The University Of Sheffield.

Anaerobic Degradation of Long Chain Phenylalkane Carboxylates by the Phototrophic Bacterium *Rhodopseudomonas palustris*

BY

Abrar A. Akbar

A thesis submitted in part fulfilment for the degree of

Doctor of Philosophy.

Department of Molecular Biology and Biotechnology,
University of Sheffield.

July 2016

Abstract

Anaerobic bacteria typically degrade many types of aromatic monomers through conversion to benzoyl-CoA which is further degraded by ring reduction, ring cleavage and β-oxidation reactions. *Rhodopseudomonas palustris* has the ability to grow on different side-chain length aromatic compounds ranging from 3-phenylpropanoic acid to 8-phenyloctanoic acid. A β-oxidation or non β-oxidation metabolic pathway could be responsible for degradation of these compounds. To study the pathways and genes involved, a full cellular proteome analysis using mass spectrometry was carried out and the expression patterns for selected genes were measured by qRT-PCR. Two putative feruloyl-CoA synthetases, Fcs1 and Fcs2, in addition to a long chain fatty acid CoA-ligase (RPA1766) were overexpressed in *E. coli* as recombinant proteins and the purified enzymes were able to catalyse the formation of CoA thioesters of phenylcarboxylates with different chain lengths. The kinetic properties of Fcs1 enzyme showed a high kcat/Km value for 7-phenyl heptanoic acid. Activity was also recorded with 5-phenyl valeric acid, 6- phenyl hexanoic acid and 8- phenyl octanoic acid, but with lower kcat/Km values, while in contrast Fcs2 only utilised 3-phenyl propionic acid and the related hydroxycinnamic acid. RPA1766 was able to utilize only 8-phenyloctanoic acid with low affinity. These data along with protein structure modelling support the idea that Fcs1 catalyses the first step in a β-oxidation pathway for long chain phenylalkane carboxylates in *R. palustris*. However, other CoA ligases can contribute, as mutants lacking Fcs1 and Fcs2 genes were able to grow on long chain phenylalkane carboxylic acids. Evidence was obtained that the solute binding proteins of an ABC-transporter previously associated with pimelic acid and long chain fatty acid transportation (RPA3718-RPA3725) could bind certain phenylalkane carboxylates with a K_d value of 1.29 μM for RPA3723 with 7-phenylheptanoic acid.

Dedicated to my daughter "Layla"

You gave me the strength to continue

Acknowledgements

I am profoundly grateful to my supervisor, Prof. Dave Kelly for his guidance and support throughout the entire course of my work. I would like to express my deepest gratitude to him for teaching me the arts of patience and self reliance. I would like to thank Kuwait Institute for scientific research for funding my research.

Next come a group of incredible individuals to whom I owe great thanks for helping me in many different ways during the various stages of my research and for helping me gain experience and knowledge in many aspects of my project. I would like to express my gratitude to Dr. Mark Collin, Dr. Robert Salmon, Dr. Nathan Adam and Mrs. Andrea Hounslow for their help in my project work, and being available. My colleagues in F1 lab thank you for everything.

I am grateful for the love, encouragement, and tolerance of my dearest friend Reem Farsi. For the last 4 years you were the only Family I had in Sheffield so thank you for the precious memories.

Last, but not the least I would like to thank my family: my Mother and Father for their support and unconditional love. To my Husband Ahmad and my little Laila for believing in me and supporting me spiritually throughout my PhD and in my life in general.

Abbreviations

3-p	3-phenylpropionic acid
4-p	4-phenylbutyric acid
5-p	5-phenylvaleric acid
6-p	6-phenylhexanoic acid
7-p	7-phenylheptanoic acid
8-p	8-phenyloctanoic acid
ABC	ATP binding Cassette
AMP	Adenosine monophosphate
APS	ammonium persulphate
ATP	adenosine triphosphate
bp	base pair
CO_2	carbon dioxide
CoA	Coenzyme A
D_2O	deuterium oxide
dH_2O	distilled water
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EAAT	excitatory amino acid transport
ECF	energy coupling factor
EDTA	ethylenediamine tetra-acetic acid
g	gram
H_2O	water
HPLC	high performance liquid chromatography
hr	hour
IPTG	isopropyl β-D-1-thiogalactopyranoside
kan	kanamycin
kb	kilobase
K_d	dissociation constant
kDa	kiloDalton
L	litre

LB	Luria-Bertani
LFQ	Label Free Quantification
LiP	lignin peroxidase
M	Molar
MFS	major facilitator superfamily
mg	milligram
mm	millimetre
mM	millimolar
MnP	manganese-dependent peroxidase
mRNA	messenger ribonucleic acid
MW	molecular weight
NAD(P)	Nicotinamide adenine dinucleotide (phosphate)
NBDs	nucleotide binding domain
nm	nanometer
NMR	Nuclear magnetic resonance
^{o}C	degrees Celsius
PAGE	polyacrylmide gel electrophoresis
PCR	polymerase chain reaction
PEP	phosphoenolpyruvate
pH	hydrogen potential
PMF	proton motive force
PO	phenol oxidases
PPi	pyrophosphate
PYE	peptone-yeast extract
RC	reaction center
RCV	Rhodobacter capsulatus vitamins
RNA	ribonucleic acid
Rnase	ribonuclease
RPM	revolutions per minute
RT-PCR	real time polymerase chain reaction
s	second

SBP	solute binding protein
SBP	solute binding protein
SDS	sodium dodecyl sulphate
TAE	tris acetate EDTA solution
TEMED	N,N,N',N'-tetramethyl- ethane-1,2-diamine
T_m	melting temperature
TMDs	two transmembrane domain
TRAP	triapartite ATP-independent periplasmic transporter
Tris	tris (hydroxymethyl)aminomethane
TSP	trimethyl-silyl propionate
UV	ultraviolet
v/v	concentration, volume/volume
VA	veratryl alcohol
VP	versatile peroxidase
w/v	concentration, weight/volume
WT	wild-type
x g	multiplied by gravitational force
µg	micrograms
µl	microlitre
µM	micromolar

1. Table of Contents

1. Introduction

1.1 Lignin

Lignin is a heterogeneous polymeric substance associated with cellulose and hemi-cellulose (Sjöström 1993). It is a main component of cell wall formation in plants and some algae (Dovgan' & Medvedeva 1983). Furthermore, lignin protects the cellulose polymer from hydrolytic attack by many microbial pathogens (Martinez et al. 2005) and it is considered to be the second most abundant carbon-based material on earth making up almost one third of the dry weight of wood (Battle et al. 2000). Lignin has a phenolic structure with a highly branched three-dimensional configuration, that is strongly joined by covalent cross-linkages and non-covalent forces (Figure 1.1) (Iiyama et al. 1994). Lignin aromatic polymers result mainly from the oxidative compinative coupling of 4-hydroxyphenylpropanoids (Ralph et al. 2004). Monolignols (or hydroxycinnamyl alcohols) such as; coniferyl alcohol, sinaphyl alcohol and p-coumaryl alcohol are the major constituents of lignin (Figure 1.2) (Vanholme et al. 2010). Lignin composition may differ between taxa, species and cell types; for example, there are three main units resulting from monolignol incorporation into the lignin polymer namely, guaiacyl, syringyl, and p-hydroxyphenyl units (Uzal et al. 2009) and different species and taxa may have variant ratios of these units in their lignin that are bound by many different ether and C-C bonds (Vanholme et al. 2010). The β- aryl ether is the most common linkage in lignin, which consists of an ether linkage to another aryl unit at C-2 position of the C3 chain. The biphenyl linkage and hetero- cyclic linkages are also found in lignin and such linkages make lignin highly resistant to breakdown (Mao et al. 2006).

Lignin is used in industry as an excellent source of fuel (Xiao et al. 2012), but burning of lignified wood has its consequence in the production of greenhouse gases. Furthermore, lignin is used as an additive in many industries such as oil well drilling,

concrete production, animal feed and as binders for other industries (Gargulak & Lebo 2000). Important chemicals such as vanillin are also produced from lignin raw material (Calvo-Flores & Dobado 2010), and recent research had shown the potential use of lignin in many industrial applications such as enzymatic production, biocide neutrialization and wood preservation (Gargulak & Lebo 2000). Also lignin removal is very important in many industrial uses such as bioethanol production and paper manufacturing (Galbe & Zacchi 2007; Sixta 2006). Research and studies on lignin degrading microorganisms and their biochemical pathways can help in the improvement of the above mentioned industrial applications that depend mainly on lignin.

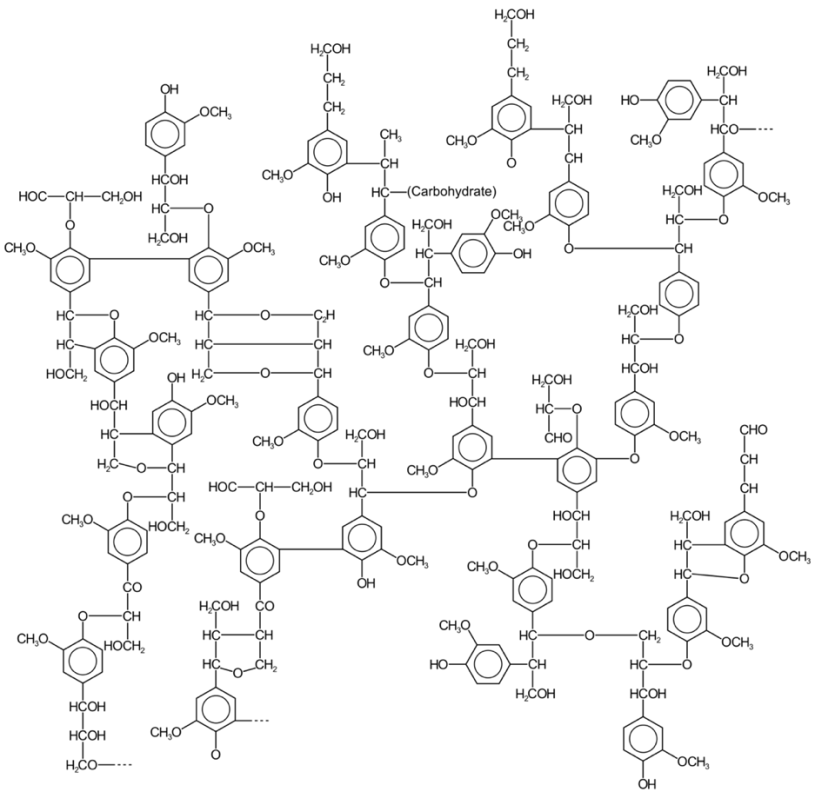


Figure 1.1: Lignin structure. Composed mainly of phenylpropane units that are linked together in a very complex manner.

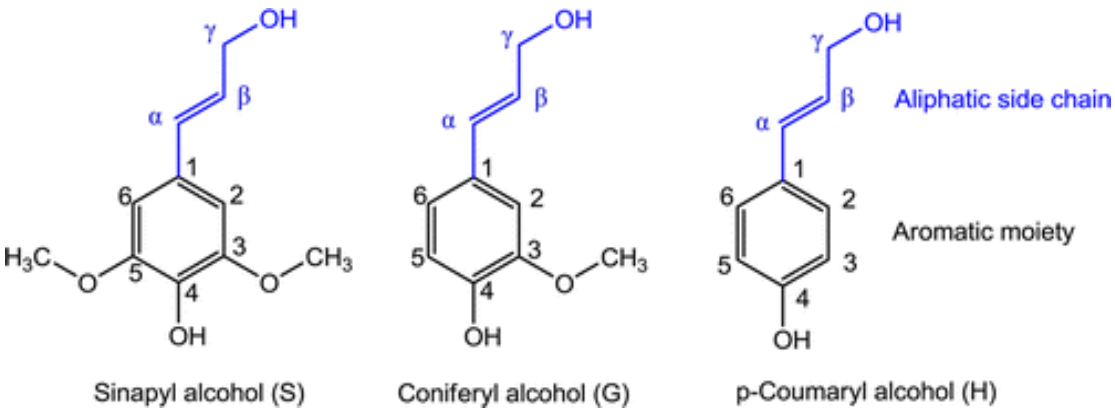


Figure 1.2. Types and Structure of Monolignols.

The initial degradation of lignin by microbial communities leads to the release of many aromatic monomers (Erdtman & Holger 1972). The most common aromatic monomers of lignin are cinnamic acid derivatives such as; caffeic acid, ferulic acid and coumarate. These compounds differ only by the functional groups positioned on the aromatic ring. Other benzoic acids are also released during lignin degradation (Hartley & Ford 1989).

1.2 Microorganisms and Lignin Degradation

Lignin degradation and decomposition are different from other regular repeating unit polymer degradation procedures because of lignin heterogeneity and the diversity of linkages, such as ether and C-C linkages, which are not susceptible to hydrolytic attack (Palmqvist & Hahn-Hägerdal 2000). Such a complex structure is hard to degrade and many microorganisms may only produce partial changes in lignin structure that then leads to full decomposition of lignin by other microorganisms. Many studies indicate that the ability of single microorganisms is highly limited and only a specific group of fungi can degrade lignin efficiently. The *Basidiomycota* and especially white rot fungi are the only organisms capable of utilizing lignin entirely

(Sánchez 2009). The rate of lignin degradation in this group may vary, but some white rot fungi such as *Phellinus nigrolimitatus* remove lignin specifically (Blanchette 1995).

1.2.1 Lignin derived aromatic compounds degradation By Fungi

The capability of white rot fungi to break down lignin depends on two sets of enzymes that are involved in ligninolysis: peroxidases and laccases (Kim et al. 1998). In fungi, the peroxidases are classified into three groups, lignin peroxidases (LiPs), manganese-dependent peroxidases (MnPs) and versatile peroxidases (VP) (Martinez et al. 2005). LiPs, MnPs and VP are heme-containing glycoproteins that needs hydrogen peroxide as an oxidant (Martínez 2002). LiP oxidizes nonphenolic lignin in multi-step electron transfers and generates intermediate radicals that are then decomposed chemically (Wong 2009). This enzyme will cleave the bond between C_α-C_β and cause aromatic ring opening (Wong 2009). LiP can also oxidize phenolic compounds, amines, aromatic ethers, and polycyclic aromatics with relevant ionization potential without the need of mediators because of LiPs high redox potential (Wong 2009). MnP catalyzes Mn(II) peroxide dependent oxidation to Mn(III) (Wong 2009). Mn(III) will then make a complex with oxalate or other chelators and act as a redox-mediator for the phenolic lignin structure, but requires a second mediator to catalyze the oxidation reaction of non-phenolic substrates (Wesenberg et al. 2003). On the other hand, Versatile peroxidases (VP), are heme glycoproteins which combine the substrate specificity of both LiP and MnP. VP has the ability to oxidize Mn(II), veratryl alcohol (VA), phenolic and non phenolic substrates (Heinfling et al. 1998) and has high catalytic properties. The high catalytic

activity of VP is due to its hybrid molecular structure, which combines different substrate binding and oxidation sites (Morgenstern et al. 2008).

Laccases are glycosylated multi-copper oxidoreductases that require the use of molecular oxygen to oxidize phenolic compounds using copper as cofactor (Claus 2004). The oxidation reactions may lead to free radical formation, which acts as transitional substrates for the enzymes (Ferraroni et al. 2007). In addition to these enzymes, there are other lignin degrading enzymes found in fungi such as oxidases which generate hydrogen peroxide, that is required by peroxidases (Martinez et al. 2005). Also enzymes such as, aryl-alcohol dehydrogenases (AAD), cellobiose dehydrogenase (CDH), and quinone reductases (QR) are associated with lignin degradation by fungi (Guillén et al. 1997). Moreover, the degradation of lignin derived aromatic compounds has also been reported in bacteria such as *Actinobacteria* from the *Streptomyces* genus (Berrocal et al. 1997).

1.2.2 Lignin derived aromatic compounds degradation by Bacteria

The importance of studying the bacterial degradation of lignin comes from bacteria's ability to grow under many different environmental conditions and biochemical diversity (Chandra et al. 2007), in contrast to poor fungal stability in extreme environments such as oxygen limitation and high environmental pH (Daniel & Nilsson 1998). Lignin biodegradation is described extensively as an aerobic process, but some researchers have shown that anaerobic microorganisms have the ability to partially degrade lignin (Akin 1980).

1.2.2.1 Aerobic degradation of lignin derived aromatic compounds

Aerobic bacteria of many different genera, such as *Pseudomonas*, *Streptomyces* and

Rhodopseudomonas palustris have the ability to degrade single ring aromatic compounds (Donnelly & Crawford 1988; Malherbe & Cloete 2002). The ability of bacteria to degrade lignin depends mainly on lignin level of crystallization and the varieties of chemical bonds. Two groups of enzymes are involved in the aerobic degradation of lignin: peroxidases and phenol oxidases (Nie et al. 1999). Peroxidases as described in fungi can also be present in bacteria and have the ability to catalyse the oxidation of many substrates such as phenol, aromatic amines, and alkyl peroxides (Sjoblad & Bollag 1981). Moreover, the enzymatic activity of peroxidases can be altered by different environmental compounds, such as heavy metals (Tuomela et al. 2005).

Phenol oxidases (PO) catalyze the oxidation of lignin aromatic compounds by using oxygen as the ultimate electron acceptor to get more readily available substrates (Cullen & Kersten 1996). The phenol oxidases can be grouped into laccases and polyphenol oxidases based on their substrate specificity. The distribution of PO in bacteria was investigated by Alexandre and Zhulin (2000), who used protein sequence database BLAST to compare and investigate the presence of fungal laccase sequences in bacteria. They found many bacterial proteins with significant similarity to the laccase enzymes found in fungi (Alexandre & Zhulin 2000).

There are four aerobic pathways to break down aromatic compounds: The Phenylacetate, Homogenistate, Protocatechuate and Homoprotocatechuate degradation pathways (Hunter et al. 2008). The four pathways are oxygenase-dependent and lead to aromatic ring cleavage. Hydroxyl groups are inserted into the aromatic ring by oxygenases that use molecular oxygen to carry out such a reaction. The insertion of hydroxyl groups by dioxygenase will permit aromatic ring cleavage

(Harayama 1992). Ring cleavage can either be *ortho* or *meta* cleavage, where *ortho* cleaved substrates such as *cis,cis*-muconate enter the β-ketoadipyl-CoA pathway to produce succinyl-CoA or acetyl-CoA (Figure 1.3) (Diaz et al. 2013) on the other hand, *meta* cleaved compounds such as hydroxymuconate-semialdehyde enter many pathways to produce pyruvate or acetyl-CoA (Figure 1.3) (Diaz et al. 2013). So after ring cleavage, the product passes through many different metabolic pathways to either produce cellular energy or valuable bi-products (Sugimoto et al. 1999). In aerobic pathways, before oxygenolytic ring cleavage the phenolic hydroxyl groups are introduced by oxygenases to activate the aromatic ring (Carmona et al. 2009). In anaerobic catabolism, the aromatic ring is reduced and the phenolic hydroxyl groups are reductively removed, leading to the formation of alicyclic compounds (Carmona et al. 2009).

Another aerobic scheme for aromatic ring cleavage depends on the use of oxygenases to produce non aromatic epoxide (Diaz et al. 2013). This pathway shares many features with the anaerobic pathway thus it's called the hybrid pathway. The aromatic compound in this pathway is activated to CoA thioesters by a CoA ligase and ring cleavage is carried out hydrolytically. Then the non-aromatic CoA thioesters metabolise through β-oxidation-like reactions to produce β-ketoadipyl-CoA (Figure 1.3) (Fuchs et al. 2011). In the hybrid pathway benzoyl-CoA is the most common intermediate (Carmona et al. 2009).

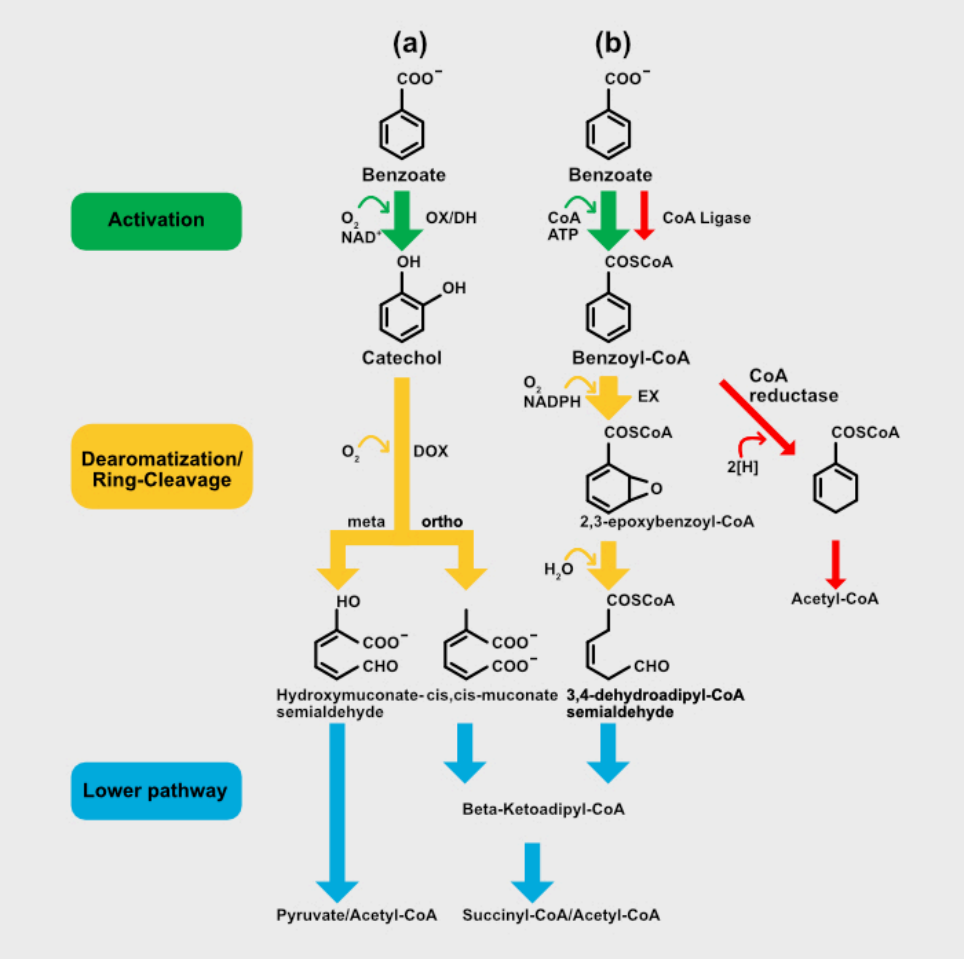


Figure 1.3. Aerobic pathway for benzoic acid degradation. a) Classical aerobic pathway. b) hybrid pathway. Both pathways have activation (green), dearomatizaton (yellow) and further reduction in lower pathways to produce the central metabolites (blue). *Ortho* cleaved catechol and benzoyl-CoA in the hybrid pathway will produce the common β-ketoadipyl-CoA intermediate. While in the anaerobic degradation pathway in orange ring reduction occurs and further β-oxidation reactions leads to the formation of acetyl-CoA (Diaz et al. 2013).

1.2.2.2 Anaerobic degradation of lignin derived aromatic compounds

Some anaerobic bacteria have the ability to degrade lignin derived compounds (Balba & Evans 1977). In aerobic bacteria, the degradation of aromatic compounds depends mainly on molecular oxygen as a common co-substrate for degradation enzymes. However, in anaerobic bacteria, the degradation of lignin derived aromatic

compounds is extremely slow and oxygen-dependent reactions are substituted by different set of reactions (Benner et al. 1984). In chemoheterotrophs, these reactions will often use Fe(III), nitrate or sulfate as the final electron acceptors (Boll & Fuchs 2005). Two distinctive enzymes are associated with phenol carboxylation: a phenylphosphate synthase and phenylphosphate carboxylase. The first enzyme can convert phenol into phenylphosphate using ATP, which in turn result in the formation of phosphate and AMP (Schmeling et al. 2004); and the second one can convert phenylphosphate into 4 -hydroxybenzoate and phosphate with the help of CO_2, using Mn as the metal cofactor (Schuhle & Fuchs 2004). Then, 4-hydroxybenzoate is activated by specific AMP-forming carboxylic acid coenzyme A ligase to form 4-hydroxybenzoyl-CoA (Biegert et al. 1993) Then, 4- hydroxybenzoyl-CoA reductase dehydroxylates 4-hydroxybenzoyl-CoA to form benzoyl-CoA (Brackmann & Fuchs 1993). Benzoyl-CoA reductase then reduces benzoyl-CoA to cyclic dienoyl-CoA compound and this is the key step in aromatic anaerobic metabolism (Boll et al. 1997). *Thauera aromatica* benzoyl-CoA reductase is the most studied benzoyl-CoA reductase; it is a heterotetramer protein that is ATP dependent (Boll & Fuchs 1998). Finally, cyclic dienoyl-CoA compound is further converted to three molecules of acetyl-CoA and one molecule of CO_2 by different downstream pathways such as the β-oxidation pathway (Carmona et al. 2009). The Acetyl-CoA molecules produced can then be used by normal metabolic pathways in the cell where it is normally oxidized via the citric acid cycle (Figure 1.4) (Harrison & Harwood 2005). Instead, the metabolism of branched chain aromatic compounds such as different side chain length phenylalkane carboxylic acids are more complex and has not been fully investigated. But what is certain about these compounds, is the need for the CoA derivative to activate the reaction. CoA thioesters are membrane impermeable (Rather et al. 2011).

Also the formation of CoA thioesters in anaerobic pathways lowers the redox potential of electron transfer by a significant amount for example the first electron

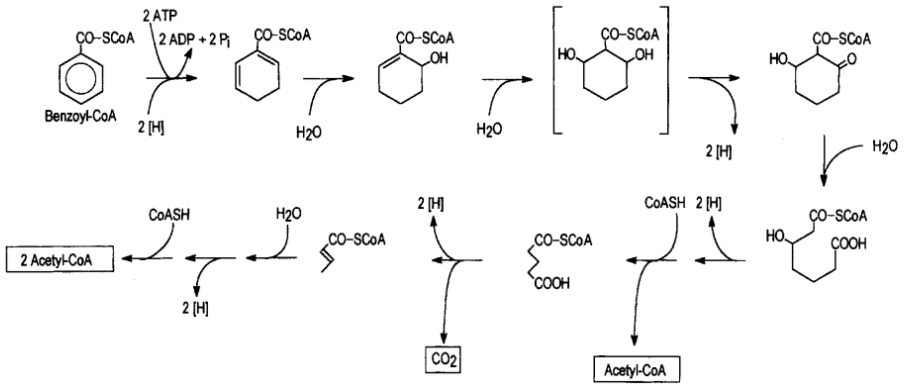


Figure 1.4. Benzoyl-CoA anaerobic degradation pathway. Benzoyl-CoA transformation to acetyl-CoA and CO_2 (Heider & Fuchs 1997).

transfer for CoA substituted benzene ring is lowered by 1.3 V compared to the unsubstituted benzene ring (Boll et al. 2000). The CoA thioester has an electron-withdrawing characteristics which can facilitate the reduction of the aromatic ring. CoA- ligase isoenzymes are responsible for the formation of CoA-thioesters and each enzyme has its substrate specificity and regulation. CoA-ligase synthesis is generally induced by the presence of its substrate (Schneider et al. 2003). Different CoA ligases were identified for anaerobic aromatic compound degradation and they mainly differ in their catalytic and structural properties. Overall CoA ligases consist of a single type of subunit ranging in size from 48 to 65 kDa and require MgATP and CoA for activity. (Mohamed & Fuchs 1993). CoA-ligases for aromatic acids cleave ATP to AMP and PP_i and they differ mainly in the type of the aromatic acid substrate they activate in addition to the substrate range. For instance, 4-hydroxybenzoate-CoA

ligase isolated from *R. palustris* can activate benzoate, 4-fluorobenzoate, cyclohex-1,4-diene-1-carboxylate and cyclohex-2,5-diene-l-carboxylate (Gibson et al. 1994), while phenylacetate-CoA ligase from *A. evansii* can only activate phenylacetate (El-Said Mohamed 2000).

Microbial communities and especially bacterial cells must obtain lignin derived aromatic compounds through dedicated transport systems. Therefore, as such transporters were also a feature of the project to be described in this thesis, I will now describe the structure and basic properties of bacterial solute transport systems.

1.3 Solute transport systems in bacteria

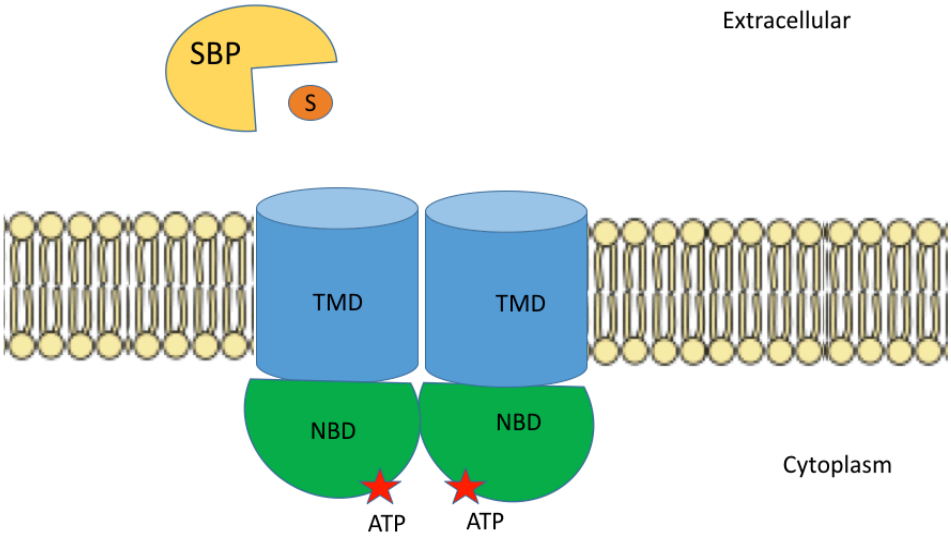
Bacterial cells usually interact with their surrounding environment by their membrane and transporters to supply the cell with external nutrients and to remove unwanted compounds. Gram-negative bacteria have an outer and inner membrane that is made mainly from phospholipids - the lipid layer gives a hydrophobic nature to the membrane that may allow for direct diffusion of some molecules such as urea and very hydrophobic fatty acids (Paula et al. 1996). On the other hand, charged molecules and those with a hydrophilic nature such as carboxylic acids cannot cross the membrane passively (Hancock & Nikaido 1978). So bacteria evolved transport systems to transport solutes across their membranes. There are two type of transport system; the primary and secondary systems. These two systems differ mainly on the basis of their energy coupling (Tchieu et al. 2001). Primary systems use the energy of hydrolysis of chemical bonds, while a secondary system uses the energy from electrochemical ion gradients (Jaehme & Slotboom 2015). In primary transport systems in bacteria, the molecules are transported mainly by the hydrolysis of ATP (Davidson et al. 2008). Other modes of transportation such as the photochemical or

electrical reactions are also present in bacteria (Davidson et al. 2008). The primary transport systems that are an ATP dependent hydrolysis type can be divided into ; the P-type ATPases, the ATP synthetases and ATP binding cassette transporters (ABC transporters) (Pedersen & Carafoli 1987). P-type ATPases function mainly to transport ions across the membrane by hydrolysing ATP through three cytoplasmic domains. (Toyoshima et al. 2013). ATP synthase transporters usually function to produce ATP, by using Na^+ ion gradients or the proton motive force (PMF) (Boyer 1997; von Ballmoos et al. 2008). The ABC transporters are often function to transport various ligands such as; sugars, amino acids and other molecules that are mostly hydrophilic, by hydrolysing ATP. ABC transporters will be discussed in detail due to their involvement in the transportation of many aromatic compounds such as coumarate (Salmon et al. 2013).

1.3.1 ATP binding Cassette (ABC) Transporters.

It is a large super family of transporters that are found in almost every living organism and are very important for the survival of most living cells. There are different types of ABC transporters (Type I, type II, exporters and ECF transporters). ABC transport systems share two important features; two nucleotide binding domain (NBDs) and two transmembrane domain (TMDs) (Rees et al. 2009). The function of the different domains is to power the active export and import of solute through the membrane in prokaryotic cells while in eukaryotic cells they can only export solutes (Pohl et al. 2005). In ABC exporters the NBD and TMD domains are joined in different ways resulting in one chain, two chains or four chains transporters. While in ABC importers the central TMD and NBD subunit are distinct proteins that will result in homo or hetero-dimeric TMDs linked to homodimeric NBDs. NBDs are positioned in the

cytoplasm at the cell membrane as a dimer and their sequence is highly conserved between many living organisms (Davidson et al. 2008). TMD sequences on the other hand are widely varied because of the wide range of ligands transported by them. ATP binding and hydrolysis in NBDs are considered to be responsible for the opening and closing of TMDs (Higgins & Linton 2004). Prokaryotic organism's ability to transport solutes is assisted by an additional component to the eukaryotic system known as the solute binding protein (SBP) which can bind to very low solute concentrations and deliver the solute to the binding site in TMDs (Figure 1.5) (Davidson & Chen 2004) In Gram negative bacteria, SBPs are located in the periplasmic layer, while in Gram positive bacteria, these SBPs are lipoproteins attached to the outer layer of the plasma membrane (Gilson et al. 1988). Recent studies also revealed that not all ABC importers in prokaryotes posses SBPs; some energy coupling factor transporters (ECF) do not require SBPs for their substrate transportation. In its place, a membrane bounded component of the transporter binds substrate. That shows how diverse ABC importers are but in general the function of ABC transporters starts with ATP hydrolysis that occurs on the NBD part which will then cause a conformational change in the TMD part. That conformational change in turn will allow for alternating access from inside and outside the cell (Davidson et al. 2008). ABC transporters catalytic cycle starts by SBP binding to the ligand (for importers) or binding of ligand to TMDs directly (for exporters). Then two MgATP molecules bind to the NBDs which will lead to NBDs and depending on transporter type the TMDs switching between the in–outward or out-inward facing conformation. After that, ATP hydrolysis, ADP, phosphate and transport substrate release associated with NBD dissociation resets the transporter (Wilkens 2015). ABC transporter systems were thought to function in a unidirectional manner but a recent study



showed that it can work in bidirectional manner, such as the histidine permease from *Salmonella typhimurium* (Hosie et al. 2001).

Figure 1.5. ABC transporter general structure. Soluble binding proteins (SBPs) that binds to substrate interact with two TMDs. Energy is obtained from ATP hydrolysis that occurs in NBDs.

1.3.1.1 Solute binding proteins (SBPs)

SBPs in Gram negative bacteria are a common feature and function by binding to desired solutes associated with a particular system with high affinity and specificity. They are made of two domains (C and N lobes) joined by a flexible joint (hinge) in which a central cleft is formed between the two domains to make the binding pocket for the ligands (Quiocho & Ledvina 1996). When ligands bind to SBPs they cause a conformational change which will close the two domains around the ligand (Berntsson et al. 2010a). The closed conformation of SBP with the ligand then transmits a signal through the transmembrane subunits to the ATPase subunit on the

cytoplasmic side of the membrane. Then ATP binding and hydrolysis leads to the opening of the pore and the release of ligands from SBP (Davidson et al. 1992). SBPs can be distinguished on the basis of their tertiary structure into six structural families that are mainly different in the hinge region topology between the two domains of the protein (Berntsson et al. 2010a). Another way to distinguish SBPs is by their substrate specificity range for example; metal SBPs, carbohydrate SBPs, amino acids SBPs and peptide SBPs) (Quiocho & Ledvina 1996; Tam & Saier Jr 1993). SBPs are not only associated with ABC transporters because recent studies have shown that SBPs in prokaryotic organisms can be part of other membrane transporters such as the tripartite ATP-independent periplasmic transporters (TRAP) (Gonin et al. 2007) and the two component regulatory systems (Neiditch et al. 2006).

1.3.1.2 ABC transporter Type I, II and ECF

Type I and Type II importers differ mainly in their transmembrane helices where Type I have fewer transmembrane helices than Type II. Type I transporters depend mainly on SBPs for their activity so Type I transporters show very low level of ATPase activity in the absent of SBPs (Davidson et al. 1992). Type I importers are the most studied system, e.g. the maltose transporter (MalFGK2) (Oldham & Chen 2011;Chen et al. 2013). Ligand transportation by Type I depends on an alternating conformational change mechanism between inward and outward facing conformation which will allow for ligand translocation to be opened to only one side of the membrane. The most studied Type I transporter mechanism is that of maltose transporter (Figure 1.6). Maltose transportation is initiated by the binding of substrate to SBP (MalE) which will bind to free TMD domain MalFGK (Oldham & Chen

2011), the binding of MalE and two ATP molecules would cause a conformational change to MalFGK. The TMD subunits of the transporter would

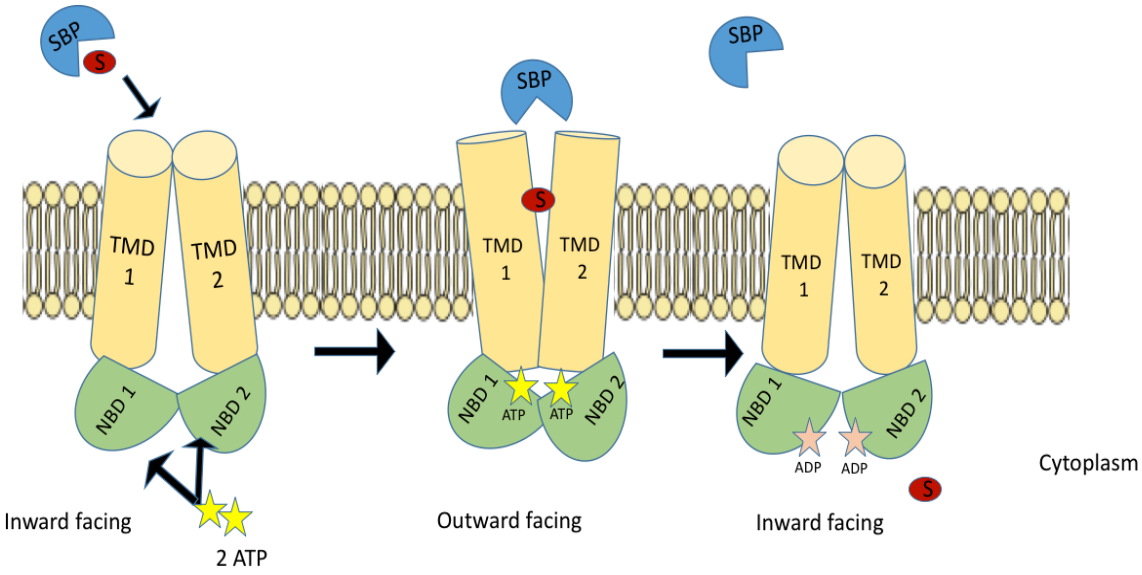


Figure 1.6. ABC transporter Type I. The inward-facing ABC transporter binds to the substrate that is delivered by solute binding protein. Two molecules of ATP binds to NBDs, which leads to an outward facing conformational change. The substrate leaves SBP and bind to TMDs. ATP hydrolysis occur followed by product release, NBD dissociation and transporter reset to the inward-facing conformation (Wilkens 2015).

rotate toward the center of the transporter and that would lead to the formation of outward facing conformational change. The maltose then dissociates from SBP and crosses the cavity of the two TM domains in their open confirmation and ATP hydrolysis occur in NBDs; this leads to the adoption of inward conformation, allowing the maltose transition to the cytoplasm and the release of SBP (Khare et al. 2009).

Type II bacterial ABC transporters have different TM domains and substrate translocation mechanism through their system when compared with other types. In

addition to that, the presence and the concentration of a substrate weakens SBP association with Type II transporter system (Vigonsky et al. 2013). So SBP form high association with their Type II transporter system in the absence of a substrate.

Whereas in Type I ABC transport system, the presence of a substrate is required to form a complex between SBP and the Type I transporter, in Type II transporters TMD conformational modifications are limited to the central helix that controls access to translocation pathway. While Type I transporters conformational modification include a domain-wide modification as described previously. The best studied model for Type II transporters is that of *E. coli* BtuCDF system that mediates the import of vitamin B_{12} into the cell (Reynolds et al. 1980). B_{12} uptake by the BtuF SBP and binding to TMDs leads to the dimerization of the NBD (Karpowich et al. 2003). This is mediated by coupling helices in the TMD, which change in conformation and alter the position of several gates. (Hollenstein et al. 2007) Cytoplasmic gate I present between TM helices 4 and 5 closes after substrate binding which in turn opens a gate to periplasm (periplasm gate) (Joseph et al. 2011). Then cytoplasmic gate II that is present between TM helices 2 and 3 closes allowing for B_{12} translocation and the formation of occluded conformation change (Korkhov et al. 2012). ATP binding and hydrolysis leads to NBD separation which in turn pull TM coupling helices. Cytoplasmic gate II then opens to create an inward facing position which allows B_{12} to enter the cytoplasm (Figure 1.7) (Korkhov et al. 2012).

The ECF type transport family comprises ABC primary transport systems that do not use typical periplasmic solute binding proteins. It is usually associated with Gram positive bacteria and has the same structure as the other importers especially the two cytosolic ATPase domains (EcfA and EcfA′) that provide the system with energy, on

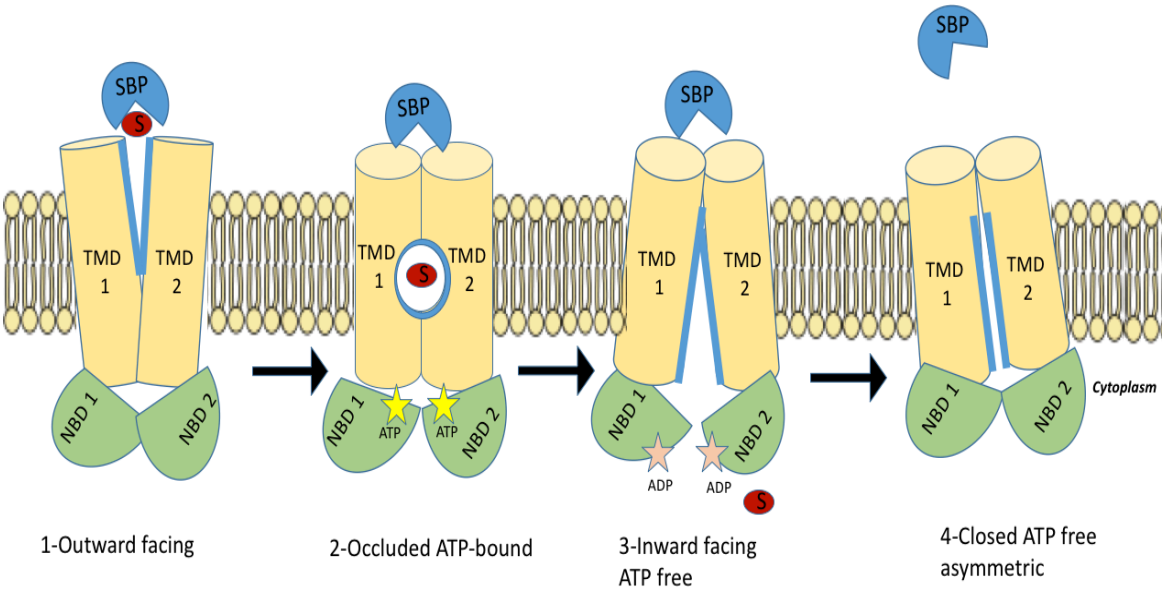


Figure 1.7. ABC transporter type II. B_{12} transportation via the BtuCDF system where the blue line shows the exact translocation of substrate that bond first to SBP that in turn binds to TMDs that posses an outward conformation. Two ATP molecules binds to NBDs which leads to the formation of occluded conformation where the substrate is sealed into the mid membrane cavity. ATP hydrolysis and NBD dissociation allow substrate release to the cytoplasm (Korkhov et al. 2012).

the other hand the two other components differ in structure and mechanisim (Slotboom 2014). EcfT mediates solute translocation across the membrane and couple the energy produced by cytosolic ATPase domains to EcfS (membrane embedded domain) that capture the solute in the periplasmic layer (Zhang 2013). The mechanistic details of solute translocation are not fully understood but the theme is common to the other ABC systems (Wang et al. 2013). EcfT helices interact with NBDs (EcfA and EcfA′) and that mediates a conformational change through to the periplasm facing EcfS proteins consequently inducing solute translocation (Figure

1.8) (Zhang 2013; Erkens et al. 2011). Other modes of solute translocation were proposed and that include pore gated mechanism and toppling mechanism (Slotboom 2014).

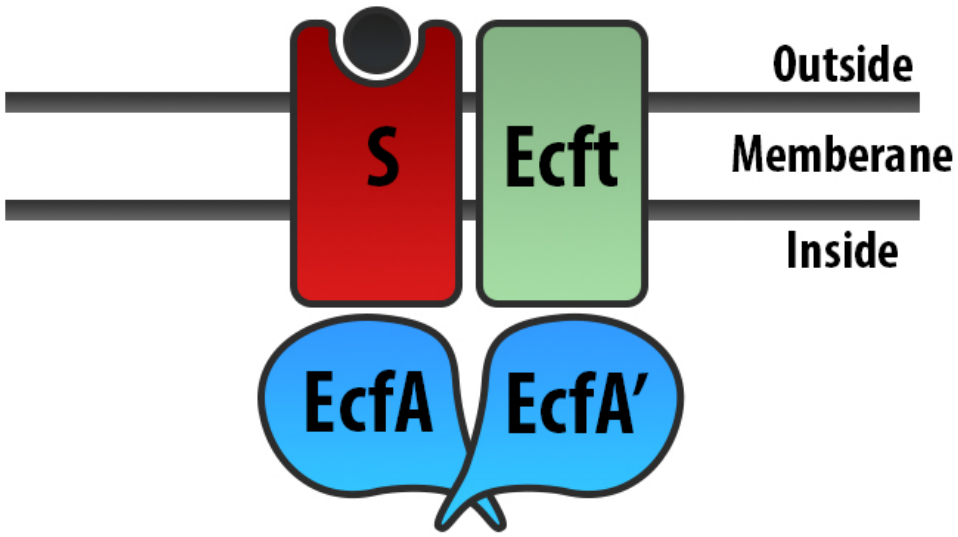


Figure 1.8. ABC transporter ECT Type. S represent the EcfS which bind to the substrate in the periplasm. Then EcfT interact with NBDs (EcfA and EcfA′) where ATP hydrolysis occurs.

1.3.2 Secondary transport systems:

Secondary transport systems such as; Uniport, Symport and Antiport transporters, transport solutes by using the energy from an electrochemical ion gradient across the membrane against a concentration gradient (Forrest et al. 2011). The system of transportation is much simpler than the ABC transport system and is coded by one gene product. They are mainly members of the major facilitator superfamily (MFS) but other distinct groups such as the excitatory amino acid transport (EAAT) that is associated with glutamatergic synaptic transportation is also part of the secondary transport system family (Mulligan & Mindell 2013). Others include; Tripartite ATP-

independent periplasmic (TRAP) transporters and Tripartite tricarboxylate transporters (TTT) which consist mainly of two membrane protein domains and a SBP. The structure of secondary transport systems that belong to the MFS family is mainly made of 12 TM helices or sometime 24 helices (Pao et al. 1998). In the Uniport system a single ion or charged molecule passes through a specific transmembrane carrier. For example, potassium ions with a positive charge pass through the Uniport systems driven by the negative charge on the inside of the cytoplasmic membrane (Figure 1.9) (Forrest et al. 2011). while, the Symport systems transport neutral, negative and positive molecules against their concentration gradient depending on the free energy obtained from opposing electrochemical gradient of a coupling ion. One example is the *E. coli* LacY permease, which brings lactose into the cell by using the proton electrochemical gradient as its driving force (Figure 1.9) (Kaczorowski & Kaback 1979). Antiport systems are a co-transporter of an ion or a molecule sharing the same or different charge in opposite directions across the membrane. The *E. coli* DcuB system is an example of an Antiport system where the intracellular concentration of succinate that results from anaerobic growth on fumarate, exchanges with the external fumarate on the basis of its solute gradient (Figure 1.9) (Six et al. 1994).

1.4 *Rhodopseudomonas palustris*

R. palustris is a gram negative purple non-sulphur phototrophic bacterium that belongs to the *Alphaproteobacteria* class. It has an extraordinary metabolic versatility and is found in soil, ponds and industrial waste. *R. palustris* can grow by photoautotrophic, photoheterotrophic, chemoautotrophic and chemoheterotrophic mode of metabolism (Larimer et al. 2004).

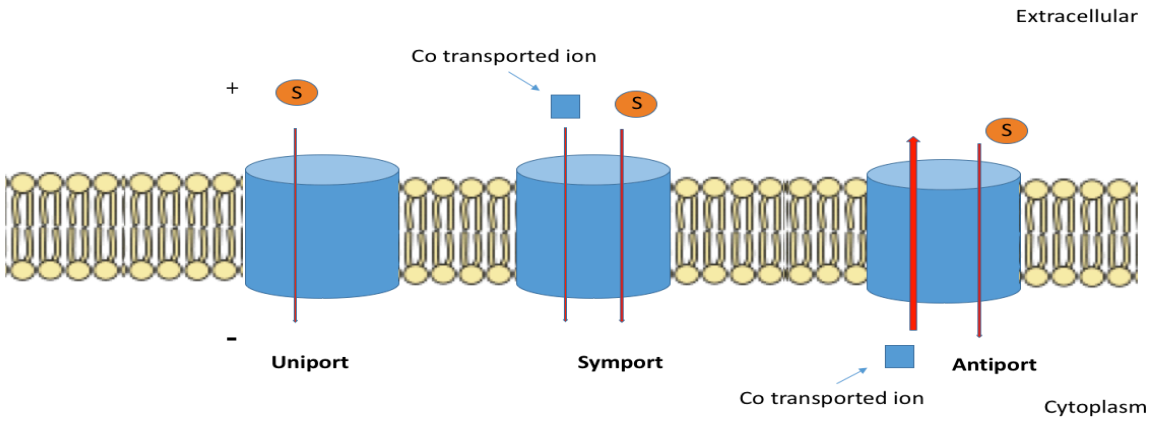


Figure 1.9. Examples of secondary transport systems.

R. palustris cells are rod shaped in general but based on the stage of cell division that occurs through budding cell morphology changes. The optimum temperature for growth is 30°C and under stressful growth conditions *R. palustris* cells form biofilms (Kanazawa et al. 2010). *R. palustris* has the capability to grow on a number of aromatic compounds, which makes it a good candidate for bioremediation processes (Dutton & Evans 1969). It also can produce energy by fixing nitrogen into ammonia and it can produce hydrogen gas as a side reaction of nitrogenase enzyme (Simpson & Burris 1984). *R. palustris* is able to grow photoheterophically on a wide range of aromatic lignin monomers such as, cinnamate, caffeate and coumarate (Harwood & Gibson 1988). *R. palustris* photosynthetic apparatus is made of four essential membrane protein complexes; first peripheral LH2 antenna complexes that function as light harvesting complex and energy transporter to the second complex (core complex), which is composed of an antenna complex (LH1) associated with photochemical reaction center (RC). LH1 funnels light energy to the RC which in turn produces a charge separation to catalyze the oxidation of water soluble carrier, usually a *c*-type cytochrome and reduce a lipid soluble electron carrier, typically ubiquinone in a site known as Q_B site (Gall & Robert 1999). After the capture of two photons, the reduced quinone leaves Q_B site and moves to the cytochrome bc_1 complex. This

complex will close the electronic circuit by re-reducing the water- soluble carriers with electrons from the lipid- soluble carrier (Hu et al. 2002). The last complex is the F_OF_1 H^+-ATP-synthetase where ATP synthesis occurs. *R. palustris* contains multiple types of LH2 complexes that absorb light at different wavelengths, which gives it a greater ability to absorb light at different intensities and wavelengths (Gall & Robert 1999).

1.5 *R. palustris* anaerobic degradation of aromatic compounds

Anaerobic degradation of many aromatic compounds have been investigated in *R. palustris*, especially those derived from phenol and benzoate. These compounds would go into several metabolic pathways to produce benzoyl-CoA which subsequently get reduced by benzoyl-CoA reductase (Fuchs 2008). The metabolism of many branched chain aromatic compounds especially those resulting from lignin degradation were investigated in the past couple of years. For example; enzymes for the conversion of *p*-coumarate to 4-hydroxybenzaldehyde and acetyl-CoA are encoded by the *couAB* operon system and regulated by CouR, a MarR family transcriptional regulator (Hirakawa et al. 2012). The *couAB* operon codes for a non β-oxidation pathway where CouB is a *p*-coumarate CoA ligase which adds a CoA group to the substrate, then enoyl-CoA hydratase (CouA) cleaves acetyl CoA and produces 4-hydroxybenzaldehyde that is used by aldehyde dehydrogenase and 4-hydroxybenzoate CoA ligase to produce 4-hydroxybenzoyl-CoA (Figure 1.10)(Pan et al. 2008). 4-hydroxybenzoyl-CoA then go through reductive dehydroxylation to produce water and benzoyl-CoA (Carmona et al. 2009). A similar non β-oxidation pathway was also seen in *Pseudomonas fluorescens* grown on ferulate (Narbad & Gasson 1998).

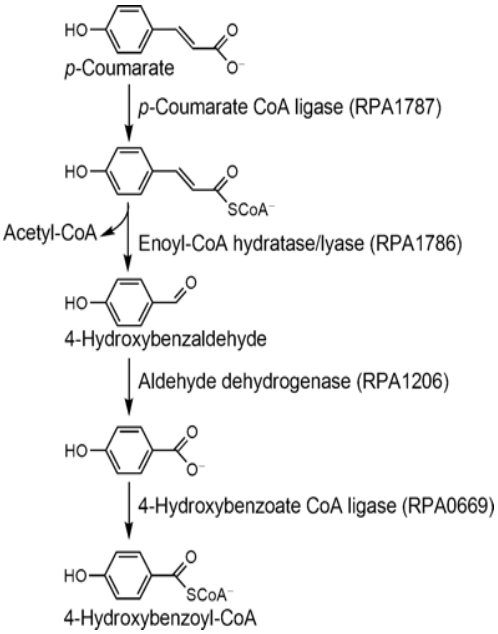


Figure 1.10. Non β-oxidation pathway. Coumarate is converted to 4-hydroxybenzoyl-CoA in *R. palustris* (Pan et al. 2008).

On the other hand, other aromatic compounds such as different chain length phenylalkane carboxylic acids that have similar structures but with a longer side chain were not utilized by CouB (Hirakawa et al. 2012). *R. palustris* has the ability to grow on a range of branched chain phenylpropenoid compounds derived from lignin (Harwood & Gibson 1988). Elder et al. (1992) investigated the growth of *R. palustris* on different side chain phenylalkane carboxylic acids ranging from 3-phenylpropanoic acid to 8-phenyloctanoic acid and suggested that these are metabolised via a β-oxidation pathway that differs from the non-β-oxidation pathway described above mainly in when the acetyl-CoA is removed. In this pathway two carbons are removed as acetyl-CoA at the end of the pathway to produce benzoyl-CoA (Figure 1.11). No genes or proteins for the degradation of different chain length phenylalkane carboxylic acids using such a β-oxidation pathway have been identified however. Other microorganisms such as the yeast *Rhodotorula rubra* uses a β-

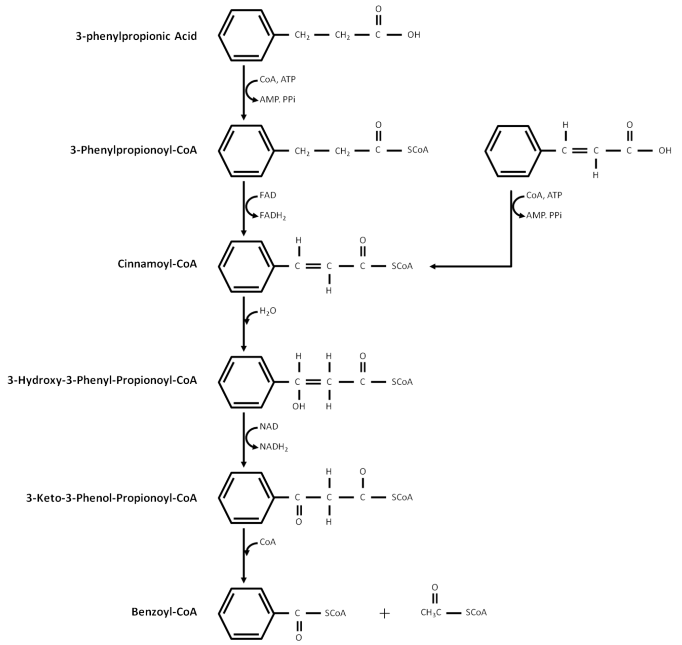


Figure 1.11. β-oxidation pathway. 3-phenylpropionic acid conversion to Benzoyl-CoA.

oxidation pathway to convert ferulate (4-hydroxy-3-methoxycinnamate) to a simple aromatic acid (vanillate) (Huang et al. 1993).

In the study done by Elder *et al* (1992), *R. palustris* was able to grow very well on phenylalkane carboxylic acids with different side chain length and the growth yield was higher on acids with an odd number in their side chain such as 3-phenylpropanoic acid, 5-phenylvaleric acid and 7-phenylheptanoic acid, in contrast to the growth on acids with even number in their side chain such as 4-phenylvaleric acid, 6-phenylhexanoic acid and 8-phenyloctanoic acid (Figure 1.12). The reason for this is the production of phenylacetyl-CoA by the β-oxidation route of even numbered side chain phenylalkane carboxylic acids, which *R. palustris* cannot grow on in comparison to benzoyl-CoA produced when cells were grown on acids with odd numbers of carbon in their side chain (Figure 1.13) (Elder et al. 1992).

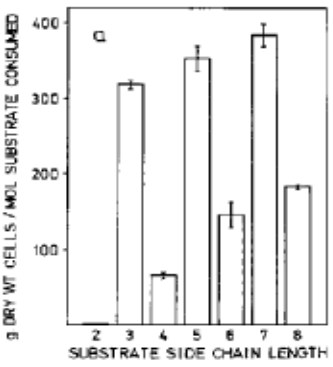


Figure 1.12: Apparent growth yields obtained from anaerobic growth of *R. palustris* on aromatic acids with side chain lengths of: 2, phenylacetic acid ; 3, 3-phenylpropionic acid ; 4. 4-phenylbutyric acid ; 5, 5-phenylvaleric acid ; 6, 6-phenylhexanoic acid ; 7, 7-phenylheptanoic acid ; 8, 8-phenyloctanoic acid (Elder et al. 1992).

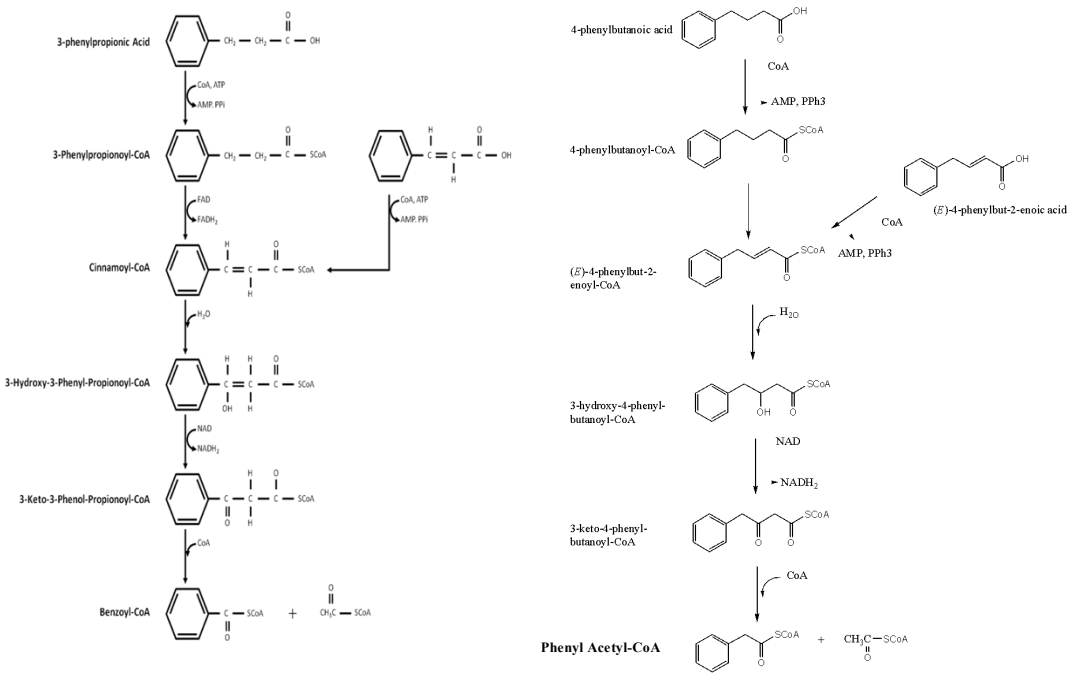


Figure 1.13 Proposed mechanism of anaerobic degradation for 3-phenylpropanoic acid and 4-phenylbutyric acid. Longer side chains will require further rounds of β-oxidation. Odd numbered side chains will produce benzoyl-CoA while even numbered side chains will produce phenylacetyl-CoA. (Elder et al. 1992).

Moreover, the *R. palustris* genome codes for almost 500 transporters some of which are relavant for understanding uptake of aromatic compounds. For example, CouPSTU that belongs to the ABC transport system and TarPQM that is a tripartite ATP-independent periplasmic (secondary transporter) were identified to transport lignin monomers such as; coumarate, ferulate and cinnamate into the *R. palustris* cell (Salmon et al. 2013). The binding affinity of the SBPs of both transport systems were in the nanomolar K_d value range and coumarate showed the highest affinity when compared with the other lignin derived aromatic monomers (Salmon et al. 2013). The molecular pathway and transporters associated with complex aromatic compound uptake and degradation in *R. palustris* is still unclear (Larimer et al. 2004). However, investigating and studying complex aromatic degradation pathways will help in the development of new biotechnological processes.

1.6 Aims and objectives

The overall aim of the study was to identify possible CoA-ligases and transport systems associated with different length side chain phenylalkane carboxylic acids uptake and metabolism in *R. palustris* and to identify the genes and proteins for other steps in the pathway.

The specific objectives of this study were:

- To identify proteins associated with 5-phenylvaleric acid metabolism in *R. palustris* and possible degradation pathway up-regulated in the presence of 5-phenylvaleric acid.

- To clone, purify and biochemically characterize the CoA-ligase enzymes being up regulated and produced during growth on 5-phenylvaleric acid.

- To investigate the substrate specificity of CoA-ligase enzymes with different length phenylalkane carboxylic acids

- To clone, purify and characterize SBPs associated with any up-regulated transport systems in 5-phenylvaleric acid grown cells.

2. Materials and Methods

2.1 Materials

Chemicals used in this study were obtained from Invitrogen, Sigma-Aldrich, Melford, Fisher Scientific, Oxoid, Becton, Agilent Technologies and Acros Organics.

2.2 Microorganisms used

Different strains of bacteria were used (Table 2.1).

Strain	Strain description	Source
***Rhodopseudomonas palustris* (CGA009)**	Genome sequenced strain of an environmental source	Prof. D. K Newman California Institute of Technology
Δ*fcs1*	Unmarked deletion of *fcs1* gene in CGA009 strain	R. Salmon University of Sheffield
Δ*fcs1* and *fcs2*	Unmarked deletion of *fcs2* gene in *Δfcs1* strain	This study
***Escherichia coli* DH5α™**	Genotype: F- Φ80*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *recA*1 *endA*1 *hsdR*17 (rk-, mk+) *phoA supE*44 λ-*thi*-1 *gyrA*96 *relA*1	Invitrogen®
***Escherichia coli* BL21 (DE3)**	F- *ompT hsdSB* (rBmB-) *gal dcm* (DE3)	Invitrogen®
***Escherichia coli* TOP 10**	*hsdR mcrA lacZ*ΔM15	Invitrogen®
S-17	RP4-2 (*Tc::Mu, Mn:Tn7*) integrated into the chromosome *thi pro hsdR* $hsdM^{+}$ *recA* . Tp^{R} (Sm^{R})	Simon *et al.* (1983)

Table 2.1 Strains used in this study

2.2.2 Media Preparation and Organism Growth.

Media was prepared using distilled water (dH_2O) following manufacturers instructions and sterilized by autoclaving at 121ºC for 20 minutes.

2.2.2.1 Antibiotics

Antibiotics (Sigma-Aldrich) stock solution (1000x concentrated) were dissolved in sterilized distilled water (dH_2O) then were filter sterilized using (0.2 μm) syringe filters and stored at 4ºC. Final concentration of antibiotic used was 50-100 μg ml^{-1} for carbenicilin and kanamycin.

2.2.2.2 *R. palustris* Growth Media

Liquid peptone yeast extract (PYE) media with succinic acid supplements was used routinely to grow *R. palustris* (5g/L each of peptone (Becton), yeast extract (Oxoid) and succinic acid (Oxoid) were mixed together with dH_2O and sterilized by autoclaving at 121 ºC for 20 min then stored at room temperature. Solid peptone yeast extract media were prepared in the same manner as the liquid PYE with the addition of 13g/L agar (Melford), sterilized by autoclaving then poured into sterile petri dishes and stored at 4 ºC. If the media needed to be supplemented with antibiotics it was allowed to cool to 50 ºC before the addition of any supplement.

Rhodobacter capsulatus vitamins (RCV) minimal media (Weaver et al. 1975) was used for the growth of *R. palustris*. The media was prepared by mixing the following (final concentrations); sodium succinate (30 mM), $(NH_4)_2SO_4$ (7.5 mM), $MgSO_4.7H_2O$ (1 mM), $CaCl_2.2H_2O$ (0.5 mM), Na_2.EDTA (0.05 mM), NaCl (1 mM), $FeSO_4$. $7H_2O$ (0.04 mM), $MnSO_4$ (0.01 mM), $NaMoO_4$. $2H_2O$ (3.3 mM), $ZnSO_4$. 7 H_2O (0.9 mM), $Cu(NO_3)_2$. 2 H_2O (0.2 mM), Thiamine HCl (6 μM), Biotin (0.2 μM) and Nicotinic acid (8 mM) with dH_2O. Media were sterilized by autoclaving at 121ºC for 20 minutes and stored at room temperature. Other components were added after autoclaving and when the media reach 50 ºC;

- Potassium phosphate buffer pH 7 (0.64 M).
- Sodium bicarbonate (10 mM final) that was sterilized by 0.2 μm filter before addition
- Vitamin mixture (p-aminobenzonoic acid (17 μM), Vitamin B_{12} (0.01 μM), Calcium pantothenate (4 μM))

Succinate was substituted with 3 mM aromatic acid in experiments which needed the aromatic acid to be the sole carbon source for *R. palustris*. the stocks were adjusted to pH 7 before filter sterilization and addition to RCV medium.

In RCV solid media. agar was added to 1.5% (w/v) concentration before autoclaving and was poured into sterilized petri-dishes. Plates were stored at 4 ºC.

Solid RCV media for mutation experiments had an additional 10% (w/v) sucrose (Sigma Aldrich) before autoclaving.

2.2.2.3 *R. palustris* Growth Experiment

R. palustris was grown anaerobically under the light (photoheterotrophically) in liquid RCV media. Energy saving 15W LED bulbs were used at 15-30 cm distance from growth vessels and the vessels were kept on top of a magnetic stirring plate with a magnetic bar (Stuart) that prevented the formation of biofilm. Cultures were grown at 30 ºC. To maintain strain integrity, *R. palustris* was only sub-cultured 2 or 3 times and then new cultures were inoculated from original glycerol stocks. *R. palustris* grown on RCV agar plates were streaked from glycerol stock and were cultured at 30 ºC inside a dark closed container.

2.2.2.4 Growth of *E. coli*

LB agar was used to grow *E. coli* and was supplemented with different types of antibiotic based on the experiment. The antibiotic was added after autoclaving and plates were stored at 4ºC before use. After inoculation plates were incubated at 37 ºC for 18 hours. For *E. coli* grown in liquid cultures, the flasks were incubated at 37ºC and were left shaking at 250 rpm.

2.3 Quantitative Reverse Transcription Polymerase Chain Reaction (*q*RT-PCR)

2.3.1 Preparation of *R. palustris* cells

R. palustris cultures were prepared in RCV media containing 3 mM benzoic acid (control) or 5-phenylvaleric acid (phenylalkane carboxylic acid) (section 2.2.2.2 and 2.2.2.3). Cultures were gown photoheterotrophically at 30°C to OD_{660} 0.6-0.7 and then samples of 10 ml were transferred into cold 15 ml tubes containing 2 ml cold 5% (v/v) phenol pH 4.3 and 95% (v/v) ethanol. The tube was then kept on ice for 15 minutes then was centrifuged at 3220 x g for 10 minutes at 4°C. The supernatant was discarded and the pellets were stored at -80°C before RNA extraction.

2.3.2 RNA Extraction

RNA samples were extracted from cell pellets using TRIzol® reagent according to manufacturer's protocol. RNA concentration was measured using a Nanodrop spectrophotometer (Thermo Fisher Scientific) with RNase free water as a blank. 1% (w/v) agarose gel electrophoresis was used to analyse RNA samples before diluting them to 20 ng/µl. RNA samples were stored at -80 °C.

2.3.3 Primer used in this study

Primers were designed using the primer3 program (http://primer3.ut.ee/). Primers were designed to be 18-22 bp in length with Tm ~60°C to amplify a fragment of 150-200 base pairs. Primers were ordered from Sigma-Aldrich. Primers sequence are listed in (Table 2.2)

Gene	Primer	Sequence
RPA1702	1702_QPCR_F	5'-CGTCAAGGCCAAGATGATCG-3'
	1702_QPCR_R	5'-TCTTGTCCGGATCGTTGTGA-3'
RPA1704	1704_QPCR_F	5'-ACTTCGGTAACAAGGACGGT-'3
	1704_QPCR_R	5'-CGTGCAGCAAGTAGTGGATC-'3
RPA1705	1705_QPCR_R	5'-CCACAACGATCGGGGAAAC-3'
	1705_QPCR_F	5'-CCTTCTCGGCATTGTCCTTG -3'
RPA1706	1706_QPCR_F	5'-CTTTCATCAAGCTCGGCCTC -3'
	1706_QPCR_R	5'-ATCCTCCATCGCGATCACC -3'
RPA1707 (Fcs1)	1707_QPCR_F	5'-TTCCTATTCACCTCCGGCTC-3'
	1707_QPCR-R	5'-CGTAGTAGGTCGGCGAGATT-3'
RPA1708	1708_QPCR-F	5'-TCGCTGTTTCTGGTCGATCT -3'
	1708_QPCR_R	5'-GATGTGCTTGATCGCCTGG -3'
RPA1709	1709_QPCR_F	5'-TGATCTGGCAGCAGGAAGAG -3'
	1709_QPCR_R	5'-CTTCTGGCCGTTGATGATCC -3'
RPA1766	1766_QPCR_F	5'-GAGTTCTACGGCTCGACTGA-'3
	1766_QPCR_R	5'-TCCGGCTTGTTGTGATAGGT-'3
RPA1787 (Fcs2)	Fcs2_QPCR_F	5'-GGATGATCTGCGCCAATCAG-'3
	Fcs2_QPCR_R	5'-CAGATTGCGGATGGTGGAAG-'3
RPA3724	3724_QPCR_F	5'-GGTGGCGATCCTTTATCAGA-'3
	3724_QPCR_R	5'-AGCGACTTCAGCTTGACGAT-'3
RPA3720	3720_QPCR_F	5'-GTTCAAGATCCTCGGCAAGA-'3
	3720_QPCR-R	5'-TCTCGATGACATACGCCTTG-'3
RpoD	RpoD_QPCR_F	5'-CGACTTCCTGCGCAACTATC -3'
	RpoD_QPCR_R	5'-GGTTGGTGTACTTCTTGGCG -3'

Table 2.2. qPCR primers

2.3.4 *q*RT-PCR Reaction

A 96-well plate was used and each reaction mixture contained: 5 µl of 20 ng/µl RNA sample, 10 µl 2x Brilliant III SYBR Green qRT-PCR mastermix (Agilent), 2 µl Primers (5 pmol/µl) 0.2 µl RT/ RNase Block enzyme DTT (100 mM) 18 µl Free-RNase water (Sigma). qRT-PCR reaction was preformed under the following conditions:

Section	Number of cycles	Duration	Temperature (°C)
1	1	10 min	50
2	1	3 min	95
3	40	15 sec	95
		20 sec	60
4	1	1 min	95
		30 sec	55
		30 sec	95

2.4 Mass spectrometry based proteomics.

R. palustris cells were grown photoheterotrophically at 30 °C to late exponential phase of OD_{660} 0.6-0.7 in Benzoic acid RCV culture (control) or 5-phenylvaleric acid culture (phenylalkane carboxylic acid). Cells were grown in 3 biological replicates for each condition and cells were harvested by centrifugation at 10,000 x g for 10 minutes. Cells were washed twice with sterile 50 mM phosphate buffer (pH 7.4) and sonicated 4 times at a frequency of 16 microns amplitude for 20 seconds then centrifuged at 18000 x g for 20 minutes. After that, 50 µg (20 µl) of protein

from each growth condition was run on an SDS gel at 160 V for 15 minutes then the gel was washed several times with dH_2O and the gel was cut into small pieces. Proteins were reduced by 10 mM dithiothreitol and 50 mM ammonium bicarbonate and alkylated with 55 mM iodoacetamide and 50 mM ammonium bicarbonate The gel piece then was dried out using a vacuum drier for 30 minutes. After that, Proteins in the dried gel pieces were digested for 4 hours with 0.5 μg trypsin at 37 °C.

For each growth condition the samples were pooled together and evaporated to dryness and reconstituted in 20 mM ammonium formate buffer pH 10. Samples were fractionated by high-pH reverse phase HPLC. Fractionated peptides were separated in a second dimension of low pH reversed phase chromatography coupled to mass spectrometry using an Orbitrap Elite system. Results were analysed using MaxQuant with default parameters (Cox et al. 2009) and a minimum of two valid intensity values for a given peptide were required in the three biological replicates of 5-phenylvaleric acid grown cell extract. Any missing intensity values were replaced by the minimum intensity value measured in benzoic acid extract samples. Mass spectrometry and data processing was performed by Dr Richard Beniston and Dr Mark Collins of the University of Sheffield BioMics facility.

2.5 DNA extraction and manipulation

Reagents were obtained from Promega, Qiagen and Sigma-Aldrich.

2.5.1 Plasmids

All plasmids used are listed in Table 2.3.

2.5.2 DNA preparation

DNA preparation was preformed as recommended by manufacturer's instructions

2.5.2.1 Extraction of genomic DNA

Wizard® Genomic DNA Purification Kit (Promega) was used to extract high quality genomic DNA

2.5.2.2 Extraction of plasmid DNA

QIAprep® Spin Miniprep Kit (Qiagen) was used to isolate Plasmids.

Plasmid	Description	Resistance	Source
pET21a (+)	IPTG induced expression under the control of T7 promoter with C-terminal-His tag. Used for protein over expression.	Ampicillin	Novagen
pK18mobsacB	An allelic exchange vector mobilized by *E. coli* S-17.	Kanamycin resistant and sucrose sensitive	Prof. N.Hunter University of Sheffield
pBAD1787	pBAD/HisB containing the *rpa1787* gene cloned into *XhoI/EcoRI* restriction sites	Ampicillin	R. Salmon University of Sheffield
pET1707	pET21a (+) containing 1707 gene cloned into *NdeI/XhoI*	Ampicillin	R. Salmon University of Sheffield
pET1766	pET21a (+) containing 1766 gene cloned into	Ampicillin	This study
pET3723	pET21a (+) containing 3723 gene cloned into	Ampicillin	This study
pET3724	pET21a (+) containing 3724 gene cloned into	Ampicillin	This study
pET3725	pET21a (+) containing 3725 gene cloned into	Ampicillin	This study
pmobsac1787	pmobsac1787 containing the up and down stream region of *rpa1787* gene cloned into *EcoRI/HindIII* restriction site	Kanamycin resistant and sucrose sensitive	This study

Table 2.3 Plasmids used in this study

2.5.3 DNA manipulation techniques

Plasmid DNA manipulation was done following (Sambrook *et al*., 1989) protocols.

2.5.3.1 Polymerase Chain Reaction (PCR) amplification

Primers were obtained from Sigma-Aldrich and stock primers were kept at 100 μM. Standard PCR reactions were carried out in 12.5 μl reaction volume using MyTaq™ master Mix to screen *R. palustris* strains and for checking the accuracy of cloning procedure in *E. coli.*

For gene cloning Accuzyme DNA polymerase (Bioline) was used and the final volume for the reaction was 20 μl. Reactions were made as described in manufacturer's guidelines (0.1 μg of template DNA was used in addition to 10 μM primer stock in addition to Accuzyme, and dNTPs). PolyMate additive (Bioline) was also used in the reaction to improve DNA yield because the *R. palustris* genome has a high G-C content. PCR was preformed using Techne Techgene Thermal Cycler (Techne) and the reaction was initiated after initial 5 minutes denaturation step at 95°C for Accuzyme PCR reaction (0.5 μl). The reaction went through the following cycle for 25 cycles.

Denaturing	96 °C	1 minute
Annealing	62 °C	2minutes
Elongation	72 °C	2 minutes

Followed by a final extension of 72 ^{o}C for 5 minutes. If the amplification was not successful different annealing temperature was tested until a suitable product was gained. QIAquick® PCR Purification Kit (Qiagen) was used to purify and concentrate

PCR product. PCR products were visualized by agarose gel electrophoresis pre stained with ethidium bromide.

To screen for the presence or the absence of a gene in *E. coli* or *R. palustris*, *Taq* DNA polymerase was used and reaction required only 1 minute denaturation step at at 96^{o}C followed by 30 cycles of

Denaturing	96 °C	15 secs
Annealing	62 °C	1 min
Elongation	72 °C	30 sec's

and finally a Final extension at 72 °C for 5 minutes. Agarose gel electrophoresis was also used to visualize PCR product. Primers used in this study are listed in (Table 2.4).

2.5.3.2 Agarose Gel Electrophoresis

DNA was analysed using 0.7 % agarose gel electrophoresis. Agarose (0.7 % [w/v]) was added to 1 x TAE buffer (40 mM Tris-acetate, 1 mM ethylenediaminetetraacetic acid (EDTA) pH 8.0) and the mixture was dissolved by heating. 200 ng/ml ethidium bromide was added to cooled agarose (~50 °C) and then poured into a gel cast with gel comb in place. 5 x DNA loading buffer (Tri-Colour, Bioline) was used to load DNA sample and Hyperladder I (Bioline) was also loaded as DNA ladder. The gel was electrophoresed in 1 x TAE buffer at 110 v for 35 minutes and then the gel was visualized using ultraviolet (UV) light gel documentation system (Syngene).

Gene	Primer sequence	Used for
RPA1766	F: 5'-TGATCACATATGTCCGAAAACCAGCTTTATCACG-'3	**Cloning & over expression**
	R: 5'- ATATATCTCGAGAATCACTCGGCCGGCATT-'3	
RPA3723	F: 5'-TGATCACATATGAGCCCTGCTGCCGGCCCC-'3	**Cloning & over expression**
	R: 5'-ATATATCTCGAGGTTGCGCAGCT-'3	
RPA3724	F: 5'-TGATCACATATGAAGAAATACGACACCGGC-3'	**Cloning & over expression**
	R: 5'-ATATATCTCGAGGTTCGACGGCG-'3	
RPA3725	F: 5'-TGATCACATATGAAGAAGTACGATCCGGGC-'3	**Cloning & over expression**
	R:5'-ATATATCTCGAGGCCGCCGACGTCGGCGCTGAT-'3	
Deletion of *rpa1787(Fcs2)*	Upstream F 5'-CCGGAATTCCCGACATGATGCTGACAGGG-'3 Upstream R 5'-GCTCTAGAGTCCATCACGGCTCCGTTCAGC-'3	**Mutation**
	Downstream F:5'-GCTCTAGAGAGTGAAGGCATCAAGATGGACC-'3 Downstream R: 5'- CCCAAGCTTCGAACATCGCGAAGTAACGC-'3	
Deletion of β-oxidation cluster *(rpa1704-rpa1712)*	Up stream region F: 5'-AAAATCGGCGGTGGGCCGGAC-3' Up stream region R: 5'-TCTTGTTTTGGTTTTCAGAGC-3'	**Mutation**
	Down stream region F: 5'-CATCAACCACGGGCCGTCATT-3' Down stream region R: 5'-GCGATCATCACCACCGATGAA -3'	
Deletion of *rpa1706*	Up stream region F: 5'-ATTGAATTC GAAGGGTGGCATCGTCGGCA-'3 Up stream region R: 5'-GCCGCTCTAGACTCGGTCACGCCATCACT-'3	**Mutation**
	Down stream region F: 5'- GCTCTAGAAACCGCTGACCAAATCAA-'3 Down stream region R: 5'-ATAAAGCTTCAGATACTTCAGCTTGAG -'3	
Deletion of *rpa1766*	Up stream region F: 5'- ATTGAATTC GTGCGCGCCGCCAAGCGAA-'3 Up stream region R:5'- GCCGCTCTAGA CGGACATTGTTACTCTCC-'3	**Mutation**
	Down stream region F: 5'- AGTCATTCTAGAATTTAGTTCGCCGC-'3 Down stream region R: 5'-CATATAAAGCTTCAGCGTTGGCGGAA-'3	

Table 2.4 Full list of Primers used in this study.

2.5.3.4 Restriction digestion of DNA

DNA digestion with restriction enzymes was preformed using (NEB) buffer and restriction enzymes as recommended by the manufacturer's guidelines. Digestion was preformed at 37 °C for 2 hours or overnight the samples were heat inactivated at 65 ^{o}C for 20 minutes. Samples were purified using QIAquick® PCR Purification Kit (Qiagen).

2.5.3.5 Phosphatase treatment of DNA

Vector self-ligation was prevented by using Antarctic Phosphatase (NEB) which removed the phosphate group at the 5´ end of plasmid DNA. Phosphatase treatment was inactivated by incubating the sample for 5 min at 65 °C.

2.5.3.6 Ligation of cohesive ended DNA fragments

Reactions were carried out in the ratio 1:1, 3:1, 5:1, and 7:1 insert to vector ratios. DNA concentration was estimated from agarose gel band intensity which was compared to the molecular weight standard of known concentration (Hyperladder I (Bioline)). 10 ng of vector DNA was used in the reaction and insert concentration was calculated using the following equation:

$$\frac{\text{ng of vector} \times \text{kb size of insert} \times \text{molar ratio of insert:vector}}{\text{kb size of vector}} = \text{ng of insert}$$

Ligation reactions contained 10 ng of vector and the calculated insert in addition to 1 µl of 10X T4 DNA ligase buffer (NEB) and 0.5 µl of T4 DNA ligase (NEB). dH_2O

was used to made the final volume up to 10 µl. A control was made with no insert to analyse the level of self-ligated and un-cleaved plasmid. Reaction mixture was kept at room temperature for 1 hour before transformation into chemically competent *E. coli* cells.

2.6 Preparation of competent cells and transformation

2.6.1 Preparation of chemically competent *E. coli*

The method of Hanhan (1983) was used to create competent *E. coli* cells and to prepare RF buffer solutions. *E. coli* was grown on 50 ml LB broth to mid-exponential phase (OD_{600} 0.6) which represent a density of ~ 4-7 x 10^7 viable cell ml^{-1}. Bacterial cell were kept on ice for 15 minutes and then centrifuged at 6000 x g for 20 minutes at 4° C. The pellets were re-suspended in 50 ml ice cold RF1 solution pH 5.8 (100 mM KCl, 50 mM $MnCl_2$ $4H_2O$, 30 mM CH_3COOK, 10 mM $CaCl_2$-$2H_2O$ and 15 % (w/v) glycerol). After that the cells were incubated on ice for 15 minutes, centrifuged as mentioned above and re-suspended in 8 ml RF2 solution pH 6.8 (10 mM MOPS, 10 mM KCl, 75 mM $CaCl_2$-$2H_2O$, and 15 % (w/v) glycerol). Finally cells were incubated on ice for 20 minutes, aliquoted as 150 µl in eppendorf tubes and stored at -80 ° C.

2.6.2 Transformation of chemically competent *E. coli* cells.

The competent cell aliquot was thawed on ice and 2 µl of purified plasmid or 5 µl of ligation mixture was added to competent cells. The mixture was incubated on ice for 30 minutes prior to heat shocking at 42 °C for 50 seconds. After that, it was incubated on ice for two minutes then 850 µl LB broth was added and incubated at 37 °C for 1 hour. Cells were harvested by centrifugation (16000 x g, 5 minutes) and the pellets were re-suspended with fresh 100 µl LB. The re-suspended cell were plated onto LB

solid media containing the antibiotic present in the plasmid. The plates were incubated at 37 °C for 18 hours. As a control a no insert DNA control was used to estimate the level of vector only transformation. The same thing was done for pK18mobscaB constructs except that the final incubation was kept for longer periods as the presence of *sacB* gene reduces the growth rate.

2.7 *R. palustris* gene deletion by pK18mobsacB mutagenesis

The targeted gene was deleted by cloning homologous upstream and downstream regions into the restriction site of pK18mobsacB vector and then transformation into *E. coli* JM109 cells, $>10^8$ cfu/ug (Promega). The constructs were purified and then transformed into *E. coli* S17-1 conjugating strain. A detailed description of the way in which this method works is given in Chapter 4.

2.7.1 Conjugation of *R. palustris* with *E. coli* S17-1 cells

Freshly grown *R. palustris* cells were inoculated into 8 ml PYE starter cultures and incubated photoheterotrophically at 30 °C for 2-3 days. After that the starter culture was used to inoculate another 100 ml PYE culture that was incubated for a another 2 days photoheterotrophically at 30 °C. 40 ml of *R. palustris* cell was centrifuged (6000 x g for 15 minutes at 30 ^{o}C) and cell pellets were re-suspended in 100 µl PYE resulting in a thick cell suspension that would be used in the conjugation experiment. A loop of the transformed S17 cells containing the pK18mobsacB vector was re-suspended in 20 µl of PYE and then added to the 100 µl *R. palustris* cell suspension. After that, the suspension was mixed by pipetting and plated as 50 µl drops on a dried PYE plate and incubated at 30 ^{o}C for 16 hours. After incubation, the entire drop was inoculated as a streak onto an RCV kanamycin plate (100 µg ml^{-}

^{1}) and the plate was incubated at 30 ^{o}C for 1 week (or until single colonies appeared).

2.7.2 Kanamycin and sucrose selection

Each single *R. palustris* colony grown on kanamycin plates was picked and streaked sequentially onto RCV kanamycin (100 μg ml^{-1}) plates and then a 10% RCV sucrose plate. Both plates were incubated at 30 ^{o}C for one week; this is done to ensure the success of the first recombination where cells must grow on RCV kanamycin plate and not the sucrose RCV plates. Cells with correct phenotype were grown photoheterotrophically at 30 ^{o}C in 8 ml RCV minimal media with 100 μg ml^{-1} kanamycin for 5 days. After that, *R. palustris* cultures were serially diluted and 100μl of the 10^{3} and 10^{5} was spread onto 10% RCV sucrose plates and incubated at 30 ^{o}C, for at least two weeks, which would allow the second recombination. The success of the second recombination was tested by plating the colony into a 10 % RCV sucrose plate first and then into an RCV containing 100 μg ml^{-1} kanamycin. The plates were incubated at 30 ^{o}C and cells that grew on sucrose but were sensitive to kanamycin are the ones that were further screened by colony PCR.

Primers used in colony PCR were designed to amplify ~100bps upstream and downstream from the selected gene that was to be deleted. PCR reactions were run on 0.7 % agarose gel and the results were compared to wild type *R. palustris* colony PCR using the same primers. Colonies that showed a band missing the correct number of base pairs after PCR were marked as a positive deletion mutant. Positive deletion mutants were subcultured onto PYE agar plates and incubated at 30 °C for 1 week. Single pure colonies were grown in Liquid PYE under the same condition and then

the culture was centrifuged at 8000 x g for 10 minutes and cell were re-suspended in 25 % glycerol PYE and stored at -80°C.

2.8 Characterization of Recombinant Proteins

2.8.1 Determination of Protein Concentration

The Bradford Assay method (Bradford,1976) was used to determine soluble protein concentration. The Bio-Rad reagent (Biorad Inc, USA) was used: 20 µl of protein solution was added to 1 ml cuvette containing 800 µl of milli-QH_2O and 200 µl of Bio-Rad Dye Reagent Concentrate. The cuvette was mixed and the concentration was measured at OD 595 nm using the following formula:

$$\frac{OD_{595} \text{ X } 15}{\text{volume of protein (µl)}} = \text{mg protein}$$

For accurate protein concentration measurement, a UV spectrophotometer (UV-24001 UV-Vis recording spectrophotometer) (Shimadzu) was used where 10 µl of purified recombinant protein was added to 990 µl dH_2O using a quartz cuvette. UV spectrum was taken from 270-330 nm and the concentration was calculated using the following formula (taking into account protein dilution 1:100)

$$\frac{\text{Absorbance at 280nm} - \text{Absorbance at 320nm}}{\text{Extinction Co-efficient of protein}} \text{ X } 100 = \text{M protein}$$

2.8.2 SDS-polyacrylamide gel electrophoresis

One dimensional SDS polyacrylamide gel electrophoresis (SDS-PAGE) was done using the Mini-Protean 3 system (Bio-Rad). The system components were cleaned by ethanol and dH_2O before use. A 12 % resolving gel was prepared by adding together 2.5 mL of 30 % (w/v) acrylamide /0.8 % (w/v) bisacrylamide diluted by 2.35 mL of 1 M Tris-HCl pH 8.8, 62.5 µl of 10 % (w/v) sodium dodecyl sulphate (SDS), 62.5 µl of 10 % (w/v) ammonium persulphate (APS), 6.25 µl of 0.01 % N,N,N',N'-Tetramethylethylenediamine (TEMED) and 1.28 ml dH_2O. Gel components were mixed gently by pipetting and then pipetted into gel casts. After gel setting, a 6 % stacking gel was prepared as follows; 0.45 mL of 1 M Tris-HCl pH 6.8, 37.5 µl of 10 % (w/v) SDS, 0.75 mL of 30 % acrylamide mixture, 37.5 µl of 10 % (w/v) ammonium persulphate (APS), 3.75 µl of 0.01 % N,N,N',N'-Tetramethylethylenediamine (TEMED) and 2.47ml dH_2O. Stacking gel components were mixed and poured into the gel casting try with a comb insert that was removed when the gel was set.

Protein samples were prepared by adding 20 µl of purified protein to 20 µl buffer (60 mM Tris-HCl pH 6.8, 2 % (w/v) SDS, 0.005 % [w/v] bromophenol blue, 5 % [v/v] ß-mercaptoethanol, 10 % (w/v) glycerol). Then samples were boiled for 5 minutes and centrifuged at 13000 x g in a microfuge for 10 minutes. After that the samples were loaded onto the gel with a PageRuler® (Plus) Prestained Protein Ladder (Fermentas) and electrophoresed at 180 V for 50 minutes or until the tracking dye reach the bottom of the gel. Coomassie brilliant blue dye (50 % [v/v] methanol, 10 % [v/v] glacial acetic acid, 0.5 % (w/v) Coomassie Brilliant blue R [Sigma-Aldrich]) was

used to stain the gel and then de-stained by 40 % (v/v) methanol and 10 % (v/v) glacial acetic acid until individual proteins were visible.

2.8.3 Over-production of Recombinant Protein

Two vectors were used in this study; the pET21 vector and pBAD vector to over-express recombinant proteins in *E. coli.*

2.8.3.1 IPTG induction using pET system

The pET21a (+) vector (Novagen) was used for recombinant protein overexpression under the control of T7 promoter that is inducible by IPTG. The recombinant protein has a 6-histidine tag incorporated at the C-terminus. The gene under study was first amplified by PCR using specific primers that included appropriate restriction sites for cloning but excluded the stop codon to allow for addition of the C-terminal 6-histidine tag in the recombinant protein, when the gene was inserted into the pET21a (+) multiple cloning site. Constructs were sequenced to confirm that the gene was in-frame and were transformed into the BL21 (DE3) overexpression strain of *E. coli.* The cells were grown under various temperature and time conditions after induction at $OD_{(600)}$ of 0.6 using 400 μg IPTG. 1 ml samples were used to visualize protein over expression using SDS-gel electrophoresis. After determining the best growth conditions, cells were grown in 2L batch cultures and harvested by centrifugation at 10000 x g for 10 minutes at 4 ^{o}C and the pellet was frozen at – 20 ^{o}C until needed.

2.8.3.2 Arabinose induction using the pBAD/HisB system

The pBAD/HisB system was used when gene overexpression did not produce appropriate amount of protein by the pET system. The recombinant protein has a 6-histidine tag at the N-terminal end and the overexpression of protein was controlled by the *araBAD* promoter, which is induced by different concentrations of arabinose.

The gene under study was amplified by PCR with primers that removed the start codon at the N- terminal end to add the 6-histidine tag. The stop codon remained in the gene. The construct was sequenced and transformed into *E. coli* (TOP10) strain. The *araBAD* promoter is highly regulated by arabinose concentration. So varying arabinose levels at induction can modify the level of recombinant protein overexpression and solubility (Guzman et al. 1995). TOP10 cells were grown under various temperature time and arabinose concentrations after induction at $OD_{(600)}$ of 0.6. 1 ml samples were used to visualize protein over expression using SDS-gel electrophoresis. After determining the best growth condition, cells were grown in 2L batch cultures harvested by centrifugation at 10000 x g for 10 minutes at 4 ^{o}C and the pellet was frozen at – 20 ^{o}C until needed.

2.8.4 Preparation of Cell free extracts

A cell free extract from *E. coli* cell was created before purifying the recombinant protein. Cells were de-frosted and re-suspended in 20 ml Binding Buffer (section 2.6.5.1). Cell re-suspension was left on ice for 15 minutes and then the cells were broken by sonication at frequency of 16 micron amplitude (Soniprep 150 ultrasonic disintegrator, SANYO) for 20 seconds and that was repeated 4 times. After sonication the cell was centrifuged at 18000 x g for 30 minutes at 4 ^{o}C. The supernatant is the cell free extract that was used for protein the purification procedure.

2.8.5 Protein purification procedures

An Akta Prime plus purification system (GE Healthcare) was used to purify proteins under study.

2.8.5.1 Nickel Affinity Chromatography

A 5 ml HisTrap^{TM} HP column (GE Healthcare, UK) was used in Akta Prime plus purification system, following manufacturers recommendation. Cell free extract was injected onto the column using binding buffer (20 mM sodium phosphate buffer, 30 mM imidazole, 500 mM NaCl, pH 7.4) and was eluted using elusion buffer (20 mM sodium phosphate buffer, 500 mM imidazole, 500 mM NaCl, pH 7.4) over 10 column volumes with a linear gradient of 0-500 mM imidazole. Different eluted protein fractions were collected in a volume of 1 ml. After protein elution the column was washed extensively with elution buffer and the column was stored in 20 % ethanol.

2.8.5.2 Protein de-salting and Concentration

To de-salt and concentrate the collected purified protein fractions, a Vivaspin column (Sartorius, UK) with an appropriate molecular weight cut off was used. Protein fractions were centrifuged at 8000 x g for 20 minutes or till the volume reduced significantly. The exchange buffer was added 3X the volume of protein fractions to wash and desalt the protein sample.

2.8.6 Unfolding and refolding of periplasmic binding proteins

Urea treatment was used to remove any ligands bounded to the purified soluble binding proteins. Purified protein was unfolded by 1 L of 6 M urea dialysis for 16 hours at 4 °C and refolded by dialysis in 2 L refolding buffer (50 mM Tris-HCl, 100 mM NaCl, pH 7.2) for a total first 8 hours dialysis which included changing the buffer every 2 hours followed by a final overnight dialysis. After that, protein was centrifuged at 13,000 × g for 10 minutes to remove unfolded and precipitated protein.

2.9 Biochemical and Biophysical Assays

2.9.1 Linked enzymatic assays of Fcs1 and Fcs2

Enzymatic assays were carried out using 96 well plates in an Omega florescence spectrophotometer (BMG lab tech) by monitoring NADH fluorescence at 340 nm excitation and 460 nm emission. Each reaction was composed of 50 mM Tris-HCl pH 8, 3 mM $MgCl_2$, 2 mM ATP, 0.4 mM CoA, 0.4 mM NADH, 2 mM phosphoenol pyruvate, 0.2 unit of myokinase (Sigma), 0.2 unit lactate dehydrogenase/pyruvate kinase (Sigma), aromatic substrates with concentrations ranging from 0.001- 0.6 µM were used. dH_2O was used to make up the final volume of 100 µl. Each reaction was set up with all the components except CoA which was injected to start the reaction. A minimum of 7 substrate concentrations was recorded and the oxidation of 2 mol of NAPH represents 1 mol of ligand. Each data set was processed and fit to Michaelis-Menton kinetics using Graph Pad Prism software.

2.9.2 Tryptophan Fluorescence spectroscopy

Experiments were done using a Cary Eclipse fluorimeter (Varian Ltd, UK) to screen fluorescence emission changes in intrinsic tryptophan residues once bound to a ligand. A 3 ml quartz cuvette with magnetic stirrer was used. Experiments were preformed at 30 °C using 10 mM Tris-HCl buffer (pH 7.4). protein concentration was 0.2 µM with various ligand concentrations.

2.9.2.1 Ligand screening

Ligand binding to proteins was screened by monitoring the quench in fluorescence emission for the protein alone compared to the protein and ligand. Samples were excited at 280 nm (5 nm slit width) and the emission recorded between 300-450 nm

(20 nm slit width). A ligand-binding event was confirmed if a quench or an enhancement of fluorescence was recorded.

2.9.2.2 Fluorescence titrations

Ligand titrations were carried out with 0.2 μM protein concentration at λ_{ex} 280 nm λ_{em} 332 nm using 5 nm excitation and 20 nm emission slit widths respectively. Ligand was titrated on the basis of changes in fluorescence emission monitored. Experiment was done done in triplicates to calculate a K_d values using Smart et al. (2009) quadratic equation.

2.9.3 Thermofluor Screens

SPYRO orange dye (invitrogen) was used to monitor protein binding and any changes in fluorescence emission was recorded by Mxpro software (Stratagene) associated with an Mx3005P RT-PCR machine (Stratagene). Each reaction contained 5 μl 10mM Tris-HCl pH 7.4, protein under study (5-25 μM), ligand of interest (66 μM) and 5 μl 10x SPYRO dye. First, a range of 5-25 μM protein concentrations were tested alone to check for the optimal dye signal and protein stability. Then once the concentration was set ligands were added in 66 μM concentration and protein stability was monitored by using a temperature ramping program on Mxpro software based SPRYO dye florescence emission. The shifts in temperature were calculated according to the protocols described in Vedadi et al. (2006). All reactions were done in triplicates.

2.9.4 Nuclear Magnetic Resonance (NMR)

Binding of RPA3724 to different ligands was detected by 1H-NMR, where the flattening and broadening of the spectral peaks of the ligands by the protein was observed. A reaction mixture was prepared using 50 mM phosphate buffer pH 7.4

with protein alone (50 μM), ligand alone (50 μM) or a protein ligand mixture in the ratio of 1:1. 5 mm NMR tubes were used and they contained 450 μL reaction mixture, 50 μL D_2O and 1 μL trimethyl-silyl propionate (TSP) as a 0 ppm reference. All 1D-NMR data were collected at 25 °C on a Bruker Advance NMR spectrometer at 800 MHz. Processing and integration was performed using Bruker Topspin v1.3 software after manual baseline correction was done. NMR spectra were acquired with the help of Mrs. Andrea Hounslow in the departmental NMR facility.

3. Proteomic and gene expression study of 5-phenylvaleric acid degradation *in vivo*

Proteomic and gene expression study of 5-phenylvaleric acid degradation *in vivo*

3.1 Introduction

R. palustris has the ability to grow photoheterophically on a wide range of aromatic lignin monomers such as, coumarate, caffeate and cinnamate which makes it a good candidate for bioremediation processes (Harwood & Gibson 1988). The molecular biochemical pathways associated with complex aromatic compound degradation are still unclear in *R. palustris* (Larimer et al. 2004). Structurally diverse single aromatic ring compounds are suggested to be degraded to a central intermediate, benzoyl-CoA, through different peripheral pathways. Then benzoyl-CoA is degraded to acetyl-CoA through a main pathway in the steps of ring reduction, ring cleavage, and β-oxidation (Carmona et al. 2009). The Acetyl-CoA molecules produced can then be used by normal metabolism pathways in the cell (Harrison & Harwood 2005). A β-oxidation pathway was suggested to be the route of degradation for different chain length phenylalkane carboxylic acids because this pathway involves the removal of two carbons of the chain each time (Elder et al. 1992). However, no genes for such enzymes have been identified to date. Identification of such genes could help in understanding how complex aromatic compounds are degraded by bacteria and hence improve biotechnological processes that deals with such compounds to produce energy or to remove them from the environment. There are several approaches that could be used to find candidate genes/enzymes for this pathway. Knowledge obtained from proteomics and transcriptomics for example, could be important not only to understand the first catabolic steps of the degradation pathway, but also to understand the molecular mechanism responsible for its in vivo regulation.

This chapter attempts to identify the different proteins and their genes, associated with anaerobic 5-phenylvaleric acid catabolism. Using a proteomic approach, protein profiles for *R. palustris* grown on 5-phenylvaleric acid as the sole carbon source were compared with the profile of *R. palustris* grown on benzoic acid. The comparison of 5-phenylvaleric acid grown cells with benzoic acid grown cells in mass spectrometry resulted in the identification of 1760 proteins of which 88 proteins were significantly up-regulated on 5-phenylvaleric acid. Many of these proteins were involved in aromatic and fatty acid degradation pathways. In *q*RT-PCR where gene expression of *R. palustris* grown on 5-phenylvaleric acid was compared to gene expression of *R. palustris* cellss grown on benzoic acid and succinic acid, the data showed that candidate genes that were not significantly up-regulated when compared with benzoic acid are significantly shifted when compared with cells grown with succinate.

3.2 Results

3.2.1 Proteomic analysis of *R. palustris* grown in the presence of 5-phenylvaleric acid by label free quantitative mass spectrometry

Application of proteomics to assess the catabolic pathway of degradation of various substrates is widely used for many bacterial species such as *Pseudomonas* (Kim et al. 2006). Mass spectrometry is a high resolution technique that can identify and quantify many proteins present within a complex mixture such as a cell-free extract (Steen & Mann 2004). In label-free quantitative mass spectrometry, the protein sample is first digested with a protease with a known cleavage site (usually trypsin), then the peptide fragments are separated by liquid chromatography coupled to a mass spectrometer for online LC-MS/MS analysis. In a tandem mass spectrometer such as a hybrid ion trap-Orbitrap used here, data-dependent analysis is performed in which the m/z of intact peptides is measured as they elute from the chromatography column in a survey or MS1 scan, then the top 20 ions by abundance (peptides) are sequentially selected for fragmentation (MS2) by collision induced dissociation and this cycle is repeated for the duration of the chromatography run. Using data from MS1 and MS2 scans, it is possible to identify peptides (and infer proteins) by matching experimental spectra to theoretical spectra predicted for all tryptic peptides in the proteome under investigation. The mass spectrometry software can also perform intensity calculations and statistical analysis, allowing quantification (Zhu et al. 2010).

Two conditions of growth were compared; cells grown on 5-phenylvaleric acid and cells grown on benzoic acid. The rationale is that this comparison should reveal specific increases in abundance of those enzymes involved in beta-oxidation of the side-chain, but not general aromatic metabolism. Wild type *R. palustris* cells were

grown anaerobically under light with 5-phenylvaleric acid or benzoic acid as the sole organic carbon source to OD_{660} 0.6, which represents mid-exponential phase (Figure 3.3).

Cell were harvested at the same OD but at different time points (third day of growth for cells grown on benzoic acid and fourth day for cells grown on 5-phenylvaleric acid) because the growth rate was different between cells grown on benzoic acid and cells grown on 5-phenylvaleric acid. Three biological replicates (separate growths) of each condition were prepared and analyzed to find proteins involved in the degradation of 5-phenylvaleric acid.

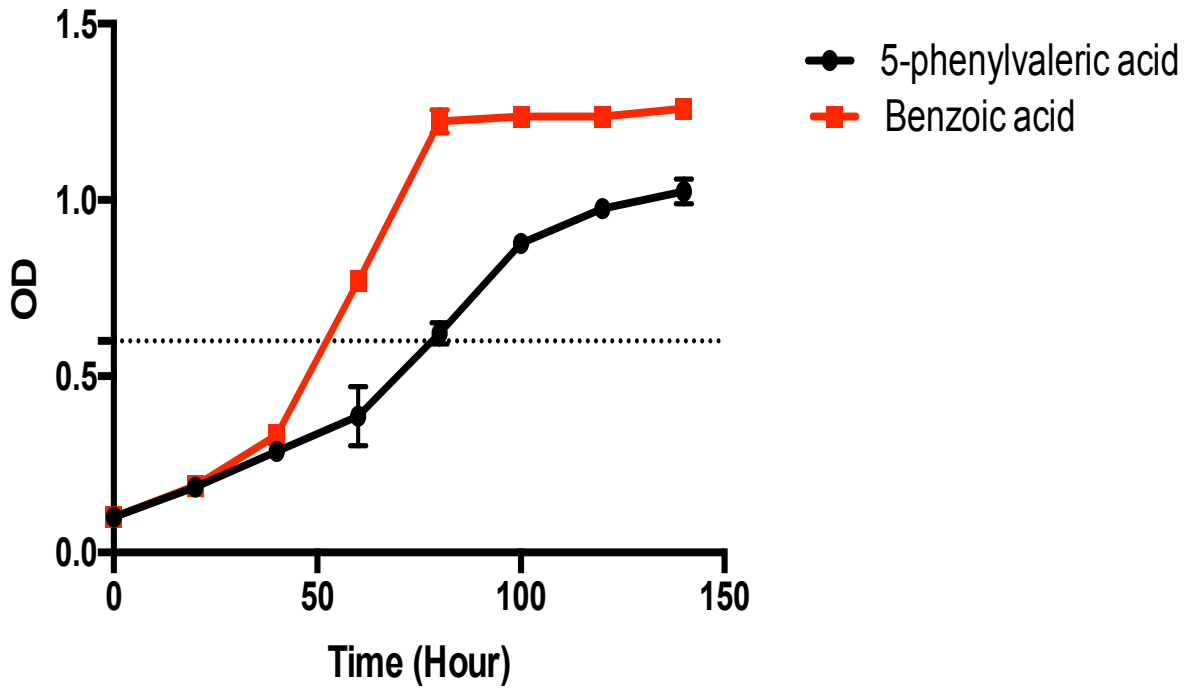


Figure 3.3. Wild type *R. palustris* growth curve for cells grown anaerobically on RCV media with 3mM benzoic acid as a carbon source in comparison to cells grown anaerobically in RCV media with 3mM 5-phenylvaleric acid. Samples for mass spectrometry and qRT-PCR were taken at OD 0.6.

Proteins from each condition were digested with trypsin and dried before being analysed by HPLC coupled to mass spectrometry.

MaxQuant was used to compare the data obtained by mass spectrometry using the default parameters (Cox et al. 2009) and the raw data for the experiment is presented in the accompanying CD (Table1). In our analysis, a minimum of two valid intensity values for a given peptide (S0=2) were required in the three biological replicates of 5-phenylvaleric acid grown cell extract and if there were any missing intensity values (i.e. the peptide was not detected in that particular sample), they were imputed using a downshifted normal distribution (width 0.3, downshift 1.8). The correlation of biological replicates was assessed and it can be seen in Figure 3.4 that the correlation coefficients between biological replicates are high.

Next t-testing was performed with correction for multiple hypothesis testing using a permutation based false discovery rate (FDR) of 0.05 and an S0 = 2. Around 88 different proteins were significantly up regulated in cells grown with 5- phenylvaleric acid as the sole carbon source (Figure 3.5). Those that were significantly different between the two sets are listed in (Table 3.1).

The results showed that many different types of proteins were up-regulated that belong to different metabolic and regulatory systems specially proteins that are involved in, for example, energy production and photosynthesis. But since many of these are not directly involved in aromatic metabolism, they are outside the scope of our research and the significance of their up-regulation has not been further investigated. Therefore, only selected proteins in Table 3.1 will be further discussed.

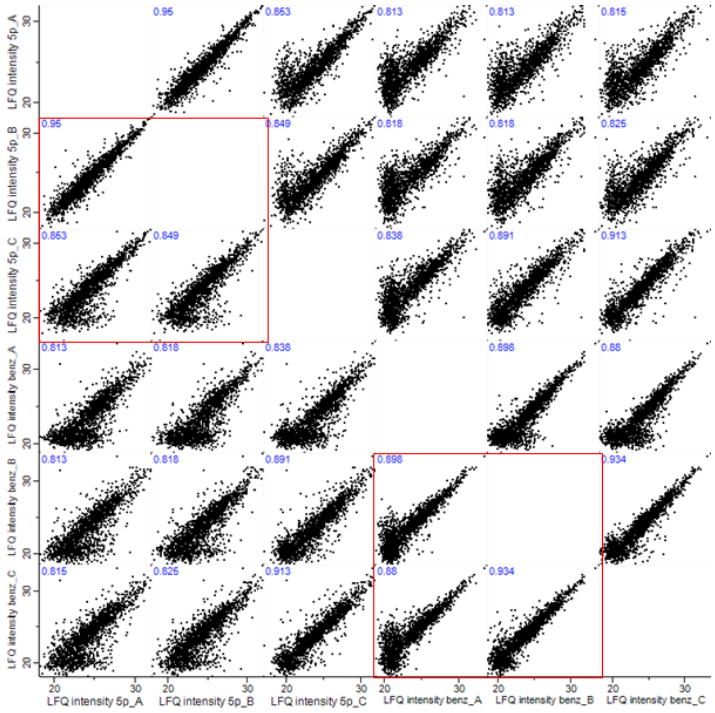


Figure 3.4. Multiscatter plot of protein LFQ intensities calculated from data obtained from proteomic analysis of *R. palustris* grown with benzoic acid versus 5-phenylvaleric acid. Pearson correlation coefficients are shown for each comparison and were >0.80 for biological replicates in each group.

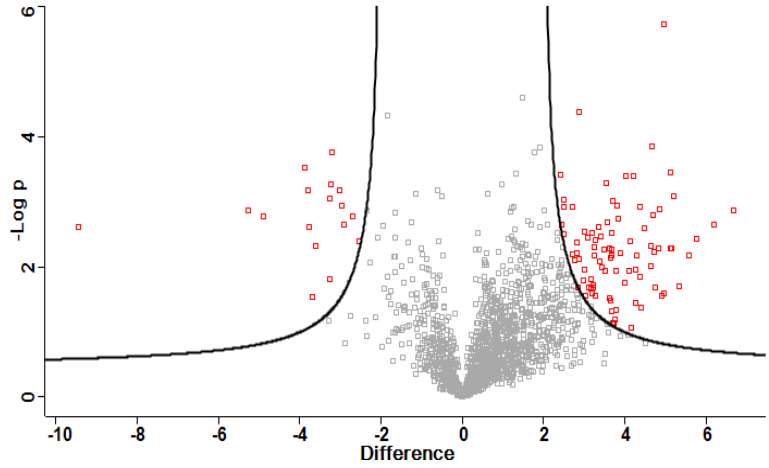


Figure 3.5. Volcano plot showing quantitative changes in protein expression for *R. palustris* grown with benzoic acid versus 5-phenylvaleric acid. Proteins with significantly altered expression (red squares) were identified from statistical analysis of triplicate cultures (grown with benzoic acid or 5-phenylvaleric acid) using t-testing and correction for multiple hypothesis testing using a Permutation based False discovery rate of 0.05.

Protein Name	Gene name	Average ratio 5-pp/benz
Histidine kinase	RPA3015	62.68
Uncharacterized protein	RPA4094	55.47
Bacteriophytochrome	RPA3016	46.27
Acetyl-coenzyme A carboxylase carboxyl transferase	RPA0508	39.76
Putative 3-hydroxybutyryl-CoA dehydratase	RPA1786	31.32
Uncharacterized protein	RPA1890	30.50
Putative CysN/CysC bifunctional enzyme,	RPA0753	27.73
Uncharacterized protein	RPA4092	27.60
Ribonuclease E	RPA2450	26.53
Uncharacterized protein	RPA0819	25.75
Amino acid ABC transporter	RPA2563	25.38
Protein HflC	RPA3490	24.54
DUF404	RPA2177	23.86
Transglutaminase-like domain	RPA3754	22.80
Uncharacterized protein	RPA3081	22.69
Amidohydrolase 2	RPA4198	22.38
Putative feruloyl-CoA synthetase	RPA1787(Fcs2)	22.05
NADH-ubiquinone dehydrogenase chain G	RPA3490	20.36
TPR repeat	RPA3211	19.82
1,4-alpha-glucan branching enzyme GlgB	RPA3644	19.77
Multidomain chemotaxis histidine kinase	RPA1676	19.67
Uncharacterized protein	RPA3011	18.63
Acetyl-coenzyme A carboxylase carboxyl transferase	RPA0071	16.73
Acetyltransferase of pyruvate dehydrogenase complex	RPA2864	15.89
Putative 3-oxoacyl-(Acyl carrier ptn) reductase	RPA0109	15.87
Protein-export protein SecB	RPA0302	15.70
GlnK, nitrogen regulatory protein P-II	RPA0272	15.46
Pyruvate dehydrogenase E1 component subunit alpha	RPA2867	14.58
Nitrogen assimilation regulatory protein	RPA2593	14.37
Indolepyruvate oxidoreductase subunit IorA	RPA1224	14.26
Probable acyl-CoA dehydrogenase	RPA3115	14.10
Possible chemotaxis cheB/cheR fusion protein	RPA3316	12.93
Cdp-glucose 4,6-dehydratase	RPA4049	12.78
Light-harvesting protein B-800-850 beta chain D	RPA3013	12.15
Uncharacterized protein	RPA1918	11.81
Transcriptional regulator PpsR2 Fis family	RPA1536	11.80
ADP-L-glycero-D-manno-heptose-6-epimerase	RPA3985	11.55

Protein Name	Gene name	Average ratio 5-pp/benz
Uncharacterized protein	RPA2656	11.45
Beta-lactamase-like	RPA4341	11.13
Putative long-chain-fatty-acid-CoA ligase	RPA1766	11.11
Chorismate mutase/prephenate dehydratase	RPA3695	10.89
Ubiquinone biosynthesis O-methyltransferase	RPA0603	10.87
Uncharacterized protein	RPA4188	10.80
NADH-ubiquinone dehydrogenase chain E	RPA2947	10.80
Ubiquinone/menaquinone biosynthesis C-methyltransferase	RPA0083	10.79
Alpha/beta hydrolase fold	RPA1212	10.12
DUF35	RPA3193	9.96
Putative UROPORPHYRINOGEN-III synthase	RPA0256	9.90
Putative 2-haloacid halidohydrolase Iva	RPA4199	9.88
Putative enoyl-CoA hydratase	RPA1706	9.45
Mg-protoporphyrin IX methyl transferase	RPA1546	9.34
Uncharacterized protein	RPA3423	9.25
Putative DEAD-box protein, ATP-independent RNA helicase	RPA4001	9.24
tRNA N6-adenosine threonylcarbamoyltransferase	RPA0255	9.10
DNA helicase	RPA3506	8.91
Related to Pyruvate ferredoxin/flavodoxin oxidoreductase	RPA3195	8.83
Possible methylmalonyl-CoA decarboxylase alpha chain	RPA3455	8.48
Putative RND efflux transporter	RPA3775	8.32
GMC-type oxidoreductase	RPA3730	8.29
Uncharacterized protein	RPA2216	8.20
TPR repeat	RPA3916	8.19
Possible coenzyme F420 hydrogenase beta subunit	RPA1501	8.10
Putative branched-chain amino acid transport system ATP-binding protein	RPA3720	7.42
Possible benzaldehyde lyase	RPA0108	7.34
Phosphoribosylaminoimidazole-succinocarboxamide synthase	RPA3821	7.07
Putative signal transduction histidine kinase with PAS/PAC domains	RPA0007	7.06
NAD(P) transhydrogenase subunit beta	RPA4180	7.02
Uncharacterized protein	RPA3733	7.00
DUF403	RPA2176	6.94
Uncharacterized protein	RPA0258	6.36

Protein Name	Gene name	Average ratio 5-pp/benz
Alpha-1,4-glucan:maltose-1-phosphate maltosyltransferase	RPA3642	6.04
Putative alpha-D-galactoside galactohydrolase	RPA0378	5.97
Putative Zn-binding dehydrogenase	RPA0107	5.96
Transcriptional regulator, TetR family	RPA4097	5.57
30S ribosomal protein S2	RPA2922	5.47
Putative signal transduction histidine kinase	RPA4513	5.44
Transcriptional regulator, XRE family	RPA4395	5.22
Light-harvesting complex 1 alpha chain	RPA1526	5.17
ABC transporter, with duplicated ATPase domains	RPA2190	5.14
Possible efflux protein	RPA1421	4.92
Uncharacterized protein	RPA0792	4.68
Uncharacterized protein	RPA4056	4.55
Putative phosphatidylethanolamine-N-methyltransferase	RPA1307	4.54
L-lactate dehydrogenase	RPA4320	4.44
DUF185	RPA4359	4.40
Uncharacterized protein	RPA2621	3.67
LexA repressor	RPA2903	3.57

Table 3.1. All proteins that are significantly up-regulated in *R. palustris* cells grown on 5-phenylvaleric acid compared to *R. palustris* cells grown on benzoic acid. The fold-changes of proteins with a t-test significance based on false discovery rate of 0.05 are shown.

One important protein is the Fcs2 (RPA1787) CoA ligase that is highly up regulated along with another long chain fatty acid CoA Ligase (RPA1766), in addition to many transporters that are up regulated such as RPA3720. RPA1706, which codes for a putative enoyl-CoA hydratase and is present in a cluster of genes that may code for a β-oxidation pathway, is also significantly up-regulated (Figure 3.6).

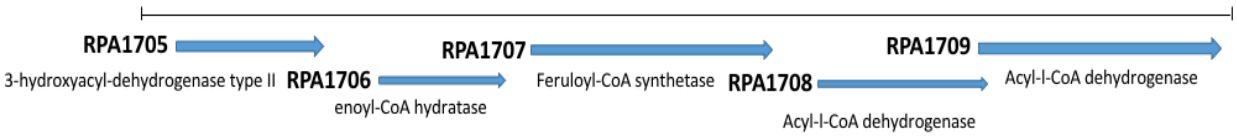


Figure 3.6 : Cluster of genes found in the *R. palustris* CGA0009 genome and thought to be associated with a β-oxidation pathway. All proteins encoded in the cluster were detected by mass spectrometry, although not all were significantly up-regulated (see text).

Many CoA ligases were detected in mass spectrometry and their average ratio in 5-phenylvaleric acid samples were > 2. They included many putative long chain fatty acid-CoA ligases and putative feruloyl-CoA synthetase 1 (Fcs1) (Table 3.2). The *fcs1* (*rpa1707*) gene is present at the same cluster of genes as *rpa1706* and almost all proteins in this cluster were detected in mass spectrometry but apart from Fcs1 only the RPA1706 protein was significantly more abundant in 5-phenylvaleric acid grown cells with a ratio of 9.45 (Table 3.1 and accompanying CD Table1). In addition to the different CoA ligases many different ABC transporters were also up-regulated (Table 3.3). Investigation into transporters and CoA ligases could give an indication by which catabolic pathway different side chain phenylalkane carboxylic acids are utilized.

Protein Name	Gene name	t-test Significant	Average ratio 5-pp/benz
Putative feruloyl-CoA synthetase	RPA1787 (Fcs2)	+	22.05
Putative long-chain-fatty-acid-CoA ligase	RPA1766	+	11.11
4-hydroxybenzoate--CoA/benzoate--CoA ligase	RPA0669		9.80
Putative feruloyl-CoA synthetase	RPA1707 (Fcs1)		7.25
Putative long-chain fatty acid-CoA ligase	RPA3657		5.19
Putative crotonobetaine/carnitine-CoA ligase	RPA4829		4.09
Putative acid-CoA ligase	RPA0743		3.79
Putative fatty-acid--CoA ligase	RPA1412		3.65
Pimeloyl-CoA ligase	RPA3716		2.97
Possible acid-CoA ligase	RPA2665		2.60
Putative 4-coumarate:CoA ligase 2	RPA4421		2.50
Putative fatty-acid--CoA ligase	RPA2142		2.30
Putative acyl-CoA ligase	RPA1702		2.08
Putative long-chain-fatty-acid--CoA ligase	RPA2714		2.03

Table 3.2 : Different CoA ligase proteins that were detected during the growth of wild type *R. palustris* cells when compared to cells grown on benzoate with a ratio higher than 2 . Both RPA1766 and Fcs2 are significantly upregulated but other CoA ligases such as Fcs1 may also play a role in the utilization of 5-phenylvaleric acid. + indicate significant shift in protein intensity.

Protein Name	Gene name	t-test Significant (down shift imputation)	Average ratio 5-pp/benz
Amino acid ABC transporter, ATP-binding protein aapP-1	RPA2563	+	25.38
Putative RND efflux transporter	RPA3775	+	8.32
Putative cation transport ATPase, possible copper transporter	RPA2333		7.84
Putative branched-chain amino acid transport system ATP-binding protein	RPA3720	+	7.42
Putative high-affinity branched-chain amino acid transport system ATP-binding protein	RPA3719		5.35
ABC transporter, with duplicated ATPase domains	RPA2190	+	5.14
Possible branched-chain amino acid transport system substrate-binding protein	RPA0106		4.79
Branched-chain amino acid transport system ATP-binding protein	RPA1791		4.42
Periplasmic mannitol-binding protein, SmoM possible Trap-T transport system, dctP subunit	RPA1975		3.84
CBS domain:Transport-associated domain	RPA4246		3.28
Putative multidrug-efflux transport protein mexF	RPA4095		2.82
Putative branched-chain amino acid transport system ATP-binding protein	RPA3294		2.61
Putative branched-chain amino acid transport system ATP-binding protein	RPA3293		2.52
Putative periplasmic binding protein for ABC transporter for branched chain amino acids	RPA1798		2.44
Possible high-affinity branched-chain amino acid transport system substrate-binding protein	RPA0806		2.35
Putative molybdenum transport system protein	RPA1002		2.23
RND multidrug efflux transporter MexD	RPA1498		2.10
Possible branched-chain amino acid transport system substrate-binding protein	RPA1415		2.00

Table 3.3 : Different proteins associated with a transport system that were detected during the growth of wild type *R. palustris* cells on 5-phenylvalerate when compared to cells grown on benzoate. The significantly changed protein includes mainly ABC- type transport system. One of interest is RPA3720. + indicate significant shift in protein intensity.

3.2.1.1 Post-translational modifications.

N-lysine acetylation / de-acetylation regulates many central metabolic enzymes such as acetyl CoA-synthetase (Peng et al. 2011). Acetylation occurs on a conserved lysine residue in the active site of acetyl CoA-synthatase and it decreases the activity of the enzyme (Starai et al. 2002). Acyl-CoA ligases are known to be regulated by acetylation in *R. palustris* grown on benzoic acid (Crosby et al. 2012). Mass spectrometry identified the acetylated proteins in *R. palustris* grown on 5-phenylvaleric acid as mentioned above. Around 64 proteins were identified to be acetylated in cells grown on 5-phenylvaleric acid (accompanying CD Table 2). Selected acetylated proteins are shown in Table 3.4, of the total proteins identified only 11 were acetylated CoA-ligases. The raw mass spectrometry data were analysed using MaxQuant (Cox et al. 2009) and the acetylated CoA ligases were mainly putative long chain Fatty Acid CoA ligases. Furthermore, other proteins were identified such as RPA0944 (Glyceraldehyde-3-phosphate dehydrogenase(GAPDH)) and RPA2138 (Putative acyl-CoA dehydrogenase). The consequences of acetylation were not analysed further in this study, but the data suggest that the activity of various acyl-CoA ligases might be regulated in this manner.

3.2.2. Transcriptional analysis of selected genes using quantitative real time PCR (*q*RT-PCR)

In order to further characterize the genes associated with 5-phenylvaleric acid utilization, *q*RT-PCR was used to identify up-regulated genes on a transcriptional level and to try to correlate the results with the proteomics. *q*RT-PCR is one of the most sensitive techniques available for mRNA detection and quantification and it was chosen to examine gene expression in wild type *R. palustris* grown with the different

carbon sources used here. The technique is composed of two steps: the synthesis of cDNA from RNA by reverse transcription and amplification of a specific cDNA by PCR

Protein name	Gene name	Modified sequence	Position
Putative long-chain-fatty-acid--CoA ligase	RPA2714	SIDFVEAIPVTGLGK*IDRK	502
Putative long-chain-fatty-acid CoA ligase	RPA3299	SVEITETPLPLSGAGK*ILK	499
Cyclohexanecarboxylate-CoA ligase	RPA0651	DAMPATPSGK*IQK	532
Acyl-CoA synthetase	RPA4443	WWMPDDIVFVEAIPHTATGK*ILK	522
Putative acyl-CoA ligase	RPA1702	TPTGK*LVK	496
2-ketocyclohexanecarboxyl-CoA hydrolase	RPA0653	AIFGQVGPK*MGSVDPGYGTAFLAR	135
Glyceraldehyde-3-phosphate dehydrogenase(GAPDH)	RPA0944	AAAMSMIPTSTGAAK*AIGLVLPELK	215
Benzoate-CoA ligase	RPA0661	TATGK*IQR	512
Acyl-CoA synthetase	RPA2302	TPSGK*LQR	539
Putative oligoendopeptidase F	RPA4185	MAK*SATSR	3
Putative crotonobetaine/carnitine-CoA ligase	RPA4829	STLEK*VAK	516
Putative long-chain fatty acid--CoA ligase	RPA1449	IIDALPMTATGK*VK	559
Putative long-chain-fatty-acid CoA ligase	RPA1763	NASGK*ILR	506
Probable transcriptional regulator, TetR family	RPA2294	LK*SAAQK*TAAK	3
Possible +E2677pyruvate-flavodoxin oxidoreductase	RPA4721	EGLEATHK*VDYDAAEFSEDAIKAA	824
Phenylacetate-coenzyme A ligase	RPA1723	LAWSLKHAYDNVAHYK*AAFDR	51
Putative acyl-CoA dehydrogenase	RPA2138	NK*SLICVPMKTK	181
Putative long-chain-fatty-acid-CoA ligase	RPA4823	LPTGK*LYK	497

Table 3.4 Selected acetylated proteins from mass spectrometry analysis by MaxQuant. Acetylated lysine residues are indicated by * with their position within the protein.

R. palustris cultures were grown anaerobically in the light to OD_{660} 0.6 in an RCV media containing 3 mM of either 5- phenylvaleric acid, succinic acid or benzoic acid as the sole carbon source. After that, RNA was extracted using TRIzol® protocol (see Materials and Methods) and quantitative real-time PCR (qRT-PCR) analyses were performed using the Syber green Agilent kit (Agilent technologies, USA). Real-time PCR included 200 ng RNA and 25 μM primers in SYBR green PCR amplification master mix. Genomic DNA was used as a standard, and RNA polymerase sigma 70 subunit *RpoD* (*rpa1288*) constitutive expression was used as an internal control. Primers used are listed in (Table 2.2).

The results showed that when *R. palustris* is grown on 5-phenylvaleric acid there is an approximately 8.7 fold increase (compared to benzoate grown cells) in *fcs2* gene expression which is a CoA ligase associated with the non-β-oxidation pathway. A 7.2 fold increase in *rpa3720* gene expression was also monitored and that codes for a putative branched-chain amino acid transport system ATP-binding protein. Furthermore, *rpa1705*, *rpa1706* and *rpa1786* showed more than a 2 fold increase but less than a 2 fold increase was observed for the other genes studied as shown in Figure 3.7.

fcs1 did not show a significant fold change when cells grown on 5-phenylvaleric acid were compared to cells grown on benzoic acid but on the other hand when cells grown on succinic acid were compared to cells grown on 5-phenylvaleric acid, an elevated level of *fcs1* and *rpa1704* which codes for a CoA ligase and a regulator respectively were recorded (Figure 3.8). In addition, cells grown on benzoic acid were compared to cells grown on succinic acid and the results showed that genes in the gene cluster associated with the β-oxidation pathway were up-regulated by more than

2 fold; that includes *rpa1704*, *rpa1705*, *rpa1706*, *rpa1707*(*fcs1*) and *rpa1766*. However, *fcs2* and *rpa3720* were not up regulated, which indicates their specific up-regulation to the presence of 5-phenylvaleric acid (Figure 3.9). *q*RT-PCR results were overall similar to the mass spectrometry results which indicate that there may be other CoA ligases involved in the initial degradation of 5-phenylvaleric acid.

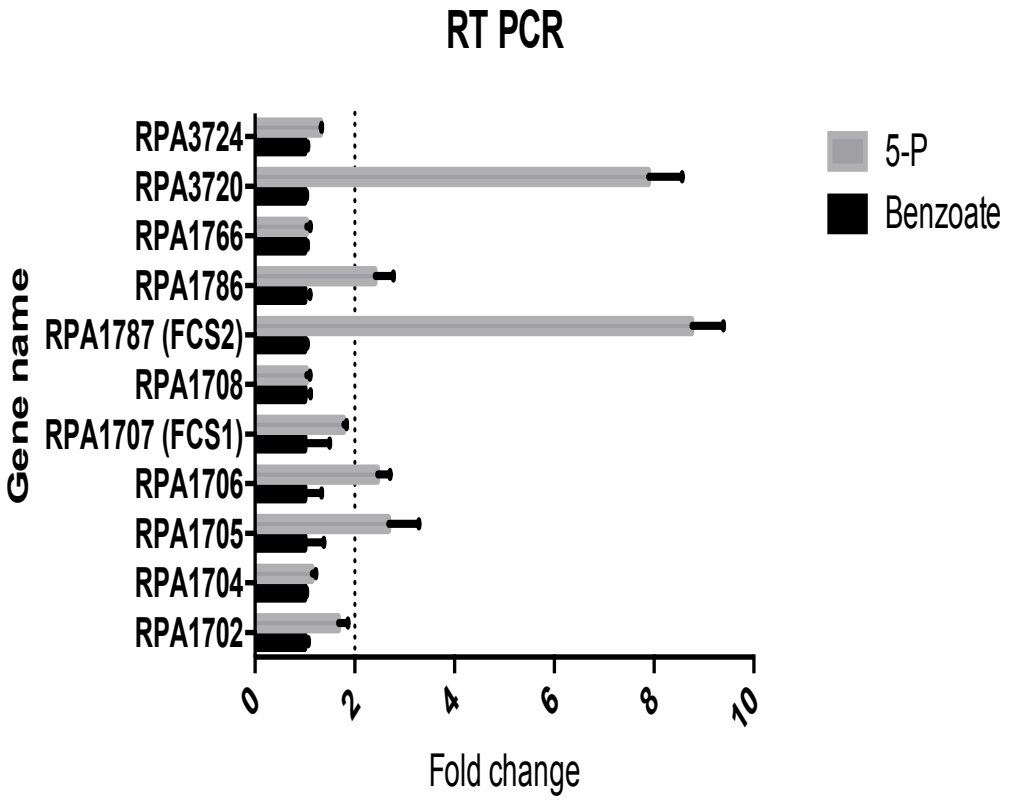


Figure 3.7 . Analysis of gene expression changes in wild type *R. palustris* grown on benzoic acid compared to *R. palustris* grown on 5-phenylvaleric acid. *R. palustris* was grown anaerobically to mid-exponential phase in RCV media; Total RNA was extracted and qRT-PCR performed using Syber green Agilent kit. Gene expression was normalized to *RpoD* and the mean value for the benzoic acid grown cells were set to 1 fold. The fold change in expression of these genes in 5-phenylvaleric acid grown cells is indicated by the bars which represent the mean and standard deviation of three independant experiments.

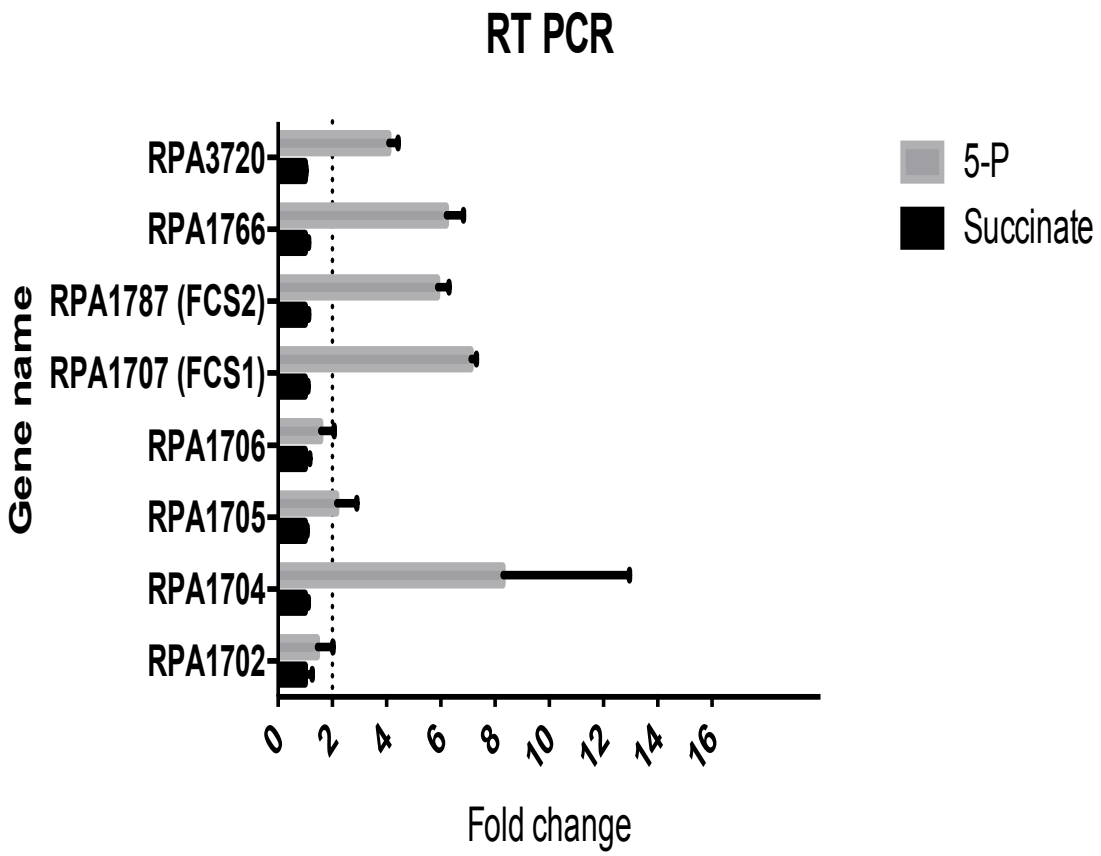


Figure 3.8. Analysis of gene expression changes in wild type *R. palustris* grown on succinic acid compared to *R. palustris* grown on 5-phenylvaleric acid. *R. palustris* was grown anaerobically to mid-exponential phase in RCV media; Total RNA was extracted and qRT-PCR performed using Syber green Agilent kit. Gene expression was normalized to *RpoD* and the mean values for the succinic acid grown cells were set to 1 fold. The fold change in expression of these genes in 5-phenylvaleric acid grown cells is indicated by the bars which represent the mean and standard deviation of three independant experiments. All CoA ligases: *fcs1*, *fcs2* and *rpa1766* under investigation showed an average fold change of more than 5 fold.

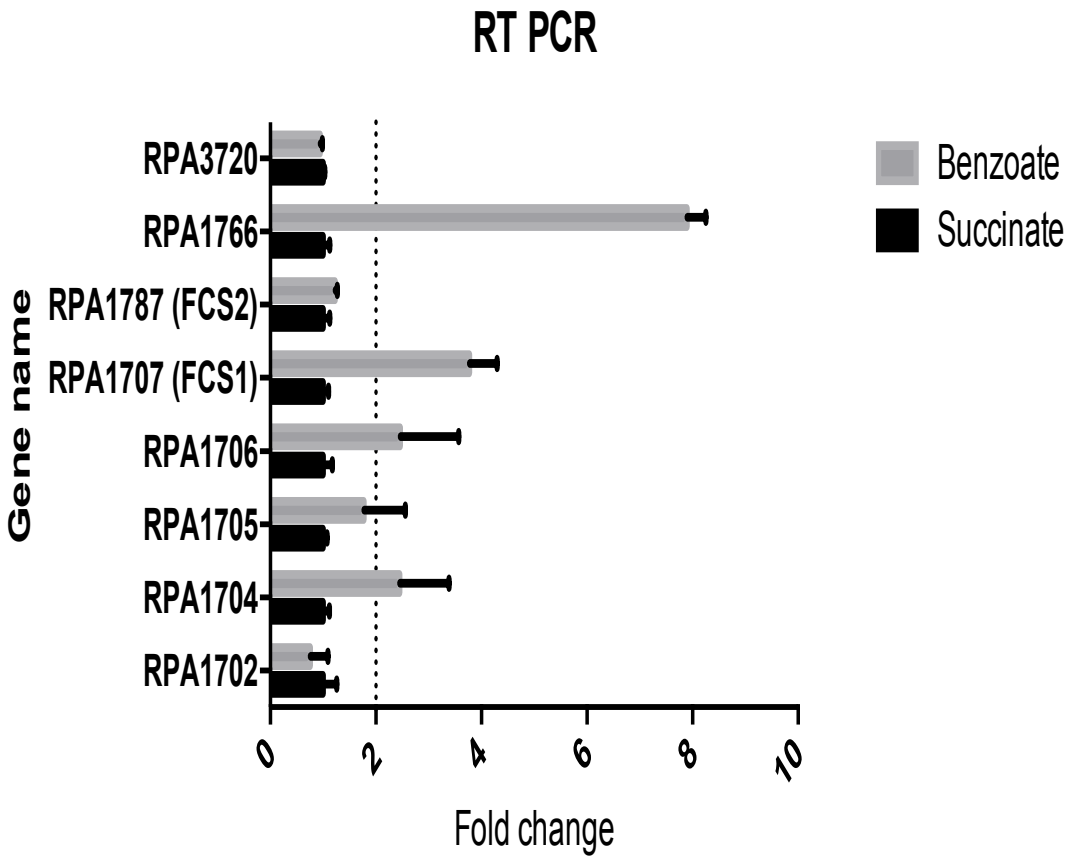


Figure 3.9 . Analysis of gene expression changes in wild type *R. palustris* grown on succinic acid compared to *R. palustris* grown on benzoic acid. *R. palustris* was grown anaerobically to mid-exponential phase in RCV media. Gene expression was normalized to *RpoD* and the mean values for the succinic acid grown cells were set to 1 fold. The fold change in expression of these genes in benzoaic acid grown cells is indicated by the bars which represent the mean and standard deviation of three independant experiments. *rpa1766* and *rpa1707* (*fcs1*) showed the higest fold change.

3.3 Discussion.

The work done in this chapter has provided a detailed prospective of possible proteins associated with anaerobic degradation of lignin derived phenylalkane carboxylic acid with medium length side-chains. To get a broader idea of routes of degradation for

phenylakanes carboxylic acids, 5-phenylvaleric acid was used as a substrate for wildtype *R. palustris* and both proteomic and gene expression techniques were used. The first step in the utilization of different chain length phenylalkane carboxylic acids is the formation of a CoA derivative by a CoA ligase (Elder et al. 1992). The *R. palustris* CGA009 genome codes for almost 44 predicted CoA ligases and it was hard to predict exactly which gene was involved in the utilization of 5-phenylvaleric acid.

Mass spectrometry results detected around 25 up-regulated CoA-ligases of which RPA1766 and Fcs2 (CouB) showed a significant shift when cells grown on benzoic acid were compared to cells grown on 5-phenylvaleric acid. In addition to that, RPA1706, which is a putative enoyl-CoA hydratase, was significantly up-regulated and the gene for this protein is present in a cluster of genes associated with a β-oxidation pathway. The gene for RPA1707 (Fcs1), that codes for a putative feruloyl-CoA synthetase 1, is also present in this cluster. Fcs2 (CouB) is the CoA-ligase for *p*-coumarate degradation and it was also significantly increased in abundance when cells grown on benzoate were compared to cells grown on *p*-coumarate (Pan et al. 2008). In addition, Fcs1 and RPA1766 showed no change in protein level when cells grown on *p*-coumarate were compared to cells grown on benzoate. Hirakawa et al. (2012) tested the substrate specificity of Fcs2 CoA ligase with different side chain phenyl carboxylic acid having side chain length of one carbon to five carbons and only the phenyl carboxylic acids with side chain lengths of three carbons were substrates for Fcs2 (Hirakawa et al. 2012). This clearly shows that Fcs2 is not the CoA ligase for 5- phenylvaleric acid as the biochemical tests done by Hirakawa *et al*., (2012) proved that Fcs2 can utilize phenyl carboxylic acids having a side chain of three carbons only.

So even though Fcs2 is significantly up regulated there must be other CoA ligases that can utilize 5-phenylvaleric acid. The significant up regulation of RPA1766 which is a putative long-chain-fatty-acid-CoA ligase in cells grown on 5-phenylvaleric acid could be due to the presence of a fatty-acid like side chain and *R. palustris* could utilize 5-phenylvaleric acid as if it was a long chain fatty acid first, and then as a phenyl carboxylic acid with 3 carbon side chain. That seems the most likely scenario by which 5-phenylvaleric acid is converted to its CoA derivative but the presence of the enoyl-CoA hydratase (RPA1706) opened a new possibility as whether other CoA ligases such as Fcs1 (RPA1707) could be expressed in the presence of both benzoic acid and 5-phenylvaleric acid and may thus play a role in the utilization of 5-phenylvaleric acid.

*q*RT-PCR was done to support the mass spectrometry results and to investigate other growth conditions. Gene expression for *R. palustris* grown on succinic acid, benzoic acid and 5-phenylvaleric acid were compared and the results showed that *fcs1* gene expression is up-regulated in both benzoic acid and 5-phenylvaleric acid grown cells when compared to succinic acid grown cells. The elevated level of *fcs1* in the presence of benzoic acid could be the reason why no significant up regulation of the protein was recorded in mass spectrometry since protein expression was compared in cells grown on 5-phenyvaleric acid versus benzoic acid only. Growth of wild type *R. palustris* with benzoate as the sole carbon source did not induce *fcs2* gene expression and this also was documented by Hirakawa *et al*. (2012), who showed that *fcs2* gene expression is elevated by the presence of coumarate and not benzoate. Almost all genes tested for gene expression correlated with mass spectrometry results.

Elder *et al*. (1992) suggested that different side chain length phenylalkane carboxylic acids could be degraded in early stages of growth through β-oxidation pathway by wild type *R. palustris*. Later studies done on *Aromatoleum aromaticum* confirmed the use of a β-oxidation pathway by proteomic study for cells grown on cinnamate, coumarate or phenylalkane carboxylates (Trautwein et al. 2012).

The mass spectrometry results not only provided us with possible CoA ligases but also provided us with evidence of different possible transport systems involved in 5-phenylvaleric acid uptake (Table 3.3). The hydrophobic nature of compounds with longer side chains could allow for simple diffusion across the membrane into the cell but that does not mean there would not be a dedicated uptake system for longer chain phenylalkane carboxylic acids. Most detected transport systems in our mass spectrometry results belong to ABC-type uptake systems. An ABC transport system was also up-regulated in a study done by Phattarasukol *et al.*, (2012), which suggested RPA1789 and RPA1791-1793 as the transport system for *p*-coumarate uptake. In the mass spectrometry results of cells grown on 5-phenylvaleric acid a different set of proteins were significantly up-regulated and that included; AapP-1, RPA3775, CtpC, RPA2190 and RPA3720. Here, each protein is a part of different transport system. One of interest is RPA3720 which is a putative high-affinity branched-chain amino acid transport system ATP-binding protein that is present in a ABC transporter cluster of genes and this system has three soluble binding proteins. The transporter genes are also next to the *pim ABCDF* operon encoding pimelate degradation enzymes. Some of the Pim proteins were detected by mass spectrometry but their up-regulation was not significant (accompanying CD Table 1). Studying this transport system could give us an insight into whether different chain phenyl alkane carboxylic acids are transported through it; this is the scope of the last results chapter in this thesis.

In addition to up-regulated genes and proteins identified in this chapter some important post-translational modifications were identified and that included around 11 CoA-ligases. Lysine acetyltransferases carry out lysine modification of proteins and *R. palustris* possesses the RpPat protein which is an acetyltransferase that regulates several acyl-CoA ligases that are involved in anaerobic degradation of aromatic compounds (Crosby et al. 2010). Acetyl-CoA ligases that are involved in anaerobic degradation of aromatic compounds such as BadA, a benzoate CoA-ligase, HbaA, a 4-hydroxy benzoate CoA-ligase and AliA, acyclohexane carboxylate CoA-ligase are regulated by RpPat protein catalysed acetylation (Crosby et al. 2010). In addition to those proteins that are highly specific to aromatic compound metabolism, other proteins that are involved in anaerobic fatty acid and dicarboxylic acid metabolism were also studied by Crosby *et al*., (2012) and they were also found to be regulated by RpPat, this includes RPA1702, RPA4504, RPA1003, RPA2302 and RPA4267. All CoA-ligases studied were acetylated at a conserved lysine residue, which is present in the active site within the conserved PX_4GK motif (Crosby & Escalante-Semerena 2014). 5-phenylvaleric acid is a combination of an aromatic ring and a 5 carbon fatty acid compound so the CoA- ligases that were identified to be regulated by acetylation in this study were a combination of benzoate CoA-ligases and fatty acid CoA-ligases (Table 3.5). When acetylation of CoA-ligases in cells grown on benzoic acid were compared to protein acetylation in cells grown on 5-phenylvaleric acid, the acetylated CoA-ligases that were only present in 5-phenylvaleric acid grown cells were mainly putative long chain fatty acid CoA-ligases (RPA2714, RPA3299 and RPA1449) in addition to one other acyl-CoA ligase, RPA4443. On the other hand, the significantly up-regulated Fcs2 and RPA1766 CoA- ligases were not detected as being acetylated also Fcs1 was not acetylated. It is possible that those CoA ligases are not subject to

lysine acetylation modification to regulate their function like the other acyl-CoA ligases since those proteins are only needed to initiate a specific reaction in a peripheral pathway. Different CoA ligases and transporters were up-regulated when grown on 5-phenylvaleric acid, and studying their biochemical characteristics could help in determining how different chain length phenylalkane carboxylic acids are degraded by *R. palustris*. Biochemical characterization of each CoA ligase is needed to assess their substrate specificity and that was the scope of the next chapter.

4. Characterization of three putative CoA-ligases from *Rhodopseudomonas palustris* and their ability to utilize phenylalkane carboxylic acids of differing side-chain lengths

Characterization of three putative CoA-ligases from *Rhodopseudomonas palustris* and their ability to utilize phenylalkane carboxylic acids of differing side-chain lengths

4.1 Introduction:

The biochemical fate of different length side chain phenylalkane carboxylic acids is thought to proceed through either a β-oxidation pathway or a non-β-oxidation pathway to produce benzoyl-CoA (Pan et al. 2008; Elder et al. 1992). In both pathways the activation of aromatic acids compounds for entrance into these pathways is believed to be through a CoA ligase activity. The two pathways differ only in acetyl-CoA cleavage : the non-β-oxidation route cleaves it off immediately after the initial CoA-thioester formation and then gains the CoA-thioester at a subsequent stage, while in the β-oxidation route the acetyl-CoA group cleavage occurs at a later stage in a transferase reaction (Figure 4.1) (Pan et al. 2008).

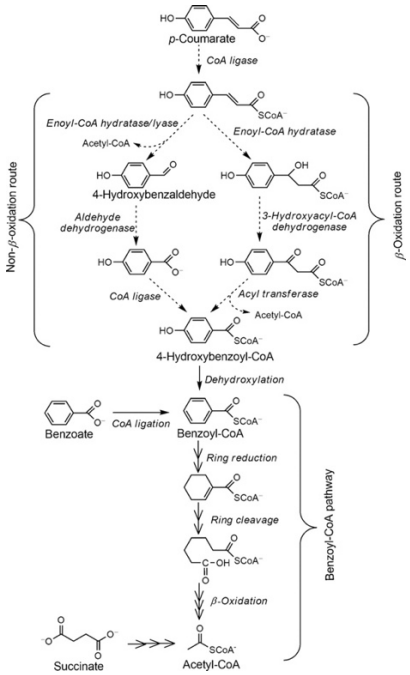


Figure 4.1 Proposed two possible pathways for coumarate degradation to produce benzoyl CoA ; The β-oxidation and the non β-oxidation pathway (Pan et al. 2008).

CoA ligases in general catalyze the formation of aromatic acyl-CoA derivatives in the presence of ATP, CoA and $MgCl_2$. In addition to the aromatic acyl-CoA formation, AMP and PPi will also be produced (Figure 4.2) (Martinez-Blanco et al. 1990). The reaction is a two-step process. First, the carboxyl group of the aromatic acid will be linked by an acyl bond to the phosphoryl group of ATP, then the second step involves the transfer of the acid acyl group to the sulfhydryl group of Coenzyme A, which in turn will release AMP (Becker-André et al. 1991).

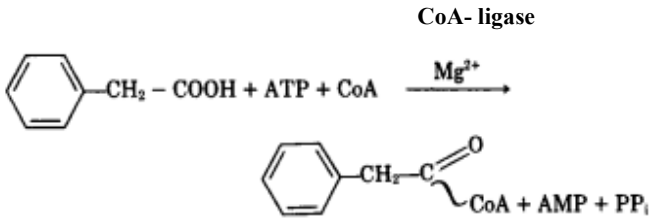


Figure 4.2 CoA ligase chemical reaction to produce aromatic Acyl-CoA. (Martinez-Blanco et al. 1990)

Based on preliminary results obtained by mass spectrometry and qRT-PCR, RPA1787 (Fcs2) and RPA1766 were significantly up-regulated when bacterial cells were grown on 5-phenylvaleric acid (5-p) compared to bacterial cells grown on benzoic acid. Furthermore, RPA1707 (Fcs1) was also detected in the proteomic study and in qRT-PCR, but it's expression and detection level was low when compared with the other two CoA ligases. Pan *et al.,* (2008) carried out a proteomic study on *R. palustris* cells grown on coumaric acid compared to succinic acid or benzoic acid and showed that CoA ligase enzymes such as RPA1787 (Fcs2) were highly expressed when grown on coumaric acid. This enzyme is present in a cluster of gene encoding enzymes of the non-ß-oxidation pathway. On the other hand, both RPA1766 which is annotated as long chain fatty acid CoA ligase and RPA1707 (Fcs1) which is encoded in a gene cluster thought to be involved in a β-oxidation pathway were also detected in the

proteomic study. Elder *et al* (1992) suggested that a β-oxidation pathway is the route of degradation for straight chain phenylalkane carboxylic acids like phenylpropionic acid (Elder et al. 1992). The aim of the work of this chapter was to determine the substrate utilization ability for the three CoA-ligases (RPA1707, RPA1787, RPA1766) and to measure their enzyme kinetics with different length side chain phenylalkane carboxylic acids, as these enzymes are the first step for their respective pathways; their substrate utilization ability could indicate which pathway is preferred. In addition to that, mutants in these CoA ligases were created to assess their phenotype and effect on *R. palustris* grown on different length side chain phenylalkane carboxylic acids.

It was discovered that Fcs1 has the highest substrate specificity and activity with 7-phenyleheptanoic acid and has a substrate range from 5-phenylvaleric acid up to 8-phenyloctanoic acid. However, no activity was measured with Fcs2 except with 3-phenylpropanoic acid. RPA1766 showed very weak activity with 8-phenyloctanoic acid *in vitro*. This supported the idea that Fcs1 and the ß-oxidation pathway encoded at the same locus could be involved in the metabolisim of different length side chain phenylalkane carboxylic acids as proposed by Elder *et al*., (1992) and further investigation into the different enzymes in the pathways is needed. On the other hand, single *fcs1* and double *fcs1* and *fcs2* mutants of *R. palustris* were still able to grow on different length side chain phenylalkane carboxylic acids and no phenotypic change was recorded. This indicates that other CoA ligases could *in vivo* contribute in the utilization of different length side chain phenylalkane carboxylic acids to produce benzoyl-CoA.

4.2 Results

4.2.1 Recombinant expression and purification of RPA1707 (Fcs1) RPA1787 (Fcs2) and RPA1766.

To assist studies on substrate utilization ability and enzyme biochemical characteristics the three CoA ligases were overexpressed in *E. coli* as his-tagged proteins and purified using nickel affinity chromatography**.**

4.2.1.1 Overproduction and purification of Fcs1, Fcs2

The genes encoding Fcs1 and Fcs2 were cloned previously in the lab by R. Salmon. Fcs1 was cloned and expressed using the pET-21 vector and BL21 *E. coli* cells while Fcs2 was cloned and expressed using the pBAD/His vector and Top10 *E. coli* cells. Expression of Fcs1 was induced with 400 μM IPTG via T7 promoter. While Fcs2 expression was induced by 0.002% arabinose via the *araC* promoter region. Different vectors were used to obtain soluble protein as in the pBAD/His vector, protein expression is more sensitive to control because the titration of arabinose concentration can change the protein expression and solubility level (Guzman *et al.*, 1995). For Fcs1 the 6x His-Tag is present in the C-terminal of the protein Figure (4.3a) on the other hand for Fcs2 the 6x His-Tag is included in the N-terminal region of the protein Figure (4.4a). Protein overexpression and solubility trials showed that the optimum protein expression time for Fcs1 was incubation at 25 ºC for 3 hours with predicted molecular weigh of 68.5 kDa and for Fcs2 it was incubation at 37 °C for 24 hour with predicted molecular weigh of 72.5 kDa. Furthermore, both proteins were purified by affinity chromatography based on His-Tag engineered in the construct, as shown in Figure (4.3b) and (4.4b). Several fractions of protein along the imidazole gradient were collected and were loaded into SDS-PAGE gels (Figure 4.3c and 4.4c). To obtain a highly pure protein and to determine the apparent molecular weight of the

protein, gel filtration chromatography was done for both Fcs1 and Fcs2 (Figure 4.5). The results showed that both proteins eluted as monomers and the apparent molecular weight for Fcs1 was 70.8 kDa while the Fcs2 apparent molecular weight was 87 kDa. The concentration of the purified proteins was measured by UV spectrophotometry and typical results for the concentration was 92µM for Fcs1 and 50 µM for Fcs2.

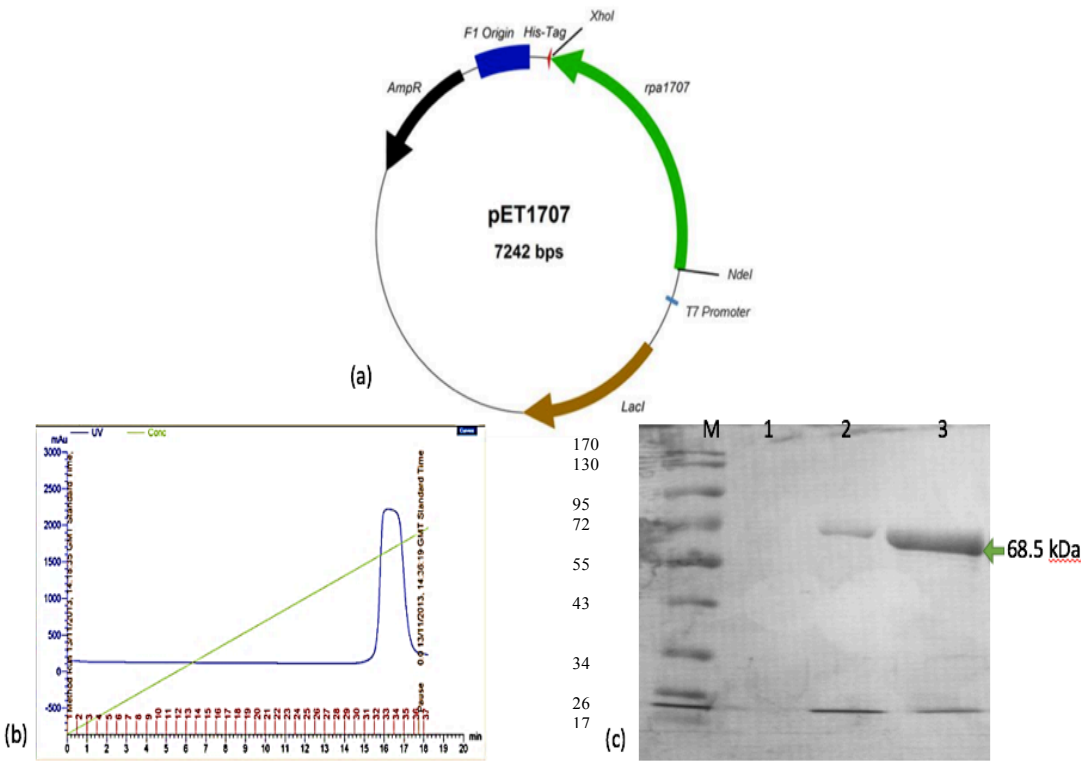


Figure 4.3. Protein overexpression construct and purification for Fcs1 protein. (a) over-expression construct pET1707 including *rpa1707* gene cloned into NdeI and XhoI sites of pET21a(+) which is induced by IPTG due to the presence of T7 promoter. (b) RPA1707 protein purification on His-trap column. (c) SDS-PAGE showing the purified protein of RPA1707(Fcs1). M= pageRuler ^{TM} prestained protein ladder (fermentas). Lanes 1-3 contain fractions eluted from the column with a predicted protein size of 68.5 kDa.

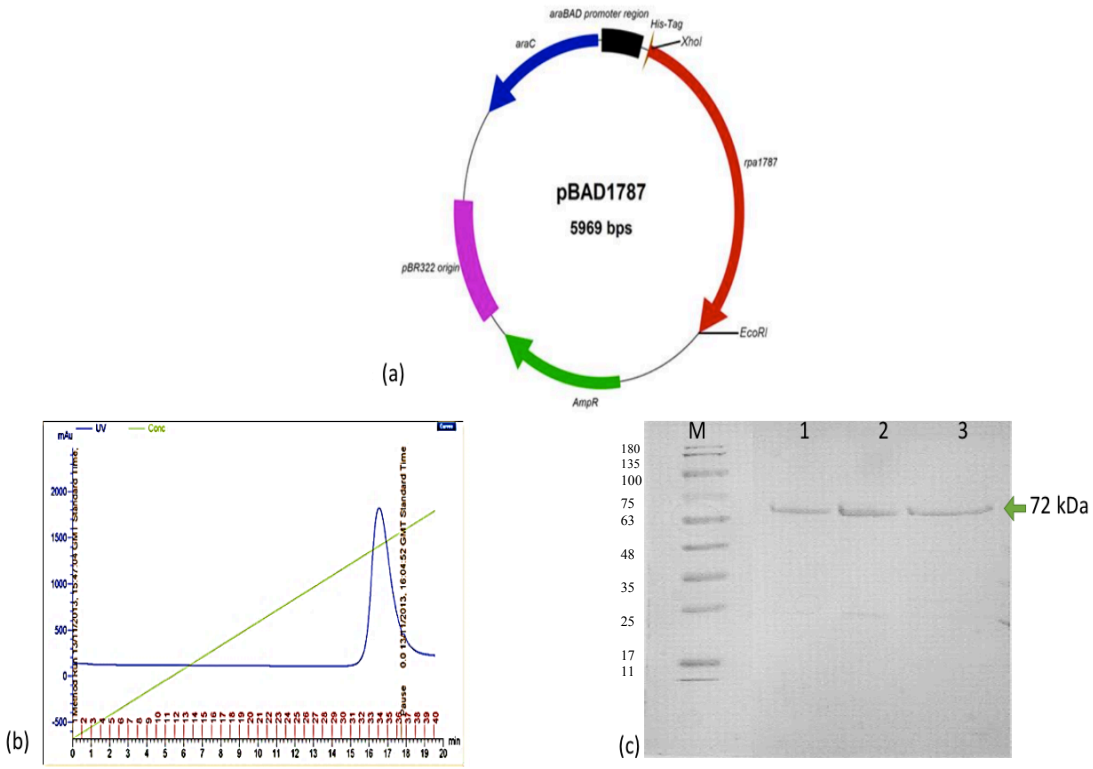


Figure 4.4 Protein overexpression construct and purification for Fcs2 protein. (a) over-expression construct pBAD1787 including *rpa1787*gene cloned into EcoRI and XhoI sites of pBAD which is induced by 0.002% arabinose due to the presence of *araC* promoter. (b) RPA1787 protein purification on His-trap column. (c) SDS-PAGE showing the purified protein of RPA1787(Fcs2). M= pageRuler ^{TM} prestained protein ladder (fermentas). Lanes 1-3 contain fraction eluted from the column with a predicted protein size of 72 kDa.

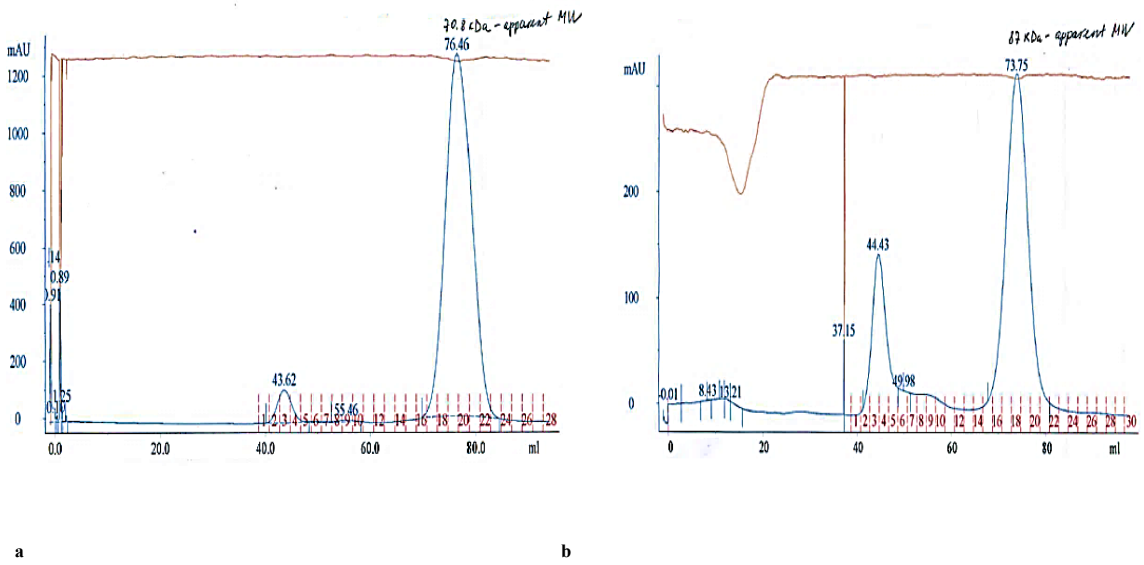


Figure 4.5. Gel filtration chromatography results of Fcs1 (a) and Fcs2 (b). Fcs1 protein apparent weight is 70.8 kDa while Fcs2 protein apparent weight is 87 kDa. This purified protein was used in enzyme assays.

4.2.1.2 Overproduction and purification of RPA1766

The gene encoding RPA1766, which is annotated as a putative long-chain-fatty-acid-CoA ligase, was amplified from *R. palustris* genomic DNA using primer pET1766_F and pET1766_R. The gene was cloned into *NdeI* and *XhoI* sites of digested pET21a (+) and expressed using 400µM IPTG via the T7 promoter. Successful cloning was confirmed by colony PCR and sequencing. pET1766 construct is shown in Figure 4.6a. Protein expression was carried out using BL21(DE3) cells and overexpression trials was preformed at 37°C and 25°C. Protein expression levels was monitored at regular time intervals (1,3,5 hours and overnight after induction). BL21(DE3) cells containing the pET1766 construct were grown to an $OD_{(600nm)}$ of 0.6 at 37 °C and induced by 400 µM IPTG. After that 10ml of the culture at each time interval and temperature was taken and centrifuged. For each time point a cell free extract was run on an SDS-PAGE gel to evaluate protein expression (Figure 4.6b). The solubility of the protein was also checked at different time intervals and the most soluble protein

was recorded at 5 hours from induction when incubated at 25 °C. RPA1766 protein cell free extracts were purified by affinity chromatography based on the His-tag engineered in the construct. A 20 ml volume of cell free extract was run through a HisTrap nickel affinity column and protein elution in 1 ml/minute samples was monitored through UV readings (Figure 4.6c). A UV single peak consisting of ~5 samples was detected and 5µl of different fractions eluted were run on an SDS–polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 4.6d) to determine their purity and predicted molecular weigh of 56 kDa. UV spectrophotometry was used to determine protein concentration; on average a concentration of 64 µM was calculated for RPA1766 protein. The protein was quite pure and in a good concentration to carry out enzymatic assays.

4.2.2 Enzymatic Analysis of Fcs1, Fcs2 and RPA1766 with different length side chain phenylalkane carboxylic acids

To determine the kinetics for Fcs1, Fcs2 and RPA1766 a coupled enzyme assay linked to NADH oxidation was used. In this assay AMP formation was enzymatically coupled to pyruvate kinase, lactate dehydrogenase, and myokinase reactions (Figure 4.7) For every mol of aromatic substrate, 2 mol of NADH is oxidized (Schühle et al. 2003) NADH oxidation was monitored by fluorescence excitation at 340 nm and emission at 460 nm using an Omega florescence spectrophotometer (BMG lab tech). The decrease of NADH florescence was monitored and the maximum slope obtained was used for the rate calculation. The experiments were conducted in a 100 µl reaction mixture containing ATP, $MgCl_2$, Tris-buffer, enzyme and substrate to be tested, and the reaction was initiated by the addition of CoA. Different substrates were

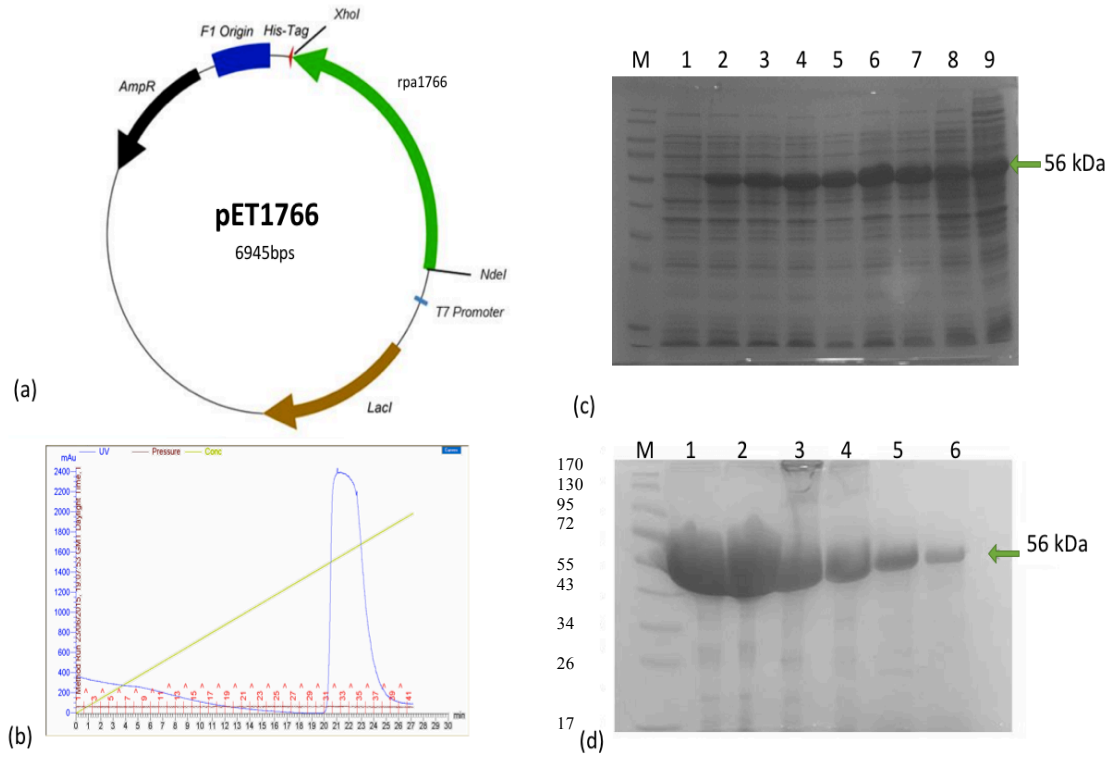


Figure 4.6 Protein overexpression construct and purification for RPA1766 protein. (a) **over-expression construct pET1766** including *rpa1766* gene cloned into NdeI and XhoI sites of pET (+) which is induced by 400 µM IPTG due to the presence of T7 promoter. (b) **RPA1766 protein purification on His-trap columns using Akta**. (c) **SDS-PAGE showing expression trials of RPA1766**. M= pageRuler ^{TM} prestained protein ladder (Fermentas). Lane 1= control without induction , lanes 2,4,6 and 8,contain samples from cells incubated at 25°C at 1,3,5,24 hour interval respectively after IPTG induction , while lane 3,5,7 and 9 , contain samples from cell incubated at 37 °C at 1,3,5,24 hour interval respectively after IPTG induction. Lane 6 shows good amount of protein expression at 56 kDa. (d)SDS-PAGE showing His-Trap purified protein of RPA1766. M= pageRuler ^{TM} prestained protein ladder (fermentas). Lanes 1-6 contain fraction eluted from the Colum with a pure protein size 56 kDa in lane 6 which was used later in enzyme kinetics assay.

$$
\begin{aligned}
&\textit{Aromatic acid} + \textit{Mg ATP} + \textit{CoASH} \rightleftarrows \textit{acyl} - \textit{CoA} + \textit{Mg AMP} + PP_i \\
&\textit{AMP} + \textit{ATP} \rightleftarrows 2\,\textit{ADP} \\
&2\,\textit{PEP} + 2\,\textit{ADP} \rightleftarrows 2\,\textit{pyruvate} + 2\,\textit{ATP} \\
&2\,\textit{pyruvate} + 2\,\textit{NADH} + 2\,H^{+} \rightleftarrows 2\,\textit{lactate} + 2\,\textit{NAD}^{+}
\end{aligned}
$$

Figure 4.7. NADH linked assay where one mol aromatic acid utilization is detected by a decrease in the absorbance corresponding to two mol of NADH (Schühle et al. 2003).

screened that included: 4-Methoxycinnamic acid, 3-Phenylpropionoic acid (hydrocinnamate), 4-Phenylbutyric acid, 5-Phenylvaleric acid, 6-Phenylhexonoic acid, 7-Phenylheptanoic acid and 8- Phenyloctanoic acid . Furthermore, straight chain fatty acids such as Propanoic acid, Butyric acid, Valeric acid, Hexanoic acid, Heptanoic acid and Octanoic acid were also used to assess their utilization.

Substrate screening using this assay showed that Fcs1 had a wide substrate utilization ability and is capable of utilizing aromatic compounds with extended side chains. Purified Fcs1 had CoA ligase activity with methoxycinnamic acid, 5-phenylevaleric acid, 6-phenylhexanoic acid, 7-phenyleheptanoic acid and 8-phenyloctanoic acid (Figure 4.8). In addition, it also had some activity with straight chain fatty acids such as hexanoic acid, heptanoic acid, and octanoic acid (Figure 4.8).

In contrast, Fcs2 enzymatic assay with different length side chain phenylalkane carboxylic acid compounds showed that only compounds with shorter side chains can be utilized by Fcs2. The straight chain fatty acids heptanoic acid and octanoic acid were both utilized by the enzyme (Figure 4.9). The assay for RPA1766 showed little activity with only 8-

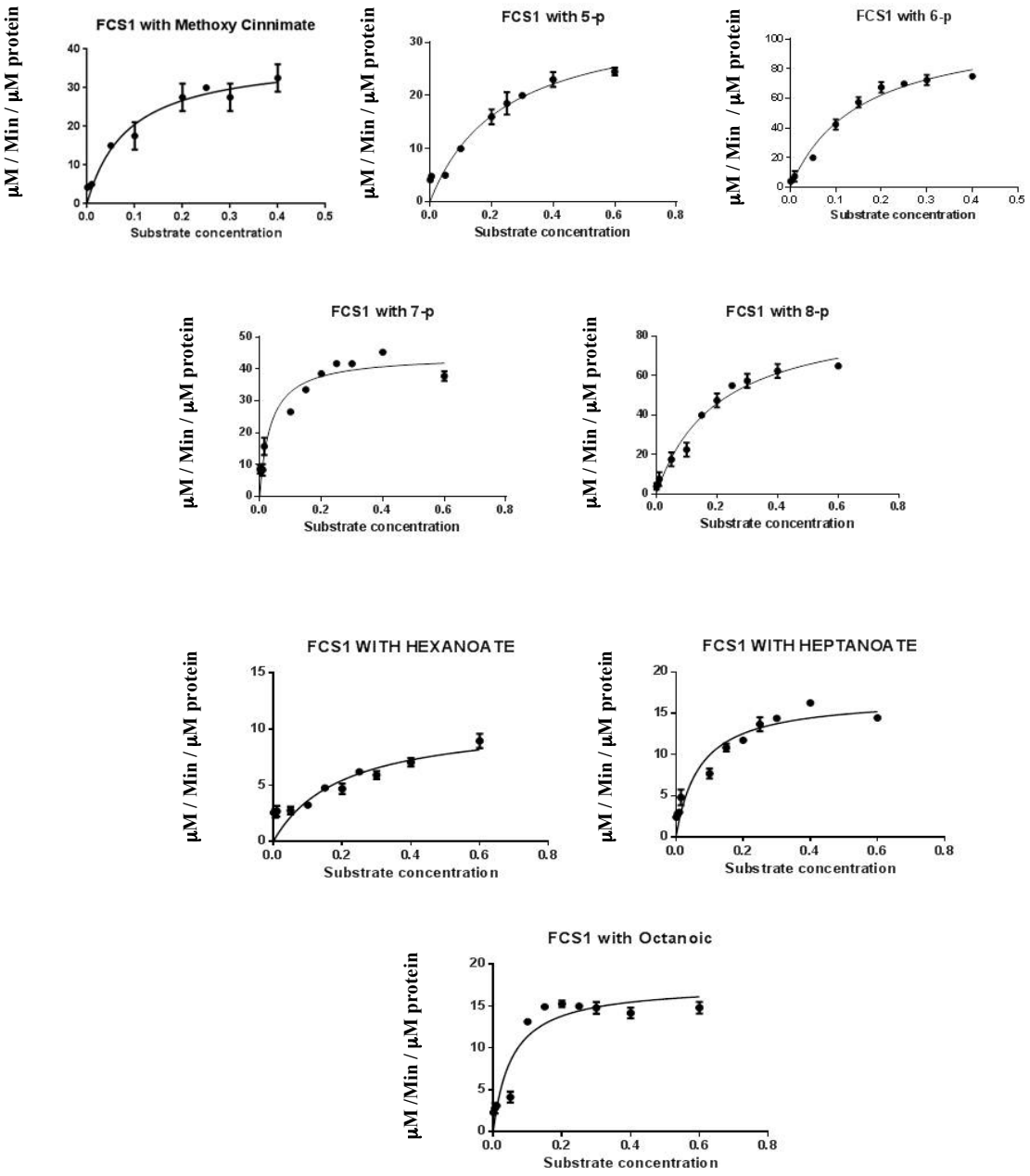


Figure 4.8 Enzyme kinetic analysis of Fcs1 with different phenylpropanoic acids and straight chain carboxylic acid was fitted to the Michaels-Menten equation by using the kinetics software in fluorescence spectrophotometer Omega (MBG LABTECH). 0.1 µM enzyme was used in each assay, which was performed in duplicate.

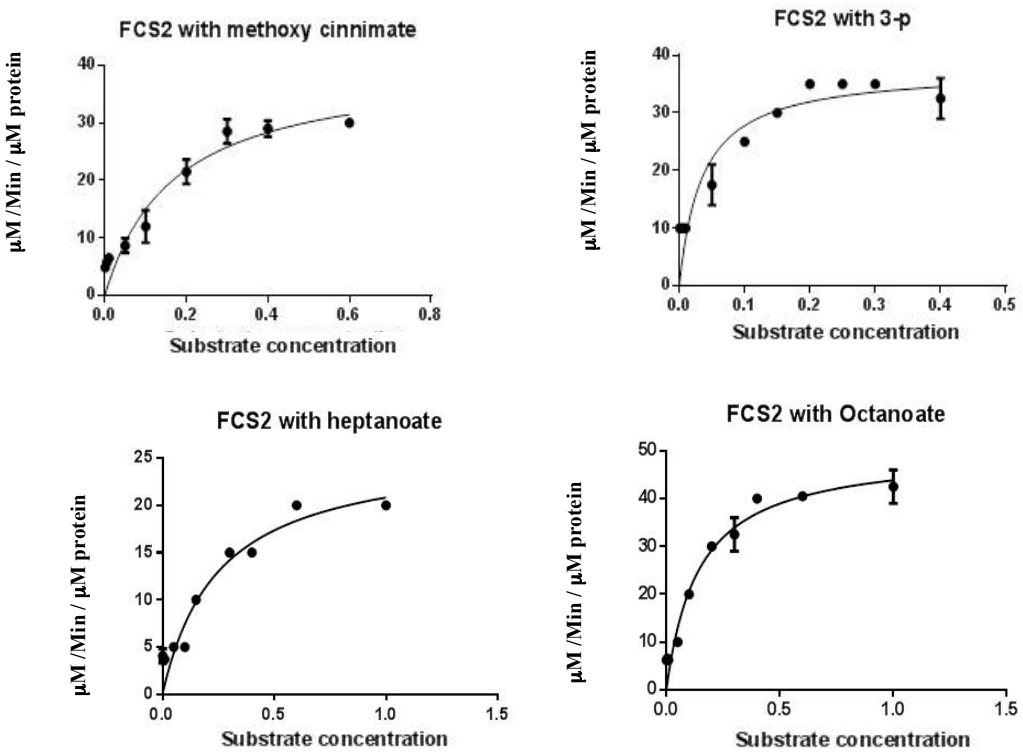


Figure 4.9: Enzyme kinetic analysis for Fcs2 with different phenylpropanoic acids and straight chain carboxylic acids was fitted to the Michaels-Menten equation by using the kinetics software in fluorescence spectrophotometer Omega (MBG LABTECH). 0.1 µM enzyme was used in each assay, which was performed in duplicate.

phenyloctanoic acid being utilized, while all the straight chain fatty acids tested showed no activity with this enzyme (Figure 4.10).

The rates for for Fcs1, Fcs2 and RPA1766 with different substrates were fitted to Michaelis–Menten kinetics by using Graph pad software (Figure 4.8,4.9,4.10). The Michaelis–Menten constants *K*m and kcat were calculated for 0.1µM enzyme

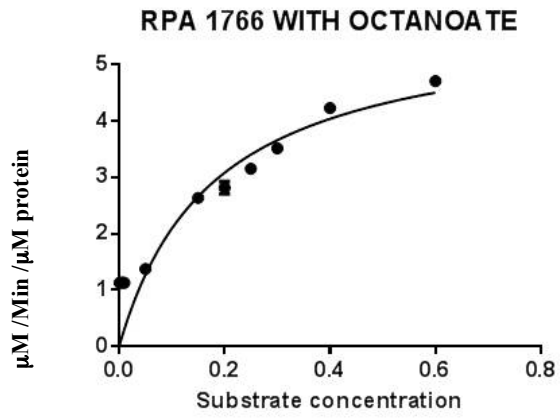


Figure 4.10. Enzyme kinetic analysis for RPA1766 with Octanoic acid. Data were fitted to the Michaelis-Menten equation by using the kinetics software in the fluorescence spectrophotometer Omega (MBG LABTECH). 0.1 µM enzyme was used in each assay, which was performed in duplicate.

concentration Table (4.1). In general, Fcs1 showed higher affinity and specificity with phenylalkane carboxylic acids compounds that have side chains greater than five carbons. 7-Phenyl heptanoic acid was the best substrate utilized by this enzyme. For methoxycinnamic acid, ferulic acid and coumaric acid, Fcs2 had almost twice the value of the specificity constant Km/kcat as Fcs1. Fcs1 enzyme was able to utilize methoxycinnamic acid, ferulic acid and coumaric acid at slower rate while Fcs2 showed no enzymatic activity with 5-phenylvaleric acid, 6-phenylhexanoic acid, 7-phenylheptanoic acid and 8-phenyloctanoic acid.

Substrate	Vmax (µM min^{-1})			Km (mM)			kcat (min^{-1})			kcat/Km mM^{-1} min^{-1}		
Substrate	**Fcs1**	**Fcs2**	**RPA1766**	**Fcs1**	**Fcs2**	**RPA1766**	**Fcs1**	**Fcs2**	**RPA1766**	**Fcs1**	**Fcs2**	**RPA1766**
3-P	0	**37.56**	0	0	**0.03519**	0	**0**	**375.6**	**0**	**0**	**10673**	0
5-P	**49.30**	0	0	**0.2446**	0	0	**493**	**0**	**0**	**2015.5**	**0**	0
6-P	**114.6**	0	0	**0.2002**	0	0	**1146**	**0**	**0**	**5724.2**	**0**	0
7-P	**50.31**	0	0	**0.077**	0	0	**503.1**	**0**	**0**	**6533.7**	**0**	**0**
8-P	**77.93**	0	**5.864**	**0.1380**	0	**0.1802**	**77.93**	**0**	**58.6**	**5644.9**	**0**	**325**
Methoxy-cinnamate	**38.17**	**16.61**	0	**0.0860**	**0.0201**	0	**381.7**	**166.1**	**0**	**4434.2**	**8239**	**0**
coumarate	**5.001**	**42.41**	0	**0.0113**	**0.0481**	0	**50.01**	**424.1**	**0**	**4417.8**	**8817**	**0**
Ferulate	**4.896**	**23.06**	0	**0.0281**	**0.0577**	0	**40**	**230.6**	**0**	**1421.9**	**3996.5**	0
Hexanoic acid	**10.87**	0	0	**0.2004**	0	0	**108.7**	**0**	**0**	**542.4**	**0**	0
Heptanoic acid	**16.99**	**26.06**	0	**0.0712**	**0.2510**	0	**169.9**	**260**	**0**	**2385**	**1038.2**	0
Octanoic acid	**17.72**	**52.68**	0	0.0576	**0.1512**	0	**177.2**	**526.8**	**0**	**3072.1**	**3484.1**	0

Table 4.1 kinetics parameters of different CoA ligase. Enzyme Kinetic calculation for FCS1, FCS2 and RPA1766 were calculated following the rate of NADH florescent at at 340 nm excitation and emission at 460 nm

4.2.3 Phenotypic analysis of deletion mutants in *fcs1* and *fcs1 fcs2*

To understand more about the route of degradation of different length side chain phenylalkane carboxylic acids compounds in *R. palustris* and the role of specific CoA ligases in the pathway, allelic exchange mutagenesis was used to create unmarked deletion mutations in *fcs1, fcs2, RPA1706* genes, β-oxidation operon and a double mutant in both *fcs1 & fcs2* genes in *R. palustris*. We used a two-step selection methodology that is used to introduce a specific mutation into the gene of interest in the genome of the bacterium under study without leaving any genetic markers (Reyrat et al. 1998). The *fcs1* mutant of *R. palustris* was previously created by R. Salmon and a double *fcs1* and *fcs2* mutant was constructed using this strain of *R. palustris* rather than a wild type *R. palustris*, that was used to create an *fcs2* mutant, β-oxidation operon and *rpa1766* mutant.

To achieve this, DNA from the homologous upstream and downstream regions of each gene missing the sequence of the gene itself was ligated into *pK18mobsacB* vector. The primers were designed to amplify a fragment of 500bp upstream and down stream of gene including 8 bp from the coding region.

Fcs2 primers were digested by *EcoRI, XbaI* and *HindIII*, the vector was digested by *EcoRI* and *XbaI* and all fragments as well as the vector were ligated into the *EcoRI* and *HindIII* site of the *pK18mobsacB* vector in a triple ligation reaction. The vector contains a kanamycin resistance gene and *sacB* gene. The kanamycin marker will select for the incorporation of the recombinant plasmid into the *fcs1* mutant *R. palustris* genome through a single recombination event, while the *sacB* gene selects for the second recombination event that excises the plasmid from the genome. The vector was introduced into *fcs1* mutant *R. palustris* by conjugation using *E. coli* S17-1

strain and the cells were grown on mineral media with the presence of 70 µg/ml kanamycin to select for the integration of the entire vector containing the mutated gene into the genome via a single cross over event. Note that as the vector carries the *sacB* gene (encoding levansucrase which makes toxic products from sucrose), those cells that successfully complete a single crossover would not grow on media that contain 10 % (w/v) sucrose. The second step involves growing the colonies on media with 10% (w/v) sucrose only to allow the reverse of the integration event to occur. In this step cells are viable only when the plasmid "loops out" (removing the *sacB* gene) and leaves the genome by allowing homologous recombination between the directly repeated sequences in the mutated gene inserted in the vector and wild type gene. The resulting colonies were screened by colony PCR using primers designed to target the upstream and downstream region of the gene to be deleted. A positive deletion mutant showed a PCR band missing the correct amount of base pairs targeted by homologous exchange mutagenesis in the vector and that is of size 1000bp for *fcs2* mutant gene and 800bp for *fcs1* mutant gene. The mutant colonies for the double mutant *fcs1* and *fcs2* were confirmed by PCR. Two separate PCR reactions were carried out for the same colony using primers for *fcs1* and *fcs2* and the results showed a lower molecular size of 1000bp and 800bp for the mutated colonies that correspond with gene deletion when compared with wild type strain of *R. palustris* (Figure 4.11).

The same thing was done to make *rpa1766* and *fcs2* mutants but the vector was transformed into wild type *R. palustris*. Unfortunately, several trials to get these mutants resulted in only getting the wild type strain after sucrose selection. Also different sucrose concentrations were used to help the looping out of the plasmid and to create a mutant but all the resulting colonies were wild type.

4.2.4 Phenotypic growth studies

R. palustris has the ability to grow on aromatic compounds such as phenylalkane carboxylic acid compounds either under aerobic or anaerobic conditions and to understand the effect of CoA ligases on the growth of *R. palustris* both wild type growth pattern and mutant growth pattern were compared. *R. palustris* strain CGA009 and its derivative mutants (*fcs1* mutant and *fcs1-fcs2* double mutant) were grown in anaerobic tubes at 30ºC in photosynthetic minimal media (RCV). The media was supplemented with different carbon sources (benzoic acid, coumaric acid, 4-phenylbutyric acid, 5- phenylvaleric acid or 6-phenylhexanoic acid) at a concentration of 3 mM. The growth medium pH was around pH 7.5 and freshly prepared 10 mM sodium bicarbonate were added to the autoclaved RCV media. All added solutions were filter sterilized.

The starting inoculum was obtained from 80 ml culture of RCV supplemented with benzoic acid as the sole carbon source so the cells were benzoic acid adapted. The cells were centrifuged and washed twice with RCV medium lacking any carbon source then the cell were re-suspended in RCV containing the carbon source under investigation. *R. palustris* growth was measured by absorbance at 660 nm using a spectrophotometer, at regular intervals. The anaerobic tubes contained glass balls to re-suspend the culture and to prevent the formation of biofilm. Before incubation in the presence of light all tubes were cultured in the dark for 2 hours at 30ºC to allow the full consumption of oxygen present in the tubes and to prevent any oxidative stress that could occur when cells are grown anaerobically in the presence of light. The resulting growth pattern showed no difference in the growth of Wild type and mutant strains for all substrates tested except for the double mutant, which showed no growth in the presence of coumarate when compared with the wild type strain (Figure

4.12). The growth pattern for mutant versus Wild type suggests that Fcs1 is not the sole CoA–ligase involved in phenylalkane carboxylate degradation and there might be other enzymes that have the capability to utilize the different side chain length of phenylalkane carboxlates as suggested by the RT-PCR and Mass spectrometry results described in Chapter 3.

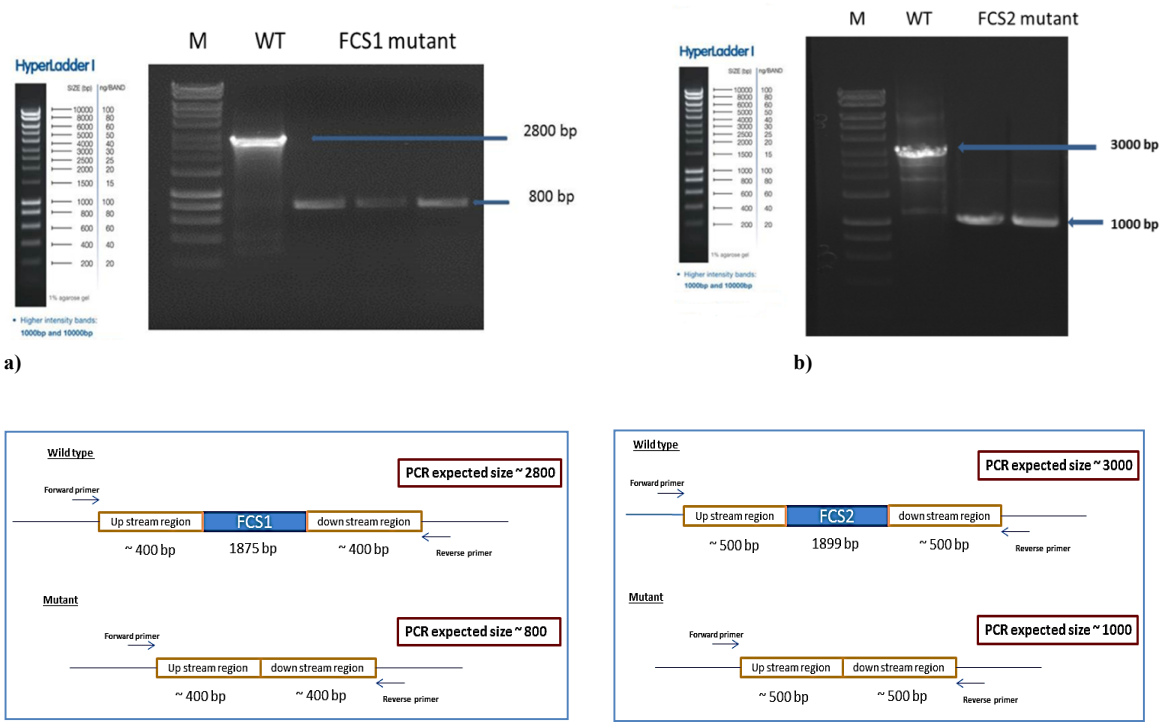


Figure 4.11. PCR amplification for *fcs1* (a) and *fcs2* (b). M= Marker, WT= wild type. For each lane 5 µl of PCR product was loaded onto a 0.7 % agarose gel and electrophoresed for 45 minutes at 110 V. (a) Δ *fcs*1 mutant PCR product at 800 bp compared with the wild type 2800 bp band (b)Δ *fcs2* mutant PCR product at 1000 bp compared with the wild type 3000bp band.

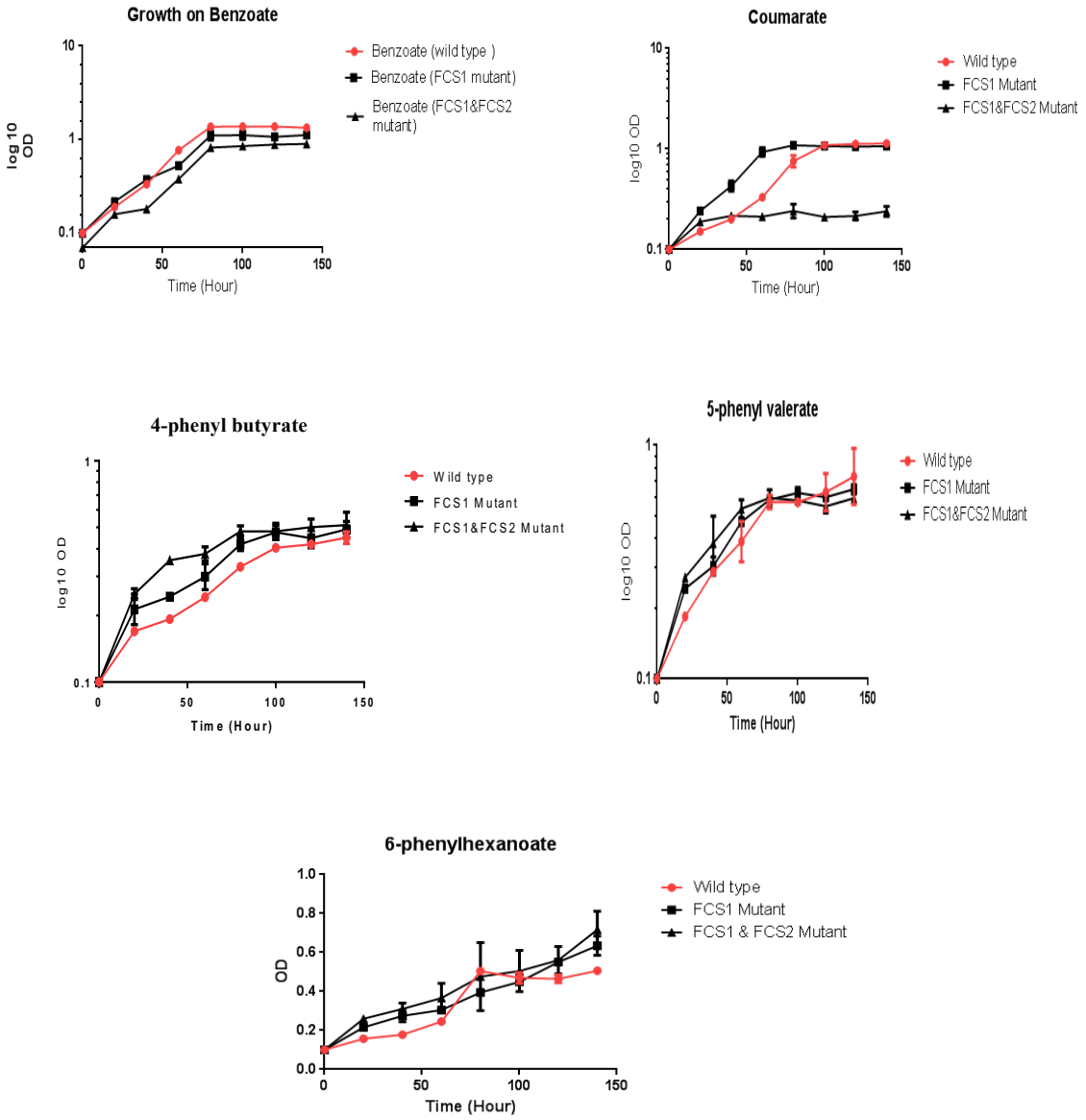


Figure 4.12. Growth of wild-type and mutant strains in RCV media under photoheterotrophic conditions on various aromatic sole carbon sources. Only the double mutant strain grown on coumarate showed a defect in growth.

4.3 Discussion

The data obtained in this chapter have provided evidence that the utilization of different side chain length phenylalkane carboxylates is a complex process and raised

the possibility that more than one CoA ligase could contribute to the first step of their utilization. In the study done by Elder *et al*., (1992) *R. palustris* was shown to grow very well on phenylalkane carboxylic acids with different side chain length and the growth yield was higher on acids with an odd number in their side chain in contrast to the growth on acids with even number in their side chain. The reason for this is the production of phenylacetate by a β-oxidation route of even numbered side chain phenylalkane carboxylic acids which *R. palustris* cannot grow on (Elder et al. 1992). *In-vitro* characterization of Fcs1, Fcs2 and RPA1766 with the different phenyl alkane carboxylates has provided evidence that Fcs1 is the main CoA ligase to convert phenylalkane carboxylates of carbon side chain length C5 to C8 to their CoA thioesters. The β-oxidation pathway seems to be the preferred pathway for degradation of phenylalkane carboxylates as Fcs1 is present in a cluster of genes thought to be associated with such a β-oxidation pathway. A similar mechanism was also suggested for *P. putida* (Garcia et al. 1999;Olivera et al. 2001) where a specific CoA ligase for medium chain length phenylalkane carboxylates was detected when cells were grown in the presence of phenylvalerate and phenylhexanoate. That CoA ligase was different from another ligase that was detected in cells grown on phenylpropanoic acid, cinnamic acid or medium chain length phenylalkanoic acids with an uneven number of carbons. The presence of a CoA ligase specific to the presence of phenylvaleric acid and phenylhexanoic acid suggests that medium chain length phenylalkane carboxylates are utilized in a different manner than the simple phenylpropanoic acid (Ward & O' Connor 2005).

For Fcs1, the substrate with the highest activity was 7-phenylheptanoic acid and the Fcs1 enzyme was able to utilize a variable side chain length of between 5-8 carbons. In addition to the different side chain lengths of phenylalkane carboxylates, straight

chain fatty acids of 6-8 carbons in the side chain were also utilized by Fcs1. Fcs2 was able to utilize a phenylpropenoid of 3 carbons in the side chain only (i.e. phenylpropionate), while RPA1766 was only active with 8-phenyloctanoic acid. Fcs2 substrate specificity was investigated previously by Hirakawa *et al.,* (2012) by monitoring CoA thioester formation using UV spectrophotometer at 400 nm. Phenylcarboxylic acids having side chains of 1-5 carbons and aromatic acids with similar structure were tested and only aromatic acids of side chain length of three carbons were the substrate of Fcs2 (CouB) (Hirakawa et al. 2012). The same results were obtained in this chapter using the NADH linked assay to determine substrate specificity for Fcs2 (Table 4.1). Fcs2 was not only specific to phenylpropanoic acid but it showed higher specificity than Fcs1 when coumaric acid and ferulic acid were tested. These data are consistent with a dedicated and specific role of Fcs2 in the degradation of lignin-derived hydroxycinnamic acids.

The *fcs1* mutant studied here was able to grow on all phenylalkane carboxylate substrates tested which indicated that other CoA ligases can substitute for the Fcs1 CoA ligase *in vivo*, while the double mutant (lacking *fcs1* and *fcs2*) showed a defect in growth on coumaric acid only. A single *fcs2* mutant was shown to be growth defective on coumaric acid by Hirakawa *et al*., (2012) and they also suggested that other CoA ligases may contribute to coumaric acid utilization but at a much slower rate (Hirakawa et al. 2012).

Structural modeling of Fcs1 using Phyre and Pymol software was carried out and compared to the *Streptomyces* malonyl-CoA ligase (MatB) structure to investigate the active site characteristics of Fcs1. First the protein was modeled by Phyre and showed a 19% similarity to 4- coumarate CoA ligase, then the modeled structured was

compared to malonyl-CoA ligase structure which showed also 19% similarity with Fcs1.The results showed an open cavity with no interfering ligands from the Fcs1 protein except an Arginine that binds to the carboxylic group of the substrate and is present at the far end of the cavity, allowing substrates like longer chain phenylalkane carboxylates to bind (Figure 4.13). This correlates with the ability of Fcs1 to bind to long chain phenylalkane carboxylates as demonstrated in this work.

In conclusion, the enzymatic assays described here and combined with what was proposed previously by Elder *et al.,* (1992) supports the idea that the degradation pathway for different chain length phenylalkane carboxylates in *R. palustris* is a β-oxidation pathway. However, the mass spectrometry, q-RTPCR and mutant growth experiments all suggest that different CoA ligases are used by *R. palustris* to utilize these compounds. The compounds seem to be essentially recognized as long chain fatty acids through the β- oxidation pathway while phenylpropanoate is seemingly metabolized through the non-β- oxidation pathway via CouB (Fcs2). Phenylpropanoic acid degradation through a non β-oxidation pathway was reported previously for *Pseudomonas flourescens* AN103, *Pedomonas* sp.HR199 and *Delftia acidovorans* (Gasson et al. 1998; Overhage et al. 1999; Plaggenborg et al. 2001). As a future goal, the full set of genes near the Fcs1 gene which are thought to be involved in a potential aromatic β-oxidation pathway could be investigated, their role in medium length phenylalkane carboxylate degradation could be studied to determine which pathway is responsible for the degradation of these compounds in more detail.

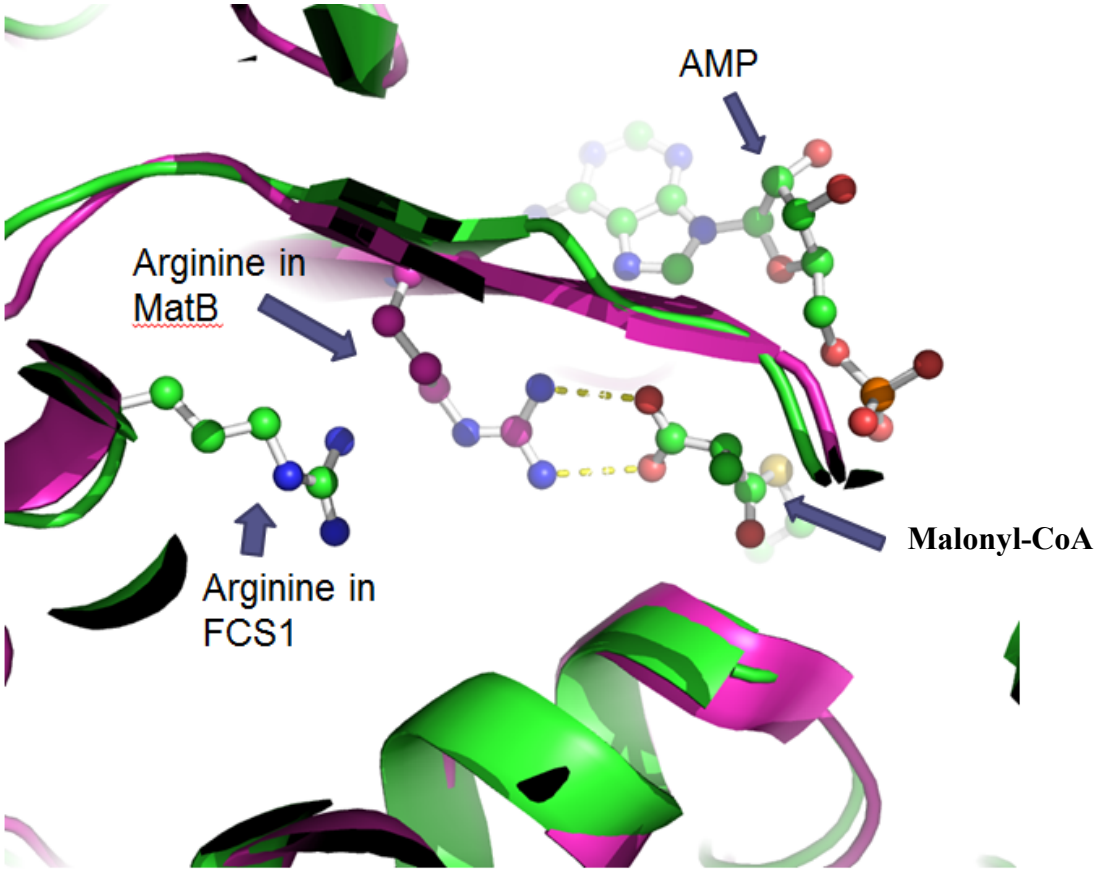


Figure 4.13. Active site modelling of Fcs1. Superimposed structure of Fcs1 and MatB; the Fcs1 is coloured green while MatB is coloured purple. The malonyl CoA carboxylic group binds to an Arginine in the active site. The position of arginine in Fcs1 could indicate why it can bind to longer chain phenylalkane carboxylates.

5. Characterization of periplasmic solute-binding proteins involved in the uptake of phenylalkane carboxylates

Characterization of periplasmic solute-binding proteins involved in the uptake of phenylalkane carboxylates

5.1 Introduction

R. palustris has the ability to grow on a wide range of aromatic compounds that it obtains from its surrounding environment (Larimer et al. 2004). For metabolism of these compounds, they must first be transported into the cell. While it is known that *R. palustris* has a large variety of solute transport systems encoded in its genome, the physiological role of many of these is still unknown. However, recent studies have identified several transport systems for aromatic compounds, particularly of the ATP-binding cassette (ABC) and tri-partite ATP-independent periplasmic (TRAP) families, which use periplasmic solute binding proteins as part of the transport mechanism (Giuliani et al. 2011; Salmon et al. 2013).

The solute binding proteins (SBPs) associated with such transport systems are the first interface that deals with those compounds, so they play a major role in regulating cellular catabolic capability as they mediate the movement of compounds across the cell membrane. The SBP structure usually shows two globular domains which are connected by a hinge region, this structure allows flexibility for the ligands to bind in the cleft and for the protein to close around the ligand (Davidson et al. 2008).

From the mass spectrometry results in Chapter 3, the protein RPA3720 was significantly more abundant when cells were grown on 5-phenylvaleric acid compared to benzoate. RPA3720 is annotated as IlvF, a putative branched-chain amino acid transport system ATP–binding protein and it is part of an ABC transport system thought to be associated with the *pimFABCDF* operon that encodes a β-

oxidation pathway for pimelate consisting of an acyl-CoA ligase, enoyl-CoA hydratase, acyl-CoA dehydrogenase, and acyl-CoA transferase enzymes. This operon allows the conversion of an odd or even chain dicarboxylic acid to glutaryl-CoA or succinyl-CoA respectively (Harrison & Harwood 2005). These enzymes are thought to be mainly involved in the degradation of aromatic compounds after the benzene ring has been reduced and opened, resulting in the formation of pimelate as an intermediate. Although the other proteins of this system were either not detected or unchanged in our proteomics analysis, we hypothesized that the increased abundance of RP3720 might suggest a possible role in phenylalkane carboxylate transport. In order to investigate this, the ligand-binding properties of the three solute binding proteins (RPA3723, RPA3724 and RPA3725) associated with this transport system were chosen to be screened and analyzed (Figure 5.1).

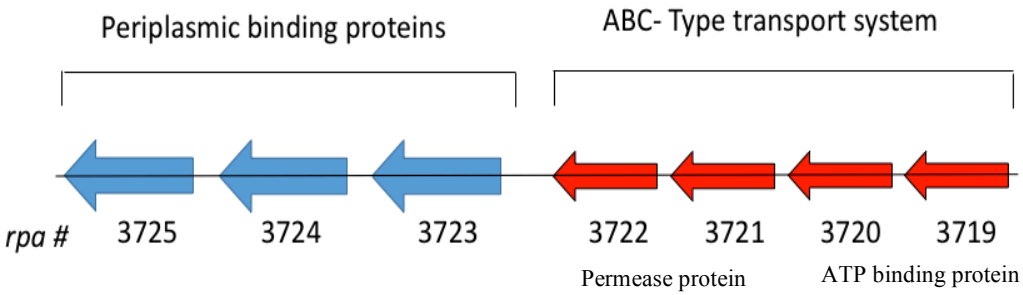


Figure 5.1. Organization of ABC transporter system with periplasmic solute binding proteins that is associated with *pimFABCDE* operon.

Using a thermal shift assay, Giuliani *et al*. (2011) studied the different solute binding proteins associated with ABC transporters in *R. palustris* and they included RPA3723, RPA3724 and RPA3725. The ligand binding profile for RPA3723 and RPA3724 showed their ability to bind to straight chain fatty acids of more than seven carbons and some dicarboxylic acids, while RPA3725 showed no ligand binding activity

(Giuliani et al. 2011). Both the former proteins were able to bind to azelaic, sebacic, undecanoic, tetradecanoic, caprylic, capric, lauric, and myristic acids but only RPA3724 was able to bind to pimelic and palmitic acids. The shift in the temperature of denaturation in the presence of ligand was around three to four degrees for octanoic acid (caprylic) for both proteins and the highest shift of four to six degree was for tetradecanoate. The presence of more than one SBP with different and wide substrate ranges could reflect the competitive environment of soil where naturally diverse nutrient acquisition is essential for microbial growth. The production of a ligand library for studying SBPs could give an indication of how metabolically diverse an organism is.

This chapter investigates the ability of the different length side chain phenylalkane carboxylic acids to bind to the periplasmic binding proteins RPA3723, RPA3724 and RPA3725 from the Pim-associated ABC transport system. The proteins were first overexpressed to assess their role in binding to different length side chain phenylalkane carboxylic acids. Then, a thermofluor screen of compounds under investigation showed a number of ligands to bind to RPA3723 and RPA3724. Data from intrinsic tryptophan fluorescence changes for RPA3723 showed that it has a binding affinity in the micromolar range for 4-phenylbutyric acid, 5-phenylvaleric acid, 6-phenylhexanoic acid, 7-phenylheptanoic acid and 8-phenyloctanoic acid. The highest $K_{\mathbf{d}}$ value was for 7-phenylheptanoic acid while the weakest binding affinity was for 4-phenylbutyric acid. In addition to the different length side chain phenylalkane carboxylic acids, straight chain fatty acids of 3 to 8 carbon chain length were tested and the binding affinity of octanoic acid was the highest. Nuclear magnetic resonance spectroscopy (NMR) was used to investigate ligand binding to RPA3724 because the fluorescence spectroscopy showed no shift in the tryptophan

florescence signal when different ligands were added. RPA3724 showed higher affinity for 8-phenyloctanoate binding.

5.2 Results

5.2.1 Cloning and Overproduction of RPA3723, RPA3724 and RPA 3725

The *rpa3725, rpa3724* and *rpa3723* genes were amplified from *R. palustris* genomic DNA using primers pET21_3725 F + R for *rpa3725*, primers pET21_3724 F + R for *rpa3724* and primers pET21_3723 F + R for *rpa3723*. All primers incorporated *NdeI* and *Xho*I sites and the products were cloned into *NdeI* and *XhoI* sites of digested pET21a (+) plasmid (Figure 5.2 a, 5.3 a and 5.4 a); the expression from pET21a (+) vector was induced by 400 μM IPTG via the T7 promoter. Colony PCR and sequencing of the constructs were done to confirm cloning and to ensure that there is no mutation in the genes under study (Core Genomic Facility, University of Sheffield Medical School, UK).

Overexpression trials for RPA3725, RPA3724 and RPA3723 were carried out in *E. coli* BL21 cells at 37 °C and 25 °C. Protein expression was monitored at regular time intervals of 1, 3, 5 and 24 hours after cells of 0.6 $OD_{(600)}$ were induced with 400 μM IPTG; a whole cell sample for each time point was taken and run on an SDS-PAGE gel to assess the best expression point (Figure 5.2b, 5.3b and 5.4b). The best expression time was 5 hours at 37 °C post induction for RPA3723, 5 hours at 25 °C post induction for RPA3724 and 5 hours at 37 °C post induction for RPA3725. This was the growth condition used to overproduce large amounts of recombinant proteins for purification. The protein solubility was also checked after sonication by running the sample on an SDS-PAGE gel.

Protein purification was carried out by passing overexpressed protein cell free extract through a nickel affinity chromatography column. 1 ml fractions were collected through different imidazole gradients that passed through the column and fractions collected were proven to contain the protein after visualization on an SDS-page gel (Figure 5.2 c, 5.3 c and 5.4 c). The concentration of purified proteins was measured by UV spectrophotometry and results for the concentration was 51 μM for RPA3723, 68 μM for RPA3724 and 92 μM for RPA3725. The purified proteins were stored at 4C° for subsequent assays.

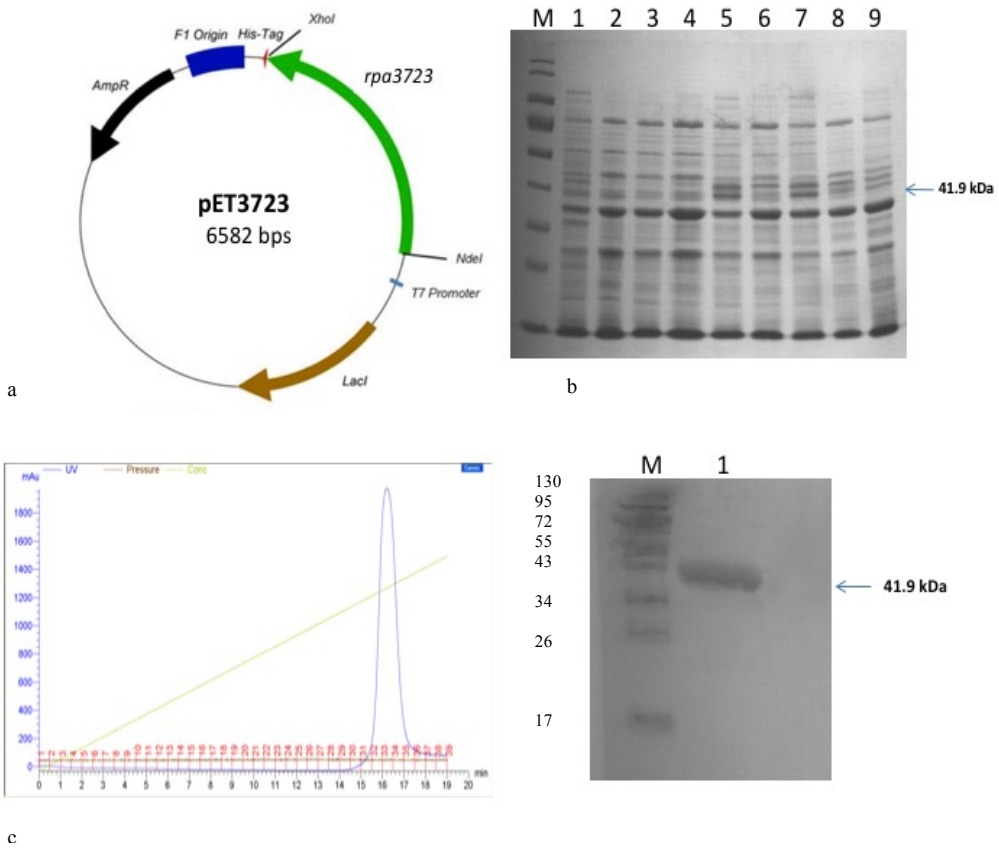


Figure 5.2. Overexpression and purification of RPA3723. (a) overexpression construct pET3723 cloned into *NdeI* and *XhoI* site and the stop codon was removed from the gene to add the 6X His-tag to the C-terminal of the protein. (b) SDS-PAGE showing the protein expression trials. M= PageRulerTM Prestained Protein Ladder (Fermentas). Lane 1,3,5,7 contain sample from cells incubated at 37C°, lane 1 is a non –induced control, Lane 3,5,7 contain sample taken at 1,3 and 5 hours after IPTG induction respectively; Lanes 2,4,6 and 8 contain sample incubated at 25C° and taken at 1,3,5 and 24 hours. Lane 7 contains a good amount of protein at 41.9 kDa. (c) UV spectrum of purified protein using Akta His-trap columns and SDS- PAGE showing the purified fraction.

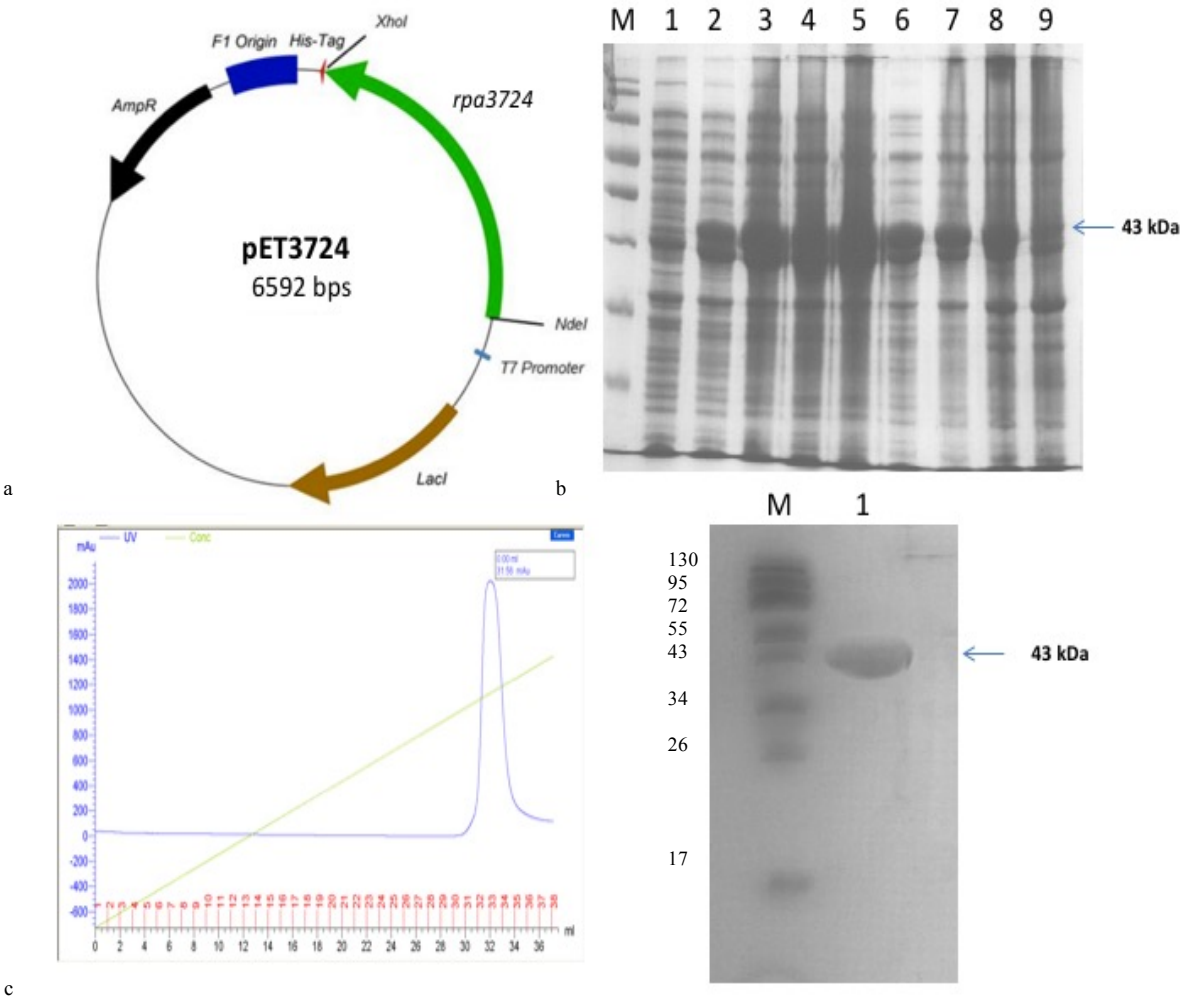


Figure 5.3. Overexpression and purification of RPA3724. (a) overexpression construct pET3723 cloned into *NdeI* and *XhoI* site and the stop codon was removed from the gene to add the 6X His-tag to the C-terminal of the protein. (b) SDS-PAGE showing the protein expression trials. M= PageRulerTM Prestained Protein Ladder (Fermentas). Lane 1,3,5,7 contain sample from cells incubated at 37C°, lane 1 is a non –induced control, Lane 3,5,7 contain sample taken at 1,3 and 5 hours after IPTG induction respectively; Lanes 2,4,6 and 8 contain sample incubated at 25C° and taken at 1,3,5 and 24 hours. Lane 6 contain a good amount of protein at 43 kDa. (c) UV spectrum of purified protein using Akta His-trap columns and SDS- PAGE showing the purified fraction.

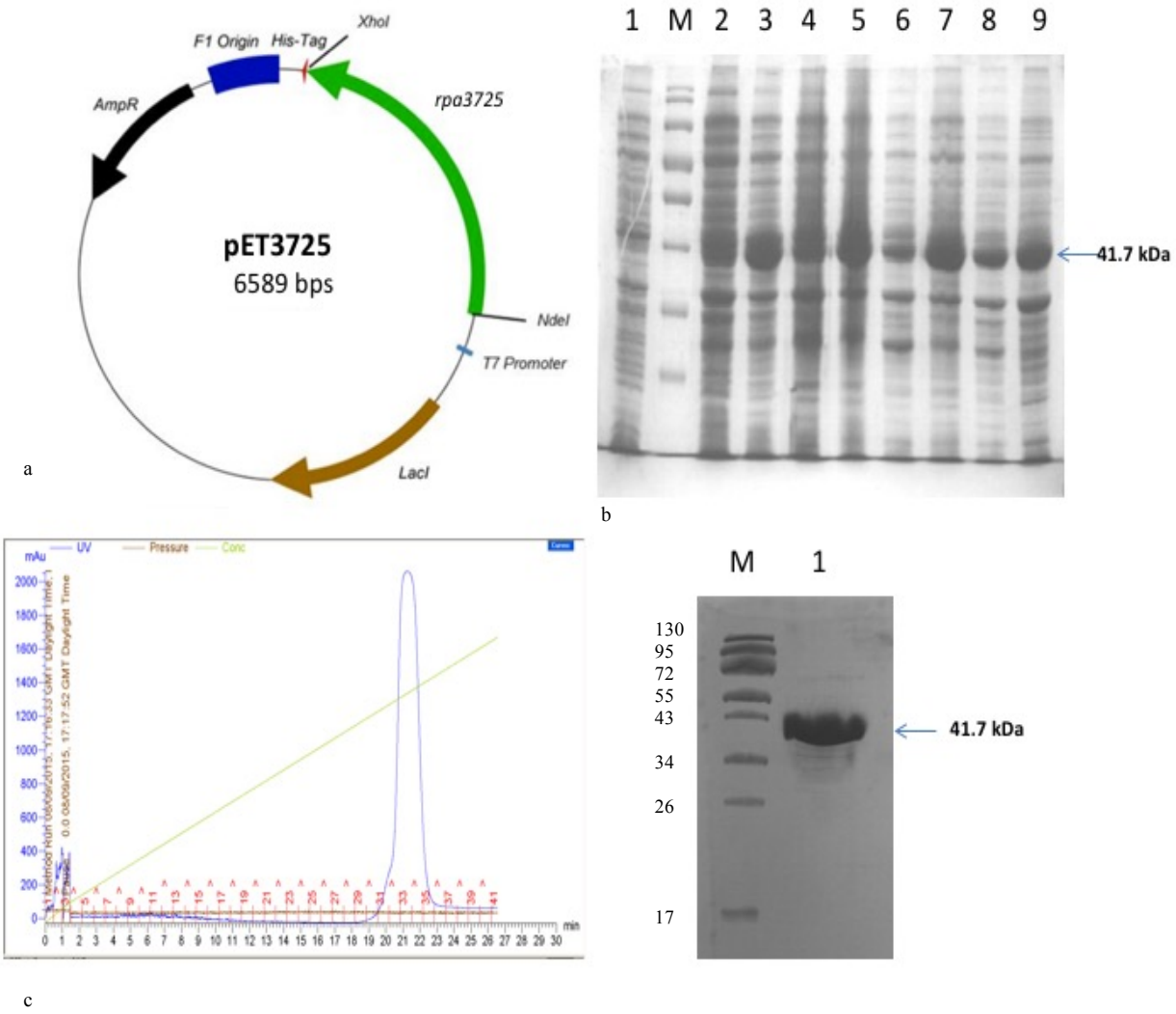


Figure 5.4. Overexpression and purification of RPA3725. (a) overexpression construct pET3725 cloned into *NdeI* and *XhoI* site and the stop codon was removed from the gene to add the 6X His-tag to the C-terminal of the protein. (b) SDS-PAGE showing the protein expression trials. M= PageRulerTM Prestained Protein Ladder (Fermentas). Lane 1,3,5,7 contain sample from cells incubated at 37C°, lane 1 is a non –induced control, Lane 3,5,7 contain sample taken at 1,3 and 5 hours after IPTG induction respectively; Lanes 2,4,6 and 8 contain sample incubated at 25C° and taken at 1,3,5 and 24 hours. Lane 7 contain a good amount of protein at 41.7 kDa. (c) UV spectrum of purified protein using Akta His-trap columns and SDS- PAGE showing the purified fraction eluted from the His-trap on an imidazole gradient.

5.2.2 Thermofluor ligand screening of RPA3723, RPA3724 and RPA3725

The RPA3723, RPA3724 and RPA3725 solute binding proteins are associated with an ABC transport system that is up-regulated when *R. palustris* was grown on 5-pheylvaleric acid so the search for ligands included different length side chain phenylalkane carboxylic acids compounds ranging from 3 carbons to 8 carbons. *R. palustris* has the ability to grow and utilize different length side chain phenylalkane carboxylic acids and thermofluor screens were used to initially identify the ligands that can bind to the purified proteins. In the thermofluor assay the protein thermal denaturation profile in the presence and absence of probable ligands is compared; the protein will be more stable when it is bound to a ligand and hence the temperature for denaturation will be higher than un-bound protein or protein without a ligand. The unfolding of the protein is monitored using a dye that fluoresces when it binds to the interior hydrophobic regions of the protein (Figure. 5.5).

To optimize the method, different concentration of protein in the range of 5 μM to 20 μM were trialed; 10 μM of all proteins tested were eventually used in the assay to obtain a good signal from the fluorescent reporter dye (SPYRO orange dye, Invitrogen). 60 μM final concentration of ligands was added in order to achieve saturation in the event of binding. The full list of ligands tested and the resulting thermal shifts for RPA3723 and RPA3724 is listed in Table 5.1 and selected results shown in Figures 5.5 and 5.6.

Ligand tested	ΔTm for RPA3723	ΔTm for RPA3724	ΔTm for RPA3725
3- phenylpropanoic acid	1.1 ± 0.04	0.6 ± 0.12	1± 0.7
4- phenylbutyric acid	**3.1 ± 0.16**	**2.5 ± 0.16**	0.9 ± 0.55
5-phenylvaleric acid	**4 ± 0.355**	1 ± 0.35	1.1± 0.22
6-phenylhexanoic acid	**4.5 ± 0.124**	1.1 ± 0.30	1.5± 0.72
7-phenylheptanoic acid	**6.2 ± 0.24**	**4 ± 0.35**	0.4±0.94
8-phenyloctanoic acid	**5.1 ± 0.047**	1.1 ± 0.17	0.9±0.140

Table 5.1: Thermal shifts produced by each ligand screened against RPA3723 RPA3724and RPA3725. ΔTm shifts were calculated by Boltzmann equation for three replicates and the ligands highlighted in bold showed the ligands with a shift of more than 2C°.

The results showed that RPA3723 binds the best to the tested ligands of carbon chain more than 4 carbons and 7-phenylheptanoic acid showed the highest shift in temperature of ΔTm 6.6 C° (Figure 5.5). RPA3724 was only able to bind to 4-phenylbutrate with ΔTm 2.5C° and 7-phenylheptanoate with ΔTm 4C° (Figure 5.6) while RPA3725 did not bind to any tested ligands. Both RPA 3724 and RPA3725 were unfolded by urea treatment and then refolded by refolding buffer dialysis. This was done to remove any endogenously bound ligands that may be present but even so both proteins gave the same results even after urea treatment. Thermofluor assay will only give us an indication of binding but it will not produce quantitative kinetics, so ligands that showed binding to RPA3723 and RPA3724 were analysed by different assays to assess their binding affinity to each protein.

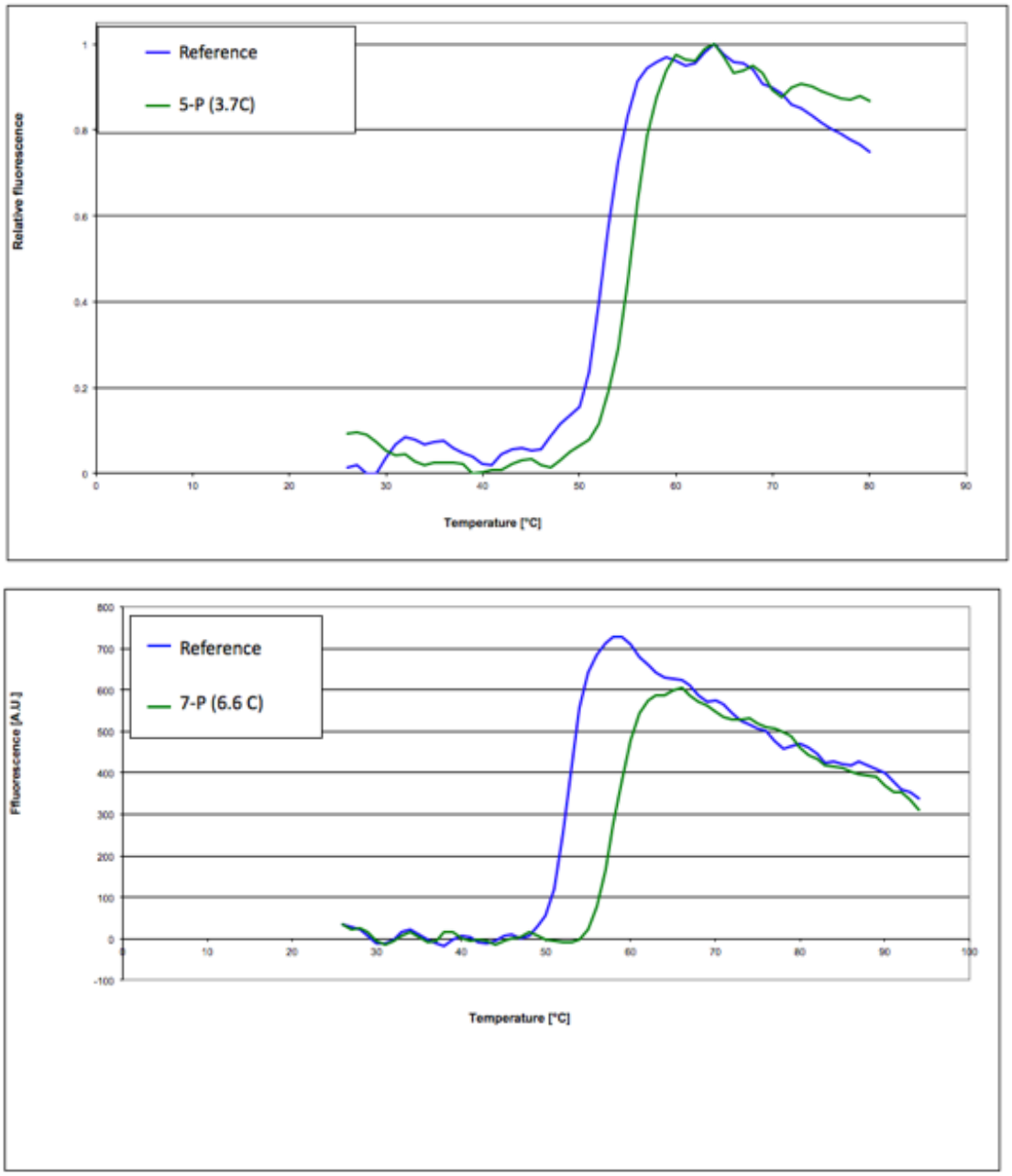


Figure 5.5: Thermal denaturation profiles of RPA3723 with 5-phenylvaleric acid and 7-phenylheptanoic acid. Thermal denaturation profile of two ligands to represent the ligands screened. For each 50 µl reaction 10 µM purified protein with 10 mM Tris-HCL pH 7.4 and 10X SPRYO dye (Invitrogen) was analysed. (a) Represents the thermal shift of RPA3723 with 5-phenylvaleric acid in green and a reference in blue containing only RPA3723. (b) Represents the thermal shift of RPA3723 with 7-phenylheptanoic acid in green and a reference in blue containing only RPA3723.

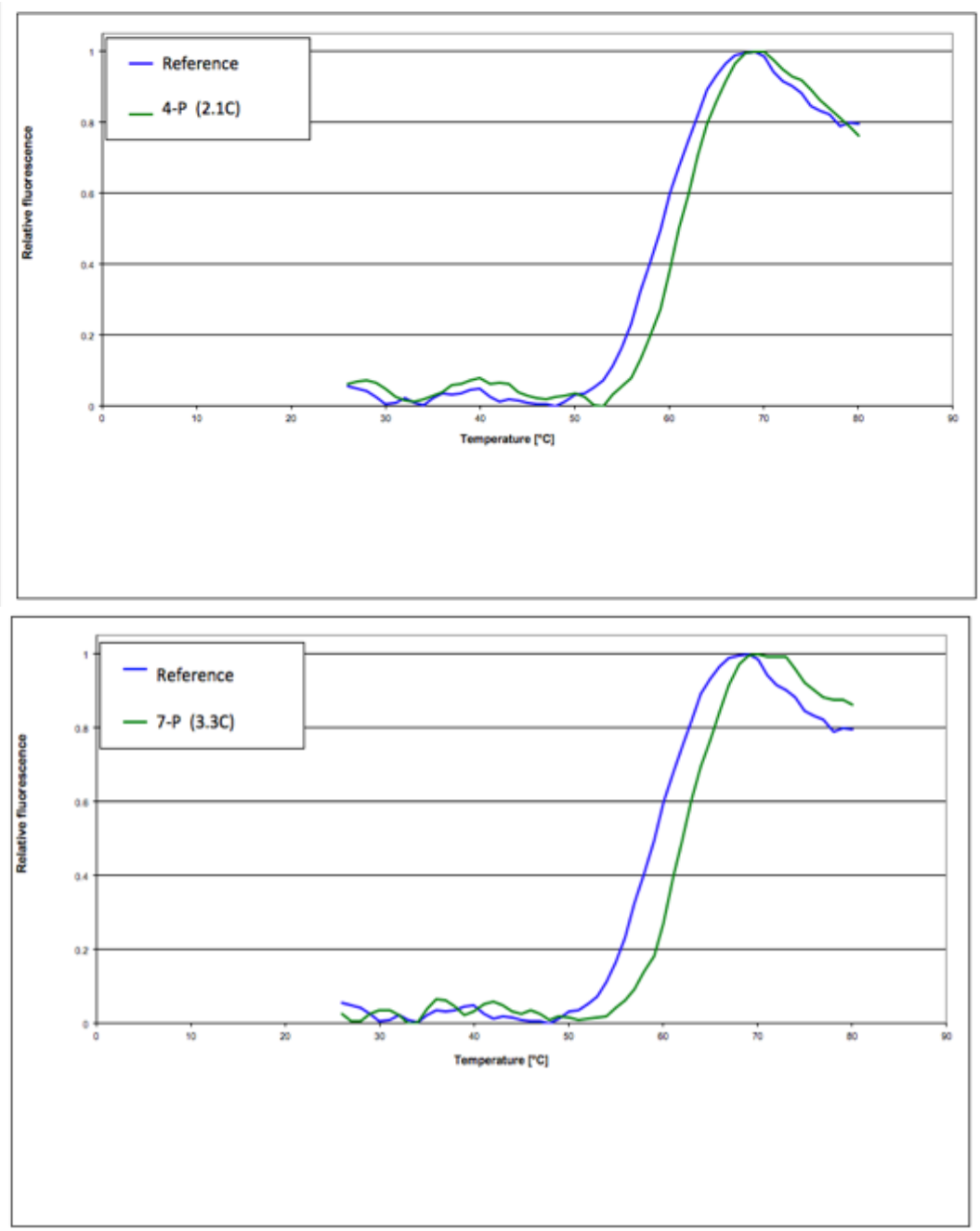


Figure 5.6: Thermal denaturation profiles of RPA3724 with 4-phenybutyric acid and 7-phenylheptanoic acid. Thermal denaturation profile of two ligands to represent the ligands screened. For each 50 µl reaction of 10 µM purified protein with 10 mM Tris-HCL pH 7.4 and 10X SPRYO dye (Invitrogen) was analysed. (a) Represents the thermal shift of RPA3723 with 4-phenylbutyric acid in green and a reference in blue containing only RPA3724. (b) Represents the thermal shift of RPA3724 with 7-phenylheptanoic acid in green and a reference in blue containing only RPA3724.

5.2.3 Tryptophan fluorescence spectroscopy for RPA3723 and RPA3724

To have more definite evidence about the ligand binding activity of RPA3723 and RPA3724 to different length side chain phenylalkane carboxylic acids, further evaluation of binding by using tryptophan fluorescence spectroscopy was carried out.

5.2.3.1 Fluorescence spectra

To determine the kinetics of solute binding protein of RPA3723 and RPA3724 to different chain length fatty acids and phenylalkane carboxylic acids, excitation of both solute binding proteins was done at 280 nm while emission was at a maximum at 332 nm (Figure. 5.7). This intrinsic tryptophan fluorescence was recorded to measure ligand induced conformational changes to the protein. The addition of 3 μM of ligand to 0.2 μM of RPA3723 resulted in a very small enhancement with all ligands tested and the percentage was ~ 5 % (Figure 5.7). The fluorescence obtained after the addition of ligands was higher than the fluorescence obtained with the protein alone, which is an indication that the protein can bind to the ligands under study. Ligands that showed an enhancement with RPA3723 were 4-phenylbutyric acid, 5-phenylvaleric acid, 6-phenylhexanoic acid, 7-phenylheptanoic acid, 8-phenyloctanoic acid, hexanoic acid, heptanoic acid and octanoic acid. On the other hand, RPA3724 showed no change in fluorescence emission. The addition of 0.2 μM concentration of ligands did not produce any change in fluorescence so a higher concentration was used which may indicate a weak interaction between protein and the ligand. To evaluate the binding affinity of RPA3723 to the different ligands tested fluorescence titrations were carried out.

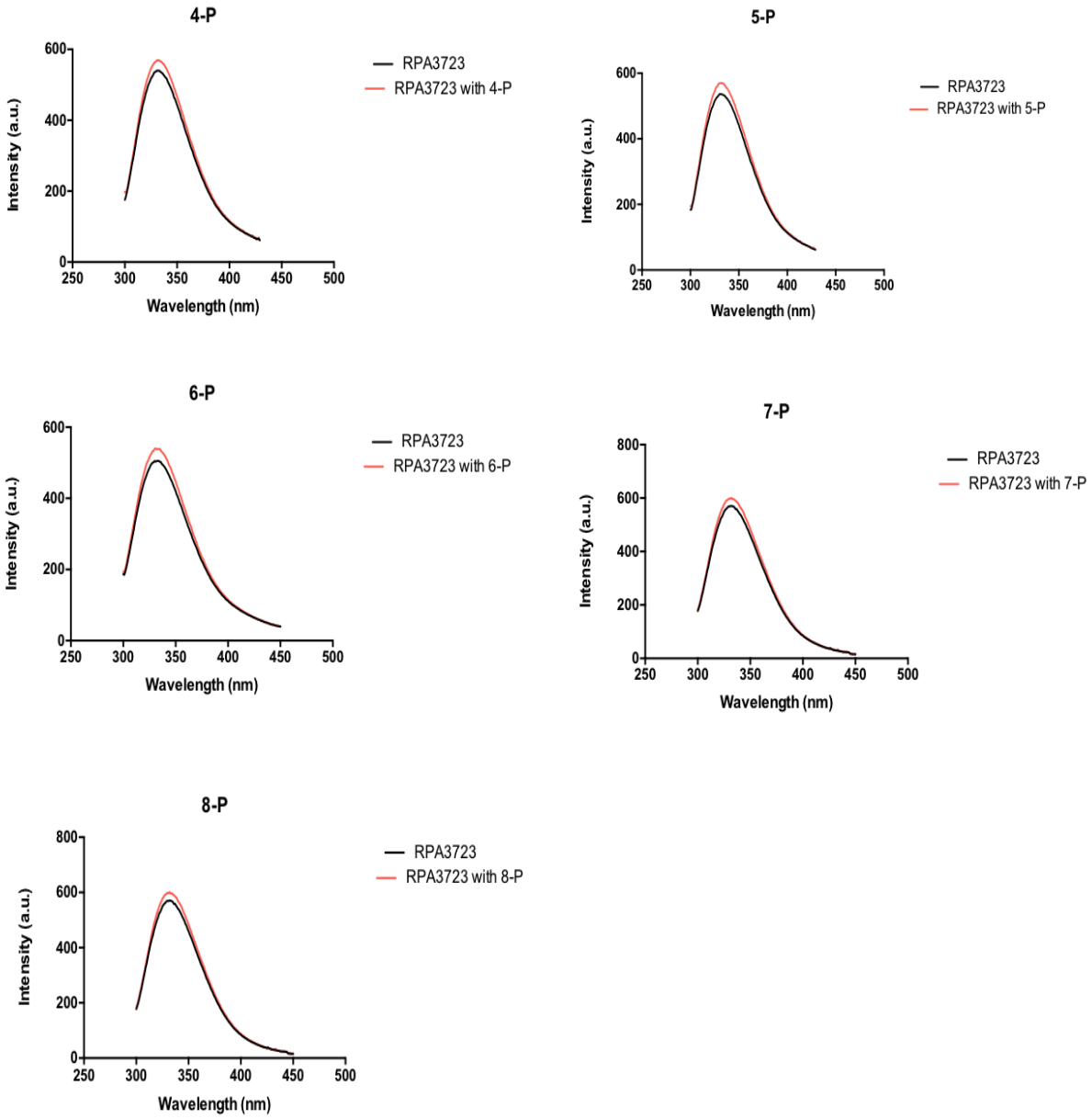


Figure 5.7: Changes in intrinsic tryptophan fluorescence of RPA3723 with and without 8-phenyloctanoic acid. 0.2 μM Purified RPA3723 was excited at 280 nm and the emission spectrum in black represents the purified protein. After the addition of 3 μM of different length side chain phenylalkane carboxylic acids ligand the emission spectrum in red represents protein binding.4-phenylbutyric acid induced 5.09% enhancement. 5-phenylvaleric acid induced 5.7% enhancement. 6-phenylhexanoic acid induced 6.4% enhancement. 7-phenylheptanoic acid induced 4.8% enhancement and 8-phenyloctanoic acid induced a 4% enhancement.

5.2.3.2 Kinetic analysis using ligand titration

The change in fluorescence emission due to ligand titration with RPA3723 produced a consistent fluorescence change as shown in Figure 5.7, which allowed for quantitative analysis for the kinetics of RPA3723 ligand-binding (Table 5.2).

Ligands	**K_d value (μM)**
4- phenylburtyric acid	17.8 ± 0.95
5-phenylvaleric acid	12.2 ± 0.56
6-phenylhexanoic acid	8.4 ± 0.28
7-phenylheptanoic acid	1.29 ± 0.1
8-phenyloctanoic acid	5 ± 0.25
Hexanoic acid	85 ± 2.9
Heptanoic acid	23.11± 0.44
Octanoic acid	4.8 ± 0.1

Table 5.2 calculated K_d value from fluorescence titration with RPA3723. Values shown are averages and errors of three independent titrations.

Dynafit software was used to fit the data obtained to a single site binding model with the majority of points fitting the curve (Kuzmic, 1996) to visualize the binding pattern for each ligand (Figure 5.8 and 5.9). For 6-phenylhexanoic acid, 7-phenylheptanoic acid and 8-phenyloctanoic acid the shape of the plot indicated a binding of k_d value 8.4, 1.29 and 5 uM respectively (Figure 5.8). Both 4-phenylbutyric acid and 5-phenylvaleric acid showed weaker binding affinity of k_d value 17.8 and 12.2 uM. The titration with straight-chain fatty acids showed a k_d value of 85 uM for hexanoic acid, 23 uM for heptanoic acid and 4.8 uM for octanoic acid (Figure 5.9). 7-phenylheptanoic acid had the lowest k_d value of binding but the values were still in the μM range. The pattern of binding seems to reach its maximum with a phenlyalkane carboxylate of side chain length of 7 carbons and then the affinity decreased with a side chain length of 8 carbons. The binding of straight chain fatty

acids and different length side chain phenylalkane carboxylic acids indicated that the possession of the side chain is essential for RPA3723 protein binding.

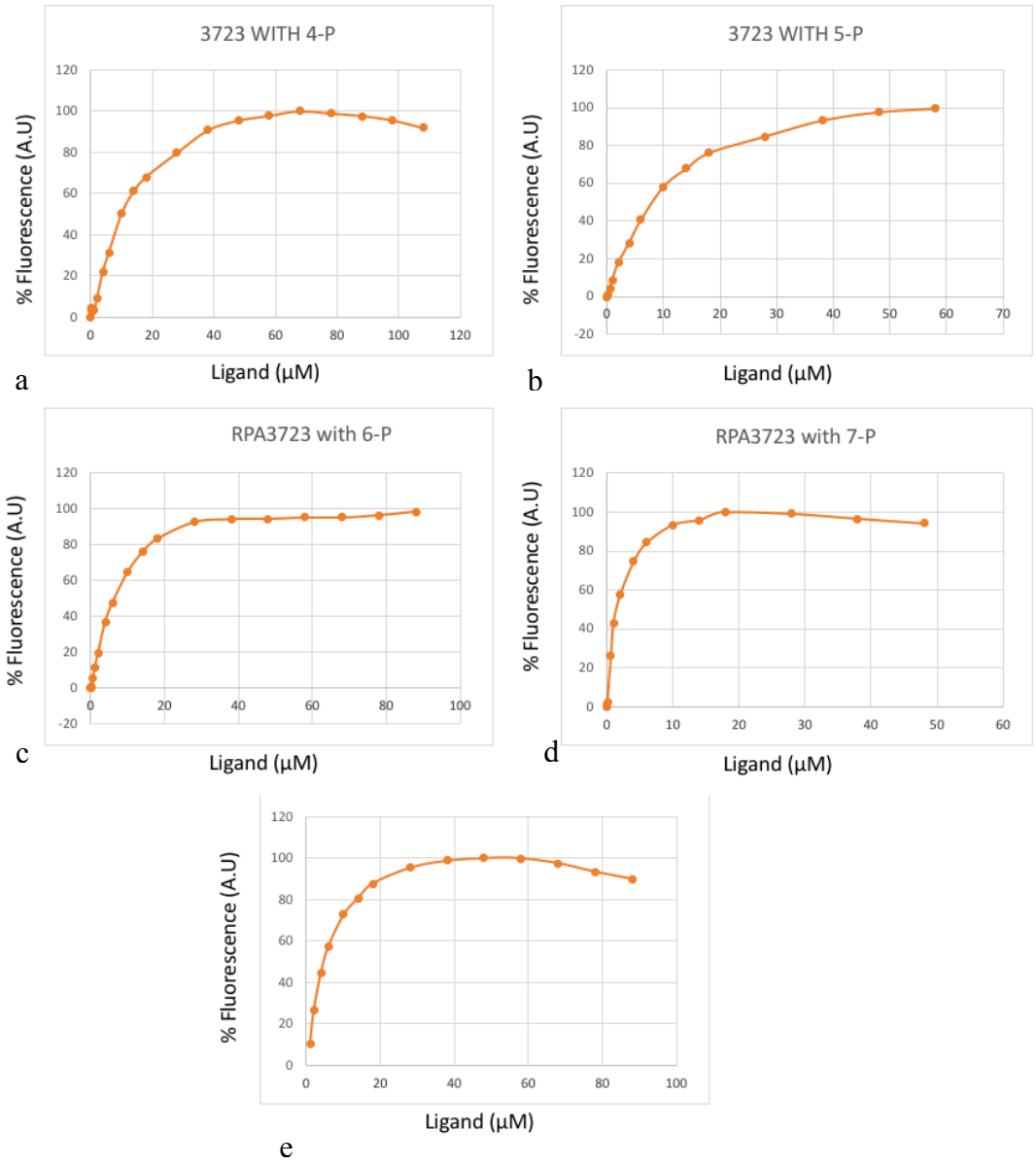


Figure 5.8: Fluorescence titration of RPA3723 with different length side chain phenylalkane carboxylic acids ligand. 0.2 µM of RPA3723 fluorescence emission was measured at 332 nm during titrations with different ligands (a) 4-phenylbutyric acid (b) 5-phenylvaleric acid (c) 6-phenylhexanoic acid (d) 7-phenylheptanoic acid (e) 8-phenyloctanoic acid. Data showed were fitted to single- site binding model and dissociation curve were plotted. The protein showed high affinity binding with 7-phenylheptanoic acid compared to 4-phenylbutyric acid.

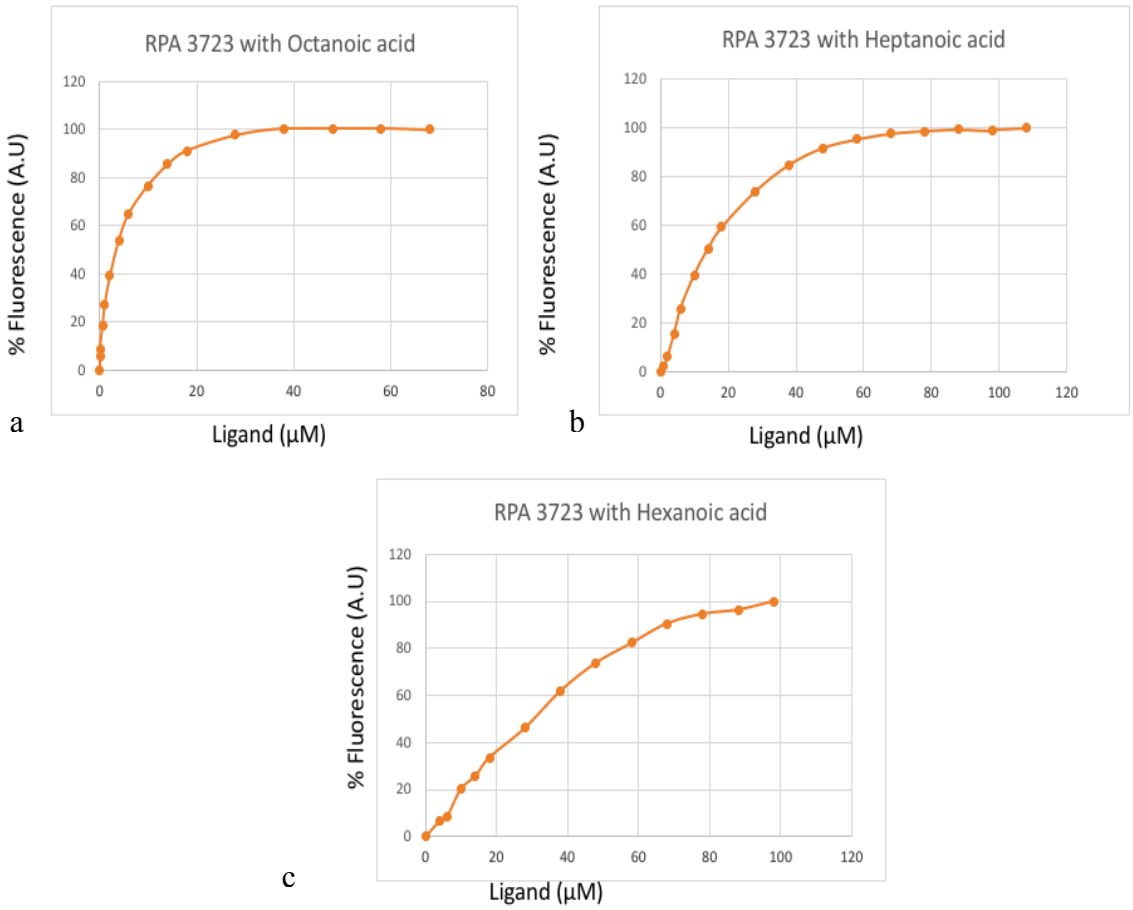


Figure 5.9: Fluorescence titration of RPA3723 with different Fatty acids. 0.2 µM of RPA3723 fluorescence emission was measured at 332 nm during titrations with different ligands (a) octanoic acid (b) heptanoic acid (c) hexanoic acid. Data showed were fitted to single- site binding model and dissociation curve were plotted where the protein had high affinity binding to octanoic acid compared to hexanoic acid.

5.2.4 Nuclear magnetic resonance (NMR) data for RPA3724 interacting with different length side chain phenylalkane carboxylic acids.

1D ^{1}H NMR Spectroscopy was used to determine the binding of RPA3724 to different ligands that showed thermal shifts because in the tryptophan fluorescence assay RPA3724 did not show any fluorescence change when different ligands were added. The chemical shift in proton NMR experiments measures changes in the properties of the hydrogen nuclei in the ligand as the marker of the binding, so the chemical shift and line width of the signals in the ligand proton spectrum can be compared in the presence and absence of the protein. If an interaction occurs, the general effect of the protein on the ligand will be to cause a broadening and flattening of the peaks in the ligand spectrum.

All reactions were carried out in 1:1 ratio, where the concentration of protein was 50 μM and the final ligand concentration was 50 μM. The reactions were carried out in 50 mM sodium phosphate buffer. A decrease in the height or width of the ligand NMR signals in the presence of the protein was observed and the highest shift was recorded with 8-phenyloctanoate followed by 7-phenylheptanoic acid (Figure 5.10). The binding pattern with the different ligand tested showed similar results as that of SBP RPA3723 which bind to the longer chained phenylalkane carboxylic acids (5-8 carbon side chain) while the shorter side chains showed week or no binding at all. Unfortunately, it was hard to calculate the affinity as the shift is too small to create a titration. However, the NMR provides a detailed and more accurate binding profile for RPA3724.

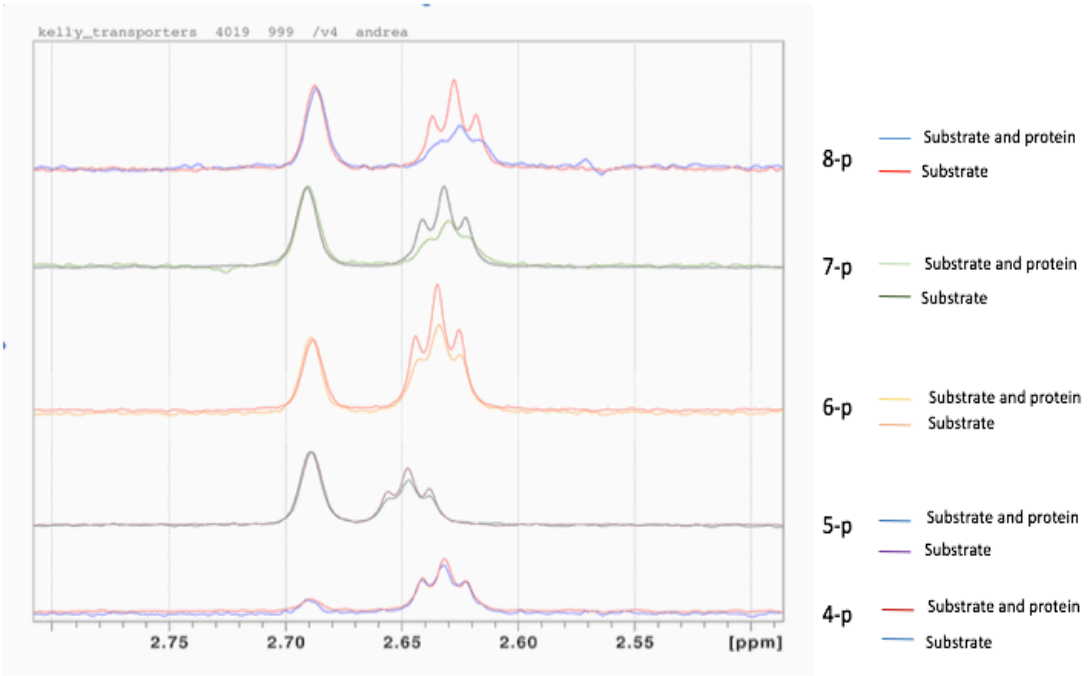


Figure 5.10. 1D ^{1}H NMR Spectrometry for RPA3724 with different length side chain phenylalkane carboxylic acid ligands. The spectrum of 50 µM ligand were used as a control and labelled as substrate compared with ligand and protein spectrum in the ratio of 1:1. The widening and shortening of the peak resemble the binding of RPA3724 with the ligand, the highest shift was recorded with 8-phenyloctanoic acid.

5.3 Discussion

This chapter provides strong evidence that the ABC transport system associated with the *pimABCDF* operon could also transport different length side chain phenylalkane carboxylic acids in addition to substrates like pimelate and other straight chain acids. Both RPA3723 and RPA3724 were shown here to be soluble binding proteins that bind to the different ligands with different affinities. Characterization of RPA3723 and RPA3724 by both thermofluor and NMR has confirmed the binding of ligands and the fluorescence spectroscopy of RPA3723 showed the binding affinity of the protein with different ligands including straight chain fatty acids and the different

length side chain phenylalkane carboxylic acids with a K_d value ranging from 85 uM to 4.8 uM for straight chain fatty acids tested and 17.9 uM to 1.29 uM for different length side chain phenylalkane carboxylic acids tested (Table 5.2). These results indicated that the solute binding protein RPA3723 will bind to the functional group of different length side chain phenylalkane carboxylic acids and perhaps will consider the aromatic ring as if it were a longer chain of a fatty acid. The difference in ligand binding is not unexpected as the amino-acid similarity between the three soluble binding protein is around 58% for RPA3723 and RPA3724. On the other hand both proteins have 61% and 71% sequence similarity with RPA3725, which showed no binding with any ligand tested. Variation of sequence similarity is a common characteristic for ABC SBPs (Berntsson et al. 2010b) and even though RPA3725 is similar in sequence to RPA3724 and RPA3723, it's binding capability may be different. So the use of protein sequence similarity to identify substrate preferences or to predict the function of SBPs is not always correct and the need for experimental data is needed.

As mentioned before, a study was done by Giuliani et al. (2011) to investigate a ligand library for binding to ABC transporter SBPs in *R. palustris* (CGA009), including the proteins under study in this chapter. Their study included the cloning and over expression of many SBPs in *E. coli* and then purifying those proteins. From more than 100 target genes they identified only about 64 SBPs that could be assigned to one or more ligands (Giuliani et al. 2011). The ligands were assigned by using a high throughput thermofluor assays that uses a broad range of ligands in a library, against each SBPs. Both RPA3723 and RPA3724 ligand binding preferences were shown to be long chain fatty acids and dicarboxylic acids (Giuliani et al. 2011). Those ligands share some structural similarity with the ligands under study in this chapter

and are used as control for all the experiments done with different length side chain phenylalkane carboxylic acids. Giuliani *et al*., (2011) reported that the pattern of binding of both RPA3723 and RPA3724 was similar and both proteins were able to bind to straight chain fatty acids with 8, 10, 12 and 14 carbon chains and only RPA3724 was able to bind to fatty acid chain of 16 carbons. Straight chain fatty acids with 6, 7 and 8 carbon chains along with the different length side chain phenylalkane carboxylic acids ligands were tested in this chapter to identify and quantify RPA3723 binding affinity. The tryptophan florescence binding assay was used and it showed that also heptanoic acid and hexanoic acid had the ability to bind to RPA3723 (Table 5.2). The binding affinity for some ligands with RPA3723 was in the high micromolar range which indicated a weak binding affinity but this work showed that this transport system may be involved in the uptake of different length side chain phenylalkane carboxylic acids as for example with 7-phenylheptanoate the affinity was around 1-2 uM. This high affinity is within the physiological range reported for many other ABC types systems and is higher than the straight-chain substrates tested.

NMR data provided evidence that RPA3724 can bind to different length side chain phenylalkane carboxylic acids of carbon 4 to carbon 8 while the thermofluor results showed that only 4-phenylbutyric acid and 7-phenylheptanoic acid can bind. The simple variation in data obtained suggests the need to use more than one technique to assess the ligand binding of a protein. The different methods used in this chapter to study the binding of SBPs indicated the need for accurate and fast methods to assign ligands to SBPs as this could help in understanding the function of adjacent gene cluster and hence the biochemical capability of the organism.

The biochemical data obtained in this chapter satisfactorily showed the binding of RPA3723 and RPA3724 SBPs and also quantified the binding ability of the RPA3723 SBP. The future work would be to investigate the other transport systems significantly up-regulated in the presence of 5-phenylvalerate and asses their affinity to find the primary transport system for the transportation of the different length side chain phenylalkane carboxylic acids. Creation of mutants in these genes to assess their phenotypic impact on the growth of *R. palustris* on these substrates would complement the *in vitro* biochemical assays. Structural analysis and crystallization to understand the binding of the ligand with protein and to elucidate the interaction formed within the binding pocket would give information of the mechanism of binding. Understanding the uptake systems of bacteria and knowing the ligands for each SBP could increase our knowledge of the complexity of soil degrading bacteria such as *R. palustris*. Furthermore, studying catabolic genes associated with these uptake systems could reveal their physiological role in utilizing different types of substrate.

6. Final Discussion

6. Final Discussion

The project aimed to understand cellular processes underlying the anaerobic degradation of lignin derived aromatic compounds and their analogues, such as different side chain length phenylalkane carboxylic acids, by *Rhodopseudomonas palustris*. Such studies are critical for the development of innovative biotechnological processes to produce valuable chemicals from biological waste (Garlapati et al. 2016) and to clean the environment from toxic contaminants (Pandey & Jain 2002). The genome sequencing of *R. palustris* revealed its biochemical diversity and the anaerobic degradation of lignin derived aromatic monomers by *R. palustris* as a model organism has been the focus of many studies (Harwood & Gibson 1988; Elder et al. 1992), but most studies have focused on phenylpropeneoids of three carbon chain length, in terms of transport systems and degradation pathways (Pan et al. 2008; Salmon et al. 2013). The two suggested pathways for degradation (β-oxidation and non β-oxidation) were studied extensively for *p*-coumarate and both proteomics and biochemical tests showed that the non β-oxidation pathway is involved in *p*-coumarate degradation (Pan et al. 2008; Hirakawa et al. 2012). However, it was unclear if different chain length phenylalkane carboxylic acids use the non β-oxidation pathway despite initial studies that showed utilization through a β-oxidation pathway (Elder et al. 1992).

The work in this study showed up-regulation of many different CoA–ligases in mass spectrometry based proteomics, which raised a question into which pathway 5-phenylvaleric acid is consumed to produce benzoyl-CoA. The biochemical characterization of the different CoA ligases cloned and expressed (Fcs1, Fcs2 and RPA1766) revealed that only Fcs1 can convert different side chain phenylalkane

carboxylic acids of more than 5 carbon side chain to their CoA derivatives in vitro. The up-regulation of proteins involved in β-oxidation and non β-oxidation was also observed in cell grown on *p*-coumarate when compared to cells grown on benzoate (Karpinets et al. 2009). A similar finding was also observed here when *R. palustris* was grown on 5-phenylvaleric acid, as many long chain fatty acid CoA ligases were up regulated in addition to the β-oxidation and non β-oxidation proteins. Karpinets et al. (2009) suggested the presence of different subpopulations with different phenotypes in *R. palustris* cultures, which might explain the presence of proteins involved in different pathways of degradation. Another explanation of the presence of proteins from different degradation pathway is that the cell identifies the different chain length phenylalkanes carboxylic acids with side chains of more than four carbons as medium and long chain fatty acid and creates the CoA derivatives on that basis. *P. putida* has the ability to utilize different side chain phenyl alkane carboxylic acids with an odd number in their chain through β-oxidation to produce

cinnamoyl-CoA that then would be accumulated as cinnamic acid (Olivera et al. 2001). The same thing was documented by (Elder et al. 1992) in *R. palustris* where cells grown on 3-phenylpropanoic acid showed an accumulation of *trans* cinnamic acid and benzoate in the media before its full utilization (Elder et al. 1992). It is possible that in an early stage of 5-phenylvaleric acid utilization, a β-oxidation pathway is used to produce cinnamic acid which then accumulates outside the cell and goes through a non β-oxidation pathway to produce benzoyl-CoA (Figure 6.1).

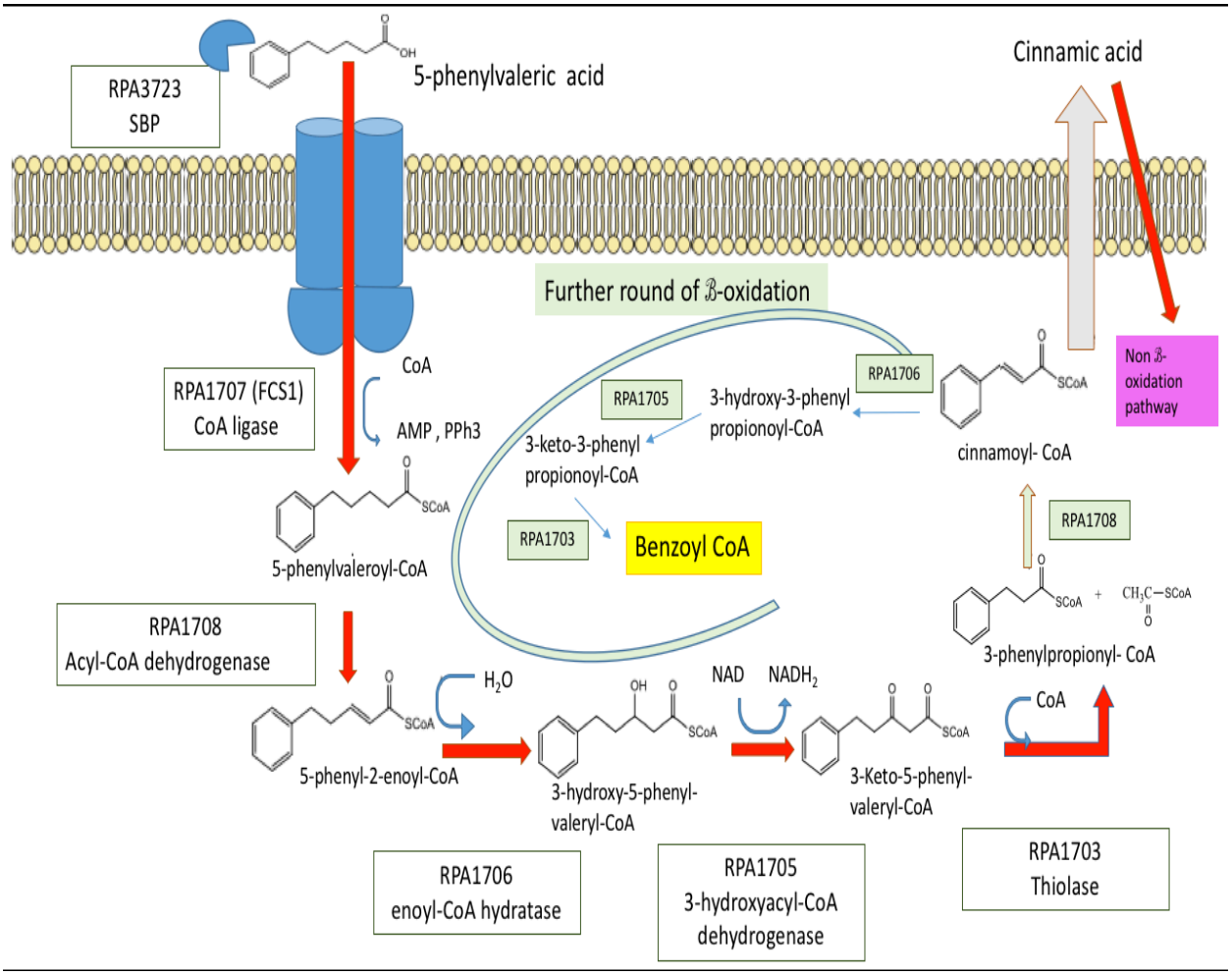


Figure 6.1. Proposed pathway for 5-phenylvaleric acid through β-oxidation and non β-oxidation.

It is noteworthy that Fcs2 did not show any activity with phenylalkane carboxylic acids having a side chain longer than 4 carbons (Hirakawa et al. 2012) while I showed that Fcs1 which is encoded in a β-oxidation operon had activity with the longer chain phenylalkane carboxylic acids. In addition, *R. palustris* mutated in both Fcs1 and Fcs2 genes were able to grow normally on 5-phenylvaleric acid which indicates that the first step in the formation of a CoA derivative can be substituted by other CoA ligases present in *R. palustris*. This finding is correlated with benzoate-CoA ligase in *R.*

palustris where inactivation of its gene did not affect the growth on benzoic acid because 4-hydroxybenzoate-CoA ligase can substitute its physiological role (Egland et al. 1995). Also when benzoate-CoA ligase and 4-hydroxybenzoate-CoA ligase were inactivated a third CoA –ligase protein was able to substitute for benzoate-CoA ligase activity (Egland et al. 1995). The β-oxidation operon encoding Fcs1 needs to be studied in more detail in the future to characterize the other proteins in the pathway biochemically and to make a conclusive statement into whether the β-oxidation pathway is the pathway for degradation of medium length phenylalkane carboxylic acids. RPA1706 that was significantly up regulated in mass spectrometry is part of this operon but the biochemical characterization was difficult to elucidate because of the need for non-commercially available CoA thioester intermediates as a substrate for this enzyme. Also the linked enzymatic assay attempted was unsuccessful. A direct future priority would be to create a mutant in the *rpa1706* gene to study its phenotypic effect on growth and to create CoA derivatives of the free acid to study the other enzymes in the pathway.

Potential phenylalkane carboxylate transporters identified in this study mainly belong to the ABC transport family. Soluble binding proteins associated with ABC transport systems in *R. palustris* were studied extensively by Giuliani *et al*. (2011) who screened all cloned soluble proteins against possible ligands but none was identified for long chain phenylalkane carboxylic acids. An ABC transport system associated with the non β-oxidation pathway (CouP) showed the ability to bind to a range of phenylpropenoid compounds such as coumarate, cinnamate, ferulate, hydrocinnamate and caffeate but showed no binding with longer chained phenylalkane carboxylic acid compounds of 4-8 carbon chain length because of the size of binding pocket that could accommodate around 3 carbon chain length to allow for optimal binding

(Salmon et al. 2013). On the other hand, as discussed in chapter 5, RPA3723 soluble binding protein associated with an ABC transporter located next to the *pim ABCDEF* operon showed binding as demonstrated by using tryptophan fluorescence spectroscopy, to 5 phenylvaleric acid, 6-phenylhexanoic acid, 7-phenyl heptanoic acid and 8-phenyloctanoic acid and this is the only tested SBP that showed binding to different length phenylalkane carboxylic acids. In conclusion this study identified several transporters and CoA ligases that have the ability to transport and utilize different side chain phenyl alkane carboxylic acids and such studies could help to broaden the knowledge about *R. palustris* utilization of complex aromatic compounds and thus exploit such knowledge to our benefit.

7. References

Akin, D.E., 1980. Attack on lignified grass cell walls by a facultatively anaerobic bacterium. *Appl. Environ. Microbiol.*, 40(4), pp.809–820.

Alexandre, G. & Zhulin, I.B., 2000. Laccases are widespread in bacteria. *Trends Biotechnol.*, 18(2), pp.41–42.

Balba, M.T. & Evans, W.C., 1977. The methanogenic fermentation of aromatic substrates. *Biochem. Soc. Trans.*, 5(1), pp.302–304.

von Ballmoos, C., Cook, G.M. & Dimroth, P., 2008. Unique rotary ATP synthase and its biological diversity. *Annu. Rev. Biophys.*, 37, pp.43–64.

Battle, M. et al., 2000. Global Carbon Sinks and Their Variability Inferred from Atmospheric O_2 and delta$_{13}$C. *Science*, 287(5462), pp.2467–2470.

Becker-André, M., Schulze-Lefert, P. & Hahlbrock, K., 1991. Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate: CoA ligase genes in potato. *J. Biol. Chem.*, 266(13), pp.8551–8559.

Benner, R., Maccubbin, A.E. & Hodson, R.E., 1984. Anaerobic biodegradation of the lignin and polysaccharide components of lignocellulose and synthetic lignin by sediment microflora. *Appl. Environ. Microbiol.*, 47(5), pp.998–1004.

Berntsson, R.P. et al., 2010a. A structural classification of substrate-binding proteins. *FEBS Lett.*, 584(12), pp.2606–2617.

Berntsson, R.P. et al., 2010b. Addendum to ``A structural classification of substrate-binding proteins. *FEBS Lett.*, 584(20), p.4373.

Berrocal, M.M. et al., 1997. Solubilisation and mineralisation of 14 C lignocellulose

from wheat straw by *Streptomyces cyaneus CECT 3335* during growth in solid-state fermentation. *Appl. Microbiol. Biotechnol.*, 48(3), pp.379–384.

Biegert, T. et al., 1993. Enzymes of anaerobic metabolism of phenolic compounds. 4-Hydroxybenzoate-CoA ligase from a denitrifying Pseudomonas species. *Eur. J. Biochem.*, 213(1), pp.555–561.

Blanchette, R.A., 1995. Degradation of the lignocellulose complex in wood. *Can. J. Bot.*, 73(S1), pp.999–1010.

Boll, M. et al., 2000. Nonaromatic Products from Anoxic Conversion of Benzoyl-CoA with Benzoyl-CoA Reductase and Cyclohexa-1,5-diene-1-carbonyl-CoA Hydratase. *J. Biol. Chem.*, 275(29), pp.21889–21895.

Boll, M., Albracht, S.S. & Fuchs, G., 1997. Benzoyl-CoA reductase (dearomatizing), a key enzyme of anaerobic aromatic metabolism. A study of adenosinetriphosphatase activity, ATP stoichiometry of the reaction and EPR properties of the enzyme. *Eur. J. Biochem.*, 244(3), pp.840–851.

Boll, M. & Fuchs, G., 1998. Identification and characterization of the natural electron donor ferredoxin and of FAD as a possible prosthetic group of benzoyl-CoA reductase (dearomatizing), a key enzyme of anaerobic aromatic metabolism. *Eur. J. Biochem.*, 251(3), pp.946–954.

Boll, M. & Fuchs, G., 2005. Unusual reactions involved in anaerobic metabolism of phenolic compounds. *Biol. Chem.*, 386(10).

Boyer, P.D., 1997. The ATP synthase- a splendid molecular machine. *Annu. Rev. Biochem.*, 66, pp.717–749.

Brackmann, R. & Fuchs, G., 1993. Enzymes of anaerobic metabolism of phenolic compounds. 4-Hydroxybenzoyl-CoA reductase (dehydroxylating) from a

denitrifying Pseudomonas species. *Eur. J. Biochem.*, 213(1), pp.563–571.

Calvo-Flores, F.G. & Dobado, J.A., 2010. Lignin as renewable raw material. *ChemSusChem*, 3(11), pp.1227–1235.

Carmona, M. et al., 2009. Anaerobic Catabolism of Aromatic Compounds: a Genetic and Genomic View. *Microbiol. Mol. Biol. Rev.*, 73(1), pp.71–133.

Chandra, R. et al., 2007. Characterisation and optimisation of three potential aerobic bacterial strains for kraft lignin degradation from pulp paper waste. *Chemosphere*, 67(4), pp.839–846.

Chen, S. et al., 2013. Carbon catabolite repression of the maltose transporter revealed by X-ray crystallography. *Nature*, 499(7458), pp.364–368.

Ciulli, A., 2013. Biophysical screening for the discovery of small-molecule ligands. *Methods Mol. Biol.*, 1008, pp.357–388.

Claus, H., 2004. Laccases: structure, reactions, distribution. *Micron*, 35(1-2), pp.93–96.

Cox, J. et al., 2009. A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat. Protoc.*, 4(5), pp.698–705.

Crosby, H.A. et al., 2010. Reversible N epsilon-lysine acetylation regulates the activity of acyl-CoA synthetases involved in anaerobic benzoate catabolism in Rhodopseudomonas palustris. *Mol. Microbiol.*, 76(4), pp.874–888.

Crosby, H.A. et al., 2012. System-wide studies of N-lysine acetylation in Rhodopseudomonas palustris reveal substrate specificity of protein acetyltransferases. *J. Biol. Chem.*, 287(19), pp.15590–15601.

Crosby, H.A. & Escalante-Semerena, J.C., 2014. The acetylation motif in AMP-

forming Acyl coenzyme A synthetases contains residues critical for acetylation and recognition by the protein acetyltransferase pat of Rhodopseudomonas palustris. *J. Bacteriol.*, 196(8), pp.1496–1504.

Cullen, D. & Kersten, P.J., 1996. Enzymology and Molecular Biology of Lignin Degradation. In R. Brambl & G. A. Marzluf, eds. *Biochemistry and Molecular Biology*. The Mycota. Springer Berlin Heidelberg, pp. 295–312.

Daniel, G. & Nilsson, T., 1998. Developments in the study of soft rot and bacterial decay. *Forest products biotechnology*, pp.37–62.

Davidson, A.L. et al., 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.*, 72(2), pp.317–64, table of contents.

Davidson, A.L. & Chen, J., 2004. ATP-binding cassette transporters in bacteria. *Annu. Rev. Biochem.*, 73, pp.241–268.

Davidson, A.L., Shuman, H.A. & Nikaido, H., 1992. Mechanism of maltose transport in *Escherichia coli*: transmembrane signaling by periplasmic binding proteins. *Proc. Natl. Acad. Sci. U. S. A.*, 89(6), pp.2360–2364.

Diaz, E., Jiménez, J.I. & Nogales, J., 2013. Aerobic degradation of aromatic compounds. *Curr. Opin. Biotechnol.*, 24(3), pp.431–442.

Donnelly, P.K. & Crawford, D.L., 1988. Production by Streptomyces viridosporus T7A of an Enzyme Which Cleaves Aromatic Acids from Lignocellulose. *Appl. Environ. Microbiol.*, 54(9), pp.2237–2244.

Dovgan', I. V & Medvedeva, E.I., 1983. Change in the structural elements of the lignin of the brown algaCystoseira barbata at different ages. *Chem Nat Compd*, 19(1), pp.81–84.

Dutton, P.L. & Evans, W.C., 1969. The metabolism of aromatic compounds by *Rhodopseudomonas palustris*. A new, reductive, method of aromatic ring metabolism. *Biochem. J*, 113(3), pp.525–536.

Egland, P.G., Gibson, J. & Harwood, C.S., 1995. Benzoate-coenzyme A ligase, encoded by badA, is one of three ligases able to catalyze benzoyl-coenzyme A formation during anaerobic growth of *Rhodopseudomonas palustris* on benzoate. *J. Bacteriol.*, 177(22), pp.6545–6551.

El-Said Mohamed, M., 2000. Biochemical and molecular characterization of phenylacetate-coenzyme A ligase, an enzyme catalyzing the first step in aerobic metabolism of phenylacetic acid in *Azoarcus evansii*. *J. Bacteriol.*, 182(2), pp.286–294.

Elder, D.J.E., Morgan, P. & 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. Arch. Microbiol., 157(2), pp.148–154.

Erdtman, H. & Holger, E., 1972. Lignins: Occurrence, formation, structure and reactions, K. V. Sarkanen and C. H. Ludwig, Eds., John Wiley & Sons, Inc., New York, 1971. *J. Polym. Sci. B*, 10(3), pp.228–230.

Erkens, G.B. et al., 2011. The structural basis of modularity in ECF-type ABC transporters. *Nat. Struct. Mol. Biol.*, 18(7), pp.755–760.

Ferraroni, M. et al., 2007. Crystal structure of a blue laccase from *Lentinus tigrinus*: evidences for intermediates in the molecular oxygen reductive splitting by multicopper oxidases. *BMC Struct. Biol.*, 7, p.60.

Forrest, L.R., Krämer, R. & Ziegler, C., 2011. The structural basis of secondary active

transport mechanisms. *Biochim. Biophys. Acta*, 1807(2), pp.167–188.

Fuchs, G., 2008. Anaerobic metabolism of aromatic compounds. *Ann. N. Y. Acad. Sci.*, 1125, pp.82–99.

Fuchs, G., Boll, M. & Heider, J., 2011. Microbial degradation of aromatic compounds - from one strategy to four. *Nat. Rev. Microbiol.*, 9(11), pp.803–816.

Galbe, M. & Zacchi, G., 2007. Pretreatment of Lignocellulosic Materials for Efficient Bioethanol Production. In L. Olsson, ed. *Biofuels*. Advances in Biochemical Engineering/Biotechnology. Springer Berlin Heidelberg, pp. 41–65.

Gall, A. & Robert, B., 1999. Characterization of the different peripheral light-harvesting complexes from high- and low-light grown cells from *Rhodopseudomonas palustris*. *Biochemistry*, 38(16), pp.5185–5190.

Garcia, B. et al., 1999. Novel biodegradable aromatic plastics from a bacterial source. Genetic and biochemical studies on a route of the phenylacetyl-coa catabolon. *J. Biol. Chem.*, 274(41), pp.29228–29241.

Gargulak, J.D. & Lebo, S.E., 2000. Commercial use of lignin-based materials. In *ACS symposium series*. pp. 304–320.

Garlapati, V.K., Shankar, U. & Budhiraja, A., 2016. Bioconversion technologies of crude glycerol to value added industrial products. *Biotechnology Reports*, 9, pp.9–14.

Gasson, M.J. et al., 1998. Metabolism of Ferulic Acid to Vanillin: A bacterial gene of the enoyl-SCoA hydratase/isomerase superfamily encodes an enzyme for the hydration and cleavage of a hydroxycinnamic acid SCoA thioester. *J. Biol. Chem.*, 273(7), pp.4163–4170.

Gibson, J. et al., 1994. 4-Hydroxybenzoate-coenzyme A ligase from

Rhodopseudomonas palustris: purification, gene sequence, and role in anaerobic degradation. *J. Bacteriol.*, 176(3), pp.634–641.

Gilson, E. et al., 1988. Evidence for high affinity binding-protein dependent transport systems in gram-positive bacteria and in Mycoplasma. *EMBO J.*, 7(12), pp.3971–3974.

Giuliani, S.E. et al., 2011. Environment sensing and response mediated by ABC transporters. *BMC Genomics*, 12 Suppl 1, p.S8.

Gonin, S. et al., 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, p.11.

Guillén, F. et al., 1997. Quinone Redox Cycling in the Ligninolytic Fungus *Pleurotus eryngii* Leading to Extracellular Production of Superoxide Anion Radical. *Arch. Biochem. Biophys.*, 339(1), pp.190–199.

Guzman, L.M. et al., 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.*, 177(14), pp.4121–4130.

Hancock, R.E. & Nikaido, H., 1978. Outer membranes of gram-negative bacteria. XIX. Isolation from *Pseudomonas aeruginosa* PAO1 and use in reconstitution and definition of the permeability barrier. *J. Bacteriol.*, 136(1), pp.381–390.

Harayama, S., 1992. Functional and Evolutionary Relationships Among Diverse Oxygenases. *Annu. Rev. Microbiol.*, 46(1), pp.565–601.

Harrison, F.H. & Harwood, C.S., 2005. The pimFABCDE operon from *Rhodopseudomonas palustris* mediates dicarboxylic acid degradation and participates in anaerobic benzoate degradation. *Microbiology*, 151(Pt 3), pp.727–

736.

Hartley, R.D. & Ford, C.W., 1989. Phenolic Constituents of Plant Cell Walls and Wall Biodegradability. In *Plant Cell Wall Polymers*. pp. 137–145.

Harwood, C.S. & Gibson, J., 1988. Anaerobic and aerobic metabolism of diverse aromatic compounds by the photosynthetic bacterium *Rhodopseudomonas palustris*. *Appl. Environ. Microbiol.*, 54(3), pp.712–717.

Heider, J. & Fuchs, G., 1997. Anaerobic metabolism of aromatic compounds. *Eur. J. Biochem.*, 243(3), pp.577–596.

Heinfling, A. et al., 1998. Transformation of industrial dyes by manganese peroxidases from *Bjerkandera adusta* and *Pleurotus eryngii* in a manganese-independent reaction. *Appl. Environ. Microbiol.*, 64(8), pp.2788–2793.

Higgins, C.F. & Linton, K.J., 2004. The ATP switch model for ABC transporters. *Nat. Struct. Mol. Biol.*, 11(10), pp.918–926.

Hirakawa, H. et al., 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(8), pp.1960–1967.

Hollenstein, K., Dawson, R.J.P. & Locher, K.P., 2007. Structure and mechanism of ABC transporter proteins. *Curr. Opin. Struct. Biol.*, 17(4), pp.412–418.

Hosie, A.H. et al., 2001. Solute-binding protein-dependent ABC transporters are responsible for solute efflux in addition to solute uptake. *Mol. Microbiol.*, 40(6), pp.1449–1459.

Hu, X. et al., 2002. Photosynthetic apparatus of purple bacteria. *Q. Rev. Biophys.*, 35(1), pp.1–62.

Huang, Z., Dostal, L. & Rosazza, J.P., 1993. Mechanisms of ferulic acid conversions to vanillic acid and guaiacol by *Rhodotorula rubra*. *J. Biol. Chem.*, 268(32), pp.23954–23958.

Hunter, C.N. et al., 2008. *The Purple Phototrophic Bacteria*, Springer Science & Business Media.

Iiyama, K., Lam, T. & Stone, B.A., 1994. Covalent Cross-Links in the Cell Wall. *Plant Physiol.*, 104(2), pp.315–320.

Jaehme, M. & Slotboom, D.J., 2015. Diversity of membrane transport proteins for vitamins in bacteria and archaea. *Biochim. Biophys. Acta*, 1850(3), pp.565–576.

Joseph, B. et al., 2011. Transmembrane gate movements in the type II ATP-binding cassette (ABC) importer BtuCD-F during nucleotide cycle. *J. Biol. Chem.*, 286(47), pp.41008–41017.

Kaczorowski, G.J. & Kaback, H.R., 1979. Mechanism of lactose translocation in membrane vesicles from *Escherichia coli*. 1. Effect of pH on efflux, exchange, and counterflow. *Biochemistry*, 18(17), pp.3691–3697.

Kanazawa, T. et al., 2010. Biochemical and physiological characterization of a BLUF protein-EAL protein complex involved in blue light-dependent degradation of cyclic diguanylate in the purple bacterium *Rhodopseudomonas palustris.* *Biochemistry*, 49(50), pp.10647–10655.

Karpinets, T. V et al., 2009. Phenotype fingerprinting suggests the involvement of single-genotype consortia in degradation of aromatic compounds by *Rhodopseudomonas palustris*. *PLoS One*, 4(2), p.e4615.

Karpowich, N.K. et al., 2003. Crystal structures of the BtuF periplasmic-binding protein for vitamin B12 suggest a functionally important reduction in protein

mobility upon ligand binding. *J. Biol. Chem.*, 278(10), pp.8429–8434.

Khare, D. et al., 2009. Alternating access in maltose transporter mediated by rigid-body rotations. *Molecular cell*, 33(4), pp.528–536.

Kim, M.-S. et al., 1998. Degradation of polycyclic aromatic hydrocarbons by selected white-rot fungi and the influence of lignin peroxidase. *J. Microbiol. Biotechnol.*, 8(2), pp.129–133.

Kim, Y.H. et al., 2006. Analysis of aromatic catabolic pathways in *Pseudomonas putida* KT 2440 using a combined proteomic approach: 2-DE/MS and cleavable isotope-coded affinity tag analysis. *Proteomics*, 6(4), pp.1301–1318.

Korkhov, V.M., Mireku, S.A. & Locher, K.P., 2012. Structure of AMP-PNP-bound vitamin B12 transporter BtuCD-F. *Nature*, 490(7420), pp.367–372.

Larimer, F.W. et al., 2004. Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. *Nat. Biotechnol.*, 22(1), pp.55–61.

Malherbe, S. & Cloete, T.E., 2002. Lignocellulose biodegradation: Fundamentals and applications. *Rev. Environ. Sci. Biotechnol.*, 1(2), pp.105–114.

Mao, J. et al., 2006. Differences between Lignin in Unprocessed Wood, Milled Wood, Mutant Wood, and Extracted Lignin Detected by 13 C Solid-State NMR. *J. Agric. Food Chem.*, 54(26), pp.9677–9686.

Martinez-Blanco, H. et al., 1990. Purification and biochemical characterization of phenylacetyl-CoA ligase from *Pseudomonas putida*. A specific enzyme for the catabolism of phenylacetic acid. *J. Biol. Chem.*, 265(12), pp.7084–7090.

Martínez, A.T., 2002. Molecular biology and structure-function of lignin-degrading heme peroxidases. *Enzyme Microb. Technol.*, 30(4), pp.425–444.

Martinez, A.T. et al., 2005. Biodegradation of lignocellulosics: microbial, chemical, and enzymatic aspects of the fungal attack of lignin. *Int. Microbiol.*, 8(3), pp.195–204.

Mohamed, M.E.-S. & Fuchs, G., 1993. Purification and characterization of phenylacetate-coenzyme A ligase from a denitrifying Pseudomonas sp., an enzyme involved in the anaerobic degradation of phenylacetate. *Arch. Microbiol.*, 159(6), pp.554–562.

Morgenstern, I., Klopman, S. & Hibbett, D.S., 2008. Molecular evolution and diversity of lignin degrading heme peroxidases in the *Agaricomycetes*. *J. Mol. Evol.*, 66(3), pp.243–257.

Mulligan, C. & Mindell, J.A., 2013. Mechanism of Transport Modulation by an Extracellular Loop in an Archaeal Excitatory Amino Acid Transporter (EAAT) Homolog. *J. Biol. Chem.*, 288(49), pp.35266–35276.

Narbad, A. & 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, pp.1397–1405.

Neiditch, M.B. et al., 2006. Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. *Cell*, 126(6), pp.1095–1108.

Nie, G., Reading, N.S. & Aust, S.D., 1999. Relative Stability of Recombinant Versus Native Peroxidases from *Phanerochaete chrysosporium*. *Arch. Biochem. Biophys.*, 365(2), pp.328–334.

Oldham, M.L. & Chen, J., 2011. Snapshots of the maltose transporter during ATP hydrolysis. *Proc. Natl. Acad. Sci. U. S. A.*, 108(37), pp.15152–15156.

Olivera, E.R. et al., 2001. Two different pathways are involved in the beta-oxidation

of n-alkanoic and n-phenylalkanoic acids in *Pseudomonas putida U*: genetic studies and biotechnological applications. *Mol. Microbiol.*, 39(4), pp.863–874.

Overhage, J. et al., 1999. Biotransformation of eugenol to vanillin by a mutant of *Pseudomonas*. *Appl. Microbiol. Biotechnol.*, 52(6), pp.820–828.

Palmqvist, E. & Hahn-Hägerdal, B., 2000. Fermentation of lignocellulosic hydrolysates. {II}: inhibitors and mechanisms of inhibition. *Bioresour. Technol.*, 74(1), pp.25–33.

Pan, C. et al., 2008. Characterization of anaerobic catabolism of p-coumarate in *Rhodopseudomonas palustris* by integrating transcriptomics and quantitative proteomics. *Mol. Cell. Proteomics*, 7(5), pp.938–948.

Pandey, G. & Jain, R.K., 2002. Bacterial chemotaxis toward environmental pollutants: role in bioremediation. *Appl. Environ. Microbiol.*, 68(12), pp.5789–5795.

Pao, S.S., Paulsen, I.T. & Saier Jr, M.H., 1998. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.*, 62(1), pp.1–34.

Paula, S. et al., 1996. Permeation of protons, potassium ions, and small polar molecules through phospholipid bilayers as a function of membrane thickness. *Biophys. J.*, 70(1), pp.339–348.

Pedersen, P.L. & Carafoli, E., 1987. Ion motive ATPases I. Ubiquity, properties, and significance to cell function. *Trends Biochem. Sci.*, 12, pp.146–150.

Peng, C. et al., 2011. The first identification of lysine malonylation substrates and its regulatory enzyme. *Mol. Cell. Proteomics*, 10(12), p.M111.012658.

Phattarasukol, S. et al., 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(19), pp.6812–6818.

Plaggenborg, R., Steinbüchel, A. & 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 Microbiol. Lett.*, 205(1), pp.9–16.

Pohl, A., Devaux, P.F. & Herrmann, A., 2005. Function of prokaryotic and eukaryotic ABC proteins in lipid transport. *Biochim. Biophys. Acta*, 1733(1), pp.29–52.

Quiocho, F.A. & Ledvina, P.S., 1996. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol.*, 20(1), pp.17–25.

Ralph, J. et al., 2004. Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl- propanoids. *Phytochem. Rev.*, 3(1-2), pp.29–60.

Rather, L.J. et al., 2011. Structure and Mechanism of the Diiron Benzoyl-Coenzyme A Epoxidase BoxB. *J. Biol. Chem.*, 286(33), pp.29241–29248.

Rees, D.C., Johnson, E. & Lewinson, O., 2009. ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.*, 10(3), pp.218–227.

Reynolds, P.R., Mottur, G.P. & Bradbeer, C., 1980. Transport of vitamin B12 in *Escherichia coli*. Some observations on the roles of the gene products of BtuC and TonB. *J. Biol. Chem.*, 255(9), pp.4313–4319.

Reyrat, J.M. et al., 1998. Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infect. Immun.*, 66(9), pp.4011–4017.

Salmon, R.C. et al., 2013. The CouPSTU and TarPQM transporters in *Rhodopseudomonas palustris*: redundant, promiscuous uptake systems for lignin-derived aromatic substrates. *PLoS One*, 8(3), p.e59844.

Sánchez, C., 2009. Lignocellulosic residues: biodegradation and bioconversion by fungi. *Biotechnol. Adv.*, 27(2), pp.185–194.

Schmeling, S. et al., 2004. Phenylphosphate Synthase: a New Phosphotransferase Catalyzing the First Step in Anaerobic Phenol Metabolism in *Thauera aromatica*. *J. Bacteriol.*, 186(23), pp.8044–8057.

Schneider, K. et al., 2003. The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase. *Proc. Natl. Acad. Sci. U. S. A.*, 100(14), pp.8601–8606.

Schühle, K. et al., 2003. Benzoate-coenzyme A ligase from *Thauera aromatica*: an enzyme acting in anaerobic and aerobic pathways. *J. Bacteriol.*, 185(16), pp.4920–4929.

Schuhle, K. & Fuchs, G., 2004. Phenylphosphate Carboxylase: a New C-C Lyase Involved in Anaerobic Phenol Metabolism in *Thauera aromatica*. *J. Bacteriol.*, 186(14), pp.4556–4567.

Simpson, F.B. & Burris, R.H., 1984. A nitrogen pressure of 50 atmospheres does not prevent evolution of hydrogen by nitrogenase. *Science*, 224(4653), pp.1095–1097.

Six, S. et al., 1994. *Escherichia coli* possesses two homologous anaerobic C4-dicarboxylate membrane transporters (DcuA and DcuB) distinct from the aerobic dicarboxylate transport system (Dct). *J. Bacteriol.*, 176(21), pp.6470–6478.

Sixta, H. ed., 2006. *Handbook of Pulp*, Weinheim, Germany: Wiley-VCH Verlag GmbH.

Sjoblad, R.D. & Bollag, J.M., 1981. Oxidative coupling of aromatic compounds by enzymes from soil microorganisms. In *Soil Biochemistry*. Dekker, pp. 113–152.

Sjöström, E., 1993. *Wood Chemistry: Fundamentals and Applications*, Gulf

Professional Publishing.

Slotboom, D.J., 2014. Structural and mechanistic insights into prokaryotic energy-coupling factor transporters. *Nat. Rev. Microbiol.*, 12(2), pp.79–87.

Smart, J.P., Cliff, M.J. & Kelly, D.J., 2009. A role for tungsten in the biology of *Campylobacter jejuni*: tungstate stimulates formate dehydrogenase activity and is transported via an ultra-high affinity ABC system distinct from the molybdate transporter. *Mol. Microbiol.*, 74(3), pp.742–757.

Starai, V.J. et al., 2002. Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine. *Science*, 298(5602), pp.2390–2392.

Steen, H. & Mann, M., 2004. The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.*, 5(9), pp.699–711.

Sugimoto, K. et al., 1999. Crystal structure of an aromatic ring opening dioxygenase LigAB, a protocatechuate 4,5-dioxygenase, under aerobic conditions. *Structure*, 7(8), pp.953–965.

Tam, R. & Saier Jr, M.H., 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.*, 57(2), pp.320–346.

Tchieu, J.H. et al., 2001. The complete phosphotransferase system in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.*, 3(3), pp.329–346.

Toyoshima, C. et al., 2013. Crystal structures of the calcium pump and sarcolipin in the Mg2+-bound E1 state. *Nature*, 495(7440), pp.260–264.

Trautwein, K. et al., 2012. Physiological and proteomic adaptation of ``*Aromatoleum aromaticum*'' EbN1 to low growth rates in benzoate-limited, anoxic chemostats. *J. Bacteriol.*, 194(9), pp.2165–2180.

Tuomela, M. et al., 2005. Influence of Pb contamination in boreal forest soil on the growth and ligninolytic activity of litter-decomposing fungi. *FEMS Microbiol. Ecol.*, 53(1), pp.179–186.

Uzal, E.N. et al., 2009. The presence of sinapyl lignin in *Ginkgo biloba* cell cultures changes our views of the evolution of lignin biosynthesis. *Physiol. Plant.*, 135(2), pp.196–213.

Vanholme, R. et al., 2010. Lignin biosynthesis and structure. *Plant Physiol.*, 153(3), pp.895–905.

Vedadi, M. et al., 2006. Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc. Natl. Acad. Sci. U. S. A.*, 103(43), pp.15835–15840.

Vigonsky, E., Ovcharenko, E. & Lewinson, O., 2013. Two molybdate/tungstate ABC transporters that interact very differently with their substrate binding proteins. *Proc. Natl. Acad. Sci. U. S. A.*, 110(14), pp.5440–5445.

Wang, T. et al., 2013. Structure of a bacterial energy-coupling factor transporter. *Nature*, 497(7448), pp.272–276.

Ward, P.G. & O' Connor, K.E., 2005. Induction and quantification of phenylacyl-CoA ligase enzyme activities in *Pseudomonas putida* CA-3 grown on aromatic carboxylic acids. *FEMS Microbiol. Lett.*, 251(2), pp.227–232.

Weaver, P.F., Wall, J.D. & Gest, H., 1975. Characterization of *Rhodopseudomonas capsulata. Arch. Microbiol.*, 105(3), pp.207–216.

Wesenberg, D., Kyriakides, I. & Agathos, S.N., 2003. White-rot fungi and their enzymes for the treatment of industrial dye effluents. *Biotechnol. Adv.*, 22(1-2), pp.161–187.

Wilkens, S., 2015. Structure and mechanism of ABC transporters. *F1000Prime Rep.*, 7, p.14.

Wong, D.W.S., 2009. Structure and action mechanism of ligninolytic enzymes. *Appl. Biochem. Biotechnol.*, 157(2), pp.174–209.

Xiao, G. et al., 2012. Catalytic Carbonization of Lignin for Production of Electrically Conductive Charcoal. *J. Biobased Materials Bioenergy*, 6(1), pp.69–74.

Zhang, P., 2013. Structure and mechanism of energy-coupling factor transporters. *Trends Microbiol.*, 21(12), pp.652–659.

Zhu, W., Smith, J.W. & Huang, C.-M., 2010. Mass spectrometry-based label-free quantitative proteomics. *J. Biomed. Biotechnol.*, 2010, p.840518.