Structural characterization of thiolase and decarboxylase enzymes for biotechnological application in microbial degradation of water pollutants

Submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

of the
Indian Institute of Technology, Bombay, India
and
Monash University, Australia

by

Sukritee Bhaskar

Supervisors:

Dr. Ruchi Anand (IIT Bombay)

Dr. Santosh Panjikar (Monash University)

The course of study for this award was developed jointly by
the Indian Institute of Technology, Bombay and Monash University, Australia
and given academic recognition by each of them.
The programme was administered by the IITB-Monash Research Academy

(Year 2019)

Dedicated to my parents

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand

that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Notice 1

Under the Copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

Notice 2

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Student Name: Sukritee Bhaskar
IITB ID: 144034005
Monash ID: 26852497
Date: 4th December, 2019

ACKNOWLEDGEMENT

Penning my scientific contributions into thesis has led me to reflect my PhD journey of five years. It has made me realized that these years, starting 2015 has been quite enduring; full of highs and lows. And, it would not have been possible without the constant help and encouragement of family, friends, colleagues, teachers and well-wishers.

Foremost, I would like to acknowledge my parents, Mrs Sunita Bhaskar and Mr. Sajal Kanti Bhaskar, who has been my backbone and my constant cheerleader. My mother has been on my side with every challenge, irrespective of the outcome. She is the person who has taught me to be my own person and has enabled me to achieve my dreams. She is the person whom I ardently admire and wish to inculcate her kindness, self-reliance and morals into my life.

I am very fortunate and proud to be my father's daughter, the man who has strived his best in giving me a better future. He has always motivated me to be independent. He has always made time for me, in-spite of his busy schedule, and has been present in every of exams, starting from kindergarten to PhD. I hope that I can be even an iota of what my parents aspire me to be.

Word cannot express my gratefulness to my friend, Dr. Rajeshwara Chary, for making my stay in IIT Bombay memorable. He has lifted my spirit when I was feeling despondent and has helped me in toughest moments of my life. He has been my constant source of support and has always urged me to enjoy both my professional and personal life.

I am especially indebted to my Monash supervisor, Dr. Santosh Panjikar, for being my mentor and believing in me. He has been an exceptional guide who taught me to look into the research problems with insights. He has worked along with me on research problems, when I have been stuck. He has provided constant encouragement and has been always available for discussions. He is a person whose hard-working nature and work ethics I will like to emulate.

I will also like to sincerely thank my IIT Bombay supervisor, Dr. Ruchi Anand for giving me freehand to work on research problems. She has provided valuable ideas and suggestions during my PhD. I am glad to be associated with her.

I am lucky to have one of the best seniors during my PhD who has constantly provided best advices and bore my rantings. I would especially like to mention Shamayeeta di, Vandana, Ruchika and Nandini di. I also thank all my lab members for creating a positive environment in the lab.

I am thankful to my RPC member Dr. Mibel Aguiler (Monash University) who has went out of her way to help me during my PhD and has evaluated my annual progress with insightful comments. I am also grateful to my other RPC Dr. Debabrata maiti (IIT Bombay) for his inputs in my PhD project.

I would like acknowledge Dr. Rachel Williamson, Dr. Alan Riboldi-Tunnicliffe, Dr. Tom Caradoc-Davies and MX staff for making my one year stay in Australian Synchrotron memorable. I would like to thank Clare Scott for her technical support at the wet lab. I would like to acknowledge Australian synchrotron for hosting me for a year and allowing me to perform experimental studies.

I thank IITB Research Academy for funding the research and providing me research scholarship.

Last but not the least, I would like to thank my friends especially, Priyanka, for the cherished memories and our outings that make my IIT Bombay.

Sukritee Bhaskar
IIT Bombay, Powai
Mumbai

ABSTRACT

Around half of the world population will be facing water scarcity in their living area by 2025, according to the WHO report, 2019. The escalating human activities such as agriculture, industrialization, mining, and the like have contributed to the rising water crisis. The aromatic water pollutants such as polycyclic aromatic hydrocarbons (PAHs) and heterocyclic aromatic derivatives are inadequately managed. These compounds are lethal, persistent, and carcinogenic and pose a threat to the environment. These xenobiotic compounds have infiltrated the terrestrial and aquatic systems and have contaminated the drinking water of millions of people. There is a dire need to eradicate these toxic compounds. Microbial degradation of toxic compounds via specific catabolic pathways is an eco-friendly approach to build a greener environment. Bacterial species such as *Pseudomonas Sp.*, *Acinetobacter Sp*. possess the catabolic ability to bind a variety of aromatic pollutants and degrade them into non-toxic tricarboxylic acid (TCA) cycle intermediates. They employ transcriptional regulatory elements, which detect the aromatic pollutants and activate the degradation pathways for the aromatic pollutants. There are four major microbial degradative pathways: the homogentisate pathway, the phenylacetate pathway, the catechol, and the protocatechuate pathway. The catechol pathway and protocatechuate pathway converge to form the β-ketoadipate central pathway. These pathways comprise of a cassette of catabolic enzymes that breakdown the specific pollutant and utilize them as their carbon source. Prospective enzymes or a series of them from the bioremediation pathways can be used to develop tools and technology to eliminate aromatic water pollutants. The central theme of this thesis is to understand the microbial degradation of toxic compounds via specific catabolic pathways with the ultimate goal of rendering the surrounding environment free from the toxic aromatics.

To get a better overview of the degradation process, two key enzymes from *Pseudomonas putida* KT2440 are chosen for structural and biochemical studied in this thesis. Both the enzymes belong to the β-ketoadipate central pathway, which is one of the key bioremediation pathways in *Pseudomonas sp.* and can assimilate both catechol and protocatechuate intermediates. The first enzyme is annotated as β-ketoadipyl-CoA thiolase (PcaF), which acts in the final step of the β-ketoadipate pathway and concludes the bioremediation process. Thiolase generally is a well-characterized family of enzymes with two distinct categories of degradative and biosynthetic type. Both the classes share identical catalytic triad but catalyze

the reactions in radically opposite directions. Hitherto, no residue or structural pattern is recognized that might help to discern the two thiolases. In this study, explicit details have been shed on PcaF, which belongs to the lesser-studied degradative thiolase, using structural studies and bioinformatic analyses. It is shown that degradative thiolases have different active site architecture compared to biosynthetic, which demonstrate the dissimilar chemical nature of the two active site architecture. Both the type of thiolases deploy different "anchoring residues" to tether the large Coenzyme A (CoA) or CoA derivatives. Interestingly, the H356 of the catalytic triad in PcaF is directly involved in tethering the CoA/ CoA derivatives into the active site in oppose to biosynthetic thiolase. H356A mutation results in a gridlocked thiolase structure where the CoA is found to be covalently linked to the catalytic cysteine residues inhibiting the overall reaction. Furthermore, H356A causes the side chain of M163 to flip inside the active site cavity disrupting the active site architecture. A long tunnel of 235 $Å^2$ surface area in PcaF is identified wherein, two longer chain CoA derivatives- Hexanal CoA and Octanal CoA were trapped for the first time. This demonstrate the substrate adaptability of PcaF, making it a promising system for biotransformation of range of aromatic pollutants. Furthermore, the dimensions and physio-chemical properties of a covering loop help to distinguish the two thiolases and also play an active role in determining the tunnel length and the nature of the binding substrate.

The second protein whose study was undertaken is PcaC which is the first common product of the two branches (protocatechuate and catechol branches) of the β-ketoadipate pathway and hence, is a strategically located enzyme. PcaC belongs to carboxymuconolactone decarboxylase (CMD) family and is branded as γ-CMD because of its sequence identity (30% ~ 50%) with the γ-CMDs of other species and the absence of the CXXC motif. PcaC catalyzes the conversion of γ-carboxymuconolactone to β-ketoadipate enol lactone and releases CO_2 as the by-product, but to date, the active site of these class of decarboxylases has remained elusive. Moreover, these enzymes catalyze metal-free decarboxylation reaction thus, they pose an interesting system to purse from a mechanistic point of view. In this study, the active site residues are characterized and the mechanism of γ-CMD is proposed by using a combination of evolutionary studies, bioinformatics and structural analysis of PcaC. Preliminary identification of the active site based on sequence conservation was proven to be quite challenging due to low sequence identity of PcaC with any other well characterized enzymes. This led to the invention of 'Bicarbonate soaking' method where the active site of PcaC was identified using sodium bicarbonate. Bicarbonate closely mimic a carboxylic group as it is deprotonated form of carbonic acid (which is an example of a straight chained saturated carboxylic acid) and interacts

with the active site residues in decarboxylases. The bicarbonate soaking method also traces the CO_2 release pathway in PcaC. The active site residues predicted by the bicarbonate soaking method is validated by mutagenesis, soaking experiment of PcaC mutant with substrate/product analogue and mass spectrometry-based studies. The studies reveal that E76 is the central residue for decarboxylation in PcaC, while H80 plays role in alignment of the substrate. The crystal structures of PcaC variant, PcaC-E76, in complex with substrates analog, homocitrate acid lactone (HCAL), and product analog, 2-Furan acetic acid (2FAA) further confirms the active site. The PcaC-H80A mutant soaked with HCAL form the reaction intermediate which is trapped in the active site. Furthermore, the γ-carboxymuconolactone is modelled into PcaC structure. Based on all the information obtained from this study, the enzymatic mechanism is elucidated. The work provides an in-depth study of the γ-CMD family with a holistic view of the CMD family. Overall, these two simultaneous projects provide us with enhanced information about the β-ketoadipate pathway and help to develop approaches that can detect and remove pollutants from the environment.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

List of Abbreviations

2FAA	2- Furan acetic acid
3D	Three dimensional
AA-mutants	H356A-C386A Mutant
ACAT	Acetyl-coenzyme A acetyl transferases
AhpC	Peroxiredoxin
AhpD	Alkyl-hydroperoxidase
A-mutant-HAL-CoA	A-mutant-hexanal CoA complex
A-mutant-Hex-CoA	A-mutant-Hexanoyl CoA complex
A-mutants	H356A mutant
ANSTO	Australian Nuclear Science and Technology Organisation
AS-mutant-OAL-CoA	AS-mutant-octanal CoA complex
AS-mutant-Oct-CoA	AS-mutant-Octanyol CoA complex
AS-mutants	H356A-C90S Mutant
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
BTEX	Benzene, toluene, ethylbenzene and xylene
CMD	Carboxymuconolactone decarboxylase
CO_2	Carbon dioxide
CO_3^{2-}	Carbonate ion
CoA	Coenzyme A
COOT	Crystallographic Object Oriented Toolkit
Da	Dalton
DALI	Distance alignment matrix method
DmpR	3,4- dimethylphenol catabolism regulator
DNA	Deoxyribo nucleic acid
dNTP	Deoxyribonucleotide triphosphate.
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Fwd	Forward
HCAL	Homocitrate acid lactone
HAL	Hexanal
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hex-CoA	Hexanoyl CoA
His	Histidine
HMMs	Hidden Markov models
IPTG	Isopropyl-β-d- thiogalactopyranoside
LB	Luria broth
LSQ	Least square fitting

MAD	Multi-wavelength anomalous diffraction
min	Minutes
MR	Molecular replacement
MRSAD	Molecular replacement single-wavelength anomalous diffraction
MSA	Multiple sequence alignment
Mt-AhpD	*Mycobacterium tuberculosis* AhpD
Mtb-thiolase	*Mycobacterium tuberculosis* thiolase
NaHCo3	Sodium bicarbonate
NCBI	National Center for Biotechnology Information
Ni-NTA	Nickel-nitrilotriacetic acid
NYSGRC	New York Structural Genomics Research Consortium
OAL	Octanal
Oct-CoA	Octanoyl CoA
OD	Optical density
PAHs	Polycyclic aromatic hydrocarbons
PcaF	β-ketoadipyl-CoA thiolase
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PEL	Permissible exposure limit
Rev	Reverse
rmsd	Root mean square deviation
RNA	Ribonucleic acid
rpm	Revolutions per minute
SAD	Single anomalous diffraction phasing
SDM	Site directed mutagenesis
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	Seconds
SSM	Secondary structure matching
TCA	Tricarboxylic acid
TEV	Tobacco etch virus
vol	Volume
WHO	World health organization
XDS	X-ray Detector Software
XylR	Xylene catabolism regulator
Zr-thiolase	*Zoogleria ramigera* thiolase
γ-CMD	γ-carboxymuconolactone decarboxylase

List of Symbols

α	Alpha
Å	Angstrom
β	Beta
°C	Degree Celsius
γ	Gamma
μ	Micro
R-factor	Residual-factor
I/σ	Signal to noise ratio
σ	Sigma

Chapter 1
Introduction

1.1 Background and Motivation

The United Nations General Assembly marked the 2005 to 2015 years as "the international Decade for Action: Water for Life" (WHO/UNICEF Joint Monitoring Programme 2005). World Health Organization (WHO) in their 2019 report predicted that the acute water shortage would affect half of the population by 2025, and one in four people will face a water crisis by 2050 (Boretti & Rosa 2019; Helmer et al. 1997; WHO 2008). These statistics indicate the looming water crisis in the near future. Water is crucial for the survival and well-being of living organisms. Water scarcity and poor quality negatively influence human health and the sustenance of the ecosystems. Pollution is one of the primary reasons for the present water crisis (Kjellstrom et al. 2006). Human activities have led to the contamination of the water bodies and have decreased the level of "safe water". The advent of industrial age is also attributed to the pollution of water sources. It has caused a startling upsurge of the polycyclic aromatic hydrocarbons (PAHs) and heterocyclic aromatics derivatives in the water bodies (WHO 1998).These natural and human-made aromatic compounds discharged from industrial, urban activities and geochemical cycles cannot be degraded or utilized by most organisms, thus creating severe water pollution (Gong et al. 2019). These pollutants are released into the water bodies where they show bioaccumulation and impact the ecology (Wild, Berrow, & Jones 1991). These aromatic compounds are toxic and carcinogenic in nature and thus pose a threat to humankind (Yang et al. 2011). One of the most predominant environmental concerns of recent times is to eradicate these lethal pollutants, but is quite a difficult task as these xenobiotic group of pollutants are highly tenacious thus making their purging from the biosphere quite challenging (Abdel-shafy & Mansour 2016; Wang et al. 2007; Yang, Zhang, Xue, & Tao 2010).

Interestingly, a plethora of bacteria exist in nature that have wide metabolic diversity and genetic plasticity. Those bacteria can adapt to limiting nutrient supplies and hostile environments that are typical of the polluted water (Díaz & Prieto 2000; Tropel & Meer 2004). These bacterial species exhibit catabolic ability to biodegrade a variety of water and soil-based aromatic pollutants into the simple, non-toxic tricarboxylic acid (TCA) cycle intermediates by using the pollutants as their sole carbon and energy source (Harayama, Kok, & Neidle 1992). The bacteria such as *Pseudomonas*, *Rhodococcus*, and *Corynebacterium* can metabolize the chemical contaminants via aromatic degradation pathways. The pathways are triggered by the presence of pollutants in the environment (Jiménez, Miñambres, García, & Díaz 2002; D. Kim et al. 2002; Shen, Huang, & Liu 2005).

Pseudomonas putida KT2440 is a well-studied strain of soil bacteria that can assimilate a variety of aromatic compounds via four major pathways, namely, the homogentisate pathway (*hmg/fah/mai* genes), the phenylacetate pathway (*pha* genes), the catechol (*cat* genes) and the protocatechuate (*pca* genes) pathway (Williams & Sayers 1994; Wu et al. 2011). The catechol and protocatechuate pathway converge to form the central β-ketoadipate pathway, which is a taxonomically widespread bioremediation pathway in bacteria (Ornston 1966c, 1966a).Usually, the pathways employ two part strategy involving ring cleavage by diverse ring alteration reactions followed by the ring fission reactions to degrade these aromatic pollutants into the non-toxic tricarboxylic acid cycle (TCA) intermediates (Buchan, Collier, Neidle, & Moran 2000). In the case of the β-ketoadipate pathway, the ring fission leads to the formation of the crucial β-ketoadipate via an ortho-cleavage reaction, which undergoes further two steps to give rise to the TCA intermediate (Nelson et al. 2002; Yamanashi, Kim, Hara, & Funa 2015). Microbial aromatic degradation pathways are robust systems to bio-transform the effluents with low environmental impact. Therefore, knowledge of the pathways and its' significant catabolic enzymes are promising means to develop effective aromatic pollutant elimination technologies. Further, comprehensive elucidation of the key enzymes will help to achieve more efficient clean-up strategies. Structural and biochemical studies of the catabolic enzymes involved in the biodegradation pathway will help in achieving the ultimate aim of making the surrounding environment free from the toxic aromatics (Figure 1-1).

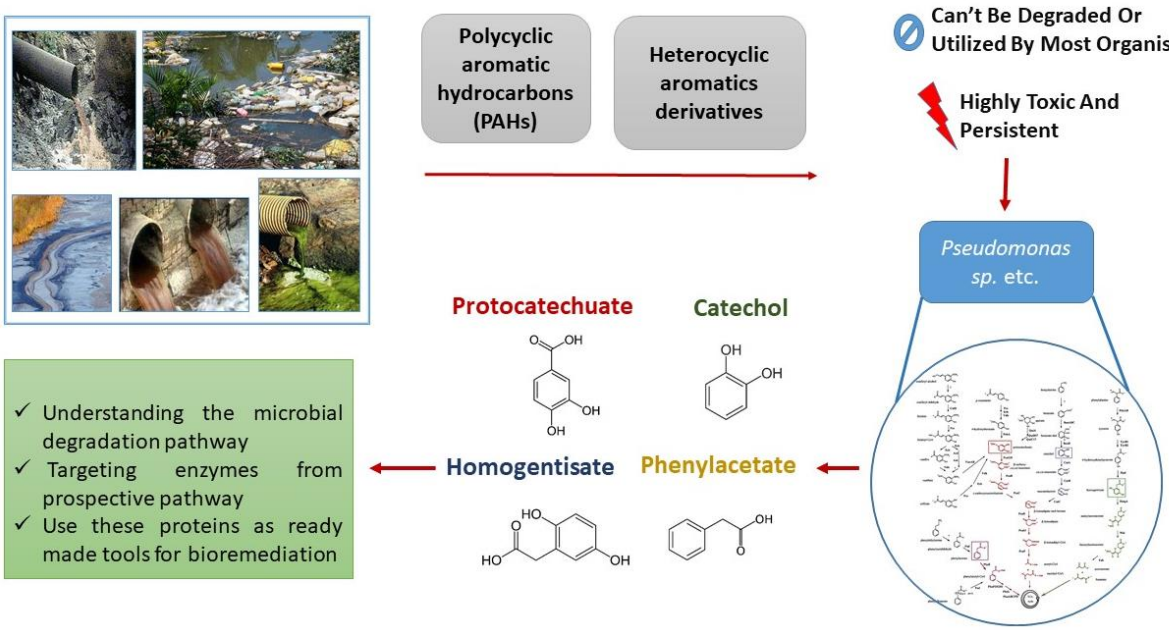

Figure 1-1 The conceptual diagram of the project in nutshell.

1.2 Scope of the present study

Pseudomonas putida belongs to the class of bacterial species which can efficiently metabolize various types of aromatic hydrocarbons such as phenol and its methylated derivatives, m, p- and o-cresol, and 3, 4 dimethyl phenol (Y. H. Kim et al. 2006; Williams & Sayers 1994). In addition, they can uptake a wide variety of aromatic compounds as the sole source of carbon and energy, including benzolate, protocatechuate, p-hydroxybenzoate and m-toluate (Jiménez et al. 2002). We have chosen *Pseudomonas putida* KT2440 as a model bacterial system to target a few important catabolic pathway enzymes for more in-depth understanding into their structure/function and activity. Selected strains of *P. putida* harbour certain regulatory proteins (DmpR, XylR) that can detect and capture the water-based aromatic pollutants such as phenol, toluene, xylene, and catechol. These regulatory proteins on binding with the respective aromatic pollutants activate transcription of specific catabolic pathway genes/operons which can finally degrade the captured pollutants to simpler non-toxic TCA cycle intermediates (Allende, Gibello, Martin, & Garrido-Pertierra 1992; Locher, Leisinger, & Cook 1991). The focus of this thesis is to understand the mechanisms of the catabolic enzymes from this bacterium and their mode of action in the presence of their substrates. This will enable us to understand various steps and the substrate scope of these intermediate enzymes. These enzymes or a combination

of them can be used in the future as a synthetic route to degrade pollutants in water. An *in vitro* synthetic cassette can be generated which can convert toxic starting compounds into the non-lethal TCA intermediates. Thus, we are chartering our way in this direction by studying the mechanism of action and the substrate scope of two of the important enzymes in the degradation pathway. Overall, the long-term aim is to develop novel approaches that help in removing these pollutants from the polluted water.

The β-Ketoadipate pathway is extensively spread across various group of eubacteria and fungi. It plays a major part in the bioremediation of the aromatic pollutants as well as it helps in degradation of naturally occurring aromatic compounds resulting from lignin, etc. This pathway coexists in bacterial strains which metabolizes chlorinated aromatic compounds, because the last step of the chlorinated catechols degradation are catalyze by the concluding two enzymes, PcaIJ and PcaF, of the β-ketoadipate pathway. This highlights the importance of the β-ketoadipate pathway. Additionally, the β-ketoadipate pathway is present in varied species of soil bacteria, and in each group the pathway has a typical regulatory characteristic and gene organization. Therefore, β-ketoadipate pathway is a classical system to study how a series of enzymes can diversify without losing its catalytic function. Such information are useful in understanding in what manner nature adapts a pathway to aid the existence of a specific bacterium.

Aromatic bioremediation enzymes have a clear-cut role in fulfilling the degradation of aromatic compounds. In this study, two key enzymes are explored. The first one is a thiolase enzyme annotated as β-ketoadipyl-CoA thiolase (PcaF) which catalyzes the breakdown of β-ketoadipyl-CoA into succinyl-CoA and acetyl-CoA (TCA cycle intermediates). PcaF is a pivotal enzymatic player of the pathway as it concludes the biocatalysis of the toxic compounds from catechol and protocatechuate pathway into the non-lethal metabolic intermediates, which enter the TCA cycle and aid the bacteria to produce energy for its permissibility(Harwood et al. 1994). The last step of phenylacetate pathway is also determined by a thiolase (PhaD) further implicating the importance of the thiolase enzymes in the microbial degradation of aromatic impurities (De Eugenio et al. 2010; Nogales et al. 2007).

γ-carboxymuconolactone decarboxylase (γ-CMD) which is annotated as PcaC is the second enzyme of interest because of its strategic positioning in the β-ketoadipate pathway. Bacteria via the catechol or protocatechuate branch of β-ketoadipate pathway degrade many aromatic

compounds. β-ketoadipate enol lactone, the catalytic product of γ-CMD is the first common product of the two branches, namely, protocatechuate and catechol branches (Lee & Yang 2009). The β-ketoadipate pathway which subsequently gets converted into TCA cycle intermediates via 3-4 more steps that are governed by enzymes common to both the pathways. γ-CMD catalyzes the conversion of γ-carboxymuconolactone to β-ketoadipate enol lactone and releases CO_2 as the by-product (Ornston 1966b).

The structure and function of these two enzymes PcaF and PcaC have been explored by using mutagenesis, activity assay, mass-spectrometry, crystallographic, and bioinformatics studies. This helps in understanding the uniqueness of the bioremediation thiolase PcaF with respect to the active site architecture and provides holistic insight about the interaction of PcaF with xenobiotic acyl CoA and in identifying active sites of PcaC (representing γ-CMD family) and their role in catalysis including understanding its enzymatic mechanism.

1.3 Organization of the thesis

This thesis is organized into six chapters. The outline of each chapter is described below.

The thesis begins with the introductory **Chapter-1**, which reflects the background and motivation of the research. Here the crucial issue of water pollution and the need for eradication of aromatic pollutants is addressed. The chapter further provides insight into the microbial degradation pathways that can be used as bioremediation tool. The scope of the study is established by highlighting the various studies that are planned. The chapter concluded with the arrangement of this thesis.

We enter **Chapter-2** wherein, the literature survey of the microbial degradation pathway and the role of regulatory proteins in triggering these pathways, is carried out. The four major pathways: homogentisate pathway (*hmg/fah/mai* genes), the phenylacetate pathway (*pha* genes), the catechol (*cat* genes) and the protocatechuate (*pca* genes) pathway are explained. The emphasis is given on the catechol and protocatechuate pathways as they converge form the central β-ketoadipate pathway of bacterial aromatic degradation pathway. Two enzymes, β-keto acyl thiolase (PcaF) and γ-carboxymuconolactone decarboxylase (PcaC), from this pathway are discussed, and their importance are explained. A background is given on the thiolase and

decarboxylase classes of enzymes. The chapter is concluded by highlighting the lacuna in current research.

Chapter-3 includes the experimental methods and the bioinformatic tools used in the study of PcaF and PcaC enzymes. The chapter is divided into two sections. The first section enumerates the details involved in cloning, site-directed mutagenesis, purification, enzymatic assay, crystallization, cryo-protection, X-ray data collection and structure determination (molecular replacement phasing, density modification, model building, and refinement) and structural analysis of PcaF. The bioinformatics software is used to identify and calculation of the tunnel present in the PcaF. The second section discusses the protein purification, mutagenesis, crystallization of PcaC and cryo-protection and data collection from native and mutant PcaC crystals, structure determination (Single anomalous diffraction (SAD) phasing, density modification, model building, and refinement) in detail. The software tools employed for structure solving and refinement of apo and ligand bound PcaC structures are explained. The novel "bicarbonate soaking" method to trace the carbon dioxide release pathway is explicitly described. The mass-based confirmatory assay for the determination of active site residue is elaborated. The substrate is modelled in the active site using bioinformatics tools, which is explained in detail.

Chapter-4 includes the explicit study done on the first enzyme of interest, PcaF. The chapter starts with a background study of thiolase group of enzymes, and the research gaps are explored. One of the apparent research gaps is the lack of knowledge about degradative thiolase, which suits our research purpose as PcaF is a degradative thiolase. In this chapter, the main focus is to understand the degradative thiolase using PcaF as a standard and pinpoint differentiating patterns between two groups of thiolase. To address the objectives, bioinformatics studies are done on the active site architecture of PcaF followed by mutation of the active site residues. The crystal structures of the PcaF variant with CoA and variety of CoA derivatives shed light on the active site pocket, longer CoA binding tunnel and novel loop present in PcaF. The multiple ligand-bound structures obtained also demonstrates the promiscuous nature of PcaF thiolase. In this chapter, the overall understanding of the thiolase family including the biosynthetic and degradative thiolase is highlighted.

Chapter-5 contains the detailed study of the second protein of interest, PcaC. The chapter begins with the basis of the protein, highlighting the γ-CMD subfamily of protein which PcaC

belong to. As not much published information is available on γ-CMD, the main aim of the project is to identify the active site, decipher the mechanism of PcaC and increase the repertoire of γ-CMD family. The research gap is addressed by using bioinformatics studies, structural analysis, and activity assay. The correlation between PcaC and other CMDs protein is highlighted. The novel bicarbonate soaking method has been developed in the identification of the active site and the product pathway of PcaC. Mutation of the active site residues followed by mass spectrometry-based assay further confirms the residues involved in the reaction. The ligand-bound structures involving PcaC mutants with substrate and product analog along with the overall analysis have assisted in postulating a mechanism for PcaC.

Chapter-6 provides the summary of this thesis along with the important conclusions. The implications of this work are mentioned in explicit details. The major contributions made by the author in unravelling the enzymes involved in the bioremediation of aromatic pollutants are described. The directions for future works that are essential to develop biosensors are provided. The publications and patent arising out of this thesis work are listed at the end of this chapter.

1.4 References

Abdel-shafy, HI & Mansour, MSM 2016, 'REVIEW A review on polycyclic aromatic hydrocarbons : Source , environmental impact , effect on human health and remediation' *Egyptian Journal of Petroleum*, vol. 25, no. 1, pp. 107–123.

Allende, JL, Gibello, A, Martin, M, & Garrido-Pertierra, A 1992, 'Transport of 4-hydroxyphenylacetic acid in *Klebsiella pneumoniae*' *Archives of Biochemistry and Biophysics*, vol. 292, no. 2, pp. 583–588.

Boretti, A & Rosa, L 2019, 'Reassessing the projections of the world water development report' *Npj Clean Water*, vol. 2, no. 15.

Buchan, A, Collier, LS, Neidle, EL, & Moran, MA 2000, 'Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine *Roseobacter* lineage' *Applied and Environmental Microbiology*, vol. 66, no. 11, pp. 4662–4672.

De Eugenio, LI, Galán, B, Escapa, IF, Maestro, B, Sanz, JM, García, JL, & Prieto, MA 2010, 'The PhaD regulator controls the simultaneous expression of the pha genes involved in polyhydroxyalkanoate metabolism and turnover in *Pseudomonas putida* KT2442' *Environmental Microbiology*, vol. 12, no. 6, pp. 1591–1603.

Díaz, E & Prieto, MA 2000, 'Bacterial promoters triggering biodegradation of aromatic pollutants' *Current Opinion in Biotechnology*, vol. 11, no. 5, pp. 467–475.

Gong, X, Shen, Z, Zhang, Q, Zeng, Y, Sun, J, Ho, SSH, Lei, Y, Zhang, T, Xu, H, Cui, S, Huang, Y, & Cao, J 2019, 'Characterization of polycyclic aromatic hydrocarbon (PAHs) source pro files in urban PM 2.5 fugitive dust : A large-scale study for 20 Chinese cites' *Science of the Total Environment*, vol. 687, pp. 188–197.

Harayama, S, Kok, M, & Neidle, EL 1992, 'Functional and evolutionary relationships among diverse oxygenases' *Annual Review of Microbiology*, vol. 46, pp. 565–601.

Harwood, CS, Nichols, NN, Kim, M, Diitty, JL, & Parales, RE 1994, 'Identification of the *pcaRKF* gene cluster from *Pseudomonas putida*: Involvement in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate' *Journal of Bacteriology*, vol. 176, no. 21, pp. 6479–6488.

Helmer, R, Hespanhol, I, Nations, U, Programme, E, & Council, SC 1997, *Water Pollution Control - A Guide to the Use of Water Quality Management Principles*.

Jiménez, JI, Miñambres, B, García, JL, & Díaz, E 2002, 'Genomic analysis of the aromatic

catabolic pathways from *Pseudomonas putida* KT2440' *Environmental Microbiology*, vol. 4, no. 12, pp. 824–841.

Kim, D, Kim, Y-S, Kim, S-K, Kim, SW, Zylstra, GJ, Kim, YM, & Kim, E 2002, 'Monocyclic aromatic hydrocarbon degradation by Rhodococcus sp. strain DK17' *Applied and Environmental Microbiology*, vol. 68, no. 7, pp. 3270–3278.

Kim, YH, Cho, K, Yun, SH, Kim, JY, Kwon, KH, Yoo, JS, & Kim, SI 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*, vol. 6, pp. 1301–1318.

Kjellstrom, T, Lodh, M, Mcmichael, T, Shrestha, R, & Kingsland, S 2006, 'Air and water Pollution : Burden and strategies for control' in D. Jamison, J. Breman, A. Measham, G. Alleyne, M. Claeson, D. Evans, P. Jha, A. Mills, & P. Musgrove (eds.), *Disease Control Priorities in Developing Countries*, 2nd editio, pp. 817–832, New York: Oxford University Press, Washington (DC): World Bank.

Lee, HY & Yang, JK 2009, 'crystallization communications Crystallization and preliminary X-ray crystallographic analysis of γ-carboxymucolactone decarboxylase from Sulfolobus solfataricus crystallization communications' *Acta Crystallographica Section F Structural Biology CommunicationsStructural Biology and Crystallization Communications*, vol. 65, pp. 1197–1199.

Locher, HH, Leisinger, T, & Cook, AM 1991, '4-Toluene sulfonate methyl-monooxygenase from *Comamonas testosteroni* T-2: Purification and some properties of the oxygenase component' *Journal of Bacteriology*, vol. 173, no. 12, pp. 3741–3748.

Nelson, KE, Weinel, C, Paulsen, IT, Dodson, RJ, Hilbert, H, Martins dos Santos, VAPP, Fouts, DE, Gill, SR, Pop, M, Holmes, M, Brinkac, L, Beanan, M, DeBoy, RT, Daugherty, S, Kolonay, J, Madupu, R, Nelson, W, White, O, Peterson, J, Khouri, H, Hance, I, Chris Lee, P, Holtzapple, E, Scanlan, D, Tran, K, Moazzez, A, Utterback, T, Rizzo, M, Lee, K, Kosack, D, Moestl, D, Wedler, H, Lauber, J, Stjepandic, D, Hoheisel, J, Straetz, M, Heim, S, Kiewitz, C, Eisen, J, Timmis, KN, Düsterhöft, A, Tümmler, B, & Fraser, CM 2002, 'Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440' *Environmental Microbiology*, vol. 4, no. 12, pp. 799–808.

Nogales, J, Macchi, R, Franchi, F, Barzaghi, D, Fernández, C, García, JL, Bertoni, G, & Díaz, E 2007, 'Characterization of the last step of the aerobic phenylacetic acid degradation

pathway' *Microbiology,* vol. 153, pp. 357–365.

Ornston, LN 1966a, 'The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida* II. Enzymes of the protocatechuate pathway' *The Journal of Biological Chemistry,* vol. 241, no. 16, pp. 3800–3810.

Ornston, LN 1966b, 'The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida* III. Enzymes of the catechol pathway' *The Journal of Biological Chemistry,* vol. 241, no. 16, pp. 3795–3799.

Ornston, LN 1966c, 'The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida* IV. Regulation' *The Journal of Biological Chemistry,* vol. 241, no. 16, pp. 3800–3810.

Shen, X, Huang, Y, & Liu, S 2005, 'Genomic analysis and identification of catabolic pathways for aromatic compounds in *Corynebacterium glutamicum*' *Microbes and Environments,* vol. 20, no. 3, pp. 160–167.

Tropel, D & Meer, JR Van Der 2004, 'Bacterial transcriptional regulators for degradation pathways of aromatic compounds' *Microbiology and Molecular Biology Reviews,* vol. 68, no. 3, pp. 474–500.

Wang, Z, Chen, J, Yang, P, Qiao, X, & Tian, F 2007, 'Polycyclic aromatic hydrocarbons in Dalian soils: Distribution and toxicity assessment' *Journal of Environmental Monitoring,* vol. 9, no. 2, pp. 199–204.

WHO/UNICEF Joint Monitoring Programme 2005, *Water for life: Making it happen WHO,* Geneva: WHO Press.

WHO 1998, *Polynuclear aromatic hydrocarbons in drinking-water World Health Organization,* vol. 2.

WHO 2008, *Guidelines for drinking-water quality,* vol. 1.

Wild, SR, Berrow, ML, & Jones, KC 1991, 'The persistence of polynuclear aromatic hydrocarbons (PAHs) in sewage sludge amended agricultural soils' *Environmental Pollution,* vol. 72, pp. 141–157.

Williams, PA & Sayers, JR 1994, 'The evolution of pathways for aromatic hydrocarbon oxidation in *Pseudomonas*' *Biodegradation,* vol. 5, no. 3–4, pp. 195–217.

Wu, X, Monchy, S, Taghavi, S, Zhu, W, Ramos, J, & van der Lelie, D 2011, 'Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*' *FEMS Microbiology Reviews,* vol. 35, pp. 299–323.

Yamanashi, T, Kim, S-Y, Hara, H, & Funa, N 2015, 'In vitro reconstitution of the catabolic

reactions catalyzed by PcaHG, PcaB, and PcaL: the protocatechuate branch of the β-ketoadipate pathway in Rhodococcus jostii RHA1.' *Bioscience, Biotechnology, and Biochemistry*, vol. 79, no. 5, pp. 830–835.

Yang, Y, Zhang, N, Xue, M, Lu, ST, & Tao, S 2011, 'Effects of soil organic matter on the development of the microbial polycyclic aromatic hydrocarbons (PAHs) degradation potentials' *Environmental Pollution*, vol. 159, no. 2, pp. 591–595.

Yang, Y, Zhang, N, Xue, M, & Tao, S 2010, 'Impact of soil organic matter on the distribution of polycyclic aromatic hydrocarbons (PAHs) in soils' *Environmental Pollution*, vol. 158, no. 6, pp. 2170–217.

Chapter 2
Microbial degradation of aromatic compounds

2.1 Introduction

One of the most prevalent organic compounds, apart from carbohydrate, is aromatic hydrocarbons. They have both natural as well as man-made origins and are part of the earth natural carbon cycle. They contribute to biosphere by the exhaust fumes of volcanic eruption, forest fires and the like (Gibson 1980; Kozak et al. 2017). Human activities such as industrialization, mining, urbanization have also added to the presence of the aromatic compounds (Ho, Show, Guo, & Norli 2012). One of the pertinacious aromatic compounds found in the nature is polycyclic aromatic compounds (PAHs) and its heterocyclic derivatives. Industries such as aluminium, petrochemical, asphalt, rubber tire etc. emit a sustainable amount of these aromatic compounds into the water bodies and terrestrial environment. These chemical compounds constitute major part of water contaminants and are registered as priority pollutants in the European Community and the U.S. Environmental Protection Agency (Gong et al. 2019). They are carcinogenic, mutagenic and lethal in nature and effect the ecosystem adversely. Often, these pollutants reach beyond the permissible exposure limit (PEL) in the aquatic system and pollute the drinking water. Thus, there is an imminent need for elimination of these toxic aromatic compound from the environment. In the last few decades there has been a growing interest to find out sustainable methods to eradicate pollutants and aromatic compounds from the environment (Abdel-shafy & Mansour 2016). Many bacterial species like *Pseudomonas Sp*., *Acinetobacter Sp*. has the metabolic capability to uptake these aromatic pollutants as their carbon and energy source (Jiménez, Miñambres, García, & Díaz 2002; Y. H. Kim et al. 2006; P. A. Williams & Sayers 1994). Utilizing these bacteria for the bioremediation of aromatic pollutants seem a rational eco-friendly approach (Smith 1990).

Both aerobic and anaerobic microorganisms are known to assimilate aromatic compounds, but detailed studies have been conducted for the aerobic pathway (Van der Meer, De Vos, Harayama, & Zehnder 1992; Widdel & Rabus 2001). Microbial degradation of these aromatic pollutants generally includes two components (Figure 2-1):

(i) Transcriptional regulatory elements -The regulatory component is made of promoter and proteins which control the expression of the enzymes to avoid uneconomical use of genetic material and employ regulator to firmly exercise control over the product formation (Cases & de Lorenzo 1998). These bacterial transcriptional regulators are involved in the expression of the genes and operons giving rise to specific downstream enzymes (Tropel & Meer 2004). They bind to the operator region of the promoter to express the genes encoding the pathway enzymes (Decker & Hinton 2013). The induction of the enzyme pathway is switched on when adequate level of inducer is present to avoid detrimental energy fluxes (Shingler 2003). The inducer are usually substrates, substrate analogue or intermediate. The regulators may behave as both transcriptional activator and repressor depending upon the inducer existence. (Browning & Busby 2004). For example, in *Pseudomonas CF600* the dmp operon encoding methyl-phenol degradative enzymes are activated by DmpR by acting on the Pdmp promoter in presence of phenol (Wise & Kuske 2000) . Different derivatives of phenol can also act as inducer, but their degree varies according to the chemical nature of the phenol ring (Sarand et al. 2001). XylR from *Pseudomonas putida mt-2* senses the presence of BTEX (benzene, toluene, ethylbenzene and xylene) to activate the Pu promoter which give rise to the enzymes involved in the conversion of m-xylene into m-methylbenzoate (M. N. Kim, Park, Lim, & Shin 2005; Ramos, Marqués, & Timmis 1997). XylR exhibits poor affinity toward phenol derivatives and degrade hydrocarbon-based aromatics. MopR from *A. calcoaceticus NCIB8250* initiate the mop operon encoding phenol hydroxylase and the *catA* encoding catechol 1,2-dioxygenase (Abril, Michan, Timmis, & Ramos 1989).

(ii) Catabolic pathway enzymes- These enzymes are required for the catabolic breakdown of the aromatic compounds and are only expressed in the presence of the aromatic compounds or their degradation pathway intermediates (Fuchs, Boll, & Heider 2011). The aromatic compounds act as inducers, bind to the regulators and trigger the synthesis of the catabolic enzymes (D. Kim et al. 2002; Navarro-Llorens et al. 2005). These enzymes mineralize these compounds into simple, non-toxic TCA- cycle intermediates and use them as their energy source, thus allowing the bacteria to

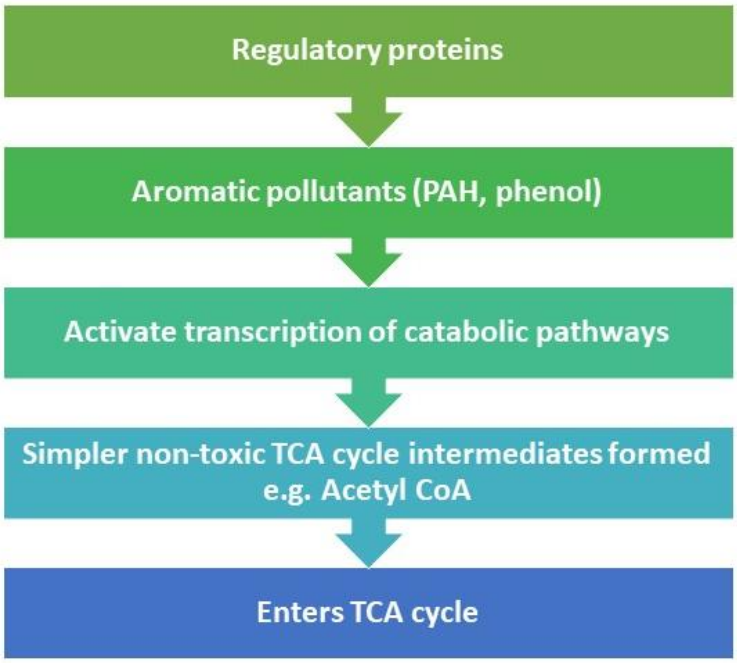

Figure 2-1 The flow diagram of the two components 'transcriptional regulatory elements' and 'catabolic pathway enzymes' of the microbial degradation of these aromatic pollutants

survive in hostile surroundings (Smith 1990). Usually the enzymes belong to degradative pathways which are characterized in detail and some of these enzymes are also outlined in the Biocatalysis/Biodegradation Database (Ellis, Roe, & Wackett 2006). There are four key degradative pathways and few minor pathways. The rule of thumb for these degradative pathways involve ring modification via a mono- or dioxygenation reaction followed by ring fission which consequently give rise to the TCA- intermediates (Schlömann 1994; Van der Meer et al. 1992).

2.2 Microbial Aromatic degradative pathways in *Pseudomonas putida* KT2440

In this thesis *Pseudomonas putida* KT2440, a soil bacterium, is chosen as the model organism. Studies and analysis of the entire genome of *P.putida* KT2440 propose that the catabolism of the various aromatic compounds occur mainly via four major pathways- the phenylacetate pathway (*paa* genes), the homogentisate pathway (*hmg/fah/mai* genes), the protocatechuate (*pca* genes) and the catechol (*cat* genes) pathways (Figure 2-1). Both protocatechuate and the catechol pathways

form the two parallel branches of a common β-ketoadipate pathway (Parales & Harwood 1992; Paralest & Harwood 1993). In addition to these central pathways, few peripheral catabolic pathways are also present in this organism which are- the *nic* cluster and the *pcm* genes both of which might be involved in the catabolism of N-heterocyclic aromatic compounds, pathways involved in the catabolism of p-hydroxybenzoate (*pob*), benzoate (*ben*), quinate (*qui*), phenylpropenoid compounds (*fcs, ech, vdh, cal, van, acd and acs*), phenylalanine and tyrosine (*phh, hpd*) and n-phenylalkanoic acids (*fad*) (Jiménez et al. 2002). The four key pathways are further explained in detail

2.2.1 Phenylacetate pathway

In *Pseudomonas putida*, the phenylacetate pathway encompasses fourteen different genes that are involved in three connecting operons (Navarro-Llorens et al. 2005). The genes comprise of five functional components namely a transport system, an activating enzyme, a ring hydroxylation complex, ring opening enzyme and a β-oxidation-type unit. Two regulatory genes are involved in encoding a transcriptional repressor and a protein of unknown function(Grishin & Cygler 2015; Teufel et al. 2010). This pathway act as the chief pathway, wherein pathways catabolizing structurally similar compounds e.g. phenylalanine, 2-phenylethylamine, styrene and ethylbenzene converge into it. This whole ensemble including the joining pathways is termed as a catabolon and the phenylacetate pathway form the catabolon core of the phenylacetyl-CoA catabolon. The catabolon thus help in biotransformation of various minor pathways through a common intermediate(Nogales et al. 2007; Song et al. 2006). Phenylacetatic acid converts to phenylacetyl-coenzyme A, followed by ring hydroxylation and the ring cleavage (Figure 2-2). This pathway is characterized in gram-negative bacteria, unlike, in gram-positive bacteria where the pathway is not well understood (Niraula, Shrestha, Oh, & Sohng 2010; Olivera et al. 1998).

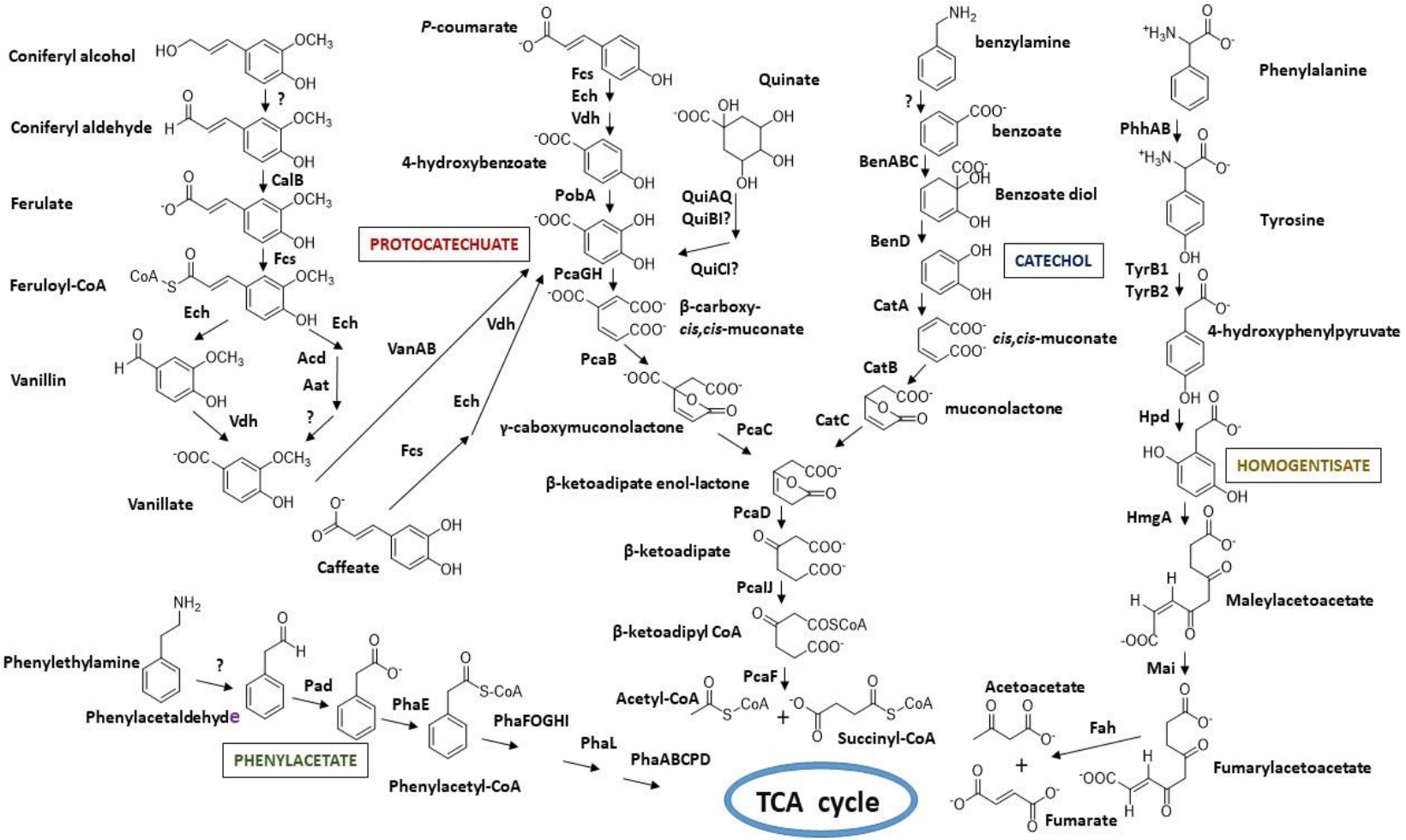

Figure 2-2 The aromatic catabolic pathways from Pseudomonas putida KT2440

.

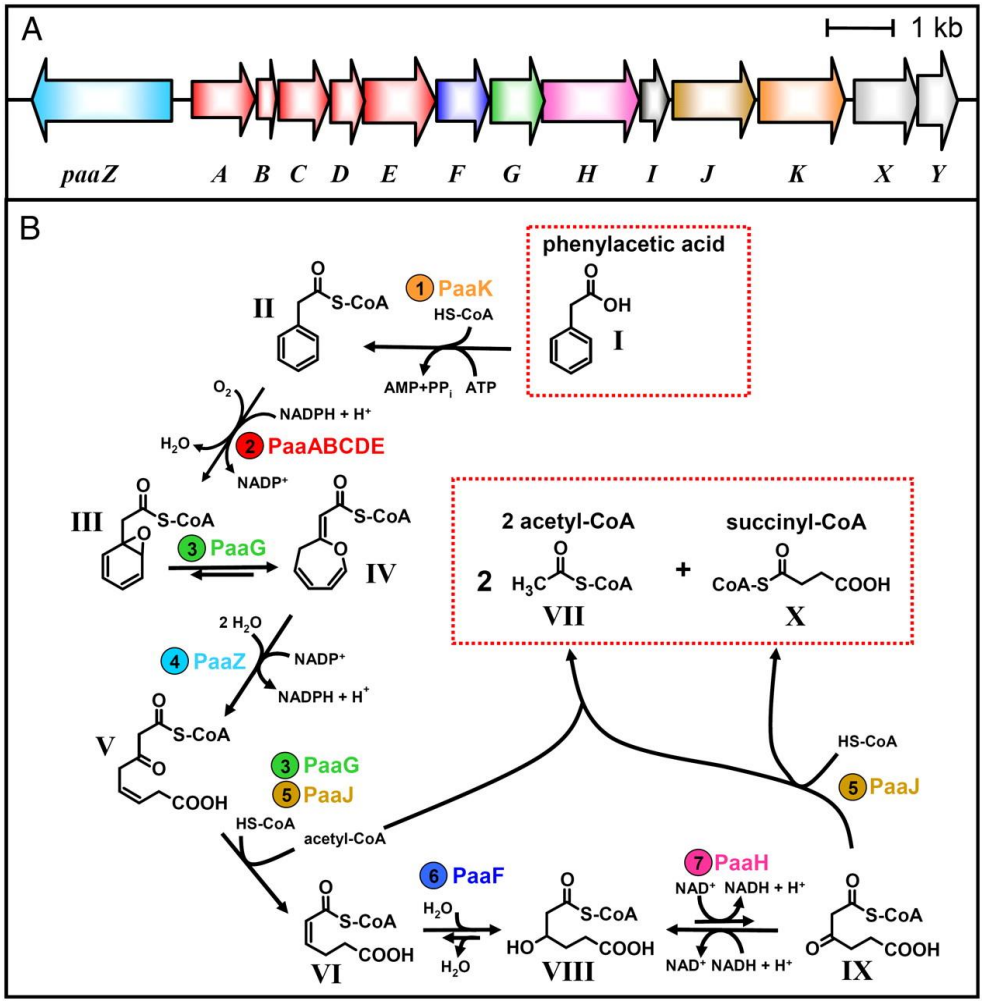

Figure 2-3 Aerobic phenylacetate catabolic pathway. (A) Catabolic gene cluster for phenylacetate degradation in E. coli K12. (B) Reactions and intermediates of the pathway as studied in E. coli K12 and Pseudomonas sp. strain Y2. Proposed enzyme names (Table S1): 1: phenylacetate-CoA ligase (AMP forming); 2: ring 1,2-phenylacetyl-CoA epoxidase (NADPH); 3: ring 1,2-epoxyphenylacetyl-CoA isomerase (oxepin-CoA forming), postulated 3,4-dehydroadipyl-CoA isomerase. 4: oxepin-CoA hydrolase/ 3-oxo-5,6-dehydrosuberyl-CoA semialdehyde dehydrogenase (NADP+); 5: 3-oxoadipyl-CoA/ 3-oxo-5,6-dehydrosuberyl-CoA thiolase; 6: 2,3-dehydroadipyl-CoA hydratase; 7: 3-hydroxyadipyl-CoA dehydrogenase (NAD+) (probably (S)-3-specific). Compounds: I, phenylacetate; II, phenylacetyl-CoA; III, ring 1,2-epoxyphenylacetyl-CoA; IV, 2-oxepin-2(3H)-ylideneacetyl-CoA; V, 3-oxo-5,6-dehydrosuberyl-CoA, VI, 2,3-dehydroadipyl-CoA; VII, acetyl-CoA; VIII, 3-hydroxyadipyl-CoA; IX, 3-oxoadipyl-CoA; X, succinyl-CoA).(Adapted from R. Teufel, V. Mascaraque, W. Ismail, M. Voss, J. Perera, W. Eisenreich, W. Haehnel, G. Fuchs, Bacterial phenylalanine and phenylacetate catabolic pathway revealed, 107 (2010). doi:10.1073/pnas.1005399107)

2.2.2 Homogentisate pathway

Phenylalanine and Tyrosine are assimilated by another pathway known as the homogentisate pathway. In this pathway the phenylalanine is hydroxylated via phenylalanine hydroxylase (PhhA) to tyrosine followed by tyrosine aminotransferase (TyrB), that catalyze the tyrosine into 4-hydroxyphenylpyruvate and lastly the 4-hydroxyphenylpyruvate is catalyzed by 4-hydroxyphenylpyruvate dioxygenase (Hpd) to homogentisate (Arias-Barrau et al. 2004; Méndez, Agulló, González, & Seeger 2011). This homogentisate form a central metabolite that is broken down by three enzymes, homogentisate dioxygenase (HmgA), fumarylacetoacetate hydrolase (HmgB), and maleylacetoacetate isomerase (HmgC) which result in final product of fumarate and acetoacetate (Arias-Barrau et al. 2005; Fernandez-Canon & Penalva 1995; Hareland, Crawford, Chapman, & Dagley 1975; Morales et al. 2004) (Figure 2-4). The *phh, tyr*, and *hpd* genes are found to be independent unlike the *hmgABC* genes which are part of one transcriptional unit (Arcos et al. 2010). The *hmgR* gene is found to encode a distinct repressor that regulates the expression of *hmgABC* genes, whereas the homogentisate form the inducer molecule for the system. HmgR bind to Phmg promoter around a region that consist of a 17-bp palindromic motif and is the first IclR-type of regulator that behave as a repressor of an aromatic degradation pathway (Morales et al. 2004; Sparnins & Dagley 1975).

2.2.3 β-ketoadipate pathway

The catechol and protocatechuate pathway converge to form the central β-ketoadipate pathway which is a taxonomically widespread bioremediation pathway in bacteria (L. N. Ornston 1966b, 1966c; L N Ornston 1966a). Usually the pathways employ two-part strategy involving ring cleavage by diverse ring alteration reactions followed by the ring fission reactions to degrade these aromatic pollutants into TCA intermediates (Buchan, Collier, Neidle, & Moran 2000; Harwood & Parales 1996). In case of the β-ketoadipate pathway, the ring fission leads to the

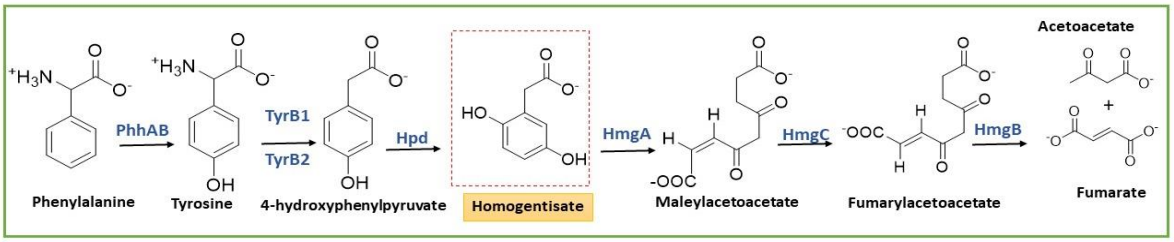

Figure 2-4 Aerobic homogentisate catabolic pathway

formation of the crucial β-ketoadipate via an ortho-cleavage reaction which undergoes further two steps to give rise to the TCA intermediate (L Nicholas Ornston 1971; Powlowski, Ingebrand, & Dagley 1985) (Figure 2-5). The protochatechute branch of the pathway metabolize phenolic compound such as p-cresol, 4-hydroxybenzoate etc. whereas the catechol branch is involved with aromatic hydrocarbons, lignin monomer and so on (Milstein et al. 1983; Nichols & Harwood 1995). Studies show that the pathway is conserved across bacterial species but demonstrates diverse combination of isozymes, branch congregation points, and regulatory proteins in different bacteria (Powlowskit & Dagley 1985). The pathway is well studied in gram negative and nitrogen fixing species and is also observed in eukaryotic *ascomycetous* and *basidiomycetous* yeasts and fungi (Parke & Ornston 1986). The catechol branch of the eukaryotic members is similar to that of the prokaryotes, unlike the protocatechuate branch where the product of protocatechuate ring cleavage is cyclized. The catechol and protocatechuate branch also join at different point i.e. they converge at β-ketoadipate instead of β-ketoadipate enol-lactone position (Middelhoven 1993; Milstein et al. 1983).

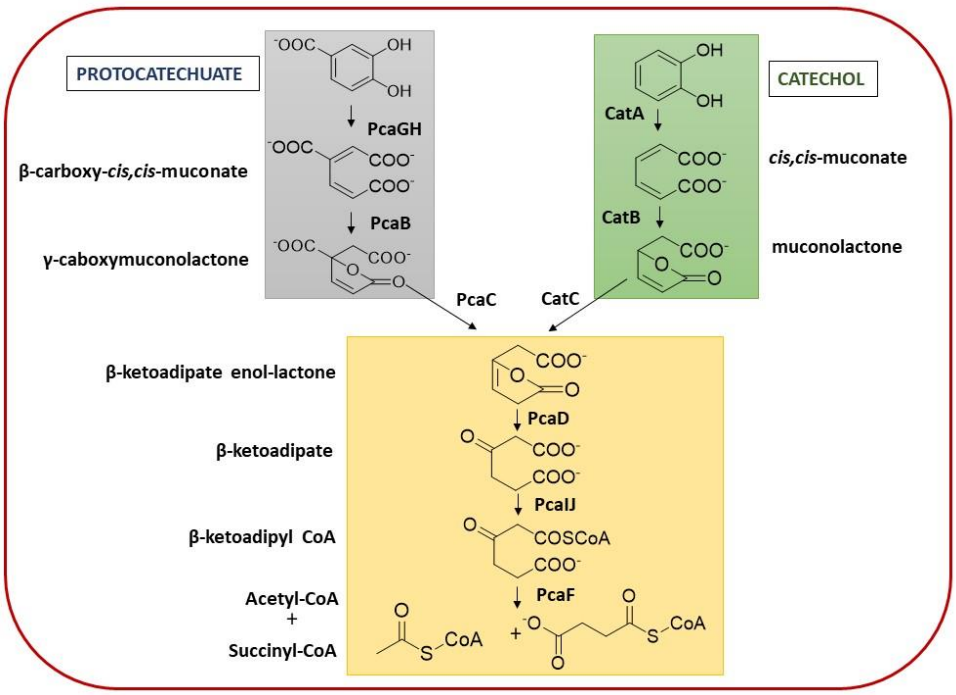

Figure 2-5 The β-ketoadipate pathway catabolic pathway

2.3 Enzymes of the β-ketoadipate pathway

The biochemistry and enzymology of β-ketoadipate pathway in *P.putida* were established by L.N. Ornston through four milestone paper in 1966. The two convergent branches, catechol and protocatechuate, were also outlined and fully characterized in these papers(L. N. Ornston 1966c, 1966b). The first enzyme of the protocatechuate pathway is protocatechuate 3, 4-dioxygenase, encoded by the *pcaGH* genes(Whittaker, Lipscombsli, Kentll, & Munck 1984). The oligomeric structure of the protein has two protomers, α (PcaG, 23kDa) and β (PcaH, 27kDa) (Ohlendorf, Orville, & Lipscomb 1994). It converts protocatechuate into β-carboxy-cis, cis-muconate by incorporation of oxygen (Bull & Ballout 1981). The next enzyme of the pathway is PcaB, a 42 kDa protein encoded by *pcaB* (Meagher & Ornston 1973; S. E. Williams et al. 1992). It is a β-carboxy-cis,cis-muconate lactonizing enzyme that yield the next product, γ-caboxymuconolactone (Kowalchuk et al. 1994; J. Yang et al. 2004). PcaC, a 15kDa γ-caboxymuconolactone decarboxylase protein encoded by *pcaC* further react with the γ-caboxymuconolactone to produce β-ketoadipate enol-lactone, which mark the first common product of catechol and protocatechuate pathways (Hartnett, Neidle, Ngai, & Ornston 1990; Parke, Meagher, & Ornston 1973; W. Yeh, Fletcher, & Ornston 1980a).

The catechol pathway begins with the 32 kDa catechol 1, 2-dioxygenase, CatA encoded by *catA* gene, that catalyzes the ring fission of the catechol to cis,cis-muconate (Nakaia, Nakazawa, & Nozaki 1988). Both CatA and PcaG enzymes share about 25% identity and contain conserved ferric iron, tyrosine and histidine at their active site (Meagher & Ornston 1973). Unlike PcaGH, CatA is usually homodimer and is classified as type I (Earhart et al. 2005). Type II catechol 1, 2-dioxygenases, is involved in the catalysis of chlorinated catechols (Nakai et al. 1995). The next enzyme in the catechol pathway is cis,cis-muconate lactonizing 40 kDa enzyme annotated as CatB which is encoded by *catB* gene(Aldrich & Chakrabarty 1988). Both the catechol and protocatechuate pathway comprise of lactonizing enzyme that has similar 3D structure but vary in their mode of action. PcaB catalyzes anti-cycloisomerization reaction and belong to class II fumarase family of proteins whereas CatB mediates a syn addition to the double bond of cis,cis-muconate and belong to mandelate racemase (Goldman, Ollis, & Steitz 1987; Meagher & Ornston 1973). The third enzyme of the catechol pathway is CatC, encode by *catC* gene(Shanley et al. 1994)(Shanley et al. 1994). It is an 11kDa muconolactone isomerase enzyme that act on muconolactone to give β-ketoadipate enol-lactone, the con-joint product of

the β-ketoadipate pathway (Katti, Katz, & Wyckoff 1989). The PcaC and CatC enzymes have opposite stereo facial specificities for their substrates (Chari, Whitman, Kozarich, et al. 1987; Chari, Whitman, & Kozarich 1987). The next enzyme is a 29kDa protein PcaD encoded by *pcaD* (Mccorkle, Yeh, Fletcher, & Ornston 1980). It forms the first common enzyme of the β-ketoadipate pathway and catalyzes the lactone hydrolyzation of β-ketoadipate enol-lactone to β-ketoadipate (W. Yeh & Ornston 1984). The penultimate step of the pathway is governed by PcaIJ enzyme, which convert β-ketoadipate to β-ketoadipyl CoA (Paralest & Harwood 1993). This is a heterodimeric protein comprising of the 24 kDa PcaI protein forming the subunit α and the 22kDa PcaJ protein giving rise to the β subunit (Parales & Harwood 1992; W. Yeh & Ornston 1981). The last step of the pathway is directed by 43kDa protein PcaF. It is encoded by *pcaF* genes and it catalyzes the breakdown of β-ketoadipyl-CoA into succinyl-CoA and acetyl-CoA(Harwood et al. 1994).

In this thesis, the enzymes PcaF and PcaC have been studied in detail. PcaF belong to the thiolase superfamily of enzyme and PcaC fall under carboxymuconolactone decarboxylase family of protein. Literature review on these two family of enzymes have been carried out in the following sections.

2.4 Overview of thiolase family

CoA is a vital co-factor involved in the transfer of acetyl group in a range of small molecules and macromolecules. It uses its terminal thiol group to form thioester linkages with carboxylic acid groups and it partakes in diverse metabolic pathways. Fatty acid metabolism is such an example of CoA-dependent pathway where the fatty acid moiety is attached to the SH of a phosphopantetheine group to form acyl CoA. In the fatty acid β-oxidation pathways, the acyl CoA undergoes chemical conversion via four steps resulting in the cleavage of a carbon–carbon bond to yield acetyl-CoA and a chain-shortened acyl-CoA. In the bacterial degradation pathway, the last step of the β-ketoadipate pathway is catalyzed by thiolase which also utilizes CoA for its reaction and produces acetyl-CoA and succinyl CoA. The acetyl CoA molecule is a key intermediate molecule that is precursor to other pathways such as TCA cycle, glycolysis. For example, in the TCA cycle, the acetyl-CoA is oxidized to generate energy.

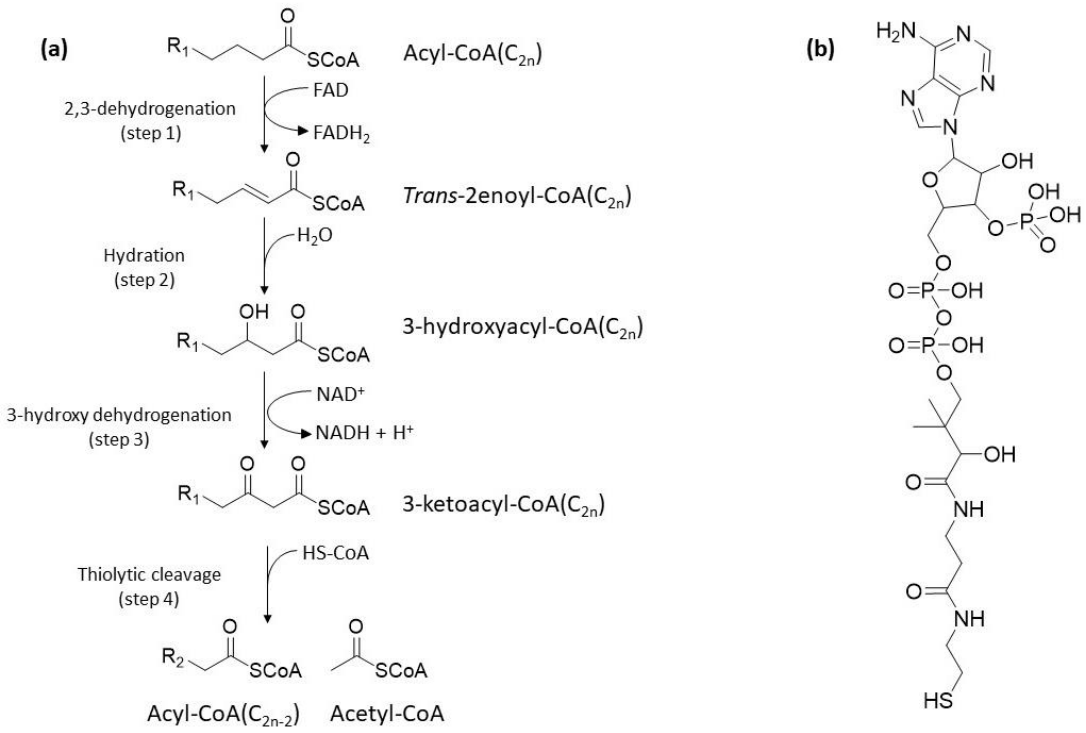

Figure 2-6(a) The overview of the fatty acids β-oxidation pathway. It is a CoA dependent fatty acyl-CoA degradation pathway involving four steps, which results in acetyl-CoA formation. Acetyl-CoA is used as a substrate for the TCA or glycoxylate cycles, or for fatty acid production. (b) The chemical structure of Coenzyme A (CoA).

Fatty acids are crucial for the permissibility of the biological spectrum and its catabolic breakdown via β-oxidation pathway is associated with energy production in cells (Bhaumik et al. 2005; Schulz 1991). The β-oxidation pathway of fatty acid is a four steps process involving shortening of the fatty acid chain by two carbon and concomitant production of one $FADH_2$ and NADH each (Houten & Wanders 2010). Enzymes regulating the significant fatty acid metabolism are of high research interest and can be used as potential drug target interest. Inhibitors against the first three enzymes of the spiral pathway are well explored, leaving the last thiolytic step to be studied in detail (Borges, Glauert, & Robertson 1993; Broadway et al. 1999; Ghisla & Thorpe 2004) Thiolases generally are a group of very significant enzymes that are found in both prokaryotes and eukaryotes and take part in many vital biochemical pathways. Thiolase enzymes, also known as acetyl-coenzyme A acetyl transferases (ACAT), takes part in the fatty acid degradation beta oxidation pathway as well as in the biosynthetic pathways involved in the fatty acid and polyketide biogenesis (Figure 2-6). There are two broad categories in which the thiolase family of enzymes can be subdivided.

i) **Degradative thiolase** (EC 2.3.1.9) or the β-ketoacyl CoA thiolase (also called thiolase I) which is mainly involved in the β-oxidation pathway of fatty acid degradation and catalyzes the cleavage of β-ketoacyl CoA into smaller acetyl-CoA and another short acyl-CoA moiety (Staack, Binstock, & Schulz 1978). This particular reaction occurs in the final step in the β- oxidation pathway, and accounts for the extraction of the metabolic energy stored in fats and fatty acids. The thiolytic cleavage reaction usually requires the presence of Coenzyme A (CoASH) as co-substrate (S.-Y. Yang et al. 1990).

ii) **Biosynthetic thiolase** (EC 2.3.1.16) or the acetoacetyl-CoA thiolase (also called thiolase II) which is primarily involved in certain biosynthetic pathways like poly β-hydroxybutyric acid synthesis and specifically catalyses the thiolysis of acetoacetyl-CoA (Davis et al. 1987; Modis & Wierenga 1999). The key step in the biosynthesis direction is the formation of a carbon–carbon bond formation via a thioester-dependent Claisen condensation reaction mechanism. This is the first step in many biosynthetic pathways and one of the basic carbon skeletal assembly patterns in biological system (Igual, González-Bosch, Dopazo, & Pérez-Ortín 1992).

2.4.1 Biosynthetic and degradative Thiolase: A Comparison

The biosynthetic and bio degradative thiolase enzymes demonstrate high level of sequence similarities and also depict a similar three-dimensional structure (Gilbert, Lennox, Mossman, & Carle 1981). It is found that both the classes of thiolase possess same active site residues and

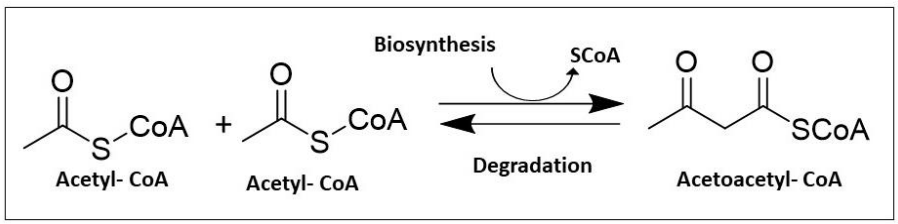

Figure 2-7 Reaction catalysed by the thiolase

mechanism of action has some common elements. Nonetheless, these two classes differ in their substrate specificity (Mathieu et al. 1997). The degradative thiolase can accommodate long acyl chain substrates ranging from 4 to 22 carbons while the biosynthetic ones prefer short acyl chains up to 4 carbon atoms (Feigenbaum & Schulz 1975; Kursula et al. 2005). It has also been observed that all thiolase has the ability to catalyze both the biosynthetic and degradative reaction however, the reaction that occurs in the direction of biosynthesis is thermodynamically unfavorable (Ithayaraja et al. 2016; Meriläinen, Poikela, Kursula, & Wierenga 2009; Modis & Wierenga 1999). It is hypothesized that the substrate binding pocket of biosynthetic thiolase is small compared to degradative thiolase as it only accepts small acyl moieties. The CoA molecule bind in a tight fit near the catalytic Cys89 and the absence of any cavities in this region provides a simple explanation for the inability of biosynthetic thiolases to utilize larger fatty acyl substrates (Kursula, Ojala, Lambeir, & Wierenga 2002).Thiolase is a well-studied family of protein but till date it remains elusive how thiolases might biosynthetic and degradative ones differ and what determine the reaction in a particular direction. Any key residues or structural pattern is not identified that might help to distinguish between biosynthetic and degradative thiolase therefore this is an important area to focus.

2.5 Overview of Carboxymuconolactone decarboxylase (CMD) family

Carboxymuconolactone decarboxylase (CMD) belong to decarboxylase family of protein and based on their function is divided into three subfamilies; the alkyl hydroperoxidases, the γ-carboxymuconolactone decarboxylase and third represented by *Thermus thermophiles* HB8. The first subfamily is alkyl hydroperoxidases or AhpD, which play a crucial part in the detoxification of peroxidase and other reactive oxygen species (Venkatesan et al. 2016). The AphD contain six helices ranging from α1 to α6 wherein, the N terminal “V-shape” motif is formed by the α1 and α2 helices. It is followed by α3 to α5 helices, which give rise to the CMD core domain that is involved in the formation of the structural core region and stable multimerizations (Chen et al. 2015; Clarke et al. 2011). The terminal α6 results in the C terminal helix. The AphD active site contain a thioredoxin like motif called the CXXC- motif which is responsible for the peroxidase like activity (Hillas et al. 2000). The CXXC-motif is located in the α4 helix. AphD is part of an AhpC/AhpD consort system, where the AhpC or alkyl hydroperoxide reductase is a non-heme peroxidase (Lessner & Ferry 2007). It has a conserved

N terminal cysteine which is oxidized during the peroxidase reaction. The AphD here acts as an adaptor protein and help in reduction of the cysteine in the final step of reaction, thus finishing the peroxidase activity (Koshkin et al. 2004). The mode of action of AphD has been studied in detail and a proton relay-based mechanism has been hypothesized. Five residues, two histidine, one glutamate and two cysteine from CXXC motif and one structural water molecule is proposed to be part of the proton relay (Koshkin, Nunn, Djordjevic, & Montellano 2003). Known structures for the AhpD subfamily includes PDB codes 1KNC, 1GU9, 1ME5, which belong to *Mycobacterium tuberculosis* (Nunn et al. 2002).

The second subfamily of CMD proteins is called γ-carboxymuconolactone decarboxylase (γ-CMD). It is part of the microbial aromatic degradation of the pathway. It is known to catalyze γ-caboxymuconolactone to β-ketoadipate enol-lactone in the protocatechuate side of the β-ketoadipate pathway (W. K. Yeh & Ornston 1982). Not much is known about the protein, except that it exists in hexametric oligomeric state and lacks the CXXC- motif (Parke et al. 1973; W. Yeh, Durham, Fletcher, & Ornston 1981). It shares very low sequence identity with the AphD subfamily (Lee & Yang 2009). The last subfamily consists of a distinct member protein TTHA0727 from *Thermus thermophiles* HB8. This protein also lacks the CXXC- motif and hence cannot perform peroxidase activity. The protein consists of 117 amino acid and has only 35% identity with the AphD core motif. The monomer comprises of seven helices (α1 to α7) and one small 3_{10}-helix instead of all six helices fold of AphD (Ito et al. 2006). It also hexamerizes via interaction of the α5 to α7 helices and a small tunnel is formed at the middle of the hexameric ring. When compared with the AphD active site, it was observed that a serine and a cysteine residue is present at the equivalent position and the site of the residues were at the bottom cleft of the protein. It has been hypothesized that this cleft region might be the supposed binding site for small molecules. The PDB code for TTHA0727 is 2CWQ. Interestingly, all the members of CMD family still undergo hexameric or trimeric arrangement using the CMD core, but the functionality of the proteins varies (W. Yeh, Fletcher, & Ornston 1980b).

2.6 Research gap and objectives

In this thesis two enzymes are selected for study and they both belong to the central β-ketoadipate pathway. The first protein is β-keto acyl CoA thiolase (PcaF), which catalyzes the

last step of the β-ketoadipate pathway and converts β-ketoadipyl-CoA into succinyl-CoA and acetyl-CoA (De Eugenio et al. 2010; Navarro-Llorens et al. 2005) (Figure 2-7). Second protein of interest, γ-caboxymuconolactone decarboxylase (PcaC), belong to the protocatechuate branch of the β-ketoadipate pathway. It is decarboxylase and act on γ-carboxymuconolactone to form β–ketoadipate enol lactone and releases CO_2 as the by-product (W. Yeh et al. 1981) (Figure 2-8). These two proteins are located at two crucial point in the pathway, thus understanding these two catabolic enzymes will aid in development of better bioremediation designs. As a first step towards for developing strategies to remove these polycyclic aromatic hydrocarbons (PAHs) and heterocyclic aromatics from the environment, understanding the mechanisms of the catabolic enzymes and its mode of action with the substrate is essential. The main aims of this project are included in below sections.

2.6.1 β-keto acyl CoA thiolase (PcaF)

- **The active-site features that differentiate the biosynthetic and degradative thiolases.** The structural differences between the two types of thiolases is not well known. The aim of the project is to identify the residues which play an important role in differentiating the two classes of enzyme.

- **How the degradative thiolases act on a wide variety of substrate.** Degradative thiolase is known to bind longer chain substrates than the biosynthetic thiolase. Till date, no studies have been performed which discuss how the long chain CoA binds to the degradative thiolase.

- **Structural feature which play role in distinguishing the two classes of thiolases.** Enumerate, different structural features that will help in identifying the two class of thiolases.

2.6.2 γ-carboxymuconolactone decarboxylase (PcaC)

- **Understanding the active site architecture.** The active site location of γ-carboxymuconolactone decarboxylase (γ-CMD) is unknown. Very few papers that has been published on γ-CMD deals with establishing the role of γ-CMD in the protocatechuate catabolism. Identification of active site and residues will vastly contribute to the existing knowledge of γ-CMD.

- **Hypothesis of Mechanism.** It will be stimulating to get ligand bound structures of PcaC as it will help in identifying the active site. And based on the active site residues mechanism can be proposed.

- **Evolutionary relationship with the other CMD proteins.** As mention, the CMD family has evolved to perform diverse function yet maintaining the hexameric or trimeric oligomeric structure. It will be interesting if connection can be made between the various CMD proteins and the functioning.

In order to achieve these goals, the enzyme need to be cloned, purified, and crystallized. Determination of the 3D structures of both the apo and ligand bound form will help in deciphering their mechanism and functions. Subsequently, high sensitivity bioassays can be developed to study the biochemical characters of these proteins. These bioassays will help to determine the activity of the protein against the ligand and thus indicating whether the protein can degrade the said ligand efficiently or not.

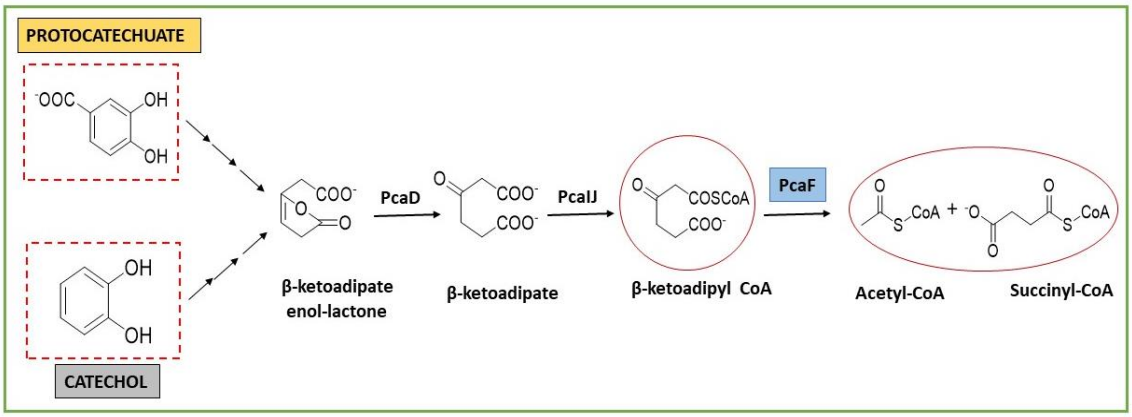

Figure 2-8 Schematic representation of β-ketoadipate pathway which is preceded by the thiolase PcaF in the last step.

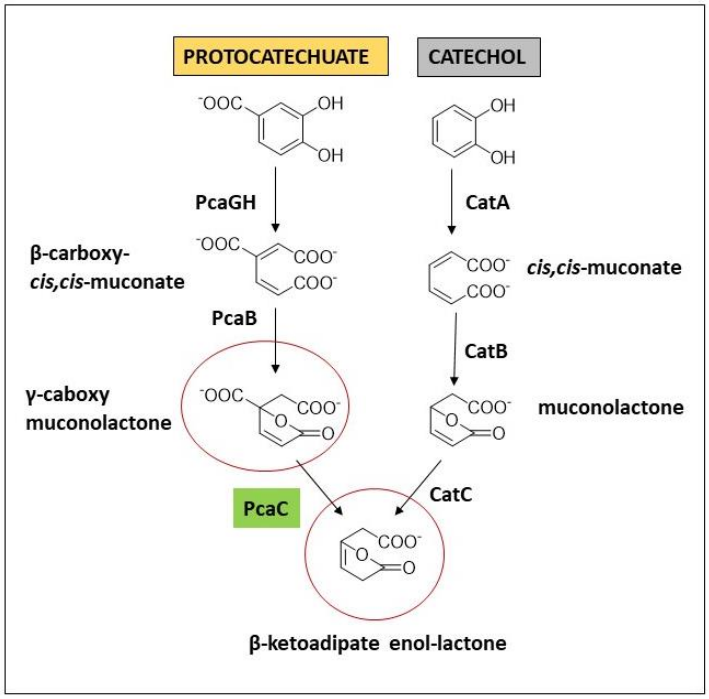

Figure 2-9 Schematic representation of the joining of protocatechuate and catechol pathways which is preceded by the γ-CMD, PcaC, in the merger step

2.7 References

Abdel-shafy, HI & Mansour, MSM 2016, 'REVIEW A review on polycyclic aromatic hydrocarbons : Source , environmental impact , effect on human health and remediation' *Egyptian Journal of Petroleum*, vol. 25, no. 1, pp. 107–123.

Abril, M-A, Michan, C, Timmis, KN, & Ramos, JL 1989, 'Regulator and enzyme specificities of the TOL plasmid-encoded upper pathway for degradation of aromatic hydrocarbons and expansion of the substrate range of the pathway' *Journal of Bacteriology*, vol. 171, no. 12, pp. 6782–6790.

Aldrich, TL & Chakrabarty, AM 1988, 'Transcriptional eegulation, nucleotide sequence, and localization of the promoter of the catBC operon in *Pseudomonas putida*' *Journal of Bacteriology*, vol. 170, no. 3, pp. 1297–1304.

Arcos, M, Olivera, ER, Arias, S, Naharro, G, & Luengo, JM 2010, 'The 3,4-dihydroxyphenylacetic acid catabolon, a catabolic unit for degradation of biogenic amines tyramine and dopamine in Pseudomonas putida U' *Environmental Microbiology*, vol. 12, no. 6, pp. 1684–1704.

Arias-Barrau, E, Olivera, ER, Luengo, JM, Ferna´ndez, C, Gala´n, B, Garcı´a, JL, Dı´az, E, & Min˜ambres, B 2004, 'The homogentisate pathway: A central catabolic pathway involved in the degradation of L-phenylalanine, L-tyrosine, and 3-hydroxyphenylacetate in *Pseudomonas putida*' *Journal of Bacteriology*, vol. 186, no. 15, pp. 5062–5077.

Arias-Barrau, E, Sandoval, Á, Naharro, G, Olivera, ER, & Luengo, JM 2005, 'A two-component hydroxylase involved in the assimilation of 3-hydroxyphenyl acetate in Pseudomonas putida' *The Journal of Biological Chemistry*, vol. 280, no. 28, pp. 26435–26447.

Bhaumik, P, Koski, MK, Glumoff, T, Hiltunen, JK, & Wierenga, RK 2005, 'Structural biology of the thioester-dependent degradation and synthesis of fatty acids' *Current Opinion in Structural Biology*, vol. 15, pp. 621–628.

Borges, T, Glauert, HP, & Robertson, LW 1993, 'Perfluorodecanoic acid noncompetitively inhibits the peroxisomal enzymes of enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase' *Toxicology and Applied Pharmacology*, vol. 118, pp. 8–15.

Broadway, NM, Engel, PC, Bahnson, BJ, Harijan, RK, Li, D, Agnihotri, G, Dakoji, S, Oh, E, Lantz, M, Liu, HW, Hiltunen, JK, Wierenga, RK, Li, D, Agnihotri, G, Dakoji, S, Oh, E, Lantz, M, & Liu, HW 1999, 'The toxicity of methylenecyclopropylglycine: Studies of the inhibitory effects of (methylenecyclopropyl)formyl-CoA on enzymes involved in fatty acid metabolism and the molecular basis of its inactivation of enoyl-CoA hydratases' *Journal of the American Chemical Society*, vol. 121, no. 39, pp. 9034–9042.

Browning, DF & Busby, SJW 2004, 'The regulation of bacterial transcription initiation' *Nature Reviews Microbiology*, vol. 2, no. 1, pp. 57–65.

Buchan, A, Collier, LS, Neidle, EL, & Moran, MA 2000, 'Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine *Roseobacter* lineage' *Applied and Environmental Microbiology*, vol. 66, no. 11, pp. 4662–4672.

Bull, C & Ballout, DP 1981, 'Purification and properties of protocatechuate 3, 4-dioxygenase from *Pseudomonas putida*: A new iron to subunit stoichiometry' *The Journal of Biological Chemistry*, vol. 256, no. 24, pp. 12673–12680.

Cases, I & de Lorenzo, V 1998, 'Expression systems and physiological control of promoter activity in bacteria.' *Current Opinion in Microbiology*, vol. 1, pp. 303–310.

Chari, RVJ, Whitman, CP, & Kozarich, JW 1987, 'Absolute stereochemical course of muconolactone Δ-isomerase and of 4-carboxymuconolactone decarboxylase: a 1H NMR "ricochet" analysis' *Journal of the American Chemical Society*, vol. 109, pp. 5520–5521.

Chari, RVJ, Whitman, CP, Kozarich, JW, Ngai, K, & Ornston, LN 1987, 'Absolute stereochemical course of the 3-carboxymuconate cycloisomerases from Ppseudomonas putida and Acinetobacter calcoaceticus: Analysis and implications' *Journal of the American Chemical Society*, vol. 109, pp. 5514–5519.

Chen, X, Hu, Y, Yang, B, Gong, X, Zhang, N, Niu, L, Wu, Y, & Ge, H 2015, 'Structure of lpg0406, a carboxymuconolactone decarboxylase family protein possibly involved in antioxidative response from Legionella pneumophila' *Protein Science*, vol. 24, pp. 2070–2075.

Clarke, TE, Romanov, V, Chirgadze, YN, Klomsiri, C, Kisselman, G, Wu-brown, J, Poole, LB, Pai, EF, & Chirgadze, NY 2011, 'Crystal structure of alkyl hydroperoxidase D like protein PA0269 from Pseudomonas aeruginosa : Homology of the AhpD-like structural family' *BMC Structural Biology*, vol. 11, no. 27.

Davis, JT, Moore, RN, Imperiali, B, Pratt, AJ, Kobayashi, K, Masamune, S, Sinskey, AJ, Walsh, CT, Fukui, T, & Tomita, K 1987, 'Biosynthetic thiolase from Zoogloea ramigera I. Preliminary characterization and analysis of proton transfer reaction' *The Journal of Biological Chemistry*, vol. 262, no. 1, pp. 82–89.

De Eugenio, LI, Galán, B, Escapa, IF, Maestro, B, Sanz, JM, García, JL, & Prieto, MA 2010, 'The PhaD regulator controls the simultaneous expression of the pha genes involved in polyhydroxyalkanoate metabolism and turnover in *Pseudomonas putida* KT2442' *Environmental Microbiology*, vol. 12, no. 6, pp. 1591–1603.

Decker, KB & Hinton, DM 2013, 'Transcription regulation at the core: Similarities among bacterial, archaeal, and eukaryotic RNA polymerases' *Annual Review of Microbiology*, vol. 67, no. 1, pp. 113–139.

Earhart, CA, Vetting, MW, Gosu, R, Michaud-soret, I, Que, L, & Ohlendorf, DH 2005, 'Structure of catechol 1,2-dioxygenase from *Pseudomonas arvilla*' *Biochemical and Biophysical Research Communications*, vol. 338, pp. 198–205.

Ellis, LBM, Roe, D, & Wackett, LP 2006, 'The University of Minnesota biocatalysis/biodegradation database: The first decade' *Nucleic Acids Research*, vol. 34, no. 90001, pp. D517–D521.

Feigenbaum, J & Schulz, H 1975, 'Thiolases of *Escherichia coli*: Purification and chain length specificities' *Journal of Bacteriology*, vol. 122, no. 2, pp. 407–411.

Fernandez-Canon, JM & Penalva, MA 1995, 'Molecular characterization of a gene encoding a homogentisate dioxygenase from *Aspergillus nidulans* and identification of its human and plant homologues' *The Journal of Biological Chemistry*, vol. 270, no. 36, pp. 21199–21205.

Fuchs, G, Boll, M, & Heider, J 2011, 'Microbial degradation of aromatic compounds — from one strategy to four' *Nature Reviews Microbiology*, vol. 9, pp. 803–816.

Ghisla, S & Thorpe, C 2004, 'Acyl-CoA dehydrogenases: A mechanistic overview' *European Journal of Biochemistry*, vol. 271, pp. 494–508.

Gibson, D. 1980, *Microbial Metabolism.* ed. O. Hutzinger, *The Handbook of Environmental Chemistry 2 Part A - Reactions and Processes*, Springer Verlag, Berlin.

Gilbert, HF, Lennox, BJ, Mossman, CD, & Carle, WC 1981, 'The relation of acyl transfer to the overall reaction of thiolase I from porcine heart' *The Journal of Biological Chemistry*, vol. 256, no. 14, pp. 7371–7377.

Goldman, A, Ollis, DL, & Steitz, TA 1987, 'Crystal structure of muconate lactonizing at 3Å resolution' *Journal of Molecular Biology*, vol. 194, no. 1, pp. 143–153.

Gong, X, Shen, Z, Zhang, Q, Zeng, Y, Sun, J, Ho, SSH, Lei, Y, Zhang, T, Xu, H, Cui, S, Huang, Y, & Cao, J 2019, 'Characterization of polycyclic aromatic hydrocarbon (PAHs) source pro files in urban PM 2.5 fugitive dust : A large-scale study for 20 Chinese cites' *Science of the Total Environment*, vol. 687, pp. 188–197.

Grishin, AM & Cygler, M 2015, 'Structural organization of enzymes of the phenylacetate catabolic hybrid pathway' *Biology*, vol. 4, no. 2, pp. 424–442.

Hareland, WA, Crawford, RL, Chapman, PJ, & Dagley, S 1975, 'Metabolic Function and Properties of 4-Hydroxyphenylacetic Acid 1-Hydroxylase from Pseudomonas acidovorans', vol. 121, no. 1, pp. 272–285.

Hartnett, C, Neidle, EL, Ngai, K, & Ornston, LN 1990, 'DNA sequences of genes encoding *Acinetobacter calcoaceticus* protocatechuate 3, 4-dioxygenase: Evidence indicating

shuffling of genes and of DNA sequences within genes during their evolutionary divergence' *Journal of Bacteriology*, vol. 172, no. 2, pp. 956–966.

Harwood, CS, Nichols, NN, Kim, M, Diitty, JL, & Parales, RE 1994, 'Identification of the *pcaRKF* gene cluster from *Pseudomonas putida*: Involvement in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate' *Journal of Bacteriology*, vol. 176, no. 21, pp. 6479–6488.

Harwood, CS & Parales, RE 1996, 'The β-ketoadipate pathway and the biology of self-identity' *Annual Review of Microbiology*, vol. 50, pp. 553–590.

Hillas, PJ, Soto, F, Oyarzabal, J, Wilks, A, & Montellano, PRO De 2000, 'The AhpC and AhpD antioxidant defense system of Mycobacterium tuberculosis' *The Journal of Biological Chemistry*, vol. 275, no. 25, pp. 18801–18809.

Ho, YC, Show, KY, Guo, XX, & Norli, I 2012, 'Industrial discharge and their effect to the environment' in K.-Y. Show & Xinxin Guo (eds.), *Industrial Waste*, pp. 1–32, InTech.

Houten, SM & Wanders, RJA 2010, 'A general introduction to the biochemistry of mitochondrial fatty acid β-oxidation' *Journal of Inherited Metabolic Disease*, vol. 33, pp. 469–477.

Igual, JC, González-Bosch, C, Dopazo, J, & Pérez-Ortín, JE 1992, 'Phylogenetic analysis of the thiolase family. Implications for the evolutionary origin of peroxisomes' *Journal of Molecular Evolution*, vol. 35, no. 2, pp. 147–155.

Ithayaraja, M, Janardan, N, Wierenga, RK, Savithri, HS, & Murthy, MRN 2016, 'Crystal structure of a thiolase from *Escherichia coli* at 1.8 Å resolution' *Acta Crystallographica Section F Structural Biology Communications*, vol. 72, no. 7, pp. 534–544.

Ito, K, Arai, R, Fusatomi, E, Kamo-Uchikubo, T, Kawaguchi, S, Akasaka, R, Terada, T, Kuramitsu, S, Shirouzu, M, & Yokoyama, S 2006, 'Crystal structure of the conserved protein TTHA0727 from *Thermus thermophilus* HB8 at 1.9 Å resolution: A CMD family member distinct from carboxymuconolactone decarboxylase (CMD) and AhpD' *Protein*

Science, vol. 15, pp. 1187–1192.

Jiménez, JI, Miñambres, B, García, JL, & Díaz, E 2002, 'Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440' *Environmental Microbiology*, vol. 4, no. 12, pp. 824–841.

Katti, SIL, Katz, BA, & Wyckoff, HW 1989, 'Crystal structure of muconolactone at 3.3Å resolution isomerase' *Journal of Molecular Biology*, vol. 205, no. 3, pp. 557–571.

Kim, D, Kim, Y-S, Kim, S-K, Kim, SW, Zylstra, GJ, Kim, YM, & Kim, E 2002, 'Monocyclic aromatic hydrocarbon degradation by Rhodococcus sp. strain DK17' *Applied and Environmental Microbiology*, vol. 68, no. 7, pp. 3270–3278.

Kim, MN, Park, HH, Lim, WK, & Shin, HJ 2005, 'Construction and comparison of *Escherichia coli* whole-cell biosensors capable of detecting aromatic compounds' *Journal of Microbiological Methods*, vol. 60, no. 2, pp. 235–245.

Kim, YH, Cho, K, Yun, SH, Kim, JY, Kwon, KH, Yoo, JS, & Kim, SI 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*, vol. 6, pp. 1301–1318.

Koshkin, A, Nunn, CM, Djordjevic, S, & Montellano, PRO De 2003, 'The mechanism of *Mycobacterium tuberculosis* alkylhydroperoxidase ahpd as defined by mutagenesis, crystallography, and kinetics' *The Journal of Biological Chemistry*, vol. 278, no. 32, pp. 29502–29508.

Koshkin, A, Zhou, X, Kraus, CN, Brenner, JM, Bandyopadhyay, P, Kuntz, ID, Barry III, CE, & Montellano, PRO De 2004, 'Inhibition of *Mycobacterium tuberculosis* AhpD , an element of the peroxiredoxin defense against oxidative stress' *Antimicrobial Agents and Chemotherapy*, vol. 48, no. 7, pp. 2424–2430.

Kowalchuk, GA, Hartnett, GB, Benson, A, Houghton, JE, Ngai, K-L, & Ornston, LN 1994, 'Contrasting patterns of evolutionary divergence within *Acinetobacter calcoaceticus pca*

operon' *Gene*, vol. 146, pp. 23–30.

Kozak, K, Ruman, M, Kosek, K, Karasiński, G, Stachnik, Ł, & Polkowska, Z 2017, 'Impact of volcanic eruptions on the occurrence of PAHs compounds in the aquatic ecosystem of the southern part of West Spitsbergen (Hornsund Fjord, Svalbard)' *Water (Switzerland)*, vol. 9, no. 1, p. 42.

Kursula, P, Ojala, J, Lambeir, A-MM, & Wierenga, RK 2002, 'The catalytic cycle of biosynthetic thiolase: A conformational journey of an acetyl group through four binding modes and two oxyanion holes' *Biochemistry*, vol. 41, no. 52, pp. 15543–15556.

Kursula, P, Sikkilä, H, Fukao, T, Kondo, N, & Wierenga, RK 2005, 'High resolution crystal structures of human cytosolic thiolase (CT): A comparison of the active sites of human CT, bacterial thiolase, and bacterial KAS I' *Journal of Molecular Biology*, vol. 347, no. 1, pp. 189–201.

Lee, HY & Yang, JK 2009, 'crystallization communications Crystallization and preliminary X-ray crystallographic analysis of γ-carboxymucolactone decarboxylase from Sulfolobus solfataricus crystallization communications' *Acta Crystallographica Section F Structural Biology CommunicationsStructural Biology and Crystallization Communications*, vol. 65, pp. 1197–1199.

Lessner, DJ & Ferry, JG 2007, 'The archaeon Methanosarcina acetivorans contains a protein disulfide reductase with an iron-sulfur cluster' *Journal of Bacteriology*, vol. 189, no. 20, pp. 7475–7484.

Mathieu, M, Modis, Y, Zeelen, JP, Engel, CK, Abagyan, RA, Ahlberg, A, Rasmussen, B, Lamzin, VS, Kunau, WH, & Wierenga, RK 1997, 'The 1.8 A crystal structure of the dimeric peroxisomal 3-ketoacyl-CoA thiolase of Saccharomyces cerevisiae: implications for substrate binding and reaction mechanism' *Journal of Molecular Biology*, vol. 273, no. 3, pp. 714–28.

Mccorkle, GM, Yeh, W, Fletcher, P, & Ornston, LN 1980, 'Repetitions in the NH2-terminal amino acid sequence of β-ketoadipate enol-lactone hydrolase from *Pseudomonas putida*'

The Journal of Biological Chemistry, vol. 255, no. 13, pp. 6335–6341.

Meagher, RB & Ornston, LN 1973, 'Relationships among enzymes of the β-ketoadipate pathway I. Properties of cis,cis-muconate-lactonizing enzyme and muconolactone isomerase from Pseudomonas putida' *Biochemistry*, vol. 12, no. 18, pp. 3523–3530.

Méndez, V, Agulló, L, González, M, & Seeger, M 2011, 'The homogentisate and homoprotocatechuate central pathways are involved in 3- and 4-hydroxyphenylacetate degradation by *Burkholderia xenovorans* LB400' *PLoS ONE*, vol. 6, no. 3, p. e17583.

Meriläinen, G, Poikela, V, Kursula, P, & Wierenga, RK 2009, 'The thiolase reaction mechanism: The importance of Asn316 and His348 for stabilizing the enolate intermediate of the Claisen condensation' *Biochemistry*, vol. 48, no. 46, pp. 11011–11025.

Middelhoven, WJ 1993, 'Catabolism of benzene compounds by ascomycetous and basidiomycetous yeasts and yeastlike fungi' *Antonie van Leeuwenhoek*, vol. 63, no. 2, pp. 125–144.

Milstein, O, Vered, Y, Shragina, L, Gressel, J, Flowers, HM, & Hfittermann, A 1983, 'Metabolism of lignin related aromatic compounds by *Aspergillus japonicus*' *Archives of Microbiology*, vol. 135, pp. 147–154.

Modis, Y & Wierenga, RK 1999, 'A biosynthetic thiolase in complex with a reaction intermediate: The crystal structure provides new insights into the catalytic mechanism' *Structure*, vol. 7, no. 10, pp. 1279–1290.

Morales, G, Linares, JF, Beloso, A, Albar, JP, Martínez, JL, & Rojo, F 2004, 'The *Pseudomonas putida* Crc global regulator controls the expression of genes from several chromosomal catabolic pathways for aromatic compounds' *Journal of Bacteriology*, vol. 186, no. 5, pp. 1337–1344.

Nakai, C, Uyeyama, H, Kagamiyama, H, Nakazawa, T, Inouye, S, Kishi, F, Nakazawa, A, & Nozaki, M 1995, 'Cloning, DNA sequencing, and amino acid sequencing of catechol

1,2-dioxygenases (pyrocatechase) from *Pseudomonas putida* mt-2 and *Pseudomonas arvilla* C-1' *Archives of Biochemistry and Biophysics*, vol. 321, no. 2, pp. 353–362.

Nakaia, C, Nakazawa, T, & Nozaki, M 1988, 'Purification and properties of catechol 1, 2-dioxygenase from *Pseudomonas putida* mt-2 in comparison with that from *Pseudomonas arvilla* C-1' *Archives of Biochemistry and Biophysics*, vol. 267, no. 2, pp. 701–713.

Navarro-Llorens, JM, Patrauchan, MA, Stewart, GR, Davies, JE, Eltis, LD, & Mohn, WW 2005, 'Phenylacetate catabolism in *Rhodococcus* sp. strain RHA1: A central pathway for degradation of aromatic compounds' *Journal of Bacteriology*, vol. 187, no. 13, pp. 4497–4504.

Nichols, NN & Harwood, CS 1995, 'Repression of 4-hydroxybenzoate transport and degradation by benzoate: A new layer of regulatory control in the *Pseudomonas putida* β-ketoadipate pathway' *Journal of Bacteriology*, vol. 177, no. 24, pp. 7033–7040.

Niraula, NP, Shrestha, P, Oh, T, & Sohng, JK 2010, 'Identification and characterization of a NADH oxidoreductase involved in phenylacetic acid degradation pathway from *Streptomyces peucetius*' *Microbiological Research*, vol. 165, no. 8, pp. 649–656.

Nogales, J, Macchi, R, Franchi, F, Barzaghi, D, Fernández, C, García, JL, Bertoni, G, & Díaz, E 2007, 'Characterization of the last step of the aerobic phenylacetic acid degradation pathway' *Microbiology*, vol. 153, pp. 357–365.

Nunn, CM, Djordjevic, S, Hillas, PJ, Nishida, CR, Paul, R, & Montellano, O de 2002, 'The crystal structure of Mycobacterium tuberculosis alkylhydroperoxidase AhpD, a potential target for antitubercular drug design' *The Journal of Biological Chemistry*, vol. 277, no. 22, pp. 20033–20040.

Ohlendorf, DH, Orville, AM, & Lipscomb, JD 1994, 'Structure of protocatechuate 3,4-dioxygenase from <i>Pseudomonas aeruginosa< at 2.15 Å resolution' *Journal of Molecular Biology*, vol. 244, pp. 586–608.

Olivera, ER, Minambres, B, Garcia, B, Muniz, C, Moreno, MA, Ferrandez, A, Diaz, E,

Garcia, JL, & Luengo, JM 1998, 'Molecular characterization of the phenylacetic acid catabolic pathway in Pseudomonas putida U: The phenylacetyl-CoA catabolon' *Proceedings of the National Academy of Sciences*, vol. 95, pp. 6419–6424.

Ornston, L. N. 1966b, 'The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida* III. Enzymes of the catechol pathway' *The Journal of Biological Chemistry*, vol. 241, no. 16, pp. 3795–3799.

Ornston, L. N. 1966c, 'The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida* IV. Regulation' *The Journal of Biological Chemistry*, vol. 241, no. 16, pp. 3800–3810.

Ornston, L N 1966a, 'The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida* II. Enzymes of the protocatechuate pathway' *The Journal of Biological Chemistry*, vol. 241, no. 16, pp. 3800–3810.

Ornston, L Nicholas 1971, 'Regulation of catabolic pathways *Pseudomonas*' *Bacteriological Reviews*, vol. 35, no. 2, pp. 87–116.

Parales, RE & Harwood, CS 1992, 'Characterization of the genes encoding 3-ketoadipate: Succinyl-coenzyme A transferase in Pseudomonas putida' *Journal of Bacteriology*, vol. 174, no. 14, pp. 4657–4666.

Paralest, RE & Harwood, CS 1993, 'Regulation of the *pcaIJ* genes for aromatic acid degradation in *Pseudomonas putida*' *Journal of Bacteriology*, vol. 175, no. 18, pp. 5829–5838.

Parke, D, Meagher, RB, & Ornston, LN 1973, 'Relationships among enzymes of the β-ketoadipate pathway. III. Properties of crystalline γ-carboxymuconolactone decarboxylase from *Pseudomonas putida*' *Biochemistry*, vol. 12, no. 18, pp. 3537–3542.

Parke, D & Ornston, LN 1986, 'Enzymes of the β-ketoadipate pathway are inducible in *Rhizobium* and *Agrobacterium* spp. and constitutive in *Bradyrhizobium* spp' *Journal of Bacteriology*, vol. 165, no. 1, pp. 288–292.

Powlowski, JB, Ingebrand, JIM, & Dagley, S 1985, 'Enzymology of the β-ketoadipate pathway in *Trichosporon cutaneum*' *Journal of Bacteriology,* vol. 163, no. 3, pp. 1136–1141.

Powlowskit, JB & Dagley, S 1985, 'β-ketoadipate pathway in *Trichosporon cutaneum* modified for methyl-substituted metabolites' *Journal of Bacteriology,* vol. 163, no. 3, pp. 1126–1135.

Ramos, JL, Marqués, S, & Timmis, KN 1997, 'Transcriptional control of the Pseudomonas TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators' *Annual Review of Microbiology,* vol. 51, pp. 341–373.

Sarand, I, Skärfstad, E, Forsman, M, Romantschuk, M, & Shingler, V 2001, 'Role of the DmpR-mediated regulatory circuit in bacterial biodegradation properties in methylphenol-amended soils' *Applied and Environmental Microbiology,* vol. 67, no. 1, pp. 162–171.

Schlömann, M 1994, 'Evolution of chlorocatechol catabolic pathways - Conclusions to be drawn from comparisons of lactone hydrolases' *Biodegradation,* vol. 5, no. 3–4, pp. 301–321.

Schulz, H 1991, 'Beta oxidation of fatty acids' *Biochimica et Biophysica Acta,* vol. 1081, pp. 109–120.

Shanley, MS, Harrison, A, Parales, RE, Kowalchuk, G, Mitchell, DJ, & Ornston, LN 1994, 'Unusual G+C content and codon usage in catIJF, a segment of the ben-cat supra-operonic cluster in the Acinetobacter calcoaceticus chromosome' *Gene,* vol. 138, pp. 59–65.

Shingler, V 2003, 'Integrated regulation in response to aromatic compounds: From signal sensing to attractive behaviour' *Environmental Microbiology,* vol. 5, no. 12, pp. 1226–1241.

Smith, MR 1990, 'The biodegradation of aromatic hydrocarbons by bacteria' *Biodegradation,*

vol. 1, pp. 191–206.

Song, F, Zhuang, Z, Finci, L, Dunaway-mariano, D, Kniewel, R, Buglino, JA, Solorzano, V, Wu, J, & Lima, CD 2006, 'Structure, function , and mechanism of the phenylacetate pathway hot dog-fold thioesterase paal' *The Journal of Biological Chemistry*, vol. 281, no. 16, pp. 11028–11038.

Sparnins, VL & Dagley, S 1975, 'Alternative routes of aromatic catabolism in Pseudomonas acidovorans and Pseudomonas putida: Gallic acid as a substrate and inhibitor of dioxygenases' *Journal of Bacteriology*, vol. 124, no. 3, pp. 1374–1381.

Staack, H, Binstock, JF, & Schulz, H 1978, 'Purification and properties of a pig heart thiolase with broad chain length specificity and comparison of thiolases from pig heart and *Escherichia coli*' *The Journal of Biological Chemistry*, vol. 253, no. 6, pp. 1827–1831.

Teufel, R, Mascaraque, V, Ismail, W, Voss, M, Perera, J, Eisenreich, W, Haehnel, W, & Fuchs, G 2010, 'Bacterial phenylalanine and phenylacetate catabolic pathway revealed' *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 32, pp. 14390–14395.

Tropel, D & Meer, JR Van Der 2004, 'Bacterial transcriptional regulators for degradation pathways of aromatic compounds' *Microbiology and Molecular Biology Reviews*, vol. 68, no. 3, pp. 474–500.

Van der Meer, JR, De Vos, WM, Harayama, S, & Zehnder, AJB 1992, 'Molecular mechanisms of genetic adaptation to xenobiotic compounds' *Microbiological Reviews*, vol. 56, no. 4, pp. 677–694.

Venkatesan, A, Palaniyandi, K, Sharma, D, Bisht, D, & Narayanan, S 2016, 'Functional characterization of PknI-Rv2159c interaction in redox homeostasis of Mycobacterium tuberculosis' *Frontiers in Microbiologyrontiers in Microbiology*, vol. 7, pp. 1–12.

Whittaker, JW, Lipscombsli, JD, Kentll, TA, & Munck, E 1984, '*Brevibacterium fuscum* protocatechuate 3,4-dioxygenase:. Purification, crystallization, and characterization' *The*

Journal of Biological Chemistry, vol. 259, no. 7, pp. 4466–4475.

Widdel, F & Rabus, R 2001, 'Anaerobic biodegradation of saturated and aromatic hydrocarbons' *Current Opinion in Biotechnology*, vol. 12, no. 3, pp. 259–276.

Williams, PA & Sayers, JR 1994, 'The evolution of pathways for aromatic hydrocarbon oxidation in *Pseudomonas*' *Biodegradation*, vol. 5, no. 3–4, pp. 195–217.

Williams, SE, Woolridge, EM, Ransom, SC, Landro, JA, Babbitt, PC, & Kozarich, JW 1992, '3-carboxy-m, m-muconate lactonizing enzyme from *Pseudomonas putida* is homologous to the class II fumarase family: A new reaction in the evolution of a mechanistic motif' *Biochemistry*, vol. 31, pp. 9768–9776.

Wise, AA & Kuske, CR 2000, 'Generation of novel bacterial regulatory proteins that detect priority pollutant phenols' *Applied and Environmental Microbiology*, vol. 66, no. 1, pp. 163–169.

Yang, J, Wang, Y, Woolridge, EM, Arora, V, Petsko, GA, Kozarich, JW, & Ringe, D 2004, 'Crystal Structure of 3-Carboxy- cis , cis -muconate Lactonizing Enzyme from Pseudomonas putida , a Fumarase Class II Type Cycloisomerase : Enzyme Evolution in Parallel Pathways', vol. 43, no. 32, pp. 10424–10434.

Yang, S-Y, Yang, X-YH, Healy-Louie, G, Schulz, H, & Elzinga, M 1990, 'Nucleotide sequence of the fadA gene. Primary structure of 3-ketoacyl-coenzyme A thiolase from *Escherichia coli* and the structural organization of the *fad*AB operon' *The Journal of Biological Chemistry*, vol. 265, no. 18, pp. 10424–10429.

Yeh, W, Durham, DONR, Fletcher, P, & Ornston, LN 1981, 'Evolutionary relationships among y-carboxymuconolactone decarboxylases' *Journal of Bacteriology*, vol. 146, no. 1, pp. 233–238.

Yeh, W, Fletcher, P, & Ornston, LN 1980a, 'Evolutionary divergence of co-selected β-ketoadipate enol-lactone hydrolases in *Acinetobacter calcoaceticus*' *The Journal of Biological Chemistry*, vol. 255, no. 13, pp. 6342–6346.

Yeh, W, Fletcher, P, & Ornston, N 1980b, ‘Homologies in the NH2-terminal amino acid sequences of γ-carboxymuconolactone decarboxylases and muconolactone isomerase’ *The Journal of Biological Chemistry*, vol. 255, no. 13, pp. 6347–6354.

Yeh, W & Ornston, LN 1981, ‘Evolutionarily homologous α2β2 oligomeric structures in beta-ketoadipate succinyl-CoA transferases from *Acinetobacter calcoaceticus* and *Pseudomonas putida*’ *The Journal of Biological Chemistry*, vol. 256, no. 4, pp. 1565–1569.

Yeh, W & Ornston, LN 1984, ‘p-Chloromercuribenzoate specifically modifies thiols associated with the active sites of β-ketoadipate enol-lactone hydrolase and succinyl CoA: β-ketoadipate CoA transferase’ *Archives of Microbiology*, vol. 138, pp. 102–105.

Yeh, WK & Ornston, LN 1982, ‘Similar structures in γ-carboxymuconolactone decarboxylase and β-ketoadipate succinyl coenzyme A transferase’ *Journal of Bacteriology*, vol. 149, no. 1, pp. 374–377.

Chapter 3
Methodologies for experimental and bioinformatic studies

3.1 Material and methods involved in PcaF study

The objective of the PcaF study was to discern the degradative and biosynthetic thiolase. To gather clues regarding the two thiolases, the PcaF reaction centre and active site was explored. The actives mutants were made to trap ligands and to understand the substrate binding in degradative thiolases. Further structural and bioinformatics analysis were performed to get a holistic difference between two groups.

The clone of the PcaF in the pSGC-His vector with N-terminal His-tag including a TEV cleavage site was kindly provided by the New York Structural Genomics Research Consortium (NYSGRC). This study began with mutation of the active site residues followed by purification and crystallization of the native and mutant proteins. Soaking of both the native and the variant with the CoA and longer chain CoA derivatives were achieved. The study was bolstered by computational studies like tunnel analysis, sequence conservation examination etc.

3.1.1 Site-directed mutagenesis of the active site residues of PcaF

The PcaF construct was used as a template to make the following single active site residue mutations: C90S, C90A and H356A. Furthers studies on the C90A and C90S mutations showed that C90A mutant form insoluble protein and C90S mutant presented above 95% activity relative to the native protein. Thus, C90S and C90A mutants were not suitable to trap ligands and further double mutations were designed to produce PcaF mutants that would have low activity that would enable trapping of ligands. The mutant construct of H356A was used as a template to make the double mutations H356A-C90S, H356A-C386A. All the mutants were made by using the "site-directed mutagenesis kit" from Kapa Biosystems (Webpage: https://www.kapabiosystems.com/productapplications/applications/pcr/mutagenesis/).The

PCR program that was used for the final site directed mutagenesis reaction start with an initial denaturation step at 95ºC for 5min; followed by 30 cycles comprising of denaturation step at 98ºC for 20sec. The next step is annealing that is usually carried at Tm -7ºC for each reaction for 20sec, this is followed by the extension step at 72ºC for 5min. The final extension is done at 72ºC for min and is kept at final hold at 4ºC.

The reagents that were used are mentioned in the Table 3.1:

Reagent	**Reagent concentrations**	**Volume used (µL)**
Kapa Hifi Buffer	5X	5.0
Fwd Primer	10 µM	1.0
Rev Primer	10 µM	1.0
dNTP	10 mM	1.0
Template DNA		4.0
Hifi Polymerase		1.0
Sterile distilled water was used to make total volume up to 50 µL		

Table 3-1: The list of regents used for the site-directed mutagenesis of PcaF

3.1.2 Expression and purification of native and mutant proteins of PcaF

The native protein and the mutants were transformed in BL21 (DE3) pLysS cells and were plated onto chloramphenicol and kanamycin (concentrations 30 µg/mL and 35 µg/mL respectively) plates followed by 16hr incubation at 37ºC. Colonies obtained in the above plates were inoculated into 5 mL LB pre-inoculum broth with the same concentration of chloramphenicol and kanamycin. Later, the 5 mL LB broth was transferred to the large-scale LB-kanamycin-chloramphenicol culture media which was then grown at 30ºC by shaking it at 250 rpm. When the OD of the culture at 600 nm reached ~0.7, the culture was induced by 0.5 mM isopropyl-β-D- thiogalactopyranoside (IPTG) and the growth was continued at a reduced temperature of 25ºC for 8 hr. Afterward the cells were harvested by centrifugation at 4000 rpm for 30min and purified using Ni-NTA resin by standard His-tagged affinity purification protocol involving lysis buffer (50 mM Tris-HCl buffer, pH 7.5; 2 mM Imidazole; 200 mM NaCl, 5 mM β-Mercaptoethanol), wash buffer (50 mM Tris-HCl buffer, pH 7.5; 30 mM Imidazole; 200 mM NaCl), and elution buffer (50 mM Tris-HCl buffer, pH 7.5; 350 mM Imidazole; 100mM NaCl).

The eluted fractions were desalted using an Econo-Pac 10 DG (Bio-Rad, CA, USA) column that was pre-equilibrated with a desalting buffer containing 25 mM Tris-HCl buffer, pH 7.5; 80 mM NaCl, 5% glycerol, and 0.5 mM DTT. The desalted protein fractions were pooled and concentrated up to 12.5 mg/ml, as determined by the Bradford assay using Bovine Serum Albumin (BSA) as a standard (Table3.2). The purity of the protein was verified by running a 10% SDS-PAGE followed by Coomassie Blue staining. The fractions were then flash-frozen in liquid N_2 and stored at −80°C until they were used.

List of mutations and their positions	Confirmed by sequencing	Purified
C90S	Yes	Yes
C90A	Yes	Insoluble
H356A	Yes	Yes
H356A C90S	Yes	Yes
H356A C386A	Yes	Yes

Table 3-2: The list of mutation that were made for the PcaF protein

3.2 Crystallography and Structure-based study of the native and mutant PcaF protein

3.2.1 Crystallization of the native and mutant PcaF protein

Several commercially available crystallization screens such as Crystal screen, PEG-Ion, SaltRx (Hampton Research), JCSG, PACT suite (Qiagen), Natrix-Sigma, Wizard Screen I and II were employed for initial crystallization screening of the native PcaF protein, and within a week crystals appeared in condition number 20 of the JCSG Suite [0.2 M MgCl2, 10% w/v PEG 8000, 0.1 M Tris (pH7.0)] and condition number 45 of the PACT Suite [0.2 M LiCl, 20% PEG 6000, 0.1 M Tris (pH-8.0)]. The crystallization trials were performed using hanging-drop vapour-diffusion method in Hampton 24-well plates. Based on initial X-ray characterization of crystals grown from the above conditions, the condition number 45 [0.2 M LiCl, 20% PEG 6000, 0.1 M Tris (pH-8.0)] of the PACT Suite was further optimized.

For the apo as well as the mutant form of the protein, 1 µL protein solution mixed with 1 µL reservoir solution and was equilibrated against 500 µL reservoir solution. All the crystallization experiments were carried out at room temperature. Temperature of the crystallization plate was

LIGANDS	**CoA**	**Acetyl CoA**	**Acetoacetyl -CoA**	**Hexanyol-CoA**	**Octanyol -CoA**	**Decanyol -CoA**
PROTEIN	***(Soaked with 1 mM and 3 mM)***	***(Soaked with 1 mM and 3 mM)***	***(Soaked with 1 mM and 3 mM)***	***(Soaked with 1 mM and 3 mM)***	***(Soaked with 1 mM and 3 mM)***	***(Soaked with 1 mM and 3 mM)***
NATIVE	0.1 M Tris pH- 7.7, 14% PEG 6000, 0.2 M LiCl	0.1 M Tris pH- 7.9, 20% PEG 6000, 0.2 M LiCl	0.1 M Tris pH-7.5, PEG 6000-20%, 0.2 M LiCl	0.1 M Tris pH-8.0,PEG 6000 16%, 0.2 M LiCl	0.1 M Tris pH-7.7, PEG 6000 22%, 0.2 M LiCl	0.1 M Tris pH- 7.7,PEG 6000 18%, 0.2 M LiCl
C90S	0.1 M Tris pH- 7.3, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH- 7.7, PEG 6000 22%, 0.2 M LiCl	0.1 M Tris pH-7.5, PEG 6000 28%, 0.2 M LiCl	0.1 M Tris pH-7.7, PEG 6000 18%, 0. 2M LiCl	0.1 M Tris pH-7.7, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH-7.7, PEG 6000 20%, 0.2 M LiCl
H356A	0.1 M Tris pH- 7.7, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH- 7.7, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH-7.7, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH- 7.7, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH- 7.7, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH- 7.7, PEG 6000 20%, 0.2 M LiCl
H356A C90S	0.1 M Tris pH- 7.3, PEG 6000 20%, 0.2M LiCl	0.1 M Tris pH- 7.7, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH-8.1, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH-7.7, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH- 7.7, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH- 7.5, PEG 6000 20%, 0.2 M LiCl
H356A C386A	0.1 M Tris pH- 8.3, PEG 6000 22%, 0.2 M LiCl	0.1 M Tris pH- 7.7, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH-7.7, PEG 6000 18%, 0.2 M LiCl	0.1 M Tris pH-8.1, PEG 6000 14%, 0.2 M LiCl	0.1 M Tris pH- 7.9, PEG 6000 20%, 0.2 M LiCl	0.1 M Tris pH- 7.7, PEG 6000 26%, 0.2 M LiCl

Table 3-3: The crystallization conditions that were optimized for the native and mutant PcaF co-crystallized with variety of ligands

maintained in a temperature-controlled cabinet at 22°C. Two different crystal forms were obtained from the same crystallization condition. The first crystal form appeared after 48 hours

of setting up crystal trays diffracted poorly and belonged to C2 space group with 16 molecules in an asymmetric unit. The second crystal form took three weeks to grow and belonged to $P2_12_12_1$ space group with 4 molecules in asymmetric unit and diffracted with higher resolution.

Ligand soaking

The orthorhombic form was further used for elucidation of native and ligand-complex structures. The native protein as well as the single and double mutant proteins (H356A, H356A-C90S, and H356A-C386A) were soaked with various substrates: Coenzyme A, Acetyl CoA, Acetoacetyl CoA, Hexanoyl CoA, Octanoyl CoA, Decanoyl CoA (Table 3.3). The crystals were soaked for 20sec, 40sec, 1min or 5min. A single crystal of each ligand-complex was cryoprotected with 20% (v/v) ethylene glycol (prepared using mother liquor) including 3 mM of ligand prior to cryo-cooling and data collection.

Data collection

X-ray diffraction experiments were performed at the micro-focus beamline (MX2) of the Australian Synchrotron (Araga et al. 2018). The crystals were flash cooled in liquid nitrogen and transferred to a stream of nitrogen gas at 100K. X-ray data were collected at a wavelength of 0.9537 Å using an EIGER-16M detector with 0.1° oscillation and 0.1sec exposure of 0% attenuated beam per frame. 1800 frames of each data set were collected in 18sec. Altogether 18 datasets were recorded in the resolution range of 2.56 to 1.37 Å. The data were indexed and integrated with XDS (Kabsch 2010) and scaled using AIMLESS (Evans & Murshudov 2013). Each of these datasets were analysed using the MR protocol of Auto-Rickshaw (as described below). After initial refinement of the native and C90S protein ligand complexes did not revealcontain any difference density for the ligand after initial refinement. The other mutants; A-mutant (H356A), AA-mutant (H356A-C386A) and AS-mutant (H356A-C90S) indicated difference density for the CoA chain and were further refined to visualize the acyl part of the CoA derivatives. Details of the data collection from the A-mutant-ligand complex, AA-mutant-ligand complex and AS-mutant-ligand crystals and its statistics from subsequent processing are presented in Table 3.4, 3.5 and 3.6.

DATA	H356A-CoA	H356A-Acetyl-CoA	H356A-Acetoacetyl-CoA	H356A Hexanoyl-CoA	H356A-Octanoyl-CoA	H356A-Decanoyl-CoA
Total frames	1800	1800	1800	1800	1800	1800
Oscillation (°)	0.10	0.10	0.10	0.10	0.10	0.10
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å, °)	a=110.49 b=115.96 c=128.87 α=90.00 β=90.00 γ=90.00	a=110.72 b=116.34 c=128.65 α=90.00 β=90.00 γ=90.00	a=110.71 b=116.16 c=129.03 α=90.00 β=90.00 γ=90.00	a=111.39 b=116.40 c=127.90 α=90.00 β=90.00 γ=90.00	a=110.54 b=116.36 c=128.73 α=90.00 β=90.00 γ=90.00	a=110.73 b=115.87 c=128.87 α=90.00 β=90.00 γ=90.00
Resolution (Å)	1.61	1.64	1.54	1.96	2.01	1.63
Mosaicity (°)	0.172	0.079	0.094	0.304	0.151	0.162
Total reflections	1445863 (218460)	1376855 (209708)	1673543 (265193)	697289 (106899	749519 (119215	1400445 (207572
Unique reflections	413663 (65788)	392209 (62830)	475824 (76173)	220195 (35163	212746 (33763)	399260 (62822)
Redundancy	3.50 (3.32)	3.51 (3.34)	3.52 (3.48)	3.17 (3.04)	3.52 (3.53)	3.51 (3.30)
Completeness (%)	99.3 (97.6)	99.4 (98.5)	99.6 (98.7)	95.8 (94.7)	99.2 (97.5)	99.2 (96.4)
I/σ(I)	11.36 (1.85)	11.35 (2.06)	14.62 (2.44)	8.52 (2.42)	9.34 (2.88)	11.74 (2.03)
R_{merge} (%)	6.8 (56.0)	8.2 (57.7)	5.3 (47.8)	11.3 (51.7)	13.7 (48.8)	7.0 (57.4)
R_{meas} (%)	8.1 (66.9)	9.7 (68.8)	6.2 (56.6)	13.6 (62.4)	16.2 (57.6)	8.3 (68.6)
Wilson-B ($Å^2$)	25.64	24.21	24.28	26.40	27.23	25.47

Table 3-4 : Crystallographic data-collection statistics for A-mutant bound with ligand (Values inside the bracket are for the outermost resolution shell).

DATA	H356A C90S -CoA	H356A C90S-Acetyl-CoA	H356A C90S-Acetoacetyl-CoA	H356A C90S-Hexanoyl-CoA	H356A C90S-Octanoyl-CoA	H356A C90S-Decanoyl-CoA
Total frames	1800	1800	1800	1800	1800	1800
Oscillation (°)	0.10	0.10	0.10	0.10	0.10	0.10
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å, °)	a=110.49 b=115.96 c=128.87 α=90.00 β=90.00 γ=90.00	a=111.06 b=116.76 c=128.33 α=90.00 β=90.00 γ=90.00	a=110.95 b=116.55 c=128.68 α=90.00 β=90.00 γ=90.00	a=110.90 b=116.78 c=128.75 α=90.00 β=90.00 γ=90.00	a=110.99 b=116.55 c=128.41 α=90.00 β=90.00 γ=90.00	a=110.90 b=116.78 c=128.75 α=90.00 β=90.00 γ=90.00
Resolution (Å)	2	2.14	1.72	1.59	1.47	1.59
Mosaicity (°)	0.172	0.127	0.188	0.223	0.058	0.223
Total reflections	1445863 (218460)	623340 (98877)	1292313 (201674)	1628430 (251830)	1933169 (307007)	1628430 (251830)
Unique reflections	413663 (65788)	175637 (27786)	340195 (53941)	432462 (68676)	547130 (87364)	432462 (68676)
Redundancy	3.50 (3.32)	3.55 (3.56)	3.80 (3.74)	3.77 (3.67)	3.53 (3.51)	3.77 (3.67)
Completeness (%)	99.3 (97.6)	99.2 (96.9)	99.4 (97.3)	99.1 (97.4)	99.6 (98.3)	99.1 (97.4)
I/σ(I)	11.36 (1.85)	10.16 (3.33)	9.64 (1.95)	10.41 (2.30)	12.87 (2.72)	10.41 (2.30)
R_{merge} (%)	6.8 (56.0)	9.2 (34.6)	11.8 (68.0)	7.2 (43.4)	6.1 (44.2)	7.2 (43.4)
R_{meas} (%)	8.1 (66.9)	10.8 (40.7)	13.7 (79.0)	8.4 (50.8)	7.2 (52.1)	8.4 (50.8)
Wilson-B ($Å^2$)	25.64	30.11	22.29	25.31	21.14	25.31

Table 3-5: Crystallographic data-collection statistics for AS-mutant bound with ligand (Values inside the bracket are for the outermost resolution shell).

DATA	H356A C386A-CoA	H356A C386A-Acetyl-CoA	H356A C386A-Acetoacetyl-CoA	H356A C386A-Octanyol-CoA	H356A C386A-Decanyol-CoA
Total frames	1800	1800	1800	1800	1800
Oscillation (°)	0.10	0.10	0.10	0.10	0.10
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å, °)	a=110.88 b=128.95 c=116.04 α=90.00 β=90.00 γ=90.00	a=110.93 b=116.07 c=128.80 α=90.00 β=90.00 γ=90.00	a=110.67 b=116.12 c=128.96 α=90.00 β=90.00 γ=90.00	a=110.61 b=116.87 c=129.07 α=90.00 β=90.00 γ=90.00	a=110.74 b=115.95 c=128.87 α=90.00 β=90.00 γ=90.00
Resolution (Å)	1.87	2.30	2.14	2.22	2.30
Mosaicity (°)	0.117	0.119	0.116	0.170	0.231
Total reflections	941221 (150266)	506792 (79128	623394 (98220)	549562 (78792)	503875 (80940)
Unique reflections	515820 (81670)	141998 (22470	174400 (27354)	156556 (23769)	141984 (22574)
Redundancy	1.82 (1.84)	3.57 (3.52	3.57 (3.59)	3.51 (3.31)	3.55 (3.59)
Completeness (%)	96.7 (94.7)	99.2 (97.2	99.1 (96.3)	98.4 (92.3)	99.3 (97.6)
I/σ(I)	8.06 (1.99)	6.91 (1.84	10.02 (3.10)	10.03 (2.30)	13.05 (4.47)
R_{merge} (%)	7.5 (39.4)	17.9 (71.9)	10.0 (40.0)	9.8 (53.9	7.3 (29.5)
R_{meas} (%)	10.4 (54.7)	21.1 (85.0)	11.8 (46.9)	11.6 (64.6)	8.6 (34.7)
Wilson-B ($Å^2$)	25.63	30.13	30.56	35.82	36.44

Table 3-6: Crystallographic data-collection statistics for AA-mutant bound with ligand (Values inside the bracket are for the outermost resolution shell).

3.3 Structure determination and refinement of the PcaF ligand complex

The apo and ligand–complex structures were solved using the MR protocol of the software pipeline Auto-Rickshaw (Panjikar et al. 2005, 2009). Structure of tt0182 (a putative acetyl-CoA acetyltransferase) from *Thermus thermophilus* HB8 (PDB code: 1ULQ) was used as a template for solving the structure of the apo PcaF. The two molecules of the thiolase (which forms a dimer) were used as the search model for molecular replacement. Within the software pipeline, MR was performed using the program using MOLREP (Vagin & Teplyakov 2000) and rigid-body, positional and B-factor refinement were carried out using the program CNS to 3.0 Å resolution. The structure was further refined using REFMAC5 (Murshudov et al. 2011) to its maximum resolution. Density modification was carried out using PIRATE (Bailey 1994) and model building was carried out using ARP/wARP (Langer, Cohen, Lamzin, & Perrakis 2008). The resulting model was further improved by rebuilding in the graphics program COOT (Emsley et al. 2010). Refinement was performed using REFMAC5 (Murshudov et al. 2011).

The native PcaF structure was used for solving all ligand complexes using MR protocol of Auto-Rickshaw (Panjikar et al. 2005, 2009). The resulting model for each complex, was refined in REFMAC5 (Murshudov et al. 2011) and water molecules were added. The resulting map of each ligand complex was analyzed and difference density for the ligand was located using the graphic program COOT. The CIF file for the ligand was prepared using the program ACEDRG (Bailey 1994) and used for the ligand building in COOT. The quality of the final model was validated with MolProbity (Chen et al. 2010). The refinement statistics for A-mutant-ligand complex, AA-mutant-ligand complex and AS-mutant-ligand crystals are shown in Table 3.7, 3.8, and 3.9.

DATA	**H356A-CoA**	**H356A-Acetyl-CoA**	**H356A-Acetoacetyl-CoA**	**H356A Hexanyol-CoA**	**H356A-Octanyol-CoA**	**H356A-Decanyol-CoA**
Resolution (Å)	20-1.61	20-1.64	20-2.00	20-1.96	20-1.8	20-1.63
No. of reflections	211120	200395	109720	112478	147959	204880

R_{work} / R_{free}	17.84/21.38	16.53/ 19.59	14.67/20.15	25.02/26.84	23.82 /28.00	24.36/25.66
r.m.s.d from target values Bond lengths (Å)	0.020	0.022	0.019	0.018	0.018	0.019
r.m.s.d from target values Bond angles (°)	2.025	2.041	1.903	1.873	1.996	2.066
Ramachandran analysis Favored regions (%)	1525 (96.09%)	1501 (95.97%)	1506 (95.62%)	1521 (95.84%)	1514 (95.76%)	1523 (95.97%)
Ramachandran analysis Allowed regions (%)	53(3.34%)	56(3.58%)	58(3.68%)	58(3.64%)	58(3.67%)	56(3.53%)
Ramachandran analysis Outliers (%)	9(0.57%)	7(0.45%)	11(0.70%)	8 (0.50%)	9(0.57%)	8(0.50%)

Table 3-7: Refinement statistics for A-mutant bound with the ligands

DATA	**C90S H356A-CoA**	**C90S H356A-Acetyl-CoA**	**C90S H356A-Acetoacetyl-CoA**	**C90S H356A-Hexanyol-CoA**	**C90S H356A-Octanyol-CoA**	**C90S H356A-Decanyol-CoA**
Resolution (Å)	20-2	20-2.14	20-2.01	20-1.59	20-1.37	20-1.48
No. of reflections	109720	90328	108273	222944	344775	265786
R_{work} / R_{free}(%)	18.8/25.2	22.85/25.99	15.32 /20.89	26.60/27.15	16.16/17.66	23.63/25.21
r.m.s.d from target values Bond lengths (Å)	0.019	0.015	0.020	0.020	0.028	0.022

r.m.s.d from target values Bond angles (°)	1.899	1.676	1.853	1.964	2.496	2.149
Ramachandran analysis Favored regions (%)	1506 (95.56%)	1513 (95.34%)	1517 (95.59%)	1514 (95.40%)	1516 (96.01%)	1521 (95.84%)
Ramachandran analysis Allowed regions (%)	59(3.74%)	65(4.10%)	59(3.72%)	61(3.84%)	55(3.48%)	56(3.53%)
Ramachandran analysis Outliers (%)	11(0.70%)	9(0.57%)	11(0.69%)	12(0.76%)	8(0.51%)	10(0.63%)

Table 3-8: Refinement statistics for AS-mutant bound with ligands

DATA	**H356A C386A-CoA**	**H356A C386A-Acetyl-CoA**	**H356A C386A-Acetoacetyl-CoA**	**H356A C386A-Octanyol-CoA**	**H356A C386A-Decanyol-CoA**
Resolution (Å)	20-1.87	20-2.30	20-2.15	20-2.23	20-2.30
No. of reflections	136213	72653	89364	80322	72976
R_{work} / R_{free}	16.34/20.77	18.30/24.13	16.21/21.42	17.70/23.21	21.11/26.61
r.m.s.d from target values Bond lengths (Å)	0.021	0.014	0.018	0.016	0.014
r.m.s.d from target values Bond angles (°)	2.061	1.672	1.780	1.768	1.704
Ramachandran analysis Favored regions (%)	1496 (95.84%)	1515 (95.46%)	1518(95.65%)	1514 (95.40%)	1503(94.71%)

Ramachandran analysis Allowed regions (%)	52(3.33%)	65(4.10%)	61(3.84%)	63(3.97%)	74(4.66%)
Ramachandran analysis Outliers (%)	13(0.83%)	7(0.44%)	8(0.50%)	10(0.63%)	10(0.63%)

Table 3-9: Refinement statistics for AA- mutant bound with ligands

3.4 Radiation damage calculation of PcaF-ligand complex

180 degree of X-ray dataset was collected from the ligand complex datasets as described in "3.2 Crystallization, ligand soaking and data collection" section. In order to analyze, radiation damage of ligand complexes, first 90 degree and last 90 degree of datasets were processed in XDS and the resulting datasets is termed as "before" and "after" dataset respectively. Overall completeness of each dataset was still about 90% due to primitive orthorhombic space group. The refined ligand complex structure was used for model phase. Isomorphous difference Fourier map was created by using the program ANODE (Thorn & Sheldrick 2011) with the model-phase and the structure factor ($F_{before} - F_{after}$) created using the program SHELXC (Uso´n & Sheldrick 2018) with 'RIP' option. The resulting map was used for radiation damage analysis along with the ligand complex structure.

3.5 Enzymatic studies of native and mutant PcaF protein

The degradative activity for PcaF was measured and the kinetic parameters for the native protein was calculated. Thiolase thiolytic cleavage activity was performed using the modified Mg^{2+} method (Ithayaraja et al. 2016; Stern 1956). The reaction mixture had 70 µM CoASH and variable amount of (10-120 µM) Acetoacetyl-CoA in 0.1 M TrisHCl (pH 8.3) buffer containing 25 mM $MgCI_2$. The loss of acetoacetic CoA enolate chromophore was followed at 303 nm after 15min of incubation at 37°C (Figure 3-1). Kinetics study of the native PcaF was followed by activity assessment of the mutants: C90S, H356A, C90S-H356A, and H356A-386A.

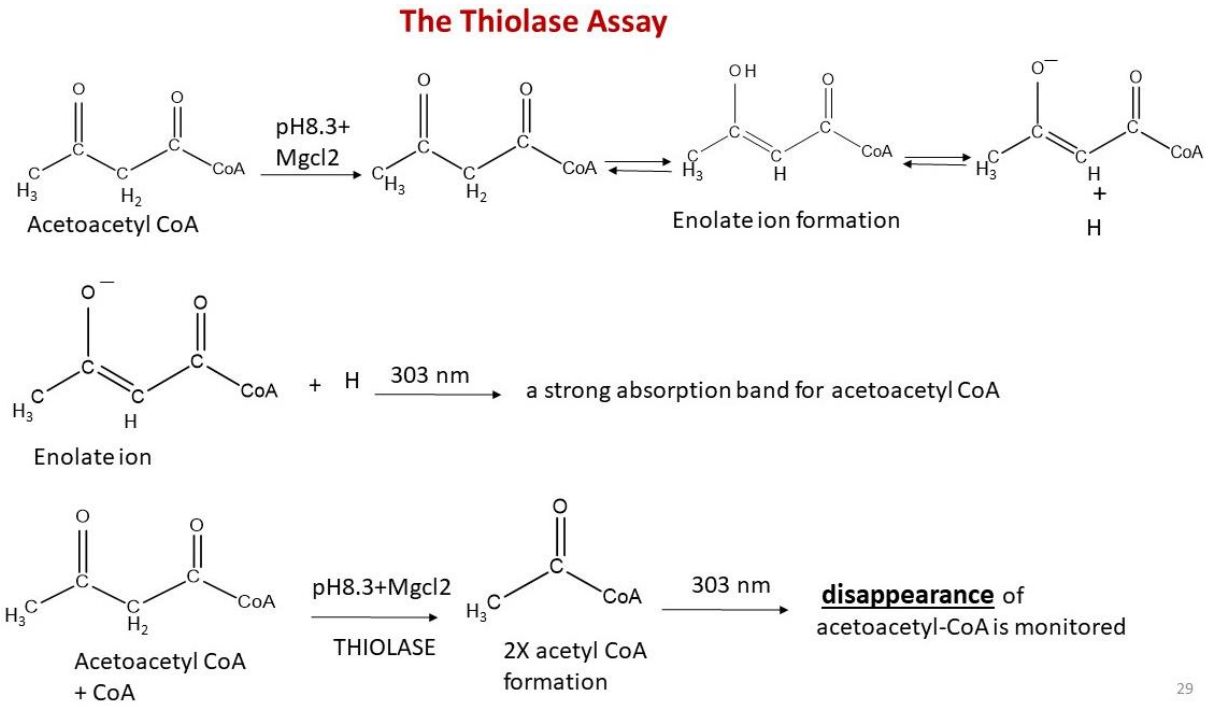

Figure 3-1 Chemical representation of the degradative assay for thiolases.

3.6 Intact mass spectrometry

The mass spectrometry was performed using the native protein and the AA mutant. The control mixture for the native protein consist of 0.25 M Tris-HCl buffer (pH 7.5), 1 mg/ml native protein. The ligand mixture for the native protein is same as control mixture but with 0.5 mM of acetoacetyl CoA. The mixture was incubated for 30min at 37°C before the experiment.

The control mixture for the AA-mutant consist of 0.25 M Tris-HCl buffer (pH 7.5), 1 mg/ml AA-mutant. Three ligand mixture for the AA-mutant were prepared with each sample having different ligand of acetoacetyl CoA/hexanoyl CoA/Octanoyl CoA. All the protein-ligand sample consist of 0.25 M Tris-HCl buffer (pH 7.5), 1 mg/ml AA-mutant, and 0.5 mM of ligand. The three reaction mixtures were incubated for 30min at 37°C. The reaction mixture then used for mass spectrometry. Protein samples were analysed by LC-MS using a quadrupole TOF mass spectrometer (MicroTOFq, Bruker Daltonics, Bremen, Germany) coupled online with a 1200 series capillary HPLC (Agilent technologies, Santa Clara, CA, USA). Samples injected onto a MabPac SEC-1 5um 300A 50x4mm (Thermo Scientific) column with 50% Acetonitrile 0.05%TFA, 0.05% FA at a flow rate of 50ul/minute. The protein is eluted with monitoring by UV detection at 254nm. The eluant is nebulised and ionised using the Bruker electrospray

source with a capillary voltage of 4500V dry gas at 180ºC, flow rate of 4l/minute and nebuliser gas pressure at 300mbar. Low concentration Tune mix (Agilent technologies, Santa Clara, CA, USA) use directly infused at the end of the run to calibrate the spectrum post acquisition. The spectra were extracted and deconvoluted using Data explorer software version 3.4 build 192 (Bruker Daltonics, Bremen, Germany).

3.7 Structural analysis and Bioinformatics study of PcaF

3.7.1 Structure analysis

To address the core aim of the projects, the crystal structures of the degradative thiolases from *Mycobacterium tuberculosis* (Mtb-thiolase) and the biosynthetic thiolase from *Zoogleria ramigera* (Zr-thiolase) were compared with the PcaF native and mutant structures. In particular, Mtb-thiolase in complex with steroid (PBD code: 4UBT) and four structures from the Zr-thiolase; the apo protein (PDB code: 1DLU), Zr-thiolase complexed with CoA (PDB code: 1DLV), C89A mutant of Zr-thiolase complexed with CoA (PDB code: 2WL4) and H348A mutant of Zr-thiolase complexed with acetoacetyl CoA (PDB code: 1M1O) were used in this study. The structures were super-positioned in the COOT software using the SSM protocol and were examined in detail (Krissinel & Henrick 2004).

COOT is a graphics software that is used for structure building and examining the macromolecular structures (Emsley et al. 2010). It can be used for visualization of crystallographic, electron microscopy and, NMR data. It is heavily used for model building and model manipulation. The SSM or secondary structure matching protocol in COOT is generally employed to compare the 3D structures of the proteins. It has a reference and a moving structure, wherein the moving PDB is aligned to the reference structure.

3.7.2 Channel analysis

Protein structures encompass pockets, crevices and channels in their 3D structures which are of significance. They usually form the site for ligand binding and provide surroundings for the reaction to occur. Understanding the chemical nature and architecture of these channel can shed light on the substrate binding. There are many software that are available online to inspect the

location of channels present in protein, namely CAVER, MOLE, POREWALKER etc. (Brezovsky et al. 2012). In this study two program CAVER and CASTp were used.

Caver is the first software tool developed to detect channels in proteins. The software requires protein structure in PDB format, coordinates for the start point and tunnel radius specifications. The output data includes channel parameters, residues around the channel etc. For this study the latest version CAVER 3.0.1 (Chovancova et al. 2012) program was used with its default parameters and with a shell radius of 3 Å in order to foretell the presence of tunnels in the protein. The C90 residue was considered as the start point. The channels were visualized by using a plugin in PyMol (DeLano 2002).

Computed Atlas of Surface Topography of proteins or CASTp program (Tian et al. 2018) was used for calculations and prediction of surface accessibility as well as internal cavities of the protein structures. The online server was used and the native PcaF monomeric structure was uploaded. The radius probe was kept default (1.4 Å). The output result was used to confirm as well as to measure area of the promising potential/cavity.

3.7.3 Sequence analysis

It presumed that the protein having analogous sequence tend to form comparable 3D structures. This theory is regularly used to search and select representative protein in structural studies. In this thesis, sequence-based studies were performed to identify conserved sequences or signature sequence that differentiate the biosynthetic and degradative thiolases. The online software that were used – Blastp, Clustwal Omega and MultiAlin.

The **Basic Local Alignment Search Tool (BLAST)** is one of the common software used for sequence analysis (https://blast.ncbi.nlm.nih.gov/). It can be used for nucleotide (DNA/RNA) and amino acid sequences (protein). The software matches the protein/ DNA/RNA sequence to the sequence database by looking at the “regions of local similarity” among the sequences. The statistical significance of the matched sequences is counted and the relevance of the found sequence can be implicated from it. BLAST searches aid in deducing the functional and phylogenetic relationships between sequences as proteins having conserved residues in active site or high sequence identify belong to same gene family (Altschu et al. 1990).

In this study, 100 sequences each for biosynthetic and degradative thiolases were sought using Blastp in NCBI. PcaF sequence was used as a reference for degradative thiolases against which 100 sequences below 95% sequence identity were curated. A similar criterion was used for the 100 biosynthetic sequences where Zr-thiolase was used as the query sequence.

Clustal Omega is a seeded guide trees and HMM profile-profile based program that is used for multiple sequence alignment. The online server on the webpage, https://www.ebi.ac.uk/Tools/msa/clustalo/, has three easy steps, wherein, in the first step, the query sequences are put in the web tool followed by the second step of optimizing the parameters of the tool and ending with the final step of submission of the query sequences. The result outputs comprises of sequence alignment file and phylogenetic tree(Sievers et al. 2011). In this study, the 100 curated sequences for each biosynthetic and degradative thiolases were alignment and the homologous regions of the alignment were analysed.

MultiAlin is another example of program that can align more than three nucleotide or protein sequences. It utilizes the programming process of pairwise alignment to form a hierarchical clustering. This program also has a web tool that can be accessed on the webpage, http://multalin.toulouse.inra.fr/multalin/, where the query sequences are entered in the input box and the alignment is obtained. There is a choice of optional parameters, where the gap penalty, consensus levels etc. can be changed from default (Corpet 1988). In this study, the 100 biosynthetic and 100 degradative sequences were aligned (independent to Clustwal Omega) to further confirm the conservation of the active site residues and the residues lining the active site. The sequence alignment figure used in this study was made from this online tool.

3.8 Material and methods involved in PcaC study

γ-carboxymuconolactone decarboxylase (γ-CMD) is a decarboxylase enzyme that belong to the aromatic compound degradation pathway in bacteria. The first structure of γ-CMD (PDB: 2AF7) was solved in 2005 by the Northeast Structural Genomics Consortium. However, detecting the active site of γ-CMD was proven to be quite challenging as they had exhibited relatively low sequence identity against the known structure. Thus, till date the actives site residues have remained to be unknown. In this study the structure of PcaC, a representative of γ-CMD, is solved to investigate the active site residues and the reaction mechanism.

The full-length clone for the PcaC, was kindly provided by the New York Structural Genomics Research Consortium (NYSGRC). The clone was present in a modified vector, namely the pNIC28-BSa4-His vector with an N terminal His-tag and having a TEV cleavage site. The active site residues of the PcaC was identified by a novel bicarbonate soaking method which was developed during the course this study. The active site was confirmed with enzymatic assay and mutagenesis studies. Finally, crystal structures of PcaC mutants in complex with various analog were determined for elucidation of the enzymatic mechanism.

3.8.1 Site-directed mutagenesis of the active site residues of PcaC

All the mutants were made by site-directed mutagenesis by employing the "site-directed mutagenesis kit" from Kapa Biosystems (Biosystems n.d.). The PcaC construct was used as a template to make the following site mutations: E76A, and H80A. The PCR program that was used for the final site directed mutagenesis reaction start with an initial denaturation step at 95ºC for 5min; followed by 30 cycles comprising of denaturation step at 98ºC for 20sec. The next step is annealing that is usually carried at Tm -7ºC for each reaction for 20sec, this is followed by the extension step at 72ºC for 5min. The final extension is done at 72ºC for 6min and is kept at final hold at 4ºC. The reagents that were used are mentioned in the Table 3.10:

Reagent	**Reagent concentrations**	**Volume used (µL)**
Kapa Hifi Buffer	5X	5.0
Fwd Primer	10 µM	1.0
Rev Primer	10 µM	1.0
dNTP	10 mM	1.0
Template DNA		4.0
Hifi Polymerase		1.0
Sterile distilled water was used to make total volume up to 50 µL.		

Table 3-10: The list of regents used for the site-directed mutagenesis of PcaC

3.8.2 Expression and purification of PcaC and its mutants

Native protein:

The full-length clone construct was transformed into Escherichia coli BL21 (DE3) pLysS cells, over expressed with 1mM IPTG (isopropyl-β-D- thiogalactopyranoside), as six-His tag fusion proteins and purified using Ni-NTA resin by standard His-tagged affinity purification protocol. The composition of the buffers used in subsequent purification steps is as follows: lysis buffer (50 mM Tris-HCl buffer, pH 7.5; 2 mM Imidazole; 200 mM NaCl, 5 mM β-MercaptoEthanol), wash buffer (50 mM Tris-HCl buffer, pH 7.5; 30 mM Imidazole; 200 mM NaCl), and elution buffer (50 mM Tris-HCl buffer, pH 7.5; 350 mM Imidazole; 100 mM NaCl). The eluted fractions were desalted using an Econo- Pac 10DG (Bio-Rad, CA, USA) column that was pre-equilibrated with a desalting buffer containing 25 mM Tris-HCl buffer, pH 7.5; 80 mM NaCl, 5% glycerol, and 0.5 mM DTT. The desalted protein fractions were pooled and concentrated up to 11 mg/ml, as determined by the Bradford assay using Bovine Serum Albumin (BSA) as a standard. The fractions were then flash-frozen in liquid N_2 and stored at −80°C until they were used. The purity of the protein was verified by running a 10% SDS-PAGE followed by Coomassie Blue (HiMedia, Mumbai, India) staining.

Selenomet protein:

The Se-Met protein of PcaC (12.5 mg/ml) was provided by the NYSGRC.

Mutant proteins:

The mutants (Table 3.11) were transformed in BL21 (DE3) pLysS cells and were plated onto chloramphenicol and kanamycin (concentrations 30 μg/mL and 35 μg/mL respectively) plates. The rest of cell growing, harvesting and protein purification is similar to the native protein.

List of mutations and their positions	Confirmed by sequencing	Purified
E76A	Yes	Yes
H80A	Yes	Yes

Table 3-11 The list of mutation that were made for the PcaC protein.

3.9 Crystallography of the PcaC and its mutant proteins

3.9.1 Crystallization of native and selenomet proteins

The purified His-Tagged PcaC (12.5 mg ml^{-1}) was first screened for crystallization using several commercially available crystallization screens such as Crystal screen, PEG/Ion (Hampton Research), Salt Rx screen and JCSG suite, PACT suite (Qiagen) using a CrystalMation (Rigaku) integrated robotic workstation and the sitting-drop vapour diffusion technique at the high-throughput crystallization facility at Monash University, Australia. Crystals from native/selenomet proteins appeared in several conditions within week. The best initial crystals were observed in the two conditions: 2 M Sodium format, 0.1 Sodium acetate, pH 4.6 (Condition-A) and 0.4 M Magnesium chloride, 0.1 M Sodium acetate, pH 4.6 (Condition-B). Both conditions were reproduced using the Hanging drop vapour-diffusion method using 1:1 ration of protein and crystallisation reservior. Prior to crystallization setup the protein was mixed with 5% glycerol to control the fast nucleation. The two conditions were optimized with grid varying pH of the acetate buffer within a range of 4.2-5.2 along with varying the Sodium formate concentration within a range of 1.6-2.2 M in Condition-A and Magnesium chloride concentration within a range of 0.3 M-0.45 M in Condition-A. The final optimised conditions for Condition-A and for Condition-B were 1.6 M Sodium formate, 0.05 M Magnesium chloride, 0.1 Sodium acetate pH 5.0 and 0.3 M Magnesium chloride, 0.1 M Sodium acetate, pH 5.0 respectively. In each condition, hexagonal-shaped crystals appeared within 2 days.

3.9.2 Ligand soaking and co-crystallization of native and mutant PcaC

Similar to native protein, 5% glycerol was added to mutant proteins prior to crystallization to control the nucleation rate. PcaC-H80A mutants were crystallized in Condition-A, whereas crystal of PcaC-E76A mutants appeared in Condition-B within a week.

Crystallization of PcaC-H80A: The H80A crystals came in the a) 1.6 M Sodium formate, 0.1 Sodium acetate (pH 5.0), 0.05 M Magnesium chloride and the drops were set up in the ratio of 1:1 of protein: well solution.

Soaking of PcaC-H80A with homocitrate acid lactone: The crystals were soaked with 5 mM Homocitrate acid lactone (HCAL) and X-ray data were collected. HCAL was found bound in

only one chain to the cavity at the identical position where acetate molecule was bound in the native crystal. The bound HCAL was decarboxylated, where the carboxylic group was broken down and the byproduct CO_2 was found in the cavity too.

Co-crystallization of PcaC-H80A with 2FAA: H80A was also co-crystallize with 5mM the product analogue 2-furan acetic acid (2FAA). The crystallographic analysis of the ligand-complexes show that the product analogue is bound to the mutants at the CO_2 corresponding position.

Crystallization of PcaC-E76A: The E76A crystals were observed in (b) 0.3 M Magnesium chloride, 0.1 M Sodium acetate (pH 5.0) and the drops were set up in the ratio 1:1 but the protein was mixed with 5% glycerol prior to set to reduce the nucleation rate. Single, big crystals appeared within 12 days.

Soaking of PcaC-E76A with HCAL: The crystals were soaked with 5 mM Homocitrate acid lactone (HCAL) and X-ray data were collected. HCAL was found bound to the cavity where acetate molecule was bound in the native crystal. This confirmed that the E76 is part of active site.

Soaking of PcaC-E76A with 2FAA: The crystals were soaked with 5 mM 2-furan acetic acid (2FAA) and X-ray data were collected. 2FAA was found bound to the cavity where CO_2 molecule was bound. This confirmed that the E76 is part of active site.

Data collection from selenomet, native and ligand complexes of PcaC:

Before data collection, the protein crystals of native and selenomet were cryoprotected by direct transfer to the mother liquor drop containing 20% (v/v) ethylene glycol as a cryoprotectant. Diffraction data from these crystals were collected under cryogenic conditions (~100 K) at a synchrotron radiation source, Australian Synchrotron (Araga et al. 2018), ANSTO at beamline MX2 at a wavelength of 0.9537Å on an ADSC Quantum 315r detector with 1° oscillation and 1sec exposure of 90% attenuated beam per frame.

Protein crystals of native protein were soaked (for 30sec to 5min) with 10 mM of sodium bicarbonate together with mother liquor containing 20% (v/v) ethylene glycol. Similarly, co-

crystallized protein crystals of mutant proteins (E76A/H80A) were soaked for 10min in presence of 5 mM of 2FAA and 5 mM of HCAL. Diffraction data from these crystals were collected under cryogenic conditions (~100 K) at a wavelength of 0.9537Å on an EIGER 16M detector with 0.1°oscillation and 0.1sec exposure of 0% attenuated beam per frame.

For all crystal structures, the diffraction data were indexed and integrated with XDS (Kabsch 2010) and scaled using AIMLESS (Evans & Murshudov 2013). Details of the data collection from apo PcaC and mutant PcaC bound with ligands and its statistics from subsequent processing are presented in Table 3.12.

DATA	**Selenomet PcaC**	**PcaC native**	**PcaC CO_3^{2-} /CO_2 bound**	**PcaC H80A_2FAA**	**PcaC E76A_HCAL**	**PcaC E76A_2FAA**
Total frames	180	180	180	1800	3600	3600
Oscillation (°)	1.00	1.00	1.00	0.10	0.10	0.10
Space group	$P6_1$	$P6_1$	$P6_1$	$P2_12_12_1$	$C2_1$	$C2_1$
Unit cell (Å, °)	a=70.33 b=70.33 c=94.22 α=90.00 β=90.00 γ=120.00	a=71.35 b=71.35 c=94.64 α=90.00 β=90.00 γ=120.00	a=71.49 b=71.49 c=94.42 α=90.00 β=90.00 γ=120.00	a=59.2030 b=116.061 c=120.9130 α=90.00 β=90.00 γ=90.00	a=165.83 b=113.32 c=103.15 α=90.00 β=101.18 γ=90.00	a=166.00 b=113.34 c=103.28 α=90.00 β=101.18 γ=90.00
Resolution (Å)	2.64	1.96	2.12	2.01	2.30	2.28
Mosaicity (°)	0.329	0.326	0.324	0.213	0.176	0.151
Total reflections	87896 (13960)	213999 (34194)	177065 (28458)	380144 (61533)	583316 (94780)	587758 (87152)
Unique reflections	15532 (2455)	37561 (6064)	30623 (4972)	106989 (17145)	163425 (26172)	167310 (26417)
Redundancy	5.65 (5.68)	5.70 (5.64)	5.78 (5.72)	3.55 (3.59)	3.5 (3.62)	3.51 (3.30)

Completeness (%)	99.6 (98.2)	97.6 (97.8)	99.9 (99.7)	99.6 (98.6)	99.4 (98.6)	99.3 (97.5)
I/σ(I)	17.83 (3.35)	18.18 (2.78)	20.24 (2.82)	11.43 (3.09)	12.12 (3.82)	9.70 (2.30)
R_{merge} (%)	7.0 (56.1)	5.2 (47.9)	5.8 (59.0)	6.6 (34.1)	7.1 (31.4)	9.1 (48.8)
R_{meas} (%)	7.7 (61.8)	5.7 (52.8)	6.3 (64.9)	7.7 (40.1)	8.4 (36.8)	10.7 (58.3)
Wilson-B ($Å^2$)	43.38	42.37	45.68	37.24	39.43	41.37

Table 3-12: : Crystallographic data-collection statistics for Selenomet PcaC, apo PcaC, PcaC bound with CO_3^{2-}/CO_2 and mutants bound with ligands ligands (Values inside the bracket are for the outermost resolution shell).

3.10 Structure determination and refinement of Selmet PcaC, PcaC and its mutant proteins

3.10.1 Structure determination of PcaC

Automated experimental phasing was carried out using single anomalous diffraction (SAD) phasing protocol of Auto-Rickshaw (Panjikar et al. 2005, 2009) with the selenomet . The input diffraction data were prepared and converted for use in Auto-Rickshaw using programs of the CCP4 suite. FA values were calculated using the program SHELXC (Uso´n & Sheldrick 2018) The maximum resolution for substructure determination and initial phase calculation was set to 4.0 Å. All of the 14 heavy atoms requested were found using the program SHELXD and the absolute configuration (hand) of the substructure was determined using the program ABS. The occupancy of all substructure atoms was refined, and initial phases were calculated using the program PHASER (Mccoy et al. 2007). Density modification and phase extension were performed to 2.64 Å resolution using PARROT (Cowtan 2010). A partial model (262 residues out of the total number of 304 residues) was produced using BUCCANEER (Cowtan 2006). Further the native data and the partial model were used in MRSAD protocol of Auto-Rickshaw (Panjikar et al. 2005, 2009) for phase enhancement and model completion. The resulting model contained 87% of the total number of residues. The model was further improved iteratively using manual model building in COOT (Emsley et al. 2010) and using refinement against the native data set at 2.64 Å using REFMAC5 (Murshudov et al. 2011).

3.10.2 Structure determination PcaC-ligand complexes

The Ligand–complex structures were solved using the MR protocol of the software pipeline Auto-Rickshaw (Panjikar et al. 2005, 2009). The native PcaC structure was used as a starting model for solving the structure of the ligand-complexes of PcaC. The two molecules of the PcaC were used as the search model for molecular replacement. Within the software pipeline, MR was performed using the program using MOLREP (Vagin & Teplyakov 2000) and rigid-body, positional and B-factor refinement were carried out using the program CNS to 3.0 Å resolution. The structure was then further refined using REFMAC5 (Murshudov et al. 2011) to its maximum resolution. The resulting model was further improved by rebuilding in the graphics program COOT (Emsley et al. 2010). The resulting model for each complex, was refined in REFMAC5 (Murshudov et al. 2011) and water molecules were added. The resulting map of each ligand complex was analysed and difference density for the ligand was located using the graphic program COOT. The CIF file for the ligand was prepared using the program ACEDRG (Bailey 1994) and used for the ligand building in COOT. The quality of the final model was validated with MolProbity (Chen et al. 2010). The refinement statistics for apo PcaC and mutant PcaC bound with ligands are shown in Table 3.13.

DATA	**PcaC native**	**PcaC CO_3^{2-} /CO_2 bound**	**PcaC H80A_2FAA**	**PcaC E76A_HCAL**	**PcaCE76A_ 2FAA**
Resolution (Å)	20-1.97	20-2.12	20-2.01	20-2.28	20-2.58
No. of reflections	18302	14465	55177	83691	56818
R_{work} / R_{free}	0.23317/0.28 628	0.22799/0.28 830	0.17606/0.22 859	0.18090/0.217 66	0.17599/0.21 988
r.m.s.d from target values Bond lengths (Å)	0.014	0.012	0.018	0.015	0.012
r.m.s.d from target values Bond angles (°)	1.669	1.574	1.833	1.756	1.542

Ramachandran analysis Favored regions (%)	243 (98.78%)	243 (98.78%)	763(98.83%)	1511(99.15%)	1511 (99.15%)
Ramachandran analysis Allowed regions (%)	3(1.22%)	3(1.22%)	9(1.17%)	13(0.85%)	13(0.85%)

Table 3-13: Refinement statistics for apo PcaC, PcaC bound with CO_3^{2-}/CO_2 and mutants bound with ligands.

3.11 Mass spectrometry assay of PcaC protein

The mass spectrometry was used to confirm the active site residues by monitoring the presence/absence of substrate peak. The reaction was carried out in 50 mM Tris buffer (pH 7.5) containing 1mM HCAL and 10 μM protein. The 'control' reaction consisted of the all the component except the protein. The reaction mixtures were incubated for 30mins at 37°C and were quenched using excess of methanol. The calculated mass of the substrate is 188.13 gm/mol and the product mass is 144.13 gm/mol. The observed value of the mass peak for the substrate was at 211 (M.W. + Na) and that of product was at 144.1. The presence of substrate peak indicates lack of the reaction, whereas the absence of substrate accompanied by the appearance of product peak, indicating the occurrence of the reaction.

3.12 Bicarbonate soaking method

For bicarbonate soaking, 100 mM sodium bicarbonate and 50% glycerol solution were prepared as stock. The final sodium bicarbonate soaking solution prepared from the stock solutions contained 10 mM of sodium bicarbonate in presence of 20 % (v/v) glycerol (prepared using mother liquor). The apo PcaC crystals were soaked for a minute, 5 minutes and 15 minutes with the sodium bicarbonate soaking solution prior to cryo-cooling in the liquid nitrogen. The frozen crystals were then used for X-ray diffraction data collection. The apo PcaC structure was used for solving the apo-bicarbonate complexes using MR protocol of Auto-Rickshaw (an automated crystal structure determination platform) (Panjikar et al. 2005, 2009). The resulting model for each complex, was refined in the program REFMAC5 (Murshudov et al. 2011) and water

molecules were added in difference map. The quality of the structure iteratively improved by refinement and model building. Finally, the highest peaks of the difference Fourier map analyzed specially those which did not look like density for water molecule. This way difference density for the ligands (CO_3^{2-} and CO_2) were located using the graphic program COOT (Emsley et al. 2010). The triangular positive difference electron density was built as CO_3^{2-} whereas the elongated positive difference electron density was modelled as CO_2.

The structural analysis from the soaked crystals (1min) indicates that bicarbonate is bound at the active site along with two carbon dioxide molecules present at two different location near the active site, marking the product pathway. Soaking of crystal for 5min shows binding of two CO_2, one at the equivalent position of bicarbonate and another at the similar CO_2 position. The longer soaked crystal (15min) shows a clear site which is devoid of bicarbonate or carbon dioxide, demonstrating the complete conversion of bicarbonate into carbon dioxide and the subsequent diffusion of carbon dioxide from the active site.

3.13 Bioinformatics study of the PcaC protein

3.13.1 Structure analysis

Crystal structures of the enzymes belonging to the CMD family from *Legionella pneumophila* (PDB code: 5DIK), *Thermotoga maritima* (PDB code: 1P8C), *Mycobacterium tuberculosis*, (PDB code: 1KNC), *Burkholderia xenovorans* (PDB code: 2Q0T) and *Pseudomonas aeruginosa* (PDB code: 2O4D) were used in this study for structural comparisons. All structural superpositions were performed using the LSQ protocol of COOT. The LSQ or Least Square Fitting protocol in COOT is generally employed when SSM protocol could not superimpose the full 3D structures of proteins. In that case, the LSQ provide the option where the range of residues for alignment can be specified (Emsley et al. 2010). All figures showing structural representation were prepared with PyMol (DeLano 2002).

3.13.2 Sequence analysis

Sequence Similarity alignment

Sequence similarity networks is used to establish relationship between proteins super families from the sequence similarity perspective which aids in visualization of the protein functioning. The reason for the development of this method is the ever-increasing repertoire of biological data, especially protein sequences. The advantage of this technique is that it employs independent pairwise alignment between protein sequences instead of multiple sequence alignments, showcasing all protein connections rather than ones which are above the user specified similarity cut off. The similarity network can also be used to decipher different features that merge into clusters (or not) when observed from sequence similarity viewpoint (Atkinson, Morris, Ferrin, & Babbitt 2009).

In this study, in order to identify active site of PcaC, a sequence similarity network of CMD family (https://pfam.xfam.org/family/CMD) was created. BLAST with an e-value cut-off of e–40 was performed using BLAST+ and BLAST2 sim plugin software from the NCBI site (Altschul et al. 1997; Camacho et al. 2009) which generated a sequence alignment for each pair (with at least 40% sequence identity), and the output was loaded into Cytoscape (Shannon et al. 2003) to visualize the network as evolutionarily distinct groups.

ConSurf is an online program (https://consurf.tau.ac.il/) that mark the biologically relevant residues of the protein using an evolution-based conservation score. From the evolution point of view, the functionally essential residues that form the active site, partake in substrate binding, and involve in protein-protein interaction are conserved and are usually solvent accessible. Whereas, the residues which help in maintaining the protein's architecture and have structural purpose are also evolutionarily conserved but are part of the protein core. Thus, based on the conservation score and solvent accessibility predictions, each residue in protein sequences are identified as having structural or functional potential (Landau et al. 2005). This webserver is used to predict the functionally relevant residues in PcaC sequence and acquire clues for identifying the active site residues.

MultiAlin (http://multalin.toulouse.inra.fr/multalin/) is another example of program that can align more than three nucleotide or protein sequences. It utilizes the programming process of

pairwise alignment to form a hierarchical clustering. This program also has a web tool where the query sequences are entered in the input box and the alignment is obtained. There is a choice of optional parameters, where the gap penalty, consensus levels etc. can be changed from default (Corpet 1988). In this study, the representative CMD protein from each cluster of sequence similarity network are aligned with the PcaC sequence to determine the equivalent active site residues in the PcaC.The sequence alignment figure used in this study was made from this online tool.

3.13.3 Tunnel analysis and calculation of CO_2 path

The folding of protein into 3D structures often results in structural feature such as tunnels, cavities, pores, grooves, clefts etc. These features are of interest, as they represent the spots where the substrates, ions and other biological molecules (DNA/RNA/protein) interact. HOLE (Smart et al. 1996) is one such program that is used to detect the presence of these biologically important features.

In order to investigate pathway of the CO_2 as a one of the products from the decarboxylation reaction into internal cavity to the protein surface in PcaC, channel/pore calculations were conducted using the program HOLE (Smart et al. 1996)as implemented in the graphic program COOT(Emsley et al. 2010). Hexamer of the native protein without acetate was generated from the symmetry and used for the channel/pore calculation. The hexameric structure was overlaid with sodium bicarbonate complex structure, and C-atom co-ordinate of the CO_3 molecule from a monomer was used as a starting point for the channel/pore calculations. Co-ordinate of the midpoint between NE2 atom of His-22 of one monomer and NE2-atom of His-115 of another monomer was used as an end point.

3.13.4 Modelling of homocitrate acid lactone in its catalytically active state

The co-ordinate and cif files of homocitrate acid lactone (HCAL) were created at the PRODRG server (Schüttelkopf & Aalten 2004). The ligand files and the native structure of the PcaC without the ACT molecule were used for docking analysis by using the software AutoDock.

Automated docking was used to locate the appropriate binding orientation and conformation of the HCL in the ACT binding pocket. To perform the task, genetic algorithm routine implemented in the program AutoDock 4.0. was employed (Morris et al. 1998). All water

molecules were removed from the original Protein Data Bank file. Polar hydrogen atoms were added and Kolllman charge, atomic solvation parameters and fragmental volumes were assigned to the protein using AutoDock Tools (ADT). For docking calculations, non-polar hydrogen atoms were merged. All torsions were allowed to rotate during docking.

The program AutoGrid used to generate the grid maps. Each grid was centered at the crystal structure of the PcaC bound ligand ACT. The grid dimensions were $31 \times 31 \times 31$ Å^3 with points separated by 0.375 Å. Lennard-Jones parameters 12-10 and 12-6, supplied with the program, were used for modeling H-bonds and van der Waals interactions, respectively. The distance-dependent dielectric permittivity of Mehler and Solmajer (Mehler & Solmajer 1991) was used for calculation of the electrostatic grid maps. For the ligand, random starting positions, random orientations and torsions were used. The translation, quaternion and torsion steps were taken from default values in AutoDock. The Lamarckian genetic algorithm and the pseudo-Solis and Wets methods were applied for minimization using default parameters. The standard docking protocol for rigid and flexible ligand docking consisted of 10 independent runs, using an initial population of 50 randomly placed individuals, with 2.5×10^6 energy evaluations, a maximum number of 27000 iterations, a mutation rate of 0.02, a crossover rate of 0.80, and an elitism value of 1. The probability of performing a local search on an individual in the population was 0.06, using a maximum of 300 iterations per local search. After docking, the 10 solutions were clustered into groups with RMS deviations lower than 1.0 Å. The clusters were ranked by the lowest energy representative of each cluster. The lowest energy docked HCAL ligand in PcaC was further energy minimized using REFMAC5 (Murshudov et al. 2011) without the X-ray term.

3.14 References

Altschu, SF, Gish, W, Miller, W, Myers, EW, & Lipman, DJ 1990, 'Basic local alignment search tool' *The Journal of Molecular Biology*, vol. 215, pp. 403–410.

Altschul, SF, Madden, TL, Schäffer, AA, Zhang, J, Zhang, Z, Miller, W, & Lipman, DJ 1997, 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs' *Nucleic Acids Research*, vol. 25, no. `7, pp. 3389–3402.

Araga, D, Aishima, J, Cherukuvada, H, Clarken, R, Clift, M, Cowieson, NP, Ericsson, DJ, Gee, CL, Macedo, S, Mudie, N, Panjikar, S, Price, JR, Riboldi-tunnicliffe, A, Rostan, R, Williamson, R, & Caradoc-Davies, TT 2018, 'MX2 : a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron' *Journal of Synchrotron Radiation*, no. 25, pp. 885–891.

Atkinson, HJ, Morris, JH, Ferrin, TE, & Babbitt, PC 2009, 'Using sequence similarity networks for visualization of relationships across diverse protein superfamilies' *Plos One*, vol. 4, no. 2, pp. 1–14.

Bailey, S 1994, 'The CCP4 suite : Programs for protein crystallography' *Acta Crystallographica Section D*, vol. D50, pp. 760–763.

Biosystems, Kapa, Website: 'https://www.kapabiosystems.com/productapplications/applications/pcr/mutagenesis/'.

Brezovsky, J, Chovancova, E, Gora, A, Pavelka, A, Biedermannova, L, & Damborsky, J 2012, 'Software tools for identification, visualization and analysis of protein tunnels and channels' *Biotechnology Advances*, vol. 31, no. 1, pp. 38–49.

Camacho, C, Coulouris, G, Avagyan, V, Ma, N, Papadopoulos, J, Bealer, K, & Madden, TL 2009, 'BLAST + : architecture and applications' *BMC Bioinformatics*, vol. 10, no. 421, pp. 1–9.

Chen, VB, Arendall, WB, Headd, JJ, Keedy, DA, Immormino, RM, Kapral, GJ, Murray, LW,

Richardson, JS, & Richardson, DC 2010, 'MolProbity: All-atom structure validation for macromolecular crystallography' *Acta Crystallographica Section D*, vol. D66, pp. 12–21.

Chovancova, E, Pavelka, A, Benes, P, Strnad, O, Brezovsky, J, Kozlikova, B, Gora, A, Sustr, V, Klvana, M, Medek, P, Biedermannova, L, Sochor, J, & Damborsky, J 2012, 'CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures' *PLoS Computational Biology*, vol. 8, no. 10, p. e1002708.

Corpet, F 1988, 'Multiple sequence alignment with hierarchical clustering' *Nucleic Acids Research*, vol. 16, no. 22, pp. 10881–10890.

Cowtan, K 2006, 'The Buccaneer software for automated model building. 1. Tracing protein chains' *Acta Crystallographica Section D*, vol. D62, pp. 1002–1011.

Cowtan, K 2010, 'Recent developments in classical density modification' *Acta Crystallographica Section D*, vol. D66, pp. 470–478.

DeLano, WL 2002, 'PyMOL: An open source molecular graphics tool, DeLano Scientific, San Carlos', DeLano Scientific, San Carlos.

Emsley, P, Lohkamp, B, Scott, WG, Cowtan, K, & Cowtand, K 2010, 'Features and development of Coot' *Acta Crystallographica Section D*, vol. D66, pp. 486–501.

Evans, PR & Murshudov, GN 2013, 'How good are my data and what is the resolution?' *Acta Crystallographica Section D*, vol. D69, pp. 1204–1214.

Ithayaraja, M, Janardan, N, Wierenga, RK, Savithri, HS, & Murthy, MRN 2016, 'Crystal structure of a thiolase from Escherichia coli at 1.8 Å resolution' *Acta Crystallographica Section F*, vol. F72, pp. 534–544.

Kabsch, W 2010, 'XDS' *Acta Crystallographica Section D*, vol. D66, pp. 125–132.

Krissinel, E & Henrick, K 2004, 'Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions' *Acta Crystallographica Section D*, vol.

D60, pp. 2256–2268.

Landau, M, Mayrose, I, Rosenberg, Y, Glaser, F, Martz, E, Pupko, T, & Ben-tal, N 2005, 'ConSurf 2005 : The projection of evolutionary conservation scores of residues on protein structures' *Nucleic Acids Research*, vol. 33, pp. W299-302.

Langer, G, Cohen, SX, Lamzin, VS, & Perrakis, A 2008, 'Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7' *Nature Protocols*, vol. 3, no. 7, pp. 1171–1179.

Mccoy, AJ, Grosse-kunstleve, RW, Adams, PD, Winn, MD, Storoni, LC, & Read, RJ 2007, 'Phaser crystallographic software' *Journal of Applied Crystallography*, vol. 40, pp. 658–674.

Mehler, EL & Solmajer, T 1991, 'Electrostatic effects in proteins : Comparison of dielectric and charge models' *Protein Engineering*, vol. 4, no. 8, pp. 903–910.

Morris, GM, Goodsell, DS, Halliday, RS, Huey, R, Hart, WE, Belew, RK, & Olson, AJ 1998, 'Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function' *Journal of Computational Chemistry*, vol. 19, no. 14, pp. 1639–1662.

Murshudov, GN, Skubák, P, Lebedev, AA, Pannu, NS, Steiner, RA, Nicholls, RA, Winn, MD, Long, F, & Vagin, AA 2011, 'REFMAC5 for the refinement of macromolecular crystal structures' *Acta Crystallographica Section D*, vol. D67, pp. 355–367.

Panjikar, S, Parthasarathy, V, Lamzin, VS, Weiss, MS, & Tucker, PA 2005, 'Auto-Rickshaw: An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment' *Acta Crystallographica Section D*, vol. D61, pp. 449–457.

Panjikar, S, Parthasarathy, V, Lamzin, VS, Weiss, MS, & Tucker, PA 2009, 'On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination' *Acta Crystallographica Section D*, vol. D65, pp. 1089–1097.

Schüttelkopf, AW & Aalten, DMF van 2004, 'PRODRG : A tool for high-throughput crystallography of protein-ligand complexes' *Acta Crystallographica Section D*, vol. D60, pp. 1355–1363.

Shannon, P, Markiel, A, Ozier, O, Baliga, NS, Wang, JT, Ramage, D, Amin, N, Schwikowski, B, & Ideker, T 2003, 'Cytoscape : A software environment for integrated models of biomolecular interaction networks' *Genome Research*, vol. 13, pp. 2498–2504.

Sievers, F, Wilm, A, Dineen, D, Gibson, TJ, Karplus, K, Li, W, Lopez, R, Thompson, JD, Higgins, DG, Mcwilliam, H, Remmert, M, & So, J 2011, 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega' *Molecular Systems Biology*, vol. 7, no. 539, pp. 1–6.

Smart, OS, Neduvelil, JG, Wang, X, Wallace, BA, & Sansom, MSP 1996, 'HOLE : A program for the analysis of the pore dimensions of ion channel structural models' *Journal of Molecular Graphics*, vol. 14, pp. 354–360.

Stern, JR 1956, 'Optical properties of acetoacetyl-X-coenzyme and its metal chelates' *Journal of Molecular Biology*, vol. 221, pp. 33–44.

Thorn, A & Sheldrick, GM 2011, 'ANODE : Anomalous and heavy-atom density calculation' *Jouurnal of Applied Crystallography*, vol. 44, pp. 1285–1287.

Tian, W, Chen, C, Lei, X, Zhao, J, & Liang, J 2018, 'CASTp 3.0: computed atlas of surface topography of proteins' *Nucleic Acids Research*, vol. 46, no. W1, pp. W363–W367.

Uso´n, I & Sheldrick, GM 2018, 'An introduction to experimental phasing of macromolecules illustrated by SHELX ; new autotracing features' *Acta Crystallographica Section D*, vol. D74, pp. 106–116.

Vagin, A & Teplyakov, A 2000, 'An approach to multi-copy search in molecular replacement' *Acta Crystallographica Section D*, vol. D56, pp. 1622–1624

Chapter 4
Structural insight into the degradative thiolase PcaF: Degradative Vs biosynthetic thiolase

4.1 Introduction

Thiolases, also known as acetyl-coenzyme A acetyltransferases, are a prevalent class of enzymes that are found in both prokaryotes and eukaryotes. The enzymes of this broad class partake in diverse biochemical pathways ranging from fatty acid metabolism to bacterial aromatic compound degradation and are subdivided into two categories: the degradative thiolases and the biosynthetic thiolases (Haapalainen, Meriläinen, & Wierenga 2006). Degradative or 3-ketoacyl-CoA thiolases are involved in the thiolytic breakdown of β-ketoacyl-CoA to acetyl-CoA and shorter acyl CoA molecule (Haapalainen et al. 2006; Magali Mathieu et al. 1997) and can accommodate long acyl chain substrates. This catalytic cleavage by degradative thiolases is vital as it accounts for the energy production in cells. Biosynthetic or the acetoacetyl-CoA thiolases catalyze the claisen condensation of two acetyl CoA molecules to give longer chain acetoacetyl CoA (Davis, Moore, et al. 1987; Modis & Wierenga 2000). The biosynthetic class prefers short acyl chains of up to 4 carbon atoms, this group predominantly occurs in the fatty acid and polyketide biogenesis (Kursula et al. 2005).Mostly, all thiolases have the ability to catalyze both the biosynthetic and degradative reactions, however, the degradative reaction is thermodynamically favourable. The two classes of thiolases have related sequences and essentially use the same active site residues to perform the relevant reaction, implying a common ancestor. Prior bioinformatics studies have shed some light into the origin of thiolases and have helped in classifying unknown thiolase sequences into classes based on signature catalytic loop sequence (Anbazhagan et al. 2014). Furthermore, phylogenetic studies of thiolase superfamily have indicated the ‘ancestral origin’ of archaeal thiolase compared to the other enzymes of the superfamily (Jiang, Kim, & Suh 2008). Evolutionary studies of eukaryotes have shown the proteobacterial origin of eukaryotic

thiolases. During the eukaryotic evolution, the loss or gain of sequences in thiolases genes is also shown to dictate the subcellular location of the thiolases as well as determine the biosynthetic and degradative classes of the thiolases (Pereto, Lo´peZ-Garcı´a, & Moreira 2005).

The active site of thiolase is characterized by the presence of two reactive cysteine residues and one histidine residue. In case of biosynthetic thiolase the active site residues are C89, H348 and C378, which corresponds to the residues C90, H356, and C386 respectively, in PcaF. The study of the architecture of the active site pocket divulge a deep cylindrical tunnel with two catalytic cysteine residues present at the bottom of the tunnel and histidine at the side of the tunnel. The active site cavity residues in PcaF comprises of conserved catalytic residues C90, H356, and C386 along with the residues N241, F325, S162, M163, M294, S253, V255, I145 and L358. The N8 atom of the CoA moiety from hydrogen bond with oxygen of S253. Two more interaction of the CoA molecule were observed, one with the T229 another with R226, which present outside the active site activity. The mechanism proposed in biosynthetic thiolase is a two-step process. In the first step the C89 is activated by the H348 residue. The C89 makes a nucleophilic attack towards the acetyl CoA molecule, resulting in an intermediate where the C89 is acylated and the CoA moiety is deprotonated. C378 donates its' proton to the CoA, allowing the CoA to leave the active site. C378, now negatively charged, is the catalytic base. In the second step, another acetyl CoA molecule enters the active site and is deprotonated by the C378; forming an enolate-acetyl-CoA intermediate. The C89 transfer the acetyl group to the enolate intermediate, leading to formation of the product acetoacetyl-CoA.

Till date a large body of work has been focused towards deciphering structural function and mechanism of biosynthetic thiolases (Davis, Chen, et al. 1987; Kursula et al. 2005; Kursula, Ojala, Lambeir, & Wierenga 2002). However, not much work has been performed on developing perspectives on the functioning of degradative thiolases. Neither efforts to fully distinguish factors that distinguish the two classes have been entailed. Hence, to develop a broader understanding on degradative thiolases, in this study, a β-ketoadipyl-CoA thiolase (PcaF) from the β-ketoadipate pathway is explored. This enzyme belongs to *Pseudomonas putida* KT2440 which is a well-studied soil bacterium that can assimilate a variety of aromatic compounds. It possesses four major pathways that degrade aromatics in homogentisate pathway (*hmg/fah/mai* genes), the phenylacetate pathway (*pha* genes), the catechol (*cat* genes) and the protocatechuate (*pca* genes) pathway (Abril, Michan, Timmis, & Ramos 1989; José Ignacio Jiménez, Miñambres, García, & Díaz 2002; Yamanashi, Kim, Hara, & Funa 2015).

The catechol and protocatechuate pathways converge to form the central β-ketoadipate pathway which is a taxonomically widespread bioremediation pathway in bacteria (Figure 4-1) (Ornston 1966b, 1966a, 1966c). In this study, a degradative thiolase, β-ketoadipyl-CoA thiolase (PcaF) from the β-ketoadipate pathway is explored. The *pcaF* gene is located in the gene cluster *pcaRKF* and encodes the last enzyme of the β-ketoadipate pathway involved in the conversion of β-ketoadipyl-CoA into succinyl-CoA and acetyl-CoA (TCA cycle intermediates) (J I Jiménez, Miñambres, García, & Díaz 2002; Kim et al. 2006; Nelson et al. 2002). PcaF is a pivotal enzymatic player of the β-ketoadipate pathway as it concludes the biocatalysis of the toxic compounds from catechol and protocatechuate pathways into non-lethal metabolic intermediates, which enter the TCA cycle (Figure 4-1) and aids in energy production in the bacteria.

To understand the active site architecture of the degradative thiolases, single and double active site mutants of PcaF were made; H356A mutant(A-mutant), H356A-C90S mutant (AS-mutant), H356A-C386A mutant (AA-mutant). Additionally, the crystal structures of the apo PcaF as

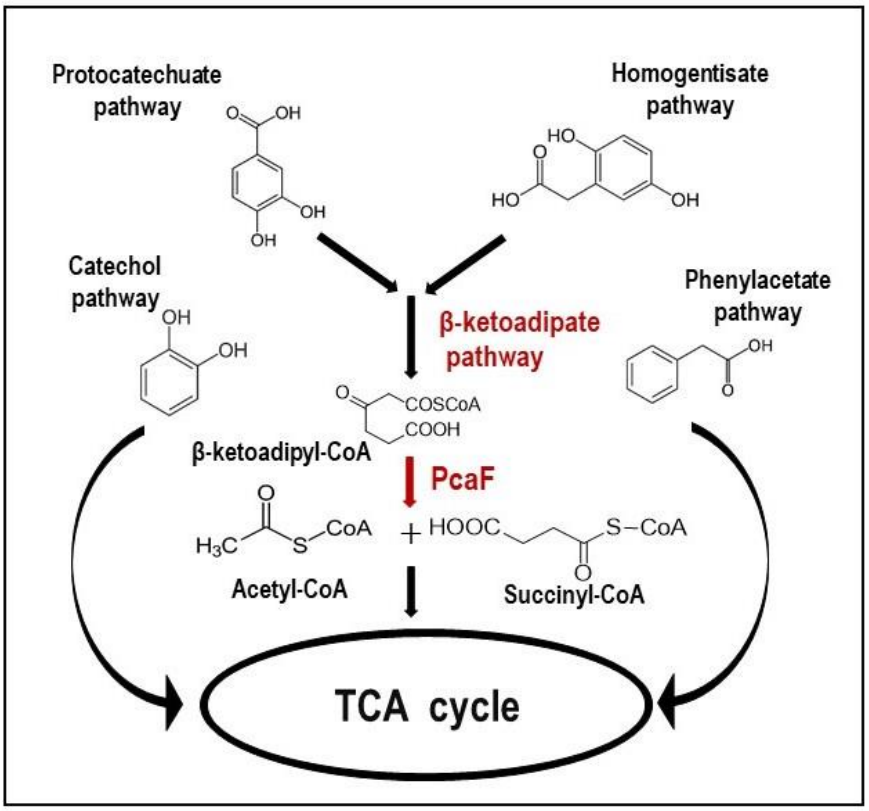

Figure 4-1 Schematic representation of the microbial aromatic degradation pathways involving protocatechuate, homogentisate, catechol and phenylacetate pathways. The protocatechuate and catechol pathways combine to form the β-ketoadipate pathway which is preceded by the thiolase PcaF in the last step.

well as the active site variants in-complex with CoA, Hexanoyl CoA (Hex-CoA) and Octanoyl CoA (Oct-CoA) were solved to deduce the structural pre-requisite required for catalysis in degradative thiolases. The structures of A-mutant complexed with CoA and in complex with Hex-CoA were determined. In both A-mutant complexes CoA was found to be covalently linked to the active site C90 residue rendering the enzymes' active site to be blocked. Further, bioinformatics analyses were performed and together the information obtained was vital in understanding the unique features of degradative thiolases. For instance, it was found that both a flexible active site architecture as well as a long tunnel in PcaF renders it capable of accommodating longer chain acyl CoA derivatives. The work encompasses in depth structural and evolutionary analyses and provides key structural evidences that differentiate degradative verse biosynthetic group of thiolases.

4.2 Overall structure and the active site architecture of PcaF

The 1.81 Å apo PcaF structure was solved by the molecular replacement (MR) method using a putative acetyl-CoA acetyltransferase (PDB code: 1ULQ, 52% sequence identity). The four molecules of PcaF in an asymmetric crystallographic unit forms a tetramer with an extensive

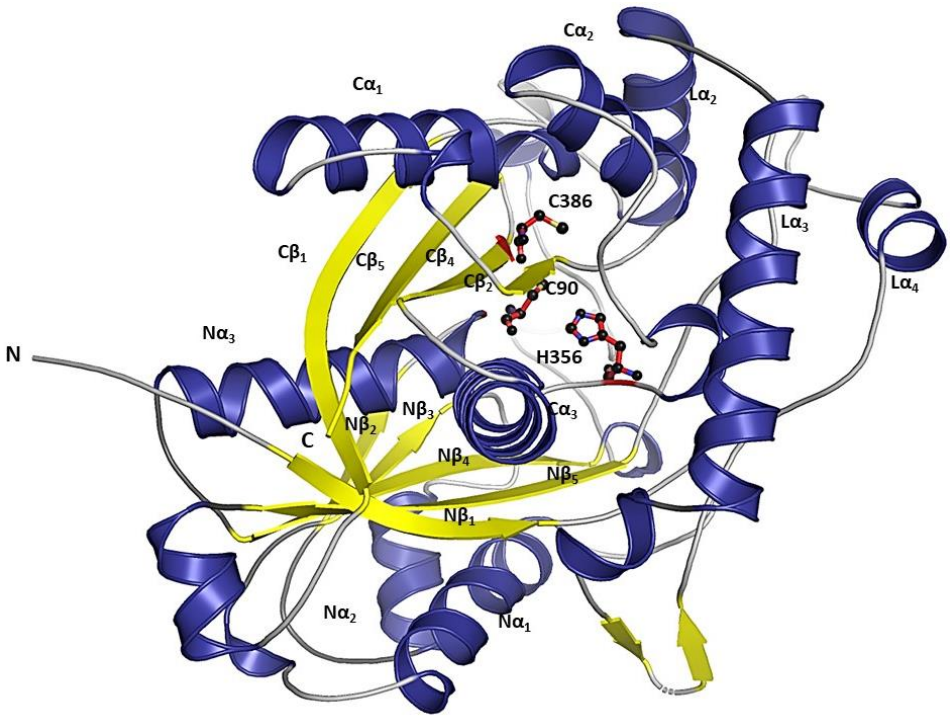

Figure 4-2 The structural assembly of PcaF resulting in a characteristic five layered thiolase superfamily arrangement.

buried surface area of 21130 $Å^2$ which is 2.5 time larger than that of PcaF dimer as calculated using PISA (Krissinel & Henrick 2007). The oligomeric state has been seen in a number of other thiolases (Ithayaraja et al. 2016; Pye et al. 2010). The monomers B and C have a continuous density for residues -2 to 400 unlike A and D subunits that lack density for residues 212-215. Each monomer is comprised of the N-domain (residues 1-119 and 256-278), the C-domain (residues 279- 400) and the loop-domain (residues 120–255). The overall structural assembly of PcaF comprises of an α/β topology β1α1β2α2β3α3β4β5 together forming a characteristic five layered thiolase superfamily arrangement (Figure 4-2) (Haapalainen et al. 2006; M Mathieu et al. 1994; Magali Mathieu et al. 1997). Active site residues (Cys90, His356 and Cys386) of PcaF are identical to that of the catalytic triad of biosynthetic thiolases.

One of the interesting questions posed was, how the same set of active site residues catalyze reactions in two opposite directions. To discern as a representative case, structural analysis of the degradative thiolase PcaF with the biosynthetic thiolase *Zoogleria ramigera* (Zr-thiolase) (PDB code: 1DLU) was conducted. PcaF active site architecture comprises of N241, 325F, S162, M163, M294, S253, V255, H356 and I145 (Figure 4-2). These residues were compared across the degradative thiolases. Surprisingly, a sequence alignment using degradative and biosynthetic thiolases showed less than 95% homology. It was noted while the catalytic triad and F325 are fully conserved, M163, M294, S253 and L358 have consensus in the range of 80 to 96%, whereas S162 and V255 are highly variable (Figure 4-2).

On further comparison of the native PcaF and Zr-thiolase, it is observed that the Zr-thiolase has an equivalent active site architectures, flanked by residues, F235, F319, H156, M157 and M288 on one side and on the other side by S247, L249, H348, L148 and I350 (Modis & Wierenga 2000) (Figure 4-2, left wall) on the other side. V255, I145, N 241, S162 and L358 are analogous residues in PcaF that form the left CoA pocket lining and corresponding L249, L148, F235, H156 and I350 residues line the other side. The residues comprising the active site architecture in Zr-thiolase are fully conserved across biosynthetic thiolase sequences, except for L249 and I350 residues which are 68% and 96% conserved. Comparison of the active site walls reveal that even though the catalytic triad is conserved in the degradative and the biosynthetic thiolases, the overall active site lining is different (Figure 4-3). The biosynthetic active site is well-demarcated and highly conserved compared to the degradative thiolase which has more flexibility in active site residues. The rigidity observed corroborates with the structural profile of the biosynthetic group which is limited to short four carbon chain acyl CoA and lacks the

versatility and range exhibited by degradative thiolases. As can be seen in Figure 4-3 the active site of degradative thiolases is wider and is adaptable to accommodate a wider substrate variety.

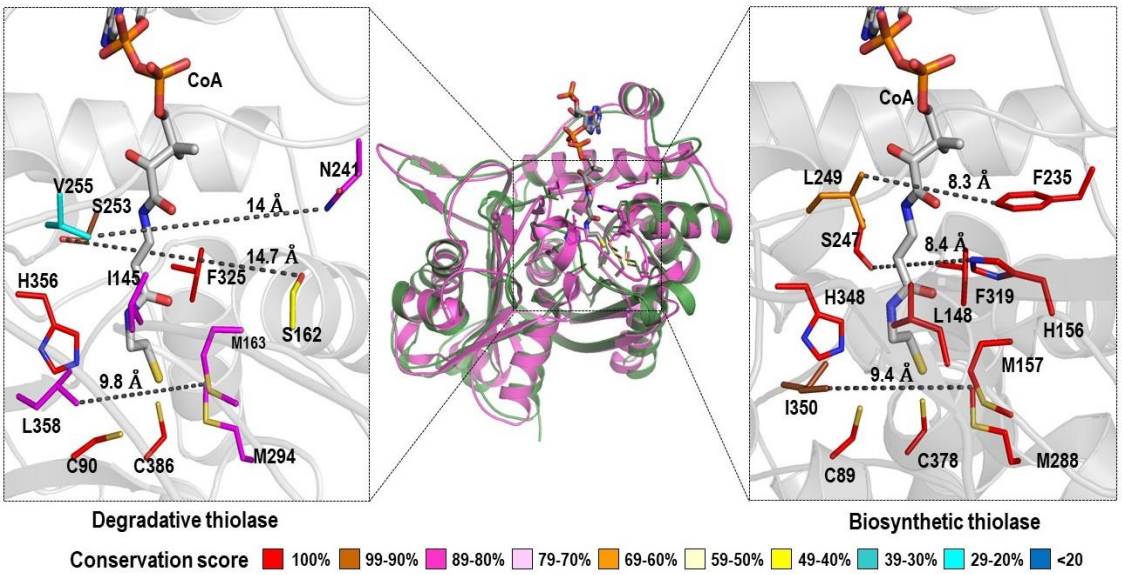

Figure 4-3 Comparison of active site cavity of a representative member of the thiolase superfamily. Superimposition of the PcaF (in green) and Zr-thiolase (PDB code: 1DLV in magenta) indicate a variation in the active site architecture. The active site architecture of the both degradative (inset B) and biosynthetic thiolase (inset C). Residues are colored according to the conservation score. The conservation score is calculated using 100 sequences of degradative and of biosynthetic thiolases respectively.

4.3 Native and mutant ligand complexes

The kinetic parameters of native PcaF were measured and Km (67.8± 8.2 µM) and kcat (0.7±0.08 S^{-1}) for acetoacetyl CoA were observed. Single (H356A: A-mutant) and double active site mutations (H356A-C90S: AS mutant, and H356A-C386A: AA-mutant) were cloned and purified (see materials and methods). The expectation was that these mutants will be inactive therefore, capturing ligand-binding complexes will be easier to undertake. Activity measurements of A-, AS- and AA-mutants revealed that these mutants exhibited 30%, 7% and 5% activity respectively, relative to native PcaF. Due to substantial reduction of activity these

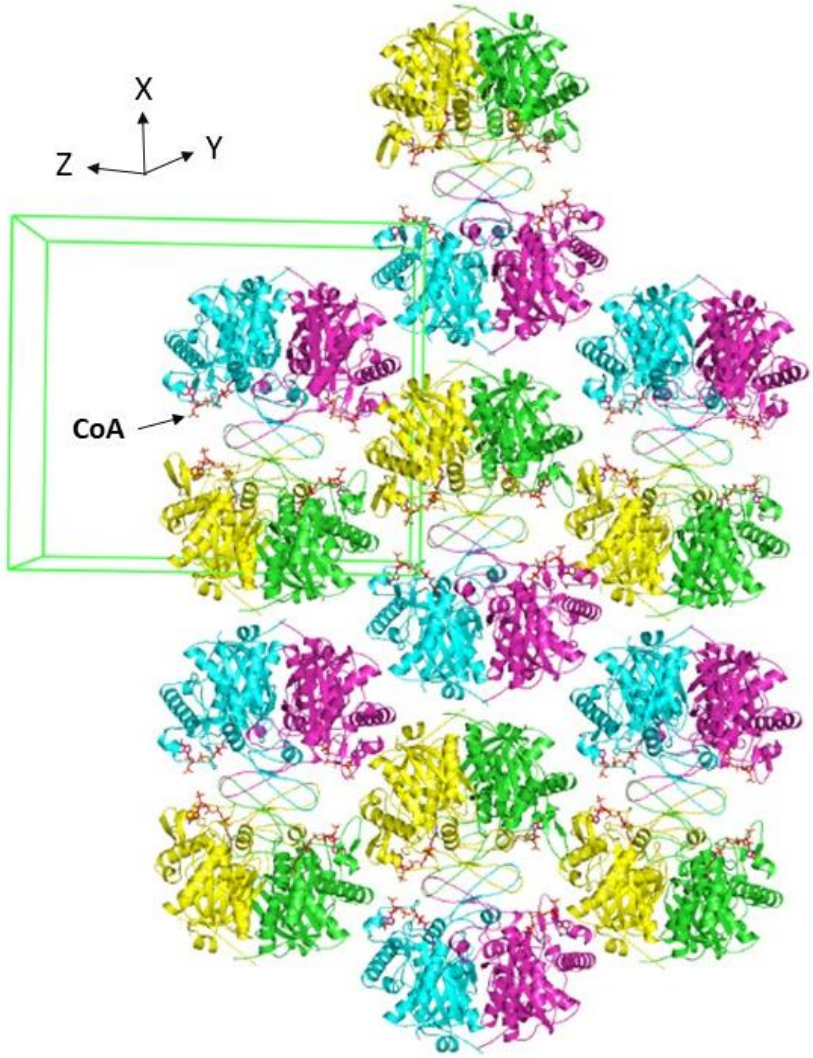

Figure 4-4 Packing diagram for the H356A (A-mutant) crystal in presence of CoA. Unit cell shown as green inline and cell axis direction in black. CoA is marked depicting the accessible channel for substrates in the crystal

three mutants therefore, turned out to be suitable candidates for ligand binding and for trapping a wide range of ligands. β-Ketoadipyl CoA is the natural substrate of PcaF which is not

commercially available. β-Ketoadipyl CoA belong to the last step of aromatic degradation pathway and any studies conducted by using the ligand would have provided insight into the bacterial aromatic bioremediation pathway. Moreover, β-Ketoadipyl CoA would have been a biologically significant substrate for determining the kinetic parameters of PcaF. Nonetheless, the closest analogue to β-Ketoadipyl CoA is the six-carbon acyl Hex-CoA and the crystals of the A-mutant, AS-mutant and AA-mutant were soaked for 20 to 300sec with Hex-CoA and other accessible varieties of CoA derivatives (CoA, Acetyl CoA, Acetoacetyl CoA, Oct-CoA,

and Decanoyl CoA. The packing (Figure 4-4) of the tetramer indicates that binding sites are accessible through the solvent channel given that solvent content of the protein crystals is 48.45%. Altogether 18 datasets were acquired in the resolution range of 2.56 to 1.37 Å and analysed in detail.

In the A-mutant and AA-mutant, it was observed that the CoA and the other longer chain fatty acyl CoAs form a disulfide bond with the C90 active site residue. Consequently, in the higher chain-length CoAs after the -SH atom there is a cleavage to form a covalently linked C90-SCoA adduct, the remaining acyl chain of CoAs exist in the aldehyde form. In the case of AS-mutant, the cleavage after–SH atom also occurred but no disulfide bond was formed. For ligands acetyl CoA, acetoacetyl CoA and decanoyl CoA, only the CoA part was trapped in all the mutants and the missing acyl parts seemed to be released from the protein. Only in A-mutant-Hexanoyl CoA complex (A-mutant-Hex-CoA) and AS-mutant-Octanoyl CoA (AS-mutant-Oct-CoA), both the CoA and the cleaved acyl part are observed and thus a total of four structures scenarios are reported in this study; the apo PcaF, A-mutant-CoA complex, A-mutant-Hex-CoA complex and AS-mutant-Oct-CoA (Table 4-1).

Dataset	PcaF, apo	PcaF, A-mutant CoA complex	PcaF, A-mutant Hex-CoA complex	PcaF, AS-mutant Oct-CoA complex
Unit cell data				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Number of subunits in the asymmetric unit	4	4	4	4
Unit-cell parameters (Å, °)	a = 110.08, b = 113.84, c = 127.42, α = β = γ = 90	a = 110.49, b = 115.96, c = 128.87 α = β = γ = 90	a = 111.39, b = 116.40, c = 127.90, α = β = γ = 90	a = 111.08, b = 116.62, c = 128.56 α = β = γ = 90
Data collection and processing (a Values in parentheses refer to the highest resolution shell)				
Beam line	Australian Synchrotron	Australian Synchrotron	Australian Synchrotron	Australian Synchrotron
Wavelength (Å)	0.9537	0.9537	0.9537	0.9537
Resolution (Å) [a]	20-1.81 (1.92-1.81)	20-1.61 (1.70-1.61)	20-1.96 (2.08-1.96	20-1.37 (1.45-1.37)

Observed reflections [a]	998852 (158022)	1445863 (218460)	697289 (106899)	2371592 (356212)
Unique reflections [a]	281317 (44875)	413663 (65788)	220195 (35163)	677103 (108090)
Data completeness (%)[a]	99.5 (98.1)	99.3 (97.6)	95.8 (94.7)	99.5 (98.1)
<I/σ(I)> [a]	10.28 (2.18)	11.36 (1.85)	8.52 (2.42)	11.96 (2.09)
Multiplicity [a]	3.55 (3.52)	3.49 (3.32)	3.16 (3.04)	3.5 (3.29)
R_{merge} (%) [a]	8.8 (59.6)	6.8 (56.0)	11.3 (51.7)	6.3 (55.0)
R_{meas} (%) [a]	10.4(70.2)	8.1(66.9)	13.6 (62.4)	7.4(65.6)
$CC_{1/2}$ (%) [a]	99.7 (69.6)	99.8 (69.5)	99.2 (74.2)	99.8 (71.0)
Wilson B factor (Å^2)	19.2	17.8	19.0	12.4
Refinement				
R_{work} (%)	15.59	16.07	17.58	13.02
R_{free} (%)	19.58	19.61	20.81	16.88
Number. of atoms				
Protein atoms	11799	11850	11813	11982
Ligand atoms	64	270	286	147
Solvent atoms	995	1324	1216	1835
Model quality				
RMS deviation from ideal value				
Bond length (Å)	0.009	0.009	0.007	0.006
Water Bond angle (°)	1.562	1.606	1.493	1.314
Average B-factor				
Protein atoms (A/B/C/D) (Å^2)	20.1/20.7/ 21.2/20.3	23.4/22.1/ 20.5/21.1	22.7/22.3/ 19.7/19.4	21.9/20.1/ 18.3/18.9
Ligand atoms (Å^2)	27.5	30.2	32.2	35.0
Waters (Å^2)	30.1	30.9	28.1	35.3
Ramachandran plot [b] (b Calculated by MOLPROBITY)				
Most favored regions (%)	96.5	96.8	95.6	95.0
Allowed regions (%)	3.1	2.8	4.0	4.2
Outlier regions (%)	0.4	0.4	0.4	0.8
PDB code entry	6PCA	6PCB	6PCC	6PCD

Table 4-1 Data-collection, processing and refinement statistics

4.4 Covalent locking of the active site

As stated above, structural analysis of the A- and AA-mutants complexed with CoA reveals a disulfide bond formation between the C90 active site residue and the thiol group (–SH atom) of the CoA leading to covalent blocking of the active site (Figure 4-6a, 4-6d). This is an uncustomary thiolase reaction that is not previously reported. It seems that the H356A mutation (A-mutant) compromised the activity of the PcaF leading to a locking of the reaction center (Figure 4-6). In the case of the AS-mutation, the disulfide bond was not detected as a serine is present instead of cysteine. However, the thiol group of CoA is found in a double conformation (Figure 4-5).

Similar mutation in the crystal structure of the Zr-thiolase was reported by Meriläinen et al. (PDB code: 2WL4), where the H348A (equivalent to the H356 position in PcaF) was mutated. However, in Zr-thiolase mutant complexed with CoA without covalent-C89 (corresponding to C90 position in PcaF) was observed (Meriläinen, Poikela, Kursula, & Wierenga 2009).

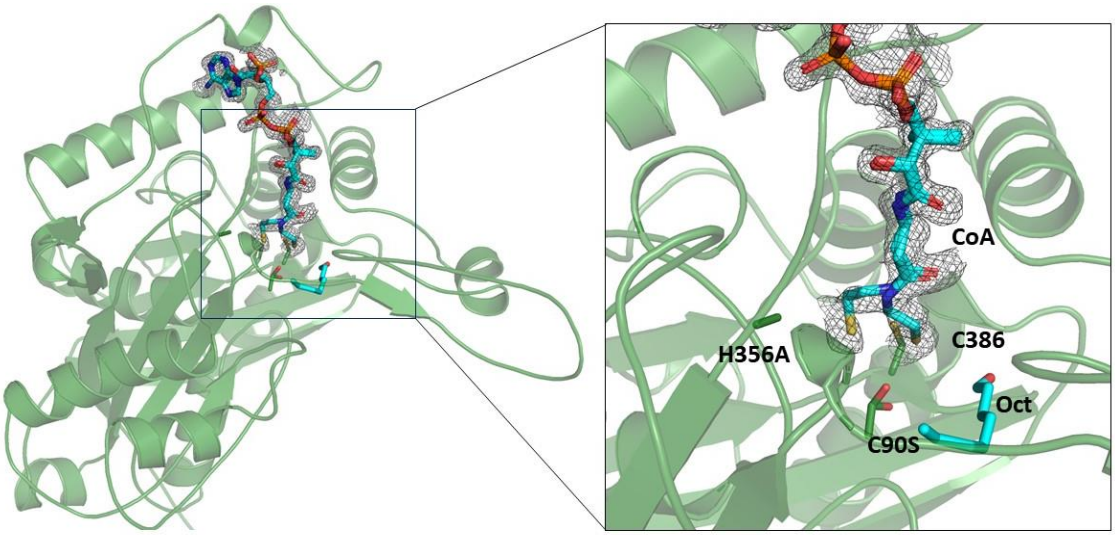

Figure 4-5 Representation of the AS-mutant complexed with Octanoyl CoA. The octanoyl CoA binds to the tunnel of PcaF, where the thiol group of CoA forms double conformation. Difference electron density (mFo − DFc) map for the CoA contoured at 3σ. AS-mutant is marked with H356A and C90S, in which C90S also adopts double conformation. The CoA and the octanoyl moiety (ball and stick model) are shown in cyan.

Therefore, the question arises, why identical mutation of the active site histidine residue in the two types of thiolases yields different binding modes. Analysis of the superimposed A-mutant-

CoA complex of PcaF and Zr-thiolase (PDB code: 2WL4) structures indicates that the presence of S162-M163-P164 (SMP) motif near the active site cavity in the degradative thiolase PcaF which corresponds to conserved H156-M157-G158 (HMG) motif of the Zr-thiolase may be responsible for the difference. The triad found in degradative thiolase appears to form a rigid architecture forcing covalent bond formation. Whereas, in Zr-thiolase this region looks to be flexible and can move to prevent the formation of this deadlock complex. Another potential difference is the occurrence of R65 residue in the degradative thiolase PcaF which corresponds to Q64 in the Zr-thiolase, (both residues are adjacent monomeric subunit of the dimer. H156 is a fully conserved residue across the biosynthetic thiolases and likely act as an "anchoring" residue by mediating interaction with the phosphopantetheine moiety of CoA and holds the substrates in the correct conformation (Figure 4-6a). This is because both the CoA and acyl CoAs are large moieties that require such an anchoring point for the active site activity. In degradative thiolases, the equivalent position is occupied by a variety of residues (Figure 4-6c), indicating the lesser conservation of the said position. In the case of PcaF, it is occupied by S162 where the residue is flipped away from the active site cavity and consequently cannot act as an anchor (Figure 4-6a). Instead, it appears that the H356 residue of the catalytic triad acts as an anchoring residue. Thus, when the H356 residue is mutated to alanine, the tethering point for the substrate is lost and the thiol bond of CoA exposed to the active site. This allows residue C90 to attack other positions instead of the usual Cα and Cβ of 3-ketoacyl CoA. The H356A mutation also causes changes in the neighbouring residues M163' and R65' (from another monomer of the dimer in PcaF). In the apo structure, M163 is located 6.72Å (CE-SG distance) away from C90 and 3.75 Å (CE-NH2 distance) away from R65, whereas in A-mutant-CoA complex structure, the side chain of M163 moves 3.78 Å away from its native conformation which is toward the active site cavity and this allows the R65 residue to undertake in double conformations (Figure 4-6b). Interestingly, side chain of the M157 residue in the biosynthetic thiolase (PDB code: 2WL4), equivalent to the M163 residue of the A-mutant-CoA complex adopts similar conformation to that of in the native PcaF (Figure 4-6a and 4-6b). The varied anchoring sites, high conservation of the active site architecture in the biosynthetic and the

additional role of H356 in the degradative thiolase PcaF emphasizes the difference between the two classes of thiolases.

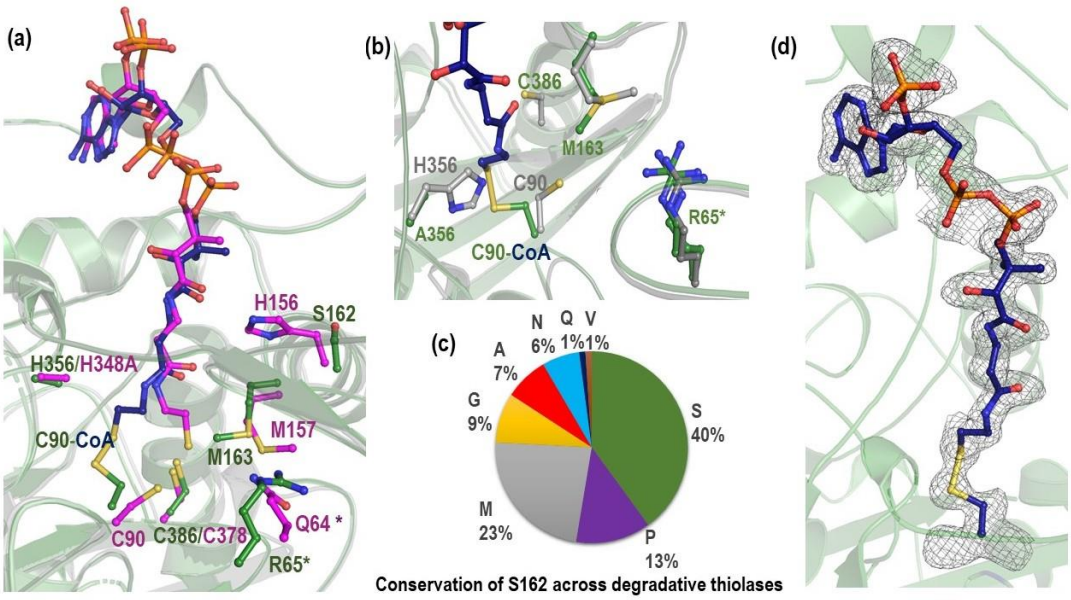

Figure 4-6 Analysis of the covalent locking of the active site in PcaF (a) Superposition of A-mutant−CoA covalent complex of PcaF (the protein is in green and the covalently bound CoA is colored in blue) with H348A mutant-CoA complex of Zr-thiolase (PDB code: 2WL4, magenta). R65 and Q64 residues marked with "" are from the adjacent subunit. (b)Superposition of native PcaF and A-mutant−CoA covalent complex of PcaF colored in grey and green respectively. (c) The pie chart shows the conservation of S162 equivalent position across 100 degradative thiolases. (d) Difference electron density (mFo − DFc) map contoured at 3σ highlighting the covalent linkage between C90 and CoA in A-mutant.*

4.5 Intact mass determination

Intact mass spectrometry was carried out on two set of proteins: the apo protein and the AA-mutant. We have chosen AA-mutant to perform mass spectrometry in presence of CoA derivatives to show that the covalent locking occurs due to the C90 only and not due to the other cysteine (C386) at the active site. The apo PcaF comprises of 400 amino acids along with an additional 22 amino acids from the vector and its molecular weight (MW) is 44454.91 Da, whereas MW of the AA-mutant is 44356.79 Da.

The observed mass spectrometry peak for apo protein and for AA-mutant in the absence of any CoA derivative was 44376.3 Da (Fig. S3a) and 44343.6 Da (Fig. S4a) respectively. There are discrepancies of 78.61 Da less and 13.19 Da more in the measured molecular weight of apo and AA mutant proteins respectively, compared with the calculated values. Such differences in the measurement are due to unknown modifications and is therefore critical to compare experimental values of the control with the modified sample and not using calculated values when determining any changes made to a protein.

The mass spectrometry was carried on the apo protein in presence of acetoacetyl CoA and the same observed peak (44376.3 Da) was detected as for the apo protein alone (Fig. S3b) indicating that the native protein did not undergo covalent locking of CoA (Table 4-2). The observed mass spectrometry peak for the AA mutant in presence of acetoacetyl CoA, hexanoyl CoA and AA-octanoyl CoA were 45110.8 Da, 45110.4 Da and 45110.4 Da indicating increases of 767.2Da, 766.8 Da and 766.8 Da respectively relative to the observed mass peak of AA mutant (Fig. S4). The calculated molecular weight of CoA is 767.5 Da and the observed molecular weight of the modification for the AA-mutant-ligand complexes were ~767 Da. Thus, the intact mass spectrometry confirmed the covalent modification of the AA-mutant and the similar observation has been made for A-mutant.

Samples	**Observed mass**	**Modification observed**
Control apo protein	44376.3 Da	-
apo-ligand complex (apo protein-acetoacetyl CoA)	44376.3 Da	No
Control AA-mutant	44343.6 Da	-
AA-mutant-ligand complex (AA-acetoacetyl CoA)	45110.8 Da	yes
AA-mutant-ligand complex (AA-hexanyol CoA)	45110.4 Da	yes
AA-mutant-ligand complex (AA-octanyol CoA)	45110.4 Da	yes

Table 4-2 Intact mass spectrometry studies using apo protein and AA mutant.

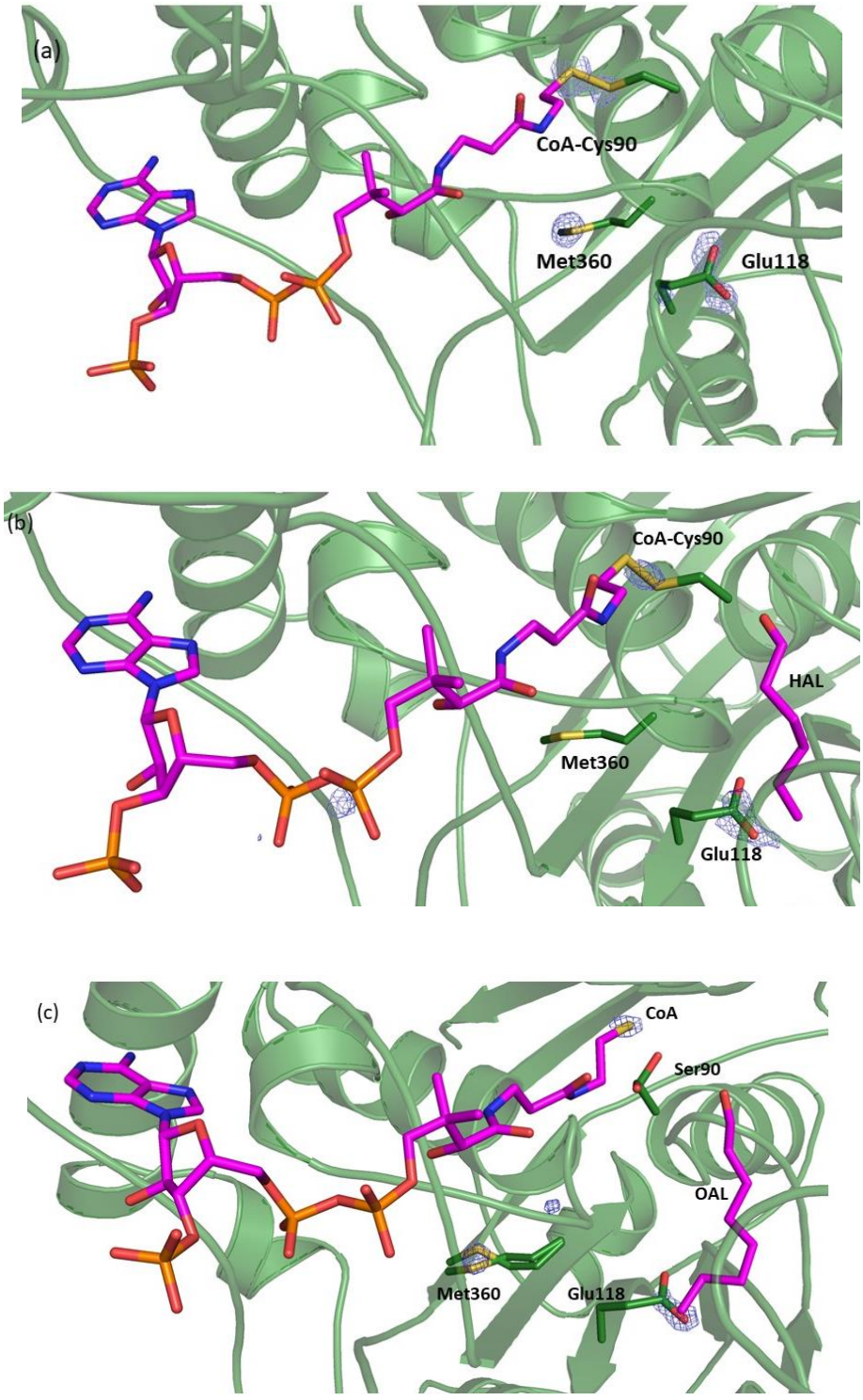

Figure 4-7 Radiation damage analysis indicates decarboxylation of aspartic acid and disordering of methionine sulfur in PcaF. Radiation damage analysis indicates the specific damage sites in (a) A-mutant complex with CoA (b) A-mutant complex with hexanoyl CoA (c) AS-mutant complex with octanoyl CoA. Carbon atoms of the ligand binding residues are shown in green and the ligands (ball and stick model) are shown in pink

4.6 Radiation damage analysis

A-mutant complex with hexanyol CoA dataset and the refined crystal structure were used for radiation damage analysis. The isomorphous difference Fourier map (model phase, $F_{before} - F_{after}$) was calculated as described in Material and Method section. The analysis of the difference Fourier map provides 20 peaks above 5 sigma and highest peak being 6.8 sigma. Most of the peaks are near the Glu and Asp residues indicating decarboxylation due to radiation damage (Figure 4-7). Sulfur atom of the methionine residues (Met-154 and Met-284 from chain A) has also suffered from radiation damage. Interestingly, significant peak was found at SG atom of Cys-90 from chain B and C as well as S1P atom of CoA molecules, which are covalently bound to Cys-90 residues. This indicates that the co-valent bond in chain B and C has suffered from the X-ray radiation damage (Figure 4-7b). However, there was no trace of radiation damage at the oxygen atom of the hexanal (Figure 4-7b). Similarly, in isomorphous difference Fourier map from A-mutant-CoA complex dataset, 6.2 – 5.6σ peaks were found at SG atom of Cys-90 from chain A, B and C and 8.1 – 9.0 σ peaks at S1P atom of CoA molecules in all chains, which are covalently bound to Cys-90 residues (Figure 4-7a), other significant peaks were near the Glu and Asp residues and sulfur atom of the methionine residues.

4.7 Longer chain CoA binding tunnel

The degradative thiolases have the ability to catalyze longer chain substrates compared to the biosynthetic thiolases (Feigenbaum & Schulz 1975). Accordingly, to accommodate the longer aliphatic chains it is expected that they have evolved to harbor long active site tunnel to bind the reactant. As hypothesized in the degradative thiolase PcaF, an interconnected network of tunnels with a long continuous channel running across the dimer is revealed using the program CAVER (Figure 4-8a). In the past several computational studies on these thiolase based systems have suggested that the substrate binding sites usually comprise of a large tunnel/ concavity in the proteins (Byska et al. 2016). Here, using structure we first locate the tunnels. Thus, the tunnel sizes of a single monomer (monomer A) were independently calculated using CASTp (Tian et al. 2018) wherein two major tunnels were observed with surface area 154 $Å^2$ and 235 $Å^2$ respectively. The remaining tunnels have surface area less than 53 $Å^2$, which are too small to form an active site tunnel. The first tunnel with 154 $Å^2$ surface area, corresponds to the tetramerization loop region. The second tunnel with 235 $Å^2$ surface area is located near the

covering loop (Figure 4-8b). Involvement of a covering loop in substrate binding have been proposed in the previous studies (Harijan et al. 2017; Kiema et al. 2014). In PcaF, the covering loop is formed by the loop-domain residues ranging from 143 to 163. The second tunnel located close to the covering loop is the substrate binding site that is validated by the ligand bound structure wherein the acyl tail of the substrate binds.

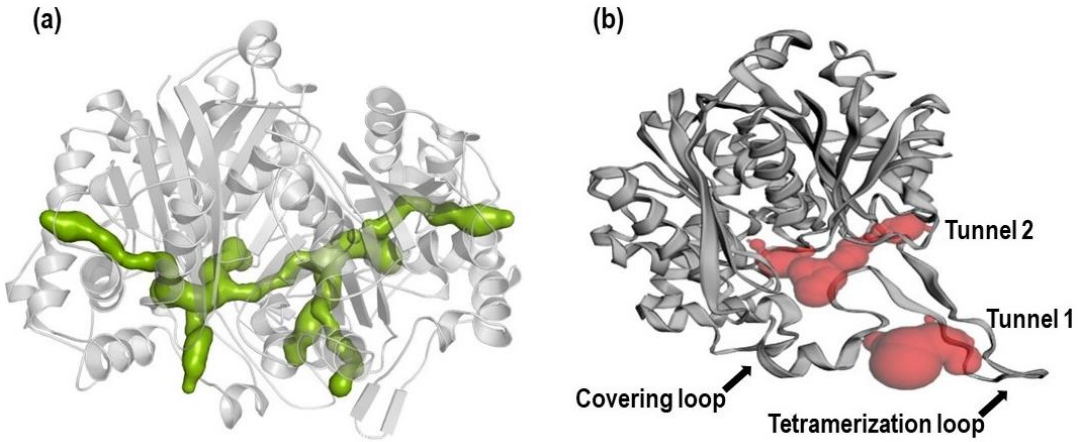

Figure 4-8(a) The interconnected tunnels in dimeric PcaF (calculated using CAVER) (b) The two tunnels (Tunnel 1 and Tunnel 2) in the monomeric PcaF are shown as red surface (calculated using CASTp). Tunnel 1 is located at the tetramerization loop while tunnel 2 located at the covering loop of the PcaF.

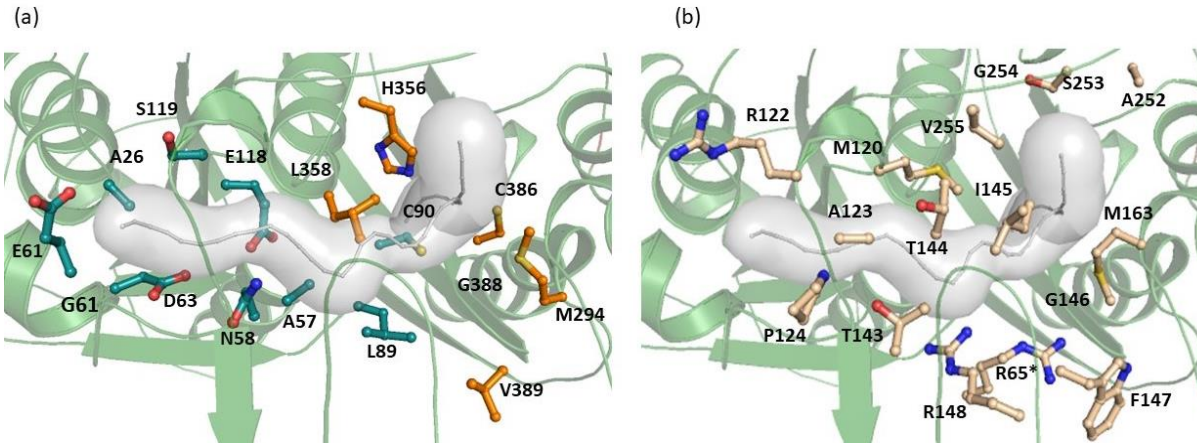

Figure 4-9 The residues partaking in tunnel formation are highlighted. (a) Residues involved in forming the tunnel from N- and C-terminal domain are colored deep teal and orange respectively. (b) The residues from the loop-domain contributing to the tunnel formation are in light yellow.

An in-depth analysis of the tunnel residues reveals that the residues participating in the tunnel formation are a combination of hydrophobic and hydrophilic amino acids. The key residues observed in the tunnel are shown in Figure 4-9. Residues from both the N -terminal and C-terminal domain constitute the tunnel. While the back end of the tunnel is formed by N-terminal residues whereas the front-end is formed by the C-terminal residues (Figure 4-9a). Majority of the tunnel residues belong to the loop-domain (Figure 4-9b). Residues T143-G146 and R148 contribute in tunnel formation belong to the covering loop. Contrastingly, in Zr-thiolase (PDB code: 1DLU) only one tunnel (with surface area of 132 Å2) was observed. Upon comparison Zr-thiolase tunnel with the equivalent tunnel in PcaF, it is observed that the surface area of the biosynthetic thiolase tunnel is almost half that of the degradative thiolase tunnel. This implies that PcaF being a degradative thiolase can accommodate longer chain substrates and is equipped with a larger active site tunnel (Figure 4-9).

4.8 Promiscuity of the degradative thiolase

The A-mutant and AS-mutant when soaked with 3 mM Hex-CoA and Oct-CoA, respectively, led to two complexes. The A-mutant-Hex-CoA complex contains the C90 that is covalently linked to CoA and the cleaved hexanoyl derivative in the form of hexanal (HAL). The AS-mutant-Oct-CoA complex comprises of cleaved Oct-CoA in which the CoA remains free and the cleaved octanoyl derivative in the form of octanal (OAL). None of the previous degradative thiolase structures have been obtained in complexed with any longer chain CoA derivatives. Here, we were able to trap both C6 and C8 acyl CoA chain bound to PcaF, A-mutant-hexanal CoA (A-mutant-HAL-CoA) and AS-mutant-octanal CoA (AS-mutant-OAL-CoA) (Figure 4-10). These structures provide insight into binding of the long chain substrates and helped in identifying the novel tunnel which accommodates longer tail of the substrates. The structures also aid in delineating the residues that partake in interacting with the acyl tail and in understanding the adaptability of the degradative thiolase to bind to a variety of substrates.

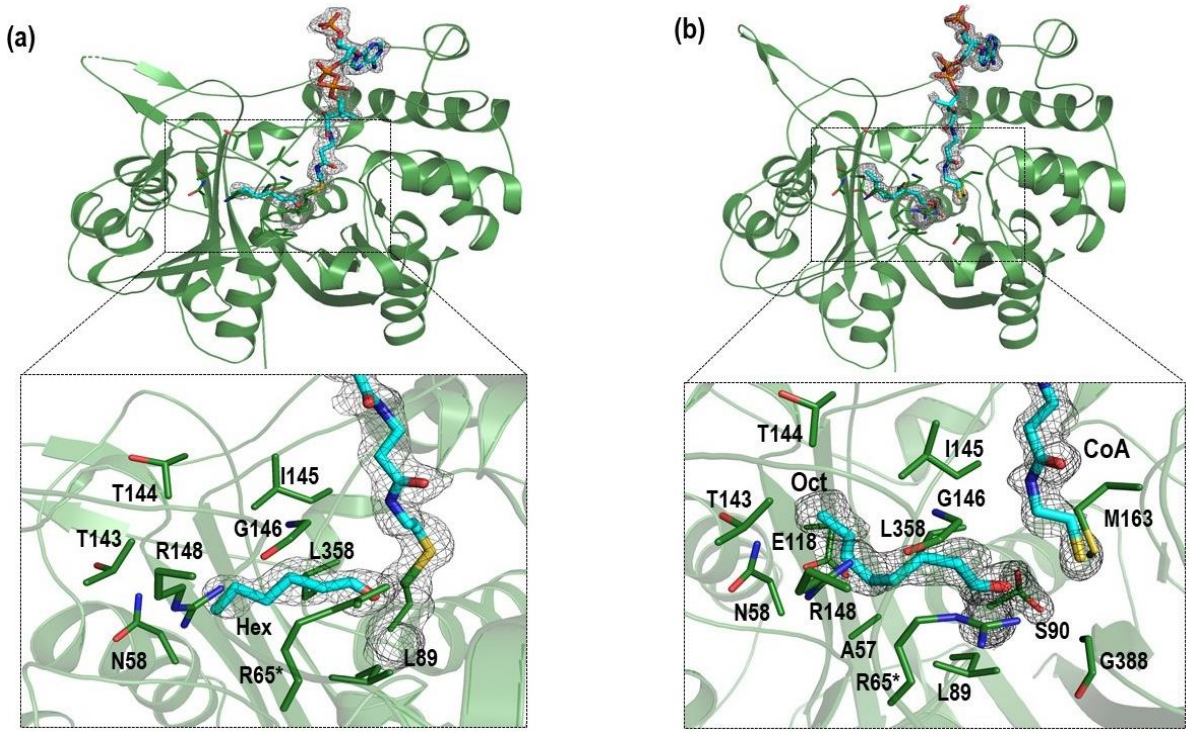

Figure 4-10 Representation of substrate binding in the long tunnel of PcaF. The interaction of the tunnel residues (a) with the hexanal part of hexanoyl CoA-mutant complex (b) with the octanal part of octanoyl CoA-AS mutant complex of PcaF. Carbon atoms of the ligand binding residues are shown in green and the ligands are shown in cyan. R65 indicates the residue from the adjacent subunit. Electron density (mFo−DFc) maps for the ligands are contoured at 3σ.*

HAL and OAL moieties are stabilized in the proposed tunnel via various hydrophobic interactions (Figure 4-10). In fact, there is more space is remaining in the tunnel after accommodating the acyl chains. The calculated surface area of the tunnel is 235Å^2 and the surface areas of the HAL and OAL are 124 Å^2 and 152 Å^2, respectively. An extra 83 Å^2 surface area is still available after the tunnel is occupied by the OAL acyl chain which implies that PcaF can bind to the substrates longer than C8 acyl CoA. As previously mentioned, the acetyl CoA, acetoacetyl CoA and decanoyl CoA ligands were also used for soaking, however the X-ray data analysis showed that only the CoA moiety of acetyl CoA, acetoacetyl CoA and decanoyl CoA were bound in all the mutants. Since the acyl tail of these longer ligands were missing, it indicates that the soaking time (20 – 300sec) might have been too long, given that each of these mutants are still active. HAL was found in all four monomeric units and the binding of the cleaved CoA from Hex-CoA is similar to that of the A-mutant-CoA. The tunnel residues that are engaged in interaction with the acyl part of HAL, are shown in the Figure 4-10a. In AS-

mutant-Oct-CoA structure, OAL was bound in the tunnel of all monomers except in the C-monomer where both CoA and acyl group were missing, and that position was occupied by a glycerol molecule. This may be due to the fact that the AS-mutant-Oct CoA complex is obtained by quick soaking (40sec) and this duration may have been too short for the ligand to bind to all four monomers. Tunnel residues involved interacting with the OAL acyl fragment are depicted (Figure 4-10b). Interestingly, both the HAL and OAL moieties interact with the T143, G146 and R148 tunnel residues which are part of the covering loop. 20 amino acids are involved in tunnel formation (Figure 4-9b), but only the covering loop forms the backbone of the tunnel and plays a key role in acyl moiety binding.

4.9 Significance of the covering loop

As discussed above, the covering loop residues are involved in tunnel formation and also interact with the acyl tail of the longer CoA derivatives in PcaF. To understand the significance of the covering loop across degradative and biosynthetic thiolases, covering loop from three structures; the Zr-thiolase in complex with acetoacetyl CoA (PDB code: 1M1O) (Kursula et al. 2002), the degradative thiolases *Mycobacterium tuberculosis* (Mtb-thiolase) in complex with steroid and CoA (PBD code: 4UBT) (Schaefer et al. 2015) and PcaF A-mutant in complex with hex and CoA were compared (Figure 4-11) The three structures superimpose with a root mean square deviation (RMSD) values ranging from 1.36 to 1.42 Å for 366 - 370 Cα pairs and excluding the covering loop, RMSD values range between 1.21 and 1.32 Å for 356 - 366 Cα pairs. The decrease in the RMSD indicates a large variation of the covering loop in these thiolases. The covering loop (residues 143 – 163) in PcaF is elongated (compared to the others) and has a narrower base. The loop comprises of both hydrophobic and hydrophilic residues and many of them are bulky amino acids Figure 4-11a). Comparison shows that the covering loop adopts a conformation tailored according to the ligand. For instance, in PcaF the natural substrate is β-Ketoadipyl CoA which has a linear acyl chain with two oxygen atoms at their tail. Thus, a covering loop that is long, narrow and amphiphilic, seems to be a perfect fit for the natural substrate of PcaF. On the other hand, the Mtb-thiolase is a steroid binding thiolase with the covering loop ranging from 128–148 amino acids. It is also amphiphilic as in PcaF but adopts a wider skeleton so that it can bind the bulky steroid substrate (Figure 4-11c). On the contrary, Zr-thiolase has a small covering loop of eight amino acid residues (148 – 156) in length, which indicates the ability of the biosynthetic thiolases to bind smaller substrates Figure

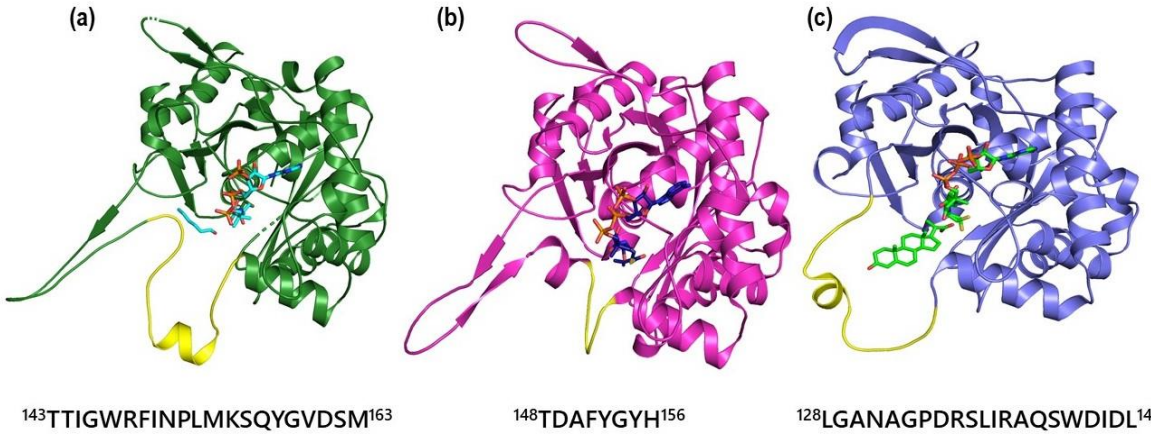

Figure 4-11 Comparison of the covering loop from representative members of the thiolase superfamily, shown in the same orientation. (a) A-mutant of PcaF complexed with non-natural substrate hexanal and CoA (b) Zr-thiolase complexed with acetoacetyl CoA (PDB code: 1M1O). (c) Mtb-thiolase complexed with steroid-CoA (PBD code: 4UBT). For each structure, the covering loop is colored in yellow and amino acid sequence with numbering of the loop are shown.

4-11b). The analysis suggests that the size and nature of the covering loop plays a significant role by providing the ligand; the right space, volume and surroundings to bind and determining its substrate specificity in thiolases.

4.10 Discussion

The comparison of the degradative thiolase PcaF with the Zr-thiolase highlights the striking structure-based differences that assist in distinguishing these two classes of thiolases. The catalytic triad of the degradative thiolase and biosynthetic thiolase are identical and structural adopt a similar orientation. Whereas, the active site architecture residues that line the wall of the cavity and the tunnel length are starkly different. In the degradative thiolases, the active site architecture residues are less conserved as they bind to a wider variety of substrates. In comparison the biosynthetic thiolases which have a restricted binding capacity of four carbon chain acyl CoA substrate, with limited substrate scope. Thus, the biosynthetic thiolases comprise of highly conserved active site architecture residues. This indicates that even though the catalytic residues are alike, there is a global difference in the amino acid residues of the active site architecture, and this is mostly likely the driving forces of thiolase reaction in the biosynthetic or degradative directions. CoA and acyl CoAs are lengthy substrates and need

tethering points to hold them in the correct configuration. The biosynthetic and degradative thiolases employ different anchoring residues to bind their substrates. In the Zr-thiolases, H157 acts as one of the anchoring residues and is conserved across the biosynthetic thiolases, but in the case of degradative thiolases, the residue at this position varies, implying that this position does not participate in substrate selection. In PcaF, H356 plays a central role in anchoring of the substrate, mutation of this residue leads to covalent locking of the reaction center. H356 also helps in retaining the active site architecture by maintaining the M163 and the R65 in correct conformation, these residues in turn are involved in anchoring the CoA in the optimal orientation. H356A mutation results in alternate conformation of Met 163, which causes the binding of CoA in an unusual orientation, exposing the thiol bond to the active C90 instead of Cα and Cβ of 3-ketoacyl CoA (consequently, covalent locking the active site). Additionally, the alternate conformation of M163 provide ample space for R65 to adopt double conformation. However, a similar histidine mutation in Zr-thiolase (PDB code: 2WL4) exhibits no change in the side chain conformation of M157 and Q64 (equivalent to M163 and R65) residues. Thus, H356 of the catalytic triad has an additional role in anchoring the CoA to the active site and preserving the functionality of the thiolase. This further emphasizes the difference of the active site architectures and the role of this conserved histidine residue of the catalytic triad in the degradative and biosynthetic thiolases.

The question is how the covalent locking of CoA is occurring in the A- and AA-mutant. One of possibilities we have investigated that if the thioester bond of the ligand is broken due to X-ray radiation damage, causing the sulfur atom of the ligand to attacks the Sγ of C90 residue to form a covalent bond. However, this possibility may be less likely as X-ray radiation damage analysis indicates breakage of the co-valent bonds between C90 and CoA, which means that the formation of covalent bond already took place during the soaking experiment prior to cryo-cooling of the protein crystals. This statement is further strengthened by the intact mass spectrometry experiment which were carried at pH 7.5. The covalent modification of the A- or AA-mutants was observed in presence of the CoA derivatives (acetoacetyl CoA, hexanoyl CoA, octanoyl CoA).

The second possibility is that the C90 residue in A- and AA-mutant is deprotonated and acts as a nucleophile attacking the thioester bond of the ligand and forming the covalent bond with CoA part of the ligand. If latter is the case, then the question is how C90 act as nucleophile in the pH range 7.5 to 8 pH (in which the intact mass spectrometry experiment and crystallization

were carried out) when the pKa of the side chain of cysteine is at pH ~8.3. It can be assumed that the pKa of the C90 may be lower than 8.3. It should be emphasized that, even if the C90 was active, the covalent locking would not have taken place if the H356 residue were present as it would had caused a steric hindrance to the C2P-S1P bond of CoA. This again emphasize a role of H356 in allowing the ligand to bind in correct orientation, thus acting as "tethering point".

Another striking observation is the presence of a substantially longer tunnel in the degradative thiolase PcaF compared to the Zr-thiolase. This longer tunnel allows PcaF to bind lengthier acyl chain CoA compounds, demonstrated by the ligand complexes of Hex-CoA/Oct-CoA. Both the HAL and OAL moieties interacts with the covering loop. Further superimposition of the degradative and biosynthetic tunnel indicates the role of the covering loop in determining tunnel length. The covering loop in biosynthetic thiolases is smaller which is found to restrict the length of the tunnel, unlike in the degradative thiolase PcaF where it is narrow and long (Figure 4-12). Comparing the degradative thiolases PcaF, Mtb-thiolase and the biosynthetic thiolase Zr-thiolase has aided in understanding the importance of the covering loop in thiolases. In PcaF, the covering loop is lengthened and amphiphilic in nature which is pinched at the base causing

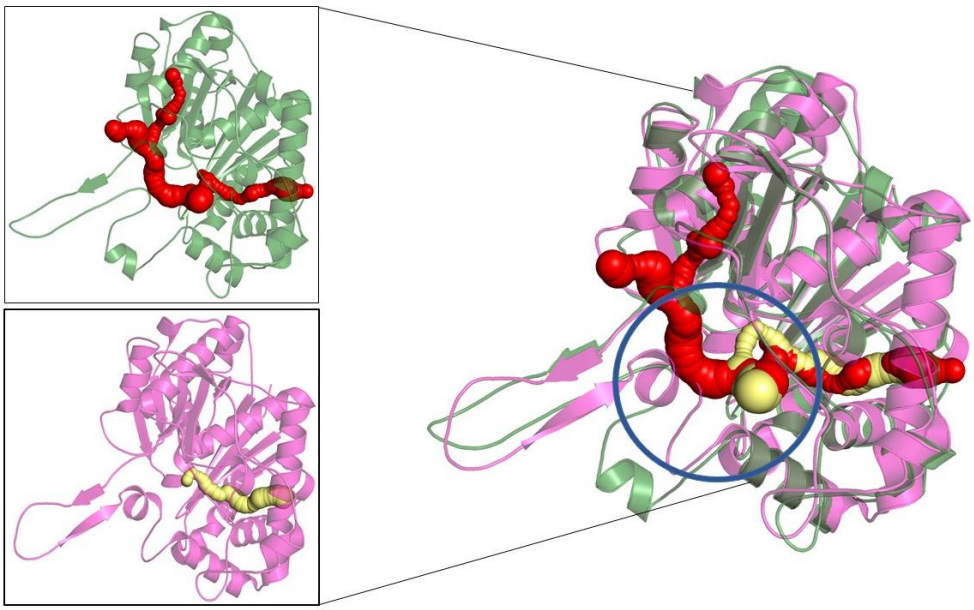

Figure 4-12 Tunnel comparison in the Zr-thiolase (PDB code: 1DLU) and the degradative thiolase PcaF structures. The structure of apo PcaF and its long tunnel are shown in green and red respectively. The tunnel of the biosynthetic thiolase is shown in yellow. Superimposition of the both structures shows that the biosynthetic covering loop block the degradative tunnel and makes it shorter indicating importance of the covering loop in tunnel. formation.

the tunnel to be restrained breadthwise. Whereas, in Mtb-thiolase to accommodate steroid groups which also requires a wide comparable circular loop of mixed amino acids, the biosynthetic thiolase has a smaller wider sized loop. Comparison clearly shows that the covering loop length and its amino acid type determine the substrate size and nature. The covering loop is the distinctive feature which plays pivotal role in determining the degradative and biosynthetic classes.

4.11 Conclusion

In this study, we have demonstrated the importance of the residues lining the active site in distinguishing the two class of thiolases. It will be fascinating to further probe these residues and investigate their role in reaction mechanism. We have also shown the additional function of H356 residue in maintaining the architecture and have provide understanding of the difference in CoA binding between the two classes. Furthermore, we have provided structural insight into substrate adaptability and binding of longer chain substrate in degradative thiolases. This knowledge can be used to modify the tunnel to allow binding of specific type of substrates. Overall, our study has presented valuable structural information that can be applied to classify uncharacterized thiolases into degradative and biosynthetic type.

4.12 References

Abril, M-A, Michan, C, Timmis, KN, & Ramos, JL 1989, 'Regulator and enzyme specificities of the TOL plasmid-encoded upper pathway for degradation of aromatic hydrocarbons and expansion of the substrate range of the pathway' *Journal of Bacteriology*, vol. 171, no. 12, pp. 6782–6790.

Anbazhagan, P, Harijan, RK, Kiema, TR, Janardan, N, Murthy, MRN, Michels, PAM, Juffer, AH, & Wierenga, RK 2014, 'Phylogenetic relationships and classification of thiolases and thiolase-like proteins of *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*' *Tuberculosis*, vol. 94, pp. 405–412.

Byska, J, Muzic, M Le, Gröller, EM, Viola, I, & Kozlíková, B 2016, 'AnimoAminoMiner : Exploration of Protein Tunnels and their Properties in Molecular Dynamics' *IEEE Trans Vis Comput Graph.*, no. 22(1), pp. 747–56.

Davis, JT, Chen, H-H, Moore, R, Nishitani, Y, Masamune, S, Sinskey, AJ, & Walsh, CT 1987, 'Biosynthetic thiolase from Zoogloea ramigera II. Inactivation with haloacetyl CoA analogs' *The Journal of Biological Chemistry*, vol. 262, no. 1, pp. 90–96.

Davis, JT, Moore, RN, Imperiali, B, Pratt, AJ, Kobayashi, K, Masamune, S, Sinskey, AJ, Walsh, CT, Fukui, T, & Tomita, K 1987, 'Biosynthetic thiolase from Zoogloea ramigera I. Preliminary characterization and analysis of proton transfer reaction' *The Journal of Biological Chemistry*, vol. 262, no. 1, pp. 82–89.

Feigenbaum, J & Schulz, H 1975, 'Thiolases of *Escherichia coli*: Purification and chain length specificities' *Journal of Bacteriology*, vol. 122, no. 2, pp. 407–411.

Haapalainen, AM, Meriläinen, G, & Wierenga, RK 2006, 'The thiolase superfamily: Condensing enzymes with diverse reaction specificities' *Trends in Biochemical Sciences*, vol. 31, no. 1, pp. 64–71.

Harijan, RK, Kiema, TR, Syed, SM, Qadir, I, Mazet, M, Bringaud, F, Michels, PAM, & Wierenga, RK 2017, 'Crystallographic substrate binding studies of *Leishmania mexicana*

SCP2-thiolase (type-2): Unique features of oxyanion hole-1' *Protein Engineering, Design and Selection*, vol. 30, no. 3, pp. 227–235.

Ithayaraja, M, Janardan, N, Wierenga, RK, Savithri, HS, & Murthy, MRN 2016, 'Crystal structure of a thiolase from Escherichia coli at 1.8 Å resolution' *Acta Crystallographica Section F*, vol. F72, pp. 534–544.

Jiang, C, Kim, SY, & Suh, D 2008, 'Molecular phylogenetics and evolution divergent evolution of the thiolase superfamily and chalcone synthase family' *Molecular Phylogenetics and Evolution*, vol. 49, pp. 691–701.

Jiménez, J I, Miñambres, B, García, JL, & Díaz, E 2002, 'Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440' *Environ Microbiol*, vol. 4, no. 12, pp. 824–841.

Jiménez, José Ignacio, Miñambres, B, García, JL, & Díaz, E 2002, 'Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440' *Environmental Microbiology*, vol. 4, no. 12, pp. 824–841.

Kiema, T-R, Harijan, RK, Strozyk, M, Fukao, T, Alexson, SEH, & Wierenga, RK 2014, 'The crystal structure of human mitochondrial 3-ketoacyl-CoA thiolase (T1): Insight into the reaction mechanism of its thiolase and thioesterase activities' *Acta Crystallographica Section D*, vol. D70, no. D70, pp. 3212–3225.

Kim, YH, Cho, K, Yun, SH, Kim, JY, Kwon, KH, Yoo, JS, & Kim, SI 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*, vol. 6, pp. 1301–1318.

Krissinel, E & Henrick, K 2007, 'Inference of macromolecular assemblies from crystalline state' *Journal of Molecular Biology*, vol. 372, no. 3, pp. 774–797.

Kursula, P, Ojala, J, Lambeir, A-MM, & Wierenga, RK 2002, 'The catalytic cycle of biosynthetic thiolase: A conformational journey of an acetyl group through four binding

modes and two oxyanion holes' *Biochemistry*, vol. 41, no. 52, pp. 15543–15556.

Kursula, P, Sikkilä, H, Fukao, T, Kondo, N, & Wierenga, RK 2005, 'High resolution crystal structures of human cytosolic thiolase (CT): A comparison of the active sites of human CT, bacterial thiolase, and bacterial KAS I' *Journal of Molecular Biology*, vol. 347, no. 1, pp. 189–201.

Mathieu, M, Zeelen, Jp, Pauptit, R, Erdmann, R, Kunau, WH, & Wierenga, R 1994, 'The 2.8 Å crystal structure of peroxisomal 3-ketoacyl-CoA thiolase of Saccharomyces cerevisiae : A five-layered αβαβα structure constructed from two core domains of identical topology' *Structure*, vol. 2, no. 9, pp. 797–808.

Mathieu, Magali, Modis, Y, Zeelen, JP, Engel, CK, Abagyan, RA, Ahlberg, A, Rasmussen, B, Lamzin, VS, Kunau, WH, & Wierenga, RK 1997, 'The 1.8 A crystal structure of the dimeric peroxisomal 3-ketoacyl-CoA thiolase of Saccharomyces cerevisiae: implications for substrate binding and reaction mechanism' *Journal of Molecular Biology*, vol. 273, no. 3, pp. 714–28.

Meriläinen, G, Poikela, V, Kursula, P, & Wierenga, RK 2009, 'The thiolase reaction mechanism: The importance of Asn316 and His348 for stabilizing the enolate intermediate of the Claisen condensation' *Biochemistry*, vol. 48, no. 46, pp. 11011–11025.

Modis, Y & Wierenga, RK 2000, 'Crystallographic analysis of the reaction pathway of *Zoogloea ramigera* biosynthetic thiolase' *Journal of Molecular Biology*, vol. 297, no. 5, pp. 1171–1182.

Nelson, KE, Weinel, C, Paulsen, IT, Dodson, RJ, Hilbert, H, Martins dos Santos, VAPP, Fouts, DE, Gill, SR, Pop, M, Holmes, M, Brinkac, L, Beanan, M, DeBoy, RT, Daugherty, S, Kolonay, J, Madupu, R, Nelson, W, White, O, Peterson, J, Khouri, H, Hance, I, Chris Lee, P, Holtzapple, E, Scanlan, D, Tran, K, Moazzez, A, Utterback, T, Rizzo, M, Lee, K, Kosack, D, Moestl, D, Wedler, H, Lauber, J, Stjepandic, D, Hoheisel, J, Straetz, M, Heim, S, Kiewitz, C, Eisen, J, Timmis, KN, Düsterhöft, A, Tümmler, B, & Fraser, CM 2002, 'Complete genome sequence and comparative analysis of the

metabolically versatile Pseudomonas putida KT2440' *Environmental Microbiology*, vol. 4, no. 12, pp. 799–808.

Ornston, LN 1966a, 'The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida* II. Enzymes of the protocatechuate pathway' *The Journal of Biological Chemistry*, vol. 241, no. 16, pp. 3800–3810.

Ornston, LN 1966b, 'The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida* III. Enzymes of the catechol pathway' *The Journal of Biological Chemistry*, vol. 241, no. 16, pp. 3795–3799.

Ornston, LN 1966c, 'The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida* IV. Regulation' *The Journal of Biological Chemistry*, vol. 241, no. 16, pp. 3800–3810.

Pereto, J, Lo´peZ-Garcı´a, P, & Moreira, D 2005, 'Phylogenetic analysis of eukaryotic thiolases suggests multiple proteobacterial origins' *Journal of Molecular Evolution*, vol. 61, pp. 65–74.

Pye, VE, Christensen, CE, Dyer, JH, Arent, S, & Henriksen, A 2010, 'Peroxisomal plant 3-ketoacyl-CoA thiolase structure and activity are regulated by a sensitive redox switch' *The Journal of Biological Chemistry*, vol. 285, no. 31, pp. 24078–24088.

Schaefer, CM, Lu, R, Nesbitt, NM, Schiebel, J, Sampson, NS, & Kisker, C 2015, 'FadA5 a thiolase from Mycobacterium tuberculosis: A steroid-binding pocket reveals the potential for drug development against tuberculosis' *Structure*, vol. 23, no. 1, pp. 21–33.

Tian, W, Chen, C, Lei, X, Zhao, J, & Liang, J 2018, 'CASTp 3.0: computed atlas of surface topography of proteins' *Nucleic Acids Research*, vol. 46, no. W1, pp. W363–W367.

Yamanashi, T, Kim, S-Y, Hara, H, & Funa, N 2015, 'In vitro reconstitution of the catabolic reactions catalyzed by PcaHG, PcaB, and PcaL: the protocatechuate branch of the β-ketoadipate pathway in Rhodococcus jostii RHA1.' *Bioscience, Biotechnology, and Biochemistry*, vol. 79, no. 5, pp. 830–835

Chapter 5
Unravelling the PcaC: identification of the active site, product pathway and mechanism of γ-CMD family

5.1 Introduction

Decarboxylase is one of the central class of enzymes in the biological pathways consisting of more than 150 subclasses that are currently identified (Broadway et al. 2004; Moss 2019). Decarboxylase catalyses a spectrum of anabolic and catabolic pathways such as amino acid decarboxylation to carbohydrate synthesis, respectively (Schörken & Sprenger 1998). The decarboxylase family can be further categorized based on their catalytic strategy into two groups; the first that uses cofactor for decarboxylation, and the second that does not require any exogenous factor for its enzymatic activity (Walsh 2019). Only a handful of decarboxylases belong to the second group and carboxymuconolactone decarboxylase (CMD) family is one among them. Interestingly, CMD family allows the survival of bacteria in a variety of critical circumstances like oxidation stress and xenobiotic environment (Clarke et al. 2011). CMD family has three distinctive members; first, the alkyl-hydroperoxidase (AhpD) that participates in bacterial antioxidant defence (Nunn et al. 2002), second, the γ-carboxymuconolactone decarboxylase (γ-CMD) which assist bacterial growth in xenobiotic habitat (Eulberg, Lakner, Golovleva, & Schlomann 1998), and third, an member whose function is unknown (Ito et al. 2006).

AhpD is the most studied member of the CMD family and a number of crystal structures from the subfamily have been reported (PDB codes: 5DIK, 1P8C, 1VKE, 3BEY, 2O4D, 3LVY, 2OYO, 2PRR, 1KNC, 1ME5, 1LW1 AND 1GU9) (Chen et al. 2015). AhpD is a part of the AhpC/AhpD-coupled system which demonstrates thioredoxin-like peroxidase activity (Hillas et al. 2000). It comprises of a conserved structural CXXC motif which is critical for the detoxification of reactive oxygen species and peroxides (Koshkin et al. 2004; Schuitz, Chivers,

& Raines 1999). The AhpD from *M. tuberculosis* (Mt-AhpD) is often taken as the representative member to study enzyme of the group as its active site and mechanism have been described (Hillas et al. 2000; Koshkin, Nunn, Djordjevic, & Montellano 2003) however, γ-CMD lacks the CXXC motif and contain relatively low sequence identity against AhpD.

γ-CMD (EC 4.1.1.44) is part of β-ketoadipate pathway that is involved in some eukaryotes such as fungi and yeast and in bacterial degradation of aromatic compounds (Iwagami, Yang, & Davies 2000; Jiménez, Miñambres, García, & Díaz 2002). β-ketoadipate pathway comprises of the catechol branch and protocatechuate branch which converge in latter stages (Ornston 1966; Yamanashi, Kim, Hara, & Funa 2015). From the two branches of the pathway, Muconolactone isomerase and γ-CMD mediate chemically analogous reactions giving rise to an identical product β-ketoadipate enol lactone (Yeh, Durham, Fletcher, & Ornston 1981) which then subsequently gets converted into TCA cycle intermediates via 3-4 more steps that are governed by enzymes common to both the pathways (Figure 5-1) (Chari, Whitman, & Kozarich 1987).

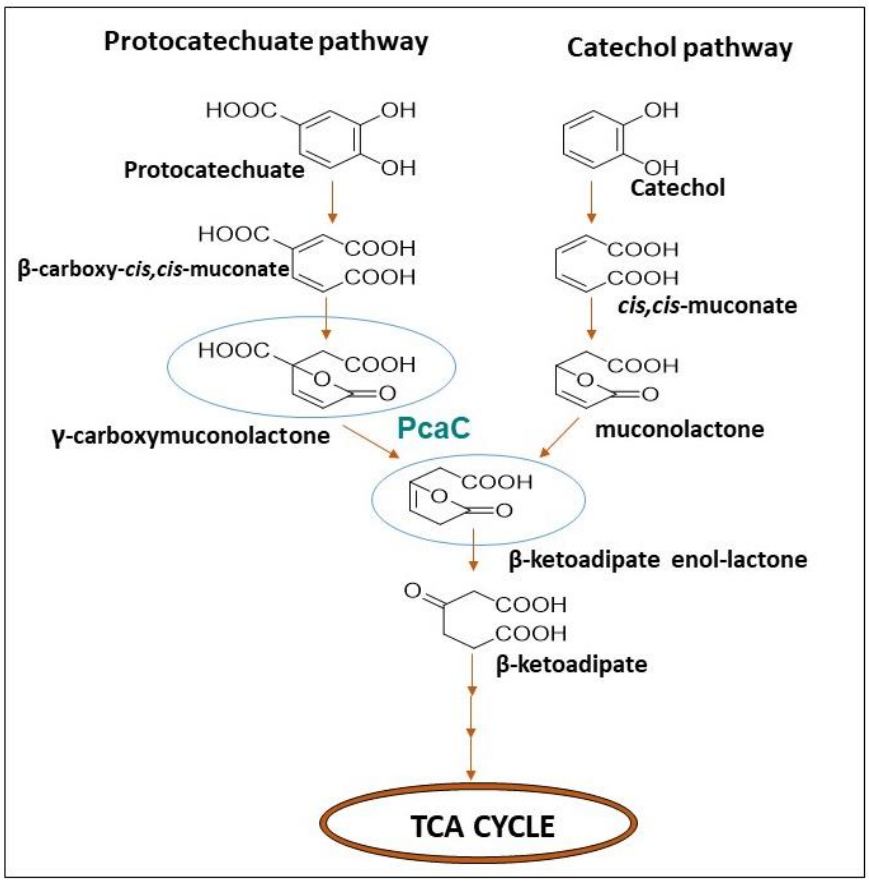

Figure 5-1 Schematic representation of the protocatechuate and catechol pathways which combine to form the β-ketoadipate pathway. The PcaC catalyze the first common product of β-ketoadipate pathway, arising out of the protocatechuate arm of the converged pathway.

Prior studies of the sequences of Muconolactone isomerase and γ-CMD shows homology in the N-terminal sequences, indicating a common ancestral origin. In some bacteria a gene fusion event leads to expression of γ-CMD with a hydrolase involved in the same (Eulberg et al. 1998; Shahbaaz, Ahmad, & Hassan 2015). Crystal structures for the γ-CMD family is available in the Protein Data Bank that contains unpublished structures (PDB codes: 4G9Q, 2AF7, 2Q0T) and also the include the latest crystallization and preliminary X-ray crystallographic analysis of γ-CMD from *Sulfolobus solfataricus* (Lee & Yang 2009). Throughout the years various trials and speculations about the possible reaction mechanism is made (Yeh, Fletcher, & Ornston 1980), however its active site and functional mechanism still remain unclear.

In this study, PcaC from the β-ketoadipate pathway in *Pseudomonas putida* KT2440 is explored. PcaC catalyzes the conversion of γ-carboxymuconolactone to β-ketoadipate enol lactone and releases CO_2 as by-product (Figure 5-1) (Parke, Meagher, & Ornston 1973). PcaC is branded as γ-CMD because of its sequence identity (30% ~ 50%) with the γ-CMDs of other species and the absence of the CXXC motif. To characterize the active sites that are involved in mechanism of γ-CMD, the crystal structure of PcaC is determined. The active site is identified using a combination of the evolutionary and bioinformatic studies and structural analysis of PcaC complexed with sodium bicarbonate. Furthermore, mutation of the proposed active site residues in conjunction with enzymatic assay confirmed the active site. The crystal structures of PcaC variant in complex with substrate-analog and product-analog together with modelling of its substrate helped in describing the enzymatic reaction. For the first time, we uncover carbon dioxide path in a decarboxylation reaction from its reaction center to way out**.** The work provides an in-depth study of the γ-CMD family with a holistic view of the CMD family

5.2 Structure determination and quality of the structures

The PcaC protein has been crystallised in with three different morphologies and yielded data in three different space groups. The hexagonal-shaped crystals of native and selenomethionine proteins that grew in sodium format belonged to space group $P6_3$ with two molecules per asymmetric unit and a calculated solvent content of 47.67%. Crystals of PcaC-H80A and PcaC-E76A mutant protein grown in the magnesium chloride condition were rod-shaped and belonged to space group $P2_12_12_1$ and C2 space group respectively. The orthorhombic crystal

form had six molecules per asymmetric unit with a calculated solvent content of 49.21 % and monoclinic crystal form contains 12 molecules per asymmetric unit with a calculated solvent content 54.18%.

The structures of the native PcaC (apo), binary complex of PcaC-E78A with 2- Furan acetic acid (2FAA), binary complex of PcaC-E78A with homocitrate acid lactone (HCAL), binary complex of PcaC-H80A with 2FAA, ternary complex of PcaC-H80A with 2-(5-oxotetrahydrofuran-2-yl) acetic acid and CO_2 and quaternary complex of native PcaC with CO_3^{2-} and CO_2 were determined. The refined model of the native PcaC includes residues from Tyr-7 to Arg-129 and Lys-4 to Arg-129 in both subunits. This model lacks only one residue from the C-terminal due to insufficient interpretable electron density. Acetate was clearly visible in the initial difference Fourier maps in both subunits.

The structure of native PcaC in complex with $NaHCO_3$ was solved at 2.12 Å resolution and the refined model the complex contains residues from Gln-5 to Arg-129 and Lys-4 to Arg-129 in both subunits. Difference Fourier map clearly shows the binding of CO_3^{2-} in place of acetate molecule and binding of CO_2 molecule near Arg-14 and binding of another CO_2 near the main chain of Asp-75 in both subunits. Binding of CO_3^{2-} and CO_2 near the substrate binding site required only minor local structural adjustment except the change in side-chain orientation of Ile-41 relative to native structure.

The structure of the PcaC_E76A-HCAL and PcaC_E76A-2FAA complex was determined at 2.2 Å and 2.58 Å resolution respectively. All 12 subunits in each of the two complexes include residues from Met-1 to Gln-130. An HCAL and 2FAA moiety were clearly visible in the initial difference Fourier maps in each subunit of each complex. Binding of 2FAA or HCAL near the substrate binding site minor local structural adjustment and change in side-chain orientation of Ile-41 similar to the quaternary complex of native PcaC with CO_3^{2-} and CO_2 was observed.

The structure of the PcaC_H80-HCAL and PcaC_H80A-2FAA complex was determined at 2.12 Å and 2.01 Å resolution respectively. All 6 subunits in each of the two complexes also include residues from Met-1 to Gln-130. Difference Fourier map distinctly showed the presence of 2FAA, in each subunit of the PcaC_H80A-2FAA complex. However, difference Fourier map did not show the presence of HCAL in any of the subunit rather presence of 2-(5-oxotetrahydrofuran-2-yl) acetic acid (decarboxylated product of HCAL) in one subunit and CO_2

bound in all six subunits. In the PcaC_H80A-2FAA complex and PcaC_H80-HCAL, change in side-chain orientation of Ile-41 relative to native structure was observed which is similar to PcaC_E76A-HCAL or PcaC_E76A-2FAA complex.

5.3 Overall structure

PcaC exists as a hexamer made of up three asymmetric units in the native protein, whereas single asymmetric unit of PcaC_H80A and PcaC_E76A mutant crystals contain one and two hexamer(s) respectively. The two monomers in an asymmetric unit of the native protein crystal are related by each other by a non-crystallographic 2-fold axis. The overall structure of PcaC monomer is an all alpha helical and contains seven α-helices (α1 – α7), N-terminal helical domain (α1 – α3) followed by C-terminal helical domain (α4 – α7). Structure homology search by the DALI server (Holm, Kääriäinen, Rosenström, & Schenkel 2008) revealed that PcaC has the highest similarity to CMD from *Methanobacterium thermoautotrophicum* (a Northeast

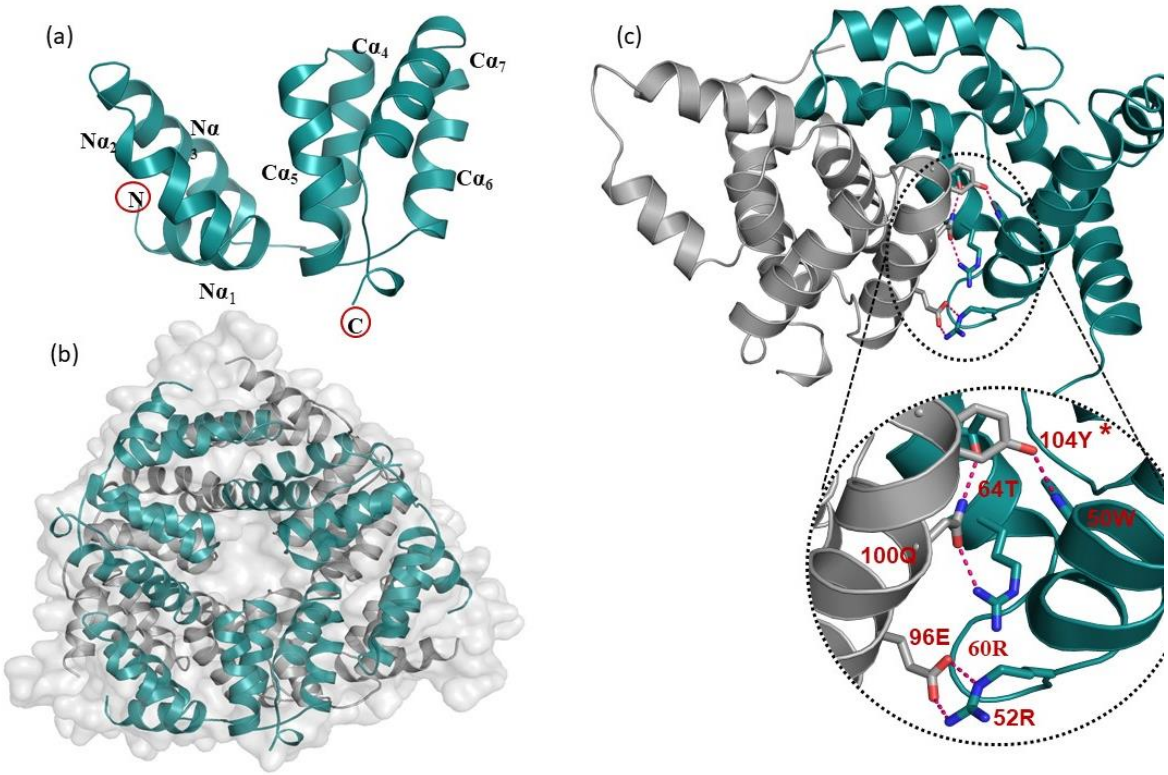

Figure 5-2 (a) Cartoon representation of PcaC structure. The N and C termini and each of the six alpha helices are labelled. (b) the hexamer organization of PcaC (c) Apo PcaC protein is present as dimeric subunit in the crystallographic asymmetric unit. The residues interacting in the dimeric interface is shown as ball and stick model.

Structural Genomics Consortium target TT747, PDB code 2AF7; unpublished), with a Z score of 16.6 and a RMSD of 1.7 Å for 117 common Ca atoms. Along with the structure similarity, the quaternary structure of PcaC protein shows similar arrangement of dimeric and hexameric structure. Despite the high structural similarity in the N-terminal domain, there are some noticeable differences in the orientation of the C-terminal part. Structure based sequence alignment indicates that PcaC contains only 32% sequence identity to the target TT747. The two molecules of PcaC in an asymmetric unit are related by two-fold symmetric dimer with an extensive buried surface area (BSA) of 1322 $Å^2$. The dimer involves all helices (α6 from one monomer and α4 and α5 from another monomer) from C-terminal domain except α7. At the dimer interface Trp-50 and Thr-64 from one monomer forms h-bonds with Tyr-104 and 100Q of the other molecule respectively, whereas Arg-52 and Arg-60 from one monomer forms salt bridge with Glu-96 and Gln-100 respectively.

The structure of the two chains is essentially similar with a root- mean-square deviation (RMSD) of only 0.36 Å for the main chain atoms. The dimer appears to trimerize by the crystallographic 3-fold symmetry axis to form a homohexamer ring. The hexameric ring does not contain tunnel passing through middle of the ring like in the MtAhpD (Nunn et al. 2002) and lpg0406 structures (Chen et al. 2015).

5.4 Identification of the active-site pocket

The CMD family includes mainly two types of CMDs one with peroxidase activity and other without. Those CMDs which lacks peroxidase activity, does not contain the CXXC motif. Most of such CMDs are part of γ-CMD subfamily. PcaC belongs to γ-CMD, because of its high sequence identity (30% ~ 50%) with the γ-CMDs of other species. In order to identify active site of PcaC, a sequence similarity network of CMD family (https://pfam.xfam.org/family/CMD) was created and the output was loaded into DALI server (Holm et al. 2008) to visualize the network as evolutionarily distinct groups (Figure 5-3). In

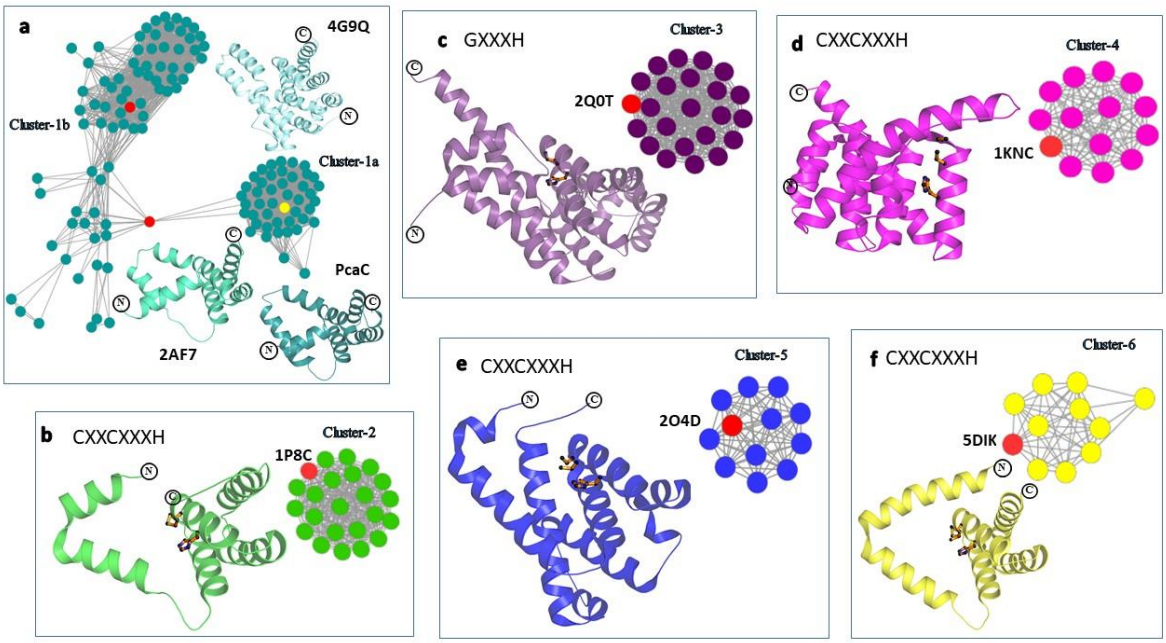

Figure 5-3 Cytoscape representation of the sequence similarity network of CMD family at a cut-off of e^{-40}. In each group, the nodes represent the proteins and the edges represent the BLASTP linkages. The representative crystal structures are coloured in red in each group.

each group, the nodes represent the proteins and the edges represent the BLASTP motif was marked as ball-and-stick (Figure 5-3). The representative crystal structure from each group shows significant variation in their C-terminal domain with respect to N-terminal domain. Three helices from the N-terminal domain (the motif containing helix, helices before and after the motif containing helix) of the representative crystal structure from each group (2 to 6) was superposed against the three helices from the N-terminal domain of PcaC in order to determine equivalent position of C/GXXCXXXH motif in the PcaC. The structurally equivalent position in PcaC is RNDELKLH, which also aligned in the protein sequences (Figure 5-4b). We then investigated electron density of the PcaC crystal structure. Surprisingly, triangular difference density was located near Glu-76 and His-80 in both monomers, which was modelled as acetate, as PcaC was crystallised in the presence of sodium acetate. The acetated molecule is located at the hydrogen bonding distance from Glu-76 and His-80 in both chains of the structure. Interestingly Glu-76 and His-80 are highly conserved within γ-CMD subfamily. Along with these two residues, Trp-46, Trp-50, Arg-14, Tyr-104 and His-22 residues were also observed to be conserved (Figure 5-4a). Surface representation of the native structure indicates that these residues are involved in forming a cavity which is solvent accessible and part of the active site architecture. The acetate molecule was clearly visible in the cavity. This enforced our suspect

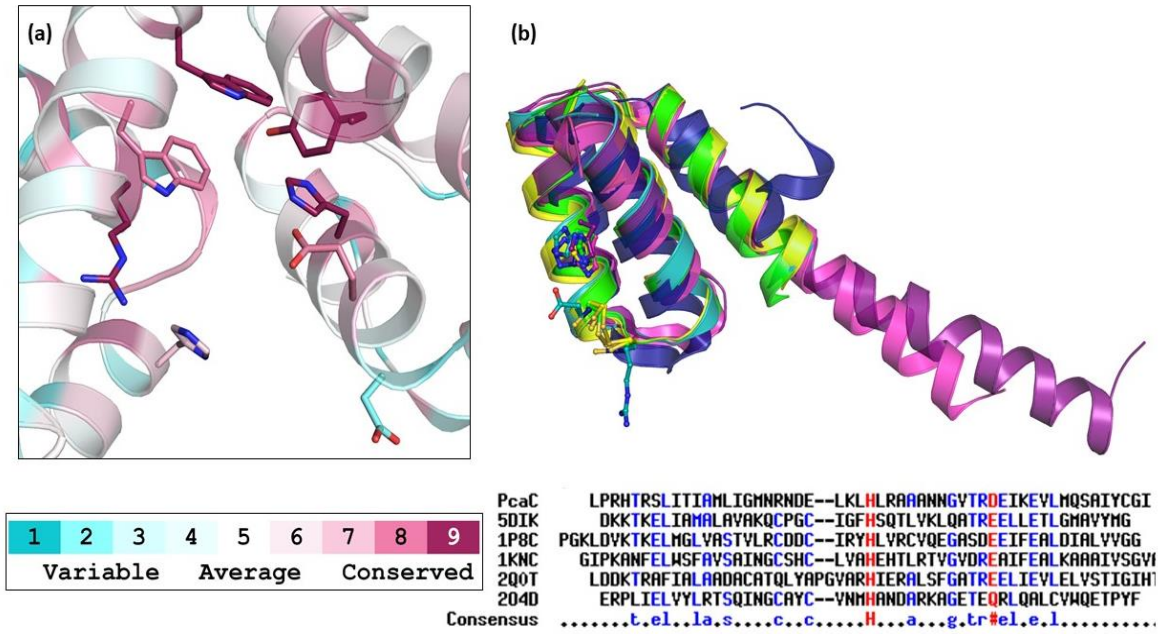

Figure 5-4 The conservation score of the individual active site residues of PcaC. It is represented by the colour gradient as shown in the conservation score strip. (b) Structural superimposition and sequence alignment of the representative structure from the 6 clusters to determine the equivalent residues at the position of C/GXXCXXXH motif. The following PDBs are taken for structural superimposition, PcaC(this study) from cluster 1 (in cyan), 1P8C from cluster-2 (in green), 2Q0T from cluster-3 (in purple) 1KNC from cluster-4 (in pink), 2O4D from cluster-5 (in blue), and 5DIK from cluster-6 (in yellow).

that these binding sites in a characteristic pocket in the center of the C-terminal domain likely mark the active-site pocket and the two residues Glu-76 and His-80 are the active site residues. The binding cavity is located at the bottom of the cleft of the N- and C-terminal helical domain.

5.5 Sodium bicarbonate complex affirms the active-site pocket

Since the putative active sites of the protein were already occupied by the acetate molecules in the native crystals, any co-crystallization or soaking attempt with product – analog did not succeed. Therefore, we carried out soaking experiment of native protein crystal with sodium bicarbonate in order to free the active site. A number of datasets were collected from the native crystals which were soaked in presence of 10 mM sodium bicarbonate between few secs to 15min. After few cycles of refinement, strong triangular positive difference electron density was found at the equivalent position of acetate in each subunit which could be unambiguously

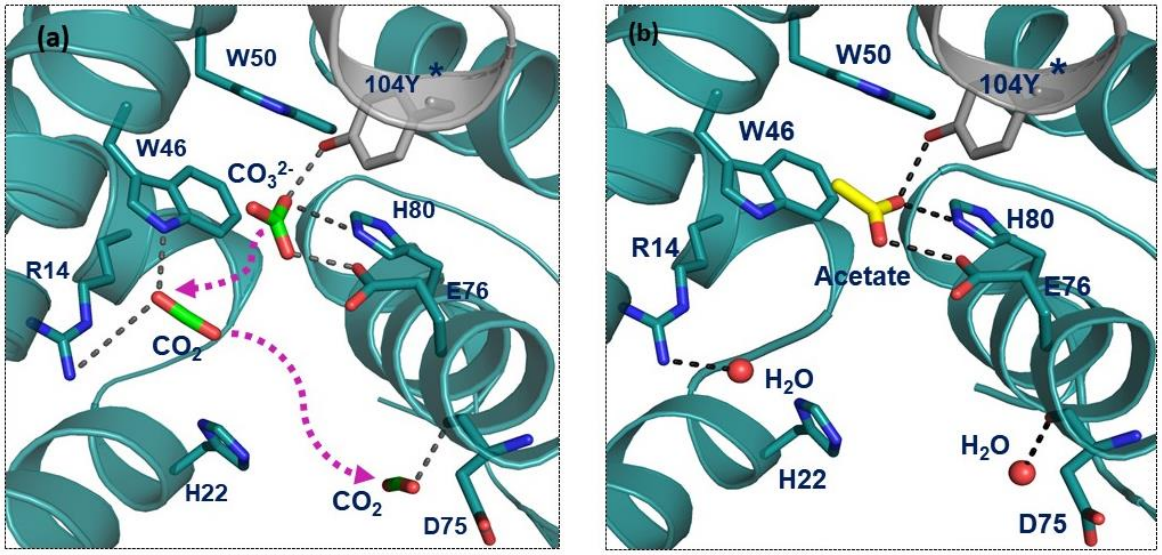

Figure 5-5 (a) the snapshot of active site of PcaC after bicarbonate soaking. The carbonate ion interacts with the active site residues, E76 and H80, validating the active sites. Carbon dioxide is formed from the car-bonate ion as a result of spontaneous dissociation and it leaves the active site, marking the product release pathway. Two carbon dioxide molecules are trapped in the crystal structures which are in-teracting with R14 and D75 residues respectively. (b) In the apo-PcaC, acetate occupy the equivalent position of carbonate ion and satisfies the interaction potential of the active site residues E76 and H80. Two water molecules occupy the corresponding positions of the two carbon dioxide molecules present in the bicarbonate soaking method by forming hydrogen bond with R14 and D75.

modelled as CO_3^{2-} In addition, elongated positive difference electron density was located in the vicinity of CO_3^{2-} in each subunit, which was modelled as CO_2.

The carbon dioxide forms hydrogen bond with NE1 atom of Trp-46 and with NH1 atom of Arg-14. Further few cycles of refinement, another elongated positive difference electron density in either subunit was observed near the carbonyl oxygen of Asp-75, which was also modelled as CO_2. The presence of the CO_2 molecule makes us believe that the triangular positive difference electron density was CO_3^{2-} not the acetate. The carboxylic group of carbonate ion forms hydrogen bonds with NE2 atom of histidine and hydroxyl group of the ion anchors with OE1 atom Glu-76 tight hydrogen bond (Figure 5-5 and Figure 5-6). The interaction is same as acetate molecule. In the native structure, the CO_2 positions are occupied by the water molecules.

In the analysis of datasets which were collected with the soaking time less than 30 sec, we have observed consistently one carbonate and one carbon dioxide whereas in longer soaking but less than 5min, we observed one carbonate and two carbon di-oxide as explained above. In longer

than 5 min soak, only two carbon di-oxide molecules were trapped, in which the one of the CO_2 molecules was found to at the equivalent position of the CO_3^{2-} and found to interact with Tyr-104 and His-80. These complexes reaffirm that the Glu-76 and His-80 are the part of the active sites and Trp-46, Arg-14 and Tyr-104 are the part of the binding pocket (Figure 5-5).

5.6 The CO_2 release pathway

The additional advantage of the bicarbonate soaking method was the identification of the pathway through which CO_2 diffuses out of the protein. The distance of the pathway was

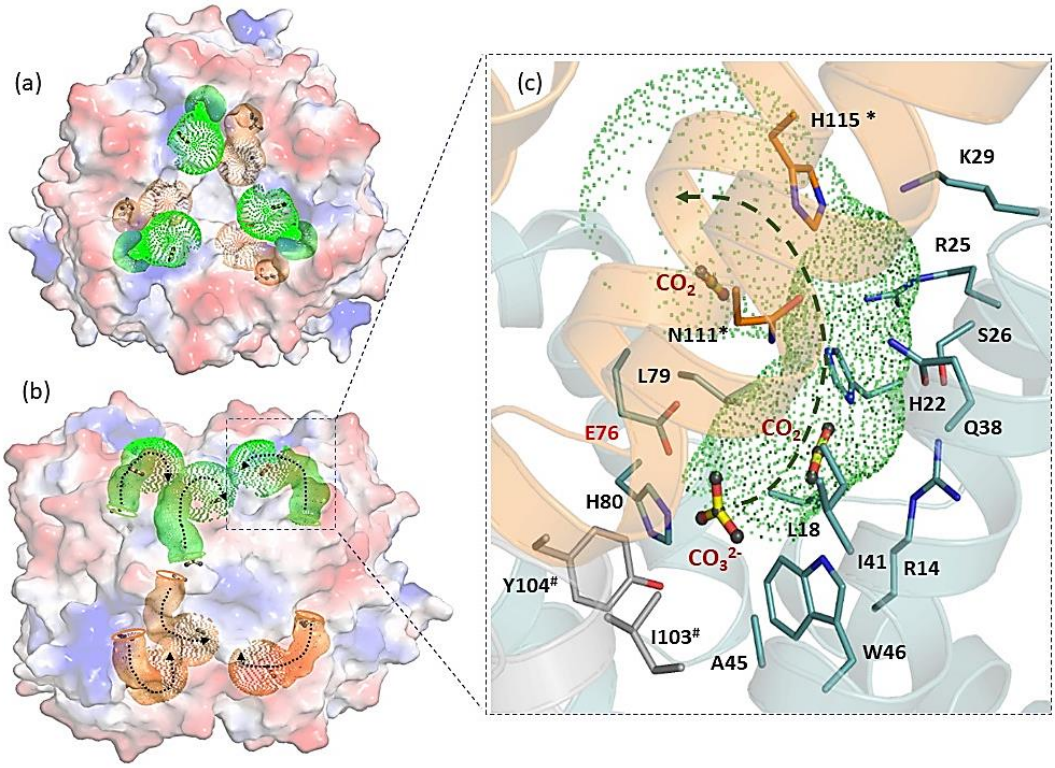

Figure 5-6 Representation of CO_2 exit tunnel in PcaC. The black dotted arrow indicate the direction of the diffusion of CO_2 from the active site: which is marked by CO_3^{2-}(a) Top view of the three CO_2 exit tunnel from the homohexamer ring (b) Side view of the six CO_2 exit tunnel, demonstrating the six active site of the hexameric PcaC. (c) Representation of the CO_2 exit tunnel present in B-chain, showcasing the residues involved in the pathway. Residues involved in forming the tunnel from B-chain, A-chain and F-chain are colored in deep teal, grey and orange respectively. The CO_3^{2-} and CO_2 molecules are marked by ball and stick model. The green dotted cylinder with globular head is the enlarged version of the CO_2 exit tunnel

calculated by HOLE (Smart et al. 1996) to be ~10.6 Å and the van der wall probe radius of the pathway is ~2.5 Å. The pathway starts from the coordinate of the carbon atom of the carbonate ion present in the active site till the protein surface.

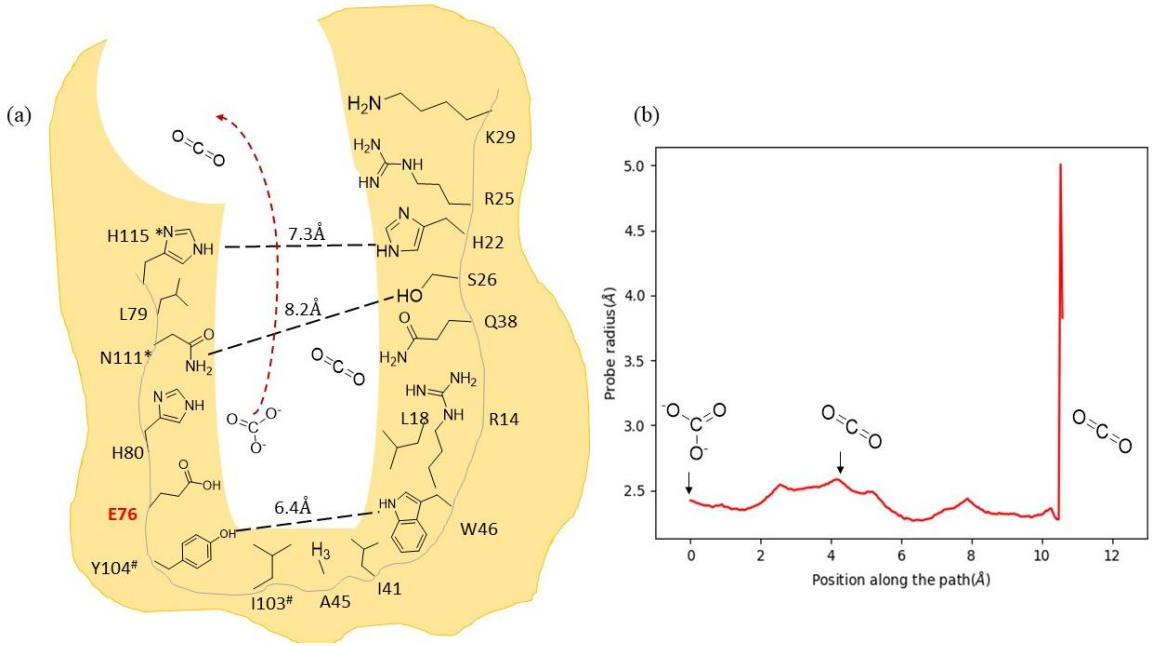

Figure 5-7(a) The 2D representation of the CO_2 exit tunnel. The tunnel shown here belong to the chain B. The H115 and N111 residues marked with "" are from the F subunit, whereas I103 and Y104 residues marked with "#" are from the A subunit. (b)The graph representing the path distance covered by the CO_2 versus Van der Wall radius.*

The pathway in one monomeric subunit is formed by residues comprising from three monomers. Most of the residues involved in pathway formation comes from the monomeric residues in which the pathway is present, except for Ile-103, Tyr-104 residues that comprises from one adjacent monomer and His-115, Asn-111 from the other neighbouring monomer (Figure 5-7). The residues forming the pathway are mostly polar and is shown in Figure5-6 and 5-7. The pathway predicted by HOLE (Smart et al. 1996) agrees with the exit path trace by the two CO_2 molecules in the ternary PcaC-bicarbonate crystal structure. The carbonate ion indicates the reaction centre and marks the start of the tunnel. The first CO_2 which~3.9Å away from bicarbonate form hydrogen bond with Trp-46W and represents the middle of the pathway. The second CO_2 that is ~10Å away from the first CO_3^{2-} is present on the outer surface of the protein shows the end of the pathway. It has been established that, as the reaction continues

CO_2 is formed from bicarbonate (by spontaneous dissociation) and CO_2 interact with residues which guides it to bind across different sites to exit the product pathway.

5.7 Substrate-analog/product analog complex

Since original substrate (γ-carboxymuconolactone) is not commercially available, we have used substrate-analog, homocitrate acid lactone (HCAL) (Nickerson et al. 2016) for the elucidation of the substrate binding mode and catalytic mechanism the inactive variant E76A was crystallized in the presence of 5mM HCAL. After few cycles of refinement, strong positive difference electron density was found in the putative active site into which the substrate could be unambiguously modelled (Figure 5-8a), whereas the lack of side-chain density confirmed

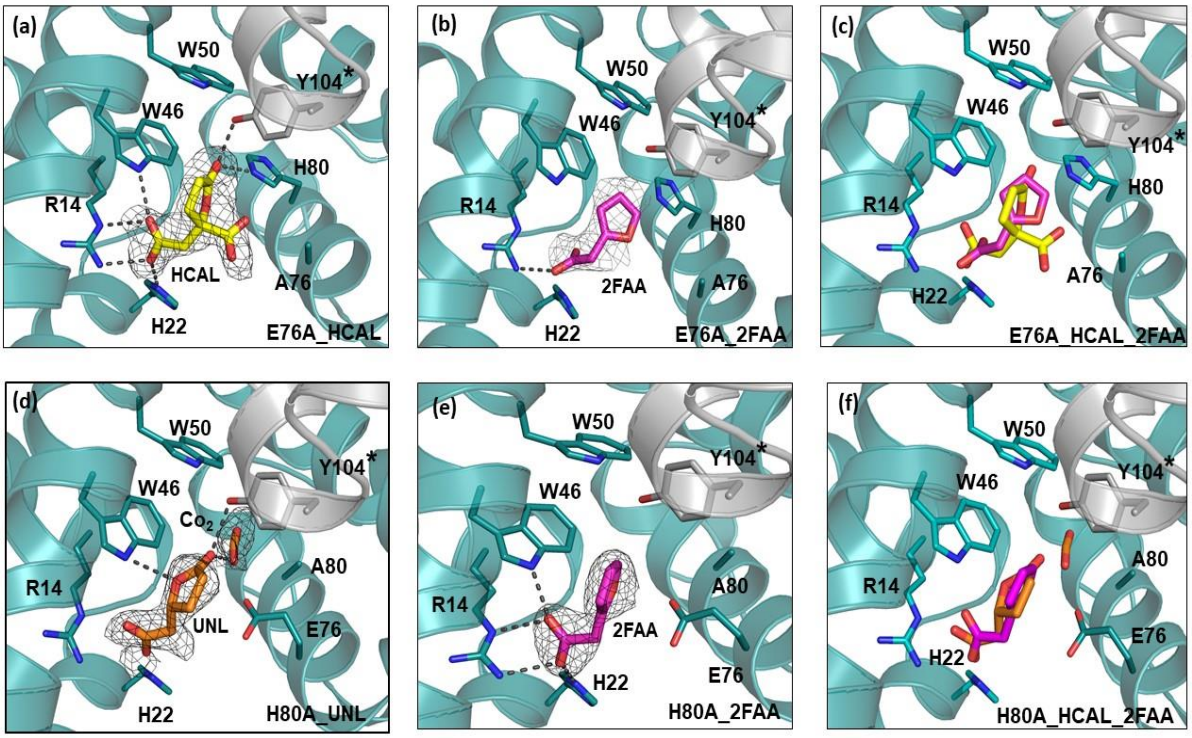

Figure 5-8 Active site representation of PcaC mutants (a)PcaC-E76A in complex with homocitrate acid lactone (HCAL) (b) PcaC-E76A in complex with 2-Furan acetic acid (2FAA).(c)The superimposition of the two PcaC-E76A mutant structures complexed with HCAL and 2FAA. Active site representation of PcaC mutants (d)PcaC-H80A in complex with 2-(5-oxotetrahydrofuran-2-yl) acetic acid (UNL)(e) PcaC-H80A in complex with2FAA. (f) The superimposition of the two PcaC-H80A mutant structures complexed with UNL and 2FAA Electron density ($F_o - F_c$) maps are contoured at 3σ for both ligands. Oxygen and nitrogen atoms are in red and blue, respectively.

the E76A mutation. The substrate is deeply buried in the hydrophobic pocket, which was occupied by an acetate molecule in the unliganded structure (Figure 5-8f). However, the leaving carboxylic group of the substrate does not bind at the position of the carboxylic group of acetate in the unliganded enzyme. Instead, the carbonyl group from the furan ring of the substrate binds at this position where it is tightly anchored through two hydrogen bonds to His-80 of the subunit and to Tyr-104 of a neighbouring subunit. Other carboxylic group of the substrate is involved in forming salt bridge with Arg14 and hydrogen bond with Trp-46 and with His-22. We noticed there is a shift in the binding of the substrate-analogue due to available space upon Glu-76 mutation to alanine.

Although H80A mutant was still active but reduced activity in comparison to the native protein, we have crystallized the variant and soaked the crystal in the presence of 5 mM HCAL. After few cycles of refinement, strong positive difference electron density was found in the putative active site of a monomer into which the product (2-(5-oxotetrahydrofuran-2-yl) acetic acid) could be unambiguously modelled (Figure 5-8d), whereas the lack of side-chain density confirmed the H80A mutation. In addition, elongated strong positive difference electron density was found in the place of His-80 side chain in each monomer into which the product CO_2 could be unambiguously modelled (Figure 5-8d). The carbonyl group of the product forms hydrogen bonds with the other product CO_2 and Tyr-104 of a neighbouring subunit. Oxygen atom from furan ring of the product is anchored through hydrogen bond Trp-46.

We have co-crystallized H80A mutant in presence of 5 mM of 2FAA. After refinement, strong positive difference density was found in the putative active site into which 2FAA could be also unambiguously modelled. The carboxylic group of the 2FAA involved in forming salt bridge with Arg14 and hydrogen bond with Trp-46 and with His-22 in a similar manner as the E76A – HCAL complex (Figure 5-8e).

5.8 Mutational analysis via mass spectrometry assay

The structures of substrate-analog/product analog complexed with mutant PcaC helped in identifying the active site residues and provided in depth understanding of the ligand binding into the active site. As no assay was previously reported for calculating the activity of the γ-

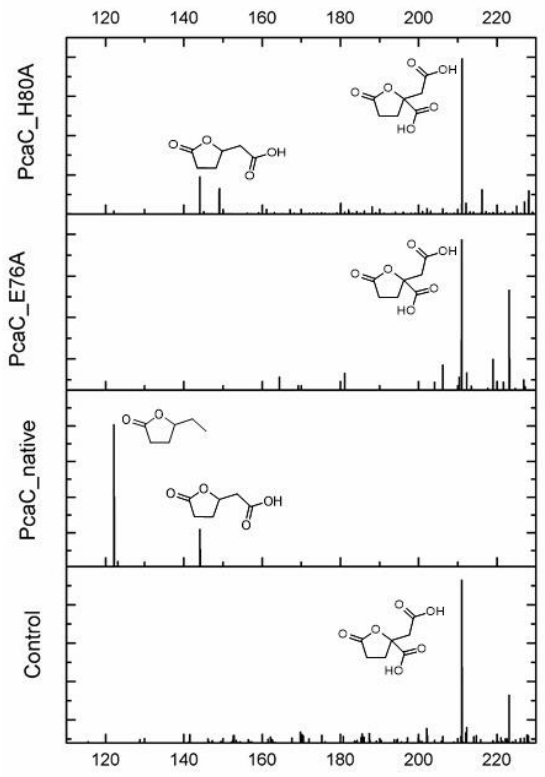

Figure 5-9 The mass spectrometry assay monitors the depletion of the substrate. In the mass of the substrate (homocitrate acid lactone) is 188.13 gm/mol and the product {(2-(5-oxotetrahydrofuran-2-yl) acetic acid)} mass 144.13 gm/mol. The presence of 188.13 mass peak is used as an indicator of whether the reaction is occurring. The mass spectrometry assay confirming the active site residues E76 and H80.

CMD mutants, a mass spectrometry-based assay was developed to validate the active site residues. In this assay, the presence/absence of substrate and product mass spectrometric peaks are monitored. In a control sample which comprises of the substrate (HCAL) and no protein, only the HCAL substrate peak is observed. The substrate mass peak corresponds to 211 which is equivalent to the mass of the HCAL (188.13gm/mol) plus mass of sodium (23 gm/mol). This demonstrates that the substrate is fully intact. In the reaction sample which contain both substrate and native PcaC protein, only the product, (2-(5-oxotetrahydrofuran-2-yl) acetic acid), mass peak at 144 is observed. An additional mass peak at 122 is also seen, which also corresponds to the product. During the ionization process in mass spectrometry, the terminal carboxylic group of the product is also ionized and resulting in the 122-mass peak. Thus, for native PcaC the HCAL peak fully disappear and two product mass peaks at 144 and 122 appears. While testing the assay for the two mutants, E76A and H80A, it was observed that the reaction sample comprising of E76A showed only one mass peak at 211, whereas the H80A containing reaction sample showed both 211 and 144 mass peaks. This indicates that the E76A

is an inactive mutant whereas the H80A mutant has some reduced activity as it gives both substrate and product peak. Nonetheless, the activity loss of H80A mutant cannot be assayed without development of an assay.

5.9 Modelling of the substrate

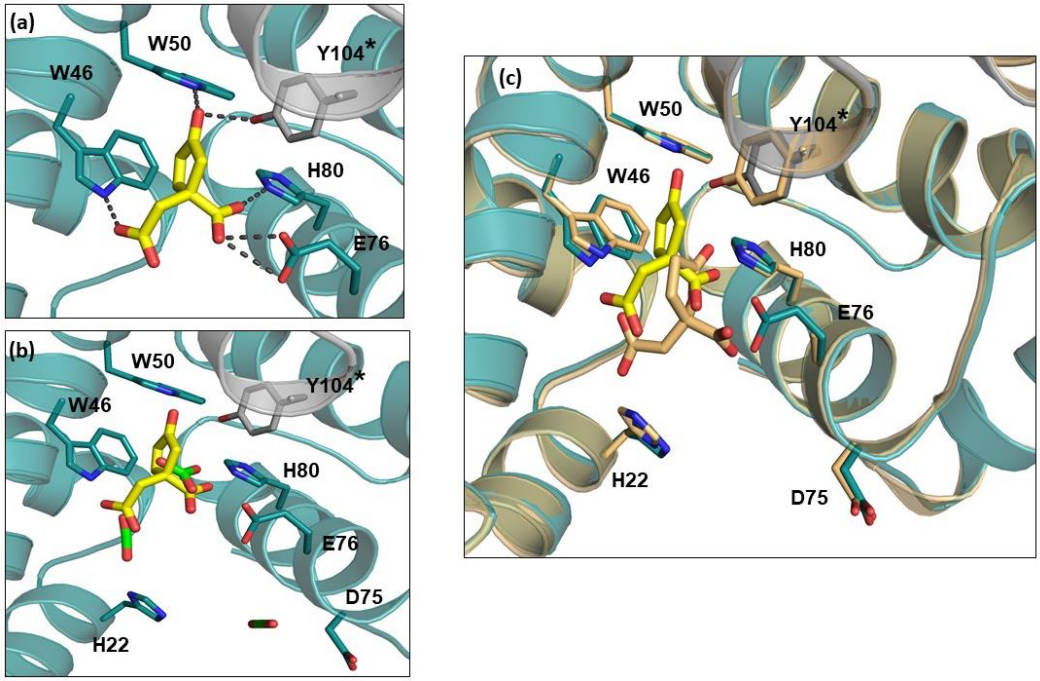

Figure 5-10 (a) Active site representation of PcaC modelled with γ-carboxymuconolactone (b)Superimposition of the modelled γ-carboxymuconolactone with the quaternary complex of native PcaC with CO_3^{2-} and CO_2, indicating that the carbonate ion bind at the equivalent site where the decarboxylation takes place. (c) Superimposition of the modelled γ-carboxymuconolactone with binary complex of PcaC-E76A with homocitrate acid lactone (HCAL).

To investigate the binding of the natural substrate of the PcaC catalysed reaction, docking of γ-carboxymuconolactone in the active site was carried out. Using the native structure of the PcaC as a starting point, the substrate was docked into the binding pocket using AutoDock (Morris et al. 1998). The substrate itself and the active site side chains were allowed freedom of movement. A total of 10 unique docked poses were generated. Coordinate of the lowest energy docked substrate was then merged with the native structure of PcaC and further energy minimized using REFMAC5 (Murshudov et al. 2011) without the X-ray term in order to optimize interaction between the substrate and the residues in the binding pocket. However, the residues in the pocket remain unmoved.

Modelling of γ-CMD reveals a shift in the position of the carbonyl oxygen of the substrate towards Tyr-104 and Trp-50 forming hydrogen bonds with these residues. This structure also shows hydrogen bonds made between the hydroxyl group and Glu-76 as well as between carboxylate group and His-80 similar to the interaction of CO_3^{2-} with these residues in the sodium bicarbonate complex (Figure 5-10).

5.10 Discussion

γ-CMD is a decarboxylase enzyme that belongs to the aromatic compound degradation pathway in bacteria (Parke et al. 1973). Although the first structure of γ-CMD (PDB: 2AF7) was solved in 2005 by the Northeast Structural Genomics Consortium, active site of the enzyme remains elusive. Detecting of the active site residue of γ-CMD was proven to be quite challenging due to low sequence identity with any other well characterized enzymes. So was the case for the γ-CMD enzyme from catabolic pathway of *P. putida*, called PcaC (Ornston 1966) of which we solved the crystal structure.

To detect the active site residue of PcaC, we invented a novel method of "Bicarbonate Soaking". Since, PcaC is a decarboxylase enzyme and it was assumed that its active site residues will interact with "carboxylic group" as decarboxylases are known to catalyze the removal of carboxylic group from the substrate by producing CO_2 as a byproduct. Bicarbonate closely mimic a carboxylic group as it is deprotonated form of carbonic acid (which is an example of a straight chained saturated carboxylic acid) and is a readily available substrate for soaking experiment. In this method the crystals of apo PcaC were soaked with sodium bi-carbonate solution with variable soaking time and a number of data sets are collected based on the different soaking time. The basis of this method is that, the bicarbonate mimics the carboxylic group of the substrate and binds to the active site at the equivalent position of the substrate's carboxylic group interacting with the exact active site residues of the PcaC enzyme. Furthermore, with increased soaking time of the crystals at room temperature, the bicarbonate sitting at the active site undergoes spontaneous dissociation to form the carbon dioxide which exits the enzyme via the protein's product route, mirroring the actual decarboxylation reaction and its product path. Using this method, we have shown that the bicarbonate interacts with amino acids 76E and 80H along with 104Y (from another monomer of the dimer) in the PcaC-bicarbonate (apo-bicarbonate) structure. We have also shown that the first CO_2 molecules is

~3.9Å away from bicarbonate and interacts with 46W residue. The CO_2 earmarks the middle of the product pathway which starts from the bicarbonate location. The second CO_2 molecule is located near the exit of the binding cavity which is ~10 Å from the first CO_3^{2-}. We have established that, as the reaction continues CO_2 is formed from bicarbonate (by spontaneous dissociation) and CO_2 interact with residues which guides it to bind across different sites to exit the product pathway. This active site residues E76 and H80, identified by bicarbonate soaking method were confirmed independently by soaking experiment of PcaC mutant with HCAL and 2FAA analogue and the mass spectrometry assay. We further modelled the biological substrate of PcaC, γ-carboxymuconolactone, into the active site to understand the interaction of the active site residue with the substrate.

From the modelled structure, it can be seen that the positively charged NE2 group of a strictly conserved H80 forms hydrogen bond to the leaving carboxylic group of γ-carboxymuconolactone and to the carboxylic group of conserved Glu-76 (Figure 5-9). This implies that role of the His-80 residue is to orient carboxylic group of the substrate γ-carboxymuconolactone with the carboxylic group of the Glu-76 residue in a plane for optimal catalysis. The modelled structure of PcaC and γ-carboxymuconolactone, suggests that at least 2 residues, Glu-76 and His-80, are close to the substrate and in a position suitable for playing a crucial role in catalysis. The imidazole ring of His-80 is at a distance of ~3 Å. And the carboxylate group of Glu-76 is close to the carboxylate of the substrate: the distance between oxygen atoms of the two negatively charged groups is ~3 Å, probably causing an electrostatic repulsion between them. This destabilize the ground state of the substrate, facilitating the release of its carboxylate group as CO_2. The decarboxylation leads to formation of a carbanion intermediate. The stabilization of the carbanion is achieved through the delocalization into aromatic systems of the generated carbanion, forming an enol intermediate which rearranges to give the final keto product. The mass spectrometry-based assay, the mutant-ligand complexes and modelled structure supports this proposed mechanism. (Figure 5-11).

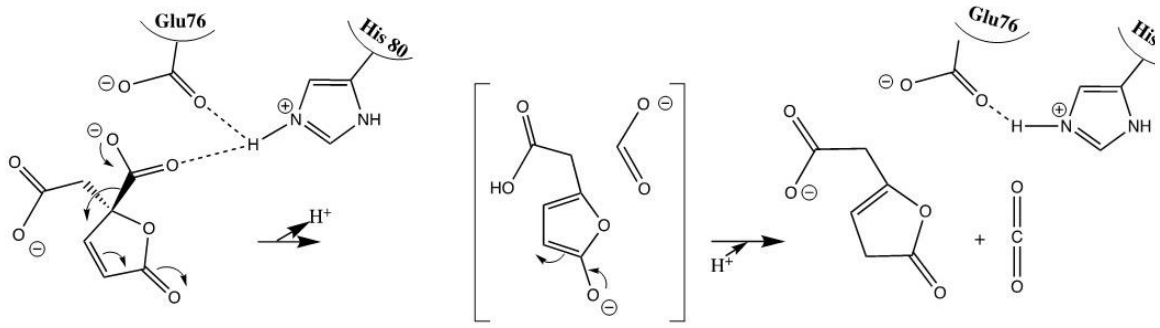

Figure 5-11 The proposed mechanism for γ-CMD family of protein. The negative charge region of the E76 positions the anionic carboxylate of the substrate, which helps to destabilize the ground state of the substrate. Meanwhile, the positively charged NE2 group of a strictly conserved His-80 is positioned close to the leaving carboxylate of the γ-carboxymuconolactone. The carbanion is formed as the CO_2 leaves the substrate, which is stabilized by charge delocalization. This leads to acidic enol formation which is finally converted to the products β-ketoadipate enol-lactone and carbon dioxide respectively.

5.11 Conclusion

The bioinformatics, structural and mass spectrometric analysis have provided new insight into the understanding of the γ-CMD superfamily. The identification of the active site residues in γ-CMD family using the novel bicarbonate soaking method has also enabled the tracing of the product release pathway for CO_2. The substrate/product analogue bound structures for γ-CMD family have been reported for the first time and the mutant crystal structure with reaction intermediate has aided in proposing the mechanism of the γ-CMD family.

5.12 References

Broadway, NM, Engel, PC, Bahnson, BJ, Harijan, RK, Li, D, Agnihotri, G, Dakoji, S, Oh, E, Lantz, M, & Liu, HW 2004, 'An atomic-resolution mechanism of 3-hydroxy-3-methylglutaryl-CoA synthase' *Proceedings of the National Academy of Sciences*, vol. 121, no. 47, pp. 16399–16400.

Chari, RVJ, Whitman, CP, & Kozarich, JW 1987, 'Absolute stereochemical course of muconolactone .DELTA.-isomerase and of 4-carboxymuconolactone decarboxylase: a 1H NMR "ricochet" analysis' *J. Am. Chem. Soc*, vol. 109, pp. 5520–5521.

Chen, X, Hu, Y, Yang, B, Gong, X, Zhang, N, Niu, L, Wu, Y, & Ge, H 2015, 'Structure of lpg0406, a carboxymuconolactone decarboxylase family protein possibly involved in antioxidative response from Legionella pneumophila' *Protein Science*, vol. 24, pp. 2070–2075.

Clarke, TE, Romanov, V, Chirgadze, YN, Klomsiri, C, Kisselman, G, Wu-brown, J, Poole, LB, Pai, EF, & Chirgadze, NY 2011, 'Crystal structure of alkyl hydroperoxidase D like protein PA0269 from Pseudomonas aeruginosa : Homology of the AhpD-like structural family' *BMC Structural Biology*, vol. 11, no. 27.

Eulberg, D, Lakner, S, Golovleva, LA, & Schlomann, M 1998, 'Characterization of a protocatechuate catabolic gene cluster from *Rhodococcus opacus* 1CP: Evidence for a merged enzyme with 4- carboxymuconolactone-decarboxylating and 3-oxoadipate enol-lactone- hydrolyzing activity' *Journal of Bacteriology*, vol. 180, no. 5, pp. 1072–1081.

Hillas, PJ, Soto, F, Oyarzabal, J, Wilks, A, & Montellano, PRO De 2000, 'The AhpC and AhpD antioxidant defense system of Mycobacterium tuberculosis' *The Journal of Biological Chemistry*, vol. 275, no. 25, pp. 18801–18809.

Holm, L, Kääriäinen, S, Rosenström, P, & Schenkel, A 2008, 'Searching protein structure databases with DaliLite v.3' *Bioinformatics*, vol. 24, no. 23, pp. 2780–2781.

Ito, K, Arai, R, Fusatomi, E, Kamo-Uchikubo, T, Kawaguchi, S, Akasaka, R, Terada, T,

Kuramitsu, S, Shirouzu, M, & Yokoyama, S 2006, 'Crystal structure of the conserved protein TTHA0727 from *Thermus thermophilus* HB8 at 1.9 Å resolution: A CMD family member distinct from carboxymuconolactone decarboxylase (CMD) and AhpD' *Protein Science*, vol. 15, pp. 1187–1192.

Iwagami, SG, Yang, K, & Davies, J 2000, 'Characterization of the protocatechuic acid catabolic gene cluster from *Streptomyces* sp. strain 2065' *Applied and Environmental Microbiology*, vol. 66, no. 4, pp. 1499–1508.

Jiménez, JI, Miñambres, B, García, JL, & Díaz, E 2002, 'Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440' *Environmental Microbiology*, vol. 4, no. 12, pp. 824–841.

Koshkin, A, Nunn, CM, Djordjevic, S, & Montellano, PRO De 2003, 'The mechanism of *Mycobacterium tuberculosis* alkylhydroperoxidase ahpd as defined by mutagenesis, crystallography, and kinetics' *The Journal of Biological Chemistry*, vol. 278, no. 32, pp. 29502–29508.

Koshkin, A, Zhou, X, Kraus, CN, Brenner, JM, Bandyopadhyay, P, Kuntz, ID, Barry III, CE, & Montellano, PRO De 2004, 'Inhibition of *Mycobacterium tuberculosis* AhpD , an element of the peroxiredoxin defense against oxidative stress' *Antimicrobial Agents and Chemotherapy*, vol. 48, no. 7, pp. 2424–2430.

Lee, HY & Yang, JK 2009, 'crystallization communications Crystallization and preliminary X-ray crystallographic analysis of γ-carboxymucolactone decarboxylase from Sulfolobus solfataricus crystallization communications' *Acta Crystallographica Section F Structural Biology CommunicationsStructural Biology and Crystallization Communications*, vol. 65, pp. 1197–1199.

Morris, GM, Goodsell, DS, Halliday, RS, Huey, R, Hart, WE, Belew, RK, & Olson, AJ 1998, 'Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function' *Journal of Computational Chemistry*, vol. 19, no. 14, pp. 1639–1662.

Moss, GP 2019, 'Enzyme Nomenclature- Recommendations', Nomenclature Committee of

the International Union of Biochemistry and Molecular Biology, (NC-IUBMB), https://www.qmul.ac.uk/sbcs/iubmb/enzyme/EC4/1/1/.

Murshudov, GN, Skubák, P, Lebedev, AA, Pannu, NS, Steiner, RA, Nicholls, RA, Winn, MD, Long, F, & Vagin, AA 2011, 'REFMAC5 for the refinement of macromolecular crystal structures' *Acta Crystallographica Section D*, vol. D67, pp. 355–367.

Nickerson, LA, Huynh, V, Balmond, EI, Cramer, SP, & Shaw, JT 2016, 'Asymmetric synthesis of homocitric acid lactone' *The Journal of Organic Chemistry*, vol. 81, pp. 11404–11408.

Nunn, CM, Djordjevic, S, Hillas, PJ, Nishida, CR, Paul, R, & Montellano, O de 2002, 'The crystal structure of Mycobacterium tuberculosis alkylhydroperoxidase AhpD, a potential target for antitubercular drug design' *The Journal of Biological Chemistry*, vol. 277, no. 22, pp. 20033–20040.

Ornston, LN 1966, 'The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida* III. Enzymes of the catechol pathway' *The Journal of Biological Chemistry*, vol. 241, no. 16, pp. 3795–3799.

Parke, D, Meagher, RB, & Ornston, LN 1973, 'Relationships among enzymes of the β-ketoadipate pathway. III. Properties of crystalline γ-carboxymuconolactone decarboxylase from *Pseudomonas putida*' *Biochemistry*, vol. 12, no. 18, pp. 3537–3542.

Schörken, U & Sprenger, GA 1998, 'Thiamin-dependent enzymes as catalysts in chemoenzymatic syntheses' *Biochimica et Biophysica Acta*, vol. 1385, pp. 229–243.

Schuitz, LW, Chivers, PT, & Raines, RT 1999, 'The CXXC motif: Crystal structure of an active-site variant of Escherichia coli thioredoxin' *Acta Crystallographica Section D*, vol. 55, pp. 1533–1538.

Shahbaaz, M, Ahmad, F, & Hassan, MI 2015, 'Structure-based functional annotation of putative conserved proteins having lyase activity from *Haemophilus influenzae*' *Biotech*, vol. 5, pp. 317–336.

Smart, OS, Neduvelil, JG, Wang, X, Wallace, BA, & Sansom, MSP 1996, 'HOLE : A program for the analysis of the pore dimensions of ion channel structural models' *Journal of Molecular Graphics*, vol. 14, pp. 354–360.

Walsh, CT 2019, 'Biologically generated carbon dioxide : Nature's versatile chemical strategies for carboxy lyases' *Natural Product Reports*.

Yamanashi, T, Kim, S-Y, Hara, H, & Funa, N 2015, 'In vitro reconstitution of the catabolic reactions catalyzed by PcaHG, PcaB, and PcaL: the protocatechuate branch of the β-ketoadipate pathway in Rhodococcus jostii RHA1.' *Bioscience, Biotechnology, and Biochemistry*, vol. 79, no. 5, pp. 830–835.

Yeh, W, Durham, DONR, Fletcher, P, & Ornston, LN 1981, 'Evolutionary relationships among y-carboxymuconolactone decarboxylases' *Journal of Bacteriology*, vol. 146, no. 1, pp. 233–238.

Yeh, W, Fletcher, P, & Ornston, N 1980, 'Homologies in the NH2-terminal amino acid sequences of γ-carboxymuconolactone decarboxylases and muconolactone isomerase' *The Journal of Biological Chemistry*, vol. 255, no. 13, pp. 6347–6354.

Chapter 6
Summary and conclusions

6.1 Summary

The recent development of the industrial sectors has largely contributed to the increased level of water pollution. Industrial and urban activities for instance, agriculture, mining, petroleum refining etc. cause discharges of pollutants into the terrestrial and aquatic bodies. These pollutants are mostly aromatic and non-degradable in nature and often accumulate in the environment, exceeding the permissible exposure limit (PEL) in the water systems. The xenobiotic pollutants such as polycyclic aromatic hydrocarbons (PAHs) and heterocyclic aromatics derivatives are of major concern. They are carcinogenic and mutagenic in nature and cause extreme environmental hazard because of their high lethality and tenacity. These chemical contaminants that are released in the environment, enter the food chain and affect the public health. There is a need to eradicate these poisonous pollutants for the well-being of the community. Keeping the suitability factor in mind, emphasis has been made to develop methodologies that are energy efficient and has low chemical impact on the environment. Soil bacteria such as *Pseudomonas sp., Acinetobacter sp*., have caught the eye of the researchers in modern times as they can assimilate these pollutants, aiding in recycling of these toxic pollutants from the surroundings. These bacteria have metabolic pathway that can breakdown the aromatic compounds into non-toxic TCA cycle intermediates. They have four major pathways: homogentisate pathway, the phenylacetate pathway, the catechol and the protocatechuatepathway. The catechol and protocatechuate pathway join to form the central β-ketoadipate pathway. Thus, microbial degradation of the aromatic pollutant is a promising ecofriendly approach to remove pollutants. Knowledge and effective understanding of these aromatic catabolic pathway will assist in developing strategies to eliminate these chemicals contaminates.

In this thesis two proteins, PcaF and PcaC, have been chosen for study, both belong to the β-ketoadipate pathway. For the first protein PcaF, a degradative thiolase, the main objectives are to understand the lesser known degradative thiolase and highlight the structural difference between degradative and biosynthetic thiolase. Initially bioinformatics studies, which included sequence-based alignment study of the active sites of two thiolases are carried on. The active residues of PcaF are also mutated to trap reaction intermediates. Single (H356A: A-mutant) and double active site mutations (H356A-C90S: AS mutant, and H356A-C386A: AA-mutant) PcaF proteins are purified and crystallized. These crystals are soaked with variety of ligands; CoA and CoA derivatives to understand the binding of degradative thiolases with longer chain CoA. From the crystal structures, it is observed that A-mutant and AA- mutants complexed with CoA form disulfide bond between the C90 active site residue and the thiol group (–SH atom) of the CoA which causes covalent locking of the active site. This study further delves into demarcation of between the biosynthetic and degradative thiolases, where a long tunnel of 235 Å2 surface area is observed in PcaF. A-mutant-Hexanoyl CoA complex (A-mutant-Hex-CoA) and AS-mutant-Octanyol CoA (AS-mutant-Oct-CoA) crystal structures confirmed this tunnel to be the longer chain binding tunnel as the acyl chain of hexanal and octanal CoA are found bound in the tunnel. Lastly, a unique covering loop is shown to play role in functioning of the degradative thiolase.

The second protein of interest is PcaC that belong to caboxymuconolactone decarboxylase (CMD) family of protein. This family has three group of protein; group one, well known for the CXXC-motif based peroxidase activity and two other groups which lacks the peroxidase activity. Group two is the γ-caboxymuconolactone decarboxylase (γ-CMD) group involved in the aromatic degradation and a group three whose function is unknown. In this study, the background on the elusive CMD is pursued which emphasis on γ-CMD (group 2), where PcaC fit into. The research gap for this protein is straightforward as nothing much is known about γ-CMD. For this class of protein, the aim is to identify the active site and the mechanism of PcaC. The chapter starts with understanding the structure of PcaC and correlating it with the known CMD structures to get a clue about the active site. A novel method of bicarbonate soaking is employed to establish the active site. PcaC crystals soaked with bicarbonate with different time intervals manages to trace the bicarbonate binding position and the trailing CO_2 pathway. The residues E76 and H80 are involved in bicarbonate binding and are mutated to E76A and H80A,

respectively. The PcaC mutants' crystals are soaked with substrate analogue homocitrate acid lactone (HCAL) and product analogue 2-Furan acetic acid (2FAA). A confirmatory massed spectrometry-based assay using HCAL confirm E76 as the active site, this observation is in concordance with the E76A-HCAL crystal structure. Based on the crystal structure and mass spectrometry assay, the original PcaC substrate (γ-carboxymuconolactone) is modelled into the active site and the mechanism for γ-CMD family is proposed.

In this thesis, detailed information on the two protein has been shed. The studies conducted successfully achieves it goals. For the PcaF protein structural differentiating point has been identified that can help in identifying the two class of thiolases. Also, it has shown light on how degradative thiolase binds to the longer chain thiolases. For the PcaC the active site architecture and the residues are identified, and mechanism is proposed.

The research work provides a detailed understanding of PcaF and PcaC which can be utilized to improve the enzyme stability, substrate specificity and kinetic properties of the two enzymes; which in turn will aid in the development of bio-sensory/ water purification machinery that can degrade the aromatic pollutants. The structural and functional characterization of the two key enzyme, PcaF and PcaC, will also help in devising novel strategies that will ultimately yield technologies focused on degradation of the xenobiotics compounds, such as protein based biosensor degradation methods etc.

6.2 Concluding Remarks

This thesis revolves around two protein PcaF and PcaC from the β-ketoadipate pathway. Both the protein has been chosen because of their tactical positioning in the pathway. The objective of the study has been not only to understand the two bioremediation proteins, but as well as get an insight into the family/class they belong. The summarization of the work is bulleted as following:

- The active site residues that lining the wall of the reaction centre are different for degradative and biosynthetic thiolases. In the degradative thiolases, the active site architecture residues are flexible compared to the highly conserved biosynthetic thiolase.

- The dual role of the H356 residue in PcaF is observed. Apart from being the part of the catalytic triad, H356 also act as an “anchoring residues” to tether the large Coenzyme A (CoA) or CoA derivatives.

- Mutation of the H356 residue to alanine led to a ‘gridlocked thiolase’ structure, where the CoA is found to be covalently linked to the catalytic C90 residue, inhibiting the overall reaction.

- Analysis of the PcaF crystal structure revealed a long amphiphilic tunnel of 235 $Å^2$ surface area. This implicated the ability of PcaF to bind longer CoA derivatives.

- X-ray structures of PcaF mutants with two long chain CoA derivatives, hexanal-CoA and octanal-CoA, exhibited the interaction of tunnel residues with the acyl tail of the CoA derivatives and showcased the promiscuity of PcaF degradative thiolase.

- The structural investigation of PcaF helped in recognising the role of covering loop in determining the tunnel length and the nature of the substrate that will bind.

- Novel bicarbonate soaking method was invented to detect the active site of PcaC protein, as it displayed low sequence identity with any other well characterized enzymes.

- The bicarbonate soaking method also traced the product release pathway for CO_2, highlighting the residues interacting with the CO_2 during its exit from the pathway.

- The active site of γ-CMD was identified and the active site residues, E76 and H80, were located. Mass spectrometry-based assay was designed that helped in validating the role of E76 and H80 as the active site residues

- Four ligand bound structure, involving substrate (HCAL) and product (2FAA) analogue, of the PcaC mutants were reported for the first time. It helped in understanding the substrate/product binding in γ-CMD. The H80A-HCAL soaked structure, had the intermediate, (2-(5-oxotetrahydrofuran-2-yl) acetic acid), in the active site that aided in proposing the reaction mechanism.

- Based on the experimental and bioinformatic studies, the mechanism was validated, where the E76 is involved in destabilisation of the substrate and the role of histidine residue is to orient carboxylic group of the substrate.

6.3 Future work

Based on our study, we have shown that despite the structural similarity and same catalytic triad both groups of thiolases have defined structure-based clues that can be used to identify the two classes of thiolases. Nonetheless, the exact role of these defined features is yet to be explored. It will be interesting to understand the function of the residue lining the active site in reaction mechanism. Our work has also provided in-depth understanding of how longer chain substrate bind into the active site tunnel, which can be further utilized to tailor-make the tunnel to permit binding of specific substrate. Overall, the structural features defined in our study can be used to modify or switch the degradative thiolase into more biosynthetic type and vice-versa.

Till date, all the enzymes of the β-ketoadipate pathway, except the PcaC, was well-studied. Our work has shed detailed light into active site architecture, active site residue, substrate binding and mechanism, however, there is still a lot left to discover. The role of the other residues in active site is yet to investigated. There little to no understanding about the promiscuity of this protein. It will be interesting to study PcaC with a variety of substrate/product analogue and gain insight into the residues that partake in the specificity/flexibility of the protein. There is also a need to establish an assay which will help in determining the activity of the residues, as our mass spectrometry-based assay could only give a yes or no answer. From the PcaC study, the role of AphD and γ-CMD proteins in the survival of bacteria in critical conditions like oxidation stress and xenobiotic environment is evident. AphD employ the CXXC motif to perform the peroxidase activity, meanwhile γ-CMD uses E76 and H80 active site residues to catalyze decarboxylase activity. Analyzing the structures of PcaC with AphD protein, from M. tuberculosis (Mt-AhpD), has shown that the E76 active site overlapped with the CXXC motif of AphD. The $R_{73}XXE_{76}XXXH_{80}$ motif of PcaC coincides with the CXXCXXXH motif in Mt-AhpD. This implies that both the groups employ same active site cavity and equivalent residue positions to perform peroxidase (AphD) and decarboxylase activity (γ-CMD). It will be interesting to mutate the R73 and E76 residues in PcaC to R73C and E76C and check the mutated PcaC protein for the gain in peroxidase activity and loss of decarboxylase activity, which will confirm the transition of the γ-CMD to AphD protein. This experiment will shed more light on the γ-CMD proteins and will help us to understand the evolution of the CMD family itself.

In the thesis, the structure and function of the two key enzyme, PcaF and PcaC, is demonstrated. The study on PcaF has shed ample light on the degradative class of thiolase. It has provided better understanding of the active site architecture of the degradative thiolase and its dissimilarity from the biosynthetic thiolase. The work has exposed the binding of the longer chain acyl CoA in degradative thiolase and has delivered in-depth information about the size and nature of the ligand binding tunnel. This knowledge can be used to modify the ligand binding tunnel to allow binding of specific type of substrates. Overall, the study has presented valuable structural information that can be applied to classify uncharacterized thiolases into degradative and biosynthetic type. The work on PcaC has restored the final piece of β-ketoadipate pathway by detecting the active site residues in PcaC and by elucidated the mechanism of the γ-CMD family of proteins. The study has also aided in the development of a novel method to identify the active site residues in decarboxylases where the active site cannot be detected based on the known structures. The research can be used to modify the active site of the PcaC to allow binding of varied substrates.

LIST OF PUBLICATIONS

Publications in Journal (PhD thesis)

Bhaskar, S, Steer, DL, Anand, R, and Panjikar, S (2020), 'Structural basis for differentiation between two classes of thiolase: Degradative vs biosynthetic thiolase' *Journal of Structural Biology: X*, doi: https://doi.org/10.1016/j.yjsbx.2019.100018

Bhaskar, S, Nickerson, L, Shaw, J, Anand, R, and Panjikar, S, Uncovering carbon dioxide path in decarboxylation reaction: Enlightening the active site and mechanism of γ-CMD family (To be submitted).

Patents

Bhaskar, S, Panjikar, S and Anand, R (2020), 'Method to identify active site on decarboxylase enzyme', patent application no 202021000473.

Other Publications

Bhaskar, S, Anand, R., and Panjikar, S. Transcriptional regulators of the NtrC family: structural insight into the aromatic effectors binding proteins (Manuscript under preparation).

Conference Presentation

Bhaskar, S, Panjikar, S, and Anand, R (2017), Crystallographic analysis of β-ketoadipyl-CoA thiolase from Pseudomonas putida, IUCR 2017.

Appendix

A1: Intact mass determination

A1.1: Native PcaF protein

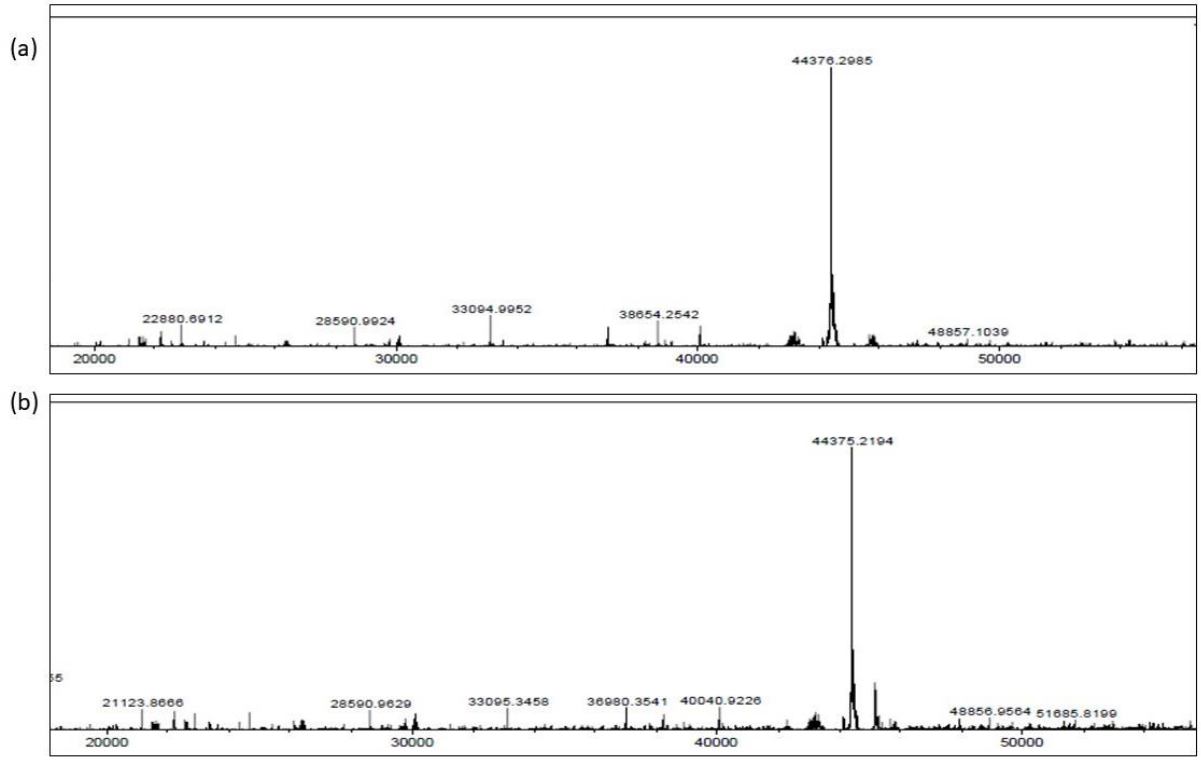

Figure A1.1 (a) The intact mass spectrometry spectrum for the APO-native (control) protein with a peak at 44376.3 Da. (b) The experimental sample is the Apo-acetoacetyl CoA complex which has a peak of 44376.3 D

A1.2: AA-mutant protein

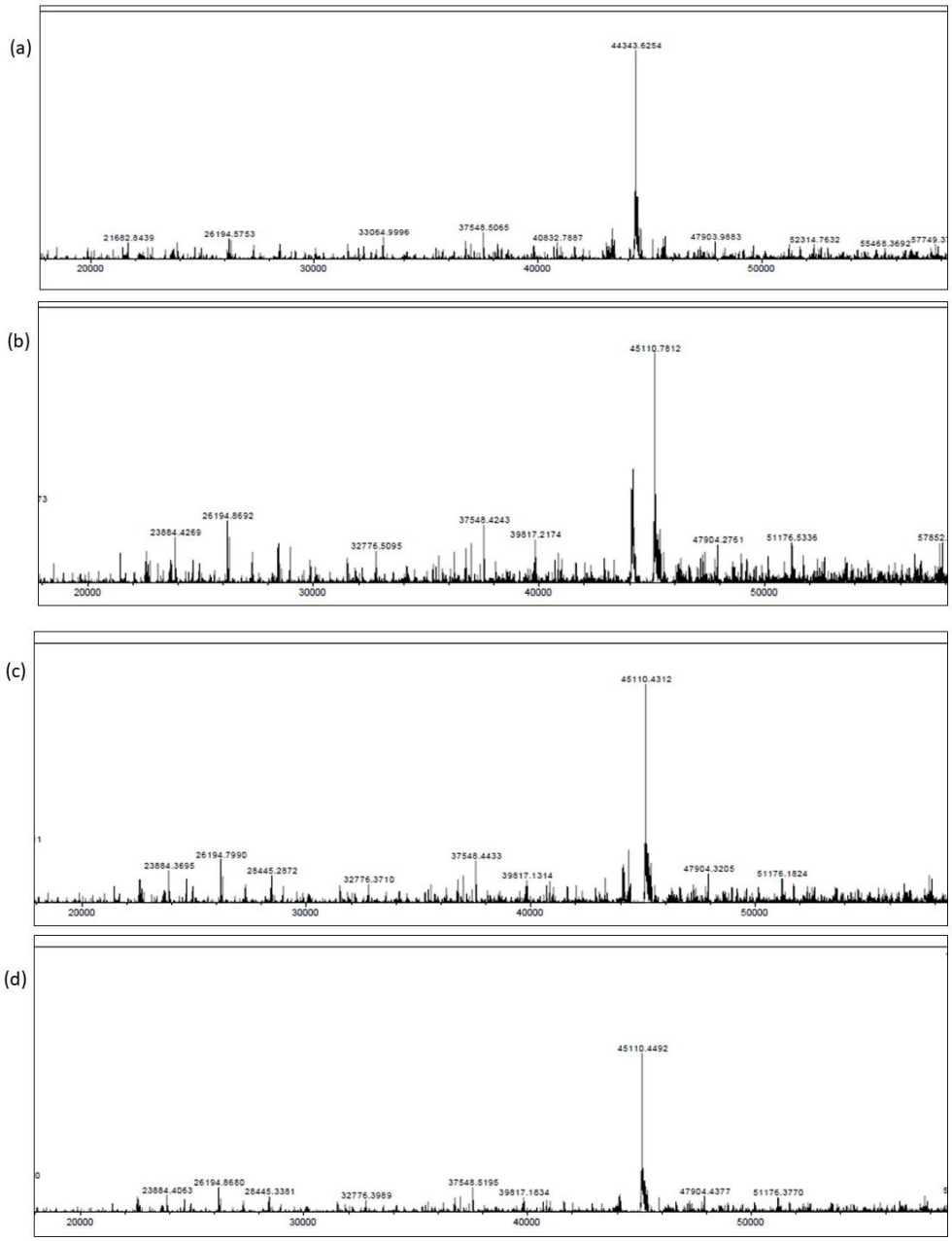

Figure A1.2 (a) The intact mass spectrometry spectrum for the AA-mutant (control) protein with a peak at 44343.6 Da (b) The experimental sample is the AA-acetoacetyl CoA complex which has a peak of 45110.8 Da (c) The experimental sample is AA-hexanyol CoA complex which has a peak of 45110.4 Da (d) The experimental sample is the AA-octanyol CoA complex which has a peak of 45110.4 Da.

A2: RSCC values for the ligand found in the PcaF-ligand complex

LIGAND	CHAIN	RSCC VALUE
COA	A	0.87
COA	B	0.85
COA	C	0.87
COA	D	0.88
GOL	A	0.98
GOL	B	0.97
GOL	C	0.96
GOL	D	0.96

Table A2-1: RSCC values for the ligands found in 6PCB (A-mutant CoA complex).

LIGAND	CHAIN	RSCC VALUE
COA	A	0.90
COA	B	0.88
COA	C	0.90
COA	D	0.88
O8Y	A	0.91
O8Y	B	0.93
O8Y	C	0.96
O8Y	D	0.92
GOL	A	0.95
GOL	B	0.94
GOL	C	0.95
GOL	D	0.96

Table A2.2: RSCC values for the ligands found in 6PCC (A-mutant-Hex-CoA complex).

LIGAND	CHAIN	RSCC VALUE
COA	A	0.71
COA	D	0.69
OYA	A	0.82
OYA	B	0.86
OYA	D	0.82
GOL	A	0.98
GOL	B	0.98
GOL	C	0.98
GOL	C	0.96
GOL	D	0.97

Table A2.3: RSCC values for the ligands found in6PCD (AS-mutant-Oct-CoA).

A3: PcaF: Enzymatic Assay

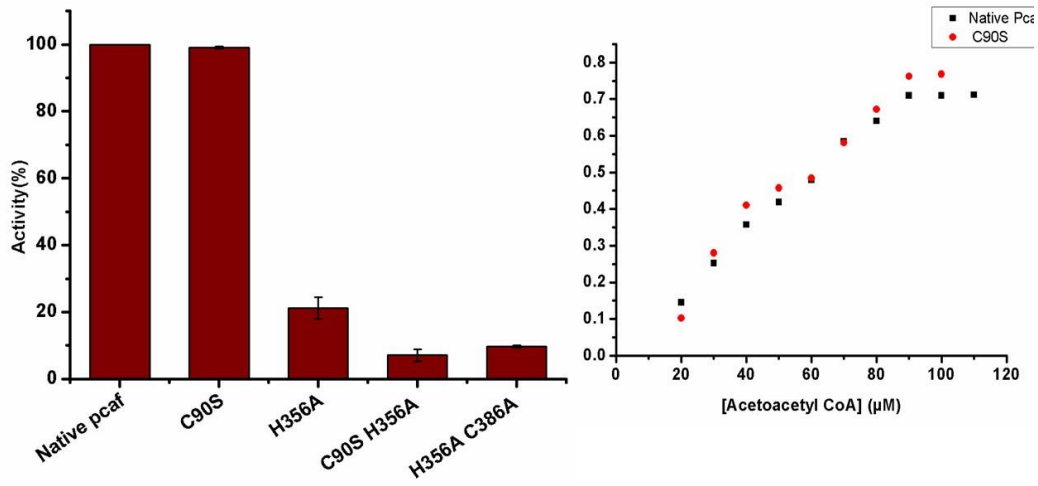

Figure A3Mutant