantification (LFQ), a minimum of two unique peptides identified per protein were required. LFQ was performed using MaxQuant LFQ intensities, and statistical analysis was performed using Perseus as follows (Tyanova et al., 2016). The data set was filtered to remove proteins with less than three valid LFQ values in at least one group leading to a set of 850 quantified proteins for statistical analysis (S. Fig 2). LFQ intensities were log2-transformed and missing values were imputed using a downshifted normal distribution (width 0.3, downshift 1.8); comparison of LFQ intensities for each replicate can be found in S. Fig 4. In order to identify changes in protein expression, t-testing of the quantitative data was performed with correction for multiple hypothesis testing using a permutation-based FDR of 0.05 and an S0 = 1.

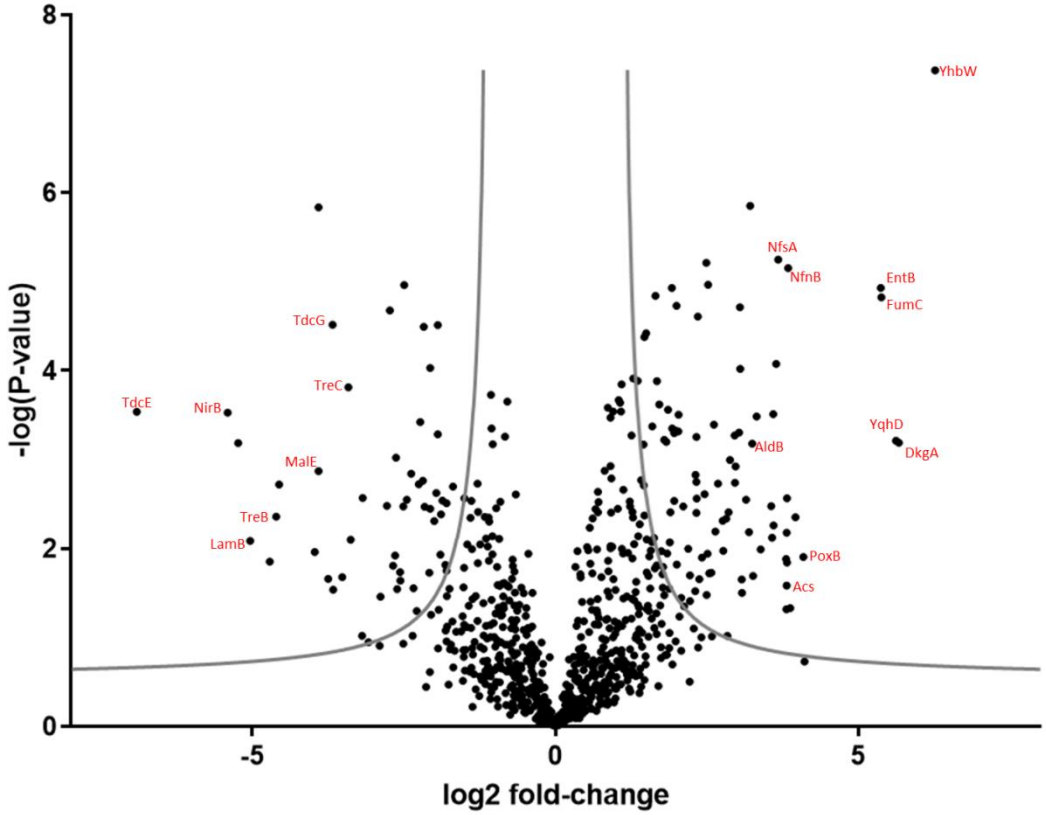


Figure 41: Volcano plot of quantifiable proteins identified by mass spectrometry. From the total dataset of proteins identified by mass-spectrometry, 850 proteins were significantly quantifiable and the result of t-testing of the quantitative data is shown here. Proteins above the grey line showed a significant fold-change in abundance on vanillin treatment.

In total, 147 proteins exhibited a significant change in abundance in response to vanillin; 93 proteins showed a ≥ 2.5-fold increase, and 54 showed a ≥ 2.5-fold decrease (Figure 41; proteins listed in Table 5 and Table 6). These sets of regulated proteins included both cytoplasmic and membrane proteins with a range of biological functions. Functional classification was performed according to the Cluster of Orthologous Groups (COG) database (NCBI) for both sets (Figure 42). This showed that a large proportion of the differentially regulated proteins were involved in amino acid and carbohydrate metabolism, and energy production; while the largest difference between upregulated and downregulated proteins groups in was in the “defence mechanisms” and “post-translational modification, protein turnover, chaperone functions” classifications, suggesting that vanillin is causing stress and eliciting a cellular adaptation response. Consistent with the slower growth seen in the presence of vanillin, there is also a pattern of down regulation of proteins involved in “translation, ribosomal structure and biogenesis”,

"replication, recombination and repair" and "cell wall/membrane/envelope biogenesis", indicating a reprogramming of metabolism away from growth and reproduction.

Table 5: Proteins detected by LC-MS/MS that show a significant increase in abundance in response to vanillin treatment. Fold-increase is calculated from the average abundance of four independent vanillin-treated replicates relative to that of the four independent untreated-replicates.

	Protein	Fold-increase	p-value
	Oxidative stress defence		
ahpC	Alkyl hydroperoxide reductase subunit C	**4.00**	1.86E-05
ahpF	Alkyl hydroperoxide reductase subunit F	**2.56**	1.30E-04
katE	Catalase HPII	**14.65**	4.65E-02
marR	Multiple antibiotic resistance protein	**7.25**	3.89E-03
osmC	Peroxiredoxin	**5.03**	2.74E-03
sodA	Superoxide dismutase [Mn]	**4.22**	3.58E-02
	Glutathione		
grxB	Glutaredoxin-2	**8.18**	4.94E-04
gshB	Glutathione synthetase	**14.15**	1.44E-02
gss	Glutathionylspermidine synthetase/amidase	**3.48**	5.99E-04
gstA	Glutathione S-transferase	**5.70**	1.14E-02
yghU	Disulfide-bond oxidoreductase	**7.38**	1.01E-03
yqjG	Glutathionyl-hydroquinone reductase	**3.89**	5.01E-04
	Low-MW compound accumulation		
gadC	Glutamate/γ-aminobutyrate antiporter	**6.78**	4.87E-03
gltB	Glutamate synthase [NADPH] large chain	**7.16**	9.58E-02
otsA	α,α-trehalose-phosphate synthase [UDP-forming]	**4.97**	1.48E-03
	Chaperones and proteases		
cbpA	Curved DNA-binding protein	**3.72**	3.89E-03
clpB	Chaperone protein	**2.83**	3.78E-05
dcp	Periplasmic serine endoprotease	**3.14**	1.44E-05
degP	Periplasmic serine endoprotease	**5.73**	1.08E-05
hchA	Molecular chaperone Hsp31 and glyoxalase 3	**14.14**	2.72E-03
hslO	33 kDa chaperonin	**8.42**	2.23E-02
pepB	Peptidase B	**5.09**	2.45E-05
	Metal homeostasis		
entB	Enterobactin synthase component B	**41.31**	1.17E-05
entE	Enterobactin synthase component E	**7.86**	1.19E-03
entF	Enterobactin synthase component F	**14.11**	6.64E-03
moaB	Molybdenum cofactor biosynthesis protein B	**9.28**	1.40E-06
sufB	FeS cluster assembly protein	**6.13**	4.04E-04
sufD	FeS cluster assembly protein	**6.24**	6.42E-03
sufS	Cysteine desulfurase	**6.43**	1.86E-03
	Aldehyde detoxification		
ahr	Aldehyde reductase	**7.80**	1.82E-03

	Protein	Fold-increase	p-value
aldB	Aldehyde dehydrogenase B	**9.49**	6.58E-04
dkgA	2,5-diketo-D-gluconic acid reductase A	**50.82**	6.46E-04
yahK	Aldehyde reductase	**4.67**	3.90E-02
yghA	Uncharacterized oxidoreductase	**5.95**	1.87E-02
yqhD	Alcohol dehydrogenase	**49.25**	6.11E-04
	Central carbon metabolism and energy production		
aceA	Isocitrate lyase	**8.46**	3.15E-02
aceB	Malate synthase A	**9.16**	6.52E-03
acnA	Aconitate hydratase 1	**11.93**	7.53E-03
acs	Acetyl-coenzyme A synthetase	**14.10**	2.60E-02
fbaB	Fructose-bisphosphate aldolase class 1	**8.86**	2.82E-03
fumC	Fumarate hydratase class II	**41.62**	1.49E-05
gltA	Citrate synthase	**4.01**	1.51E-02
mdh	Malate dehydrogenase	**4.09**	3.12E-04
mtlA	PTS system mannitol-specific EIICBA component	**3.95**	4.68E-04
pfkB	ATP-dependent 6-phosphofructokinase isozyme 2	**4.65**	2.00E-02
ppk	Polyphosphate kinase	**3.55**	6.31E-04
poxB	Pyruvate dehydrogenase [ubiquinone]	**17.04**	1.24E-02
sad	Succinate semialdehyde dehydrogenase	**10.48**	1.02E-02
talA	Transaldolase A	**6.82**	1.06E-02
tktB	Transketolase 2	**9.59**	2.03E-02
zwf	Glucose-6-phosphate 1-dehydrogenase	**3.80**	4.46E-04
	Nitro-compound detoxification		
nfsA	Oxygen-insensitive NADPH nitroreductase	**12.78**	5.64E-06
nfnB	Oxygen-insensitive NAD(P)H nitroreductase	**14.34**	7.03E-06
	Amino acid biosynthesis		
astC	Succinylornithine transaminase	**14.05**	4.80E-02
hisB	Histidine biosynthesis bifunctional protein	**7.06**	4.59E-03
hisC	Histidinol-phosphate aminotransferase	**5.19**	2.21E-02
hisD	Histidinol dehydrogenase	**3.89**	2.90E-03
hisG	ATP phosphoribosyltransferase	**8.27**	9.55E-05
ilvB	Acetolactate synthase isozyme 1 large subunit	**5.65**	3.32E-02
lysC	Lysine-sensitive aspartokinase 3	**3.03**	4.23E-04
pheA	P-protein	**3.59**	1.13E-02
trpB	Tryptophan synthase beta chain	**2.74**	6.73E-04
yecD	Isochorismatase family protein	**7.76**	5.35E-04
	Folate metabolism		
folE	GTP cyclohydrolase 1	**3.20**	1.31E-04
panB	3-methyl-2-oxobutanoate hydroxymethyltransferase	**3.70**	8.48E-03
ybgI	Putative GTP cyclohydrolase 1 type 2	**3.09**	7.58E-03
ygfZ	tRNA-modifying protein	**5.51**	2.44E-03
	β-D-glucuronide and D-glucuronate degradation		
uidA	Beta-glucuronidase	**5.02**	1.78E-03
uxaC	Uronate isomerase	**12.49**	8.36E-05
uxuA	Mannonate dehydratase	**4.57**	1.00E-02

	Protein	Fold-increase	p-value
	Uncharacterised		
ybjP	Uncharacterized lipoprotein	**4.36**	4.46E-02
ydhJ	Uncharacterized protein	**3.36**	1.06E-02
yeiR	Uncharacterized protein	**5.81**	1.90E-02
yhbW	Uncharacterized protein	**76.96**	4.18E-08
yjhC	Uncharacterized oxidoreductase	**15.57**	4.43E-03
yncE	Uncharacterized protein	**14.00**	1.31E-02
	Others		
add	Adenosine deaminase	**3.63**	2.75E-04
amn	AMP nucleosidase	**5.62**	6.10E-06
anmK	Anhydro-N-acetylmuramic acid kinase	**3.28**	2.39E-04
cfa	Cyclopropane-fatty-acyl-phospholipid synthase	**4.07**	4.81E-04
codA	Cytosine deaminase	**4.96**	2.99E-02
curA	NADPH-dependent curcumin reductase	**5.02**	3.97E-03
dcyD	D-cysteine desulfhydrase	**4.32**	3.35E-03
emrA	Multidrug export protein	**5.35**	1.26E-02
ghrA	Glyoxylate/hydroxypyruvate reductase A	**5.02**	5.56E-04
hdhA	7-alpha-hydroxysteroid dehydrogenase	**12.10**	3.08E-04
hemB	Delta-aminolevulinic acid dehydratase	**11.81**	3.32E-03
lysU	Lysine--tRNA ligase, heat inducible	**3.78**	1.18E-05
mdaB	Modulator of drug activity B	**8.25**	1.93E-05
metG	Methionine--tRNA ligase	**2.76**	4.19E-05
mnmA	tRNA-specific 2-thiouridylase	**3.43**	1.60E-02
torA	Trimethylamine-N-oxide reductase 1	**12.13**	5.48E-03
yadG	Uncharacterized ABC transporter ATP-binding protein	**10.00**	3.26E-04

Table 6: Proteins detected by LC-MS/MS that show a significant decrease in abundance in response to vanillin treatment. Fold-increase is calculated from the average abundance of four independent vanillin-treated replicates relative to that of the four independent untreated-replicates.

	Protein name	Fold-decrease	p-value
	Anaerobic response and FeS-cluster-containing proteins		
dmsA	Dimethyl sulfoxide reductase	**23.557**	1.90E-03
frdA	Fumarate reductase flavoprotein subunit	**4.192**	9.34E-05
frdB	Fumarate reductase iron-sulfur subunit	**6.650**	2.09E-05
fumB	Fumarate hydratase class I, anaerobic	**5.493**	2.79E-03
glpA	Anaerobic glycerol-3-phosphate dehydrogenase subunit A	**37.628**	6.51E-04
glpB	Anaerobic glycerol-3-phosphate dehydrogenase subunit B	**4.463**	3.39E-03
miaB	tRNA-2-methylthio-N(6)-dimethylallyladenosine synthase	**6.394**	1.56E-02
narG	Respiratory nitrate reductase 1 alpha chain	**9.130**	9.52E-02
narH	Respiratory nitrate reductase 1 beta chain	**15.661**	1.09E-02
nirB	Nitrite reductase (NADH) large subunit	**42.508**	2.96E-04
preT	NAD-dependent dihydropyrimidine dehydrogenase subunit	**15.003**	1.46E-06
sdaB	L-serine dehydratase 2	**8.473**	1.13E-01
tdcG	L-serine dehydratase	**12.773**	3.04E-05
	Glycyl radical proteins		
grcA	Autonomous glycyl radical cofactor	**4.772**	1.88E-03
tdcE	PFL-like enzyme	**119.882**	2.89E-04
	Maltose and trehalose		
lamB	Maltoporin	**32.814**	8.15E-03
malE	Maltose-binding periplasmic protein	**15.000**	1.34E-03
malP	Maltodextrin phosphorylase	**4.511**	3.21E-05
malQ	4-alpha-glucanotransferase	**5.201**	1.44E-03
treB	PTS system trehalose-specific EIIBC component	**24.355**	4.38E-03
treC	Trehalose-6-phosphate hydrolase	**10.666**	1.53E-04
	Porins		
fadL	Long-chain fatty acid transport protein	**10.380**	7.97E-03
ompF	Outer membrane protein F	**13.429**	2.18E-02
tsx	Nucleoside-specific channel-forming protein	**5.468**	2.82E-03
	Replication and cell division		
deaD	ATP-dependent RNA helicase	**26.241**	1.40E-02
mnmG	tRNA uridine 5-carboxymethylaminomethyl modification enzyme	**12.723**	2.89E-02
murA	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	**3.645**	2.89E-03
parE	DNA topoisomerase 4 subunit B	**11.465**	2.10E-02
rimO	Ribosomal protein S12 methylthiotransferase	**3.522**	1.52E-02
rplD	50S ribosomal protein L4	**3.469**	1.78E-02
rplP	50S ribosomal protein L16	**5.895**	2.28E-02
rplT	50S ribosomal protein L20	**5.067**	2.78E-02
rplU	50S ribosomal protein L21	**3.707**	4.12E-03
rplY	50S ribosomal protein L25	**7.385**	3.47E-02

	Protein name	Fold-decrease	p-value
rpmB	50S ribosomal protein L28	**4.230**	1.86E-02
ttcA	tRNA 2-thiocytidine biosynthesis protein	**6.103**	2.84E-02
	LPS biosynthesis		
rfbB	dTDP-glucose 4,6-dehydratase 1	**3.838**	3.06E-05
rfbD	dTDP-4-dehydrorhamnose reductase	**3.727**	1.17E-02
	Aspartate/asparagine catabolism		
ansB	L-asparaginase 2	**6.255**	1.19E-02
asnA	Aspartate--ammonia ligase	**6.872**	3.30E-03
aspA	Aspartate ammonia-lyase	**5.642**	1.09E-05
	Glyoxalate degrdation		
garD	D-galactarate dehydratase	**4.186**	3.55E-03
garR	2-hydroxy-3-oxopropionate reductase	**9.052**	2.70E-03
	Uncharacterised		
ydhQ	Uncharacterized protein YdhQ	**6.192**	9.51E-04
ygeV	Uncharacterized σ54-dependent transcriptional regulator	**3.229**	2.01E-03
yjjI	Uncharacterized protein YjjI	**5.695**	3.36E-03
	Others		
ftnA	Bacterial non-heme ferritin	**5.918**	1.83E-02
gatZ	D-tagatose-1,6-bisphosphate aldolase subunit	**3.910**	2.37E-03
guaC	GMP reductase	**4.548**	1.72E-03
hupA	DNA-binding protein HU-alpha	**3.991**	4.90E-03
hypB	Hydrogenase isoenzymes nickel incorporation protein	**4.878**	5.01E-02
nanA	N-acetylneuraminate lyase	**3.477**	3.08E-03
pepE	Peptidase E	**3.840**	5.16E-04
rpoS	RNA polymerase sigma factor	**4.694**	3.77E-04

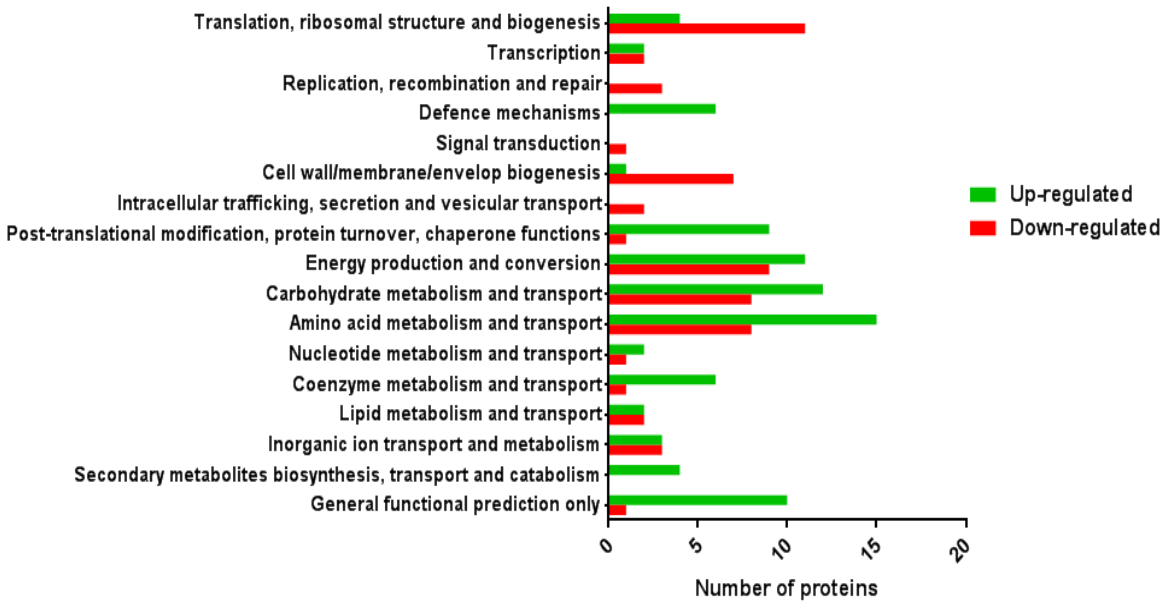


Figure 42: Functional classification of proteins that exhibited a significant change in abundance in response to vanillin treatment. Proteins were assigned a functional category according to the Cluster of Orthologous Groups (COGs) database for *E. coli*. Proteins that showed an increase in abundance in response to vanillin treatment are shown in green bars, and proteins that showed a decrease in red bars. Six proteins that were upregulated do not have an assigned COG: NfnB, HdhA YghA, YbjP and YbgI; and one protein that was downregulated is also unassigned: YjjL.

The proteomic analysis identified several groups of proteins related by membership of the same regulon or by being encoded in the same operon, with similar changes in abundance, suggesting this approach reliably reports the effects of vanillin on the proteome. A clear pattern was observed in the upregulated group of proteins involved in various stress responses, in particular the oxidative stress response. There were also changes in proteins mediating uptake and processing of metal ions, upregulation of enzymes capable of converting aldehydes to potentially less toxic compounds, and perturbation of carbon metabolism and amino acid biosynthesis. Therefore, four major areas of cellular physiology were identified as being most affected by vanillin: oxidative stress, metal homeostasis, direct detoxification and metabolism. These are examined in more detail below.

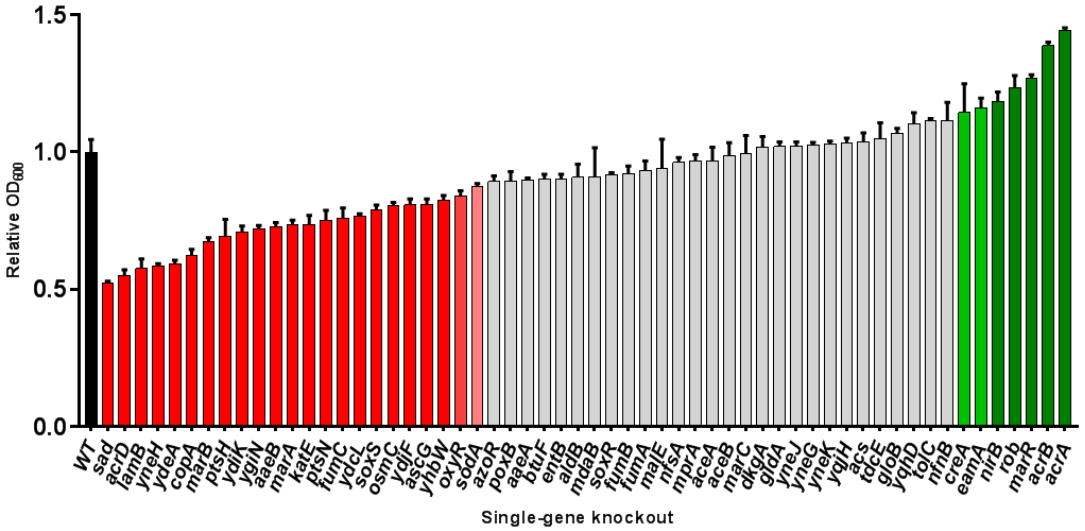


Figure 43: Susceptibility of single-gene knock-out Keio *E. coli* mutants to vanillin. A selection of 58 strains from the Keio collection of single knock-out *E. coli* strains were obtained to assess their growth phenotype in the presence of 10 mM vanillin relative to wildtype BW25113. This selection was based on genes of interest as a result of the proteomic data, and of the genome sequencing of the cultivated tolerance strains. Bars represent the mean OD_{600} relative to the WT OD_{600} of two cultures grown for 6 h, standard deviations are shown as error bars. Coloured bars represent a growth phenotype was detected for that strain; red bars show an increased sensitivity to vanillin and green bars an increased tolerance, with a paler colour indicating a less significance difference.

4.4.2. Vanillin-dependent oxidative stress

Growth with vanillin increased the expression of a large number of proteins involved in the oxidative stress response in *E. coli* (Table 5). Multiple proteins that are members of the superoxide-induced SoxRS regulon were identified as being significantly upregulated (Blanchard et al., 2007): YhbW (77.0-fold increase), FumC (41.6-fold), PoxB (17.0-fold), NfnB (14.3-fold), NfsA (12.8-fold), AcnA (11.9-fold), HemB (11.8-fold), MarR (7.2-fold), YgfZ (5.5-fold), and Zwf (3.8-fold). However, other SoxRS regulon proteins were identified that showed no significant change in abundance (GlnD, Map, AcrA, AcrB, YbaI, FldA, Cmk, PqiB, TolC and Pgi) but no members of the SoxRS regulon exhibited a decrease in abundance during growth with vanillin (S. Fig 2). RT-PCR experiments showed that transcription of *soxS* itself was upregulated by vanillin (Figure 44A), while a *soxS*-deletion strain showed increased sensitivity to vanillin (Figure 43).

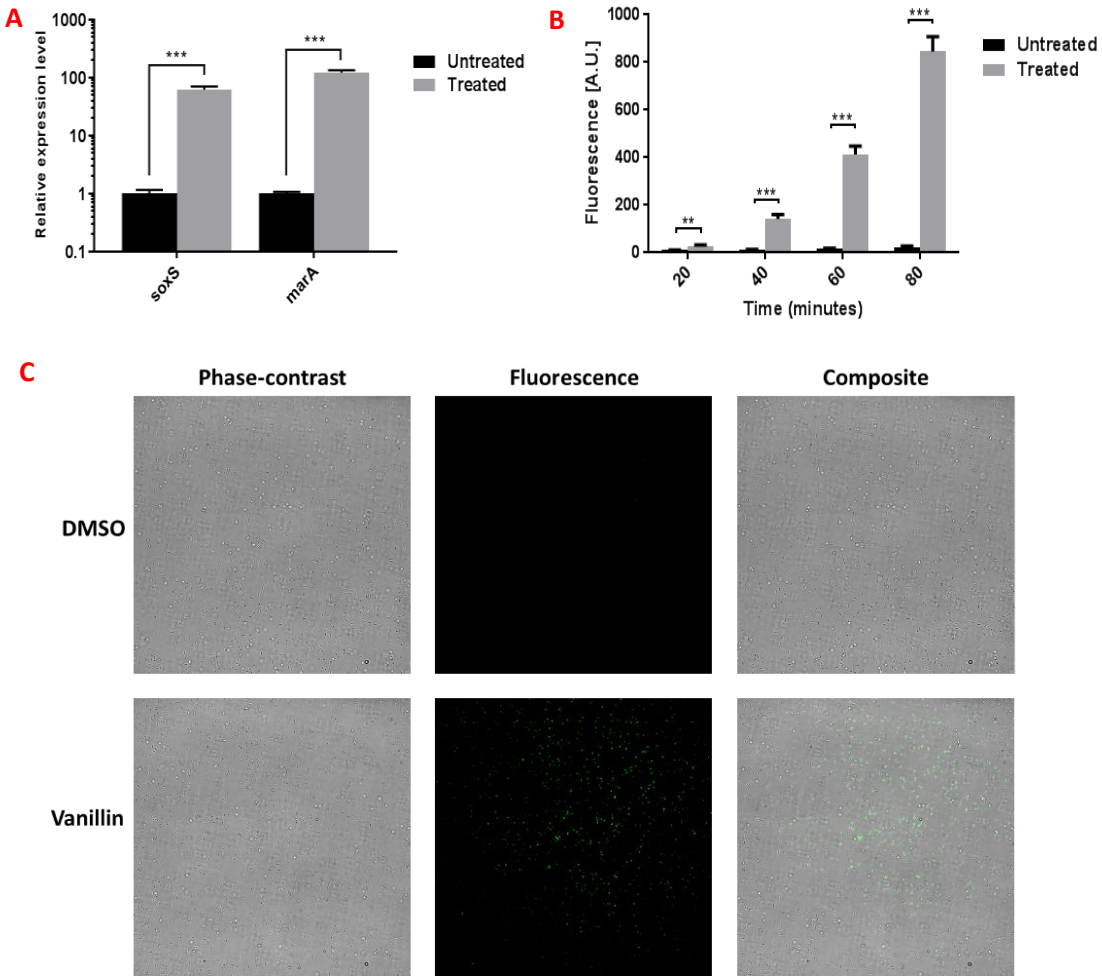


Figure 44: Vanillin treatment elicits an oxidative stress response in *E. coli.* A) Real-time PCR showing transcript levels of oxidative stress response transcriptional activators when grown with and without 10 mM vanillin. Expression levels are relative to untreated cells and normalised to the *rrsA* housekeeping gene. Data are plotted as means of three biological replicates (each consisting of three technical replicates) with standard deviations as error bars. B) Endogenous reactive-oxygen species (ROS) levels in *E. coli* in response to vanillin treatment detected by use of the ROS-activated fluorescent dye 2', 7'-dihydrodichlorofluorescein diacetate (H_2DCFDA). Data are plotted as means of three biological replicates with standard deviations as error bars. C) Fluorescence microscopy of DMSO- (control) and vanillin-treated *E. coli* cells stained with H_2DCFDA. *, ** and *** denote statistical p-values of 0.01 – 0.05, 0.001 – 0.01, and < 0.001, respectively.

E. coli also exhibits an extensive SoxRS-independent oxidative stress response. The redox state of OxyR is affected by peroxide exposure, with $OxyR^{ox}$ activating transcription of several genes. These include alkyl hydroperoxide reductases AhpC and AhpF, both of which were found to be more abundant during growth in the presence of vanillin, by 4.0- and 2.6-fold respectively (Table 5). The manganese-containing superoxide dismutase SodA is positively regulated by both SoxS and OxyR and showed a 4.2-fold increase in abundance, while the iron-containing superoxide dismutase SodB appeared unregulated by vanillin treatment. Transcription of the glucose dehydrogenase Gcd is known to be promoted by both SoxS and OxyR but repressed by cAMP(Yamada et al., 1993); here Gcd showed no change in abundance. OxyR, along with the ferric uptake regulatory protein, Fur, induces the *suf* operon that encodes proteins for iron-sulfur cluster synthesis (Blanchard et al., 2007). SufB, SufD and SufS were upregulated 6.1-, 6.2- and 6.4-fold respectively by growth with vanillin.

There were other proteome indicators of vanillin-dependent oxidative stress including proteins that are regulated indirectly or entirely independently of the SoxRS and OxyR regulons. The peroxiredoxin OsmC was upregulated 5.0-fold; the catalase KatE 14.7-fold; and NfnB and MdaB, typically upregulated as a response to oxidative stress via MarR/MarA, showed a 14.3- and 8.2-fold increase in abundance respectively. Additionally, proteins key to reduced glutathione (GSH) cycling - a process that that is important in response to oxidative, osmotic and other stresses - also showed upregulation (Figure 45). The GSH synthetase GshB, which catalyses one of the steps for GSH synthesis from the amino acids glycine, glutamate and cysteine, was upregulated 14.2-fold. Several enzymes that utilise GSH for alleviating the effects of oxidative stress – GstA, YghU, Gss, YqjG – also showed an increase in abundance. GSH donates reducing equivalents to ROS to neutralise them, undergoing oxidation and dimerization to form glutathione disulfide (GSSG) itself; glutaredoxin-2, GrxB, is involved in the regeneration of GSH through reduction of GSSG and was upregulated 8.2-fold. The GSH-dependent glyoxalases involved in methylglyoxal detoxification, GloA, GloB and GloC, were not quantifiable in our analysis, while a *gloB*-deficient strain showed no change in vanillin sensitivity compared to wildtype (Figure 43).

The set of downregulated proteins also includes indicators that oxidative stress is an important feature of vanillin toxicity. Several anaerobically-induced proteins, mostly controlled by Fnr, showed a decrease in abundance: NirB, GlpA, GlpB, DmsA, NarH, NarG,

AnsB, AspA, FumB, FrdA, and FrdB. A common feature in these and some other downregulated proteins, such as TdcG and SdcB, is that they contain oxygen- and ROS susceptible FeS clusters. Two proteins that generate the oxygen-sensitive glycyl radical in catalysis, GrcA and TdcE, were downregulated 4.8- and 119.9-fold respectively.

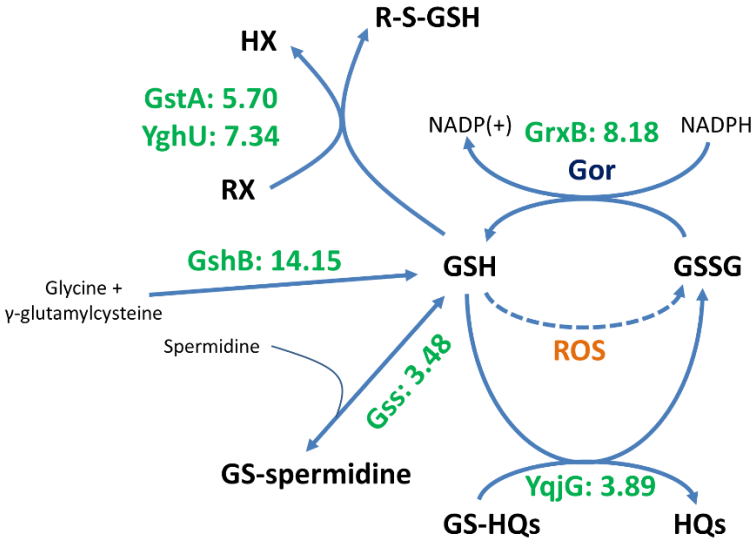


Figure 45: Changes to glutathione cycling caused by vanillin. Enzymes in green showed an increase in abundance; the numbers represent the fold increase they exhibited in the presence of vanillin. GSH: glutathione; GSSG: glutathione disulfide; GS-spermidine: glutathionylspermidine; GS-HQ: glutathionyl-hydroquinone; HQ: hydroquinone.

A cell-permeant indicator of ROS, 2',7'-dichlorodihydrofluorescein diacetate (H_2DCFDA), was used in an assay to directly measure endogenous levels of ROS production in response to vanillin and the related compounds vanillic acid and vanillyl alcohol. Treatment with vanillin causes a significant increase in ROS production relative to untreated cells (Figure 44B), reinforcing the notion that this is one of the mechanisms of vanillin toxicity. Imaging of the cells stained with H_2DCFDA visualised by fluorescence microscopy confirmed that vanillin caused increased ROS production (Figure 44C). Moreover, vanillin dependent ROS production was shown to be greater in *E. coli* strains with deletions in key genes involved in the oxidative stress defence; *sodA, osmC* and *katE* (S. Fig 5). Vanillic acid and vanillyl alcohol treatment resulted in no ROS production (S. Fig 6).

4.4.3. Vanillin stress affects fumarase regulation

The class II fumarate hydratase FumC showed the fourth-highest fold-increase in abundance (41.6-fold) of any protein in response to vanillin, while its class I counterparts FumA and FumB were unchanged and downregulated, respectively. FumC has been shown

to act as a contingency enzyme for FumA, and is upregulated as a response to iron deficiency and superoxide radicals due to the fact that it lacks the [4Fe-4S] co-factor that FumA requires (Liochev and Fridovich, 1992; Varghese et al., 2003). The importance of this was further highlighted by the fact that a Δ*fumC* mutant strain showed a significant increase in vanillin sensitivity compared to wildtype but Δ*fumA* and Δ*fumB* strains showed no significant change in sensitivity (Figure 43). Fumarase activities in cell-free extracts derived from *E. coli* cultures treated with vanillin were compared to those of untreated cultures (Figure 46). Treated wildtype, Δ*fumA* and Δ*fumB* cells possessed a significantly increased rate of *in vitro* fumarate hydration with untreated cells showing low activity. However, Δ*fumC* cells showed no difference in activity between treated and untreated samples. Taken together, the proteomic data, mutant sensitivity assays and enzyme activities show that vanillin dependent up-regulation of FumC is a key physiological response during growth in the presence of this compound.

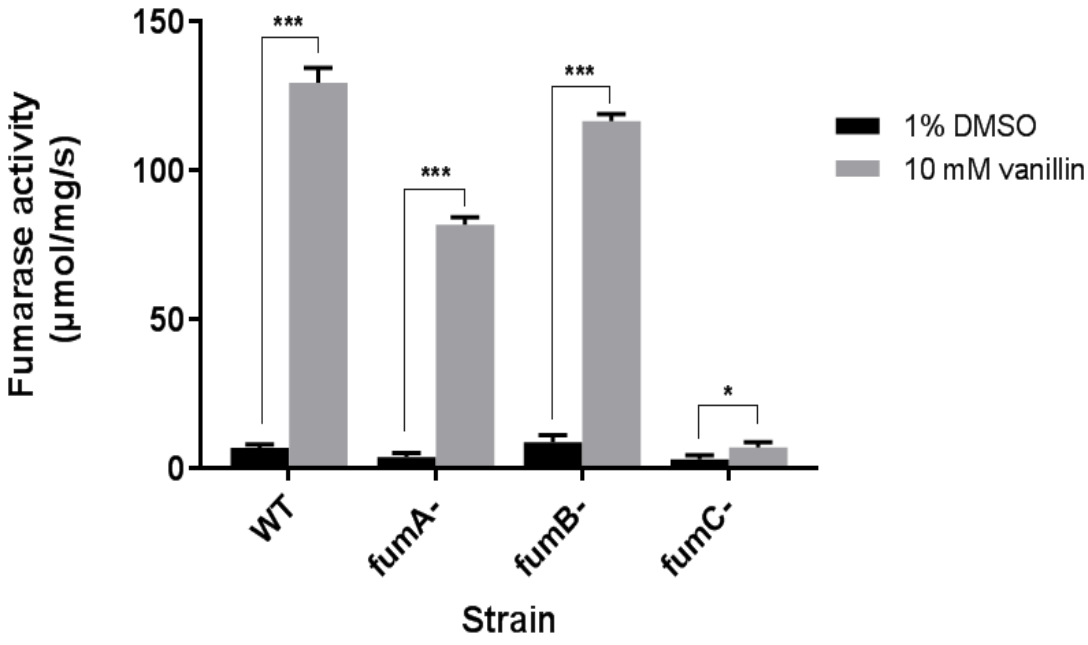


Figure 46: Fumarase C is the prominent fumarase during vanillin treatment. Fumarase activity of *E. coli* cell-free extracts from wildtype and fumarase-deficient cells grown in the presence of DMSO or 10 mM vanillin was measured spectrophotometrically at 240 nm. Columns represent the mean activity of three cell-free extracts, with the standard deviation shown as error bars. *, ** and *** denote statistical p-values of 0.01 – 0.05, 0.001 – 0.01, and < 0.001, respectively.

4.4.4. Proteome indicators of osmotic stress during growth with vanillin

A number of other indicators of general stress responses to growth with vanillin were observed. There was an upregulation of proteins that possess chaperone capabilities: ClpB, 2.8-fold, CbpA, 3.72, HslO, 8.4-fold; and HchA, which also exhibits type-III (GSH-independent) glyoxalase activity, 14.1-fold. The proteolytic enzymes Dcp, PepB and DegP were also upregulated, 3.1-, 5.1- and 5.7-fold respectively; DegP can also act as a chaperone. The upregulation of AcnA, HchA, HdHA, KatE, OsmC, OtsA, TalA and TktB are indicators of osmotic stress. OtsA is UDP-forming α, α-trehalose-phosphate synthase involved in the biosynthesis of trehalose, a low molecular weight compound accumulated in response to high osmolarity (Horlacher and Boos, 1997). Furthermore, TreB and TreC, two proteins vital to the degradative, catabolic utilisation of trehalose as a carbon source, were 24.4- and 10.7-fold less abundant. Other indicators of osmoprotectant accumulation are the increases in abundance of glutamate synthase, GltB, and the glutamate/gamma-aminobutyrate antiporter, GadC. However, we found that supplementing LB medium with additional potassium glutamate up to 50 mM did not improve growth of *E. coli* in the presence of vanillin. The uncharacterised oxidoreductase YjhC was upregulated 15.6-fold; this enzyme has putative glucose-fructose oxidoreductase activity and is potentially involved in production of sorbitol, another osmoprotective compatible solute (Burk et al., 2009).

4.4.5. Changes in metal homeostasis in response to vanillin

Metal homeostasis and oxidative stress in bacteria are intrinsically linked; in particular, careful iron management is a key requirement due to intracellular Fenton chemistry facilitating production of damaging hydroxyl radicals and other ROS (Cornelis et al., 2011). Therefore, the vanillin-dependent increase in oxidative stress might be expected to be linked to a down-regulation of the iron uptake machinery. However, the proteomic data suggest the contrary (Table 5). Three enzymes of the pathway responsible for biosynthesis of the iron siderophore enterobactin, EntB, EntE and EntF, were 41.3-, 7.9- and 14.1-fold more abundant in cells grown with vanillin compared to without. Other indicators of an absence of repressive Fur regulation were an increase in abundance of the uncharacterised periplasmic protein YncE, and examples of RyhB-mediated downregulation. RyhB is a noncoding RNA repressed by Fur, that itself represses expression of iron-utilising proteins including FtnA, NirB, SodB and FeS-containing enzymes of the TCA cycle (Huang et al., 2012). FtnA and NirB were shown to be downregulated in this study (Table 6). The *suf*

operon, the products of which are involved in FeS assembly, is repressed by iron-bound Fur and induced by OxyR. SufB, SufD and SufS were shown to be ~6-fold more abundant during growth with vanillin.

In order to directly assess changes in metal homeostasis, intracellular metal content in response in vanillin treatment was assessed via ICP-MS (Figure 47A). There was a significant increase in the intracellular levels of several transition metals detected in cells grown with vanillin. In particular, iron content increased ~2-fold, suggesting the upregulation of the iron uptake machinery was having a detectable effect. Enterobactin synthesis and secretion was assessed in response to vanillin treatment by the chrome azurol S assay, to determine if this was responsible for this increased iron accumulation, but no change in enterobactin secretion levels was detected (Data not shown).

The largest vanillin-dependent change in metal content was a 17.5-fold increase in manganese; a redox-active metal that acts as an antioxidant itself and is vital as a co-factor for the manganese-containing superoxide dismutase SodA. The proteomic data did not detect any proteins involved in manganese uptake, potentially due to the lower coverage of membrane proteins.

Copper homeostasis is essential in bacteria that use it as a vital enzymatic co-factor, due to its cytotoxicity. Copper efflux via the Cus and CopA transport systems are important means of coping with Cu-mediated damage to Fe-S clusters (Chillappagari et al., 2010; Tan et al., 2014). The copper exporting P-type ATPase encoding gene *copA* was shown to be upregulated 5.4-fold in response to vanillin by RT-PCR (Figure 47B); the effect of this would be an increase of transport of Cu^{+} from the cytoplasm to the periplasm. CopA is a central component of copper homeostasis and is regulated by CueR activated by Cu^{+} (or Ag^{+}) accumulation (Rademacher and Masepohl, 2012). A Δ*copA* strain showed an increased susceptibility to vanillin (Figure 43), highlighting that CopA and copper homeostasis are involved in vanillin toxicity. Furthermore, *E. coli* was unable to grow in minimal media supplemented with 5 µM Cu^{2+} in the presence of vanillin (Figure 47C). Unfortunately, ICP-MS analyses detected only very low copper levels, which were not considered reliable.

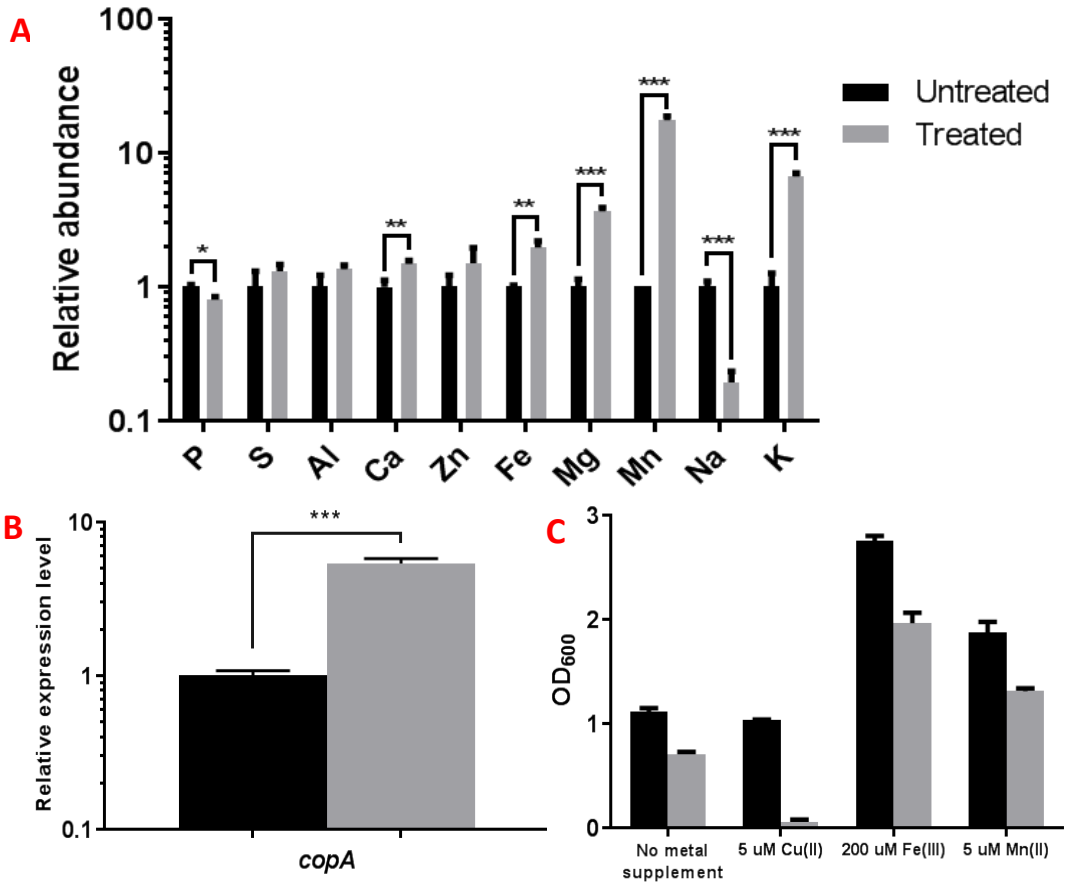


Figure 47: Vanillin treatment causes a perturbation of metal homeostasis. A) Inductively coupled plasma mass spectrometry (ICP-MS) analysis of intracellular metal concentrations of *E. coli* in response to treatment with 10 mM vanillin. Abundance is the mean of three cultures relative to the untreated cells. B) RT-PCR showing expression of *copA* in response to 10 mM vanillin relative to expression in untreated cells. Expression levels are normalised to the *rrsA* housekeeping gene. Data are plotted as means of three biological replicates (each consisting of three technical replicates) with standard deviations as error bars. C) Cell-density after 24 h growth of *E. coli* BW25113 in minimal media supplemented with metal concentrations shown. The mean OD_{600} from cultures grown in triplicate are shown. *, ** and *** denote statistical p-values of 0.01 – 0.05, 0.001 – 0.01, and < 0.001, respectively.

4.4.6. Enzymes involved in direct detoxification of vanillin

Several enzymes previously shown to be involved in reducing aromatic aldehydes to alcohols were identified in the proteomics data and were all more abundant during growth with vanillin: the aldo-keto reductase DkgA was 50.8-fold more abundant (the second largest increase), and the aldehyde reductases YqhD, Ahr, and YahK were 49.3-fold (the

third largest increase), 7.80-fold and 4.67-fold more abundant, respectively (see Table 5). YjbG was not identified by MS and DkgB and YeaE were identified but were not included in the final set of quantified proteins used for statistical analysis. The oxidoreductase YghA is annotated as uncharacterised but has been shown to have aldehyde reductase activity and was 5.95-fold more abundant (Kunjapur et al., 2014a). The aldehyde dehydrogenase AldB, which oxidises aldehydes to carboxylic acids, was 9.49-fold more abundant during growth with vanillin. Single-gene knock-outs of DkgA, YqhD, and AldB all resulted in no significant change to vanillin tolerance (Figure 43). Purified YqhD and AldB were assessed for reductive and oxidative activity, respectively, with vanillin. YqhD was confirmed as capable of reducing vanillin to vanillyl alcohol (S. Fig 7); no activity was detected with AldB.

CurA, the NADPH-dependent curcumin reductase, also showed a 5.02-fold increase in abundance. Curcumin has a similar structure to vanillin with two phenolic rings with 4-hydoxy, 3-methoxy substitutions and is toxic to *E. coli* (Shlar et al., 2017); CurA plays a key role in its detoxification. The uncharacterised protein YhbW exhibited the greatest increase in abundance during growth with vanillin; its sequence shows homology to alkanal monooxygenases and its potential to oxidise vanillin to vanillic acid using $FMNH_2$ and molecular oxygen was investigated but no activity or binding was discovered with overexpressed recombinant protein *in vitro* (Data not shown). RT-PCR confirmed that *yhbW* was upregulated in *E. coli* grown in the presence of vanillin; despite *yhbW* having been reported to be part of the SoxRS-regulon, this upregulation was identified at a similar level in a *soxS*-deficient strain (S. Fig 8).

4.4.7. Perturbation in central metabolic pathways during growth in the presence of vanillin

Several of the enzymes described above that potentially catalyse the reduction of vanillin to vanillyl alcohol, and those involved in glutathione regeneration, require NADPH. We found evidence that a number of enzymes in central catabolic pathways that regenerate NADPH were more abundant in response to the presence of vanillin (Figure 48). This was particularly apparent for the pentose phosphate pathway (PPP) that generates both NADPH and precursors for nucleotide and aromatic amino acid biosynthesis. The glucose-6-phosphate 1-dehydrogenase, Zwf, is the first step in this pathway, is regulated by SoxS and directly produces NADPH. Zwf, along with other key PPP enzymes such as TalA and TktB were all significantly increased in abundance during growth with vanillin (Table 5). Interestingly, although TktA in *E. coli* is the major transketolase, its abundance was not

significantly altered by vanillin treatment, whereas the abundance of TktB, which is responsive to stress conditions via RpoS regulation, was increased. A similar situation occurs with the fructose-bisphosphate aldolase enzymes FbaA and FbaB, where FbaA is downregulated by osmotic stress (Webber et al., 2009). The proteomic data show an 8.86-fold increase in FbaB abundance with no change noted in FbaA during growth with vanillin. There appeared to be no changes in the Entner-Doudoroff pathway enzymes phosphogluconate dehydratase and 2-keto-3-deoxygluconate-6-phosphate aldolase.

Another clear pattern we observed was of a change in enzyme abundance that suggests a shift from the complete oxidative citric-acid cycle to use of the glyoxylate bypass. The glyoxalate bypass enzymes isocitrate lyase, AceA, and malate synthase, AceB, were upregulated 8.5- and 9.2-fold respectively during growth with vanillin (Table 5). Enzymes shared by both the citric-acid and glyoxylate cycles – citrate synthase, aconitase and malate dehydrogenase – were also increased in abundance, while Icd, SucAB and SucCD were unchanged. The glyoxalate cycle requires two molecules of acetyl-CoA rather the one required for the TCA cycle. While AceE and AceF were unchanged, Acs increased in abundance with vanillin (14.1-fold). The only additional source of acetate in vanillin-treated cells detected appears to be from pyruvate via pyruvate oxidase PoxB, reducing ubiquinone to ubiquinol in the process. The enzyme GhrA that catalyses the reversible conversion of glycolate to glyoxylate was shown to be 5.0-fold more abundant, while the enzymes GarR and GarD involved in glyoxylate metabolism were less abundant (Table 6). The overall importance of this process is unclear, with knock-out of AceA, AceB, Acs and PoxB having no significant effect on *E. coli*'s susceptibility to vanillin (Figure 43).

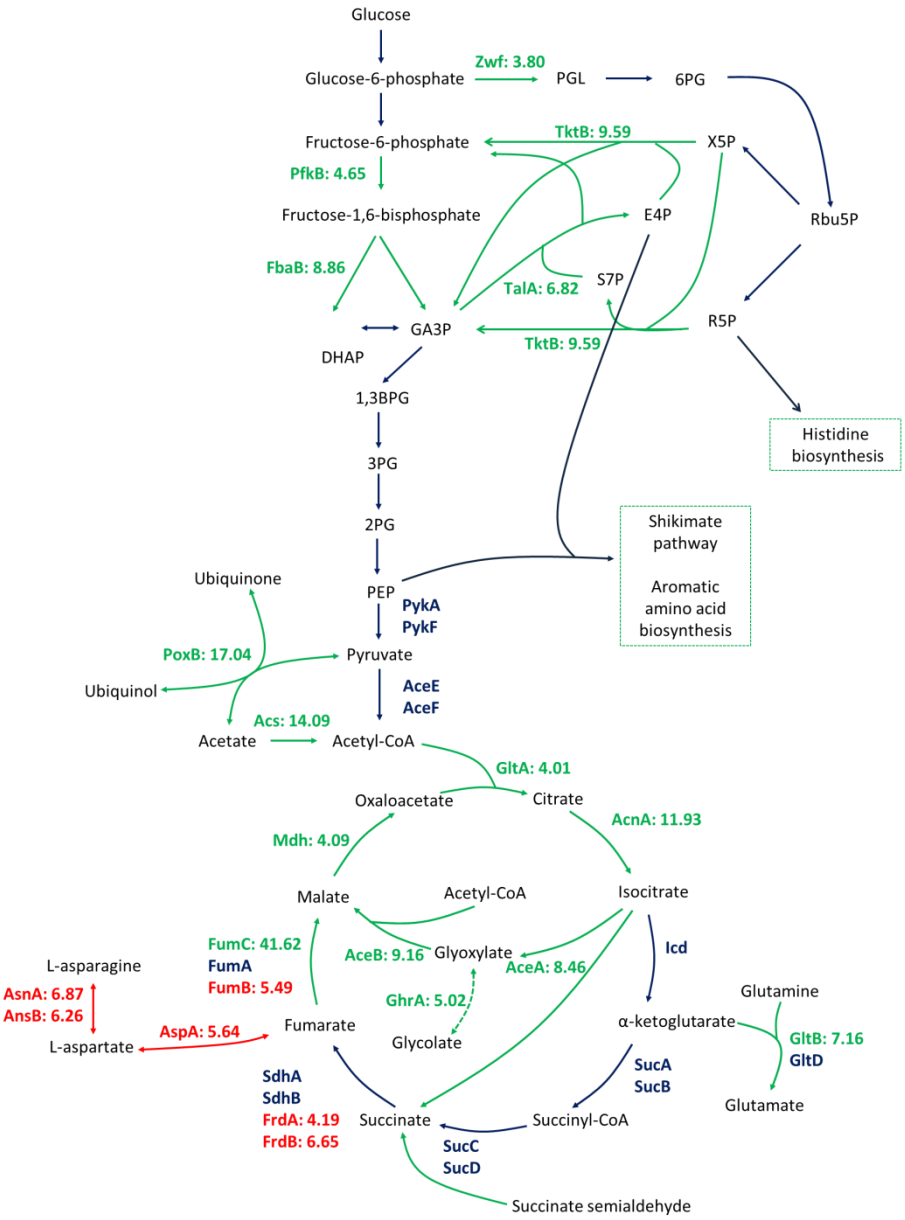


Figure 48: Vanillin treatment perturbs central carbon metabolism. Enzymes in green showed an increase in abundance, and enzymes in red showed a decrease; the numbers represent the fold-increase/decrease they exhibited. Enzymes in blue showed no significant change in abundance. PGL: 6-phosphogluconolactone; 6PG: 6-phosphogluconate; Rbu5P: ribulose-5-phosphate; R5P; ribose-5-phosphate; X5P: xylulose-5-phosphate; E4P: erythrose 4-phosphate; F6P: fructose 6-phosphate; S7P: sedoheptulose 7-phosphate; DHAP: dihydroxyacetone phosphate; GA3P: glyceraldehyde 3-phosphate; 1,3BPG: 1,3-bisphosphoglyceric acid; 3PG: 3-phosphoglyceric acid ; 2PG: 2-Phosphoglyceric acid ; PEP: phosphoenolpyruvate.

Several proteins involved peripherally to these central pathways also appear to be involved in *E. coli's* vanillin response. The succinate semialdehyde dehydrogenase, Sad (or YneI), was 10.5-fold more abundant, and knock-out of the *sad* gene resulted in a greater increase in sensitivity to vanillin than shown by any other knock-out mutant (Figure 43). The knock-out of the neighbouring gene that encodes an aminotransferase, *yneH*, also resulted in a significant increase in vanillin sensitivity. These two enzymes link glutamate metabolism pathways and the TCA cycle. AstC, which is involved in the degradation of arginine to glutamate and succinate, was also 14.1-fold more abundant in response to vanillin.

4.4.8. Identification of putative vanillin-efflux systems

One cellular strategy for vanillin tolerance might involve increased efflux of this compound. Few membrane transport proteins were detected directly from the proteomic data as significantly regulated in response to vanillin. The multidrug efflux protein EmrA, which is part of the tripartite efflux system EmrAB-TolC, and the uncharacterized ABC transporter ATP-binding protein YadG, are the only proteins that showed an increase in abundance that might be involved in vanillin efflux, based on their annotations. However, this study has shown vanillin to elicit a MarA-mediated response (Table 5), which regulates antibiotic efflux proteins at a transcriptional level (Ruiz and Levy, 2014). RT-PCR confirmed that transcription of the multidrug efflux genes *acrA* and *acrB* was upregulated (Figure 49); however, knock-out strains lacking either of these components of the AcrAB-TolC efflux system were less sensitive to vanillin than wild-type *E. coli* (Figure 43). The related protein AcrD, an aminoglycoside efflux pump, was also upregulated at the transcriptional level (Figure 49), and a Δ*acrD* strain did show an increased sensitivity to vanillin (Figure 43). Conversely, AcrF, part of the AcrEF-TolC tripartite efflux system, was down-regulated at the transcriptional level. The AaeAB aromatic amino acid efflux system also increased in abundance; a Δ*aaeB* strain showed an increased sensitivity to vanillin compared to wildtype, but less so than Δ*acrD* (Figure 43) AaeAB in *E. coli* acts as a method for relieving toxicity from accumulation of aromatic acids from metabolism such as *p*-hydroxybenzoate, and expression has been shown to be induced by salicylate (Van Dyk et al., 2004); no efflux of aromatic aldehydes has been reported.

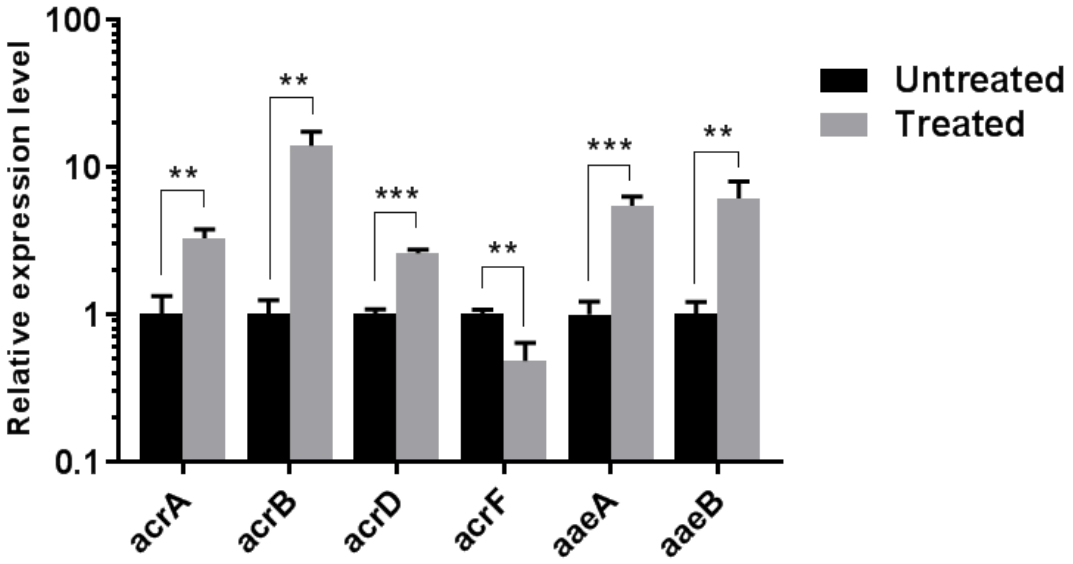


Figure 49: Gene expression studies indicate putative vanillin efflux systems. RT-PCR showing expression levels of several transporter genes in *E. coli* BW25113 cells treated with 10 mM vanillin relative to untreated cells, normalised to expression of *rrsA*. Data plotted as the mean of three biological replicates (each consisting of three technical replicates) with standard deviations shown as error bars. *, ** and *** denote statistical p-values of 0.01 – 0.05, 0.001 – 0.01, and < 0.001, respectively.

4.4.9. Vanillin affects maltose operon regulation

The proteomic data showed that two proteins involved in maltose uptake, the maltoporin, LamB, and the maltose periplasmic binding protein, MalE, were 33- and 15-fold less abundant in *E. coli* cells treated with vanillin. Down regulation of maltose transport in *E. coli* has previously been observed in response to toxic phenylpropanoids (Zhou et al., 2015). Based on this it was hypothesised that these transporters may be involved in vanillin's movement into the bacteria; however, the *lamB* and *malE* knock-mutants showed an increase and no change in sensitivity, respectively, rather than an increased tolerance (Figure 43). Two other maltose-related proteins, MalQ and MalP, were also less abundant in response to vanillin, indicating down regulation of the *mal* operon. *E. coli* growing with maltose as the sole carbon source appeared significantly more sensitive to vanillin compared to growth with glucose (S. Fig 9). Vanillin's effect on metabolic regulation could render *E. coli* unable to utilise maltose as a carbon source efficiently.

4.4.10. Electron microscopy showing vanillin-induced membrane perturbation

Transmission electron microscopy (TEM) was used to assess any morphological changes in *E. coli* cells as a result of vanillin treatment. A culture of *E. coli* BW25113 was grown in the

presence of 5 mM vanillin over a time period of 6 h with samples taken periodically. Samples were stained with uranyl acetate for examination. Visualisation of the cells matched the effect previously shown by an antibacterial arylamide foldamer on *E. coli* (Mensa et al., 2011). After 1 h of incubation with vanillin the cells appearance after staining darkened dramatically (Figure 50) compared to untreated cells that remained unchanged. This indicated increased accessibility of the stain due to increased membrane permeability and persisted through 3 h of treatment. Also apparent was ruffling of the membrane, which appears much less well-defined. By 6 h, it appeared that most cells had recovered and appeared regular in shape and staining. Cells treated with 10 mM were also studied using TEM, and exhibited heavy staining, osmotic swelling and a diffuse halo was observed.

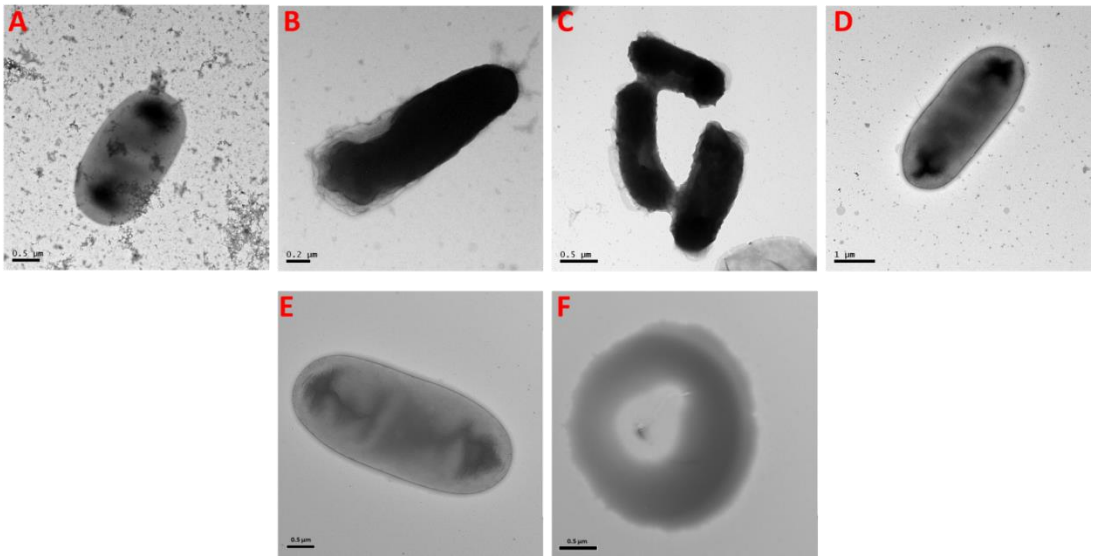


Figure 50: Electron microscopy reveals changes in cell morphology during incubation with vanillin. *E. coli* cells were grown the presence of vanillin and visualised using transmission electron microscopy. A) Membrane is uniform at the time of 5 mM vanillin addition, though cells already appeared truncated B) After 1 h the membrane exhibits severe ruffling with outer membrane vesiculation apparent; the cytoplasm is heaving stained. C) Membrane perturbation and increased strain accessibility persists at 3 h. D) By 6 h surviving cells had returned to normal. E) Untreated *E. coli* after 4 h of incubation showing normal staining F) After 3 h of 10 mM vanillin treatment diffuse halos of stain and no distinctive membrane were visible.

TEM also showed increased clumping of *E. coli* cells treated with vanillin, potentially due to increased biofilm formation — a typical stress response. There is some evidence of increased biofilm formation in the proteomic data: the uncharacterised protein YdhQ was

6.2-fold less abundant, and its knock-out in *E. coli* has been shown to result in increased biofilm production (Vlasblom et al., 2015).

4.4.11. Adaptive laboratory evolution (ALE) of vanillin tolerant E. coli

Four strains of *E. coli* were isolated following repeated, parallel sub-culturing of *E. coli* BW25113 in gradually increasing concentrations of vanillin in rich media. These strains displayed an increased tolerance to vanillin compared to the parent strain, BW25113, and were nominated Evolved-vanillin tolerance (EVT) strains 1–4, with EVT1 appearing to be more tolerant than the other three strains (S. Fig 10). The whole genomes of these strains were sequenced and compared to the parent strain, with the identification of several single nucleotide polymorphisms (SNPs) and deletions in coding regions (summary in Table 7).

All four strains had distinct non-synonymous SNPs in the citrate synthase, *gltA*, gene. GltA in *E. coli* forms a hexameric structure, and the R119F and G126S mutations in EVT1 and EVT3, respectively, occur at the interface between GltA monomers (S. Fig 11); these could result in impaired hexamer formation. The A348T mutation occurs in the region of the acetyl-CoA binding site and is next to a phenylalanine residue (Phe383) that forms part of this site, with a F383A mutation having been previously shown to reduce substrate turnover (Pereira et al., 1994). The A160V mutation in EVT4 occurs next to an isoleucine residue (Ile159) that is part of the allosteric NADH binding site. Additionally, an A161V mutant GltA has been characterised that has an increased K_m for oxaloacetate, and decreased activity (Quandt et al., 2015; Stokell et al., 2003). All four of these mutations could feasibly result in decreased citrate synthase activity, impacting the entire TCA cycle.

Table 7: Coding SNPs identified in evolved vanillin tolerant (EVT) *E. coli* strains.

Strain	Position	Annotation	Gene	Description
EVT1	749,566	R119L (CGT→CTT)	*gltA*	Citrate synthase
	4,625,115	Δ1 bp, coding (12/870 nt)	*rob*	Right origin-binding protein
EVT2	748,772	A384T (GCA→ACA)	*gltA*	Citrate synthase
	1,612,766	Δ11 bp, coding (389–399/666 nt)	*marC*	Inner membrane protein
EVT3	749,516	G136S (GGT→AGT)	*gltA*	Citrate synthase
	4,624,518	Q203stop (CAG→TAG)	*rob*	Right origin-binding protein
	3,169,702	L162del (CTGCAT→CAT)	*cpdA*	cAMP phophodiesterase
EVT4	749,443	A160V (GCC→GTC)	*gltA*	Citrate synthase
	1,605,879	Δ9,699 bp	12 genes	*dgcF – eamA*

Two strains were shown to have a mutation in the *rob* gene that encodes an AraC-XylS-family transcriptional regulator, with its regulatory effects intrinsically coordinated with those of SoxS and MarA and involved in oxidative stress (Chubiz et al., 2012). The frameshift caused by the single nucleotide deletion in the *rob* gene in EVT1 results in a premature stop codon and a severely truncated product of just nine amino acids. The SNP in EVT3 also causes a premature stop codon, preventing translation of the 85 C-terminal amino acids. Both these mutations likely severely impair Rob-dependent regulation, and as the *Δrob* knock-out mutant also showed increased tolerance (Figure 43) this regulation appears to be detrimental to *E. coli*'s tolerance to vanillin. EVT3 was also shown to have a 3 bp deletion in *cpdA*, which encodes a 3',5'-cyclic AMP phosphodiesterase. This results in an in-frame deletion of a leucine residue (Leu162) that has been shown to be semi-conserved in bacterial cAMP phosphodiesterases.

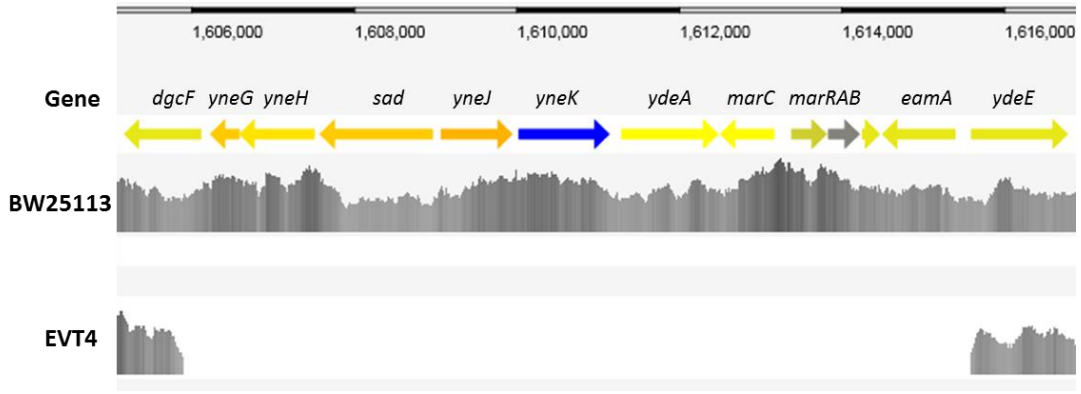


Figure 51: The ~10,000 bp deletion in EVT4. Genome sequencing revealed that a 9,699 bp region had been deleted in EVT4, this region encompassed the genes shown: *dgcF* – *eamA*.

An 11 bp deletion was identified in the inner membrane protein *marC* gene in EVT2. Additionally, EVT4 also had a ~10 kb deletion that covered 12 genes (11 entirely deleted, and partial N-terminal deletion of *dgcF*), including *marC* and the entire *marRAB* operon (Figure 51). Unexpectedly, the single-gene knock-out of some of these genes — *sad*, *yneH*, *ydeA* — were actually more sensitive to vanillin rather than more tolerant. Increased tolerance was seen with single-gene knock-outs of *marR* and *eamA*. The effect of deleting *marR* reinforces the role of *marRAB*-mediated regulation in vanillin stress. The gene *eamA* encodes a putative cysteine (and cysteine-pathway-metabolite) efflux pump, the deletion of which could increase the intracellular cysteine pool. The vastly conflicting effects of separately deleting each of the genes in the ~ 10 kb region makes this genome sequence hard to interpret.

4.4.12. Analysis of the interactions of vanillin with metals

Spectrophotometric assays using the cuprous-ion-sensitive BCS showed that 0.5 mM vanillin appeared to physically reduce Cu^{2+} ions to Cu^{+} ions *in vitro* in buffer (S. Fig 12). This activity was also shown with the structurally related compounds vanillic acid, vanillyl alcohol and ferulic acid, but not with benzaldehyde, benzoate, 3-anisaldehyde (3-methoxy-benzaldehyde) and 4-hydroxy-benzaldehyde. This indicates that the presence of both the methoxy group on C4 and the hydroxyl group on C3 are required for this activity, and it is unrelated to the aldehyde moiety or the aromatic ring itself. A similar experiment with vanillin was repeated with iron using ferrozine as an indicator, but detected no analogous effect on Fe^{3+} ions.

4.5. Discussion

Biotransformation to produce vanillin using microbes is an area of biotechnology that has blossomed in interest over recent years due to the abundance and origin of the precursors, the relatively straightforward enzymatic pathway involved, and high value of and growing desire for the naturally sourced product. As well as problems such as product recovery and the undesired further conversion of vanillin to unwanted products, the toxicity of this aromatic aldehyde to the organisms being utilised to produce it is a hurdle that must, and can, be overcome to increase overall yields. This study offers a broad view of how bacteria, in this case *E. coli,* react to the presence of inhibitory levels of vanillin in their environment.

Mass spectrometry-based proteomic analysis identified sets of proteins that were significantly more and significantly less abundant in *E. coli* BW25113 cells in response to vanillin treatment. The proteomics results are validated by the consistent effect on the abundance of proteins encoded in distinct operons and regulons; the identification of proteins known to be involved in vanillin and related compound toxicity/defence; and the follow-on physiological and growth experiments conducted in this study. Protein identification and the statistical parameters for determining a significant change in abundance was conducted stringently, thus providing a robust data set for analysis.

The major conclusion of these proteomic data is the discovery of a clear oxidative stress response on vanillin treatment, which was somewhat unforeseen considering vanillin's reported antioxidant characteristic (Tai et al., 2011). A perturbation of metal homeostasis and central metabolic pathways was also revealed; the former is potentially intrinsically linked to the oxidative stress response. Furthermore, potential detoxification and efflux systems have been identified.

Aldehydes are reactive and are known to be toxic, and there were expectations that vanillin would elicit a response akin to that displayed by *E. coli* in response to other aldehydes. However, there were no indications from the proteomics that *E. coli*'s defence against vanillin is similar to its means of coping with endogenously occurring aliphatic aldehydes such as malondialdehyde, methylglyoxal and formaldehyde. While GSH plays a role in their detoxification, key enzymes involved such as glyoxalase I and II for methylglyoxal were not detected and the formaldehyde dehydrogenase FrmA showed no significant change in abundance; furthermore, ICP-MS showed an increase in intracellular K^+ levels in response to vanillin that contrasts the KefBC-mediated defence against methylglyoxal (MacLean et al., 1998).

It has been suggested that inhibition of growth by aldehydes is caused by an NADPH drain in the cell, with aldehyde reductases consuming these cofactors at excessive rates (Jarboe, 2011; Miller et al., 2009b); silencing of genes encoding such enzymes – YqhD and DkgA – was shown to actually increase furfural tolerance in *E. coli* (Miller et al., 2009b; Wang et al., 2013). NADPH has a role in multiple biological processes in bacteria as an essential electron donor in biosynthetic pathways and buffering reactive oxygen species (Singh et al., 2007; Spaans et al., 2015). The group investigating furfural tolerance linked NADPH-depletion to sulfur assimilation based on supplementation with cysteine and methionine increasing tolerance, and the NADPH requirement of sulfite reductase (Miller et al., 2009a). However, the actual significance of NADPH depletion was uncertain as a strain of *E. coli* specifically engineered for aromatic aldehyde accumulation, by knock-out of genes encoding NADPH- (and NADH-) utilising enzymes, experienced aldehyde-induced toxicity to a similar degree (Kunjapur et al., 2014a). The proteomic data from this study identified several NADPH-utilising enzymes of this nature amongst the list of proteins that showed the greatest increase in abundance in the presence of vanillin – DkgA, YqhD, Ahr and YghA. While the protein that showed the greatest change in abundance, YhbW, would not directly utilise NADPH for its putative monooxygenase function, it would require a reduced flavin cofactor produced by an NADPH-dependent flavin reductase. The recycling of the antioxidant glutathione also requires NADPH for the reduction of glutathione disulfide.

The NADPH-dependent curcumin reductase was also found to be more abundant during growth with vanillin; while curcumin does not possess an aldehyde moiety, the side groups on its phenol rings are identical to vanillin. Studies into the antibacterial effects of curcumin have shown similar results to those found in this study with vanillin with curcumin shown to cause oxidative stress, a disruption iron homeostasis and a dysregulation of FeS-cluster biosynthesis in *E. coli*, as well as an increase in membrane permeabilisation of several microbes (Shlar et al., 2017; Tyagi et al., 2015). Due to *E. coli* predominantly being located in the intestines of animals they do not encounter plant-derived phenolic compounds to the degree that soil bacteria do, and thus have not evolved a means of metabolising them. However, it is expected that they would encounter such phenolics at some level due to the ingestion of plant matter by their hosts, hence the evolution of a curcumin detoxification mechanism.

That the two compounds produce a strikingly similar stress response despite curcumin lacking the aldehyde functional group of vanillin brings into question the level of contribution that the aldehyde nature of vanillin plays in the toxic effect indicated in this

study. The metal interaction assays conducted in this study using a variety of substituted phenolic compounds have shown that the presence of 3-methoxy and 4-hydroxy side groups – a feature vanillin and curcumin do share – in combination result in interactions with copper *in vitro*. Ferulic acid and sinapic acid, both of which have these same side groups, have also been shown to reduce Fe^{3+} *in vitro* (Hynes and O'Coinceanainn, 2004). The significance of this to the overall antibacterial effect of vanillin is somewhat diminished by the relatively low toxicity of ferulic acid and vanillyl alcohol. Furthermore, as neither vanillic acid nor vanillyl alcohol treatment resulted in an increase in endogenous ROS, it must be the aldehyde moiety that is responsible for eliciting an oxidative stress response (S. Fig 6). However, ferulic acid has been shown to act as a siderophore (Schrey et al., 2012); if vanillin can also function in this way then it could be responsible for the increased uptake of iron, which in turn has a synergistic effect on the production of ROS.

Vanillin's interactions with copper could be key to its toxicity. The role of CopA, the apparent reductive capabilities of vanillin and the copper sensitivity data suggest that the presence of vanillin results in an accumulation of intracellular Cu^{+}. It has been shown that Cu^{+} ions destabilise FeS clusters, inhibit FeS assembly, and promote FeS cluster biogenesis, iron acquisition and sulfur acquisition in bacteria (Chillappagari et al., 2010; Tan et al., 2014). This fits the apparent derepression of the Fur regulon shown with vanillin that occurs despite the accumulation of ROS. Another speculation is that intracellular vanillin is binding iron preventing its interaction with Fur, resulting in a "false" low iron signal. Glutathione has also been identified as key to *E. coli*'s defence against copper toxicity (Große et al., 2014). Additionally, *eamA* was deleted in EVT4 and the *ΔeamA* Keio mutant showed an increased tolerance to vanillin. EamA is annotated as a probable cysteine efflux pump, and maintaining an adequate cysteine pool (in addition to an iron pool) is essential to FeS maintenance and copper defence (Chillappagari et al., 2010).

CopA production has also been proposed to be induced by cell envelope stress via the Cpx regulon (Outten et al., 2000). The EM images of vanillin-treated *E. coli* indicate perturbation of the membrane, while the increase in abundance of DegP and decrease in abundance of OmpF, and increased expression of AcrD, are proteomic and transcriptomic indicators of an envelope-stress-induced response, respectively (Mensa et al., 2011; Price and Raivio, 2009). Furthermore, the significant change in the intracellular Na^{+}:K^{+} ratio elicited by vanillin treatment could be an indication of loss of membrane integrity. Lignocellulose-derived compounds with similar structures to vanillin, including ferulic acid and syringaldehyde, have been shown to inhibit enzymes such as the UDP-glucose 4-epimerase

GalE resulting in cell wall defects; however, no GalE response was exhibited in response to vanillin (Forsberg et al., 2016). The decrease in intracellular Na^+ could be linked to copper in an alternative fashion. Cu^+ ions are similar in charge and ionic diameter to Na^+ ions and may enter bacteria via sodium uptake systems (Große et al., 2014); while not observed in the proteomic data here, *E. coli* may adapt to vanillin-induced copper stress by downregulation of sodium importers resulting in reduced intracellular Na^+.

A recent study has also linked copper stress to glutamate/glutamine metabolism in *E. coli* (Djoko et al., 2017). Copper stress was shown to disrupt glutamate synthesis by glutamate synthase, which could explain the increase in abundance GltB, GadC and AstC shown in response to vanillin as these proteins are all involved in glutamate metabolism. Furthermore, the increased abundance of Sad, and the increased vanillin sensitivity of the *sad* and *yneH* knock-out strains, also points towards a key role of glutamate metabolism in the response to vanillin. As the compatible ion for glutamate, the increase in intracellular potassium may also indicate increased glutamate accumulation. Glutamate and gamma-aminobutyrate pathways are also known to involved in acid stress resistance, anaplerosis and oxidative stress (Feehily and Karatzas, 2013).

It is clear from the study that growth in the presence of vanillin perturbs central carbon metabolism in *E. coli*, despite no alterations in the levels of carbon source in the growth media. While the increase in the abundance of key enzymes of the pentose phosphate pathway could be attributable to the need to increase production of NADPH via the activity of the glucose-6-phosphate 1-dehydrogenase Zwf, the reason for the apparent shift to the glyoxalate cycle is more ambiguous. One of the roles of a shift to the glyoxylate cycle is to allow gluconeogenesis and synthesis of carbohydrates from acetate (Renilla et al., 2012); however, the enzyme responsible for the committed gluconeogenesis step, phosphoenolpyruvate carboxykinase (PckA), showed no change at the proteomic level. As the level of acetate initially available exogenously was consistent between treated and untreated cultures, it is unlikely that the glyoxylate cycle is being activated to allow growth on acetate; intracellular acetate levels could be increasing due to the significant increase in PoxB abundance. PoxB reduces ubiquinone to ubiquinol as part of the dehydrogenation of pyruvate to acetate and this reduced form has been shown to be an essential redox component and antioxidant defence in *E. coli* (Soballe and Poole, 2000). Therefore, the increase in PoxB may be a direct defence against vanillin-induced oxidative stress, with the glyoxalate cycle an indirect response to the acetate produced. The glyoxylate cycle has also been proposed as a method of forming malate during oxidative stress, due to the FeS-

containing fumarases being compromised. Overall the importance of the glyoxylate cycle in response to vanillin is unclear, with the single-gene knockouts in either AceA or AceB, and PoxB, having no significant effect on *E. coli* vanillin sensitivity.

Furthermore, the impairment of ubiquinone biosynthesis in an *ubiCA* mutant has been shown to cause oxidative stress (Soballe and Poole, 2000). The chorismate pyruvate-lyase UbiC catalyses the initial step of the ubiquinone biosynthesis, and has been shown to be inhibited by vanillic acid and benzaldehyde (Holden et al., 2002). The structural similarity of these two compounds to vanillin indicates that vanillin is likely to have a similar inhibitory effect that would contribute to oxidative stress. Vanillin could also feasibly inhibit other chorismate enzymes such as isochorismate synthase (MenF), p-protein (PheA) or aminodeoxychorismate synthase (PabAB). This could potentially provide an explanation for the increase in abundance of enzymes involved in processes such as folate biosynthesis and aromatic amino acid synthesis; *E. coli* must increase the levels of these enzymes to compensate for the inhibition.

A study into resveratrol, naringenin and rutin tolerance in *E. coli* identified several putative transporters of these phenylpropanoids (Zhou et al., 2015). Their results suggested that OmpA and FadL were involved in the movement of phenylpropanoids across the outer membrane, while the maltose uptake system participated in their movement across the cytoplasmic membrane. The proteomic data in this study support the idea that vanillin might enter the cell cytoplasm by similar route; the porin FadL was less abundant in the presence of vanillin along with two proteins involved in maltose uptake, outer membrane maltose/maltodextrin porin, LamB, and the periplasmic maltose binding protein, MalE (Table 6). OmpA in contrast showed no significant change in abundance, though another outer membrane porin, OmpF, was amongst the proteins that showed the greatest decrease in abundance, in addition to FadL, LamB and MalE, indicating they may be responsible for uptake of vanillin. The products of two other *mal* genes, MalP and MalQ, showed decreases in abundance suggesting negative regulation of the maltose regulon (Boos and Shuman, 1998). The regulation of maltose and trehalose metabolic genes is linked, and TreB has been shown to also uptake maltose; TreB and TreC were also less abundant in the presence of vanillin. The growth of the *ΔlamB* and *ΔmalE* mutants on vanillin doesn't corroborate with the idea that these are involved in vanillin's movement into the cell. However, *E. coli* appeared significantly more susceptible to vanillin when maltose was the carbon source instead of glucose. Rather than this being a result of the maltose facilitating vanillin uptake due to hijacking of the transport systems, the effects of

vanillin stress on catabolite repression and metabolism regulation may reduce the bacteria's ability to utilise the maltose as a sole carbon source effectively.

The identified alterations in the genomes of the evolved vanillin-tolerant *E. coli* strains also add to the overall picture of vanillin toxicity, while showing there are multiple ways to achieve vanillin tolerance. The increase in ROS production exhibited during exposure to vanillin could be a direct result of the increased metabolic and respiratory activity, e.g. the TCA cycle, indicated by the increase in abundance of metabolic enzymes shown in the proteomic analysis. The mutations in *gltA* could result in reduced citrate synthase activity, carbon flux into the TCA cycle and respiration, and reduced production of ROS as a result. Increased respiration was proposed to be the origin of oxidative stress elicited in *n*-butanol toxicity, while *E. coli* strains evolved for tolerance to isopropanol, which also causes oxidative stress, showed decreased respiratory levels (Horinouchi et al., 2017; Rutherford et al., 2010).

The detection of mutations in the *rob* gene and the deletion of the *marRAB* operon, plus the increased vanillin tolerance shown by the *Δrob* and *ΔmarR* Keio strains, is somewhat surprising considering their role in antimicrobial resistance and oxidative stress responses. A major outcome of regulation via the MarA regulon is increased production of the AcrAB efflux pump, with this upregulation previously shown to be Rob mediated (Rosenberg et al., 2003). This could be an undesirable response during vanillin stress, as the *ΔacrA* and *ΔacrB* strains showed the greatest increase in tolerance. The knock-out of the *acrA* and/or *acrB* genes has been shown to increase tolerance to other biotechnologically relevant compounds such as isoprenol and isobutanol (Ankarloo et al., 2010; Minty et al., 2011a; Wang et al., 2015). It has been proposed that this could be due to the absence of AcrAB resulting in increased levels of other efflux systems such as AcrD and AcrF, with the former potentially identified as a vanillin efflux pump here (Blair et al., 2015; Webber et al., 2009).

Interestingly, two of the six evolved isopropanol-tolerant strains of *E. coli* also showed large (25 and 34 genes) deletions in the same region of the genome covering the ~10,000 bp deleted here; a third isopropanol-tolerant strain also had a mutation in *marC* (Horinouchi et al., 2017). It was also shown that the single-gene knock-out of *marC* resulted in increased isopropanol, isobutanol and *n*-butanol tolerance; while the function of the membrane protein MarC is unknown (it has been shown not to be regulated by MarR or involved in antibiotic susceptibility), it putatively could be an alcohol channel or pore protein (Dragosits and Mattanovich, 2013; McDermott et al., 2008). The lack of increased

tolerance to vanillin shown by the *ΔmarC* mutant implies MarC is not necessarily the key to vanillin toxicity.

Through expanding the knowledge of vanillin's mechanism of toxicity we have provided physiological targets for manipulation to improve tolerance, vitally this includes gene targets for overexpression that do not achieve tolerance through conversion of vanillin to less toxic compounds. The significant increase in intracellular ROS implies that overexpression of oxidative stress defences, such as SodA, could prove effective, while overexpression of CopA and CueO could potentially alleviate accumulation of Cu^{+}. Engineering of the regulatory cAMP receptor protein (CRP) has been proven as a method of improving oxidative stress tolerance in *E. coli* and could be applied to vanillin (Basak and Jiang, 2012). The mutation in *cpdA* in EVT3 indicates cAMP levels could have an effect on tolerance. Identifying an effective endogenous efflux system for vanillin in *E. coli* would provide a target that could be manipulated to result in an increased tolerance to synthesised vanillin, while also potentially improving product recovery. While this study has not provided definitive proof of a vanillin efflux system, the proteomic data and vanillin sensitivity of mutants suggest that the AaeAB efflux pump or AcrD may be involved. This study supports, and should be combined with, the findings of Kunjapur et al. (2014), with deletion of aldo-keto reductase and alcohol dehydrogenases genes facilitating vanillin accumulation and dampening the NADPH drain (Kunjapur et al., 2014b).

This study has helped to identify a range of targets for genetic manipulation to enhance vanillin tolerance. Importantly, overexpression of many of these would not lead to a reduction in accumulation of biotransformed vanillin, which would deleteriously impact on its industrial potential. The mechanism of vanillin's toxicity, and *E. coli*'s defence against it, was previously poorly understood. These results could also potentially be applied to other areas of biotechnology such as microbial production of benzaldehyde, which has the second largest market share of flavours after vanillin, or dealing with furfural and lignin-feedstock toxicity in bacteria (Kunjapur and Prather, 2015). Furthermore, aldehydes are used as a building blocks for further industrial conversion; vanillin is a precursor for the production of polymers such as polyesters, phenolic resins and thermosetting plastics (Fache et al., 2015). This study has shed light on the toxic effects of vanillin in *E. coli*, which arguably could be expanded to the effects of other phenolic aldehydes in other microbial species, though more focused investigation on specific targets – e.g. efflux systems – is needed to work towards engineering a proficiently tolerant strain.

4.6. Acknowledgements

Genome sequencing was provided by MicrobesNG (http://www.microbesng.uk), which is supported by the BBSRC (grant number BB/L024209/1). Many thanks to Robert Poole (University of Sheffield) for providing Keio mutants, Craig MacGregor-Chatwin (University of Sheffield) for assistance with the TEM, Neil Bramall for running the ICP-MS and Adam Brooks (University of Sheffield) for help with fluorescence microscopy.

4.7. References

Alsaker, K.V., Paredes, C., and Papoutsakis, E.T. (2010). Metabolite stress and tolerance in the production of biofuels and chemicals: gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe Clostridium acetobutylicum. Biotechnol Bioeng 105, 1131-1147.

Ankarloo, J., Wikman, S., and Nicholls, I.A. (2010). Escherichia coli mar and acrAB mutants display no tolerance to simple alcohols. Int J Mol Sci 11, 1403-1412.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2, 2006 0008.

Barghini, P., Di Gioia, D., Fava, F., and Ruzzi, M. (2007). Vanillin production using metabolically engineered Escherichia coli under non-growing conditions. Microb Cell Fact 6, 13.

Basak, S., and Jiang, R. (2012). Enhancing E. coli Tolerance towards Oxidative Stress via Engineering Its Global Regulator cAMP Receptor Protein (CRP). PLOS ONE 7, e51179.

Blair, J.M., Bavro, V.N., Ricci, V., Modi, N., Cacciotto, P., Kleinekathfer, U., Ruggerone, P., Vargiu, A.V., Baylay, A.J., Smith, H.E., et al. (2015). AcrB drug-binding pocket substitution confers clinically relevant resistance and altered substrate specificity. Proc Natl Acad Sci U S A 112, 3511-3516.

Blanchard, J.L., Wholey, W.Y., Conlon, E.M., and Pomposiello, P.J. (2007). Rapid changes in gene expression dynamics in response to superoxide reveal SoxRS-dependent and independent transcriptional networks. PLoS One 2, e1186.

Boos, W., and Shuman, H. (1998). Maltose/Maltodextrin System of Escherichia coli: Transport, Metabolism, and Regulation. Microbiol Mol Biol Rev 62, 204-229.

Burk, J., Weiche, B., Wenk, M., Boy, D., Nestel, S., Heimrich, B., and Koch, H.G. (2009). Depletion of the signal recognition particle receptor inactivates ribosomes in Escherichia coli. J Bacteriol 191, 7017-7026.

Cetin-Karaca, H., and Newman, M.C. (2015). Antimicrobial efficacy of plant phenolic compounds against Salmonella and Escherichia Coli. Food Bioscience 11, 8-16.

Chillappagari, S., Seubert, A., Trip, H., Kuipers, O.P., Marahiel, M.A., and Miethke, M. (2010). Copper Stress Affects Iron Homeostasis by Destabilizing Iron-Sulfur Cluster Formation in Bacillus subtilis. J Bacteriol 192, 2512-2524.

Chong, H., Huang, L., Yeow, J., Wang, I., Zhang, H., Song, H., and Jiang, R. (2013). Improving ethanol tolerance of Escherichia coli by rewiring its global regulator cAMP receptor protein (CRP). PLoS One 8, e57628.

Chubiz, L.M., Glekas, G.D., and Rao, C.V. (2012). Transcriptional cross talk within the mar-sox-rob regulon in Escherichia coli is limited to the rob and marRAB operons. J Bacteriol 194, 4867-4875.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6, 80-92.

Cornelis, P., Wei, Q., Andrews, S.C., and Vinckx, T. (2011). Iron homeostasis and management of oxidative stress response in bacteria. Metallomics 3, 540-549.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Molecular & cellular proteomics : MCP 13, 2513-2526.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26, 1367-1372.

Di Gioia, D., Luziatelli, F., Negroni, A., Ficca, A.G., Fava, F., and Ruzzi, M. (2011). Metabolic engineering of Pseudomonas fluorescens for the production of vanillin from ferulic acid. J Biotechnol 156, 309-316.

Ding, W., Si, M., Zhang, W., Zhang, Y., Chen, C., Zhang, L., Lu, Z., Chen, S., and Shen, X. (2015). Functional characterization of a vanillin dehydrogenase in Corynebacterium glutamicum. Sci Rep 5, 8044.

Djoko, K.Y., Phan, M.-D., Peters, K.M., Walker, M.J., Schembri, M.A., and McEwan, A.G. (2017). Interplay between tolerance mechanisms to copper and acid stress in Escherichia coli. Proceedings of the National Academy of Sciences 114, 6818-6823.

Down, T.A., Piipari, M., and Hubbard, T.J.P. (2011). Dalliance: interactive genome viewing on the web. Bioinformatics (Oxford, England) 27, 889-890.

Dragosits, M., and Mattanovich, D. (2013). Adaptive laboratory evolution – principles and applications for biotechnology. Microbial Cell Factories 12, 64.

Duckworth, H.W., Nguyen, N.T., Gao, Y., Donald, L.J., Maurus, R., Ayed, A., Bruneau, B., and Brayer, G.D. (2013). Enzyme-substrate complexes of allosteric citrate synthase: evidence for a novel intermediate in substrate binding. Biochim Biophys Acta 1834, 2546-2553.

Dunlop, M.J. (2011). Engineering microbes for tolerance to next-generation biofuels. Biotechnol Biofuels 4, 32.

Esterbauer, H., Schaur, R.J., and Zollner, H. (1991). Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. Free radical biology & medicine 11, 81-128.

Fache, M., Boutevin, B., and Caillol, S. (2015). Vanillin, a key-intermediate of biobased polymers. European Polymer Journal 68, 488-502.

Feehily, C., and Karatzas, K.A. (2013). Role of glutamate metabolism in bacterial responses towards acid and other stresses. J Appl Microbiol 114, 11-24.

Fitzgerald, D.J., Stratford, M., Gasson, M.J., Ueckert, J., Bos, A., and Narbad, A. (2004). Mode of antimicrobial action of vanillin against Escherichia coli, Lactobacillus plantarum and Listeria innocua. J Appl Microbiol 97, 104-113.

Fleige, C., Hansen, G., Kroll, J., and Steinbuchel, A. (2013). Investigation of the Amycolatopsis sp. strain ATCC 39116 vanillin dehydrogenase and its impact on the biotechnical production of vanillin. Appl Environ Microbiol 79, 81-90.

Forsberg, K.J., Patel, S., Witt, E., Wang, B., Ellison, T.D., and Dantas, G. (2015). Identification of Genes Conferring Tolerance to Lignocellulose-Derived Inhibitors by Functional Selections in Soil Metagenomes. Appl Environ Microbiol 82, 528-537.

Forsberg, K.J., Patel, S., Witt, E., Wang, B., Ellison, T.D., and Dantas, G. (2016). Identification of Genes Conferring Tolerance to Lignocellulose-Derived Inhibitors by Functional Selections in Soil Metagenomes. Appl Environ Microbiol 82, 528-537.

Furuya, T., Miura, M., Kuroiwa, M., and Kino, K. (2015). High-yield production of vanillin from ferulic acid by a coenzyme-independent decarboxylase/oxygenase two-stage process. N Biotechnol 32, 335-339.

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing, Vol 1207.

Ghosh, S., Sachan, A., Sen, S.K., and Mitra, A. (2007). Microbial transformation of ferulic acid to vanillic acid by Streptomyces sannanensis MTCC 6637. J Ind Microbiol Biotechnol 34, 131-138.

Graf, N., and Altenbuchner, J. (2014). Genetic engineering of Pseudomonas putida KT2440 for rapid and high-yield production of vanillin from ferulic acid. Appl Microbiol Biotechnol 98, 137-149.

Grimsrud, P.A., Xie, H., Griffin, T.J., and Bernlohr, D.A. (2008). Oxidative stress and covalent modification of protein with bioactive aldehydes. J Biol Chem 283, 21837-21841.

Große, C., Schleuder, G., Schmole, C., and Nies, D.H. (2014). Survival of Escherichia coli Cells on Solid Copper Surfaces Is Increased by Glutathione. Appl Environ Microbiol 80, 7071-7078.

Holden, M.J., Mayhew, M.P., Gallagher, D.T., and Vilker, V.L. (2002). Chorismate lyase: kinetics and engineering for stability. Biochim Biophys Acta 1594, 160-167.

Horinouchi, T., Sakai, A., Kotani, H., Tanabe, K., and Furusawa, C. (2017). Improvement of isopropanol tolerance of Escherichia coli using adaptive laboratory evolution and omics technologies. J Biotechnol 255, 47-56.

Horlacher, R., and Boos, W. (1997). Characterization of TreR, the major regulator of the Escherichia coli trehalose system. J Biol Chem 272, 13026-13032.

Huang, S.H., Wang, C.K., Peng, H.L., Wu, C.C., Chen, Y.T., Hong, Y.M., and Lin, C.T. (2012). Role of the small RNA RyhB in the Fur regulon in mediating the capsular polysaccharide biosynthesis and iron acquisition systems in Klebsiella pneumoniae. BMC microbiology 12, 148.

Hynes, M.J., and O'Coinceanainn, M.n. (2004). The kinetics and mechanisms of reactions of iron(III) with caffeic acid, chlorogenic acid, sinapic acid, ferulic acid and naringin. Journal of Inorganic Biochemistry 98, 1457-1464.

Jarboe, L.R. (2011). YqhD: a broad-substrate range aldehyde reductase with various applications in production of biorenewable fuels and chemicals. Appl Microbiol Biotechnol 89, 249-257.

Jonsson, L.J., Alriksson, B., and Nilvebrant, N.O. (2013). Bioconversion of lignocellulose: inhibitors and detoxification. Biotechnol Biofuels 6, 16.

Karlíčková, J., Macáková, K., Říha, M., Pinheiro, L.M.T., Filipský, T., Horňasová, V., Hrdina, R., and Mladěnka, P. (2015). Isoflavones Reduce Copper with Minimal Impact on Iron In Vitro. Oxidative Medicine and Cellular Longevity 2015.

Kaur, B., and Chakraborty, D. (2013). Biotechnological and molecular approaches for vanillin production: a review. Appl Biochem Biotechnol 169, 1353-1372.

Koma, D., Yamanaka, H., Moriyoshi, K., Ohmoto, T., and Sakai, K. (2012). Production of aromatic compounds by metabolically engineered Escherichia coli with an expanded shikimate pathway. Appl Environ Microbiol 78, 6203-6216.

Kunjapur, A.M., and Prather, K.L. (2015). Microbial engineering for aldehyde synthesis. Appl Environ Microbiol 81, 1892-1901.

Kunjapur, A.M., Tarasova, Y., and Prather, K.L. (2014a). Synthesis and accumulation of aromatic aldehydes in an engineered strain of Escherichia coli. J Am Chem Soc 136, 11644-11654.

Kunjapur, A.M., Tarasova, Y., and Prather, K.L.J. (2014b). Synthesis and Accumulation of Aromatic Aldehydes in an Engineered Strain of Escherichia coli. Journal of the American Chemical Society 136, 11644-11654.

Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome biology 15, R84.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, Vol 1303.

Liochev, S.I., and Fridovich, I. (1992). Fumarase C, the stable fumarase of Escherichia coli, is controlled by the soxRS regulon. Proceedings of the National Academy of Sciences 89, 5892-5896.

Ma, X.K., and Daugulis, A.J. (2014). Effect of bioconversion conditions on vanillin production by Amycolatopsis sp. ATCC 39116 through an analysis of competing by-product formation. Bioprocess Biosyst Eng 37, 891-899.

MacLean, M.J., Ness, L.S., Ferguson, G.P., and Booth, I.R. (1998). The role of glyoxalase I in the detoxification of methylglyoxal and in the activation of the KefB K+ efflux system in Escherichia coli. Mol Microbiol 27, 563-571.

Maddox, C.E., Laur, L.M., and Tian, L. (2010). Antibacterial activity of phenolic compounds against the phytopathogen Xylella fastidiosa. Curr Microbiol 60, 53-58.

Martínez-Cuesta, M.d.C., Payne, J., Hanniffy, S.B., Gasson, M.J., and Narbad, A. (2005). Functional analysis of the vanillin pathway in a vdh-negative mutant strain of Pseudomonas fluorescens AN103. Enzyme and Microbial Technology 37, 131-138.

Marzan, L.W., Barua, R., Akter, Y., Arifuzzaman, M., Islam, M.R., and Shimizu, K. (2017). A single metabolite production by Escherichia coli BW25113 and its pflA.cra mutant cultivated under microaerobic conditions using glycerol or glucose as a carbon source. Journal of Genetic Engineering and Biotechnology 15, 161-168.

McDermott, P.F., McMurry, L.M., Podglajen, I., Dzink-Fox, J.L., Schneiders, T., Draper, M.P., and Levy, S.B. (2008). The marC gene of Escherichia coli is not involved in multiple antibiotic resistance. Antimicrob Agents Chemother 52, 382-383.

Mensa, B., Kim, Y.H., Choi, S., Scott, R., Caputo, G.A., and DeGrado, W.F. (2011). Antibacterial Mechanism of Action of Arylamide Foldamers. Antimicrob Agents Chemother 55, 5043-5053.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat Protocols 8, 1551-1566.

Miller, E.N., Jarboe, L.R., Turner, P.C., Pharkya, P., Yomano, L.P., York, S.W., Nunn, D., Shanmugam, K.T., and Ingram, L.O. (2009a). Furfural inhibits growth by limiting sulfur assimilation in ethanologenic Escherichia coli strain LY180. Appl Environ Microbiol 75, 6132-6141.

Miller, E.N., Jarboe, L.R., Yomano, L.P., York, S.W., Shanmugam, K.T., and Ingram, L.O. (2009b). Silencing of NADPH-dependent oxidoreductase genes (yqhD and dkgA) in furfural-resistant ethanologenic Escherichia coli. Appl Environ Microbiol 75, 4315-4323.

Minty, J.J., Lesnefsky, A.A., Lin, F., Chen, Y., Zaroff, T.A., Veloso, A.B., Xie, B., McConnell, C.A., Ward, R.J., Schwartz, D.R., et al. (2011a). Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in Escherichia coli. Microbial Cell Factories 10, 18.

Minty, J.J., Lesnefsky, A.A., Lin, F., Chen, Y., Zaroff, T.A., Veloso, A.B., Xie, B., McConnell, C.A., Ward, R.J., Schwartz, D.R., et al. (2011b). Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in Escherichia coli. Microb Cell Fact 10, 18.

Nguyen, T.T., Iwaki, A., Ohya, Y., and Izawa, S. (2014). Vanillin causes the activation of Yap1 and mitochondrial fragmentation in Saccharomyces cerevisiae. J Biosci Bioeng 117, 33-38.

Odoux, E., and Grisoni, M. (2010). Vanilla, 1st edn (CRC Press).

Outten, F.W., Outten, C.E., Hale, J., and O'Halloran, T.V. (2000). Transcriptional activation of an Escherichia coli copper efflux regulon by the chromosomal MerR homologue, cueR. J Biol Chem 275, 31024-31029.

Pereira, D.S., Donald, L.J., Hosfield, D.J., and Duckworth, H.W. (1994). Active site mutants of Escherichia coli citrate synthase. Effects of mutations on catalytic and allosteric properties. J Biol Chem 269, 412-417.

Plaggenborg, R., Overhage, J., Loos, A., Archer, J.A., Lessard, P., Sinskey, A.J., Steinbuchel, A., and Priefert, H. (2006). Potential of Rhodococcus strains for biotechnological vanillin production from ferulic acid and eugenol. Appl Microbiol Biotechnol 72, 745-755.

Price, N.L., and Raivio, T.L. (2009). Characterization of the Cpx Regulon in Escherichia coli Strain MC4100. J Bacteriol 191, 1798-1815.

Pugh, S., McKenna, R., Halloum, I., and Nielsen, D.R. (2015). Engineering Escherichia coli for renewable benzyl alcohol production. Metabolic Engineering Communications 2, 39-45.

Quandt, E.M., Gollihar, J., Blount, Z.D., Ellington, A.D., Georgiou, G., and Barrick, J.E. (2015). Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. eLife 4.

Rademacher, C., and Masepohl, B. (2012). Copper-responsive gene regulation in bacteria. Microbiology 158, 2451-2464.

Renilla, S., Bernal, V., Fuhrer, T., Castano-Cerezo, S., Pastor, J.M., Iborra, J.L., Sauer, U., and Canovas, M. (2012). Acetate scavenging activity in Escherichia coli: interplay of acetyl-CoA synthetase and the PEP-glyoxylate cycle in chemostat cultures. Appl Microbiol Biotechnol 93, 2109-2124.

Rodriguez-Perez, C., Quirantes-Pine, R., Uberos, J., Jimenez-Sanchez, C., Pena, A., and Segura-Carretero, A. (2016). Antibacterial activity of isolated phenolic compounds from cranberry (Vaccinium macrocarpon) against Escherichia coli. Food Funct 7, 1564-1573.

Rosenberg, E.Y., Bertenthal, D., Nilles, M.L., Bertrand, K.P., and Nikaido, H. (2003). Bile salts and fatty acids induce the expression of Escherichia coli AcrAB multidrug efflux pump through their interaction with Rob regulatory protein. Mol Microbiol 48, 1609-1619.

Ruiz, C., and Levy, S.B. (2014). Regulation of acrAB expression by cellular metabolites in Escherichia coli. J Antimicrob Chemother 69, 390-399.

Rutherford, B.J., Dahl, R.H., Price, R.E., Szmidt, H.L., Benke, P.I., Mukhopadhyay, A., and Keasling, J.D. (2010). Functional genomic study of exogenous n-butanol stress in Escherichia coli. Appl Environ Microbiol 76, 1935-1945.

Sanhueza, L., Melo, R., Montero, R., Maisey, K., Mendoza, L., and Wilkens, M. (2017). Synergistic interactions between phenolic compounds identified in grape pomace extract with antibiotics of different classes against Staphylococcus aureus and Escherichia coli. PLoS One 12, e0172273.

Sardessai, Y., and Bhosle, S. (2002). Tolerance of bacteria to organic solvents. Research in microbiology 153, 263-268.

Schrey, S.D., Erkenbrack, E., Früh, E., Fengler, S., Hommel, K., Horlacher, N., Schulz, D., Ecke, M., Kulik, A., Fiedler, H.-P., et al. (2012). Production of fungal and bacterial growth modulating secondary metabolites is widespread among mycorrhiza-associated streptomycetes. BMC Microbiology 12, 164.

Seemann, T. (2015). Snippy: fast bacterial variant calling from NGS reads. (https://github.com/tseemann/snippy).

Shen, C.R., and Liao, J.C. (2008). Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways. Metab Eng 10, 312-320.

Shlar, I., Droby, S., and Rodov, V. (2017). Modes of antibacterial action of curcumin under dark and light conditions: A toxicoproteomics approach. J Proteomics 160, 8-20.

Simon, O., Klaiber, I., Huber, A., and Pfannstiel, J. (2014). Comprehensive proteome analysis of the response of Pseudomonas putida KT2440 to the flavor compound vanillin. J Proteomics 109, 212-227.

Singh, N.P., and Khan, A. (1995). Acetaldehyde: genotoxicity and cytotoxicity in human lymphocytes. Mutation research 337, 9-17.

Singh, R., Mailloux, R.J., Puiseux-Dao, S., and Appanna, V.D. (2007). Oxidative stress evokes a metabolic adaptation that favors increased NADPH synthesis and decreased NADH production in Pseudomonas fluorescens. J Bacteriol 189, 6665-6675.

Soballe, B., and Poole, R.K. (2000). Ubiquinone limits oxidative stress in Escherichia coli. Microbiology 146 (Pt 4), 787-796.

Spaans, S.K., Weusthuis, R.A., van der Oost, J., and Kengen, S.W. (2015). NADPH-generating systems in bacteria and archaea. Front Microbiol 6, 742.

Stokell, D.J., Donald, L.J., Maurus, R., Nguyen, N.T., Sadler, G., Choudhary, K., Hultin, P.G., Brayer, G.D., and Duckworth, H.W. (2003). Probing the roles of key residues in the unique regulatory NADH binding site of type II citrate synthase of Escherichia coli. J Biol Chem 278, 35435-35443.

Tai, A., Sawano, T., Yazama, F., and Ito, H. (2011). Evaluation of antioxidant activity of vanillin by using multiple antioxidant assays. Biochim Biophys Acta 1810, 170-177.

Tan, G., Cheng, Z., Pang, Y., Landry, A.P., Li, J., Lu, J., and Ding, H. (2014). Copper binding in IscA inhibits iron-sulfur cluster assembly in Escherichia coli. Mol Microbiol 93, 629-644.

Tyagi, P., Singh, M., Kumari, H., Kumari, A., and Mukhopadhyay, K. (2015). Bactericidal activity of curcumin I is associated with damaging of bacterial membrane. PLoS One 10, e0121313.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nature methods 13, 731-740.

Van Dyk, T.K., Templeton, L.J., Cantera, K.A., Sharpe, P.L., and Sariaslani, F.S. (2004). Characterization of the Escherichia coli AaeAB Efflux Pump: a Metabolic Relief Valve? J Bacteriol 186, 7196-7204.

Varghese, S., Tang, Y., and Imlay, J.A. (2003). Contrasting Sensitivities of Escherichia coli Aconitases A and B to Oxidation and Iron Depletion. J Bacteriol 185, 221-230.

Vlasblom, J., Zuberi, K., Rodriguez, H., Arnold, R., Gagarinova, A., Deineko, V., Kumar, A., Leung, E., Rizzolo, K., Samanfar, B., et al. (2015). Novel function discovery with GeneMANIA: a new integrated resource for gene function prediction in Escherichia coli. Bioinformatics (Oxford, England) 31, 306-310.

Walton, N.J., Mayer, M.J., and Narbad, A. (2003). Vanillin. Phytochemistry 63, 505-515.

Wang, C., Yang, L., Shah, A.A., Choi, E.S., and Kim, S.W. (2015). Dynamic interplay of multidrug transporters with TolC for isoprenol tolerance in Escherichia coli. Sci Rep 5.

Wang, X., Yomano, L.P., Lee, J.Y., York, S.W., Zheng, H., Mullinnix, M.T., Shanmugam, K.T., and Ingram, L.O. (2013). Engineering furfural tolerance in Escherichia coli improves the fermentation of lignocellulosic sugars into renewable chemicals. Proc Natl Acad Sci U S A 110, 4021-4026.

Webber, M.A., Bailey, A.M., Blair, J.M., Morgan, E., Stevens, M.P., Hinton, J.C., Ivens, A., Wain, J., and Piddock, L.J. (2009). The global consequence of disruption of the AcrAB-TolC efflux pump in Salmonella enterica includes reduced expression of SPI-1 and other attributes required to infect the host. J Bacteriol 191, 4276-4285.

Yamada, M., Asaoka, S., Saier, M.H., Jr., and Yamada, Y. (1993). Characterization of the gcd gene from Escherichia coli K-12 W3110 and regulation of its expression. J Bacteriol 175, 568-571.

Zaldivar, J., Martinez, A., and Ingram, L.O. (1999). Effect of selected aldehydes on the growth and fermentation of ethanologenic Escherichia coli. Biotechnol Bioeng 65, 24-33.

Zhang, F., Qian, X., Si, H., Xu, G., Han, R., and Ni, Y. (2015). Significantly improved solvent tolerance of Escherichia coli by global transcription machinery engineering. Microb Cell Fact 14, 175.

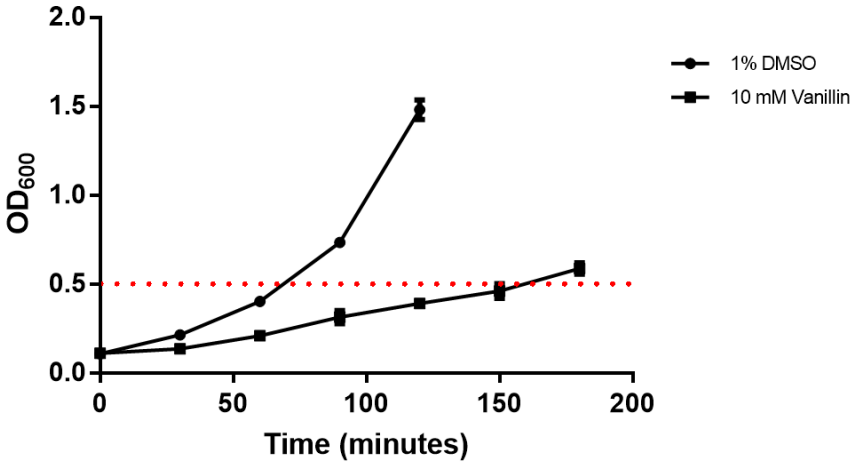
Zhou, J., Wang, K., Xu, S., Wu, J., Liu, P., Du, G., Li, J., and Chen, J. (2015). Identification of membrane proteins associated with phenylpropanoid tolerance and transport in Escherichia coli BL21. J Proteomics 113, 15-28.

4.8. Supplementary Tables and Figures

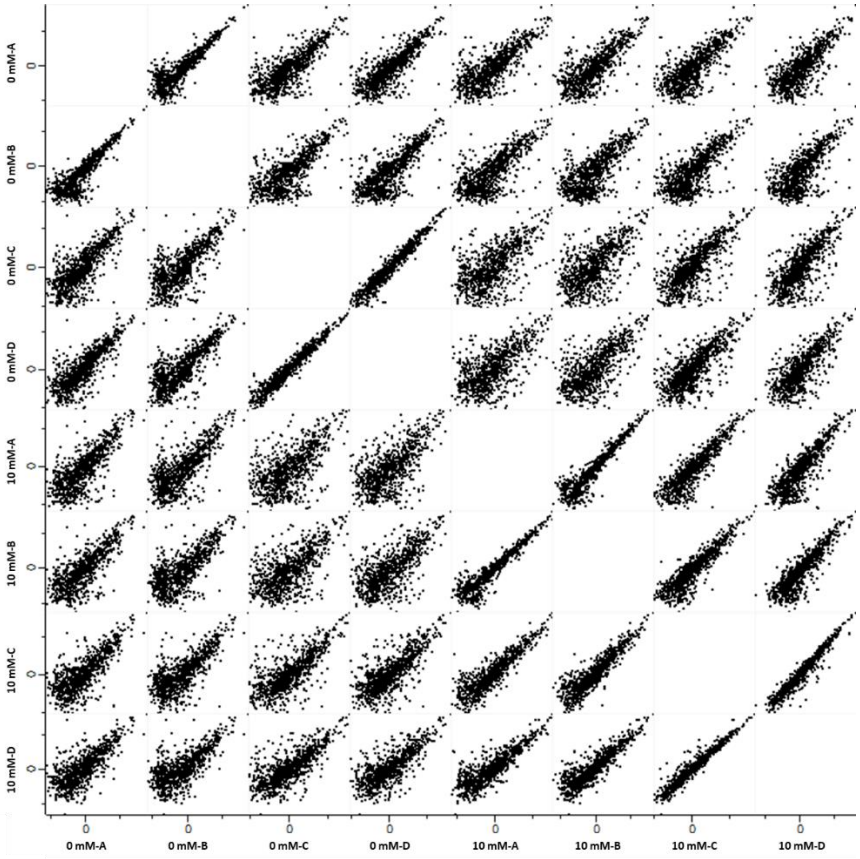
S. Fig 1: Proteins identified from label-free shotgun proteomic analysis by LC-MS/MS. The LC-MS/MS raw data listing the complete set of 1885 proteins identified by peptide detection is provided in a Microsoft Excel spreadsheet on the CD at the back of this thesis.

S. Fig 2: Dataset of the 850 proteins that were reproducibly quantifiable. The list of 850 proteins that were reproducibly quantifiable with their raw, calculated and normalised LFQ intensities are included on the CD provided. In the “Imputed” tab, highlighted cells indicate imputed values. In the “Summary” tab, highlighted cells and a ‘+’ indicate quantification of that protein was significant using that column’s parameters.

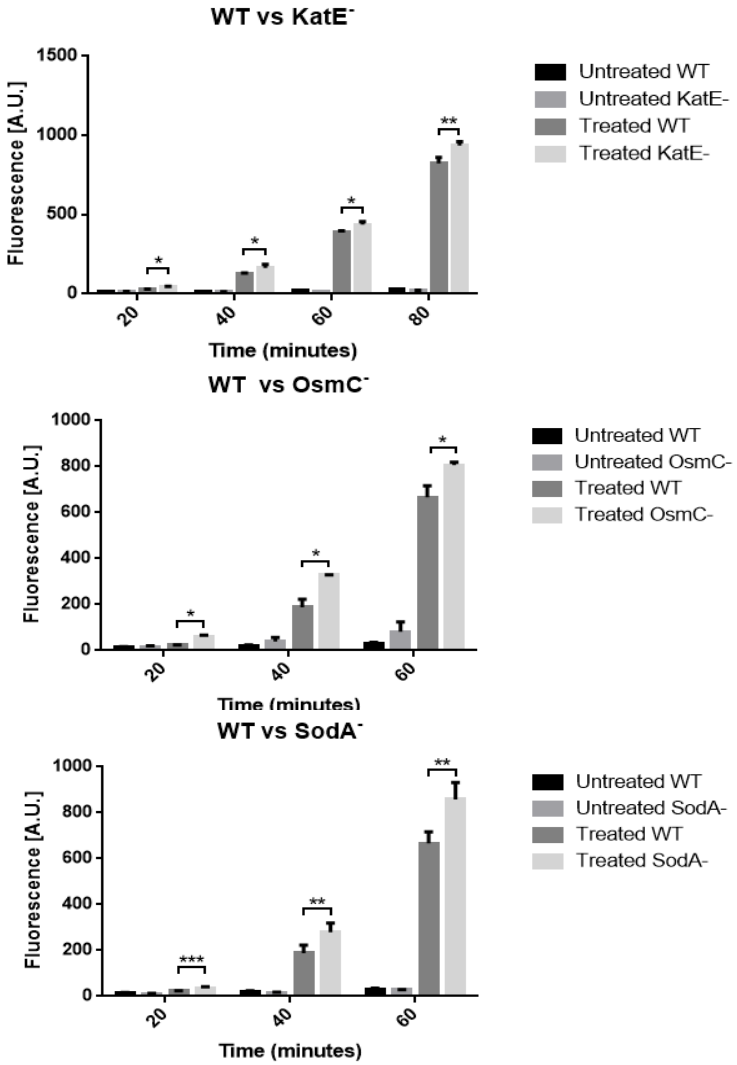
S. Fig 3: Growth curve of *E. coli* BW25113 with vanillin showing at what point cells were harvested for proteomics. *E. coli* cells were harvested at an equivalent OD_{600} of 0.5 for both the control and treated samples, rather than after an equivalent time period.



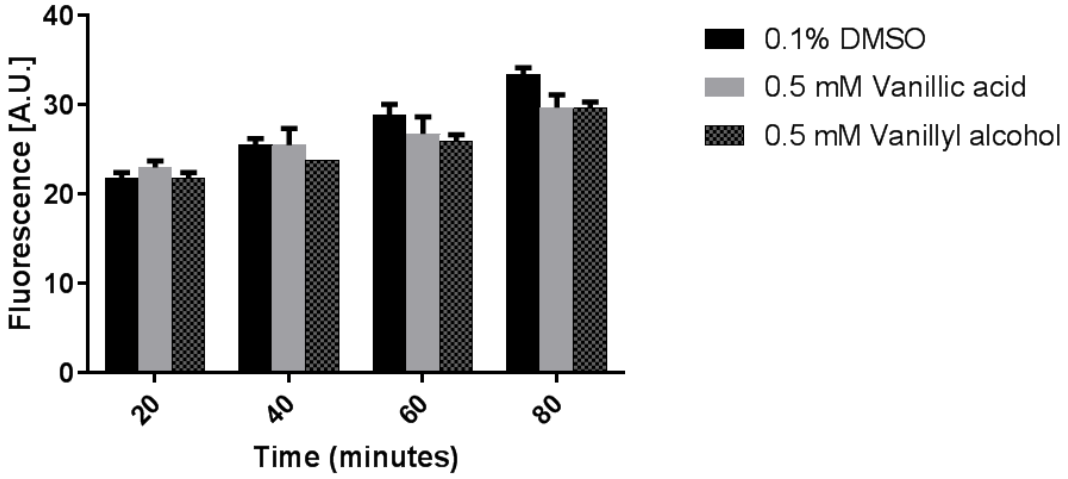
S. Fig 4: Multiscatter graphs of proteomic data replicates. Label-free quantification (LFQ) intensities calculated from the proteomic analysis of each sample are plotted against each other.



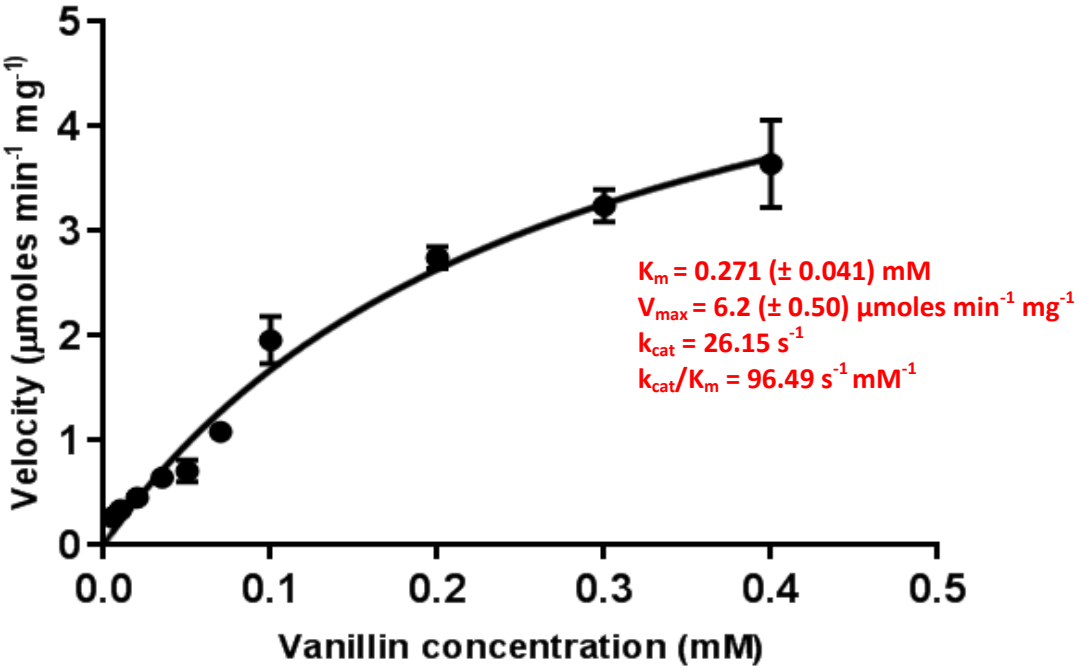
S. Fig 5: Knock-out of key oxidative stress defence genes exacerbates ROS production. The increase in **e**ndogenous reactive-oxygen species (ROS) levels as a result of vanillin treatment was exacerbated in *E. coli* mutants lacking genes that contribute to reducing oxidative-stress defence compared to wildtype cells. ROS levels were detected by use of the ROS-activated fluorescent dye 2', 7'-dihydrodichlorofluorescein diacetate (H_2DCFDA). Data are plotted as means of three biological replicates with standard deviations as error bars. *, ** and *** denote statistical p-values of 0.01 – 0.05, 0.001 – 0.01, and < 0.001, respectively.



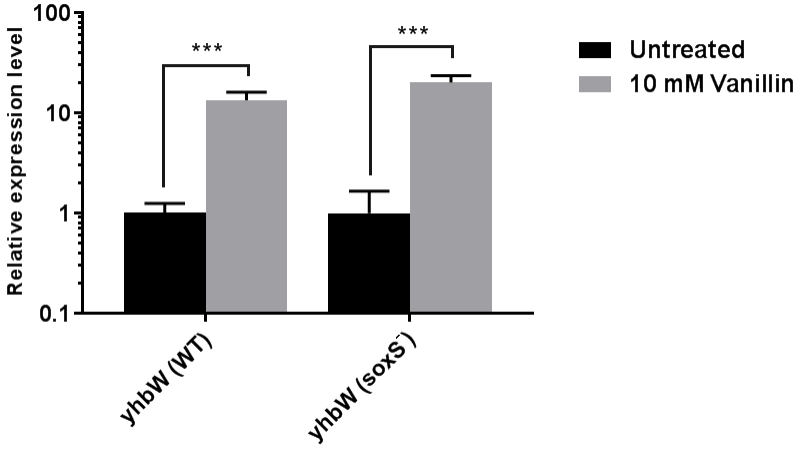
S. Fig 6: Vanillic acid and vanillyl alcohol do not result in endogenous ROS production in *E. coli*. Wildtype *E. coli* cells exhibited no significant increase in endogenous ROS levels on treatment with vanillic acid or vanillyl alcohol at concentrations equivalent to that used with vanillin. ROS levels were detected by use of the ROS-activated fluorescent dye 2', 7'-dihydrodichlorofluorescein diacetate (H_2DCFDA). Data are plotted as means of three biological replicates with standard deviations as error bars.



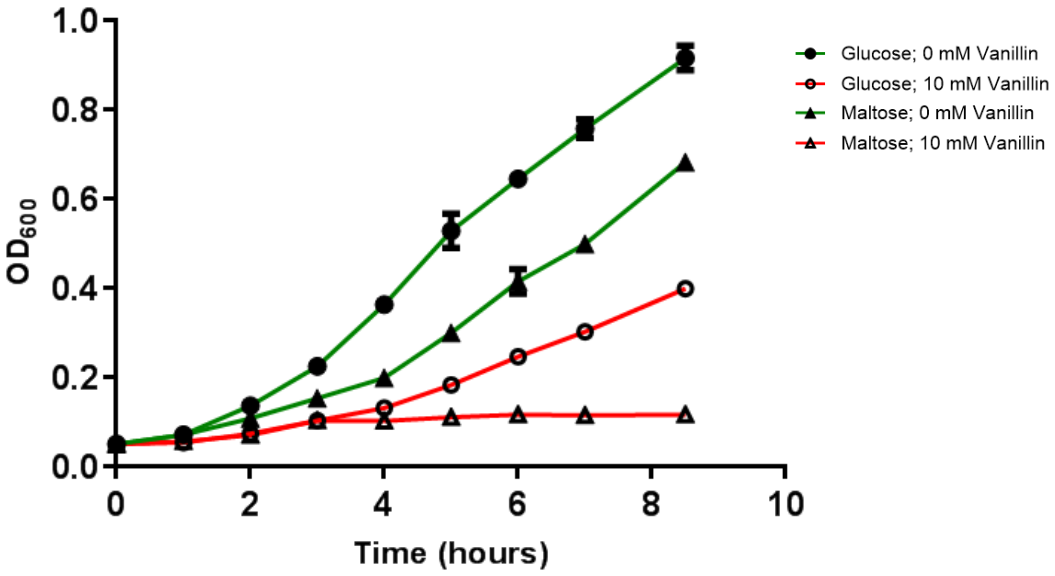
S. Fig 7: Vanillin reductase activity shown by YqhD. YqhD activity with vanillin as the substrate was determined by measurement of NADPH oxidation at 340 nm at vanillin concentrations ranging from 5 µM to 0.4 mM, in triplicate at each concentration. Enzyme kinetics were determined according to Michaelis-Menten kinetics using GraphPad Prism 7 software.



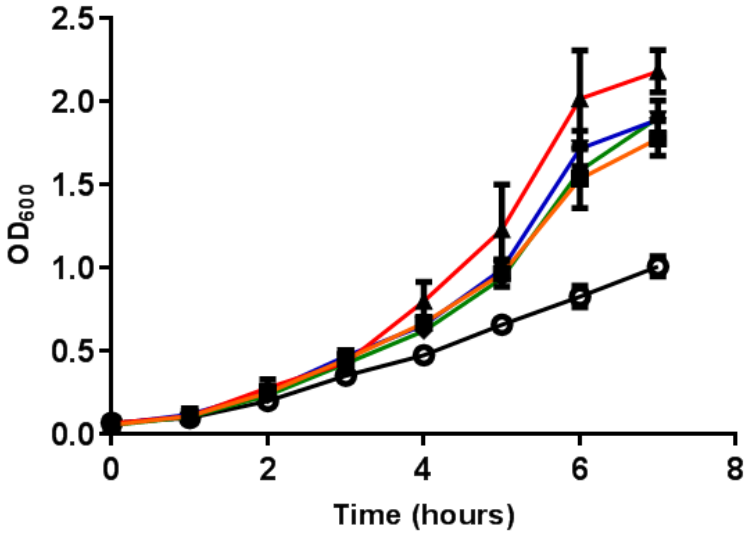
S. Fig 8: Expression level of the gene encoding the uncharacterised protein YhbW. RT-PCR assessed expression of the *yhbW* gene in both wildtype and *ΔsoxS* mutant cells grown in the presence of 10 mM vanillin relative to untreated cells. Expression was normailised to that of the *rrsA* housekeeping gene. Data are plotted as means of three biological replicates (each consisting of three technical replicates) with standard deviations as error bars. *, ** and *** denote statistical p-values of 0.01 – 0.05, 0.001 – 0.01, and < 0.001, respectively



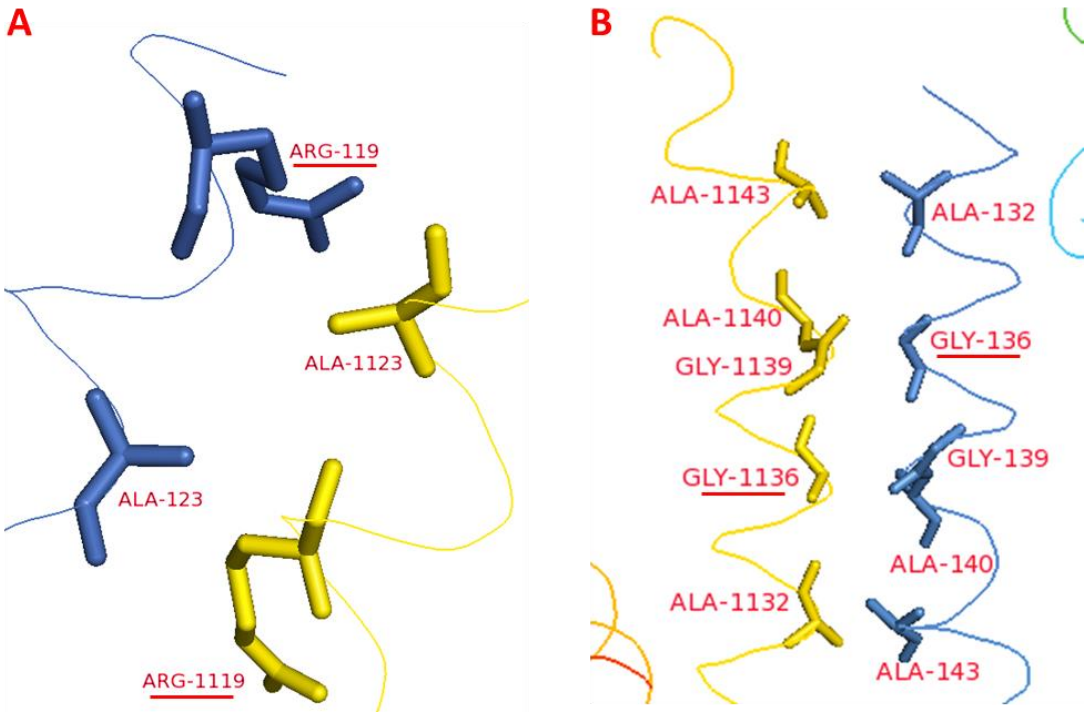
S. Fig 9: Sensitivity of *E. coli* grown in minimal media to vanillin. *E. coli* BW25113 cells were grown in M9 minimal media with either glucose or maltose as the sole carbon source, with or without 10 mM vanillin. Growth on maltose was reduced relative to growth on glucose, but completely inhibited in the presence of vanillin. Data plotted are the mean of three cultures with the standard deviation shown as error bars.



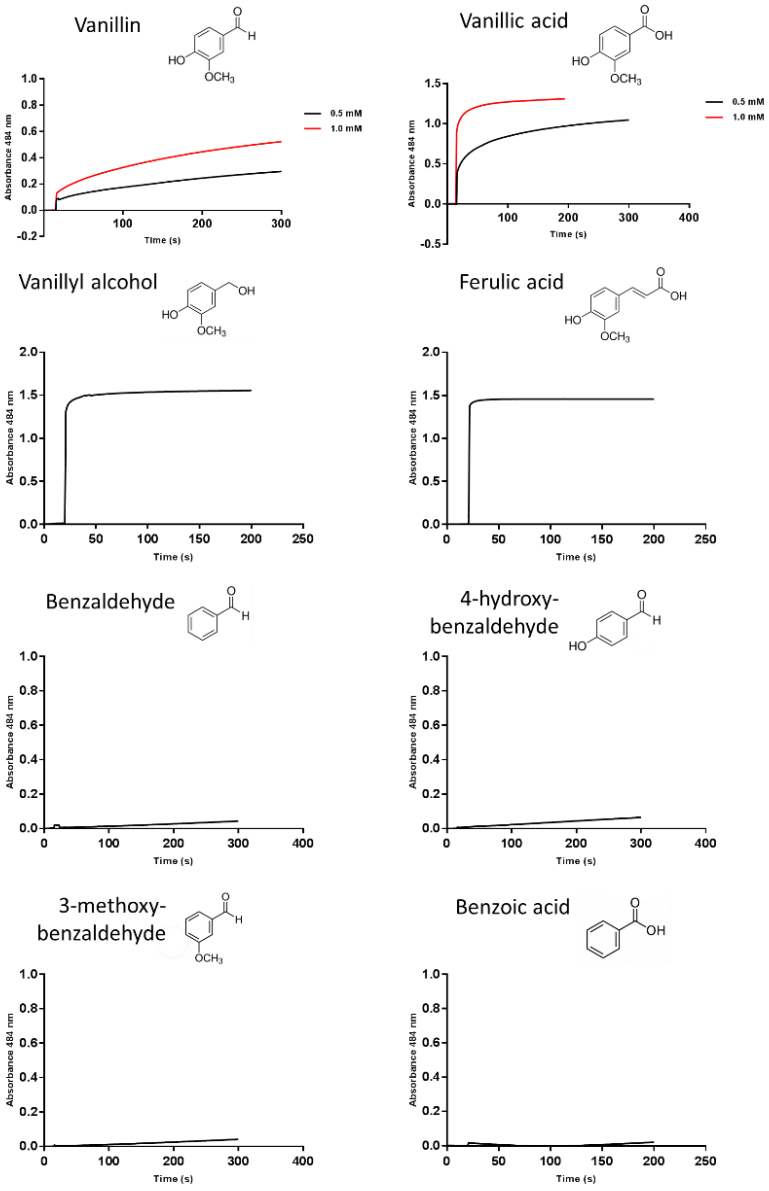
S. Fig 10: Increased vanillin tolerance shown by evolved *E. coli* strains. *E. coli* strains were grown in LB media with 10 mM vanillin for 7 h with their cell density monitored. The black line is wildtype, the red line is EVT1, the blue line EVT2, the green line EVT3 and the orange line EVT4. Data plotted are the mean values of cultures grown in triplicate with the standard deviation shown as error bars.



S. Fig 11: Location of mutations in GltA in EVT strains. Analysis of the structure of the *E. coli* citrate synthase GltA (PDB entry 4JAD) in PyMOL shows two of the EVT strains had mutations in GltA residues that occur at the interface between monomers (Duckworth et al., 2013). A) Arg119 is substituted for a phenylalanine in EVT1, possibly preventing interaction with Ala123 of the adjacent monomer. B) Gly136 is substituted for a serine in EVT3. The addition of the polar side chain of the serine to a hydrophobic region could disrupt hexamer formation. Protein structure is PDB entry 4JAD shown in PyMOL. The blue and yellow coloured amino acid ribbons represent separate monomers.



S. Fig 12: Copper reduction activities exhibited by vanillin and structurally-related compounds. The Cu^{+}-specific BCS was used to assess the reductive effect of several aromatic compounds structurally related to vanillin *in vitro*. Activity was measured with compounds of interest at a final concentration of 0.5 mM in all cases, except with vanillin and vanillic acid where activity was also measured at 1 mM showing concentration-dependent reduction. The structures of the aromatic compounds assays are included to highlight the effect the difference in chemical structure has on copper reduction.



S. Fig 13: Primers used in this study.

Primer	Sequence 5’ → 3’
aaeA_RTPCR_F	GATGGTGAACAGTATCTGTCCT
aaeA_RTPCR_R	GTCGTATTAGTCATTCTGGCC
aaeB_RTPCR_F	GCAATACAGCCAATAAATGTG
aaeB_RTPCR_R	TGCTAACCAACATATTCGCT
acrA_RTPCR_F	TAGTAACAGTCAAAACTGAACC
acrA_RTPCR_R	TTTCGCACTGTCGTATGT
acrB_RTPCR_F	TAATTTCTTTATCGATCGCC
acrB_RTPCR_R	GTCACTGTTAGAGGACATGTA
acrD_RTCPR_F	TCTTTATTGATCGCCCCATT
acrD_RTPCR_R	TATCGAGGCCGGTCATATTT
acrF_RTPCR_F	TATTCGACGACCGATATTTGC
acrF_RTPCR_R	TGTTCGATAACCTGCGTCAC
aldB_OE_F	ATTATCATATGACCAATAATCCCCCTTCAGCA
aldB_OE_R	ATTATAAGCTTGAACAGCCCCAACGGTTTATC
copA_RTPCR_F	AAACTATCGACCTGACCCT
copA_RTPCR_R	TTTGGGTGGCTTACAGAT
marA_RTPCR_F	CAATACTGACGCTATTACCA
marA_RTPCR_R	GGTTTCTTTTTTAAACATCC
marR_RTPCR_F	GTTCAATGAAATTATTCCATT
marR_RTPCR_R	CAGTTCAACCGGAGTAATAC
soxS_RTPCR_F	GCATATTGACCAGCCGCTTAA
soxS_RTPCR_R	TTACAGGCGGTGGCGATAATCGCT
yhbW_OE_F	ATTATGCTAGCACTGATAAAACCATTGCGTTTT
yhbW_OE_R	ATTATCTCGAGTCCCAACAACTCTTCCTTAAC
yhbW_RTPCR_F	GGACTGATAAAACCATTGCGT
yhbW_RTPCR_R	AGATGCAGCGTGGTGGTAT
yqhD_OE_F	ATATTAGCTAGCAACAACTTTAATCTGCACACCCCA
yqhD_OE_R	ATATTACTCGAGGCGGGCGGCTTCGTATAT

5. Exploration of the function and role of TRAP transport systems in *Rhodopseudomonas palustris*

5.1. Introduction

The photosynthetic purple bacterium *Rhodopseudomonas palustris* has been shown to have an extremely versatile metabolism allowing it to grow in a wide range of environments. Central to this is the expansive repertoire of transport systems it possesses that make up ~15% of its genome, allowing it to be proficient at the uptake of essential substrates for its metabolism. The ATP-binding cassette (ABC) and tripartite ATP-independent periplasmic (TRAP) transport systems both employ high-affinity extracytoplasmic solute-binding proteins (SBPs) to facilitate the scavenging and uptake of substrates at low concentrations, but are driven by contrasting energy-coupling methods (Section 1.2). Multiple studies have been conducted where these *R. palustris* SBPs have been recombinantly produced and purified for biochemical characterisation to determine the substrates of the transport systems. This includes high-throughput screening of a wide range of ABC SBPs using fluorescence-based thermal shift assays and thermodynamic characterisation of aromatic-compound uptake systems, with multiple crystal structures of *R. palustris* SBPs solved and deposited in the PDB (Giuliani et al., 2011; Pietri et al., 2012; Salmon et al., 2013). As shown with the characterisation of CouP and TarP, an ABC and a TRAP SBP that both bind the same hydroxycinnamates with K_d values in the same order of magnitude, *R. palustris* retains a level of redundancy in its transport capabilities.

These studies have been vital to the understanding of *R. palustris*'s physiology, highlighting the diverse compounds the bacterium can uptake and utilise, which contributes to a competitive ecological advantage over other microbes in the environments it has been isolated from. Furthermore, the characterisation of the hydroxycinnamate uptake systems highlights the contribution of transport in *R. palustris* to industrial biotechnology (as discussed in Chapter 3).

As the work of Giulani et al. (2011) focused on identifying the substrates of *R. palustris*'s ABC transport systems the initial focus of the current study was on the relatively neglected TRAP transport systems (Giuliani et al., 2011). There are eight TRAP systems annotated in the *R. palustris* genome as determined by Larimer et al. (2004), including the aforementioned TarPQM system (Chapter 3) that has been previously characterised

(Larimer et al., 2004; Salmon et al., 2013). The SBPs of two other TRAP systems, RPA2543 and RPA2782 have been partially characterised in the alternative strains of *R. palustris* BisB18 and HaA2, respectively; a crystal structure of the RPA2782 homolog has been solved showing binding of 2-acetolactate (PDB entry 4OAN). The genomic locations of the five remaining uncharacterised systems investigated in this chapter are displayed in Figure 52. Three of the SBPs are members of the DctP-family, RPA1975, RPA2047 and RPA4556; and two are members of the TAXI-family of TRAP SBPs, RPA3458 and RPA4510, and these latter two systems consist of a fused QM polypeptide domain. Gene clusters that share a related overall function are frequently found in *R. palustris*, with the cluster containing the TarPQM and CouPSTU transporters a prime example of this. Therefore the annotated function of the adjacent gene products in the *R. palustris* genomic database could potentially be used to predict the substrate of the TRAP transporters being investigated. However, Figure 52 highlights that the *R. palustris* genome sequence is still poorly annotated with a high frequency of gene products of unknown function and hypothetical proteins.

The aim of this chapter was to purify the SBPs of these five TRAP transport systems for biochemical characterisation of their binding properties, and assess the overall physiological role and importance of the transporters in *R. palustris*.

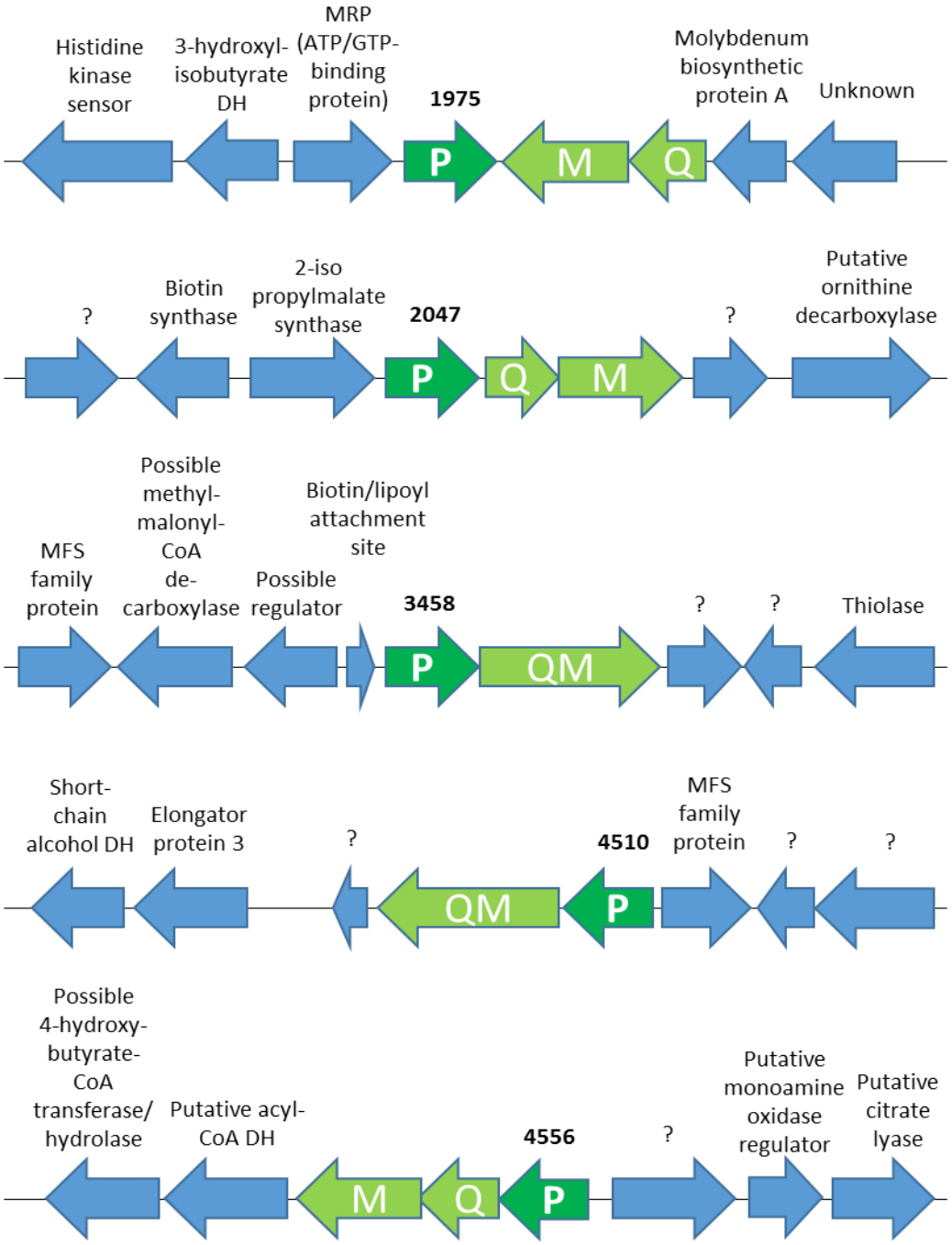


Figure 52: Uncharacterised TRAP transport systems in *R. palustris*. The genes encoding the SBP (P), and the transmembrane (Q and M) components of five TRAP transport systems that had not previously been characterised are shown in green, with the neighbouring genes shown in blue. RPA1975, RPA2047 and RPA4556 are DctP-family proteins, and RPA3458 and RPA4510 are TAXI-family proteins and their respective transmembrane domains, and the genes encoding them, are fused as shown. DH = dehydrogenase; MRP = multidrug resistance protein; ? = unknown function, or putative protein.

5.2. Results

5.2.1. Initial overproduction and purification of RPA1975

Primers were designed for the PCR amplification of the *rpa1975* gene from *R. palustris* CGA009 genomic DNA, excluding the nucleotide sequence that encodes the N-terminal signal sequence as predicted by the SignalP 4.1 program (Petersen et al., 2011). Two pairs of primers were used; 1975_21a_F and 1975_21a_R for the amplification of *rpa1975* with *Nde*I and *Not*I restriction digestion cut sites for ligation into the pET21a(+) vector, and 1975_21b_F and 1975_21b_R with *Nco*I and *Not*I cut sites for ligation into the pET22b(+) vector. pET21a:*1975* was used for cytoplasmic production of RPA1975, and pET22b:*1975* was used for periplasmic production, harnessing the vector's *pelB* signal sequence for periplasmic localisation. The amino acid sequence shows RPA1975 contains 4 cysteine residues indicating two potential disulphide bonds that could cause problems with solubility during overproduction; the oxidising environment of the periplasm and the presence of two periplasmic foldases, DsbA and DsbC, facilitates disulphide bond formation. The *rpa1975 s*top codon was not included to allow for production of the incorporated C-terminal 6 x histidine tag. Completed constructs were transformed into *E. coli* BL21, with cloning confirmed through colony PCR screening and subsequent DNA sequencing of the constructs.

Protein production and protein solubility trials were conducted and assessed by SDS-PAGE using the whole-cells and the soluble fraction of the CFEs, respectively, with both pET21a(+) and pET22b(+). Throughout trials using *E. coli* BL21(DE3) the yield of the protein production appeared high, but the protein continuously proved largely insoluble. Figure 53 displays the results of cytoplasmic production with a large amount of RPA1975 produced and identified using the whole-cell samples, but only small amounts visible in the soluble fractions; lowering the temperature did not alleviate this problem. The use of pET22b(+) produced similar results as shown in Figure 54.

Due to the solubility problems an alternative approach was pursued by using the *E. coli* protein production strain Origami B. The *E. coli* Origami B has two mutations, in the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, which facilitates cytoplasmic disulphide bond formation. The pET21a(+) vector was utilised as before, but low solubility remained. Therefore, the pET22b(+) construct was transformed into *E. coli* Origami B resulting in an oxidative cytoplasm combined with periplasmic targeting of the produced RPA1975. This approach was successful and resulted in a higher yield of soluble

protein, at both 25 °C and 37 °C, as can be seen in Figure 55. This strain was subsequently used for larger-scale protein production, with RPA1975 purified from the cell-free extract of 1 l of cells via nickel-affinity chromatography, with a yield of 2.6 mg l^{-1} pure protein. The protein sample was subjected to urea treatment and concentrated for use in assays.

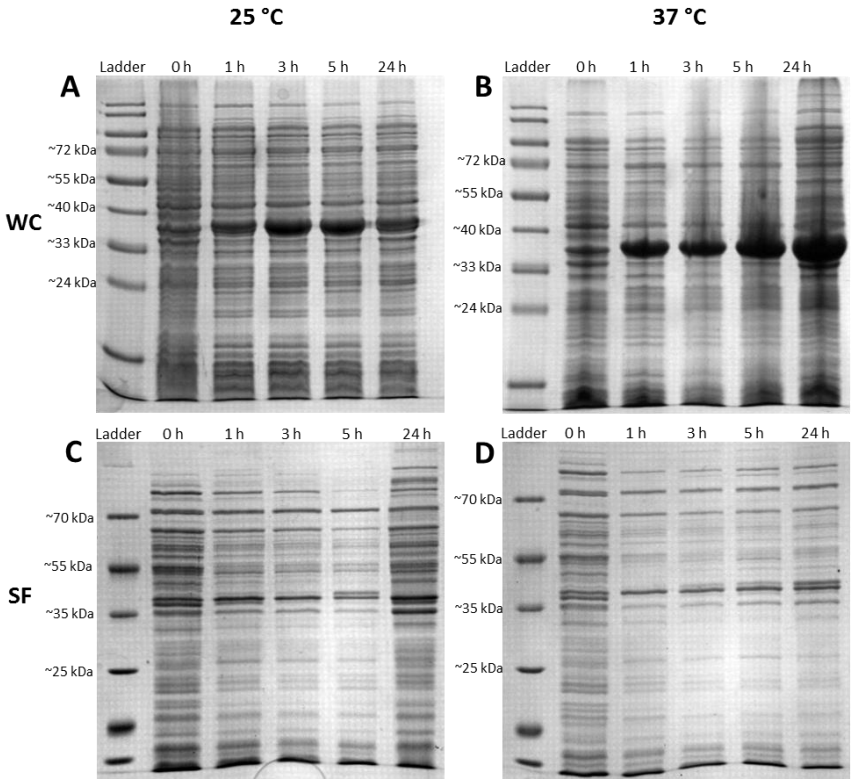


Figure 53: SDS-PAGE showing attempted production of RPA1975 using pET21a(+) in *E. coli* BL21. A+B) Whole-cell (WC) samples showing cytoplasmic production of RPA1975 at 25 °C and 37 °C, respectively, showing a high yield of protein as indicated by the large band at ~38 kDa. C+D) Soluble fraction (SF) samples of the same cultures showing the low solubility of the produced protein, with only a faint band visible at ~38 kDa. The ladder for A and B is shown on page 257; the ladder for C and D is shown on page 256.

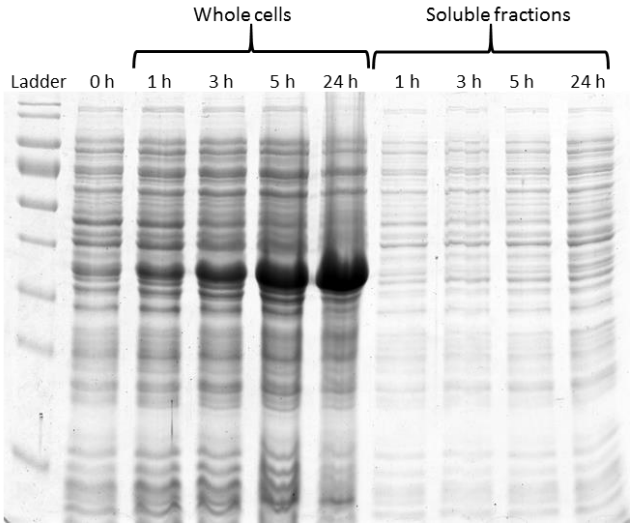


Figure 54: SDS-PAGE showing attempted production of RPA1975 using pET22b(+) in *E. coli* BL21(DE3). The addition of the N-terminal pelB leader sequence for periplasmic targeting of the recombinantly produced protein was trialed to improve the solubility of RPA1975; the large band at ~38 kDa in the whole cell sample lanes shows production yield remained high, while the lack of a significant equivalent band in the soluble fractions shows this approach had not alleviated the problem. The ladder used is shown on 255.

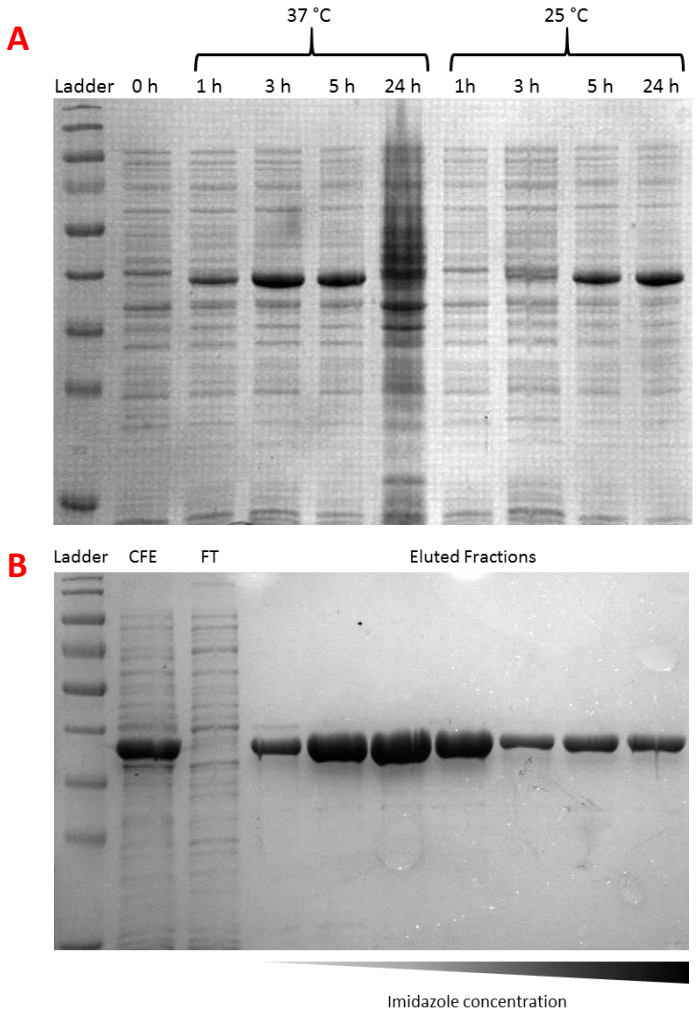


Figure 55: SDS-PAGE showing RPA1975 production using the pET22b(+) vector in *E. coli* Origami B(DE3). A) SDS-PAGE of soluble factions from protein production trials shows a clearly distinguishable band of soluble protein at both growth temperatures from 1 h onwards. The largest band was seen with 3 hours of production at 37 °C suggesting this was the optimal condition for large-scale production for purification. The ladder used shown on page 257. B) SDS-PAGE showing the purification by nickel-affinity chromatography of RPA1975, with the cell-free extract (CFE) lane showing a large band at ~38 kDa highlighting the presence of soluble RPA1975 prior to its binding to the column; the flow-through (FT) lane shows the proteins in the CFE that did not bind to the column, with the prominent band absent. An imidazole gradient was used to elute the bound protein from the column, and the eluted fractions that contained a significant amount of protein, as indicated by an increase in A280, are shown, highlighting the high yield and purity of the RPA1975 protein. The ladder used is shown on page 255.

5.2.2. Ligand screening of RPA1975 by thermal shift assays

The purified RPA1975 was initially screened using thermal shift (or thermofluor) assays, which involves denaturing the protein using an increasing temperature gradient. Denaturation is quantified as the protein's melting temp (T_m), and determined by the presence of a fluorescent dye, SYPRO Orange, which fluoresces on binding of hydrophobic surfaces non-specifically. As the protein denatures it unfolds, revealing its hydrophobic surfaces for the dye to bind and the fluorescence increases allowing for plotting of a melting curve as see in Figure 56, with the T_m defined as the midpoint of the fluorescence peak. The binding of a substrate by the protein results in a stabilisation of the protein structure that increases the temperature necessary for denaturation; therefore, a difference in T_m (ΔT_m) on addition of a substrate indicates binding.

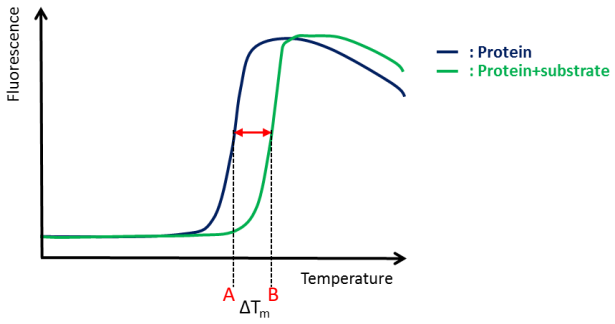


Figure 56: Principles of thermal shift assays. The blue curve shows the fluorescence plotted as temperature increases in the protein only control, with the T_m shown as A. The green curve shows the result when the temperature is increased with a substrate present, with binding resulting in stabilisation of the protein and an increased Tm, shown as B. The ΔT_m is calculated as B - A.

The RPA1975 screening process was conducted against a wide range of biologically relevant compounds as there was little evidence suggesting what its substrate specificity might be. This library of potential ligands included aliphatic carboxylic acids, aromatic acids, amino acids, and sugars. Ligands were added in excess to ensure saturation of the protein if binding occurs. This screening process identified two dicarboxylates, adipate and suberate, which produced a thermal shift. Therefore a range of dicarboxylates were screened and five medium- to long-chain dicarboxylates showed a significant ΔT_m, indicating binding, with pimelic and suberate showing the greatest shift with a ΔT_m of > 10 °C. Other dicarboxylates of shorter and longer carbon chain lengths, tricarboxylates, unsaturated

dicarboxylates and substituted dicarboxylates were screened but showed no indication of binding.

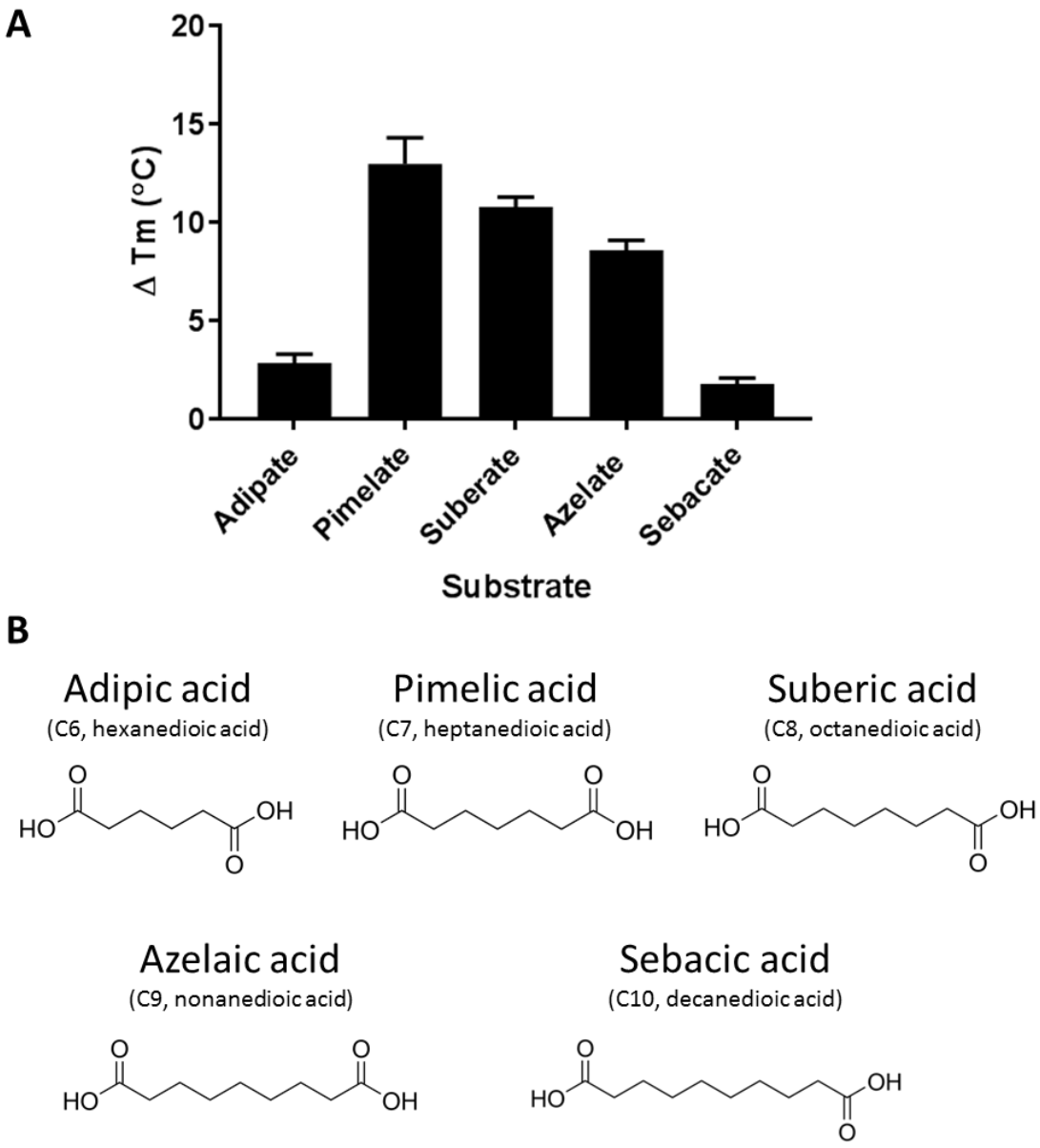


Figure 57: Thermal shift assay indicates binding of medium-/long-chain dicarboxylates by RPA1975. A) Five medium/long chain dicarboxylates in excess were shown to cause a shift in the T_m of RPA1975 using a fluorescent dye, with the degree of shift apparently dependent on the carbon-chain length of the substrate. B) The chemical structures of the five dicarboxylates, with their carbon-chain (C) length and IUPAC name given.

5.2.3. Binding affinity of RPA1975 with medium-/long-chain dicarboxylates

Thermal shift assays provide an indication of substrate binding but cannot accurately measure binding quantitatively; tryptophan fluorescence titrations assays allow for a quantitative characterisation of a protein's specificity by determination of the dissociation constant (K_d) with a given substrate. Every protein containing tryptophan and tyrosine residues has an intrinsic fluorescence level on excitation at maximal absorption at 280 nm and emission levels peaking between 300 and 350 nm, predominantly as a result of the tryptophan residues. The fluorescence intensity is dependent on the protein conformation and is influenced by the spatial arrangement and degree of hydration of the tryptophan residues relative to other residues. Substrate binding results in a conformational change that elicits a significant quench (or occasionally an enhancement) in the intrinsic fluorescence of the protein in solution. This is highly sensitive, allowing for the titration of nanomolar amounts of substrate against the protein and the plotting of fluorescence against substrate concentration to determine the K_d.

RPA1975 has nine tryptophan residues, which is a relatively high proportion of its sequence, resulting in a high intrinsic fluorescence. Proteins with many tryptophan residues can be problematic as the change in overall fluorescence on substrate binding is smaller proportional to the initial total fluorescence. However, addition of four of the potential substrates identified from thermal shift screening — adipate, pimelate, suberate and azelate —did elicit a noticeable change in fluorescence; Figure 58 shows an enhancement of fluorescence, rather than a quench, was observed. Due to RPA1975's high fluorescence the protein concentration for the assays had to be reduced to 0.025 µM, with maximal fluorescence emission at 330 nm. Protein-free controls were conducted to confirm that the substrates did not produce artefactual fluorescent effects.

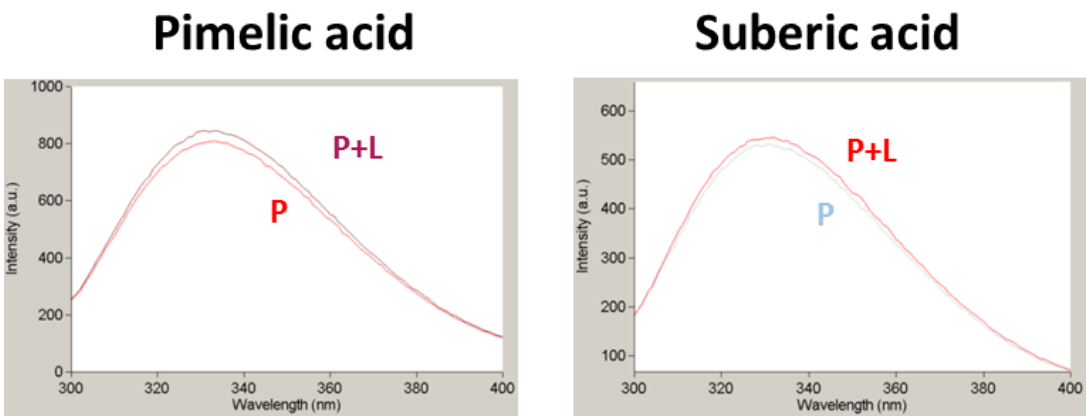


Figure 58: Changes in intrinsic fluorescence of RPA1975 on addition of dicarboxylate substrates. The fluorescence emission spectrum of RPA1975 was measured when excited at 280 nm, showing maximal emission at 330 nm. Due to RPA1975's high intrinsic fluorescence the protein concentration was reduced to 0.025 µM to allow for measurement. Shown are the shifted emission spectra on addition of two dicarboxylate ligands, pimelate amd suberate. Addition of the ligands (10 µM) to the protein (P+L) shows an enhancement in fluorescence compared to the protein only (P) spectra.

Titrations of the four dicarboxylates were then conducted and the fluorescence emission measured at 330 nm (Figure 59). The substrate concentration range used for pimelate and suberate was 10 nM – 2 µM, and for the other substrates was 10 nM – 10 µM. A protein-only control, where buffer was added in equivalent volumes, was conducted and the change subtracted from the substrate values to account for dilution effects and fluorescence drift. These adjusted values were used to plot the change in fluorescence against substrate concentration, and input into an iterative fitting process using the Levenberg-Marquadt non-linear "least squares" method to solve the tight-binding quadratic equation and determine the K_d. The titration and subsequent analysis was performed for each substrate in triplicate, and the calculated binding affinities shown in Table 8.

These data show that RPA1975 exhibits tight binding with pimelate and suberate, with K_d values of 13 and 31 nM, respectively. The binding shown with azelate was significantly weaker, with a K_d value of 659 nM. The titrations using adipate were hindered by the extremely low enhancement in fluorescence displayed on addition of the substrate. This enhancement was actually less than the decrease in fluorescence from drift in the protein-only control; with this factored in the enhancement reached only 2% above substrate-free

level. The data for adipate was used to calculate a high K_d value of 2.05 µM, though the accuracy of this is uncertain due to the extremely low changes shown.

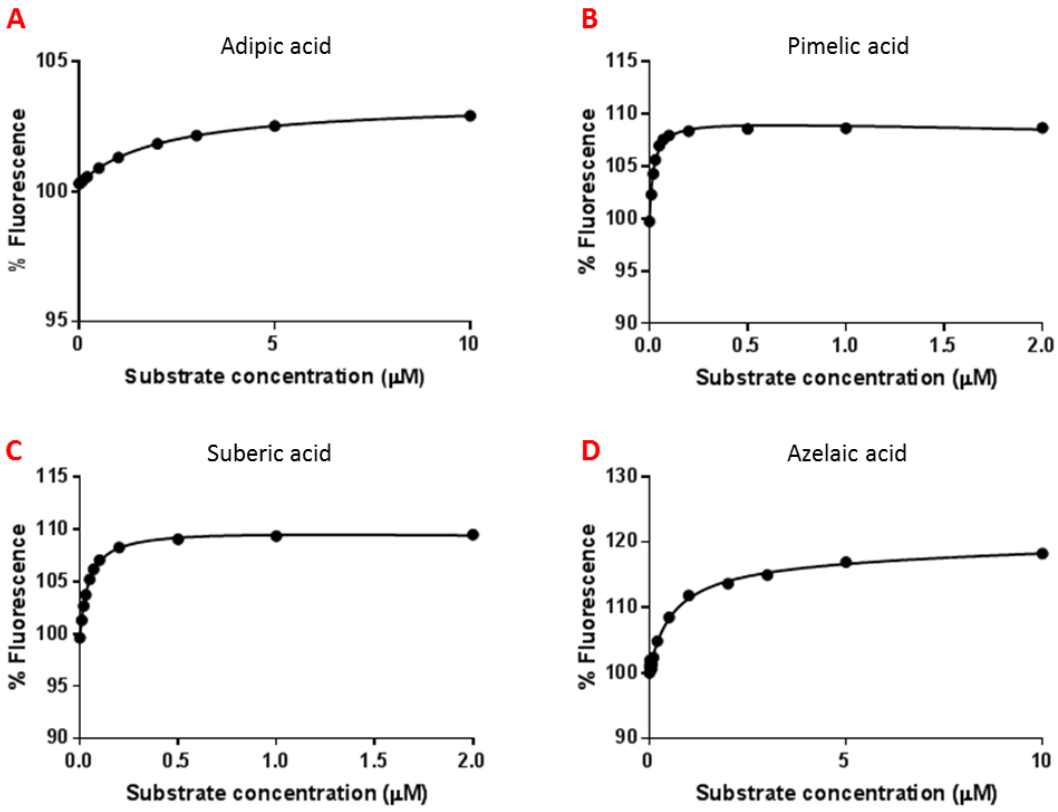


Figure 59: Fluorescence titrations of RPA1975 with dicarboxylates. The intrinsic fluorescence of a protein solution of RPA1975 was measured at 330 nm while adipate (A), pimelate (B), suberate (C), and azelate (D) were titrated in. All four substrates elicited a concentration-dependent enhancement of fluorescence, indicating binding. Shown are the calculated values for one independent titration experiment, with a non-linear fitting for single-site saturation binding.

Substrate	K_d (nM)
Adipate	**2,054** ± 429
Pimelate	**13** ± 6
Suberate	**31** ± 6
Azelate	**659** ± 55
Sebaceate	**N/A**

Table 8: Binding characteristics of RPA1975 determined by tryptophan fluorescence titrations. Five dicarboxylates identified as potential substrates of RPA1975 were titrated against the protein and the change in intrinsic fluorescence measured. K_d values are given as a mean of three independent experiments, with the standard deviation shown.

These data suggest that the seven-carbon dicarboxylate, pimelate, is the preferred substrate of RPA1975, with carbon-chain length a determining factor in substrate affinity.

5.2.4. Purification of MBP-fused RPA1975

In order to further characterise RPA1975 and for determination of its structure, more purified protein was required. However, subsequent attempts to purify RPA1975 failed due to a recurrence of the solubility problems previously encountered. With all conditions and parameters varied, i.e. expression plasmid, *E. coli* production strain, growth temperature (18 °C–37 °C), IPTG concentration for induction, induction timing, no significant amount of soluble RPA1975 could be produced for purification. Therefore an alternative approach was used in an attempt to purify the protein for structural determination by x-ray crystallography.

The RPA1975 gene was amplified and ligated, as before, into the pMAL-p2e expression vector. This vector was used for production of RPA1975 with an N-terminal fusion with maltose-binding protein (MBP; encoded by the *malE* gene), with the periplasmic signal sequence intact for targeting of the fused proteins to the periplasm. Fusion with MBP has been shown to increase solubility of proteins, and allows for purification by using amylose affinity chromatography and elution by using a maltose concentration gradient. The linker region between the MBP and the RPA1975 includes an enterokinase cut site for cleavage.

Following cloning, the RPA1975:MBP-fusion proved soluble and enabled purification. Gel filtration chromatography was then used to separate the fusion protein from unfused MBP that bound the amylose column. Unfortunately, cleavage using enterokinase rendered the RPA1975 product insoluble again. Therefore, crystallisation trials using the purified

RPA1975:MBP-fusion product are planned in an attempt to determine the structure with the fusion intact.

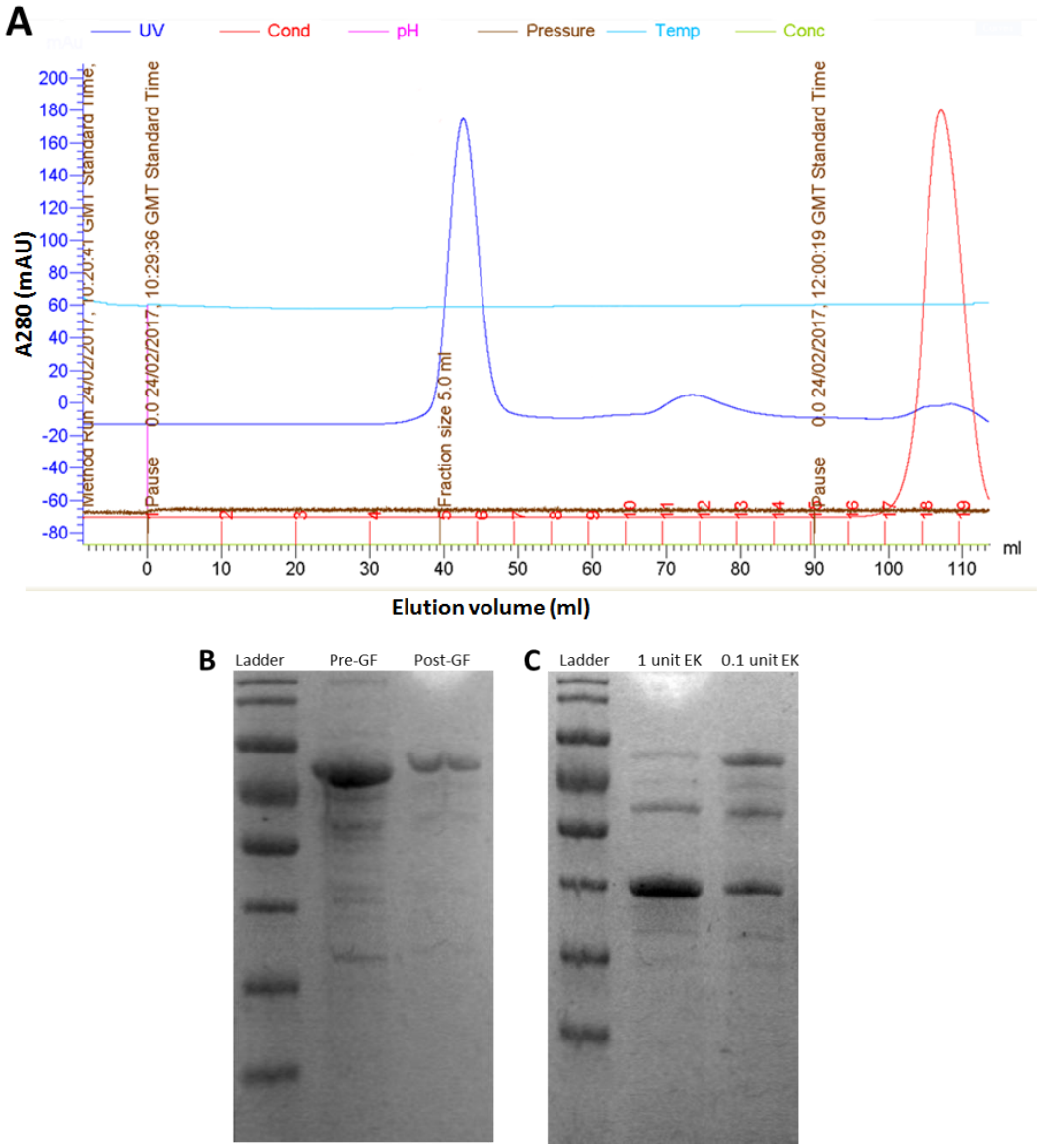


Figure 60: Purification of a RPA1975:MBP-fusion. A) Plot showing elution of protein from the gel filtration column over 110 ml of buffer. The A280 (blue line) peak after ~38 – 48 ml of elution shows the large (78 kDa) RPA1975:MBP fusion eluting; the peak ~70 – 80 ml shows the smaller (43 kDa) unfused MBP (MalE) protein. B) SDS-PAGE showing the protein nature of the sample pre-gel filtration (pre-GF) and post-gel-filtration (post-GF). The band at 78 kDa represents the fused product; gel filtration resulted in a purer product. C) SDS-PAGE showing attempts to cleave the fusion product to produce RPA1975 (and MBP). Different amounts of enterokinase (EK) were used; though cleavage was successful with a band at 43 kDa showing the MBP, no band representing RPA1975 was visible at 37.7 kDa.

5.2.5. Comparison of RPA1975 and previously characterised R. palustris ABC SBPs

The thermal shift and tryptophan fluorescence assays for RPA1975 clearly indicate a possible role in the uptake of pimelate, and potentially other dicarboxylates. Based on the findings of previous studies this is likely another example of redundancy in *R. palustris*; downstream of the *pimFABCDE* operon that encodes the enzymes of the β-oxidation pathway for metabolism of pimeloyl-CoA, as shown in Figure 16, are eight genes encoding components of an ABC transport system (Figure 61).

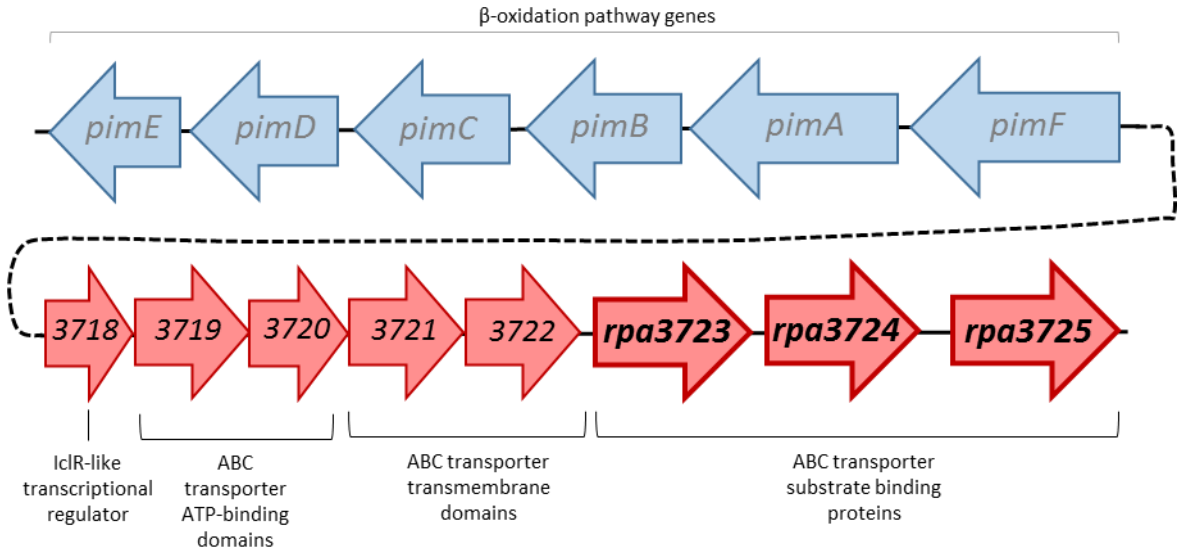


Figure 61: Location of genes encoding the ABC transport system that contains three SBPs. Adjacent to the *pim* operon of enzymes is a complete ABC transport system with three distinct SBPs that have been shown to bind dicarboxylic and fatty acid substrates of the pathway by thermal shift assays.

The work of Giulani et al (2011) into *R. palustris*'s ABC transport systems screened these three ABC system SBPs against six-carbon to fourteen-carbon chain dicarboxylates and eight-carbon to eighteen-carbon chain fatty acids using thermal shift assays (Giuliani et al., 2011). They reported T_m shifts for RPA3723 with suberate, azelate, sebaceate, undecandioate and tetradecanedoate, and the fatty acids caprylate, caprate, laurate and myristate; for RPA3724 with the same substrates plus pimelate and the fatty acid palmitate; and found no indication of binding for any of the substrates screened with RPA3725. Therefore, there is likely crossover in the substrates of RPA1975 with these three ABC SBPs based on these published data, with RPA3724 potentially exhibiting the most similar substrate range compared to the TRAP SBP.

A comparison of the amino acid sequences of these ABC SBPs (seen in Figure 62) shows a high degree of homology, with the sequences being nearly identical excluding the N-terminal signal sequence; this suggests a gene duplication event is responsible. Comparing these sequences to that of RPA1975 shows little sequence identity, though there are several conserved residues. Based on these comparisons it is interesting that the three ABC SBPs appear to display different affinities for different dicarboxylic and fatty acids based on the substrates carbon chain length despite the homology. Therefore quantitative characterisation of the ABC SBPs was pursued for comparison with each other and RPA1975, while determination of the protein structure was sought to elucidate the reason for the different affinities.

Figure 62: Multiple sequence alignment of RPA1975 and the three ABC SBPs. The T-Coffee webserver was used for multiple sequence alignment of the amino acid sequences of RPA1975, RPA3723, RPA3724 and RPA3725(Di Tommaso et al., 2011). Residues are highlighted in the colour based on degree of alignment indicated in the key; an (*****) denotes conserved residues, (**:**) conservative residues, (**.**) semi-conservative residues, and ( ) non-conservative residues. The red box indicates the N-terminal periplasmic targeting signal sequence.

5.2.6. Overproduction and purification of the ABC SBPs: RPA3723, RPA3724 and RPA3725

A similar approach as used for purification of RPA1975 was used for the three ABC SBPs. Genes encoding each protein were amplified from *R. palustris* CGA009 gGNA excluding the N-terminal
R. palustris periplasmic signal sequence, and the stop codon. PCR products were ligated into both pET21a(+) and pET22b(+) expression vectors using double-digest restriction sites added by PCR; the restriction sites used were *Nde*I and *Xho*I for ligation into pET21a(+), and *Nco*I and *Xho*I for *rpa3723* and *rpa 3724*, and *Nco*I and *Bam*HI for *rpa3725* for ligation into pET22b(+). Constructs were transformed into *E. coli* BL21(DE3) for production trials.

Protein production was assessed at different growth temperatures, with different concentrations of IPTG used for induction, and at different time points post-induction. The results of these trials using the pET21a(+) constructs are shown in Figure 63 for RPA3723, Figure 64 for RPA3724, and Figure 65 for RPA3725. RPA3723 showed consistent yields across conditions and proved soluble. The optimal conditions were determined to be 37 °C, 0.4 mM IPTG and 5 h, as shown by the band ~41 kDa in Figure 63; these conditions were used for purification of the protein using nickel-affinity chromatography with a yield of ~15 mg l^{-1} *E. coli* BL21(DE3) culture. RPA3724 showed high yields but low solubility, with use of the pET22b(+) construct producing no improvement. Solubility of the RPA3724 was achieved by re-cloning of the gene from *R. palustris* gDNA with the inherent signal sequence intact, and the pET21a(+) construct with this insert showed high yields and solubility as shown by the large band at ~45 kDa in Figure 64; purification achieved a yield of ~4 mg l^{-1} culture. Figure 64C shows that the protein eluted from the column appeared to be a mixture of a protein of size ~45 kDa, and a protein of size ~42.5 kDa; this corresponds to RPA3724 with and without the N-terminal signal sequence and suggests that *E. coli* is capable of cleaving the signal sequence from a proportion of the recombinant protein produced. RPA3725 showed high yield and solubility. The optimal conditions were determined to be 37 °C, 0.4 mM IPTG and 3 h as shown by the band at ~41 kDa in Figure 65; these conditions were used for purification of the protein with a yield of ~5 mg l^{-1} culture. All protein samples eluted from nickel-affinity chromatography were urea treated to remove endogenous ligands and concentrated for biochemical analysis.

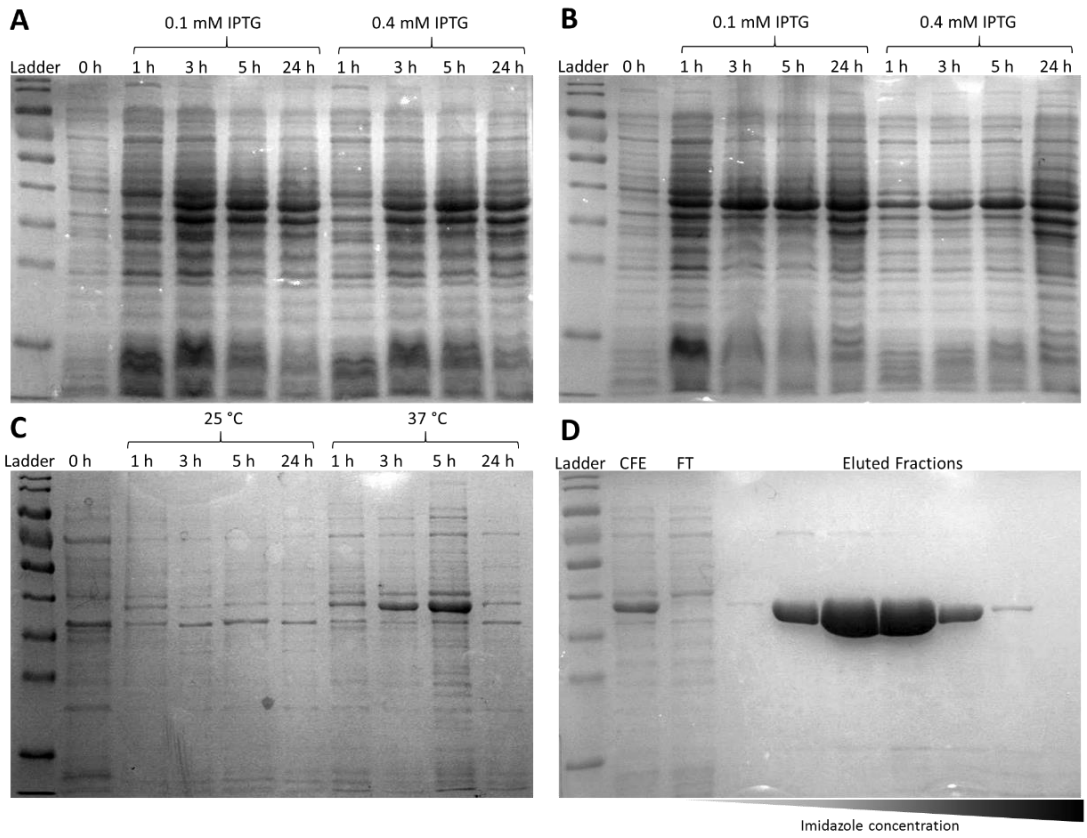


Figure 63: Overproduction and purification of RPA3723. A) SDS-PAGE of whole-cell samples of BL21(DE3) harbouring pET21a:*rpa3723* grown at 25 °C. B) SDS-PAGE of whole-cell samples of BL21(DE3) harbouring pET21a:*rpa3723* grown at 37 °C. C) SDS-PAGE of soluble-fraction samples of BL21(DE3) harbouring pET21a:*rpa3723* with induction by addition of 0.4 mM IPTG. D) SDS-PAGE showing purification of RPA3723 via nickel-affinity chromatography. RPA3723 is shown as a band at ~41 kDa in the cell-free extract (CFE) that is not presence in the unbound elution of the flow-through (FT); increasing the imidazole concentration results in purification of RPA3723 in the eluate. The ladder used can be found on page 255.

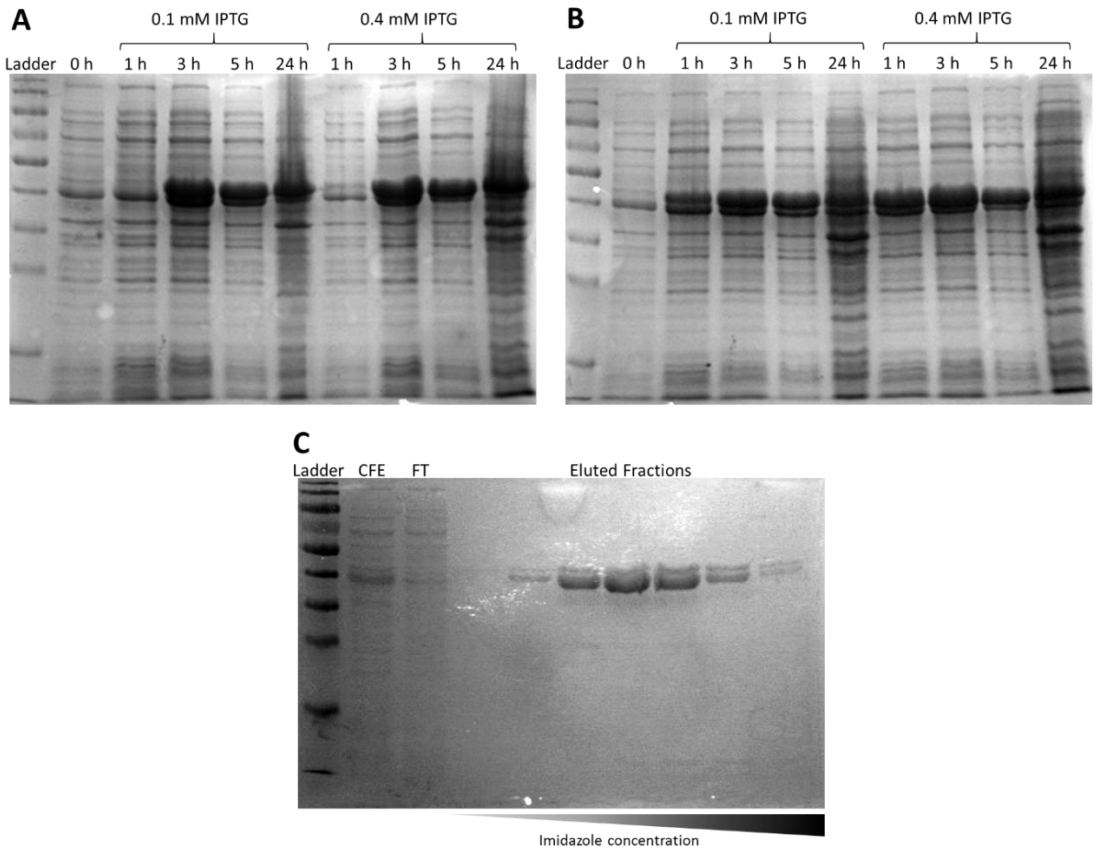


Figure 64: Overproduction and purification of RPA3724. A) SDS-PAGE of whole-cell samples of BL21(DE3) harbouring pET21a:*rpa3724* grown at 25 °C. B) SDS-PAGE of whole-cell samples of BL21(DE3) harbouring pET21a:*rpa3724* grown at 37 °C. C) SDS-PAGE showing purification of RPA3724 via nickel-affinity chromatography. The cell-free extract (CFE) shows a strong band at ~42.5 kDa, with a fainter band above at ~45 kDa, potentially indicating RPA3724 without and with the signal sequence, respectively; these bands do not appear in the unbound flow-through (FT). Following elution with an imidazole gradient both forms of protein were eluted from the column together, with no other impurities. The ladder used can be found on page 255.

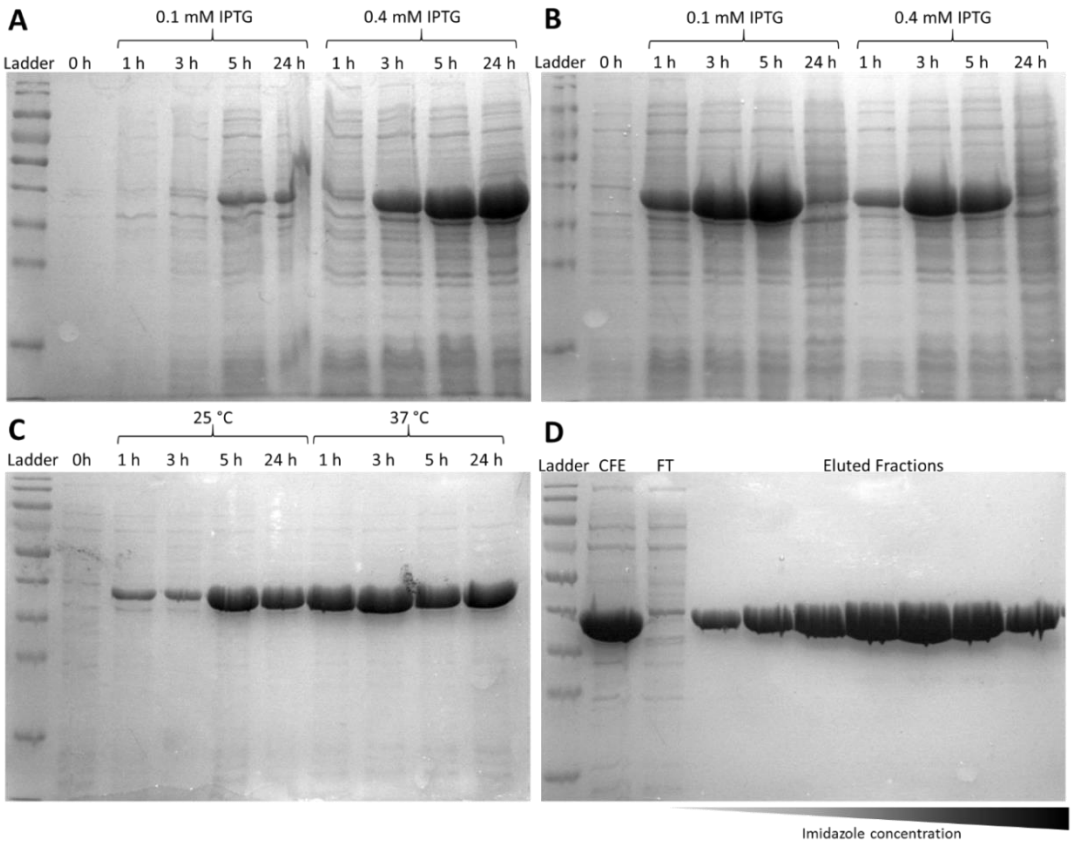


Figure 65: Overproduction and purification of RPA3725. A) SDS-PAGE of whole-cell samples of BL21(DE3) harbouring pET21a:*rpa3725* grown at 25 °C. B) SDS-PAGE of whole-cell samples of BL21(DE3) harbouring pET21a:*rpa3725* grown at 37 °C. C) SDS-PAGE of soluble-fraction samples of BL21(DE3) harbouring pET21a:*rpa3725* with induction by addition of 0.4 mM IPTG. D) SDS-PAGE showing purification of RPA3725 via nickel-affinity chromatography. RPA3725 is shown as a band at ~41 kDa in the cell-free extract (CFE) that is not present in the unbound elution of the flow-through (FT); increasing the imidazole concentration results in purification of RPA3725 in the eluate. The ladder used can be found on page 255.

5.2.7. Biochemical analysis of the three ABC transporter SBPs.

The ligands for the three ABC transporters had been previously indicated by the work of Giulani et al. (2011), and several of these dicarboxylic and fatty acids were used in thermal shift assays with the purified SBPs for confirmation of binding and comparison with the results from RPA1975 (Giuliani et al., 2011). 10 µM of purified protein was used with substrates added in excess at a final concentration of 60 µM. The C6 – C10 dicarboxylates adipate, pimelate, suberate, azelate and sebaceate, and the C8, C10 and C14 fatty acids caprylate, capric and myristate were used as substrates and ΔT_m for each with each protein is shown in Figure 66. The results for the ABC SBPs are similar to those shown previously, and the data suggests that there is crossover in the substrate ranges for RPA1975 and the ABC SBPs, though the latter potentially have higher affinity for longer carbon-chain substrates. In contrast to the previous study, shifts were detected with RPA3725 indicating binding of three fatty acid substrates.

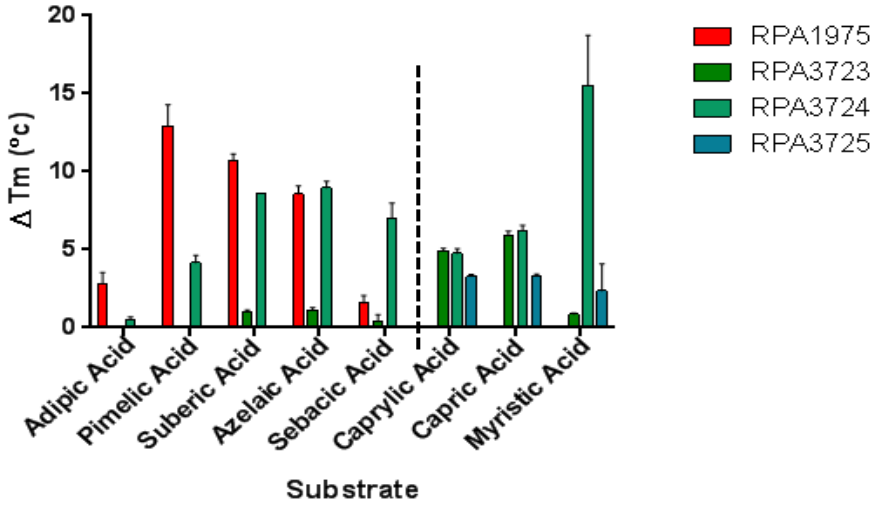


Figure 66: Comparison of thermal shifts with dicarboxylic and fatty acids between RPA1975, RPA3723, RPA3724 and RPA3725. The ΔT_m of the three purified ABC transporters with potential ligands as determined here are shown for comparison with those shown by RPA1975. RPA1975 appears to show a preference for the dicarboxylate substrates to the left of the dotted partition, while the ABC SBPs a preference for the fatty acids to the right.

To provide a more comprehensive comparison between the SBPs, tryptophan fluorescence titration assays were undertaken with the three purified ABC SBPs, using the dicarboxylate substrates and methodology described in Section 5.2.3. Quenches were displayed with RPA3723 but these were extremely small relative to the total fluorescence and resulted in

variable binding titration data (Figure 67); however, sub-micromolar affinity was indicated for all five dicarboxylates (Table 9). The fluorescence drift exhibited by RPA3723 was also large and variable, which ultimately limits the validity of the data. Based on these data RPA3723's affinity is similar for all five substrates, and they also indicate the binding of pimelic and suberate is not as tight as shown with RPA1975. The thermal shift data had suggested that RPA3724 binds pimelate, suberate, azelate and sebaceate; however, on addition of these substrates to an RPA3724 solution no significant change in fluorescence was observed. At the time these experiments were ongoing, another PhD student in the lab, Abrar Akbar, was also assessing binding of phenylalkane carboxylates by RPA3724 based on the results of a proteomic investigation in *R. palustris* grown on phenylvalerate. Binding assays titrating these alternative substrates, that had exhibited thermal shifts, also showed no fluorescence change. Titrations were also attempted with purified RPA3725 with all five dicarboxylate substrates but no reliable binding data could be obtained, suggesting RPA3725 does not bind any of these substrates. Due to time restraints and conflicts, no binding titration data was obtained for any of these ABC SBPs with fatty acid substrates.

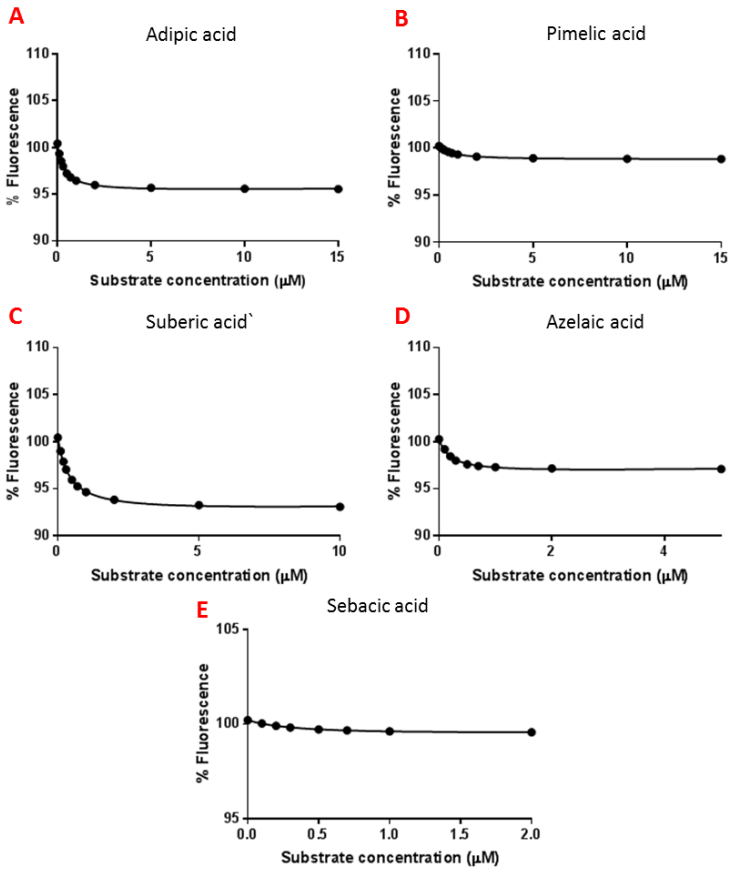


Figure 67: Fluorescence titrations of RPA3723 with dicarboxylates. The fluorescence of a protein solution of RPA3723 was measured at 330 nm while adipate (A), pimelate (B), suberate (C), azelate (D), and sebaceate (E) were titrated in. Shown are the calculated values for one independent titration experiment, with a non-linear fitting for single-site saturation binding.

As mentioned previously the major source of intrinsic fluorescence in proteins when excited at 280 nm are tryptophan residues. RPA3724 has seven tryptophan residues (compared to RPA3723's five); this large number may result in the fluorescence of the protein being very stable on binding, and any quenching (or enhancing) is diluted. This suggests that fluorescence titrations are not a feasible means of determining RPA3724's binding kinetics.

Substrate	K_d (nM)
Adipate	**172** ± 46
Pimelate	**208** ± 65
Suberate	**145** ± 100
Azelate	**70, 184**
Sebaceate	**140, 221**

Table 9: Binding data for RPA3723 determined by tryptophan fluorescence. While quenching of the fluorescence during titration of dicarboxylates into a solution of RPA3723 was observed, the percentage of total fluorescence quenched was very small resulting in an extremely high variability in the data as shown by the large standard deviations. K_d values are a mean of three titrations for adipate, pimelate and suberate; for azelaic and sebaceate the changes in fluorescence were extremely variable and values of two independent titrations are shown.

5.2.8. Structural determination of RPA3724

Ultimately, structural determination of all three ABC SBPs would be ideal for comparing the structure of their binding pockets and the relationship between this and binding specificity. Initial efforts to obtain a crystal structure were pursued with RPA3724 based on the thermal shift data indicating binding of pimelate, to allow for comparison with a RPA1975 structure. Preliminary crystallography trials were conducted using purified 10 mg ml^{-1} RPA3724 with and without additional pimelate using the PACT, JCSG+ and Proplex plates in 96-well sitting-drop experiments. These trials identified crystals conditions for crystal growth: 20% (w/v) polyethylene glycol 6000 in 0.1 M MES buffer (pH 6), 0.1 M HEPES buffer (pH 7) or 0.1 M Tris-HCl buffer (pH 8), with either 0.2 M lithium chloride, magnesium chloride or calcium chloride salt added. RPA3724 crystals were successfully used to obtain an x-ray diffraction data set. However, at the time of writing a structure has not been able to be solved by molecular replacement.

5.2.9. Creation of R. palustris ABC SBP gene knock-out mutants

To assess the importance of the TRAP and ABC transport systems, the genes encoding their respective SBPs were knocked out. A *Δ1975* mutant of *R. palustris* CGA009 had previously been created by the work of Rob Salmon (University of Sheffield); therefore a *Δ3723-3725* mutant was created from *R. palustris* CGA009, and a double *Δ1975Δ3723-3725* mutant created from the Δ1975 mutant. The pK18mobsacB suicide vector was used for this

mutagenesis, with ~500 bp sequences of DNA corresponding to the flanking regions of the three ABC SBP genes ligated into this vector by isothermal assembly. The pK18mobsacB:Δ3723-3725 construct was conjugated into wildtype and *Δ1975 R. palustris* cultures. These cultures were sub-cultured and grown in the presence of kanamycin to select for cells that had undergone the first recombination event. Sucrose selection was subsequently used to select cells that had undergone a second recombination event, with markerless knock-out mutants formed. Colony PCR was used to screen for revertants to wildtype during the second recombination, and the positive colonies from this were grown up for gDNA isolation. This gDNA was then used as a template for PCR amplification to confirmed knock-out of the target, with the results shown in Figure 68.

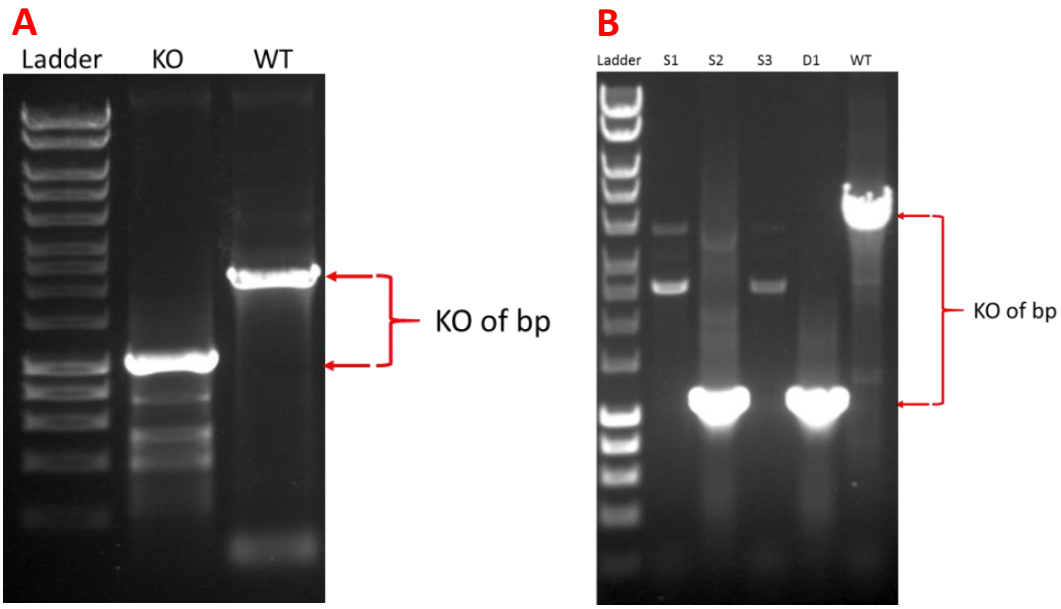


Figure 68: Confirmation of knock out of targeted gDNA in *R. palustris*. A) PCR amplification using primers outside the flanking regions with template gDNA from the *Δ1975* mutant provided by Rob Salmon (KO), and WT *R. palustris*; the removal of ~1100 bp is apparent. B) PCR amplification using primers outside the flanking regions of the three ABP SBP genes with template gGNA from three potential single *Δ3723-3725* mutants (S1-S3) and one potential double *Δ1975Δ3723-3725* mutant (D1), and WT. S2 and D1 produced a PCR product ~3000 bp shorter, indicating knock out occurred. "KO of bp" = knock-out of base pairs indicating targeted knock out.

5.2.10. Growth phenotype of mutants

The growth rates of these three mutant strains of *R. palustris* were compared with each other and WT on pimelate, and succinate as a control, to assess their physiological role and their importance for growth on the substrates they bind. WT, Δ1975, *Δ3723-3725* and *Δ1975Δ3723-3725 R. palustris* cells were grown in RCV minimal medium supplemented with succinate for 5 days before inoculation into fresh medium containing either succinate or pimelate as the sole carbon source. Growth was monitored and results are shown in Figure 69.

When an excess of pimelate (6 mM) was provided, there was no significant growth phenotype displayed by any of the mutants relative to the wildtype. Therefore growth was also monitored at a 10-fold lower pimelate concentration in order to deduce whether these high-affinity SBPs were required for growth when the carbon source was limited. However, there was still no significant difference in growth between *R. palustris* strains. These data suggest that *R. palustris* is still capable of uptake of pimelate when both the TRAP and ABC transport systems' SBPs are knocked out.

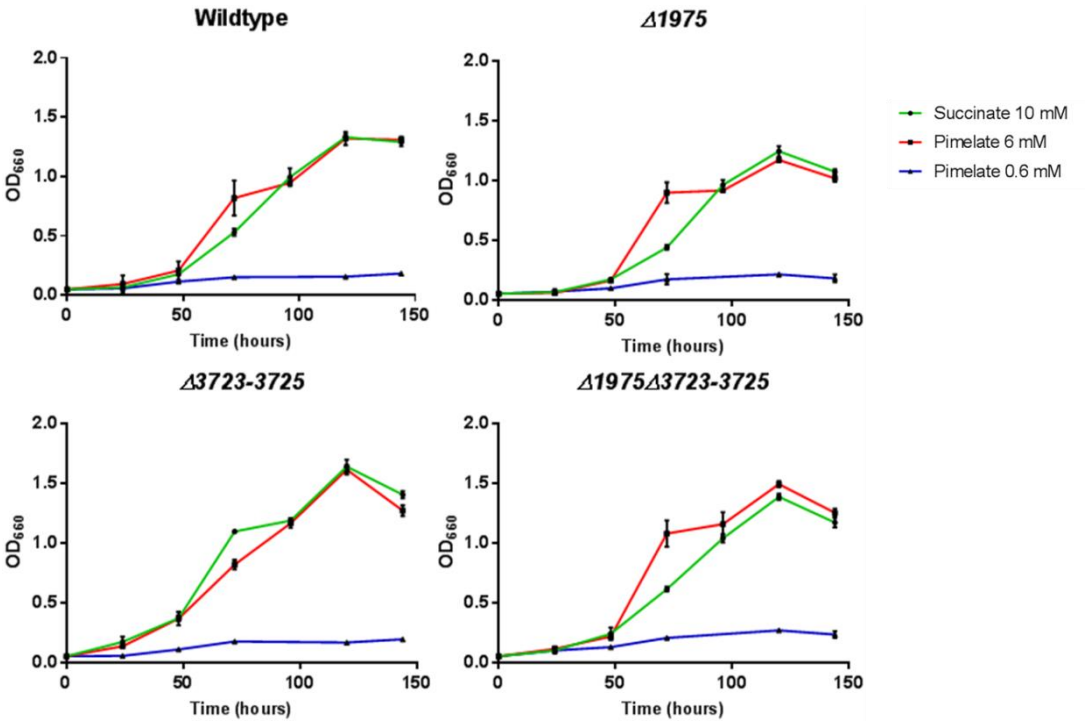


Figure 69: Growth of *R. palustris* transporter mutants. Wildtype, *Δ1975*, *Δ3723-3725*, and *Δ1975Δ3723-3725* cultures were grown photoheterotrophically in RCV minimal medium with succinate (as a control) and pimelate as the sole carbon source provided. Growth was monitored daily by measurement of the OD_{660} of the cultures. The mean of cultures grown in triplicate are shown, with the SD displayed as error bars.

5.2.11. Transcriptional regulation of rpa1975 and rpa3724 genes determined by RT-PCR

In order to investigate the regulation of the transporters at a transcriptional level, quantitative RT-PCR was used to determine the abundance of SBP mRNA from *R. palustris* cells growing with the dicarboxylate substrates as the carbon source. Regulation of transcription of the *rpa1975* and *rpa3724* was assessed for a comparison in regulation of the TRAP and ABC transport system. WT *R. palustris* cells were grown to mid-logarithmic phase on succinate before addition of either succinate or, pimelate or sebaceate; cells were harvested for RNA extraction after growth for 4 h on the substrate of interest. Pimelate was used as the previous data suggest it is a substrate for both binding proteins, while sebaceate only binds RPA3724.

The RT-PCR data shown in Figure 70 indicate that there is no significant regulation of *rpa1975* as a result of growth on either of these substrates, while there was a 16.2-fold and 11.8-fold down-regulation of *rpa3724* when grown on pimelate and sebaceate relative to growth on succinate, respectively. These results were contrary to expectations, as it suggests neither the TRAP nor the ABC transport system is positively regulated by the substrates they bind.

Based on the similar substrate ranges of the TRAP and ABC transport systems, but different source of energy-coupling, RT-PCR was used to assess any regulation based on sodium ion concentration. *R. palustris* cells were grown with pimelate as the carbon source in two types of minimal media; "Low Na^+" RCV medium was prepared using potassium salts substituted instead of sodium salts, and "High Na^+" RCV medium was the same but with 120 mM NaCl added. Figure 71 shows that there was no significant regulation of these transporters based on the sodium concentration of the media.

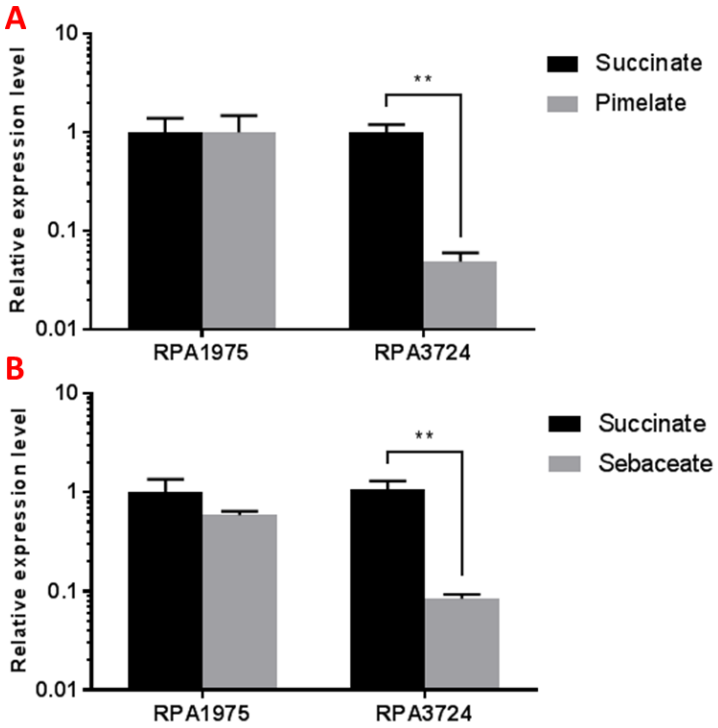


Figure 70: Transcription levels of SBP genes during growth on different carbon sources. *R. palustris* cells were grown anaerobically in minimal media the with carbon sources — succinate, pimelate or sebaceate — shown. RNA was extracted and mRNA levels quantified by RT-PCR. Expression level was normalised to *rpoD,* and shown as fold-difference relative to expression during growth on succinate. The mean value of three biological and technical replicates is shown with the standard deviation as error bars. Statistical t-tests were performed on data sets; p-values: <0.001 = ***, 0.001–0.01 = **, 0.01–0.05 = *, >0.05 = not significant.

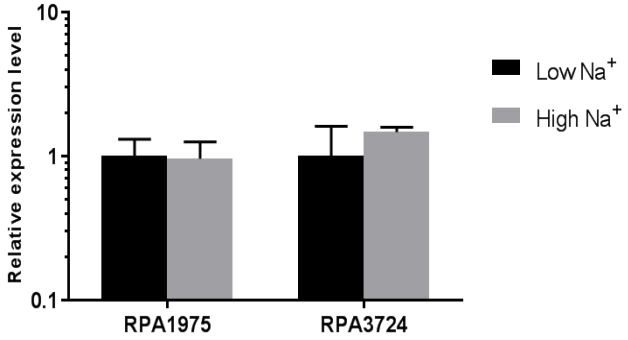


Figure 71: Transcription levels of SBPs during growth at different sodium levels. *R. palustris* cells were grown anaerobically in "Low Na^+" and "High Na^+" minimal media with pimelate. RNA was extracted and mRNA levels quantified by RT-PCR. Expression level was normalised to *rpoD,* and shown as fold-difference relative to expression during growth in "Low Na^+" media. The mean value of three biological and technical replicates is shown with the standard deviation as error bars. Statistical t-tests were performed on data sets; p-values: <0.001 = ***, 0.001–0.01 = **, 0.01–0.05 = *, >0.05 = not significant.

5.2.12. Overproduction and purification of RPA2047

The *rpa2047* gene, excluding the N-terminal sequence encoding the signal sequence, and the stop codon, was amplified by PCR from *R. palustris* gDNA using the 2047_21a_F and 2047_21a_R primers, for ligation into pET21a(+) using the *Nde*I and *Hind*III restriction sites. The construct was then transformed into *E. coli* BL21(DE3) cells for protein overproduction. RPA2047 was produced with a good yield and proved soluble during growth at 37°C as shown in Figure 72.

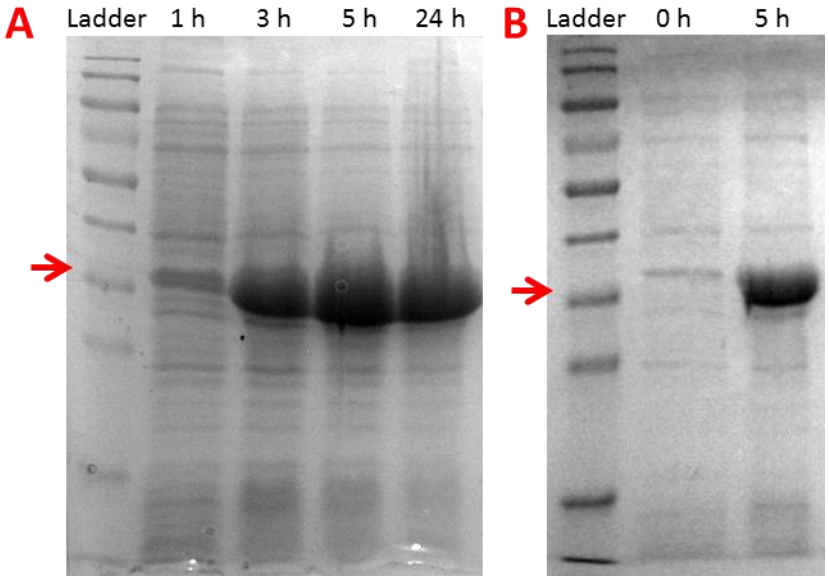


Figure 72: Overproduction trials of RPA2047. A) SDS-PAGE of whole-cells of BL21(DE3) transformed with pET21a:2047 and induced with 0.4 mM IPTG at t = 0 h. A high yield of RPA2047 was visualised from 3 h of production onwards. B) SDS-PAGE of soluble fractions of the same strain; the overproduced RPA2047 proved highly soluble. The ladder used can be found on page 257.

Overproduction for purification used the optimal conditions identified in the trials: induction with 0.4 mM IPTG, and overproduction for 5 h at 37°C. Nickel-affinity chromatography was used to purify recombinant 6xHis-tagged RPA2047 with a high yield of relatively pure protein as shown in Figure 73. Collected elution fractions were pooled together and concentrated during buffer exchange to remove the imidazole, which also removed the low mw impurities.

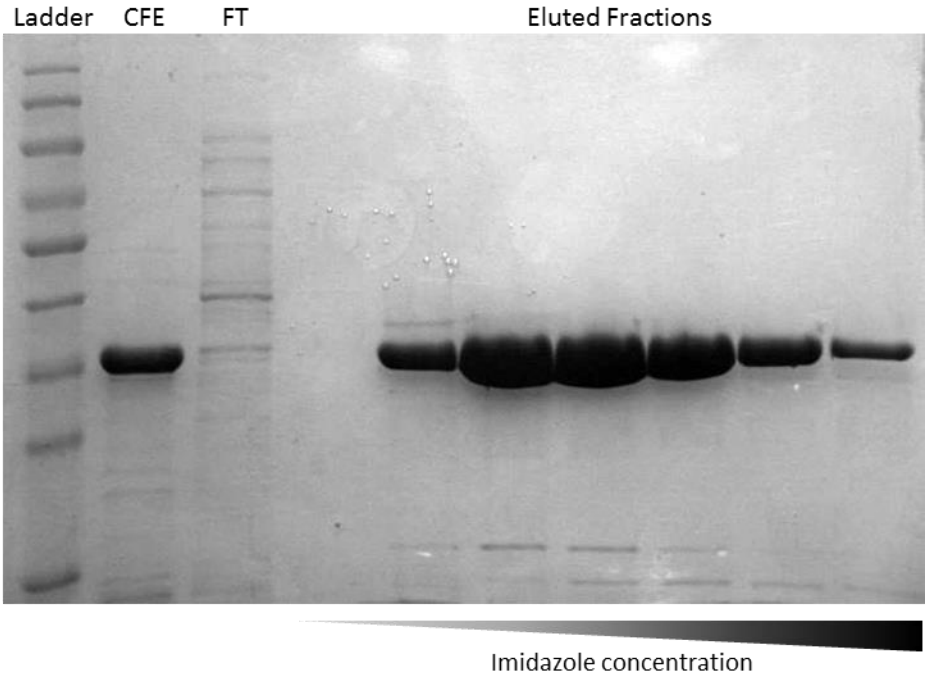


Figure 73: Purification of RPA2047. SDS-PAGE shows the stages of the purification by nickel-affinity chromatography of RPA2047. The cell-free extract (CFE) lane shows a large band at ~36 kDa highlighting the presence of soluble RPA2047 prior to its injection, and the flow-through (FT) lane shows the proteins in the CFE that did not binding to the column. As the imidazole concentration increased, a large rise in A280 indicated elution of RPA2047, and these collected samples are shown as the eluted fractions confirming the purification. The ladder used can be found on page 257.

5.2.13. Ligand screening of RPA2047 by thermal shift assays

To identify possible substrates for RPA2047, the purified protein was used in 96-well plate thermal shift screening assays. RPA2047 was screened at final concentration of 10 µM against a library of potential ligands including carboxylic, amino, aromatic and keto acids at a final concentration of 60 µM to provide an excesses of ligand. Five compounds resulted in an increase in the T_m of RPA2047, as detected by fluorescence, of more than 5 °C: malate, fumarate, succinate, mesaconate and itaconate; all these are dicarboxylates with a very similar structure (Figure 74). Three other carboxylates, citrate, oxalacetate and aconitate showed a small shift of less than 2 °C. No thermal shift was seen with the following ligands: any proteinogenic amino acids, lactate, isopropylmalate, valerate, acetate, butyrate, oxalate, propionate, pyruvate, hydroxypyruvate, α-ketoglutarate, methyl 2-oxobutanoate, α-ketobutyrate, malonate, tartarate, crotonate, cinnamate, ferulate, coumrate, caffeate,

benzoate, biotin, ascorbate, fucitol and mannitol.

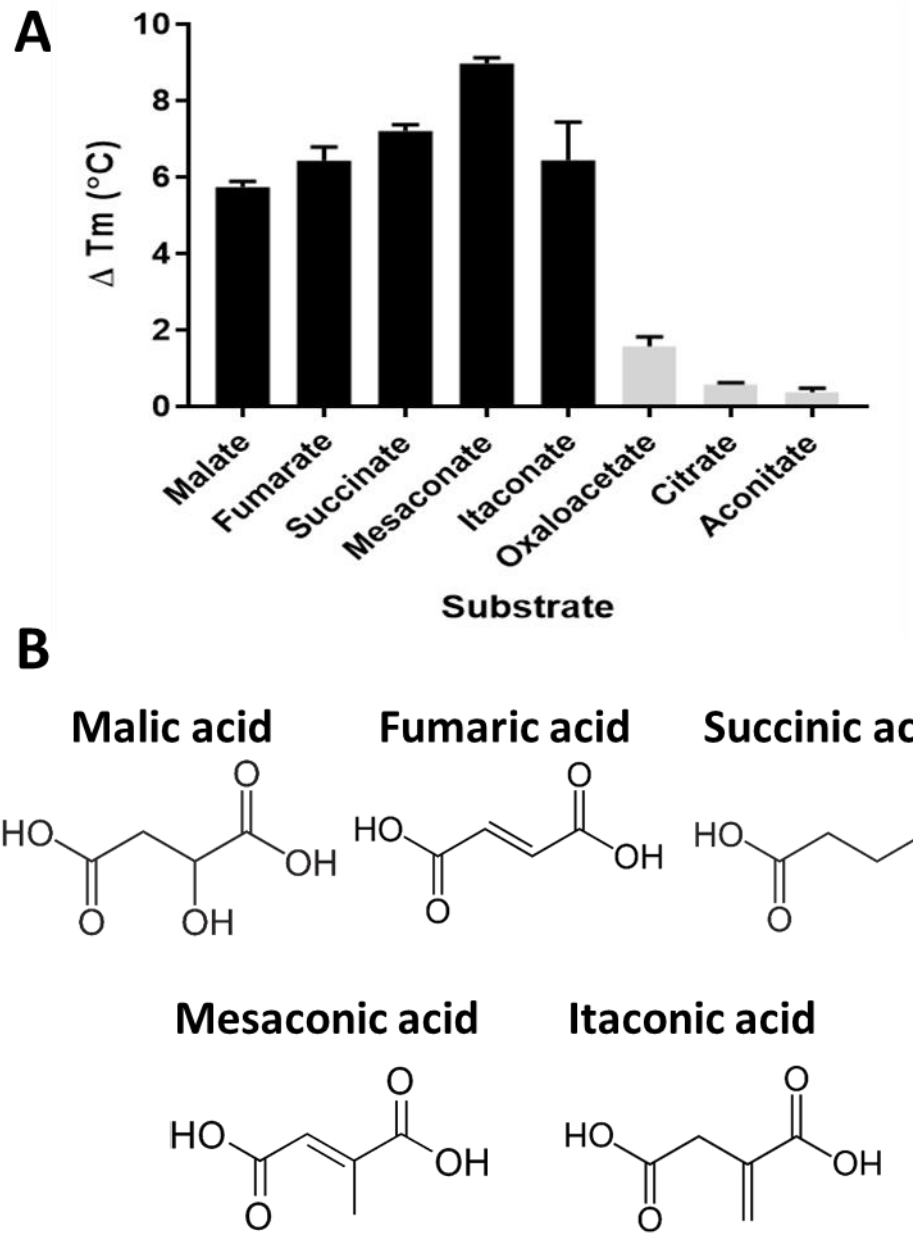


Figure 74: Thermal shift assays show C4/C5-dicarboxylate binding by RPA2047. A) The increase in melting temperature of RPA2047 when mixed with five dicarboxylates, at a final concentration of 60 µM, during thermal shift assay screening indicates binding of malate, fumarate, succinate, mesaconate and itaconate, a substrate specificity range archetypal of DctP proteins. Another C4-dicarboxylate oxaloacetate showed a slight thermal shift, while the shift with other TCA cycle intermediates was insignificant. B) The chemical structures of RPA2047's proposed substrates.

5.2.14. Binding affinity of RPA2047 determined by tryptophan fluorescence

The thermal shift screening assay had clearly indicated that RPA2047 bound several dicarboxylates. The four-carbon dicarboxylates malate, fumarate and succinate were the substrates shown to bind the first characterised TRAP SBP, DctP from *Rhodobacter capsulatus*, which the DctP-family of proteins is named after (Forward et al., 1997). RPA2047 shows a high degree of sequence homology to DctP from *R. capsulatus* with 65% identity. In order to provide more quantitative data on RPA2047's binding characteristics with these substrates, the purified protein was subsequently used in tryptophan fluorescence titration assays.

RPA2047 only contains two tryptophan residues and its intrinsic fluorescence level is relatively low. With an excitation wavelength of 280 nm, RPA2047 showed a maximal fluorescent emission at 330 nm. When an excess of each of the five dicarboxylates shown above was added to a 0.2 µM solution of the protein, a large quench in this fluorescence was seen confirming binding as can be seen in Figure 75. All the candidate substrates elicited at least a 10% decrease in fluorescence, with fumarate showing the largest proportional quench of ~33%.

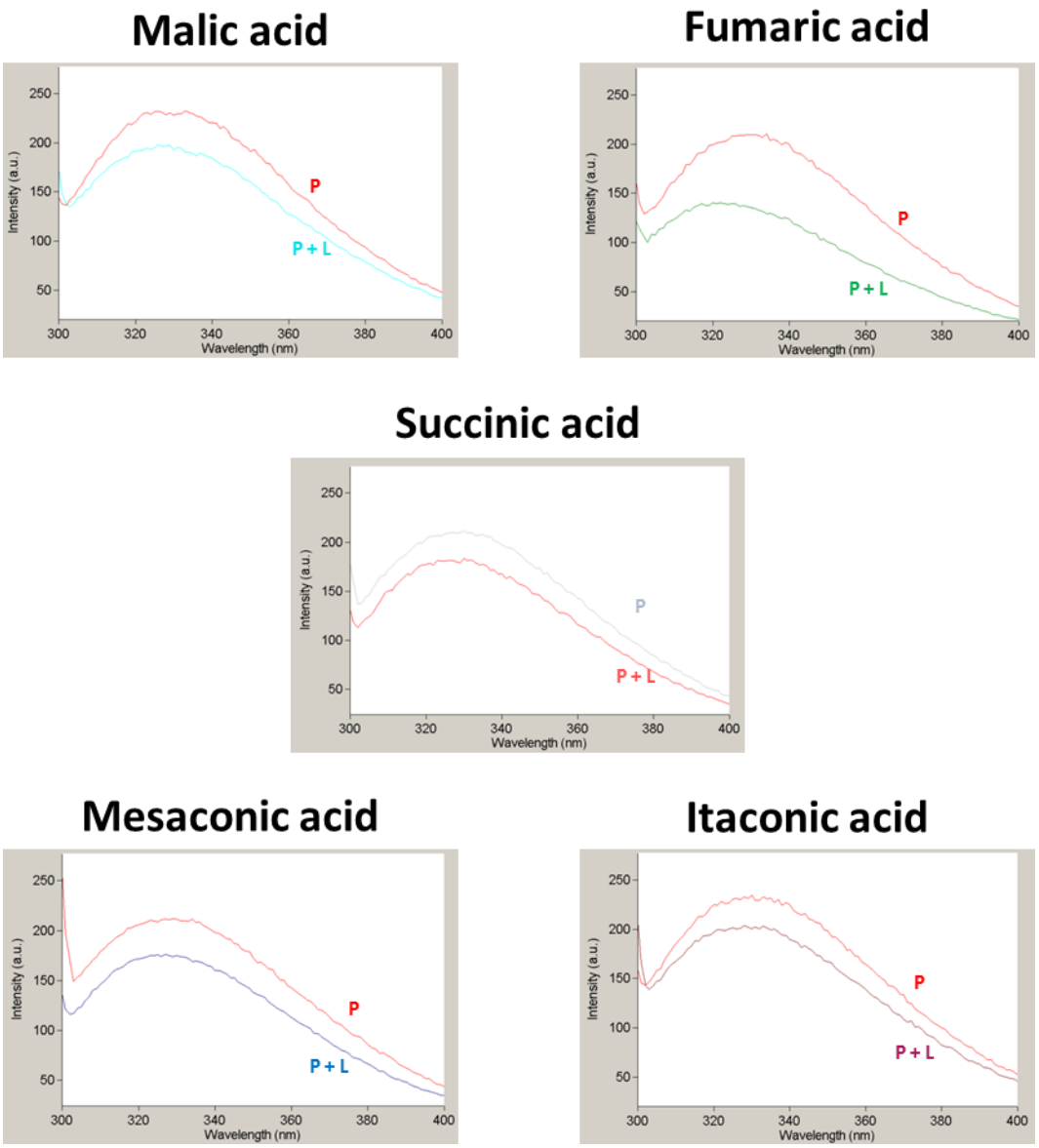


Figure 75: Intrinsic fluorescence quenching effect of addition of ligands to RPA2047. The five candidate ligands (10 µM) were added to a 0.2 µM solution of RPA2047 in buffer and the intrinsic fluorescence of the protein measured. The solution was excited at 280 nm and a spectrum of fluorescence emission measured, with maximal emission for RPA2047 at 330 nm. Each substrate showed a > 10% quench in fluorescence at 330 nm, indicating binding. **P** indicates the fluorescence emission of the ligand-free protein solution; **P + L** indicates the fluorescence emission of the solution after the ligand had been added to the protein.

The fluorescence emission of an RPA2047 solution was then measured while each dicarboxylate was independently titrated in from a concentration of 10 nM up to 5 µM. The decreases in fluorescence were used to plot substrate-concentration-dependent binding curves; a protein-only control whereby buffer was added in identical volumes to the substrates was used as a baseline with the change subtracted from the change shown by each substrate. The K_d for each substrate was determined by iteratively fitting the fluorescence quenches to the quadratic equation for tight-binding. This was repeated for each substrate to produce data in triplicate, and the values are given in Table 10 and representative fluorescence titration curves shown in Figure 76.

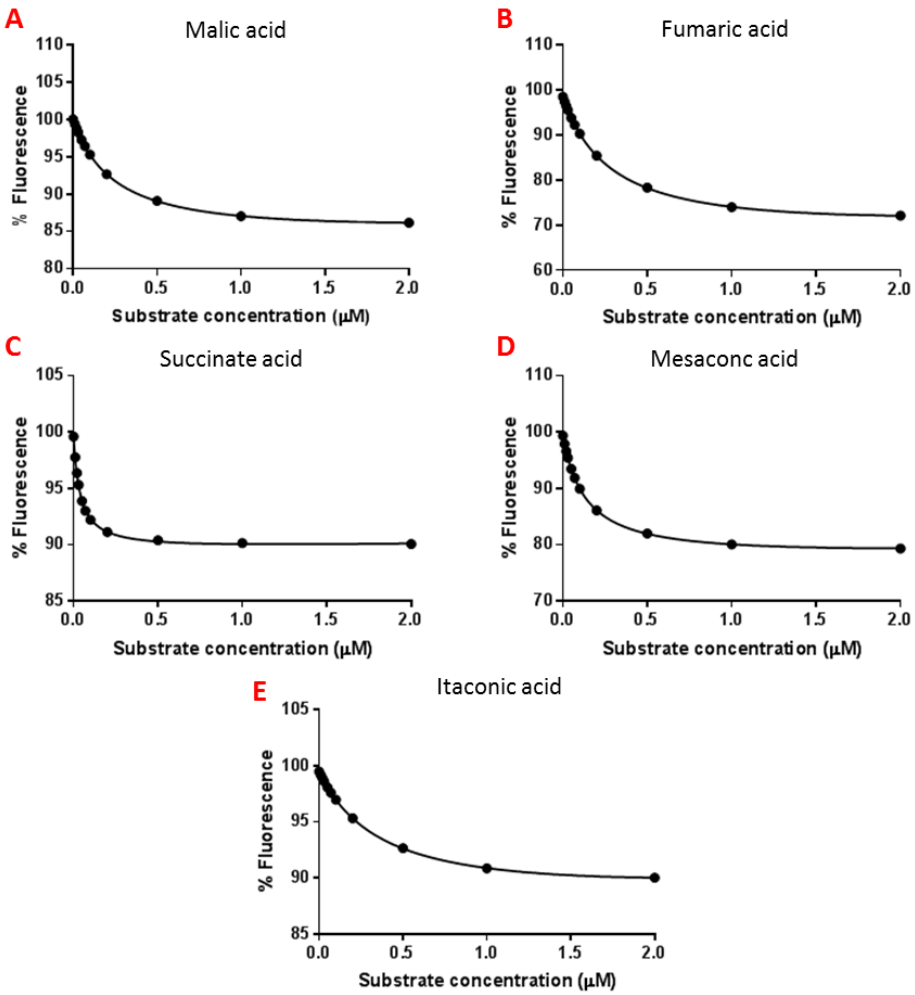


Figure 76: Fluorescence titrations of RPA2047 with dicarboxylates substrates. The intrinsic fluorescence of RPA2047 was measured at 330 nm while malate (A), fumarate (B), succinate (C), mesaconate (D), and itaconate (E) were titrated in. Addition of each substrate elicited a concentration-dependent quench of fluorescence. Shown are the calculated values for one representative titration experiment, with a non-linear fitting for single-site saturation binding.

Substrate	K_d (nM)
Malate	**93.7** ± 6.51
Fumarate	**60.4** ± 38.9
Succinate	**2.80** ± 1.46
Mesaconate	**37.1** ± 4.30
Itaconate	**24.2** ± 12.1

Table 10: Binding characteristics of RPA2047 determined by tryptophan fluorescence titrations. Five carboxylates identified as substrates for RPA2047 were titrated against the protein and the quench in intrinsic fluorescence measured. K_d values are given as a mean of three independent experiments, with the standard deviation shown.

These values show that RPA2047 shows binding affinities for its substrates in the same range as many other TRAP SBPs. The highest specificity was for succinate, with a K_d of just 2.80 nM, indicating very tight binding, while the other four-carbon dicarboxylates appear to exhibit less tight binding. Fumarate is the closest structurally to succinate with no side groups on C2 and C3, but showed a >20-fold higher K_d; however, the replicates for fumarate were highly variable. RPA2047 showed a similar affinity for the two branched isomers mesaconate and itaconate, though the data for the latter was highly variable.

5.2.15. Attempted characterisation of RPA3458, RPA4510 and RPA4556

Characterisation of the three other *R. palustris* TRAP SBPs was attempted through the same approach as with RPA1975 and RPA2047. RPA4510 was cloned using the pET21a(+) vector, while RPA3458 and RPA4556 were cloned using the pET22b(+) vector for periplasmic expression as they contain two and three cysteine residues, respectively. The *R. palustris* signal sequence was not included and a C-terminal polyhistidine tag was incorporated for purification.

During production trials temperature, IPTG concentration and production time were varied. All three showed high production yields but low solubility; lowering the incubation temperature did not improve solubility. Figure 77A+B shows that purification of RPA3458 was achieved, though the yield was low (2 mg l^{-1} culture). Purification of RPA4510 was not successful, with a high abundance of non-specific *E. coli* proteins relative to the yield of recombinant RPA4510 obtained. Purification of RPA4556 resulted in a high yield of recombinant protein, though the SDS-PAGE showed significant impurities including similar non-specific proteins that co-eluted from the column with RPA4510. A potential

explanation of this occurrence is that both RPA4510 and RPA4556 appeared to elute from the nickel column at a lower imidazole concentration than other polyhistidine-tagged recombinant proteins, resulting in the *E. coli* proteins that bound to the column co-eluting as they had not been sufficiently washed off. With RPA4556 the eluted fractions from the higher imidazole concentrations only were combined and concentrated using a 30,000 kDa MW-cut off spin column to remove the smaller protein impurities.

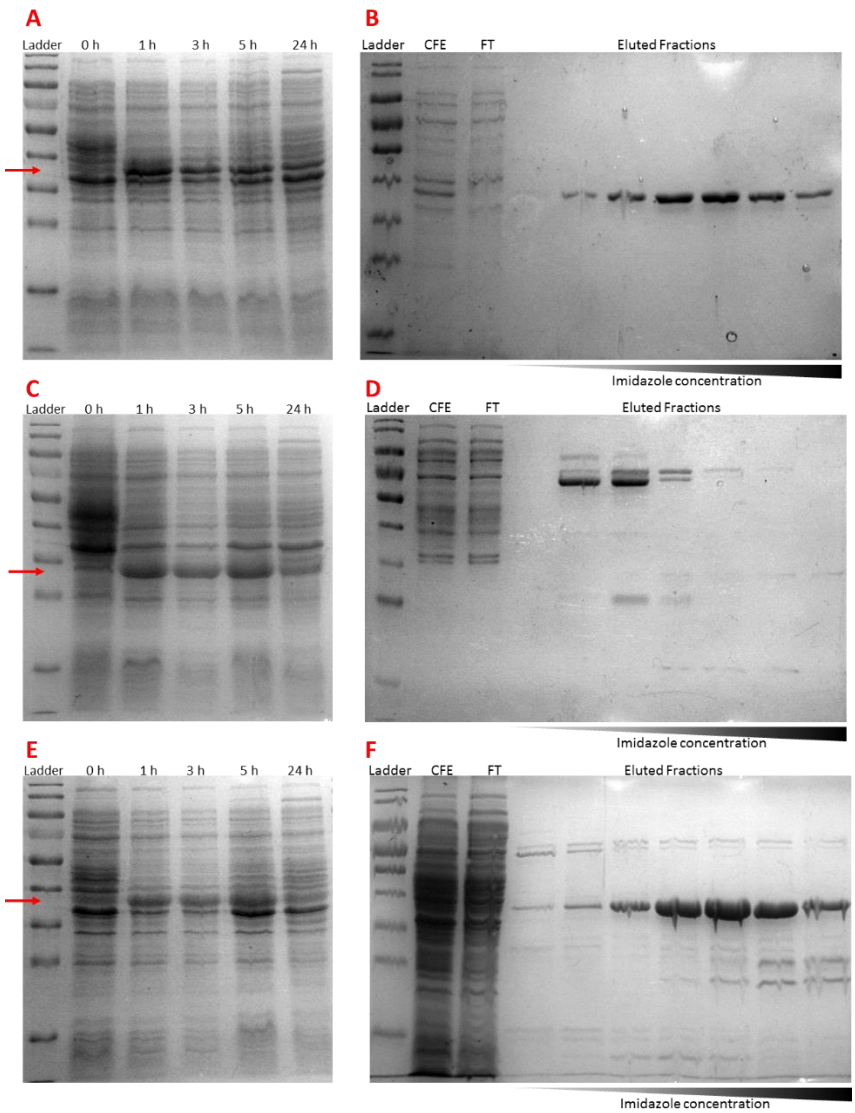


Figure 77: Attempted purification of recombinantly produced RPA3458, RPA4510 and RPA4556. A) SDS-PAGE of whole-cell samples showing production of ~40 kDa RPA3458 at 37 °C with induction by 0.1 mM IPTG. B) SDS-PAGE showing purification of RPA3458 using these conditions; a soluble band is clear in the cell-free extract (CFE) that was purified cleanly using an imidazole gradient. C) SDS-PAGE of whole-cell samples showing production of ~31 kDa RPA3458 at 37 °C with induction by 0.1 mM IPTG. D) SDS-PAGE showing the attempted purification of RPA4510. While RPA4510 had proved to be soluble during trials, only a very faint band was visible at ~ 31 kDa following elution from the column; the majority of the protein appears to be non-specific *E. coli* proteins ~65 – 72 kDa. E) SDS-PAGE of whole-cell samples showing production of ~37.5 kDa RPA3458 at 37 °C with induction by 0.4 mM IPTG. F) SDS-PAGE showing attempted purification of RPA4556 using these conditions; while a reasonable yield of RPA4556 was achieved there are clearly

mainly non-specific *E. coli* proteins being eluted simultaneously. The red arrows indicate the size of the respective proteins.

The purified RPA3458 and RPA4556 were both screened using thermal shift assays, as described previously in this chapter, against a range of compounds but no binding was indicated with any of these potential substrates. The compounds screened were:

- Dicarboxylates: oxylate, malate, fumarate, succinate, adipate, pimelate.
- Amino acids: glutamate, aspartate, leucine, isoleucine, tryptophan.
- Other carboxylates: acetate, lactate, pyruvate, citrate, propionate, butyrate.
- Aromatic compounds: cinnamate, coumarate, ferulate, caffeate, benzoate,
- Others: biotin, taurine, ascorbate, fucitol, mannitol.

Additionally, RPA3458 was screened against malonate and methylmalonate, and RPA4556 against 2-hydroxybutyrate and 3-hydroxybutyrate. These substrates were chosen based on the genes neighbouring the transporters.

Within the time constraints of this study, and with other areas of research prioritised, further attempts to purify RPA4510, improve the yield of purified RPA3458 and RPA4556, and expand the ligand screening process to determine the function of these TRAP SBPs were not attempted.

5.3. Discussion

The genome of the biotechnologically-relevant bacterium *R. palustris* is still relatively poorly annotated, including characterisation of the large array of transporters that its genome encodes. Furthermore, TRAP transporters are under-characterised compared to ABC transporters, despite their occurrence in most bacteria and archaea. In this chapter partial characterisation of two *R. palustris* TRAP transporters has been achieved, RPA1975–RPA1977 and RPA2047–RPA2049, while redundancy in the *R. palustris* genome has been further explored.

The characterisation of RPA1975 showed that this SBP binds medium-chain dicarboxylates, with the highest affinity for pimelate. The results suggest that the carbon-chain length, and thus size, of the dicarboxylate substrate determines the binding affinity. The tight binding of pimelate, and suberate, suggests that the size of RPA1975's binding pocket preferably accommodates binding of C7/C8 dicarboxylates, with substrates shorter or longer in carbon-chain length not capable of forming the optimal interactions with the protein.

As described in Section 1.6.3, the CoA thioester derivative of pimelate, pimeloyl-CoA, is an intermediate in the degradation pathway of aromatic compounds in *R. palustris*; Figure 16 shows the β-oxidation pathway responsible for its metabolism. The pimeloyl-CoA ligase PimA is capable of producing the CoA-thioester derivative of pimelate, and therefore pimelate transported into the cell by RPA1975 can be easily diverted into this metabolic pathway. Suberate and azelate could also be metabolised in a similar fashion as PimA has been shown to have CoA-ligase activity with both of these substrates, and the downstream pathway can act cyclically to remove one acetyl-CoA at a time. Pimelate is also a precursor for biotin synthesis in bacteria; the enzymes BioF (RPA2972) BioA (RPA2970) BioD (RPA2971) and BioB(RPA2045) are responsible for conversion of pimeloyl-CoA and alanine to biotin(Manandhar and Cronan, 2017). Most bacteria have to synthesise pimelate from acetate rather than utilise extracellular pimelate, with no high affinity pimelate uptake systems reported (Cronan and Lin, 2011). Biotin is an essential vitamin that acts as a cofactor for enzymes in fatty acid and amino acid biosynthesis, and gluconeogenesis(Salaemae et al., 2016). It is also used in molecular biology and biotechnology due to its extremely tight binding of avidin and streptavidin (K_d of ~ 10^{-15} M) allowing it to be used to conjugate proteins and as a biosensor. Therefore, there is some interest in microbial production of biotin, and patents for this have been filed in the past (Kishimoto et al., 1993; Kumagai et al., 1995; Revuelta et al., 2016).

In order to complete the comparison of RPA1975 and the ABC transporters, additional binding data and structural determination of the TRAP SBP is required. The persistent low solubility problems experienced have prevented purification of the protein for initiation of crystallography trials, and therefore different protein overproduction methods are required going forward. The use of the pMAL vector here has proved successful in producing MBP-fused RPA1975, though the inability to cleave off the protein fusion could make x-ray crystallography and structural determination more difficult. The linking region of the fusion is flexible, and therefore the MBP and the RPA1975 domains could be unstable and moving relative to each other. Nevertheless, structural determination of MBP-fused proteins has been achieved with 102 fusion structures deposited in the PDB as of March 2016, with MBP-mediated structural distortions rare (Waugh, 2016). While with the majority of these structures the target protein is less than 150 amino acids in size, 10 of these structures were over 300 amino acids. Crystallisation trials are currently ongoing using the purified MBP:RPA1975 fusion.

One potential solution would be to overexpress the *rpa1975* gene using the pBBRBB vector utilised for *couAB* overexpression in Chapter 3. This vector can be used for incorporation of a C-terminal polyhistidine tag during production of the SBP to allow for purification. Large-scale photoheterotrophic growth of *R. palustris* to the degree required to achieve sufficient yield has not been attempted in this laboratory and would have to be optimised. However, solubility of the produced RPA1975 should be improved compared to the recombinant protein produced in *E. coli*, while induction using arabinose would allow for fine-tuning of expression for optimal yield.

The identification of the range of dicarboxylates as the substrates for RPA1975 is consistent with TRAP SBP's propensity for binding compounds with carboxylic acid moieties. It also appeared to be another example of *R. palustris* possessing both a TRAP and an ABC transport system for the same range of substrates, as with TarPQM and CouPSTU, based on previous screening of the three ABC SBPs RPA3723, RPA3724 and RPA3725 (Giuliani et al., 2011; Salmon et al., 2013). Ligand-screening assays conducted in this chapter reinforced the evidence that each of these ABC SBPs has a different binding affinity range despite their similar sequence identities. Unfortunately, further biochemical analysis to confirm binding characterisation of the three ABC SBPs has not been successful so far. Use of isothermal titration calorimetry (ITC) with all three proteins could be used to provide thermodynamic characterisation of binding for a comprehensive comparison of binding affinities between the ABC SBPs, and with RPA1975. If crystal structures for RPA3723, RPA3724 and RPA3725

are obtained then a comparison of the sizes and arrangements of their binding sites might explain the differing affinities for different length dicarboxylic and fatty acids.

A proteomics study conducted in lab by Abrar Akbar (University of Sheffield) studied *R. palustris*'s proteomic response to growth on different aromatic compounds. Interestingly, both RPA1975 and the two ABC transporter NBDs, RPA3719 and RPA3720, showed an increase in abundance during growth on 5-phenylvalerate compared to benzoate. Based on these results, characterisation of the three ABC SBPs was also attempted with phenylalkane carboxylate substrates, such as 4-phenylbutyrate, 5-phenylvalerate, 6-phenylhexanoate, 7-phenylheptanoate and 8-phenyloctanoate. Thermal shift assays indicated binding of all of these substrates by RPA3723, with RPA3724 also potentially binding 7-phenyheptanoate. However, fluorescence titration attempts were hampered by the same problems encountered in this chapter. The results of her work suggest that RPA3723 is capable of binding to this range of substrates; these compounds consist of a six-carbon aromatic ring joined to an aliphatic carboxylic chain via a keto group and it is this aliphatic moiety (that resembles the dicarboxylate substrates binding assays in this chapter) that the protein likely binds. This is more evidence that the three ABC SBPs bind different carboxylic substrates to noticeably different degrees, despite their high sequence identity. Unfortunately the inability to re-purify RPA1975 has prevented assessment of phenylalkane carboylates as potential substrates for RPA1975.

While the omics-based study into *R. palustris* conducted by Pan et al. (2008) focused on the genes identified as involved in the metabolism of coumarate, their expansive dataset did include data on the transporters discussed in this chapter (Pan et al., 2008). RPA1975 was shown to be more abundant in *R. palustris* grown on succinate than benzoate, which corroborates the binding data here that showed no binding of aromatic compounds in the thermal shift assays. RPA3723 was not detected in the proteomics study, but RPA3724 and RPA3725 were found to be 6- and 26-fold more abundant on succinate than benzoate, with transcriptomics data reinforcing this relative upregulation of the genes encoding these two SBPs and RPA3719, RPA3721 and RPA3722. The IclR-family regulator RPA3718 showed similar regulation, suggesting that its transcription is repressed by growth on benzoate compared to succinate, reducing the effect of its function as an activator for the transcription of the ABC transport genes. In contrast the *pim* cluster genes encoding the β-oxidation pathway enzymes were regulated divergently and were positively regulated by benzoate, presumably due to the requirement to metabolise the pimelyl-CoA derived from degradation of the aromatic ring. RPA3724 was also shown to be upregulated during

growth on coumarate compared to benzoate, based on the transcriptomic and proteomic data. Thermal shift assays show no binding of either of these compounds, supporting the theory that benzoate may cause a down-regulation of these genes.

The pimelyl-CoA ligase PimA actually showed significantly less activity with pimelate than the longer-chain dicarboxylates (Harrison and Harwood, 2005). The binding data for the ABC SBPs produced in this chapter support what was previously shown by Giulani et al (2011), with the higher affinity for longer-chain substrates matching the activity of PimA (Giuliani et al., 2011). The role of this transport system is likely to scavenge any carboxylic substrate available to feed in into the pathway catalysed by the *pim* operon gene products, ultimately as sources of carbon.

The RT-PCR analysis of regulation of these SBPs (Figure 70) did not show the expected results. Regulation of the TRAP transporter gene appears to be independent of growth on pimelate, while the expression levels of *rpa3724*, which should be indicative of the complete set of ABC transporter genes, indicated its expression is significantly higher on succinate compared to pimelate and sebaceate (to both of which the thermal-shift assays suggest it binds). The RNA analysed was isolated from *R. palustris* cells grown with an excess of the carbon sources investigated; these SBPs have a high affinity for their substrates and therefore may only be upregulated when extracellular concentrations are low. This suggests that there may be alternative transport systems – potentially SBP-independent permeases – in *R. palustris* that are capable of mediating uptake of pimelate and other dicarboxylate substrates. The lack of a significant growth phenotype with pimelate as the carbon source with the transporter gene knock-out mutants (Figure 69) does indicate this could be the case. Regulation of these transport systems could also be controlled at post-transcriptional level, with *R. palustris* known to employ post-translational modifications such as N-lysine acetylation, ADP-ribosylation and phosphorylation as means of regulation (Crosby et al., 2012; Heiniger et al., 2012).

BLAST searching reveals that closest homolog of RPA1975 is an uncharacterised protein annotated as an ABC transporter SBP in *Bradyrhizobium* species, a genus of soil bacteria fairly closel related to *R. palustris*. Analysing the genomic region around the homolog in *Bradyrhizobium japonicum*, which is 88% identical to RPA1975, reveals that this binding protein appears to be an orphan, with no neighbouring membrane proteins. However, three of its neighbouring genes are homologs of RPA1974 (a MRP ATP-binding protein), RPA1973 (an isobutyrate dehydrogenase) and RPA1972 (a histidine kinase sensor), as

shown in Figure 78. SBPs of transport systems are known to act as sensors in signalling pathways and interact with histidine kinases giving them a dual function (Piepenbreier et al., 2017). It is a possibility these SBPs here are involved in an extracytoplasmic pimelate-triggered signal.

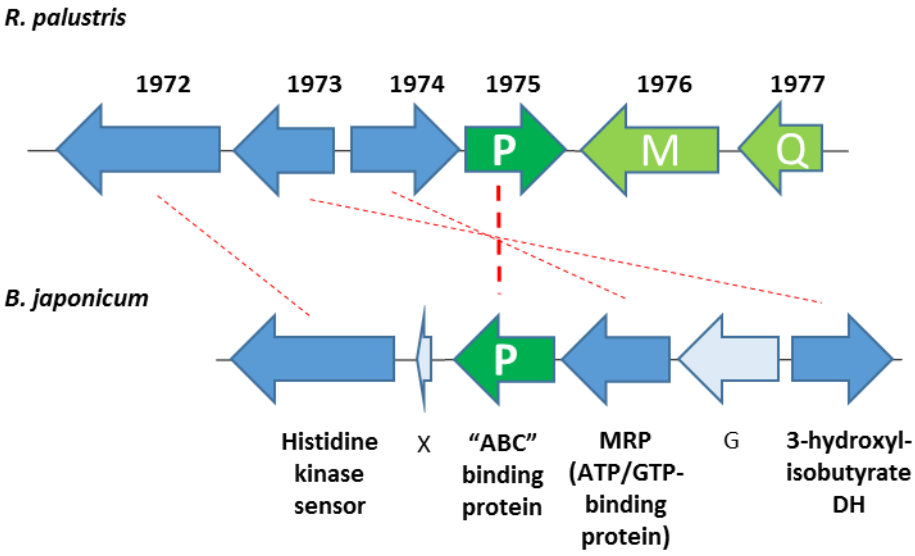


Figure 78: Comparison of RPA1975 genomic region and its homolog in *B. japonicum*. The outlines of the genes surrounding RPA1975 in *R. palustris* and its homolog in *B. japonicum* identified by BLAST searching. The red dotted lines indicate homology in the gene products. X is a small, hypothetical protein; G is a glyoxalase.

Due to the location of the ABC transport genes next to the *pim* operon, it may have been expected that pimelate would be the major substrate of the three ABC SBPs; however, it is clear that their substrate range is broad, with no data suggesting the tightest binding is with pimelate. In contrast, RPA1975 has shown the greatest affinity for pimelate and its binding appears more specific. Therefore it is proposed that the RPA1975-RPA1977 transporter be designated PimPQM, based on this substrate specificity and the conventional nomenclature of TRAP transporter subunits (RPA1975, as the pimelate-binding periplasmic protein subunit, PimP; RPA1977 as the small transmembrane subunit, PimQ; and RPA1976 as the large transmembrane subunit, PimM).

The characterisation of RPA2047 showed its binding affinity range to match that of the first characterised TRAP SBP, DctP from *R. capsulatus.* Considering this, and sequence homology, it is proposed that RPA2047, RPA2048 and RPA2049 be designated DctP, DctQ and DctM, respectively. DctPQM homologs exist in many bacteria and archaea and are one of the most studied TRAP transport systems (Fischer et al., 2010). Many structures of these

homologs have been deposited in the PDB, including some with succinate bound, and therefore no further structural studies for RPA2047 are planned. There is also likely redundancy in the *R.* palustris genome for uptake of C4-dicarboxylates; work conducted by Leonardo Rosa (University of Sheffield) has shown a TTT SBP (RPA3494) that binds these substrates including malate and fumarate. Therefore, it is unlikely that a *Δ2047 R. palustris* mutant would exhibit a growth phenotype with these substrates as sole carbon sources. The study screening *R. palustris*'s ABC SBPs only detected one protein that showed a thermal shift with one of these substrates; this was RPA2043 that showed a 4 °C shift with malate (Giuliani et al., 2011). While they designated their own data confidence for this binding as "Low", the location of the gene encoding this protein and its proximity to RPA2047 suggests a gene cluster/operon for uptake of malate. No binding of succinate, fumarate, mesaconate or itaconate was detected with RPA2043, or any other ABC SBP.

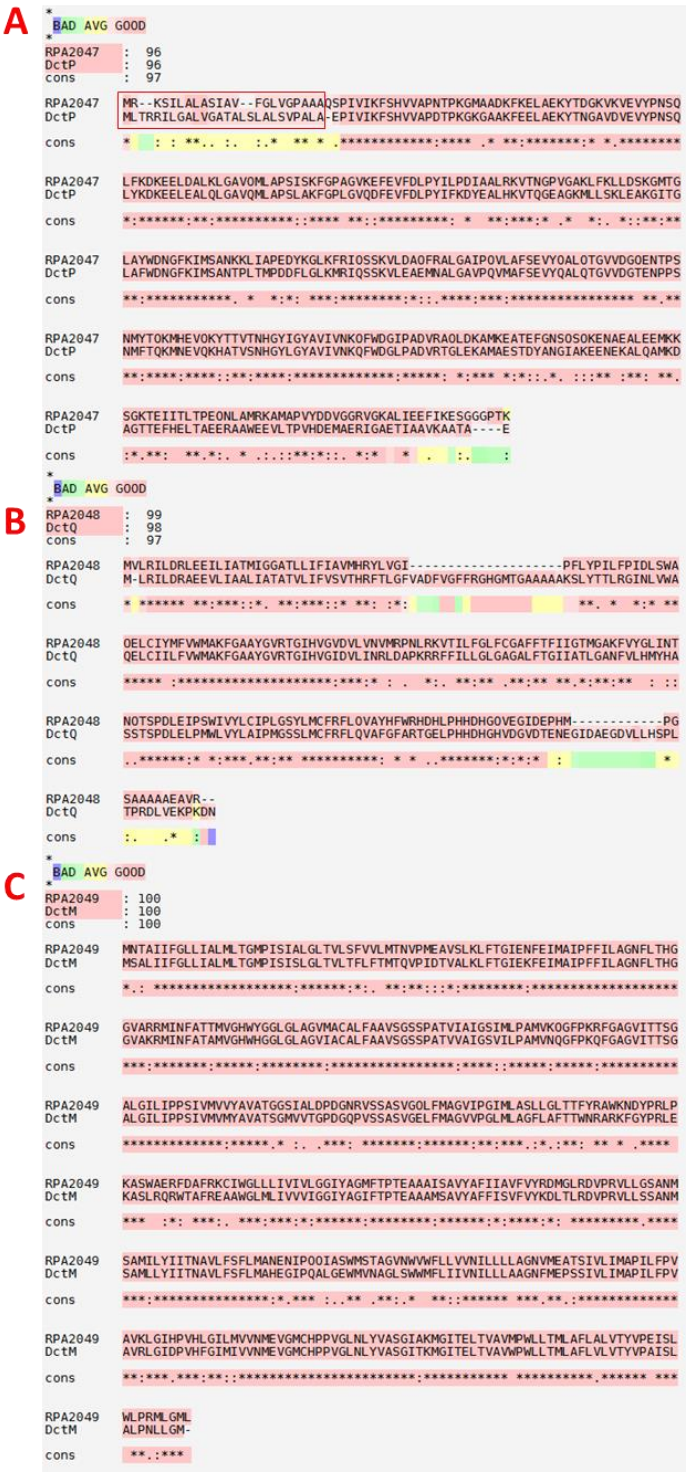


Figure 79: Sequence alignment of RPA2047-2049 with DctPQM from *R. capsulatus*. The amino acid sequences compared using the T-coffee sequence alignment server for A) RPA2047 and DctP; B) RPA2048 and DctQ; and C) RPA2049 and DctM (Di Tommaso et al., 2011). The high proportion of identical and conserved residues highlights the homology. The red box highlights the N-terminal signal sequences of the SBPs.

As the experimental programme focused on RPA1975 and the comparison with the three ABC SBPs, and RPA2047, attempts to improve the purification of the other three TRAP SBPs were limited in comparison. A clear path of future research would be to complete characterisation of *R. palustris*'s range of TRAP transporters through focus on RPA3458, RPA4510 and RPA4556, as they may prove capable of binding novel substrates while offering further insight into the diverse roles and evolutionary origins of TRAP SBPs. It is noteworthy that the study by Pan et al. (2008) showed transcriptomic and proteomic evidence that RPA3458 (and RPA3459) was upregulated during growth on succinate relative to coumarate and benzoate, while RPA4556 was more abundant during growth on coumarate than succinate; however, no binding of these respective compounds was identified in the thermal-shift assays investigated here (Pan et al., 2008).

Despite homology between TRAP binding proteins being generally low, it is well established that a conserved arginine residue – R147 – in TRAP SBPs is responsible for forming a salt-bridge with the carboxyl group of its substrate. The DctP-family TRAP SBPs studied in this chapter all possess this residue, with the SBPs, RPA1975, RPA2047 and RPA4556 all having high homology in this region of their amino acid sequences. Multiple alignment of these SBPs with TarP and the two most comprehensively studied TRAP DctP-family SBPs, DctP from *R. capsulatus* and SiaP, from *Haemophilus influenza,* shows a conserved $DX_3GLK\Phi R$ motif (Figure 80).

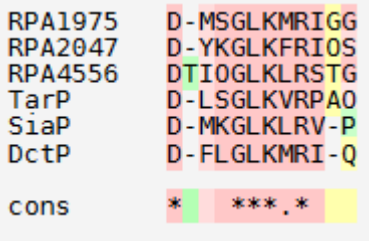


Figure 80: Conserved motif of DctP-family SBPs. Multiple sequence alignment of the amino acid sequence of DctP-family SBPs from *R. palustris* studied in this chapter, the aromatic-binding *R. palustris* TRAP SBP TarP, and two other well characterised DctP-family SBPs, in the region of the conserved arginine. This suggests a conserved motif of $DX_3GLK\Phi R$.

The TAXI-family SBPs, RPA3458 and RPA4510, do not share this motif, nor do they appear to possess the conserved arginine residue. Multiple sequence alignment analysis of these two *R. palustris* proteins and three other TAXI SBPs from different species shows a reasonable degree of homology between TAXI SBPs, with two highly conserved motifs

(Figure 81A). However, analysis of the structure of the *Thermus thermoophilus* TAXI SBP, TT1099 (PDB 1US5), reveals these areas of conserved residues are on the periphery of the protein structure and not involved in binding of the glutamate substrate (Figure 81B); analysis of the binding site shows no involvement of any arginine residues (Takahashi et al., 2004).

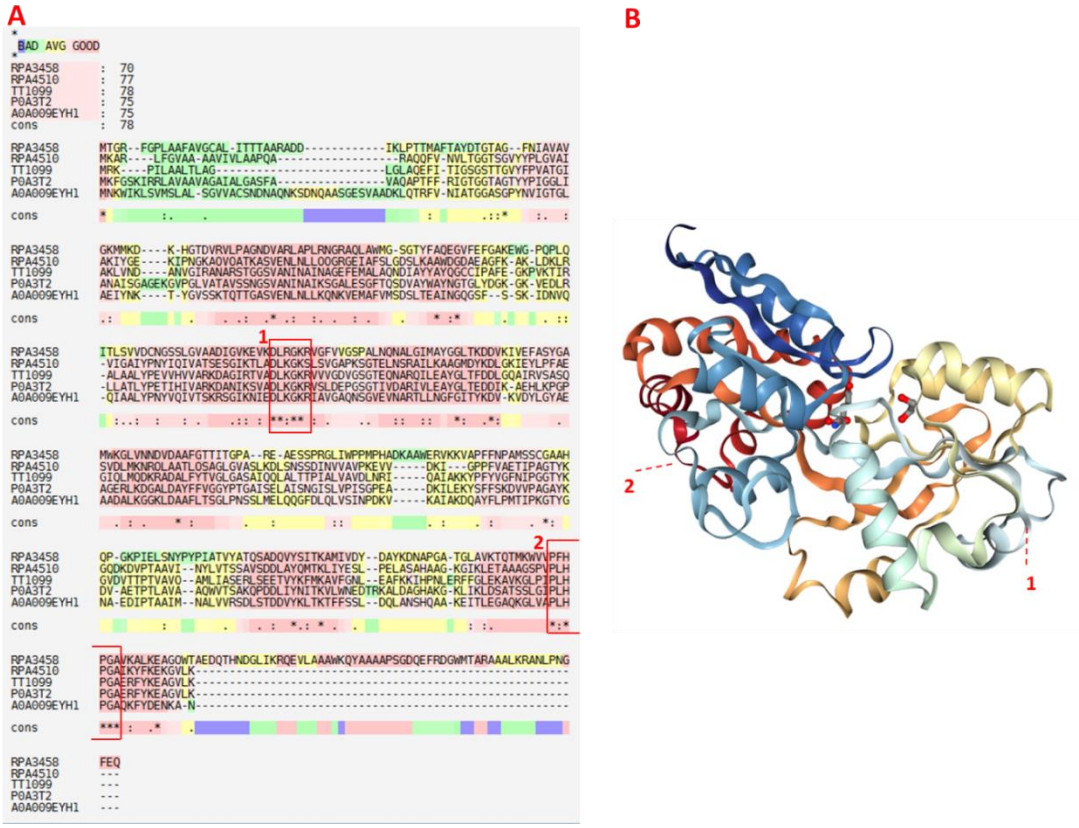


Figure 81: Analysis of TAXI-family SBPs. A) Multiple sequence alignment using the T-coffee webserver of the TAXI-family TRAP SBPs RPA3458 and RPA4510 from *R. palustris,* TT1099 from *T. thermophilus*, P0A3T2 from *Brucella melitensis*, and A0A009EYH1 from *Acinetobacter baumannii* (Di Tommaso et al., 2011). No conserved arginine residue was detected, though two highly-conserved motifs were identified and are highlighted in red boxes (designated *1* and *2*). B) The structure of TT1099 that is deposited in the PDB database. The dashed lines indicate approximate locations of the highly-conserved motifs. Bound-glutamate is shown in the centre of the structure.

The TAXI-family of TRAP SBPs are clearly characterised less extensively than their DctP-family relatives. Limited time did not allow further work to be conducted on the characterisation of RPA3458 and RPA4510, but if high yields of soluble protein could be

achieved then a binding and structural analysis would provide interesting information on the differences between the two families of TRAP transporters in bacteria.

6. Conclusions and planned future research

The primary aim of this work was to use microbial physiology approaches for the purpose of industrial applications with regards to the use of lignin. Central to this was exploring means of enhancing the microbial production of vanillin and increasing the understanding of the physiological capabilities of *R. palustris*.

Previous studies on the metabolism of hydroxycinnamates in *R. palustris* had indicated that the enzymes of the CouR-regulated gene cluster would be capable of catalysing the biotransformation of ferulate to vanillin shown in other species. Therefore, it was expected that these enzymes could be utilised as a basis for creating a recombinant strain of *E. coli* that could efficiently produce vanillin, with this process being enhanced by the additional engineering of increased ferulate uptake. In addition, investigating vanillin toxicity in *E. coli* would help work towards a robust vanillin-producing microbial platform. Ultimately the attempts to utilise the *R. palustris* enzymes failed as shown by the NMR data, but did unveil an apparent novel characteristic of the CouA protein. This resulted in a modified pathway for metabolism of hydroxycinnamates in *R. palustris* being proposed, but suggests that neither *R. palustris*, nor recombinant production of its enzymes, is capable of being harnessed for vanillin production. While more work is needed to confirm parts of this pathway, such as characterisation of RPA4198, these findings do add depth to the understanding of the metabolism of lignin-derived compounds in bacteria such as *R. palustris*.

Clear outcomes that should enhance microbial vanillin production were achieved through the investigation of vanillin's toxicity. The proposed means by which vanillin tolerance can be engineered into *E. coli* could be applied to other microbial species being harnessed for production of vanillin, such as *Pseudomonas* sp., *Streptomyces* sp. and *Amycolatopsis* sp., all of which have been shown to utilise ferulate, and other lignin- and lignocellulose-derived compounds, as substrates. The antimicrobial effects of aldehydes, and aromatic aldehydes, are relatively poorly understood and biotransformation of other compounds such as benzaldehyde and cinnamaldehyde could be enhanced by these findings. It is worth noting that the novel activity of CouA discovered would not prevent the use of *R. palustris* or recombinant CouA and CouB in the biotransformation of lignin-derived cinnamate to benzaldehyde, the global market of which is expected to continually grow due to increased demand. These data regarding the mechanisms of vanillin's toxicity can also be a useful

resource for the characterisation of the antibiotic properties of essential oils and other natural products at a molecular and physiological level.

While useful in their own right, these data offer the foundation for a wide-range of follow up investigations. Potential means by which vanillin exposure results in the increase in production of ROS have been proposed, such as increased respiration or through its interactions with metal ions, but this needs to be clarified. An in-depth analysis of the TCA cycle and NAD(P)H formation in relation to vanillin toxicity might provide an answer, as a similar approach has shown bactericidal antibiotics' effectivity is related to ROS produced from increased respiration (Kohanski et al., 2007). Putative vanillin-efflux systems have been identified such as AcrD and AaeAB, specific investigations into their role in vanillin export should be conducted. A caveat in the methodology of the proteomic analysis was that vanillin was added exogenously, while a more appropriate approach might be to assess changes in the proteome in response vanillin being produced intracellularly, as this could be more representative of its toxic effects in a biotechnological application. Furthermore, a microarray or RNAseq approach would provide complementary data on gene expression levels in response to vanillin to provide an even more comprehensive understanding.

R. palustris is a biotechnologically-relevant bacterium with its use in areas such as biohydrogen production and bioremediation in addition to its metabolism of lignin-derived aromatics. Furthermore, efforts are being made to engineer recombinant strains of *R. palustris* as a biotechnological platform for biofuel production (Doud et al., 2017). A greater understanding of its overall physiology will facilitate its industrial application going forwards, with exploitation of its extensive transportome key to this. While novel transporters of aromatic compounds were not identified, characterisation of two transport systems was achieved. Further effort into characterisation of RPA3458, RPA4510 and RPA4556 is desirable to provide a complete view of *R. palustris*'s TRAP transporter capabilities.

This study also identified further examples of metabolic redundancy in *R. palustris*. Efforts to investigate differential regulation of different transporters that appear to uptake the same substrates could provide an insight into the physiological importance of, and evolutionary basis for, TRAP transporters. An expanded structural study into the binding pockets of SBPs that bind the same substrate with different affinities could provide insight into the relationship between protein structure and specificity.

In conclusion, this study has: discovered *R. palustris* enzymes cannot be used for the microbial production of vanillin, while suggesting modifications to the currently accepted metabolic pathway of lignin-derived aromatic compounds; provided clear means by which to engineer a vanillin-tolerant strain of *E. coli* through investigating its toxicity; and further characterised *R. palustris*'s transport capabilities, expanding the knowledge of this biotechnologically-relevant bacterium's physiology.

7. References

Achterholt, S., Priefert, H., and Steinbuchel, A. (2000). Identification of *Amycolatopsis* sp. strain HR167 genes, involved in the bioconversion of ferulic acid to vanillin. Appl Microbiol Biotechnol *54*, 799-807.

Adebusuyi, A.A., Smith, A.Y., Gray, M.R., and Foght, J.M. (2012). The EmhABC efflux pump decreases the efficiency of phenanthrene biodegradation by *Pseudomonas fluorescens* strain LP6a. Appl Microbiol Biotechnol *95*, 757-766.

Agrawal, A., Kaushik, N., and Biswas, S. (2014). Derivatives and Applications of Lignin – An Insight. The Scitech Journal *Online*.

Ahmad, M., Roberts, J.N., Hardiman, E.M., Singh, R., Eltis, L.D., and Bugg, T.D. (2011). Identification of DypB from *Rhodococcus jostii* RHA1 as a lignin peroxidase. Biochemistry *50*, 5096-5107.

Akiyama, N., Takeda, K., and Miki, K. (2009). Crystal structure of a periplasmic substrate-binding protein in complex with calcium lactate. J Mol Biol *392*, 559-565.

Alejandro-Marin, C.M., Bosch, R., and Nogales, B. (2014). Comparative genomics of the protocatechuate branch of the beta-ketoadipate pathway in the *Roseobacter* lineage. Marine Genomics *17*, 25-33.

Alekshun, M.N., and Levy, S.B. (2007). Molecular mechanisms of antibacterial multidrug resistance. Cell *128*, 1037-1050.

Anfelt, J., Hallström, B., Nielsen, J., Uhlén, M., and Hudson, E.P. (2013). Using Transcriptomics To Improve Butanol Tolerance of *Synechocystis* sp. Strain PCC 6803. Appl Environ Microbiol *79*, 7419-7427.

Ankarloo, J., Wikman, S., and Nicholls, I.A. (2010). *Escherichia coli mar* and *acrAB* mutants display no tolerance to simple alcohols. Int J Mol Sci *11*, 1403-1412.

Anne, J., Vrancken, K., Van Mellaert, L., Van Impe, J., and Bernaerts, K. (2014). Protein secretion biotechnology in Gram-positive bacteria with special emphasis on *Streptomyces lividans.* Biochim Biophys Acta *1843*, 1750-1761.

Antoine, R., Huvent, I., Chemlal, K., Deray, I., Raze, D., Locht, C., and Jacob-Dubuisson, F. (2005). The periplasmic binding protein of a tripartite tricarboxylate transporter is involved in signal transduction. J Mol Biol *351*, 799-809.

Antoine, R., Jacob-Dubuisson, F., Drobecq, H., Willery, E., Lesjean, S., and Locht, C. (2003). Overrepresentation of a Gene Family Encoding Extracytoplasmic Solute Receptors in *Bordetella*. Journal of Bacteriology *185*, 1470-1474.

Ashwini, R., Vijayanand, S., and Hemapriya, J. (2017). Photonic Potential of Haloarchaeal Pigment Bacteriorhodopsin for Future Electronics: A Review. Curr Microbiol *74*, 996-1002.

Balsera, M., Buey, R.M., and Li, X.D. (2011). Quaternary structure of the oxaloacetate decarboxylase membrane complex and mechanistic relationships to pyruvate carboxylases. J Biol Chem *286*, 9457-9467.

Baqueiro-Peña, I., and Guerrero-Beltrán, J.Á. (2016). Vanilla (*Vanilla planifolia* Andr.), its residues and other industrial by-products for recovering high value flavor molecules: A review. Journal of Applied Research on Medicinal and Aromatic Plants.

Barghini, P., Di Gioia, D., Fava, F., and Ruzzi, M. (2007). Vanillin production using metabolically engineered *Escherichia coli* under non-growing conditions. Microb Cell Fact *6*, 13.

Batista, R. (2014). Uses and potential applications of ferulic acid. In Ferulic acid: antioxidant properties, uses and potential health benefits, B. Warren, ed. (New York, USA: Nova Science Publishers, Inc), pp. 39-70.

Batool, K., Tuz Zahra, F., and Rehman, Y. (2017). Arsenic-Redox Transformation and Plant Growth Promotion by Purple Nonsulfur Bacteria *Rhodopseudomonas palustris* CS2 and *Rhodopseudomonas faecalis* SS5. Biomed Res Int *2017*, 6250327.

Baucheron, S., Tyler, S., Boyd, D., Mulvey, M.R., Chaslus-Dancla, E., and Cloeckaert, A. (2004). AcrAB-TolC directs efflux-mediated multidrug resistance in *Salmonella enterica* serovar *typhimurium* DT104. Antimicrob Agents Chemother *48*, 3729-3735.

Bennett, J.P., Bertin, L., Moulton, B., Fairlamb, I.J., Brzozowski, A.M., Walton, N.J., and Grogan, G. (2008). A ternary complex of hydroxycinnamoyl-CoA hydratase-lyase (HCHL) with acetyl-CoA and vanillin gives insights into substrate specificity and mechanism. Biochem J *414*, 281-289.

Berg, M., Hilbi, H., and Dimroth, P. (1997). Sequence of a gene cluster from *Malonomonas rubra* encoding components of the malonate decarboxylase Na+ pump and evidence for their function. European Journal of Biochemistry *245*, 103-115.

Berntsson, R.P., Smits, S.H., Schmitt, L., Slotboom, D.J., and Poolman, B. (2010). A structural classification of substrate-binding proteins. FEBS Lett *584*, 2606-2617.

Bertz, M., and Rief, M. (2009). Ligand binding mechanics of maltose binding protein. J Mol Biol *393*, 1097-1105.

Bienert, G.P., and Chaumont, F. (2014). Aquaporin-facilitated transmembrane diffusion of hydrogen peroxide. Biochim Biophys Acta *1840*, 1596-1604.

Blair, J.M., Bavro, V.N., Ricci, V., Modi, N., Cacciotto, P., Kleinekathfer, U., Ruggerone, P., Vargiu, A.V., Baylay, A.J., Smith, H.E*., et al.* (2015). AcrB drug-binding pocket substitution confers clinically relevant resistance and altered substrate specificity. Proc Natl Acad Sci U S A *112*, 3511-3516.

Blanchard, J.L., Wholey, W.Y., Conlon, E.M., and Pomposiello, P.J. (2007). Rapid changes in gene expression dynamics in response to superoxide reveal SoxRS-dependent and independent transcriptional networks. PLoS One *2*, e1186.

Bloem, A., Bertrand, A., Lonvaud-Funel, A., and de Revel, G. (2007). Vanillin production from simple phenols by wine-associated lactic acid bacteria. Letters in Applied Microbiology *44*, 62-67.

Bobrov, A.G., Kirillina, O., Fetherston, J.D., Miller, M.C., Burlison, J.A., and Perry, R.D. (2014). The Yersinia pestis siderophore, yersiniabactin, and the ZnuABC system both contribute to zinc acquisition and the development of lethal septicaemic plague in mice. Mol Microbiol *93*, 759-775.

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. Annu Rev Plant Biol *54*, 519-546.

Bokinsky, G., Peralta-Yahya, P.P., George, A., Holmes, B.M., Steen, E.J., Dietrich, J., Lee, T.S., Tullman-Ercek, D., Voigt, C.A., Simmons, B.A.*, et al.* (2011). Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. Proc Natl Acad Sci U S A *108*, 19949-19954.

Bose, A., Gardel, E.J., Vidoudez, C., Parra, E.A., and Girguis, P.R. (2014). Electron uptake by iron-oxidizing phototrophic bacteria. Nat Commun *5*, 3391.

Bose, A., and Newman, D.K. (2011). Regulation of the phototrophic iron oxidation (pio) genes in Rhodopseudomonas palustris TIE-1 is mediated by the global regulator, FixK. Mol Microbiol *79*, 63-75.

Boudker, O., and Verdon, G. (2010). Structural perspectives on secondary active transporters. Trends Pharmacol Sci *31*, 418-426.

Boyce, C.K., Zwieniecki, M.A., Cody, G.D., Jacobsen, C., Wirick, S., Knoll, A.H., and Holbrook, N.M. (2004). Evolution of xylem lignification and hydrogel transport regulation. Proc Natl Acad Sci U S A *101*, 17555-17558.

Brunati, M., Marinelli, F., Bertolini, C., Gandolfi, R., Daffonchio, D., and Molinari, F. (2004). Biotransformations of cinnamic and ferulic acid with actinomycetes. Enzyme and Microbial Technology *34*, 3-9.

Brunschwig, C., Rochard, S., Pierrat, A., Rouger, A., Senger-Emonnot, P., George, G., and Raharivelomanana, P. (2016). Volatile composition and sensory properties of *Vanilla x tahitensis* bring new insights for vanilla quality control. J Sci Food Agric *96*, 848-858.

Bullerman, L.B., Lieu, F.Y., and Seier, S.A. (1977). Inhibition of growth and aflatoxin production by cinnamon and clove oils, cinnamic aldehyde and eugenol. Journal of Food Science *42*, 1107-1109.

Busch, W., and Saier, M.H. (2003). The IUBMB-Endorsed Transporter Classification System. In Membrane Transporters: Methods and Protocols, Q. Yan, ed. (Totowa, NJ: Humana Press), pp. 21-36.

Callaway, T.R., Adams, K.A., and Russell, J.B. (1999). The ability of "low G + C gram-positive" ruminal bacteria to resist monensin and counteract potassium depletion. Curr Microbiol *39*, 226-230.

Campillo, T., Renoud, S., Kerzaon, I., Vial, L., Baude, J., Gaillard, V., Bellvert, F., Chamignon, C., Comte, G., Nesme, X.*, et al.* (2014). Analysis of Hydroxycinnamic Acid Degradation in *Agrobacterium*

fabrum Reveals a Coenzyme A-Dependent, Beta-Oxidative Deacetylation Pathway. Appl Environ Microbiol *80*, 3341-3349.

Carlozzi, P., and Sacchi, A. (2001). Biomass production and studies on *Rhodopseudomonas palustris* grown in an outdoor, temperature controlled, underwater tubular photobioreactor. J Biotechnol *88*, 239-249.

Carmona, M., Zamarro, M.T., Blázquez, B., Durante-Rodríguez, G., Juárez, J.F., Valderrama, J.A., Barragán, M.J.L., García, J.L., and Díaz, E. (2009). Anaerobic Catabolism of Aromatic Compounds: a Genetic and Genomic View. Microbiol Mol Biol Rev *73*, 71-133.

Chae, J.C., and Zylstra, G.J. (2006). 4-Chlorobenzoate uptake in *Comamonas* sp. strain DJ-12 is mediated by a tripartite ATP-independent periplasmic transporter. J Bacteriol *188*, 8407-8412.

Chandrasekhar, K., Lee, Y.J., and Lee, D.W. (2015). Biohydrogen production: strategies to improve process efficiency through microbial routes. Int J Mol Sci *16*, 8266-8293.

Chaptal, V., Kwon, S., Sawaya, M.R., Guan, L., Kaback, H.R., and Abramson, J. (2011). Crystal structure of lactose permease in complex with an affinity inactivator yields unique insight into sugar recognition. Proc Natl Acad Sci U S A *108*, 9361-9366.

Chen, C., and Beattie, G.A. (2008). *Pseudomonas syringae* BetT is a low-affinity choline transporter that is responsible for superior osmoprotection by choline over glycine betaine. J Bacteriol *190*, 2717-2725.

Chen, J. (2013). Molecular mechanism of the *Escherichia coli* maltose transporter. Curr Opin Struct Biol *23*, 492-498.

Choquet, G., Jehan, N., Pissavin, C., Blanco, C., and Jebbar, M. (2005). OusB, a broad-specificity ABC-type transporter from *Erwinia chrysanthemi*, mediates uptake of glycine betaine and choline with a high affinity. Appl Environ Microbiol *71*, 3389-3398.

Chubiz, L.M., Glekas, G.D., and Rao, C.V. (2012). Transcriptional cross talk within the *mar-sox-rob* regulon in *Escherichia coli* is limited to the *rob* and *marRAB* operons. J Bacteriol *194*, 4867-4875.

Cogdell, R.J., Isaacs, N.W., Howard, T.D., McLuskey, K., Fraser, N.J., and Prince, S.M. (1999). How photosynthetic bacteria harvest solar energy. J Bacteriol *181*, 3869-3879.

Collins, R.F., and Derrick, J.P. (2007). Wza: a new structural paradigm for outer membrane secretory proteins? Trends Microbiol *15*, 96-100.

Conrad, T.M., Lewis, N.E., and Palsson, B. (2011). Microbial laboratory evolution in the era of genome-scale science. Mol Syst Biol *7*.

Cooper, V., and Lenski, R. (2000). The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature *407*.

Cortés-Rojas, D.F., Souza, C.R.F., and Oliveira, W.P. (2014). Encapsulation of eugenol rich clove extract in solid lipid carriers. Journal of Food Engineering *127*, 34-42.

Cronan, J.E., and Lin, S. (2011). Synthesis of the alpha,omega-dicarboxylic acid precursor of biotin by the canonical fatty acid biosynthetic pathway. Curr Opin Chem Biol *15*, 407-413.

Crosby, H.A., Pelletier, D.A., Hurst, G.B., and Escalante-Semerena, J.C. (2012). System-wide Studies of N-Lysine Acetylation in *Rhodopseudomonas palustris* Reveal Substrate Specificity of Protein Acetyltransferases. Journal of Biological Chemistry *287*, 15590-15601.

Cruz, J.M., Dominguez, J.M., Dominguez, H., and Parajo, J.C. (2001). Antioxidant and antimicrobial effects of extracts from hydrolysates of lignocellulosic materials. Journal of Agricultural and Food Chemistry *49*, 2459-2464.

Dashtban, M., Schraft, H., Syed, T.A., and Qin, W. (2010). Fungal biodegradation and enzymatic modification of lignin. International journal of Biochemistry and Molecular Biology *1*, 36-50.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A *97*, 6640-6645.

Davidson, A.L., Dassa, E., Orelle, C., and Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol Mol Biol Rev *72*, 317-364.

Day, R.E., Kitchen, P., Owen, D.S., Bland, C., Marshall, L., Conner, A.C., Bill, R.M., and Conner, M.T. (2014). Human aquaporins: regulators of transcellular water flow. Biochim Biophys Acta *1840*, 1492-1506.

de Bruijn, F.J. (2016). Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria, 2 Volume Set (Wiley).

de Cristobal, R.E., Vincent, P.A., and Salomon, R.A. (2006). Multidrug resistance pump AcrAB-TolC is required for high-level, Tet(A)-mediated tetracycline resistance in *Escherichia coli*. J Antimicrob Chemother *58*, 31-36.

de Gonzalo, G., Colpa, D.I., Habib, M.H., and Fraaije, M.W. (2016). Bacterial enzymes involved in lignin degradation. J Biotechnol *236*, 110-119.

Denger, K., Smits, T.H., and Cook, A.M. (2006). Genome-enabled analysis of the utilization of taurine as sole source of carbon or of nitrogen by *Rhodobacter sphaeroides* 2.4.1. Microbiology *152*, 3197-3206.

Deutscher, J., Aké, F.M.D., Derkaoui, M., Zébré, A.C., Cao, T.N., Bouraoui, H., Kentache, T., Mokhtari, A., Milohanic, E., and Joyet, P. (2014). The Bacterial Phosphoenolpyruvate:Carbohydrate Phosphotransferase System: Regulation by Protein Phosphorylation and Phosphorylation-Dependent Protein-Protein Interactions. Microbiology and Molecular Biology Reviews *78*, 231-256.

Di Gioia, D., Luziatelli, F., Negroni, A., Ficca, A.G., Fava, F., and Ruzzi, M. (2011). Metabolic engineering of *Pseudomonas fluorescens* for the production of vanillin from ferulic acid. J Biotechnol *156*, 309-316.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.M., Taly, J.F., and Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and

RNA sequences using structural information and homology extension. Nucleic Acids Research *39*, W13-17.

Diaz, E. (2004). Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. International microbiology : The Official Journal of the Spanish Society for Microbiology *7*, 173-180.

Dipasquale, L., Adessi, A., d'Ippolito, G., Rossi, F., Fontana, A., and De Philippis, R. (2015). Introducing capnophilic lactic fermentation in a combined dark-photo fermentation process: a route to unparalleled H2 yields. Appl Microbiol Biotechnol *99*, 1001-1010.

Dobson, P.D., and Kell, D.B. (2008). Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? Nature reviews Drug discovery *7*, 205-220.

Doud, D.F.R., Holmes, E.C., Richter, H., Molitor, B., Jander, G., and Angenent, L.T. (2017). Metabolic engineering of *Rhodopseudomonas palustris* for the obligate reduction of n-butyrate to n-butanol. Biotechnology for Biofuels *10*, 178.

Dragosits, M., and Mattanovich, D. (2013). Adaptive laboratory evolution – principles and applications for biotechnology. Microbial Cell Factories *12*, 64.

Dragosits, M., Mozhayskiy, V., Quinones-Soto, S., Park, J., and Tagkopoulos, I. (2013). Evolutionary potential, cross-stress behavior and the genetic basis of acquired stress resistance in *Escherichia coli*. Mol Syst Biol *9*.

Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci USA *88*.

Du, D., Wang, Z., James, N.R., Voss, J.E., Klimont, E., Ohene-Agyei, T., Venter, H., Chiu, W., and Luisi, B.F. (2014). Structure of the AcrAB-TolC multidrug efflux pump. Nature *509*, 512-515.

Dunlop, M.J., Dossani, Z.Y., Szmidt, H.L., Chu, H.C., Lee, T.S., Keasling, J.D., Hadi, M.Z., and Mukhopadhyay, A. (2011). Engineering microbial biofuel tolerance and export using efflux pumps. Mol Syst Biol *7*, 487.

Duval, V., McMurry, L.M., Foster, K., Head, J.F., and Levy, S.B. (2013). Mutational analysis of the multiple-antibiotic resistance regulator MarR reveals a ligand binding pocket at the interface between the dimerization and DNA binding domains. J Bacteriol *195*, 3341-3351.

Edlin, D.A.N., Narbad, A., Dickinson, J.R., and Lloyd, D. (1995). The biotransformation of simple phenolic compounds by *Brettanomyces anomalus*. FEMS Microbiology Letters *125*, 311-315.

Egland, P.G., Gibson, J., and Harwood, C.S. (2001). Reductive, coenzyme A-mediated pathway for 3-chlorobenzoate degradation in the phototrophic bacterium *Rhodopseudomonas palustris.* Appl Environ Microbiol *67*, 1396-1399.

Egland, P.G., Pelletier, D.A., Dispensa, M., Gibson, J., and Harwood, C.S. (1997). A cluster of bacterial genes for anaerobic benzene ring biodegradation. Proc Natl Acad Sci U S A *94*, 6484-6489.

Eitinger, T., Rodionov, D.A., Grote, M., and Schneider, E. (2011). Canonical and ECF-type ATP-binding cassette importers in prokaryotes: diversity in modular organization and cellular functions. FEMS Microbiol Rev *35*, 3-67.

Ekberg, K., Wielandt, A.G., Buch-Pedersen, M.J., and Palmgren, M.G. (2013). A conserved asparagine in a P-type proton pump is required for efficient gating of protons. J Biol Chem *288*, 9610-9618.

Elder, D.J., Morgan, P., and Kelly, D.J. (1992). Anaerobic degradation of trans-cinnamate and omega-phenylalkane carboxylic acids by the photosynthetic bacterium *Rhodopseudomonas palustris*: evidence for a beta-oxidation mechanism. Archives of microbiology *157*, 148-154.

Erkens, G.B., and Slotboom, D.J. (2010). Biochemical characterization of ThiT from *Lactococcus lactis:* a thiamin transporter with picomolar substrate binding affinity. Biochemistry *49*, 3203-3212.

Evolva (2017). Vanillin (http://www.evolva.com/vanillin/).

Fache, M., Boutevin, B., and Caillol, S. (2016). Vanillin Production from Lignin and Its Use as a Renewable Chemical. ACS Sustainable Chemistry & Engineering *4*, 35-46.

Falconnier, B., Lapierre, C., Lesage-Meessen, L., Yonnet, G., Brunerie, P., Colonna-Ceccaldi, B., Corrieu, G., and Asther, M. (1994). Vanillin as a product of ferulic acid biotransformation by the white-rot fungus *Pycnoporus cinnabarinus* I-937: Identification of metabolic pathways. Journal of Biotechnology *37*, 123-132.

Fargues, C., Mathias, Á., and Rodrigues, A. (1996). Kinetics of Vanillin Production from Kraft Lignin Oxidation. Industrial & Engineering Chemistry Research *35*, 28-36.

Ferguson, G.P., Totemeyer, S., MacLean, M.J., and Booth, I.R. (1998). Methylglyoxal production in bacteria: suicide or survival? Archives of Microbiology *170*, 209-218.

Fischer, M., Zhang, Q.Y., Hubbard, R.E., and Thomas, G.H. (2010). Caught in a TRAP: substrate-binding proteins in secondary transport. Trends Microbiol *18*, 471-478.

Fisher, M.A., Boyarskiy, S., Yamada, M.R., Kong, N., Bauer, S., and Tullman-Ercek, D. (2014). Enhancing tolerance to short-chain alcohols by engineering the *Escherichia coli* AcrB efflux pump to secrete the non-native substrate n-butanol. ACS Synth Biol *3*, 30-40.

Fleige, C., Hansen, G., Kroll, J., and Steinbuchel, A. (2013). Investigation of the *Amycolatopsis* sp. strain ATCC 39116 vanillin dehydrogenase and its impact on the biotechnical production of vanillin. Appl Environ Microbiol *79*, 81-90.

Foo, J.L., and Leong, S.S. (2013). Directed evolution of an E. coli inner membrane transporter for improved efflux of biofuel molecules. Biotechnol Biofuels *6*, 81.

Forward, J.A., Behrendt, M.C., Wyborn, N.R., Cross, R., and Kelly, D.J. (1997). TRAP transporters: a new family of periplasmic solute transport systems encoded by the dctPQM genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. J Bacteriol *179*, 5482-5493.

Fuchs, G. (2008). Anaerobic Metabolism of Aromatic Compounds. Annals of the New York Academy of Sciences *1125*, 82-99.

Furuya, T., Miura, M., Kuroiwa, M., and Kino, K. (2015). High-yield production of vanillin from ferulic acid by a coenzyme-independent decarboxylase/oxygenase two-stage process. N Biotechnol *32*, 335-339.

Gall, A., and Robert, B. (1999). Characterization of the different peripheral light-harvesting complexes from high- and low-light grown cells from *Rhodopseudomonas palustris.* Biochemistry *38*, 5185-5190.

Gallage, N.J., Hansen, E.H., Kannangara, R., Olsen, C.E., Motawia, M.S., Jorgensen, K., Holme, I., Hebelstrup, K., Grisoni, M., and Moller, B.L. (2014). Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. Nat Commun *5*, 4037.

Gallage, N.J., and Moller, B.L. (2015). Vanillin-bioconversion and bioengineering of the most popular plant flavor and its de novo biosynthesis in the vanilla orchid. Mol Plant *8*, 40-57.

Gao, R., Wang, Y., Zhang, Y., Tong, J., and Dai, W. (2017). Cobalt(II) bioaccumulation and distribution in *Rhodopseudomonas palustris*. Biotechnology & Biotechnological Equipment *31*, 527-534.

Gao, X., Lu, F., Zhou, L., Dang, S., Sun, L., Li, X., Wang, J., and Shi, Y. (2009). Structure and mechanism of an amino acid antiporter. Science *324*, 1565-1568.

Genee, H.J., Bali, A.P., Petersen, S.D., Siedler, S., Bonde, M.T., Gronenberg, L.S., Kristensen, M., Harrison, S.J., and Sommer, M.O. (2016). Functional mining of transporters using synthetic selections. Nat Chem Biol *12*, 1015-1022.

Giraud, E., Zappa, S., Vuillet, L., Adriano, J.M., Hannibal, L., Fardoux, J., Berthomieu, C., Bouyer, P., Pignol, D., and Vermeglio, A. (2005). A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in *Rhodopseudomonas palustris.* J Biol Chem *280*, 32389-32397.

Giuliani, S.E., Frank, A.M., Corgliano, D.M., Seifert, C., Hauser, L., and Collart, F.R. (2011). Environment sensing and response mediated by ABC transporters. BMC Genomics *12*, S8.

Glaenzer, J., Peter, M.F., Thomas, G.H., and Hagelueken, G. (2017). PELDOR Spectroscopy Reveals Two Defined States of a Sialic Acid TRAP Transporter SBP in Solution. Biophys J *112*, 109-120.

Global Market Insights, I. (2016). Bio Vanillin Market Size, I. Global Market Insights, ed. (https://www.gminsights.com/industry-analysis/bio-vanillin-market).

Gogada, R., Singh, S.S., Lunavat, S.K., Pamarthi, M.M., Rodrigue, A., Vadivelu, B., Phanithi, P.B., Gopala, V., and Apte, S.K. (2015). Engineered *Deinococcus radiodurans* R1 with NiCoT genes for bioremoval of trace cobalt from spent decontamination solutions of nuclear power reactors. Appl Microbiol Biotechnol *99*, 9203-9213.

Gonin, S., Arnoux, P., Pierru, B., Lavergne, J., Alonso, B., Sabaty, M., and Pignol, D. (2007). Crystal structures of an Extracytoplasmic Solute Receptor from a TRAP transporter in its open and closed forms reveal a helix-swapped dimer requiring a cation for alpha-keto acid binding. BMC Struct Biol *7*, 11.

Gonzalez, C.F., Proudfoot, M., Brown, G., Korniyenko, Y., Mori, H., Savchenko, A.V., and Yakunin, A.F. (2006). Molecular basis of formaldehyde detoxification. Characterization of two S-formylglutathione hydrolases from *Escherichia coli*, FrmB and YeiG. J Biol Chem *281*, 14514-14522.

Goodarzi, H., Bennett, B., Amini, S., Reaves, M., Hottes, A., and Rabinowitz, J. (2010). Tavazoie S: Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*. Mol Syst Biol *6*.

Grimsrud, P.A., Xie, H., Griffin, T.J., and Bernlohr, D.A. (2008). Oxidative stress and covalent modification of protein with bioactive aldehydes. J Biol Chem *283*, 21837-21841.

Guan, L., and Kaback, H.R. (2006). Lessons from lactose permease. Annu Rev Biophys Biomol Struct *35*, 67-91.

Gurujeyalakshmi, G., and Mahadevan, A. (1987). Dissimilation of ferulic acid by*Bacillus subtilis*. Current Microbiology *16*, 69-73.

Haghighi, O., Shahryari, S., Ebadi, M., Modiri, S., Zahiri, H.S., Maleki, H., and Noghabi, K.A. (2017). Limnothrix sp. KO05: A newly characterized cyanobacterial biosorbent for cadmium removal: the enzymatic and non-enzymatic antioxidant reactions to cadmium toxicity. Environmental Toxicology and Pharmacology *51*, 142-155.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology *166*, 557-580.

Hao, Z., Lou, H., Zhu, R., Zhu, J., Zhang, D., Zhao, B.S., Zeng, S., Chen, X., Chan, J., He, C.*, et al.* (2014). The multiple antibiotic resistance regulator MarR is a copper sensor in *Escherichia coli.* Nat Chem Biol *10*, 21-28.

Hara, H., Masai, E., Katayama, Y., and Fukuda, M. (2000). The 4-oxalomesaconate hydratase gene, involved in the protocatechuate 4,5-cleavage pathway, is essential to vanillate and syringate degradation in *Sphingomonas paucimobilis* SYK-6. J Bacteriol *182*, 6950-6957.

Harrison, F.H., and Harwood, C.S. (2005). The *pimFABCDE* operon from *Rhodopseudomonas palustris* mediates dicarboxylic acid degradation and participates in anaerobic benzoate degradation. Microbiology *151*, 727-736.

Harwood, C.S., and Gibson, J. (1988). Anaerobic and aerobic metabolism of diverse aromatic compounds by the photosynthetic bacterium *Rhodopseudomonas palustris*. Appl Environ Microbiol *54*, 712-717.

Hatakka, A. (1994). Lignin-modifying enzymes from selected white-rot fungi: production and role from in lignin degradation. FEMS Microbiology Reviews *13*, 125-135.

Hayashi, K., Fukushima, A., Hayashi-Nishino, M., and Nishino, K. (2014). Effect of methylglyoxal on multidrug-resistant *Pseudomonas aeruginosa*. Front Microbiol *5*, 180.

Hearn, E.M., Dennis, J.J., Gray, M.R., and Foght, J.M. (2003). Identification and Characterization of the emhABC Efflux System for Polycyclic Aromatic Hydrocarbons in *Pseudomonas fluorescens* cLP6a. J Bacteriol *185*, 6233-6240.

Heiniger, E.K., Oda, Y., Samanta, S.K., and Harwood, C.S. (2012). How posttranslational modification of nitrogenase is circumvented in *Rhodopseudomonas palustris* strains that produce hydrogen gas constitutively. Appl Environ Microbiol *78*, 1023-1032.

Hernandez, M.E., and Newman, D.K. (2001). Extracellular electron transfer. Cellular and molecular life sciences : CMLS *58*, 1562-1571.

Higgins, C.F., and Linton, K.J. (2004). The ATP switch model for ABC transporters. Nat Struct Mol Biol *11*, 918-926.

Higuchi-Takeuchi, M., Morisaki, K., Toyooka, K., and Numata, K. (2016). Synthesis of High-Molecular-Weight Polyhydroxyalkanoates by Marine Photosynthetic Purple Bacteria. PLoS One *11*, e0160981.

Hirakawa, H., Hirakawa, Y., Greenberg, E.P., and Harwood, C.S. (2015). BadR and BadM Proteins Transcriptionally Regulate Two Operons Needed for Anaerobic Benzoate Degradation by *Rhodopseudomonas palustris*. Appl Environ Microbiol *81*, 4253-4262.

Hirakawa, H., Schaefer, A.L., Greenberg, E.P., and Harwood, C.S. (2012). Anaerobic p-coumarate degradation by *Rhodopseudomonas palustris* and identification of CouR, a MarR repressor protein that binds p-coumaroyl coenzyme A. J Bacteriol *194*, 1960-1967.

Holland, I.B., and Blight, M.A. (1999). ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. J Mol Biol *293*, 381-399.

Horinouchi, T., Sakai, A., Kotani, H., Tanabe, K., and Furusawa, C. (2017). Improvement of isopropanol tolerance of *Escherichia coli* using adaptive laboratory evolution and omics technologies. J Biotechnol *255*, 47-56.

Hua, D., Ma, C., Lin, S., Song, L., Deng, Z., Maomy, Z., Zhang, Z., Yu, B., and Xu, P. (2007a). Biotransformation of isoeugenol to vanillin by a newly isolated *Bacillus pumilus* strain: identification of major metabolites. J Biotechnol *130*, 463-470.

Hua, D., Ma, C., Song, L., Lin, S., Zhang, Z., Deng, Z., and Xu, P. (2007b). Enhanced vanillin production from ferulic acid using adsorbent resin. Appl Microbiol Biotechnol *74*, 783-790.

Hua, F., Wang, H.Q., Li, Y., and Zhao, Y.C. (2013). Trans-membrane transport of n-octadecane by *Pseudomonas* sp. DG17. J Microbiol *51*, 791-799.

Huang, J.J., Heiniger, E.K., McKinlay, J.B., and Harwood, C.S. (2010). Production of hydrogen gas from light and the inorganic electron donor thiosulfate by *Rhodopseudomonas palustris*. Appl Environ Microbiol *76*, 7717-7722.

Huang, X.F., Santhanam, N., Badri, D.V., Hunter, W.J., Manter, D.K., Decker, S.R., Vivanco, J.M., and Reardon, K.F. (2013). Isolation and characterization of lignin-degrading bacteria from rainforest soils. Biotechnol Bioeng *110*, 1616-1626.

Huder, J.B., and Dimroth, P. (1993). Sequence of the sodium ion pump methylmalonyl-CoA decarboxylase from *Veillonella parvula*. J Biol Chem *268*, 24564-24571.

Hunte, C., Screpanti, E., Venturi, M., Rimon, A., Padan, E., and Michel, H. (2005). Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH. Nature *435*, 1197-1202.

Hvorup, R.N., Goetz, B.A., Niederer, M., Hollenstein, K., Perozo, E., and Locher, K.P. (2007). Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF. Science *317*, 1387-1390.

Imlay, J.A. (2013). The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. Nat Rev Microbiol *11*, 443-454.

Ishibashi, K., Kondo, S., Hara, S., and Morishita, Y. (2011). The evolutionary aspects of aquaporin family. Am J Physiol Regul Integr Comp Physiol *300*, R566-576.

Ishikawa, H., Schubert, W.J., and Nord, F.F. (1963). Investigations on lignins and lignification. Archives of Biochemistry and Biophysics *100*, 140-149.

Itel, F., Al-Samir, S., Oberg, F., Chami, M., Kumar, M., Supuran, C.T., Deen, P.M., Meier, W., Hedfalk, K., Gros, G.*, et al.* (2012). CO2 permeability of cell membranes is regulated by membrane cholesterol and protein gas channels. FASEB J *26*, 5182-5191.

Iyer, R., Williams, C., and Miller, C. (2003). Arginine-Agmatine Antiporter in Extreme Acid Resistance in *Escherichia coli*. Journal of Bacteriology *185*, 6556-6561.

Jacso, T., Grote, M., Daus, M.L., Schmieder, P., Keller, S., Schneider, E., and Reif, B. (2009). Periplasmic loop P2 of the MalF subunit of the maltose ATP binding cassette transporter is sufficient to bind the maltose binding protein MalE. Biochemistry *48*, 2216-2225.

Jebbar, M., Sohn-Bosser, L., Bremer, E., Bernard, T., and Blanco, C. (2005). Ectoine-induced proteins in *Sinorhizobium meliloti* include an Ectoine ABC-type transporter involved in osmoprotection and ectoine catabolism. J Bacteriol *187*, 1293-1304.

Jiao, Y., and Newman, D.K. (2007). The pio operon is essential for phototrophic Fe(II) oxidation in *Rhodopseudomonas palustris* TIE-1. J Bacteriol *189*, 1765-1773.

Jones, C.M., Hernández Lozada, N.J., and Pfleger, B.F. (2015). Efflux Systems in Bacteria and their Metabolic Engineering Applications. Appl Microbiol Biotechnol *99*, 9381-9393.

Joshi, G.S., Zianni, M., Bobst, C.E., and Tabita, F.R. (2013). Regulatory twist and synergistic role of metabolic coinducer- and response regulator-mediated CbbR-cbbI interactions in *Rhodopseudomonas palustris* CGA010. J Bacteriol *195*, 1381-1388.

Julsing, M.K., Schrewe, M., Cornelissen, S., Hermann, I., Schmid, A., and Buhler, B. (2012). Outer membrane protein AlkL boosts biocatalytic oxyfunctionalization of hydrophobic substrates in *Escherichia coli.* Appl Environ Microbiol *78*, 5724-5733.

Jurková, M., and Wurst, M. (1993). Biodegradation of aromatic carboxylic acids by *Pseudomonas mira*. FEMS Microbiology Letters *111*, 245-250.

Kaback, H.R., Smirnova, I., Kasho, V., Nie, Y., and Zhou, Y. (2011). The alternating access transport mechanism in LacY. J Membr Biol *239*, 85-93.

Kabelitz, N., Santos, P.M., and Heipieper, H.J. (2003). Effect of aliphatic alcohols on growth and degree of saturation of membrane lipids in *Acinetobacter calcoaceticus*. FEMS microbiology letters *220*, 223-227.

Kai, L., and Kaldenhoff, R. (2014). A refined model of water and CO(2) membrane diffusion: effects and contribution of sterols and proteins. Sci Rep *4*, 6665.

Kandori, H. (2015). Ion-pumping microbial rhodopsins. Front Mol Biosci *2*, 52.

Kang, Y.S., Lee, Y., Jung, H., Jeon, C.O., Madsen, E.L., and Park, W. (2007). Overexpressing antioxidant enzymes enhances naphthalene biodegradation in *Pseudomonas* sp. strain As1. Microbiology *153*, 3246-3254.

Karmakar, B., Vohra, R.M., Nandanwar, H., Sharma, P., Gupta, K.G., and Sobti, R.C. (2000). Rapid degradation of ferulic acid via 4-vinylguaiacol and vanillin by a newly isolated strain of *Bacillus coagulans.* J Biotechnol *80*, 195-202.

Kaur, B., and Chakraborty, D. (2013). Biotechnological and molecular approaches for vanillin production: a review. Appl Biochem Biotechnol *169*, 1353-1372.

Kell, D.B., Dobson, P.D., and Oliver, S.G. (2011). Pharmaceutical drug transport: the issues and the implications that it is essentially carrier-mediated only. Drug Discov Today *16*, 704-714.

Kell, D.B., and Oliver, S.G. (2014). How drugs get into cells: tested and testable predictions to help discriminate between transporter-mediated uptake and lipoidal bilayer diffusion. Front Pharmacol *5*, 231.

Kell, D.B., Swainston, N., Pir, P., and Oliver, S.G. (2015). Membrane transporter engineering in industrial biotechnology and whole cell biocatalysis. Trends Biotechnol *33*, 237-246.

Kelly, D.J., and Thomas, G.H. (2001). The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. FEMS Microbiol Rev *25*, 405-424.

Khare, D., Oldham, M.L., Orelle, C., Davidson, A.L., and Chen, J. (2009). Alternating access in maltose transporter mediated by rigid-body rotations. Mol Cell *33*, 528-536.

Kim, J.W., Kim, S., Kim, S., Lee, H., Lee, J.-O., and Jin, M.S. (2017). Structural insights into the elevator-like mechanism of the sodium/citrate symporter CitS. Scientific Reports *7*, 2548.

Kishimoto, J., Haze, S., and Ifuku, O. (1993). Process for producing biotin vitamers using novel microorganisms (Google Patents).

Koebnik, R., Locher, K.P., and Van Gelder, P. (2000). Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol Microbiol *37*, 239-253.

Kohanski, M.A., Dwyer, D.J., Hayete, B., Lawrence, C.A., and Collins, J.J. A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics. Cell *130*, 797-810.

Korshunov, S., and Imlay, J.A. (2010). Two sources of endogenous hydrogen peroxide in *Escherichia coli*. Mol Microbiol *75*, 1389-1401.

Krings, U., and Berger, R.G. (1998). Biotechnological production of flavours and fragrances. Appl Microbiol Biotechnol *49*, 1-8.

Kuhlbrandt, W. (2004). Biology, structure and mechanism of P-type ATPases. Nat Rev Mol Cell Biol *5*, 282-295.

Kumagai, K., Miki, M., Kawano, E., and Mitsuda, S. (1995). Process for production of biotin using bacteria belonging to the genus *Sphingomonas* (Google Patents).

Kunjapur, A.M., and Prather, K.L. (2015). Microbial engineering for aldehyde synthesis. Appl Environ Microbiol *81*, 1892-1901.

Kunjapur, A.M., Tarasova, Y., and Prather, K.L. (2014). Synthesis and accumulation of aromatic aldehydes in an engineered strain of *Escherichia coli*. J Am Chem Soc *136*, 11644-11654.

Kusaka, I., and Matsushita, T. (1987). Characterization of a Ca2+ uniporter from *Bacillus subtilis* by partial purification and reconstitution into phospholipid vesicles. J Gen Microbiol *133*, 1337-1342.

Labuda, I.M., Goers, S.K., and Keon, K.A. (1992). Bioconversion process for the production of vanillin (Google Patents).

Lampman, G.M., Andrews, J., Bratz, W., Hanssen, O., Kelley, K., Perry, D., and Ridgeway, A. (1977). Preparation of vanillin from eugenol and sawdust. Journal of Chemical Education *54*, 776.

Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S*., et al.* (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. Nat Biotechnol *22*, 55-61.

Law, C.J., Maloney, P.C., and Wang, D.N. (2008). Ins and outs of major facilitator superfamily antiporters. Annu Rev Microbiol *62*, 289-305.

Lecher, J., Pittelkow, M., Zobel, S., Bursy, J., Bonig, T., Smits, S.H., Schmitt, L., and Bremer, E. (2009). The crystal structure of UehA in complex with ectoine-A comparison with other TRAP-T binding proteins. J Mol Biol *389*, 58-73.

Lee, C., Kang, H.J., von Ballmoos, C., Newstead, S., Uzdavinys, P., Dotson, D.L., Iwata, S., Beckstein, O., Cameron, A.D., and Drew, D. (2013). A two-domain elevator mechanism for sodium/proton antiport. Nature *501*, 573-577.

Lee, E.G., Yoon, S.H., Das, A., Lee, S.H., Li, C., Kim, J.Y., Choi, M.S., Oh, D.K., and Kim, S.W. (2009). Directing vanillin production from ferulic acid by increased acetyl-CoA consumption in recombinant *Escherichia coli*. Biotechnol Bioeng *102*, 200-208.

Lee, S.K., Lur, H.S., Lo, K.J., Cheng, K.C., Chuang, C.C., Tang, S.J., Yang, Z.W., and Liu, C.T. (2016). Evaluation of the effects of different liquid inoculant formulations on the survival and plant-growth-promoting efficiency of *Rhodopseudomonas palustris* strain PS3. Appl Microbiol Biotechnol *100*, 7977-7987.

Leffingwell, J.C., and Leffingwell, D. (2015). Flavours & fragrances - Recent advances in biotechnology. In Speciality Chemicals Magazine (www.specchemonline.com), pp. 32-34.

Lennen, R.M., Politz, M.G., Kruziki, M.A., and Pfleger, B.F. (2013). Identification of transport proteins involved in free fatty acid efflux in *Escherichia coli*. J Bacteriol *195*, 135-144.

Lepore, B.W., Indic, M., Pham, H., Hearn, E.M., Patel, D.R., and van den Berg, B. (2011). Ligand-gated diffusion across the bacterial outer membrane. Proc Natl Acad Sci U S A *108*, 10121-10126.

Lesage-Meessen, L., Stentelaire, C., Lomascolo, A., Couteau, D., Asther, M., Moukha, S., Record, E., Sigoillot, J.-C., and Asther, M. (1999). Fungal transformation of ferulic acid from sugar beet pulp to natural vanillin. Journal of the Science of Food and Agriculture *79*, 487-490.

Li, H., Zhang, G., and Dang, Y. (2016). Adaptive laboratory evolution of *Klebsiella pneumoniae* for improving 2,3-butanediol production. Bioengineered *7*, 432-438.

Li, T., and Rosazza, J.P. (2000). Biocatalytic synthesis of vanillin. Appl Environ Microbiol *66*, 684-687.

Li, X., and Chapple, C. (2010). Understanding lignification: challenges beyond monolignol biosynthesis. Plant Physiol *154*, 449-452.

Li, X., Yang, J., Li, X., Gu, W., Huang, J., and Zhang, K.-Q. (2008). The metabolism of ferulic acid via 4-vinylguaiacol to vanillin by *Enterobacter* sp. Px6-4 isolated from Vanilla root. Process Biochemistry *43*, 1132-1137.

Lin, L., Ji, Y., Tu, Q., Huang, R., Teng, L., Zeng, X., Song, H., Wang, K., Zhou, Q., Li, Y.*, et al.* (2013). Microevolution from shock to adaptation revealed strategies improving ethanol tolerance and production in Thermoanaerobacter. Biotechnol Biofuels *6*, 103.

Linger, J.G., Vardon, D.R., Guarnieri, M.T., Karp, E.M., Hunsinger, G.B., Franden, M.A., Johnson, C.W., Chupka, G., Strathmann, T.J., Pienkos, P.T.*, et al.* (2014). Lignin valorization through integrated biological funneling and chemical catalysis. Proc Natl Acad Sci U S A *111*, 12013-12018.

Liochev, S.I., Benov, L., Touati, D., and Fridovich, I. (1999). Induction of the *soxRS* regulon of *Escherichia coli* by superoxide. J Biol Chem *274*, 9479-9481.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of next-generation sequencing systems. J Biomed Biotechnol *2012*.

Liu, S., Zhang, G., Zhang, J., Li, X., and Li, J. (2016). Performance, carotenoids yield and microbial population dynamics in a photobioreactor system treating acidic wastewater: Effect of hydraulic retention time (HRT) and organic loading rate (OLR). Bioresour Technol *200*, 245-252.

Llanes, C., Hocquet, D., Vogne, C., Benali-Baitich, D., Neuwirth, C., and Plesiat, P. (2004). Clinical strains of *Pseudomonas aeruginosa* overproducing MexAB-OprM and MexXY efflux pumps simultaneously. Antimicrob Agents Chemother *48*, 1797-1802.

Locher, K.P. (2016). Mechanistic diversity in ATP-binding cassette (ABC) transporters. Nat Struct Mol Biol *23*, 487-493.

Lolkema, J.S., Chaban, Y., and Boekema, E.J. (2003). Subunit composition, structure, and distribution of bacterial V-type ATPases. Journal of bioenergetics and biomembranes *35*, 323-335.

LoPachin, R.M., and Gavin, T. (2014). Molecular mechanisms of aldehyde toxicity: a chemical perspective. Chemical Research in Toxicology *27*, 1081-1091.

Lopez-Marques, R.L., Theorin, L., Palmgren, M.G., and Pomorski, T.G. (2014). P4-ATPases: lipid flippases in cell membranes. Pflugers Arch *466*, 1227-1240.

Lu, H., Chen, J., Jia, Y., Cai, M., and Lee, P.K. (2016). Transcriptomic Responses of the Interactions between *Clostridium cellulovorans* 743B and *Rhodopseudomonas palustris* CGA009 in a Cellulose-Grown Coculture for Enhanced Hydrogen Production. Appl Environ Microbiol *82*, 4546-4559.

Luan, G., Cai, Z., Li, Y., and Ma, Y. (2013). Genome replication engineering assisted continuous evolution (GREACE) to improve microbial tolerance for biofuels production. Biotechnology for Biofuels *6*, 137.

Luer, L., Moulisova, V., Henry, S., Polli, D., Brotosudarmo, T.H., Hoseinkhani, S., Brida, D., Lanzani, G., Cerullo, G., and Cogdell, R.J. (2012). Tracking energy transfer between light harvesting complex 2 and 1 in photosynthetic membranes grown under high and low illumination. Proc Natl Acad Sci U S A *109*, 1473-1478.

Ma, P., Mori, T., Zhao, C., Thiel, T., and Johnson, C.H. (2016). Evolution of KaiC-Dependent Timekeepers: A Proto-circadian Timing Mechanism Confers Adaptive Fitness in the Purple Bacterium *Rhodopseudomonas palustris*. PLoS Genet *12*, e1005922.

Ma, X.K., and Daugulis, A.J. (2014). Effect of bioconversion conditions on vanillin production by *Amycolatopsis* sp. ATCC 39116 through an analysis of competing by-product formation. Bioprocess Biosyst Eng *37*, 891-899.

Madigan, M.T., and Gest, H. (1988). Selective enrichment and isolation of Rhodopseudomonas palustris using trans-cinnamic acid as sole carbon source. FEMS Microbiology Letters *53*, 53-58.

Mahipant, G., Paemanee, A., Roytrakul, S., Kato, J., and Vangnai, A.S. (2017). The significance of proline and glutamate on butanol chaotropic stress in *Bacillus subtilis* 168. Biotechnol Biofuels *10*, 122.

Manandhar, M., and Cronan, J.E. (2017). Pimelic acid, the first precursor of the *Bacillus subtilis* biotin synthesis pathway, exists as the free acid and is assembled by fatty acid synthesis. Mol Microbiol *104*, 595-607.

Martin, R.G., and Rosner, J.L. (2004). Transcriptional and translational regulation of the *marRAB* multiple antibiotic resistance operon in *Escherichia coli*. Mol Microbiol *53*, 183-191.

Masai, E., Harada, K., Peng, X., Kitayama, H., Katayama, Y., and Fukuda, M. (2002). Cloning and Characterization of the Ferulic Acid Catabolic Genes of *Sphingomonas paucimobilis* SYK-6. Applied and Environmental Microbiology *68*, 4416-4424.

Masai, E., Katayama, Y., Nishikawa, S., and Fukuda, M. (1999). Characterization of *Sphingomonas paucimobilis* SYK-6 genes involved in degradation of lignin-related compounds. J Ind Microbiol Biotechnol *23*, 364-373.

Mathew, S., and Abraham, T.E. (2004). Ferulic acid: an antioxidant found naturally in plant cell walls and feruloyl esterases involved in its release and their applications. Critical reviews in biotechnology *24*, 59-83.

McGrath, J.E., and Harfoot, C.G. (1997). Reductive dehalogenation of halocarboxylic acids by the phototrophic genera *Rhodospirillum* and *Rhodopseudomonas*. Appl Environ Microbiol *63*, 3333-3335.

McKinlay, J.B., Oda, Y., Ruhl, M., Posto, A.L., Sauer, U., and Harwood, C.S. (2014). Non-growing *Rhodopseudomonas palustris* increases the hydrogen gas yield from acetate by shifting from the glyoxylate shunt to the tricarboxylic acid cycle. J Biol Chem *289*, 1960-1970.

Mehrabi, S., Ekanemesang, U.M., Aikhionbare, F.O., Kimbro, K.S., and Bender, J. (2001). Identification and characterization of *Rhodopseudomonas* spp., a purple, non-sulfur bacterium from microbial mats. Biomolecular engineering *18*, 49-56.

Merugu, R., Rajya laxmi, K., Girisham, S., and Reddy, S.M. (2013). Chromate Reduction by Purple Non Sulphur Phototrophic Bacterium *Rhodopseudomonas palustris* KU003 Isolated from Tannery Effluent. International Journal of Environment and Bioenergy *6*, 187-192.

Messner, K.R., and Imlay, J.A. (2002). Mechanism of superoxide and hydrogen peroxide formation by fumarate reductase, succinate dehydrogenase, and aspartate oxidase. J Biol Chem *277*, 42563-42571.

Mingardon, F., Clement, C., Hirano, K., Nhan, M., Luning, E.G., Chanal, A., and Mukhopadhyay, A. (2015). Improving Olefin Tolerance and Production in *E. coli* Using Native and Evolved AcrB. Biotechnology and Bioengineering *112*, 879-888.

Muheim, A., and Lerch, K. (1999). Towards a high-yield bioconversion of ferulic acid to vanillin. Applied Microbiology and Biotechnology *51*, 456-461.

Mukhopadhyay, A. (2015). Tolerance engineering in bacteria for the production of advanced biofuels and chemicals. Trends in Microbiology *23*, 498-508.

Mulder, K.C., Bandola, J., and Schumann, W. (2013). Construction of an artificial *secYEG* operon allowing high level secretion of alpha-amylase. Protein Expr Purif *89*, 92-96.

Muller, V., and Gruber, G. (2003). ATP synthases: structure, function and evolution of unique energy converters. Cellular and Molecular Life Sciences : CMLS *60*, 474-494.

Mulligan, C., Fenollar-Ferrer, C., Fitzgerald, G.A., Vergara-Jaque, A., Kaufmann, D., Li, Y., Forrest, L.R., and Mindell, J.A. (2016). The bacterial dicarboxylate transporter VcINDY uses a two-domain elevator-type mechanism. Nat Struct Mol Biol *23*, 256-263.

Mulligan, C., Fischer, M., and Thomas, G.H. (2011). Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. FEMS Microbiol Rev *35*, 68-86.

Mulligan, C., Geertsma, E.R., Severi, E., Kelly, D.J., Poolman, B., and Thomas, G.H. (2009). The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. Proceedings of the National Academy of Sciences *106*, 8398-8398.

Mulligan, C., Kelly, D.J., and Thomas, G.H. (2007). Tripartite ATP-independent periplasmic transporters: application of a relational database for genome-wide analysis of transporter gene frequency and organization. Journal of Molecular Microbiology and Biotechnology *12*, 218-226.

Murata-Kamiya, N., and Kamiya, H. (2001). Methylglyoxal, an endogenous aldehyde, crosslinks DNA polymerase and the substrate DNA. Nucleic acids research *29*, 3433-3438.

Murata, T., Yamato, I., Kakinuma, Y., Shirouzu, M., Walker, J.E., Yokoyama, S., and Iwata, S. (2008). Ion binding and selectivity of the rotor ring of the Na+-transporting V-ATPase. Proc Natl Acad Sci U S A *105*, 8607-8612.

Narbad, A., and Gasson, M.J. (1998). Metabolism of ferulic acid via vanillin using a novel CoA-dependent pathway in a newly-isolated strain of *Pseudomonas fluorescens*. Microbiology *144 (Pt 5)*, 1397-1405.

Nazareth, S., and Mavinkurve, S. (1986). Degradation of ferulic acid via 4-vinylguaiacol by *Fusarium solani* (Mart.) Sacc. Canadian Journal of Microbiology *32*, 494-497.

Nguyen, T.T., Iwaki, A., Ohya, Y., and Izawa, S. (2014). Vanillin causes the activation of Yap1 and mitochondrial fragmentation in *Saccharomyces cerevisiae*. J Biosci Bioeng *117*, 33-38.

Ni, J., Tao, F., Du, H., and Xu, P. (2015). Mimicking a natural pathway for de novo biosynthesis: natural vanillin production from accessible carbon sources. Sci Rep *5*, 13670.

Nikaido, H. (1993). Transport across the bacterial outer membrane. Journal of Bioenergetics and Biomembranes *25*, 581-589.

Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. Microbiol Mol Biol Rev *67*, 593-656.

Numata, K., and Morisaki, K. (2015). Screening of Marine Bacteria To Synthesize Polyhydroxyalkanoate from Lignin: Contribution of Lignin Derivatives to Biosynthesis by *Oceanimonas doudoroffii*. ACS Sustainable Chemistry & Engineering *3*, 569-573.

Nygaard, T.P., Rovira, C., Peters, G.H., and Jensen, M.O. (2006). Ammonium recruitment and ammonia transport by *E. coli* ammonia channel AmtB. Biophys J *91*, 4401-4412.

Oda, Y., de Vries, Y.P., Forney, L.J., and Gottschal, J.C. (2001). Acquisition of the ability for *Rhodopseudomonas palustris* to degrade chlorinated benzoic acids as the sole carbon source. FEMS Microbiology Ecology *38*, 133-139.

Oda, Y., Samanta, S.K., Rey, F.E., Wu, L., Liu, X., Yan, T., Zhou, J., and Harwood, C.S. (2005). Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bacterium *Rhodopseudomonas palustris*. J Bacteriol *187*, 7784-7794.

Oda, Y., Star, B., Huisman, L.A., Gottschal, J.C., and Forney, L.J. (2003). Biogeography of the Purple Nonsulfur Bacterium *Rhodopseudomonas palustris*. Applied and Environmental Microbiology *69*, 5186-5191.

Odoux, E., and Grisoni, M. (2010). Vanilla, 1st edn (CRC Press).

Oh, Y.-K., Seol, E.-H., Lee, E.Y., and Park, S. (2002). Fermentative hydrogen production by a new chemoheterotrophic bacterium *Rhodopseudomonas Palustris* P4. International Journal of Hydrogen Energy *27*, 1373-1379.

Oh, Y. (2004). Photoproduction of hydrogen from acetate by a chemoheterotrophic bacterium *Rhodopseudomonas palustris* P4. International Journal of Hydrogen Energy.

Okochi, M., Kurimoto, M., Shimizu, K., and Honda, H. (2007). Increase of organic solvent tolerance by overexpression of manXYZ in *Escherichia coli*. Appl Microbiol Biotechnol *73*, 1394-1399.

Oldham, M.L., and Chen, J. (2011). Snapshots of the maltose transporter during ATP hydrolysis. Proc Natl Acad Sci U S A *108*, 15152-15156.

Orelle, C., Alvarez, F.J., Oldham, M.L., Orelle, A., Wiley, T.E., Chen, J., and Davidson, A.L. (2010). Dynamics of alpha-helical subdomain rotation in the intact maltose ATP-binding cassette transporter. Proc Natl Acad Sci U S A *107*, 20293-20298.

Overhage, J., Priefert, H., and Steinbuchel, A. (1999). Biochemical and genetic analyses of ferulic acid catabolism in *Pseudomonas* sp. Strain HR199. Appl Environ Microbiol *65*, 4837-4847.

Overhage, J., Steinbuchel, A., and Priefert, H. (2003). Highly Efficient Biotransformation of Eugenol to Ferulic Acid and Further Conversion to Vanillin in Recombinant Strains of *Escherichia coli*. Applied and Environmental Microbiology *69*, 6569-6576.

Pakpour, F., Najafpour, G., Tabatabaei, M., Tohidfar, M., and Younesi, H. (2014). Biohydrogen production from CO-rich syngas via a locally isolated *Rhodopseudomonas palustris* PT. Bioprocess Biosyst Eng *37*, 923-930.

Pan, C., Oda, Y., Lankford, P.K., Zhang, B., Samatova, N.F., Pelletier, D.A., Harwood, C.S., and Hettich, R.L. (2008). Characterization of anaerobic catabolism of p-coumarate in *Rhodopseudomonas palustris* by integrating transcriptomics and quantitative proteomics. Molecular & cellular proteomics : MCP *7*, 938-948.

Pandey, A., Pandey, A., Srivastava, P., and Pandey, A. (2006). Using reverse micelles as microreactor for hydrogen production by coupled systems of *Nostoc*/*R. palustris* and *Anabaena/R. palustris*. World Journal of Microbiology and Biotechnology *23*, 269-274.

Pechter, K.B., Gallagher, L., Pyles, H., Manoil, C.S., and Harwood, C.S. (2015). Essential Genome of the Metabolically Versatile Alphaproteobacterium *Rhodopseudomonas palustris*. J Bacteriol *198*, 867-876.

Pelletier, D.A., and Harwood, C.S. (2000). 2-Hydroxycyclohexanecarboxyl coenzyme A dehydrogenase, an enzyme characteristic of the anaerobic benzoate degradation pathway used by *Rhodopseudomonas palustris*. J Bacteriol *182*, 2753-2760.

Peng, X., Egashira, T., Hanashiro, K., Masai, E., Nishikawa, S., Katayama, Y., Kimbara, K., and Fukuda, M. (1998). Cloning of a *Sphingomonas paucimobilis* SYK-6 gene encoding a novel oxygenase that cleaves lignin-related biphenyl and characterization of the enzyme. Appl Environ Microbiol *64*, 2520-2527.

Peralta-Yahya, P.P., Zhang, F., del Cardayre, S.B., and Keasling, J.D. (2012). Microbial engineering for the production of advanced biofuels. Nature *488*, 320-328.

Pérez-Pantoja, D., González, B., and Pieper, D.H. (2010). Aerobic Degradation of Aromatic Hydrocarbons. In Handbook of Hydrocarbon and Lipid Microbiology, K.N. Timmis, ed. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 799-837.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth *8*, 785-786.

Phattarasukol, S., Radey, M.C., Lappala, C.R., Oda, Y., Hirakawa, H., Brittnacher, M.J., and Harwood, C.S. (2012). Identification of a p-Coumarate Degradation Regulon in *Rhodopseudomonas palustris* by Xpression, an Integrated Tool for Prokaryotic RNA-Seq Data Processing. Appl Environ Microbiol *78*, 6812-6818.

Piepenbreier, H., Fritz, G., and Gebhard, S. (2017). Transporters as information processors in bacterial signalling pathways. Mol Microbiol *104*, 1-15.

Pietri, R., Zerbs, S., Corgliano, D.M., Allaire, M., Collart, F.R., and Miller, L.M. (2012). Biophysical and structural characterization of a sequence-diverse set of solute-binding proteins for aromatic compounds. J Biol Chem *287*, 23748-23756.

Plácido, J., and Capareda, S. (2015). Ligninolytic enzymes: a biotechnological alternative for bioethanol production. Bioresources and Bioprocessing *2*.

Plaggenborg, R., Overhage, J., Steinbuchel, A., and Priefert, H. (2003). Functional analyses of genes involved in the metabolism of ferulic acid in *Pseudomonas putida* KT2440. Appl Microbiol Biotechnol *61*, 528-535.

Plaggenborg, R., Steinbuchel, A., and Priefert, H. (2001). The coenzyme A-dependent, non-beta-oxidation pathway and not direct deacetylation is the major route for ferulic acid degradation in *Delftia acidovorans*. FEMS microbiology letters *205*, 9-16.

Porter, A.W., and Young, L.Y. (2014). Benzoyl-CoA, a universal biomarker for anaerobic degradation of aromatic compounds. Advances in Applied Microbiology *88*, 167-203.

Pott, R.W., Howe, C.J., and Dennis, J.S. (2014). The purification of crude glycerol derived from biodiesel manufacture and its use as a substrate by *Rhodopseudomonas palustris* to produce hydrogen. Bioresour Technol *152*, 464-470.

Quistgaard, E.M., Low, C., Guettou, F., and Nordlund, P. (2016). Understanding transport by the major facilitator superfamily (MFS): structures pave the way. Nat Rev Mol Cell Biol *17*, 123-132.

Rabenhorst, J., and Hopp, R. (1991). Process for the preparation of vanillin (Google Patents).

Rabie, E., Serem, J.C., Oberholzer, H.M., Gaspar, A.R., and Bester, M.J. (2016). How methylglyoxal kills bacteria: An ultrastructural study. Ultrastructural pathology *40*, 107-111.

Rahmanpour, R., and Bugg, T.D. (2013). Assembly in vitro of *Rhodococcus jostii* RHA1 encapsulin and peroxidase DypB to form a nanocompartment. FEBS J *280*, 2097-2104.

Rahouti, M., Seigle-Murandi, F., Steiman, R., and Eriksson, K.E. (1989). Metabolism of Ferulic Acid by *Paecilomyces variotii* and *Pestalotia palmarum*. Appl Environ Microbiol *55*, 2391-2398.

Raimunda, D., Long, J.E., Sassetti, C.M., and Arguello, J.M. (2012). Role in metal homeostasis of CtpD, a Co(2)(+) transporting P(1B4)-ATPase of *Mycobacterium smegmatis*. Mol Microbiol *84*, 1139-1149.

Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P.F., Marita, J.M., Hatfield, R.D., Ralph, S.A., Christensen, J.H.*, et al.* (2004). Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl- propanoids. Phytochemistry Reviews *3*, 29-60.

Rensing, C., Fan, B., Sharma, R., Mitra, B., and Rosen, B.P. (2000). CopA: An *Escherichia coli* Cu(I)-translocating P-type ATPase. Proc Natl Acad Sci U S A *97*, 652-656.

Revuelta, J.L., Buey, R.M., Ledesma-Amaro, R., and Vandamme, E.J. (2016). Microbial biotechnology for the synthesis of (pro)vitamins, biopigments and antioxidants: challenges and opportunities. Microb Biotechnol *9*, 564-567.

Rey, F.E., Heiniger, E.K., and Harwood, C.S. (2007). Redirection of metabolism for biological hydrogen production. Appl Environ Microbiol *73*, 1665-1671.

Reyes, N., Ginter, C., and Boudker, O. (2009) Transport mechanism of a bacterial homologue of glutamate transporters. Nature *462*(7275) 880-885.

Reyes, L.H., Almario, M.P., and Kao, K.C. (2011). Genomic library screens for genes involved in n-butanol tolerance in *Escherichia coli*. PLoS One *6*, e17678.

Reyes, L.H., Almario, M.P., Winkler, J., Orozco, M.M., and Kao, K.C. (2012). Visualizing evolution in real time to determine the molecular mechanisms of n-butanol tolerance in *Escherichia coli*. Metab Eng *14*.

Rice, A.J., Park, A., and Pinkett, H.W. (2014). Diversity in ABC transporters: type I, II and III importers. Crit Rev Biochem Mol Biol *49*, 426-437.

Rodriguez, G.M., and Smith, I. (2006). Identification of an ABC transporter required for iron acquisition and virulence in *Mycobacterium tuberculosis*. J Bacteriol *188*, 424-430.

Rojas, A., Duque, E., Mosqueda, G., Golden, G., Hurtado, A., Ramos, J.L., and Segura, A. (2001). Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. J Bacteriol *183*, 3967-3973.

Romagnoli, S., and Tabita, F.R. (2006). A novel three-protein two-component system provides a regulatory twist on an established circuit to modulate expression of the cbbI region of *Rhodopseudomonas palustris* CGA010. J Bacteriol *188*, 2780-2791.

Roosild, T.P., Castronovo, S., Healy, J., Miller, S., Pliotas, C., Rasmussen, T., Bartlett, W., Conway, S.J., and Booth, I.R. (2010). Mechanism of ligand-gated potassium efflux in bacterial pathogens. Proc Natl Acad Sci U S A *107*, 19784-19789.

Ruhl, J., Schmid, A., and Blank, L.M. (2009). Selected *Pseudomonas putida* strains able to grow in the presence of high butanol concentrations. Appl Environ Microbiol *75*, 4653-4656.

Rutherford, B.J., Dahl, R.H., Price, R.E., Szmidt, H.L., Benke, P.I., Mukhopadhyay, A., and Keasling, J.D. (2010). Functional genomic study of exogenous n-butanol stress in *Escherichia coli*. Appl Environ Microbiol *76*, 1935-1945.

Saha, B.C. (2003). Hemicellulose bioconversion. J Ind Microbiol Biotechnol *30*, 279-291.

Sainsbury, P.D., Hardiman, E.M., Ahmad, M., Otani, H., Seghezzi, N., Eltis, L.D., and Bugg, T.D. (2013). Breaking down lignin to high-value chemicals: the conversion of lignocellulose to vanillin in a gene deletion mutant of *Rhodococcus jostii* RHA1. ACS Chem Biol *8*, 2151-2156.

Salaemae, W., Booker, G.W., and Polyak, S.W. (2016). The Role of Biotin in Bacterial Physiology and Virulence: a Novel Antibiotic Target for *Mycobacterium tuberculosis*. Microbiology spectrum *4*.

Salmon, R.C., Cliff, M.J., Rafferty, J.B., and Kelly, D.J. (2013). The CouPSTU and TarPQM transporters in *Rhodopseudomonas palustris*: redundant, promiscuous uptake systems for lignin-derived aromatic substrates. PLoS One *8*, e59844.

Sato, T., Yokota, S., Uchida, I., Okubo, T., Usui, M., Kusumoto, M., Akiba, M., Fujii, N., and Tamura, Y. (2013). Fluoroquinolone resistance mechanisms in an *Escherichia coli* isolate, HUE1, without quinolone resistance-determining region mutations. Front Microbiol *4*.

Scheffel, F., Fleischer, R., and Schneider, E. (2004). Functional reconstitution of a maltose ATP-binding cassette transporter from the thermoacidophilic gram-positive bacterium *Alicyclobacillus acidocaldarius*. Biochim Biophys Acta *1656*, 57-65.

Scheuring, S., Goncalves, R.P., Prima, V., and Sturgis, J.N. (2006). The photosynthetic apparatus of *Rhodopseudomonas palustris*: structures and organization. J Mol Biol *358*, 83-96.

Schmitt, L., Benabdelhak, H., Blight, M.A., Holland, I.B., and Stubbs, M.T. (2003). Crystal Structure of the Nucleotide-binding Domain of the ABC-transporter Haemolysin B: Identification of a Variable Region Within ABC Helical Domains. Journal of Molecular Biology *330*, 333-342.

Sharma, R., Rensing, C., Rosen, B.P., and Mitra, B. (2000). The ATP hydrolytic activity of purified ZntA, a Pb(II)/Cd(II)/Zn(II)-translocating ATPase from *Escherichia coli*. J Biol Chem *275*, 3873-3878.

Singh, D., Mishra, K., and Ramanthan, G. (2015). Bioremediation of Nitroaromatic Compounds. In Wastewater Treatment Engineering, M. Samer, ed. (Rijeka: InTech), p. Ch. 02.

Singh, N.P., and Khan, A. (1995). Acetaldehyde: genotoxicity and cytotoxicity in human lymphocytes. Mutation Research *337*, 9-17.

Slotboom, D.J. (2014). Structural and mechanistic insights into prokaryotic energy-coupling factor transporters. Nat Rev Microbiol *12*, 79-87.

Steed, P.R., and Fillingame, R.H. (2014). Residues in the polar loop of subunit c in *Escherichia coli* ATP synthase function in gating proton transport to the cytoplasm. J Biol Chem *289*, 2127-2138.

Steeves, V., Forster, H., Pommer, U., and Savidge, R. (2001). Coniferyl alcohol metabolism in conifers -- I. Glucosidic turnover of cinnamyl aldehydes by UDPG: coniferyl alcohol glucosyltransferase from pine cambium. Phytochemistry *57*, 1085-1093.

Strachan, C.R., Singh, R., VanInsberghe, D., Ievdokymenko, K., Budwill, K., Mohn, W.W., Eltis, L.D., and Hallam, S.J. (2014). Metagenomic scaffolds enable combinatorial lignin transformation. Proc Natl Acad Sci U S A *111*, 10143-10148.

Strugatsky, D., McNulty, R., Munson, K., Chen, C.K., Soltis, S.M., Sachs, G., and Luecke, H. (2013). Structure of the proton-gated urea channel from the gastric pathogen *Helicobacter pylori*. Nature *493*, 255-258.

Sutcliffe, I.C., and Russell, R.R. (1995). Lipoproteins of gram-positive bacteria. J Bacteriol *177*, 1123-1128.

Sutherland, J.B., Crawford, D.L., and Pometto, A.L., 3rd (1983). Metabolism of cinnamic, p-coumaric, and ferulic acids by *Streptomyces setonii*. Canadian Journal of Microbiology *29*, 1253-1257.

Sweet, G.D., Kay, C.M., and Kay, W.W. (1984). Tricarboxylate-binding proteins of *Salmonella typhimurium*. Purification, crystallization, and physical properties. J Biol Chem *259*, 1586-1592.

Takahashi, H., Inagaki, E., Kuroishi, C., and Tahirov, T.H. (2004). Structure of the *Thermus thermophilus* putative periplasmic glutamate/glutamine-binding protein. Acta crystallographica Section D, Biological crystallography *60*, 1846-1854.

Takatsuka, Y., Chen, C., and Nikaido, H. (2010). Mechanism of recognition of compounds of diverse structures by the multidrug efflux pump AcrB of *Escherichia coli*. Proc Natl Acad Sci U S A *107*, 6559-6565.

Tenaillon, O., Barrick, J.E., Ribeck, N., Deatherage, D.E., Blanchard, J.L., Dasgupta, A., Wu, G.C., Wielgoss, S., Cruveiller, S., Médigue, C.*, et al.* (2016). Tempo and mode of genome evolution in a 50,000-generation experiment. Nature *536*, 165-170.

ter Beek, J., Guskov, A., and Slotboom, D.J. (2014). Structural diversity of ABC transporters. J Gen Physiol *143*, 419-435.

Tikh, I.B., Held, M., and Schmidt-Dannert, C. (2014). BioBrick compatible vector system for protein expression in *Rhodobacter sphaeroides*. Appl Microbiol Biotechnol *98*, 3111-3119.

Toms, A., and Wood, J.M. (1970). Degradation of trans-ferulic acid by *Pseudomonas acidovorans*. Biochemistry *9*, 337-343.

Torres, B.R., Aliakbarian, B., Torre, P., Perego, P., Domínguez, J.M., Zilli, M., and Converti, A. (2009). Vanillin bioproduction from alkaline hydrolyzate of corn cob by *Escherichia coli* JM109/pBB1. Enzyme and Microbial Technology *44*, 154-158.

Uzdavinys, P., Coincon, M., Nji, E., Ndi, M., Winkelmann, I., von Ballmoos, C., and Drew, D. (2017). Dissecting the proton transport pathway in electrogenic Na+/H+ antiporters. Proc Natl Acad Sci U S A *114*, E1101-E1110.

Valadares, L.F. (2014). Biobased fibers and materials in Brazil. Chemical and Biological Technologies in Agriculture *1*, 16.

van den Berg, B. (2005). The FadL family: unusual transporters for unusual substrates. Curr Opin Struct Biol *15*, 401-407.

van den Berg, B., Black, P.N., Clemons, W.M., Jr., and Rapoport, T.A. (2004). Crystal structure of the long-chain fatty acid transporter FadL. Science *304*, 1506-1509.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010). Lignin biosynthesis and structure. Plant Physiol *153*, 895-905.

Verhoef, S., Ballerstedt, H., Volkers, R.J., de Winde, J.H., and Ruijssenaars, H.J. (2010). Comparative transcriptomics and proteomics of p-hydroxybenzoate producing *Pseudomonas putida* S12: novel responses and implications for strain improvement. Appl Microbiol Biotechnol *87*, 679-690.

Wainhouse, D., Cross, D.J., and Howell, R.S. (1990). The role of lignin as a defence against the spruce bark beetle *Dendroctonus micans*: effect on larvae and adults. Oecologia *85*, 257-265.

Walton, N.J., Mayer, M.J., and Narbad, A. (2003). Vanillin. Phytochemistry *63*, 505-515.

Wang, J., Fulford, T., Shao, Q., Javelle, A., Yang, H., Zhu, W., and Merrick, M. (2013a). Ammonium transport proteins with changes in one of the conserved pore histidines have different performance in ammonia and methylamine conduction. PLoS One *8*, e62745.

Wang, J.F., Xiong, Z.Q., Li, S.Y., and Wang, Y. (2013b). Enhancing isoprenoid production through systematically assembling and modulating efflux pumps in *Escherichia coli*. Appl Microbiol Biotechnol *97*, 8057-8067.

Wang, L., Nie, Y., Tang, Y.Q., Song, X.M., Cao, K., Sun, L.Z., Wang, Z.J., and Wu, X.L. (2016). Diverse Bacteria with Lignin Degrading Potentials Isolated from Two Ranks of Coal. Front Microbiol *7*, 1428.

Wang, X., Cheng, X., Sun, D., and Qi, H. (2008). Biodecolorization and partial mineralization of Reactive Black 5 by a strain of *Rhodopseudomonas palustris*. Journal of Environmental Sciences (China) *20*, 1218-1225.

Wang, Y., and Tajkhorshid, E. (2010). Nitric oxide conduction by the brain aquaporin AQP4. Proteins *78*, 661-670.

Waugh, D.S. (2016). Crystal structures of MBP fusion proteins. Protein Sci *25*, 559-571.

Weaver, P.F., Wall, J.D., and Gest, H. (1975). Characterization of *Rhodopseudomonas capsulata*. Archives of Microbiology *105*, 207-216.

Webber, M.A., Bailey, A.M., Blair, J.M., Morgan, E., Stevens, M.P., Hinton, J.C., Ivens, A., Wain, J., and Piddock, L.J. (2009). The global consequence of disruption of the AcrAB-TolC efflux pump in *Salmonella enterica* includes reduced expression of SPI-1 and other attributes required to infect the host. J Bacteriol *191*, 4276-4285.

Wiener, M.C., and Horanyi, P.S. (2011). How hydrophobic molecules traverse the outer membranes of gram-negative bacteria. Proc Natl Acad Sci U S A *108*, 10929-10930.

Wijekoon, C.J., Udagedara, S.R., Knorr, R.L., Dimova, R., Wedd, A.G., and Xiao, Z. (2017). Copper ATPase CopA from *Escherichia coli*: Quantitative Correlation between ATPase Activity and Vectorial Copper Transport. J Am Chem Soc *139*, 4266-4269.

Wilkens, S. (2015). Structure and mechanism of ABC transporters. F1000prime reports *7*, 14.

Winkler, J.D., and Kao, K.C. (2014). Recent advances in the evolutionary engineering of industrial biocatalysts. Genomics *104*, 406-411.

Winnen, B., Hvorup, R.N., and Saier, M.H. (2003). The tripartite tricarboxylate transporter (TTT) family. Research in Microbiology *154*, 457-465.

Wolken, W.A., Lucas, P.M., Lonvaud-Funel, A., and Lolkema, J.S. (2006). The mechanism of the tyrosine transporter TyrP supports a proton motive tyrosine decarboxylation pathway in *Lactobacillus brevis*. J Bacteriol *188*, 2198-2206.

Woodruff, L.B., Boyle, N.R., and Gill, R.T. (2013). Engineering improved ethanol production in *Escherichia coli* with a genome-wide approach. Metab Eng *17*, 1-11.

Wu, S.C., Liou, S.Z., and Lee, C.M. (2012). Correlation between bio-hydrogen production and polyhydroxybutyrate (PHB) synthesis by *Rhodopseudomonas palustris* WP3-5. Bioresour Technol *113*, 44-50.

Wyborn, N.R., Alderson, J., Andrews, S.C., and Kelly, D.J. (2001). Topological analysis of DctQ, the small integral membrane protein of the C4-dicarboxylate TRAP transporter of *Rhodobacter capsulatus*. FEMS microbiology letters *194*, 13-17.

Xing, D., Zuo, Y., Cheng, S., Regan, J.M., and Logan, B.E. (2008). Electricity generation by *Rhodopseudomonas palustris* DX-1. Environmental Science & Technology *42*, 4146-4151.

Xu, W., Chai, C., Shao, L., Yao, J., and Wang, Y. (2016). Metabolic engineering of *Rhodopseudomonas palustris* for squalene production. Journal of Industrial Microbiology & Biotechnology *43*, 719-725.

Yang, W., Tang, H., Ni, J., Wu, Q., Hua, D., Tao, F., and Xu, P. (2013). Characterization of two *Streptomyces* enzymes that convert ferulic acid to vanillin. PLoS One *8*, e67339.

Yardley-Jones, A., Anderson, D., and Parke, D.V. (1991). The toxicity of benzene and its metabolism and molecular pathology in human risk assessment. British Journal of Industrial Medicine *48*, 437-444.

Ye, J., and van den Berg, B. (2004). Crystal structure of the bacterial nucleoside transporter Tsx. The EMBO journal *23*, 3187-3195.

Yildiz, O., Vinothkumar, K.R., Goswami, P., and Kuhlbrandt, W. (2006). Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. The EMBO journal *25*, 3702-3713.

Yoon, S.H., Li, C., Kim, J.E., Lee, S.H., Yoon, J.Y., Choi, M.S., Seo, W.T., Yang, J.K., Kim, J.Y., and Kim, S.W. (2005). Production of vanillin by metabolically engineered *Escherichia coli*. Biotechnol Lett *27*, 1829-1832.

Yung, P.Y., Grasso, L.L., Mohidin, A.F., Acerbi, E., Hinks, J., Seviour, T., Marsili, E., and Lauro, F.M. (2016a). Global transcriptomic responses of *Escherichia coli* K-12 to volatile organic compounds. Sci Rep *6*, 19899.

Yung, T.W., Jonnalagadda, S., Balagurunathan, B., and Zhao, H. (2016b). Transcriptomic Analysis of 3-Hydroxypropanoic Acid Stress in *Escherichia coli*. Appl Biochem Biotechnol *178*, 527-543.

Zachariae, U., Kluhspies, T., De, S., Engelhardt, H., and Zeth, K. (2006). High resolution crystal structures and molecular dynamics studies reveal substrate binding in the porin Omp32. J Biol Chem *281*, 7413-7420.

Zaitseva, J., Jenewein, S., Oswald, C., Jumpertz, T., Holland, I.B., and Schmitt, L. (2005). A molecular understanding of the catalytic cycle of the nucleotide-binding domain of the ABC transporter HlyB. Biochem Soc Trans *33*, 990-995.

Zaldivar, J., Martinez, A., and Ingram, L.O. (1999). Effect of selected aldehydes on the growth and fermentation of ethanologenic *Escherichia coli*. Biotechnol Bioeng *65*, 24-33.

Zamzuri, N.A., Abd-Aziz, S., Rahim, R.A., Phang, L.Y., Alitheen, N.B., and Maeda, T. (2014). A rapid colorimetric screening method for vanillic acid and vanillin-producing bacterial strains. J Appl Microbiol *116*, 903-910.

Zeth, K., and Thein, M. (2010). Porins in prokaryotes and eukaryotes: common themes and variations. Biochem J *431*, 13-22.

Zhang, L., Chen, J., Chen, N., Sun, J., Zheng, P., and Ma, Y. (2013). Cloning of two 5-aminolevulinic acid synthase isozymes HemA and HemO from *Rhodopseudomonas palustris* with favorable characteristics for 5-aminolevulinic acid production. Biotechnol Lett *35*, 763-768.

Zhao, C., Zhang, Y., Chan, Z., Chen, S., and Yang, S. (2015). Insights into arsenic multi-operons expression and resistance mechanisms in *Rhodopseudomonas palustris* CGA009. Front Microbiol *6*, 986.

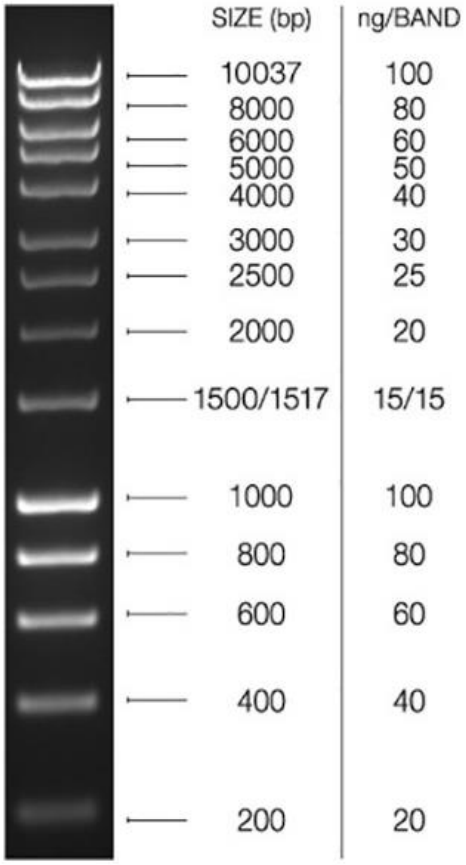
Zhao, L., Zhao, C., Han, D., Yang, S., Chen, S., and Yu, C.P. (2011). Anaerobic utilization of phenanthrene by *Rhodopseudomonas palustris*. Biotechnol Lett *33*, 2135-2140.

Zhou, J., Wang, K., Xu, S., Wu, J., Liu, P., Du, G., Li, J., and Chen, J. (2015). Identification of membrane proteins associated with phenylpropanoid tolerance and transport in *Escherichia coli* BL21. J Proteomics *113*, 15-28.

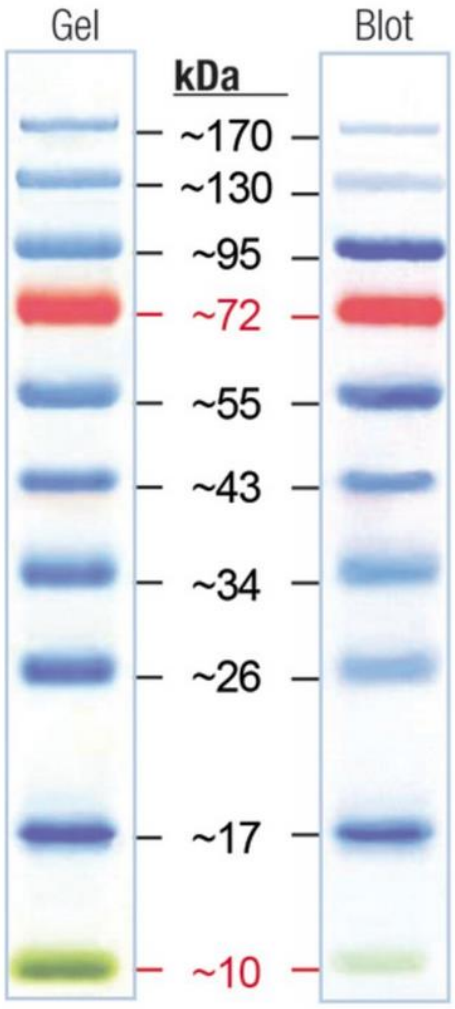
Zhu, L., Li, Y., and Cai, Z. (2015). Development of a stress-induced mutagenesis module for autonomous adaptive evolution of *Escherichia coli* to improve its stress tolerance. Biotechnology for Biofuels *8*, 93.

8. Appendices

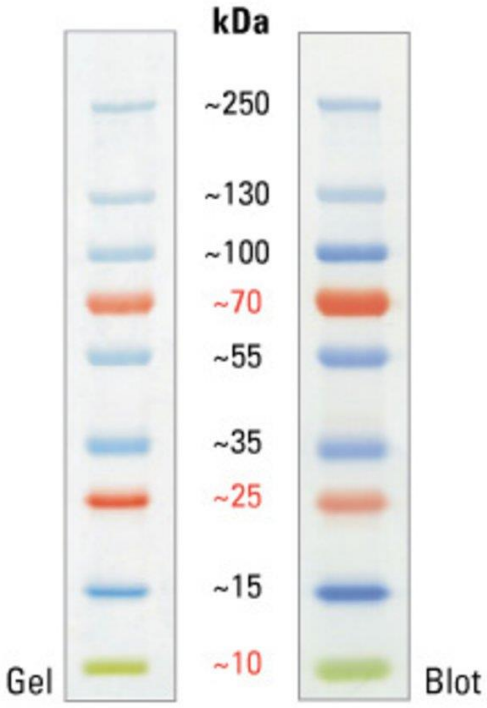
8.1. DNA Ladder : Hyperladder 1



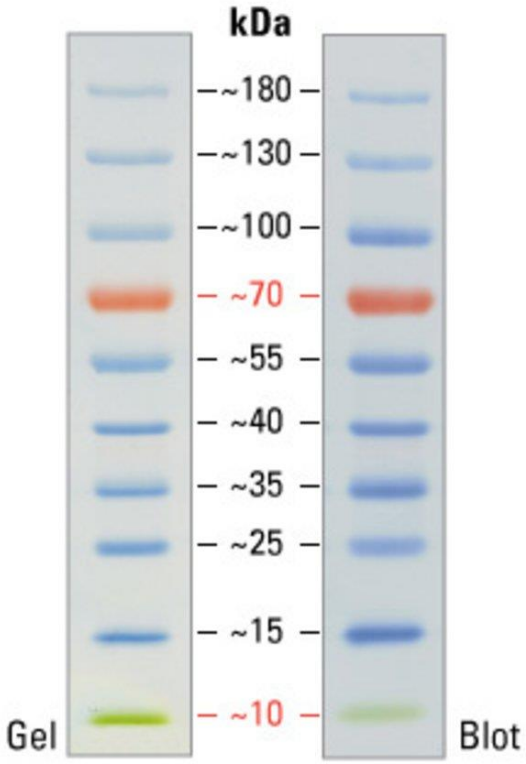
8.2. Protein Ladder A : Fisher BioReagents™ EZ-Run™ Prestained Rec Protein Ladder



8.3. Protein Ladder B : ThermoFisher PageRuler™ Plus Prestained Protein Ladder (10-250)



8.4. Protein Ladder C : ThermoFisher PageRuler™ Prestained Protein Ladder (10-180)



8.5. ^{1}H-NMR reference spectra