Structural and Functional Characterization of PaaK1 & PaaK2: Phenylacetate CoA Ligase paralogs from *Burkholderia cenocepacia* J2315

by

Adrienne Law
BSc. University of Guelph, 2007

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

Supervisory Committee

Structural and Functional Characterization of PaaK1 & PaaK2: Phenylacetate CoA Ligase paralogs from *Burkholderia cenocepacia* J2315

by

Adrienne Law
BSc. University of Guelph, 2007

Supervisory Committee

Dr. Martin J. Boulanger, Department of Biochemistry and Microbiology
Supervisor

Dr. Alisdair B. Boraston, Department of Biochemistry and Microbiology
Departmental Member

Dr. John S. Taylor Department of Biology
Outside Member

Abstract

Supervisory Committee
Dr. Martin J. Boulanger, Department of Biochemistry and Microbiology
Supervisor
Dr. Alisdair B. Boraston Department of Biochemistry and Microbiology
Departmental Member
Dr. John S. Taylor Department of Biology
Outside Member

Aromatic compounds comprise approximately 25% of the Earth`s biomass. Accordingly, turnover of these thermostable compounds is essential to the biogeochemical carbon cycle. Specialized microbial enzymatic pathways are largely responsible for mineralization of aromatic compounds, and are extensively studied for bioremediation purposes. The phenylacetic acid degradation pathway (PAA) is of particular interest as it is utilized for degradation of a multitude of aromatic compounds, including environmental pollutants, and appears to be widely distributed among bacteria. Intriguingly, the PAA pathway has also been implicated as a virulence factor in the cystic fibrosis pathogen, *Burkholderia cenocepacia*. As such, detailed biochemical characterization of the PAA pathway holds great potential for improving bioremediation strategies for aromatic pollutants, as well as understanding carbon source utilization during *B. cenocepacia* infection.

A striking feature of the PAA pathway in *B. cenocepacia* is the presence of two genes encoding the phenylacetate CoA ligase (PCL) enzyme (*paaK1 and paaK2*), responsible for the initial CoA activation of phenylacetic acid. PCLs are members of the adenylate forming enzyme superfamily. However, sequence alignments reveal several intriguing features, including a potentially novel microdomain consisting of the initial ~80 N-terminal residues, which possesses percent identity to any structurally characterized family member. Furthermore, this superfamily utilizes a complicated catalytic mechanism, exploiting several conserved motifs during the reaction process. The precise roles of many key conserved residues are not yet well understood, especially during the pre-adenylation, ATP bound state, for which few high quality crystal structures exist.

In order to define the early stages of the catalytic mechanism, and to assess how the divergent polypeptide region may impact the PaaK enzymes, we have pursued a detailed structural characterization of the paralogs, complemented with functional assessments. Specifically, we have produced a 1.6 Å resolution crystal structure of PaaK1 in complex with ATP, which reveals a novel helical bundle arrangement at the N-terminal domain never before seen in this superfamily. Remarkably, homodimerization of PaaK1 appears to reconstitute potentially important β sheet interactions observed in the classical N-terminal arrangement of family members. Moreover, our structure is one of few which contain well ordered β and γ phosphates, allowing for detailed examination of significant protein-ATP interactions with conserved catalytic residues.

To better comprehend the roles of these residues over the course of the reaction, we have produced additional crystal structures of PaaK1 and PaaK2 in complex with the phenylacetyl adenylate intermediate. Notably, PaaK2 was captured following the domain reorientation, poised to catalyze thioesterification of phenylacetyl adenylate, providing insight into the later stages of the reaction process. Furthermore, the intermediate co-structures divulge the location of the aryl substrate binding pocket for both paralogs. Detailed comparisons of the binding pockets accompanied kinetic characterizations for both paralogs, demonstrating that PaaK2 possesses an apparent K_M of 150 μM for phenylacetic acid, more than double that of PaaK1 (62 μM). Our findings provide preliminary evidence for distinct functional roles of the PaaK paralogs in *B. cenocepacia,* while imparting additional insight into catalytic roles of conserved residues within the adenylate forming superfamily at large.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to gratefully acknowledge my supervisor, Dr. Martin Boulanger, for his patience, support, expertise during the experimental portion of this project, as well as his tactfully worded critiques and advice. I would also like to thank my committee members, Dr. Al Boraston and Dr. John Taylor their input and advice. Much thanks to my lab members, both past and present: Joanna Crawford, Ognjen Grujic, Jasleen Bains, Katia Bruic, Jeremy Mason, Michelle Tonkin, and Ben Farnell, with special thanks to Jasleen for her involvement in my training. I also would like to acknowledge Dr. Evans and lab members for the use of their equipment and moral support and encouragement.

Chapter 1: Introduction

1.1 Aromatic compounds: Abundant and Un-reactive

Aromatic compounds are generally considered planar compounds with a ring of conjugated bonds and defined as containing 'closed circuits of mobile electrons' (Schleyer 2001). This cyclic delocalization, or resonance, imparts a significant degree of stabilization, much more than would be expected from conjugation alone. Such compounds are ubiquitous in the environment and a major portion of the Earth's biomass. The benzene ring, the archetype of aromatic compounds, is one of the most widely distributed chemical units in the biosphere, second only to the glucosyl residue (Elder & Kelly 1994). Indeed, benzene containing compounds constitute 25% of the Earth's biomass (Gibson & Harwood 2002).

While the simple benzene ring may be the exemplar for aromatic compounds in general, an impressive diversity of aromatics exists in nature. A variety of functional groups may be substituted for one or more hydrogen atoms surrounding the aromatic nucleus, bestowing additional chemical properties to the molecule. Compounds containing elements other than carbon can also be incorporated into the aromatic ring. Thus, the term 'aromatic compound' embodies a diverse array of omnipresent chemical compounds, which are especially recalcitrant to mineralization and turnover.

In terms of sheer mass, plants are responsible for the synthesis and assembly of the bulk of aromatic compounds on Earth. As lignin makes up 20-30% of the dry weight of

vascular plant tissue (Colberg & Young 1982), understandably, the depolymerisation and turnover of lignin is responsible for generating much of the aromatic compounds found within the biosphere (Harwood & Parales 1996). In fact, the degradation of lignin and resulting aromatic compounds, which relies heavily on the actions of microorganisms, is believed to be the rate limiting step in the global carbon cycle (Orth *et al.* 1991).

1.2 Aromatic Degradation Pathways - Feeding the Global Carbon Cycle

The resonance stabilization within the aromatic nucleus presents a significant challenge for carbon turnover. The inherent chemical stability of aromatic compounds necessitates the use of specialized enzymatic pathways for the catalytic feat of ring cleavage; for the bacteria and fungi which possess these *bio-degradative* pathways, aromatic compounds represent a rich carbon and energy source.

1.2.1 Peripheral Enzymatic Pathways Facilitate Conversion to a Common Aromatic Metabolite

Microbes accomplish the degradation of a wide array of aromatic compounds to produce acyclic central intermediates such as pyruvate, acetate, succinate, or their corresponding CoA thioesters (Harwood & Parales 1996; Cao *et al.* 2009). Compound specific peripheral pathways are often first required to convert aromatic compounds to a common intermediate; functional groups on the aromatic nucleus must be removed, modified or added, where necessary (Elder & Kelly 1994; Harwood & Parales 1996; Cao *et al.* 2009). This efficient strategy channels the degradation of a multitude of compounds to converge at a single central pathway that is then utilized for the difficult task of dearomatization and ring cleavage (Figure 1).

1.2.2 Central Aromatic Degradation Pathways

The choice of common intermediate depends on which peripheral and central pathways are at the disposal of the microbe. Traditionally, aromatic degradation pathways have been divided into two classes that rely on aerobic or anaerobic degradation. This distinction is based on whether oxidative or reductive steps are performed on the aromatic nucleus in order to accomplish dearomatization and ring fission and depends largely on what type of microbe and environment the event takes place in.

Classical Aerobic Aromatic Degradation

Classical aerobic aromatic degradation is performed by aerobic bacteria as well as facultative anaerobes. An astounding variety of aromatic compounds, such as lignin degradation products as well as many other plant associated phenolic compounds, are first converted to catechol or protocatechuate (Harwood & Parales 1996) (Figure 1). Molecular oxygen is utilized for dearomatization and *meta* or *ortho* oxygenolytic ring cleavage (Harwood & Parales 1996). The newly acyclic compounds are then converted to compounds such as pyruvate, succinyl-CoA, and acetyl-CoA which enter the TCA cycle and central metabolism (Cao *et al.* 2009). Thus, the *meta* and *ortho* cleavage pathways play a central role in the degradation of dozens of aromatic compounds, with several peripheral pathways funnelling substrate towards a common catechol metabolite.

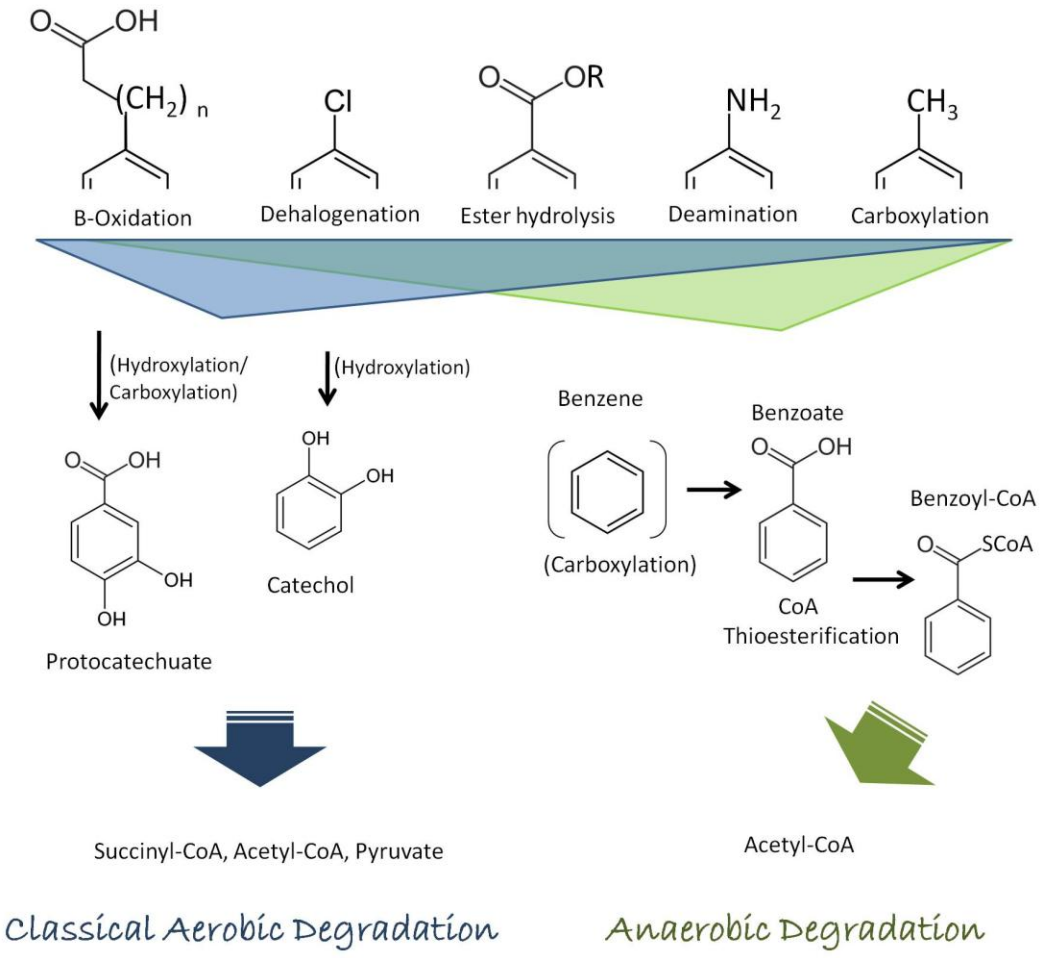


Figure 1: Peripheral Pathways for Conversion to a Common Intermediate
Removal, addition or modification of functional groups may be required for complete conversion to common intermediates, which then undergo dearomatization and ring cleavage. In the case of aerobic degradation pathways, addition of hydroxyl groups and/or carboxyl groups may be necessary to complete conversion to catechol or protocatechuate common intermediates. Benzoyl-CoA is the favoured common intermediate for anaerobic degradation; however, carboxylation prior to CoA activation may be necessary.

Anaerobic Aromatic Degradation Pathways

The anaerobic degradation of many aromatic compounds relies on the conversion to the common metabolite benzoyl-CoA (Figure 1) (Altenschmidt & Fuchs 1991; Dangel *et al.* 1991). The subsequent steps of metabolism necessitate a very different mechanism from that of aerobic metabolism: reduction of the aromatic moiety, followed by non-oxygenolytic ring cleavage (Elder & Kelly 1994). A second startling difference in classical anaerobic aromatic metabolism is that compounds are first rapidly activated to a CoA thioester (Harwood & Gibson 1986).

CoA is a well known acyl carrier in central metabolism, especially β-oxidation of fatty acids. As such, CoA activation may also serve several functions in aromatic degradation pathways. The CoA handle likely improves substrate recognition, maximizing affinity between the enzyme and substrate while the electron withdrawing properties of CoA may stabilize transition states and facilitate ring reduction (Copley & Crooks 1992). It has also been proposed to help trap the hydrophobic aromatic compound within the cell by increasing polarity, thereby decreasing membrane permeability and improving uptake (Harwood & Gibson 1986; Copley & Crooks 1992).

1.2.3 Hybrid Aromatic Degradation Pathways

While decades of extensive characterization has taken place for anaerobic and classical aerobic aromatic metabolic pathways, the hybrid aerobic aromatic degradation pathways are much more recently discovered. The concept of hybrid aromatic degradation pathways was first proposed in 1989 by Zeigler *et al*. Presently, the *aerobic* metabolism of benzoate, 2-aminobenzoate, and phenylacetic acid is understood to proceed via activation to their respective CoA thioesters within several bacterial strains and species

(Martinez-Blanco *et al.* 1990; Ziegler *et al.* 1989; Ismail *et al.* 2003). As in anaerobic degradation pathways, the thioester handle is maintained for all subsequent steps of the hybrid pathways (Ismail *et al.* 2003; Martinez-Blanco *et al.* 1990), imparting all the same advantages as with anaerobic aromatic degradation pathways. Aerobic growth conditions are required for utilization of any of the hybrid pathways as molecular oxygen is critical for the de-aromatization step (Ismail *et al.* 2003); however, many additional mechanistic details for these hybrid pathways remain poorly understood.

1.3 The Phenylacetic Acid Degradation Pathway

The phenylacetic acid degradation pathway (PAA) is one of only three known hybrid aromatic degradation pathways, and catalyzes the conversion of phenylacetic acid to acetyl-CoA and succinyl-CoA (Nogales *et al.* 2007). At the most basic level, the PAA pathway enables growth on phenylacetic acid as a sole carbon and energy source in aerobic conditions (Martinez-Blanco *et al.* 1990). However, the PAA pathway is also a major point of convergence for the degradation of many structurally related aromatic compounds, both natural and human-made (Luengo *et al.* 2001; Corkery *et al.* 1994; Hartmans *et al.* 1990; Long *et al.* 1997; Olivera *et al.* 1994; Parrott *et al.* 1987; Velasco *et al.* 1998). Furthermore, the PAA pathway is not limited to a single clade of bacteria but appears to be widely distributed in both Gram positive and Gram negative bacteria (Erb *et al.* 2008; Mohamed *et al.* 2002; Navarro-Llorens *et al.* 2005; Navarro-Llorens *et al.* 2008; Olivera *et al.* 1998), making this a pathway of broad importance in terms of both carbon turnover and bacterial distribution in the biosphere.

1.3.1 The PAA Pathway is Central to the Degradation of many Aromatic Compounds

The PAA pathway is essential for the degradation of a multitude of aromatic compounds; many derivatives of phenylacetic acid are enzymatically converted to phenylacetate or phenylacetyl-CoA *via* peripheral enzymatic pathways (Luengo *et al.* 2001). The PAA pathway is then responsible for de-aromatization and ring cleavage to acyclic central metabolites succinyl-CoA and acetyl-CoA (Figure 2). Exploitation of the PAA pathway holds great potential for engineered bioremediation strategies for environmental pollutants, such as styrene and ethylbenzene (Luengo *et al.* 2001). The PAA pathway is also utilized for the metabolism of the aromatic amino acid phenylalanine. Phenylalanine can be enzymatically converted to phenylpyruvate, then phenylacetaldehyde, and finally phenylacetate prior to entering the PAA pathway (Navarro-Llorens *et al.* 2005; Law *et al.* 2008).

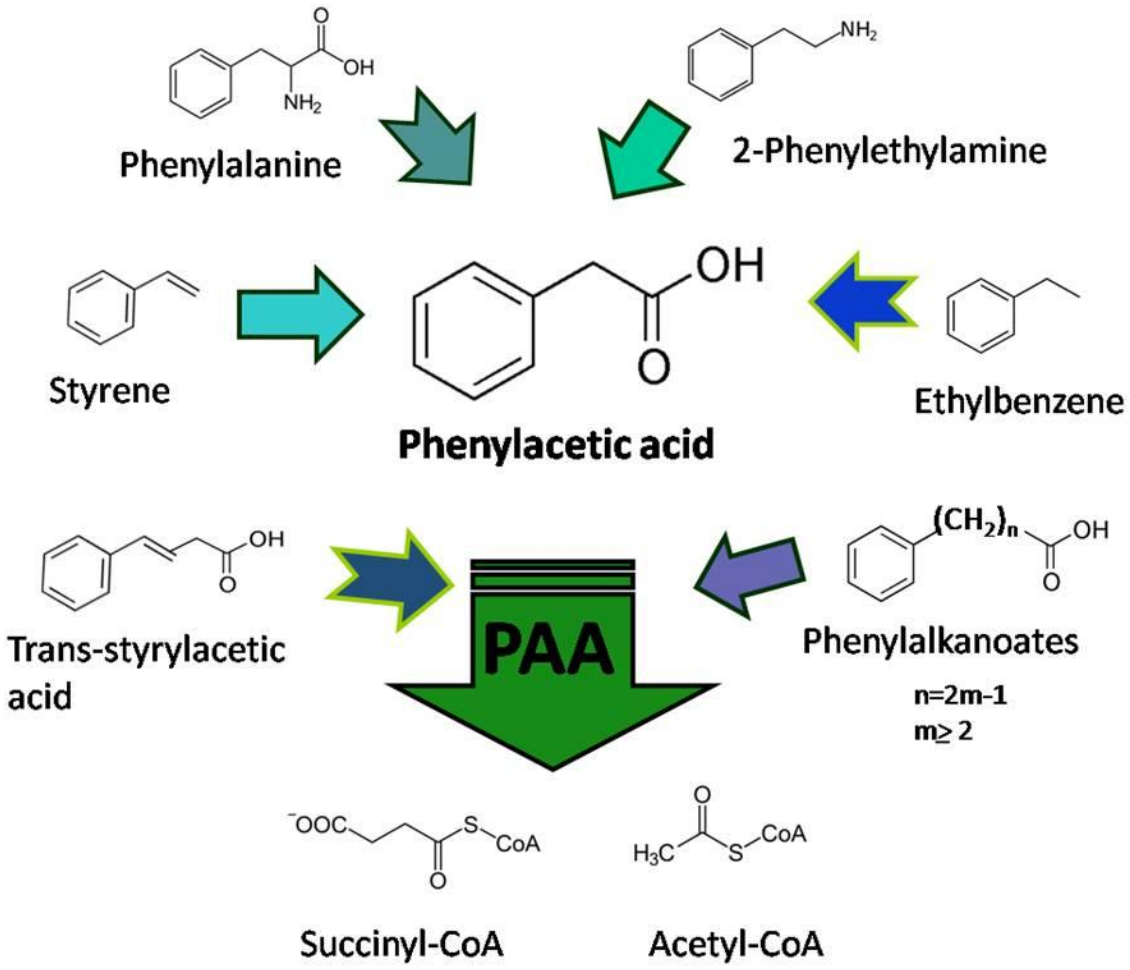


Figure 2: Phenylacetic acid: A Common Intermediate for Aromatic Degradation
Phenylalanine, as well as aromatic pollutants 2-phenylethylamine, styrene, and ethylbenzene are converted to phenylacetic acid by bacteria, prior to degradation via the PAA pathway. Trans-styrylacetic acid and phenylalkanoates undergo rounds of β-oxidation before entering the PAA pathway as the first intermediate, phenylacetyl-CoA.

1.3.2 The PAA Pathway is a Common Degradation Strategy among Bacteria

In addition to its extensive applications in aromatic compound degradation, the PAA pathway is also widely distributed among bacteria. Initially proposed in *P. putida* in 1990 (Martinez-Blanco *et al.* 1990), functional PAA pathways have since been identified in several bacterial species and strains, both Gram positive and Gram negative (Erb *et al.* 2008; Mohamed *et al.* 2002; Navarro-Llorens *et al.* 2008; Navarro-Llorens *et al.* 2005; Olivera *et al.* 1998) (Figure 3). Many of these species are environmental strains in possession of an exceptional aromatic metabolic inventory. *Rhodococcus* RHA1, which was isolated from heavily lindane contaminated soils, possesses a 9.7Mbp genome affording it an astounding ability to degrade hydrophobic pollutants, many of which are aromatic compounds (Navarro-Llorens *et al.* 2005). The impressive ability of *Pseudomonas putida* and sp. strain Y2 to degrade pollutants such as styrene and ethylbenzene relies on the capacity of the PAA pathway to perform the ring cleavage and conversion to central metabolites (Olivera *et al.* 1998; Alonso *et al.* 2003).

As more bacterial genomes are sequenced, *in silico* analyses have identified many *paa* homologs and putative PAA pathways. PAA gene organization and regulation strategies appear to differ greatly among bacteria; the genes may be clustered together (Navarro-Llorens *et al.* 2005), or distributed throughout one or more chromosomes (Erb *et al.* 2008; Law *et al.* 2008). Already, three different transcriptional regulator families have been identified in PAA gene regulation (del Peso-Santos *et al.* 2006; Hamlin *et al.* 2009; Erb *et al.* 2008). These differences in gene arrangement and regulation strategies demonstrate the versatility of the PAA pathway in terms of catabolon organization among bacteria, but make extrapolation between bacterial strains and species especially difficult.

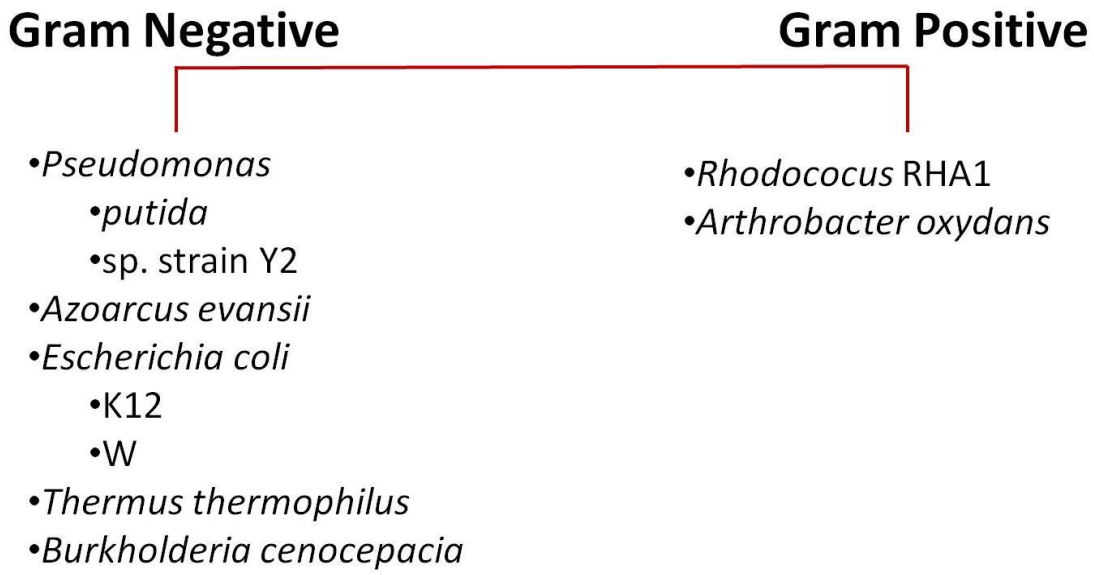


Figure 3: PAA Pathway Distribution in Bacteria
A functional PAA pathway has been identified in both Gram negative and Gram positive bacteria. Although PAA gene orthologs have since been identified in numerous bacteria, only those species or strains in which a functional pathway has been demonstrated are listed here.

1.4 PAA Pathway in *Burkholderia cenocepacia*

Among the bacteria harbouring the PAA pathway, *Burkholderia cenocepacia* is especially interesting. Its impressive 8 Mbp genome affords it the metabolic versatility to inhabit a variety of niches. In addition to its environmental role, where it colonizes water, soil, and the rhizosphere (Vermis *et al.* 2003; Mahenthiralingam *et al.* 2005; Baldwin *et al.* 2007), *B. cenocepacia* is also a significant human pathogen, causing life threatening pulmonary infections in cystic fibrosis patients (Mahenthiralingam *et al.* 2005; Mahenthiralingam *et al.* 2001). It is the most clinically important member of the *Burkholderia cepacia* complex (Bcc) (Holden *et al.* 2009), a group of closely related *Burkholderia* strains and species with impressive colonization capabilities (Mahenthiralingam *et al.* 2005; Coenye *et al.* 2001).

1.4.1 *B. cenocepacia* and The *Burkholderia cepacia* complex

The Bcc is ubiquitous in the environment, where it may impart many beneficial, or detrimental, effects wherever it colonizes (Figure 4). Many strains, even *B. cenocepacia*, have the ability to act as plant pathogens, causing diseases such as 'sour skin onion rot', while others are used as bio-control agents to prevent fungal damage to food crops (Mahenthiralingam *et al.* 2005; Parke & Gurian-Sherman 2001). The complex is also remarkable in its ability to metabolize and subsist on an astounding array of carbon sources. Environmental and clinical isolates, including *B. cenocepacia*, were able to utilize unconventional compounds such as phthalate, salicin, and most astonishingly, penicillin G as sole carbon sources (Vermis *et al.* 2003).

The metabolic versatility of *B. cenocepacia* is no doubt what allows it to colonize such a variety of niches (Vermis *et al.* 2003). The ability to successfully colonize a host clearly rests on the ability of the pathogen to efficiently extract and utilize carbon and energy sources. *B. cenocepacia* is extremely qualified in this regard with its extensive repertoire of metabolic pathways. In light of these studies, it is not so surprising that a metabolic pathway may be involved in pathogenesis of *B. cenocepacia*.

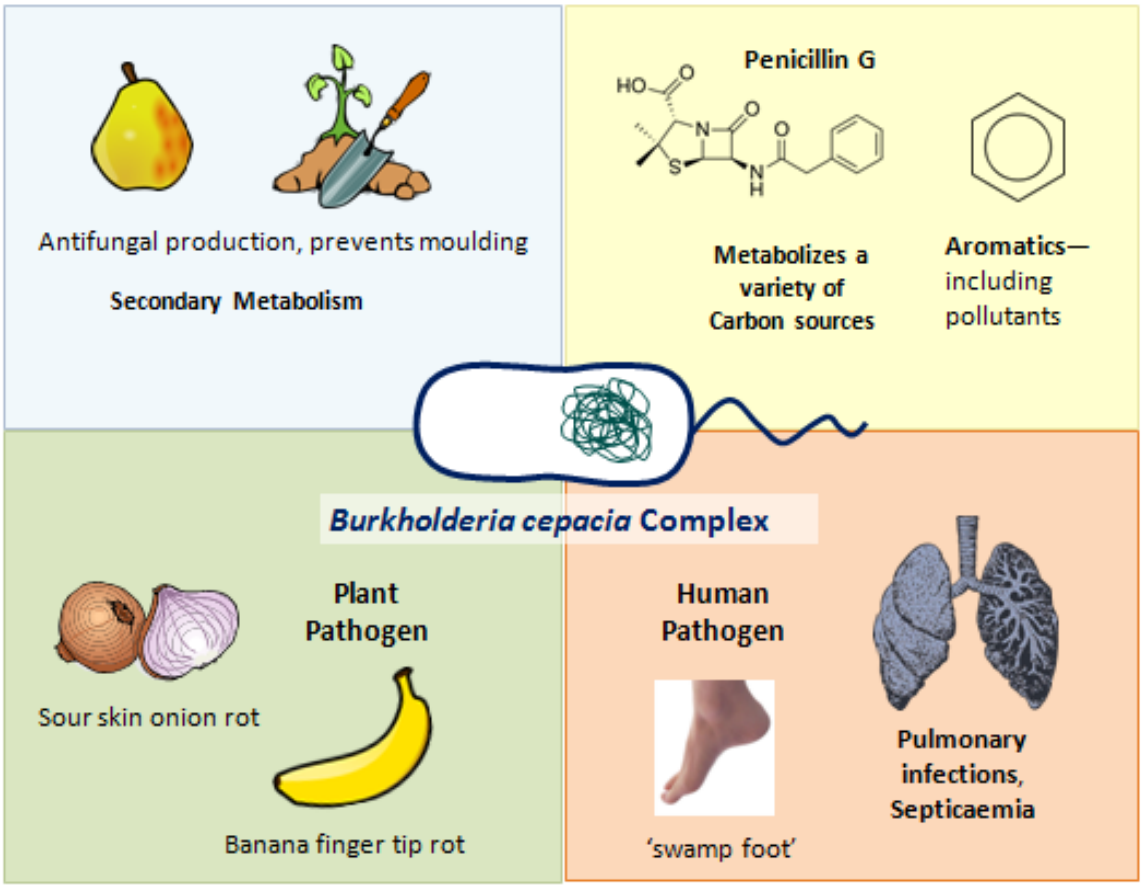


Figure 4: the metabolically versatile *Burkholderia cepacia* complex
The Bcc inhabits a wide array of niches, utilizing a variety of compounds as sole carbon sources. Where it inhabits the rhizosphere or colonizes fruit and crops, the complex may impart beneficial effects. Other members of the complex are plant and human pathogens, causing life threatening infections in immunocompromised individuals.

1.4.2 PAA Pathway and Pathogenesis in *B. cenocepacia*

An important outcome of a signature tagged mutagenesis study completed in 2004 by Hunt *et al.* was that a functional PAA pathway appears to be involved in *Burkholderia cenocepacia* pathogenesis. It was found that an insertion within *paaE,* encoding a component of the ring epoxidation enzyme of the PAA pathway, rendered the *B. cenocepacia* unable to colonize and cause infection in a rat lung infection model (Hunt *et al.* 2004).

Follow up studies undertaken by Dr. Cardona's group demonstrated a decrease in pathogenicity in the nematode, *C. elegans*, when any of the genes encoding the ring epoxidation complex were knocked out by insertional mutations (Law *et al.* 2008). Complementation in *trans* was then found to restore pathogenicity. The authors first hypothesized that the loss of pathogenicity resulted from decreased growth resulting from the inability to utilize important carbon sources such as phenylalanine and phenylacetic acid (Law *et al.* 2008). However, the mutants were found to populate the gut of *C. elegans* equally as well as the wildtype (Law *et al.* 2008). More interestingly, insertions within genes involved further along the PAA pathway rendered the bacteria more pathogenic with higher nematode kill rates (Law *et al.* 2008). This unexpected pattern of loss followed by gain of pathogenicity raises the possibility that pathogenicity may not be entirely related to nutrition. Indeed, Law *et al*. hypothesize that perhaps some pathway intermediates may also be important for signalling between pathogen and host or even among pathogens (2008). Detailed genetic, biochemical, and structural characterization of the pathway enzymes and intermediates may eventually explain these intriguing

observations as well lend some clues to as the precise role of the PAA pathway during *B. cenocepacia* infection.

1.4.3 Organization and Regulation of PAA genes in *B. cenocepacia* J2315

B. cenocepacia J2315 (hereafter *B. cenocepacia*) harbours the complete repertoire of enzymes for the phenylacetic acid pathway. The *paa* genes are arranged in three separate clusters and split between the two chromosomes. Genes encoding the ring epoxidation complex (*paaA, paaB, paaC, paaD,* and *paaE*) are clustered together on chromosome one and transcribed contiguously under control of the *paaA* promoter (Hamlin *et al.* 2009). PaaR, a TetR type transcriptional regulator responsible for the negative regulation of the PAA pathway, is encoded BCAL 0210 (now *paaR*). Phenylacetyl-CoA, the first pathway intermediate, has recently been identified as the inducer molecule for *PaaR* (unpublished, communications with Dr. Cardona). BCAL 0211 is positioned between *paaE* and *paaR*, and encodes a protein of unknown function. Its involvement, if any, in the regulation of the PAA pathway remains unclear; however, it is co-transcribed with *paaR* and co-regulated at the transcriptional level.

Further along chromosome 1, *paaF*, *paaZ*, *paaJ*, *paaG*, *paaI*, and one copy of *paaK* (*paaK1*) are located. Intriguingly, *B. cenocepacia* encodes two genes for the initial enzyme, PaaK, but only a single gene copy for all subsequent enzymes. This second cluster appears to be under control of the *paaZ* promoter (Hamlin *et al.* 2009). Finally, on chromosome 2, a second copy of *paaK* (*paaK2*) is located adjacent to *paaH*, both of which appear to be regulated through the *paaH* promoter (Hamlin *et al.* 2009).

It is interesting to note that *B. cenocepacia* encodes two genes for the initial PAA enzyme, *paaK1* and *paaK2*, but only one gene copy of all subsequent enzymes (Figure 5 & 6). This intriguing feature, a duplicate gene copy for the initial enzyme *paaK*, is not entirely unique to *B. cenocepacia*. *Pseudomonas sp*. strain Y2, as previously mentioned, contains two functional copies of the PAA pathway, but three copies of the initial enzyme (del Peso-Santos *et al.* 2006). In *Pseudomonas sp*. strain Y2, like *B. cenocepacia*, phenylacetyl-CoA, the product of PaaK, phenylacetyl-CoA, acts as the inducer molecule for the transcriptional regulator (del Peso-Santos *et al.* 2006). Fascinatingly, a single gene copy for the initial enzyme is co-regulated with the styrene degradation pathway which is responsible for the conversion of styrene to phenylacetic acid (del Peso-Santos *et al.* 2006). This allows for immediate production of phenylacetyl-CoA during styrene degradation, expediting de-repression and transcription of the remaining PAA genes, and the full conversion of styrene into citric acid cycle intermediates *via* the PAA pathway (del Peso-Santos *et al.* 2006). While this story sets an interesting precedent for complex co-ordination schemes with peripheral metabolic pathways, a biochemical rationale for two copies of *paaK* within *B. cenocepacia* remains to be determined.

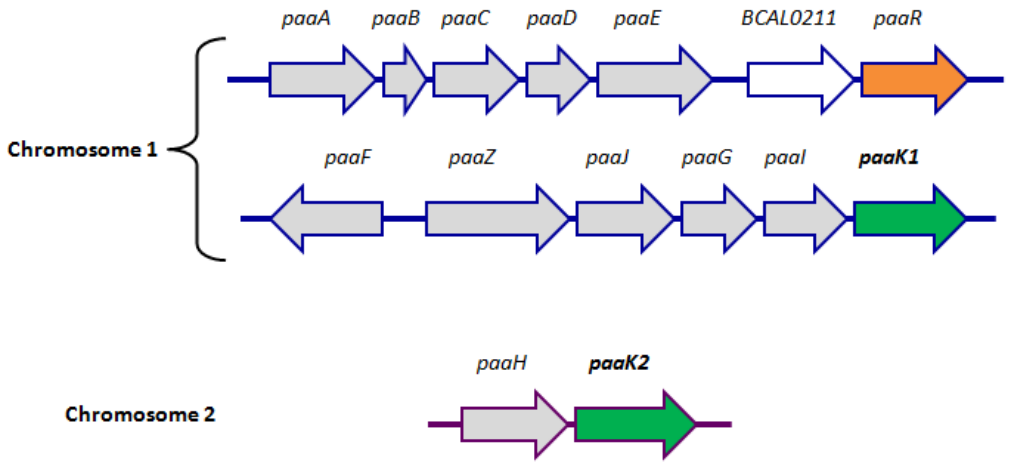


Figure 5: PAA Gene Arrangement in *B. cenocepacia* J2315
Genes encoding enzymes of the PAA pathway are split in three clusters, two of which are found on chromosome 1, the other on chromosome 2. The gene encoding the transcriptional regulator for the pathway, *paaR*, is clustered together with the genes for the ring epoxidation complex, *paaA, paaB, paaC, paaD, and paaE. paaK2*, a second gene encoding the initial pathway enzyme is found on Chromosome 2.

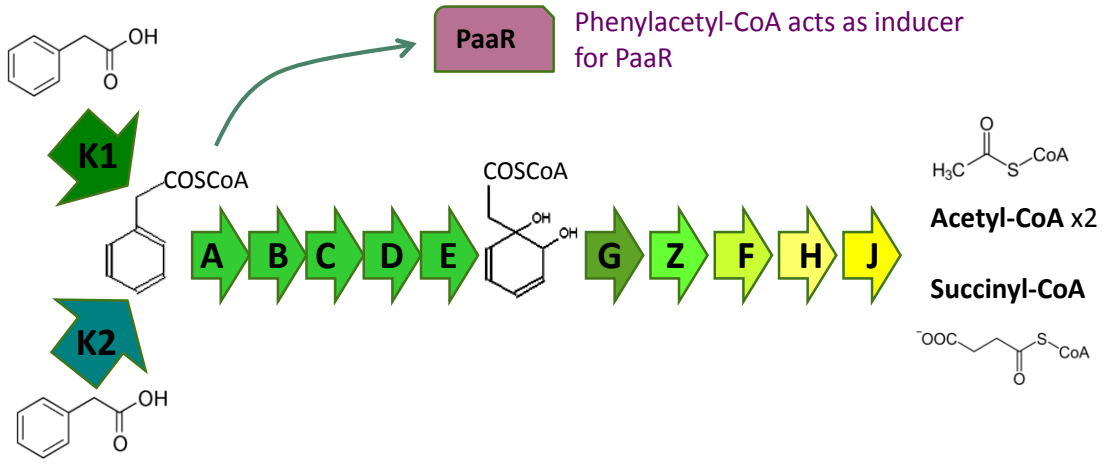


Figure 6: The PAA Pathway in *B. cenocepacia*
B. cenocepacia incorporates an additional copy of the initial enzyme, PaaK, but only a single copy of all subsequent enzymes. The product formed by PaaK1 & PaaK2, phenylacetyl-CoA acts as the inducer molecule for PaaR, a TetR type transcriptional regulator, derepressing transcription of the pathway.

1.4.4 Multi-step, Multi-Enzyme Mechanism of the PAA Pathway

B. cenocepacia incorporates all eleven genes required for a functional PAA pathway (Ismail *et al.* 2003), with an additional gene copy for the initial enzyme: phenylacetate CoA ligase (PCL) designated PaaK1 and PaaK2, respectively. While PCLs are responsible for the CoA activation of phenylacetate as the initial, committed step, common to both anaerobic and hybrid degradation pathways, subsequent steps have been subject to much speculation and uncertainty. Several putative pathway schemes have been reported, however, the latest investigation using recombinantly purified PAA enzymes conducted in 2010 is presented here (Teufel *et al*. 2010) (Figure 7). Molecular oxygen presently understood to participate in the epoxidation of the phenylacetyl-CoA ring at positions 1 and 2, catalyzed by the five-subunit PaaABCDE oxygenase (Teufel *et al*. 2010). Isomerization of the epoxide is performed by PaaG, followed by PaaZ catalyzed ring cleavage. The newly linear 8-C CoA thioester undergoes cleavage to 2,3-dehydroadipyl-CoA and acetyl CoA (Teufel *et al*. 2010). A second round of β-oxidation, catalyzed by PaaJ, PaaH and PaaF, an additional molecule of acetyl-CoA and one of succinyl-CoA, thus completing the breakdown of phenylacetic acid to common cellular metabolites (Teufel *et al*. 2010; Ismail *et al.* 2003; Nogales *et al.* 2007).

The structural and mechanistic details of many PAA enzymes are not yet well characterized. Crystal structures are only available for PaaG, the putative ring opening enzyme (Kichise *et al.* 2009); PaaB, a subunit of the five component oxygenase/reductase (PDB ID 3EGR); and PaaI, a thioesterase not required for phenylacetate degradation

(Kunishima *et al.* 2005). Structural and biochemical studies of remaining enzymes have great potential to further illuminate this novel pathway.

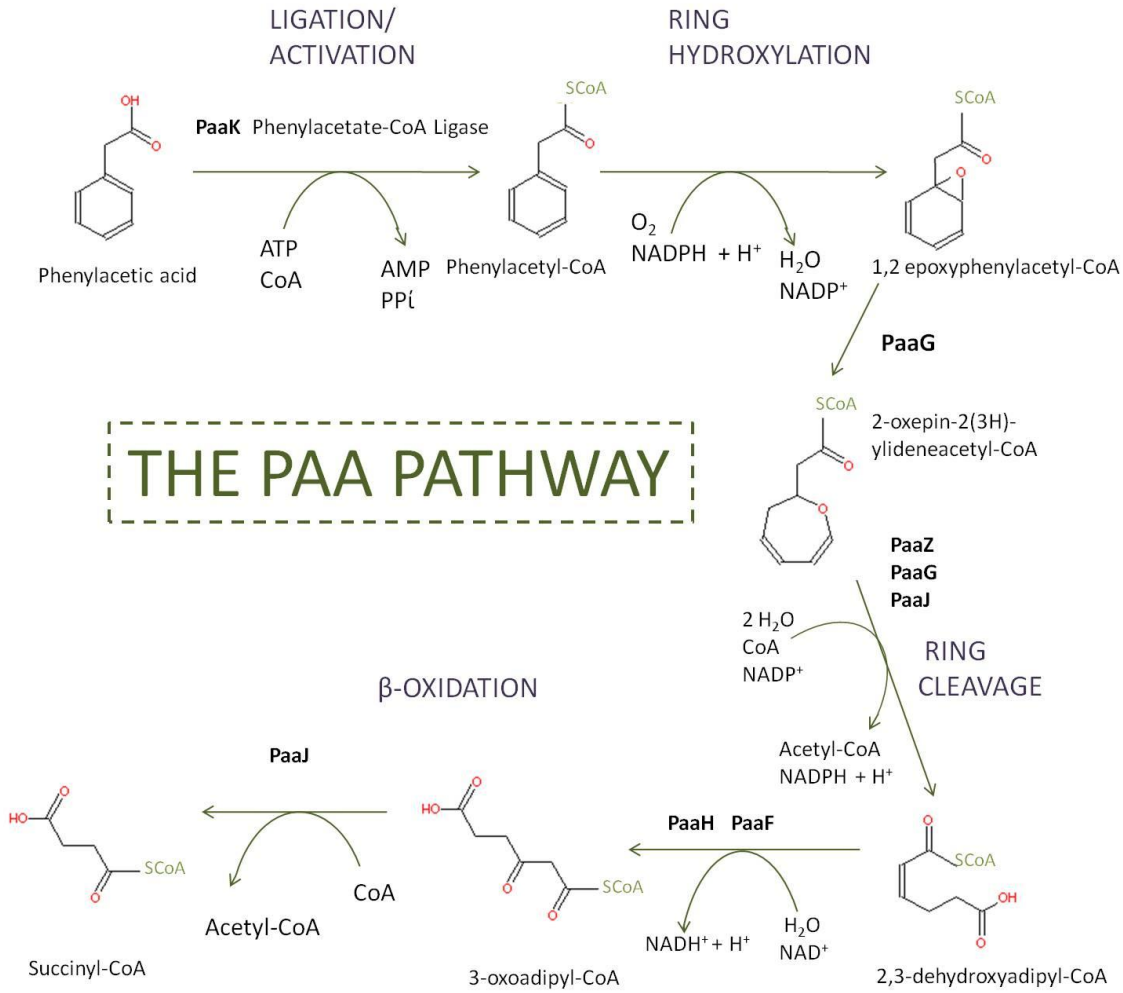


Figure 7: The Putative General Enzymatic Mechanism of the PAA Pathway
The PAA pathway is understood to require 11 separate gene products. Phenylacetic acid is first activated to Phenylacetyl-CoA by PaaK. The CoA moiety is retained for the duration of the pathway. Molecular oxygen is utilized by a multi-component oxygenase (PaaA, PaaB, PaaC, PaaD, and PaaE) for dearomatization of the aromatic nucleus. Isomerization, followed by ring cleavage and dehydrogenation are understood to be performed by PaaG and PaaZ. PaaJ, a putative thioesterase, cleaves the first of the acetyl-CoA products. β-oxidation is performed by PaaH, PaaF and PaaJ to produce the final pathway products: succinyl-CoA and a second acetyl-CoA.

1.5 PaaK: A Phenylacetate CoA Ligase

In the context of the PAA pathway, PaaK is the initial enzyme and is responsible for CoA activation of phenylacetate, thereby lowering the activation energy of subsequent reactions and imparting additional benefits as previously discussed. PaaK catalyzes a two step processes which consumes one molecule of ATP and proceeds through a phenylacetyl adenosine monophosphate intermediate, followed by thioesterification to phenylacetyl-CoA and release of AMP (Martinez-Blanco *et al.* 1990) (Figure 8). The formation of the phenylacetyl adenylate is noteworthy, as this is a distinguishing feature of the enzyme superfamily to which PaaK belongs.

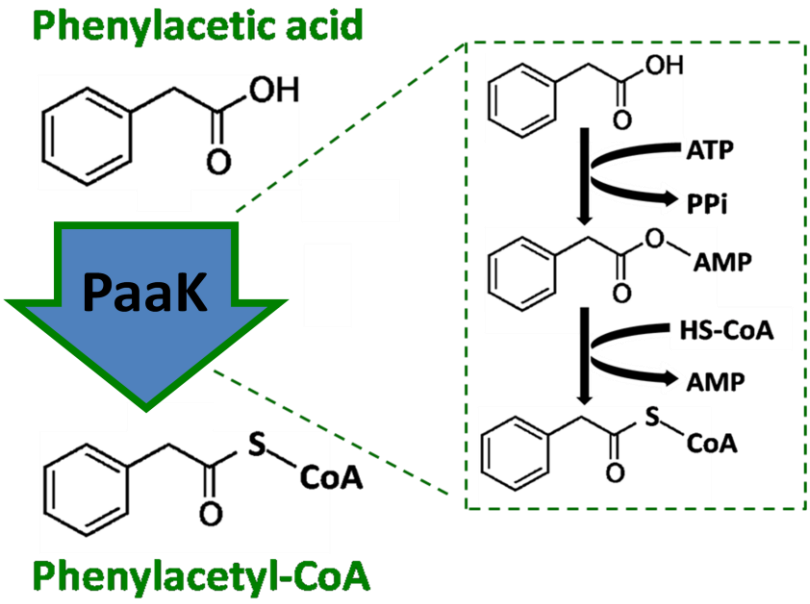


Figure 8: Coenzyme A Activation of Phenylacetic Acid by PaaK
PaaK catalyzes a two-step reaction generating a phenylacetyl adenylate intermediate. One molecule of ATP is consumed in the process; pyrophosphate is released during following formation of the intermediate, while AMP is released following the second thioesterification part reaction which utilizes one molecule of Co-Enzyme A. The final product of the overall reaction is phenylacetyl-CoA, the first intermediate of the PAA pathway.

1.5.1 PaaK: A Member of the Adenylate Forming Enzyme Superfamily

Bacterial PCLs belong to the aryl/acyl CoA ligase subfamily of the Adenylate Forming Enzyme Superfamily. This family is also comprised of two additional subfamilies, namely luciferases and the adenylation domain of non-ribosomal peptide synthetases (NRPS) (Figure 9) (McElroy *et al.* 1967). While catalyzing altogether dissimilar overall reactions, superfamily members catalyze a similar first part reaction, the adenylation of a carboxylate substrate, and differ only in the second part reaction.

Acyl and Aryl CoA ligases catalyze, overall, a thioesterification using one molecule of CoA per carboxylate substrate. Likewise, the adenylation domain of NRPS catalyzes the linkage of the substrate to a peptidyl carrier domain via a thioester linkage with the pantetheine cofactor. Luciferases are most distinct in that they catalyze an oxidative decarboxylation as the second part reaction, producing the unstable oxyluciferin which then decays to emit light. Interestingly, the presence of CoA enhances the efficiency of this reaction through conversion of an inhibitory side product, dehydroluciferyl adenylate to dehydroluciferyl-CoA (Airth *et al.* 1958; Fraga *et al.* 2005). Moreover, a study by Oba *et al.* (2003) demonstrated an additional ability of firefly luciferase to catalyze the formation of fatty acyl-CoA (Oba *et al.* 2003). Thus it appears that even the luciferase subfamily maintains the ability to catalyze a thioesterification part reaction.

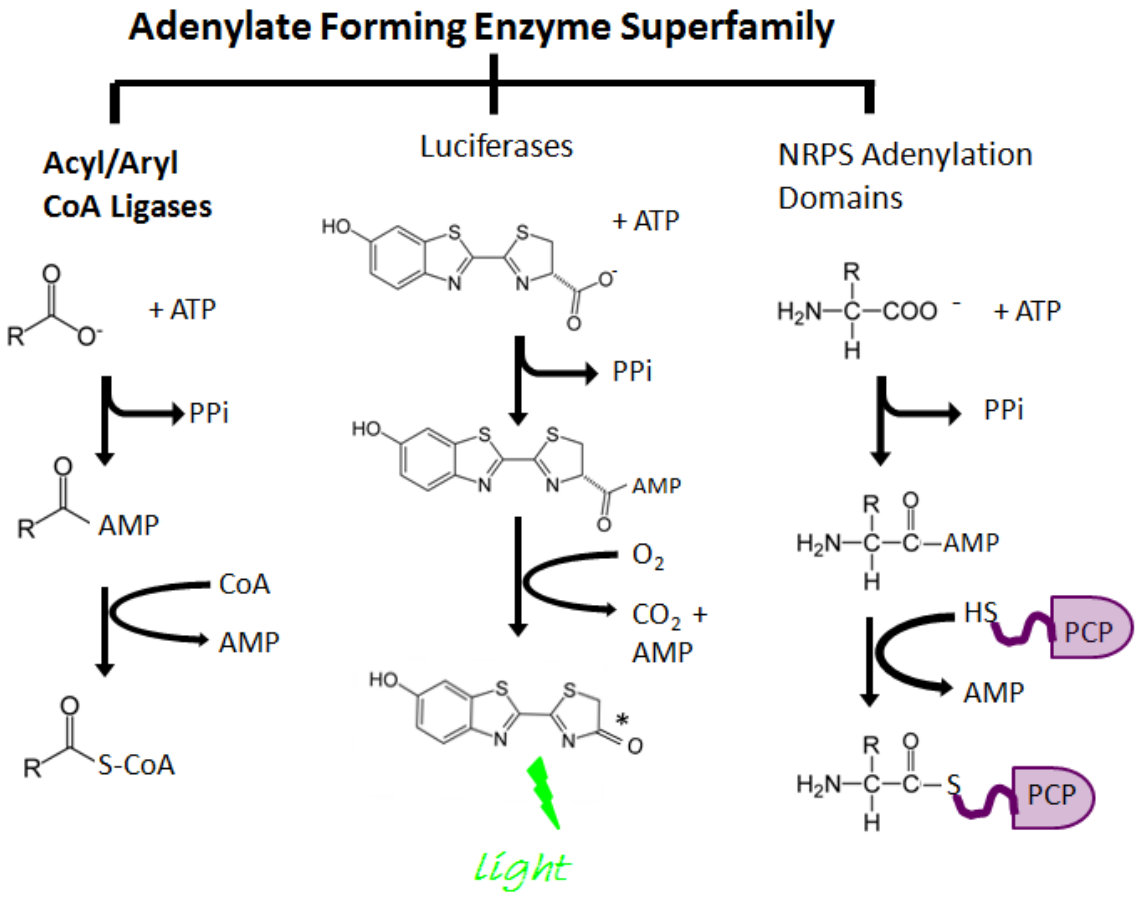


Figure 9: The Adenylate Forming Enzyme Superfamily
The superfamily is comprised of three separate subfamilies, each of which catalyzes the overall reaction in two distinct phases. The first part reaction uses one molecule of ATP to adenylate a carboxylate substrate generating an adenylate. The subfamilies differ greatly in the second part reaction: CoA ligases utilize free cofactor coenzyme A to form a thioester bond, while NRPS often use a prosthetic pantetheine group attached to a peptidyl carrier domain/protein (PCP) for the second part reaction. Oxidative decarboxylation takes place as the second part reaction in firefly luciferases utilizing molecular oxygen.

Fascinatingly, family members share only approximately 20-40% sequence identity, even within subfamilies, and yet are structurally homologous (Gulick *et al.* 2003). The enzymes are comprised of two separate α/β domains: a much larger N-terminal domain with a small C-terminal domain joined by a solvated linker. This superfamily undergoes an extreme conformational change during the complete course of the two part reaction (Gulick *et al.* 2003; Gulick *et al.* 2004). As each member catalyses a similar adenylation part reaction, a C-terminal domain lysine which is presented to the active site and essential for the first part reaction is invariably retained. Other conserved sequence motifs, such as the flexible solvated linker that joins the two domains and motifs involved in ATP binding, are also used in identifying family members (Marahiel *et al.* 1997; Chang *et al.* 1997; Babbitt *et al.* 1992).

Superfamily members have been crystallized with a variety of substrates, reaction intermediates, and product complexes or analogs, and have been captured in both the adenylation (conformation 1) and thioester (conformation 2) positions. It is unusual that there has been a distinct paucity of ATP bound structures. The crystal structures of *Bacillus cereus* D-alanyl Carrier Protein Ligase (DltA) in complex with ATP (PDB ID 3FCE, 3FCC) are the only published accounts of an ATP co-structure for this superfamily (Osman *et al.* 2009). Unfortunately, regions that are believed to form vital ATP interactions remain unmodelled and substitution of the catalytically required Mg^{2+} cation for Ca^{2+} was required in order for observable cation density (Osman *et al.* 2009). The structure of human medium chain acyl-CoA synthetase (2A ACS) (PDB ID 3C5E *to be published*) is the only available ATP co-structure to date with none of these afflictions.

1.5.2 Biochemistry of Bacterial Phenylacetate CoA Ligases

Following the discovery of the PAA pathway, biochemical studies of bacterial PCLs were undertaken. Although only a few bacterial PCLs have been purified and biochemically characterized, the orthologs have exhibited distinct differences in terms of catalytic efficiency and substrate specificity. Initial kinetic determinations for *P. putida* PCL showed a startlingly high K_M value of 16.5 mM with phenylacetic acid (Martinez-Blanco *et al.* 1990). More recent biochemical characterizations of PCLs from a number of other bacteria show much better agreement, with K_Ms ranging from 50-60 µM (Erb *et al.* 2008; Mohamed & Fuchs 1993). However, the K_M for native PCL purified from *Azoarcus evansii* was determined to be 14 µM (El-Said Mohamed 2000). The specific activity for the PCLs are also in close agreement for *Thermus thermophilus* and *P. putida*, ranging from 21-24.4 µmol/min/mg (Erb *et al.* 2008) (Martinez-Blanco *et al.* 1990; Mohamed & Fuchs 1993), however the *A. evansii* PCL again showed a vastly different activity of 0.48 µmol/min/mg (Mohamed *et al.* 2002).

Perhaps even more interesting than these kinetic parameters is the variety of additional substrates tested during the characterizations of these enzymes. While some showed high activity towards short chain acyl-acid substrates, others showed negligible activity (Erb *et al.* 2008; Martinez-Blanco *et al.* 1990). The highly purified, recombinantly produced the PCL of *T. thermophilus* was somewhat active towards 4-hydroxy-phenylacetic acid, while others have displayed negligible activity towards any substituted phenylacetic acid (Erb *et al.* 2008; Mohamed & Fuchs 1993). While these discrepancies make for an interesting case for functional diversity among bacterial PCLs, no rationale for these differences has been forthcoming. Without a high resolution structure detailing the

substrate binding site and catalytic machinery, the structural basis for these differences remains undetermined. As such, it is difficult to predict whether functional differences between the PaaK1 and PaaK2 PCL paralogs in *B. cenocepacia* are what necessitate the presence of both copies.

As the initial enzyme and only committed step of the pathway (Mohamed *et al.* 2002), the PaaKs act as gatekeepers for the entire pathway. The product of this initial reaction, phenylacetyl-CoA, has an additional role within the regulation of the pathway: phenylacetyl-CoA acts as the inducer molecule for PaaR. Only when bound to phenylacetyl-CoA will PaaR release the promoter regions and permit transcription of *Paa* genes. Potentially, the presence of two isoforms could allow for differing rates of controlled entry to the pathway based on availability of substrate, in addition to impacting the rate of transcription of the *Paa* genes.

1.6 Objectives and Hypotheses

Burkholderia cenocepacia has been shown to possess impressive metabolic capabilities, including the ability to metabolize aromatic compounds. These extensive metabolic abilities allow *B. cenocepacia* to inhabit a variety of niches, as well as cause deadly opportunistic pulmonary infections in CF patients. *B. cenocepacia* possesses a functional PAA pathway, which is understood to be involved in pathogenesis. Intriguingly, the pathway incorporates two copies of the initial enzyme PaaK, a phenylacetate CoA ligase, but only a single copy of all other PAA enzymes.

Bacterial phenylacetate CoA ligases, such as PaaK1 & PaaK2, are members of the adenylate forming enzyme superfamily. Sequence alignments and analysis reveal that it includes several of the superfamily conserved motifs, although the first ~80 residues show negligible homology to structurally characterized family members (Appendix A). Prior to this work, no structural characterization of bacterial phenylacetate CoA ligases has been undertaken. Based on the characterizations of other superfamily members, we hypothesize that:

1. Structural characterization of the PaaKs from *B. cenocepacia* J2315 will reveal that the ligases will conform to the overall superfamily fold, with a larger N-terminal domain, smaller C-terminal domain, conserved linker region, and motifs classically positioned, with gross rearrangements confined to the initial 80 residues.

Previous biochemical studies of bacterial PCL orthologs have revealed distinct differences in specific activity and K_M. *B. cenocepacia* J2315 PaaK isoforms share only 69% sequence identity and appear to be no more identical to each other than other PCL orthologs found in other phenylacetate degrading bacteria, possibly resulting in differing functional profiles. Based on these observations and the presence of paralogous phenylacetate CoA ligases in *B. cenocepacia*, we hypothesise that:

2. The PaaK1 and PaaK2 paralogs may possess differing functional profiles. Residue substitutions within the aryl substrate binding pocket may confer varying abilities to bind aromatic substrate or complete the domain reorganization process resulting in different apparent K_Ms or specific activities.

Chapter 2: Materials and Methods

2.1 Materials

B. cenocepacia J2315 genomic DNA (generously provided by Dr. Silvia Cardona); primers prepared by IDT, PCR mix (Pwo SuperYield DNA polymerase, 4 mM MgCl2, dATP, dCTP, dGTP, dTTP, each 0.4 mM) and PCR grade water (Pwo Master, Roche); agarose (EMD), ethidium bromide (BioRad), restriction enzymes (NEB), BSA (NEB); T4 DNA ligase (NEB); 10x T4 DNA ligase buffer (NEB); LB broth (EMD), Competent cells (Novagen); QIAprep® Spin Miniprep Kit; QIAquick ® Gel Extraction Kit (Qiagen), QIAquick® PCR Purification Kit (Qiagen), pET28a(+) vector (Novagen); antibiotics: kanamycin (SIGMA), ampicillin (Fisher Bioreagents); Hepes (Sigma), Sodium Chloride (SIGMA); Imidazole (SIGMA); filters (Millipore); Centricon spin concentrators (Millipore); thrombin (Invitrogen), Calcium Chloride (SIGMA) ; WIZARD I/II (Emerald BioSystems), 96-well crystallization plates (Emerald Biosystems), KSCN (SIGMA); ATP (Calbiochem), $MgCl_2$ (SIGMA), Phenylacetic acid (SIGMA), NADH (Calbiochem), Myokinase (Calbiochem), Lactate Dehydrogenase (SIGMA), Pyruvate Kinase (SIGMA), Phosphoenolpyruvate (SIGMA), Coenzyme A (Calbiochem).

2.2 General Methods

2.2.1 DNA manipulation

The *paaK1* gene was amplified from genomic DNA using a PCR reaction that contained 500 n*M* of each forward and reverse primer, 20 ng of template DNA, 22 µl of PCR grade water and 25 µl of PCR mix (Pwo SuperYield DNA polymerase, 4 mM $MgCl_2$, dATP, dCTP, dGTP, dTTP, each 0.4 mM). PCR product was visualized on a 1% (w/v) agarose

gel with ethidium bromide (EtBr) using the EagleEye II system. A second DNA product was amplified along with *paaK1* necessitating gel extraction and purification of the major PCR product with QIAquick ® Gel Extraction Kit (Qiagen).

Restriction digestion

Purified insert and pET28a(+) vector were digested with appropriate enzymes (2 U of enzyme per 1 μg of DNA) in 10x NEBuffer supplemented with 10X bovine serum albumin (BSA). The digestion reaction was incubated for at 37 ° C 2 hours (hr), and in the case of digested vector, was followed by treatment with Antarctic Phosphatase to remove 5'phosphates and prevent recircularization. Five U Antarctic Phosphatase was used per 1 μg of the vector DNA in the presence of 10X Antarctic Phosphatase Buffer and incubated 30 minutes (min) at 37° C followed by heat inactivation of the enzyme at 65° C for 5 min. Digested insert and vector DNA were then purified with QIAquick® PCR Purification Kit according to the manufacturer's instructions.

Ligation

A ligation reaction as carried out overnight at 16° C with the digested and de-phosphorylated pET28a(+) vector (100 ng) for each digested insert (three molar excess of insert:vector) using 10 U of T4 DNA ligase and 10x T4 DNA ligase buffer in the final volume of 20 μl.

Transformation and Isolation of Constructs

E. coli DH5α competent cells (50 μl) were transformed with the ligation mixtures as follows: cells were thawed on ice for 10min and incubated with 5 μl of the ligation mixture on ice for 30 min. This incubation was followed by a heat shock at 42° C for 45

seconds (sec) and incubation on ice for an additional 2 min. For the recovery period, LB broth (500 µl) was added to the transformation mixture and incubated shaking at 37° C for 45 min with shaking. 100 µl of the bacterial suspension was spread plated on LB agar plates with selective antibiotic, followed by incubation at 37°C overnight. Resulting colonies were screened by growing bacterial cultures in 5ml LB media with appropriate antibiotic overnight with shaking at 200 rpm at 37° C. Plasmids were then isolated using the QIAprep® Spin Miniprep Kit and screened by digestion with appropriate enzymes with visual inspection on a 1% (w/v) agarose gel with EtBr and using the EagleEye II system. Plasmids which contained appropriate size inserts were then sent for sequencing to the DNA sequencing facility at the Center for Biomedical Research (University of Victoria, BC). DNA sequence data was analysed using BLAST server (Sayers *et al.* 2008). The correct plasmid constructs were then used for transformation of *E. coli* BL21 (DE3).

2.3 Cloning of *paaK* Paralogs

2.3.1 PaaK1

The *paaK1* gene was amplified using the forward primer 5′ CTACGCTAGCACTACCCCGCTACCGCTC 3’ and the reverse primer 5′ TCAGAAGCTTTTATCAGCCCTTGCGCTTGTCG. The amplified DNA fragments were digested with *Nhe*I and *Hin*dIII, and subcloned into pET-28a(+) (Novagen, Mississauga, ON, Canada) in-frame with an N-terminal hexa-histidine tag to facilitate purification.

2.3.2 PaaK2

A synthesized gene, synonymous with *paaK2* of *B. cenocepacia* J2315, was purchased from GenScript. This synthetic gene had been codon optimized for expression in *E. coli* BL21 but most importantly, alternate choices of codons decreased the CG content of the gene, facilitating subsequent manipulations (Figure 10). The codon optimized *paaK2* gene was provided cloned into the pUC57 vector between *Nde*I and *Hin*dIII sites. *E. coli* DH5α competent cells were transformed with the *paaK2*-pUC57 plasmid and plated on selected media. Plasmids were isolated from overnight cultures of resulting colonies, as described above. The *paaK2* insert was isolated by restriction digestion with *Nde*I and *Hin*dIII, followed gel extraction and purification with QIAquick ® Gel Extraction Kit (Qiagen). The *paaK2* gene was then inserted into pET-28a(+) in frame with the hexa-histidine tag using *Nde*I and *Hin*dIII restriction sites. Construct sequencing confirmed that no mutations were introduced during amplification or subcloning.

A

```
ATGACTCACCCGACGCATCCCGCCGCCGCCCTCGAGCCGATCGAGACCGCCAGCCGCGACGAACTGCAGGCGCTGCA
GCTCGAACGCCTCAAGTGGTCGCTGCGCCACGCGTACGACAACGTCCCGCACTATCGCCGCACGTTCGACGCGGCCG
GCGTGCATCCGGACGACCTGAAATCGCTCGCCGATCTCGCGAAGTTCCCGTTCTCGACCAAGAACGACCTGCGCGAC
AACTATCCGTTCGGCCTCTTCGCGGTGCCGCGCGAGCAGGTCGTACGCGTGCATGCGTCGAGCGGCACGACCGGCAA
GCCGACCGTGGTCGGCTACACCGCGCGCGACATCGACACGTGGGCGAACGTGACCGCGCGCTCGATCCGCGCGGCCG
GCGGCCGCCCGGGCGATACGCTGCACAACGCGTTCGGCTACGGCCTCTTCACCGGCGGCCTCGGGATTCACTACGGC
GCGGAGCGGCTCGGCTGCATGGTCGTGCCGATGTCGGGTGGCCAGACCGAGAAGCAGGTGCAACTGATTCGCGACTT
CGAGCCGAAGATCATCCTCGTCACGCCGTCGTACATGCTGAACCTGATCGACGAGATGGTGCGGCAGGGGATGGACC
CGGCCGAATCGTCGCTGAAGATCGGCATCTTCGGCGCCGAGCCGTGGACGCAGGCGCTGCGCAATGAAGTGGAGACG
CGCGTGGGCATCGACGCGCTCGACATCTACGGGCTGTCGGAAGTGATGGGCCCGGGCGTCGCGTGCGAATGCGTCGA
GACGAAGGACGGCCCGGTGATCTGGGAAGACCATTTCTACCCGGAGATCATCGATCCCGTCACCGGCGAGGTGCTGC
CCGACGGCAGCCAGGGCGAGCTCGTGTTCACGTCGCTGACGAAGGAGGCGATGCCGGTGATCCGCTACCGCACGCGC
GACCTCACCGCGCTGCTGCCGCCGACCGCGCGCGCGATGCGCCGGCTCGCGAAGATCACGGGCCGCTCCGACGACAT
GCTGATCGTGCGCGGCGTGAACGTGTTTCCGAGCCAGATCGAGGAGATCGTCGTCGCGCTGCCGCTGTTGTCGGGCC
AGTTCCAGATCACGCTGTCGCGCGACGGTCACATGGACCGGCTCGACCTCGCGGTCGAGTTGCGCTCCGAGGCGGCG
GCGTCCGTCACCGACGGCGAGCGCGCGGCGCTCGCGCGCGAGTTGCAGCACCGGATCAAGACGATGGTCGGCGTGTC
GTCCGGCGTGACGGTGCTGGCAGCCGGCGGCATTCCGGCGACCGCGACCGGCAAGGCGCGCCGCGTGATCGATCGCC
GTCAGGCCGCCTGA
```

B

```
ATGACCCATCCGACGCACCCGGCAGCAGCACTGGAACCGATTGAAACCGCATCTCGTGATGAACTGCAGGCGCTGCA
GCTGGAACGTCTGAAATGGAGTCTGCGCCATGCCTATGATAACGTTCCGCACTACCGTCGCACCTTTGATGCGGCCG
GCGTGCATCCGGATGATCTGAAAAGCCTGGCAGATCTGGCGAAATTTCCGTTCTCTACCAAAAACGATCTGCGCGAT
AATTATCCGTTTGGTCTGTTCGCCGTTCCGCGCGAACAGGTGGTTCGTGTGCACGCAAGCTCTGGCACCACGGGTAA
ACCGACCGTGGTTGGCTACACGGCGCGTGATATTGATACCTGGGCCAACGTGACGGCACGTAGCATCCGTGCAGCAG
GTGGTCGTCCGGGTGATACCCTGCATAATGCATTTGGCTATGGTCTGTTCACGGGCGGTCTGGGCATTCACTACGGT
GCGGAACGTCTGGGCTGCATGGTGGTTCCGATGAGCGGCGGTCAGACCGAAAAACAGGTTCAGCTGATTCGCGATTT
TGAACCGAAAATTATCCTGGTGACGCCGTCTTATATGCTGAACCTGATCGATGAAATGGTTCGTCAGGGTATGGATC
CGGCGGAAAGTAGCCTGAAAATTGGCATCTTCGGTGCAGAACCGTGGACCCAGGCCCTGCGTAATGAAGTGGAAACG
CGCGTTGGCATTGATGCGCTGGATATCTATGGTCTGAGCGAAGTGATGGGCCCGGGTGTTGCCTGCGAATGTGTGGA
AACCAAAGATGGCCCGGTGATTTGGGAAGATCATTTTTACCCGGAAATTATCGATCCGGTGACCGGTGAAGTTCTGC
CGGATGGCAGTCAGGGTGAACTGGTTTTCACCAGCCTGACGAAAGAAGCGATGCCGGTGATCCGTTACCGCACCCGT
GATCTGACGGCACTGCTGCCGCCGACCGCACGTGCAATGCGTCGCCTGGCAAAAATTACGGGCCGTAGTGATGATAT
GCTGATCGTTCGCGGTGTGAATGTTTTTCCGAGCCAGATTGAAGAAATCGTGGTTGCCCTGCCGCTGCTGTCTGGCC
AGTTCCAGATTACCCTGAGTCGTGATGGTCACATGGATCGCCTGGATCTGGCCGTTGAACTGCGCAGCGAAGCCGCA
GCGTCTGTGACCGATGGTGAACGTGCAGCACTGGCACGTGAACTGCAGCACCGCATCAAAACCATGGTGGGCGTTTC
TAGTGGTGTGACGGTTCTGGCAGCAGGTGGTATTCCGGCAACCGCAACGGGTAAAGCACGTCGTGTGATCGATCGTC
GCCAGGCAGCGTAA
```

Figure 10: Synonymous Genes for PaaK2

GC content is highlighted
A: Native *paaK2* contains 69% GC
B: Codon Optimized *paaK2* with a reduction to 56% GC

2.4 Protein Expression and Purification

2.4.1 PaaK1

Expression of recombinant *paaK1* was performed in *Escherichia coli* BL21 Star (DE3) cells (Invitrogen) grown in Overnight Express Instant TB media (EMD chemicals) supplemented with 50 µg ml^{-1} kanamycin (Sigma). After 22 hours, cells were harvested by centrifugation at 8000 rev min^{-1} for 15 min and the pellet resuspended in Buffer A (20 mM Hepes pH 8.3, 15 mM imidazole, 3 mM β-mercaptoethanol, 1 M NaCl, 1% glycerol). The cells were lysed using a French press (SLM-Instruments) and the insoluble fraction was removed by centrifugation at 16000 rev min^{-1} for 45 min. The supernatant containing the soluble protein was diluted to approximately 150mL and sequentially filtered through 5 µm, 1 µm and 0.45 µm filters (Millipore) to remove cellular debris. The sample was loaded onto an FPLC HisTrapFF Ni affinity column (GE Healthcare). Fractions were eluted with a concentration gradient of Buffer B (20 mM HEPES pH 8.3, 1 M NaCl, 500 mM imidazole, 3 mM β-mercaptoethanol, 10% glycerol) and analyzed by SDS-PAGE. Fractions containing PaaK1 were pooled, concentrated using Centricon (Millipore) spin concentrators, thrombin digested and centrifuged at 16,000 x g, prior to manual injection onto an FPLC Superdex 200 Hi-load 16/60 size-exclusion column (GE Healthcare) in Buffer C (20 mM HEPES pH 8.3, 200 mM NaCl, 3 mM β-mercaptoethanol, 5% glycerol). Fractions were then pooled based on purity, buffer exchanged into buffer D (20 mM HEPES pH 8.3, 10 mM NaCl, 3 mM β-mercaptoethanol, 2% glycerol), manually loaded onto Source 30Q anion exchange column (GE Healthcare) and eluted with a concentration gradient of buffer E (20 mM HEPES pH 8.3, 500 mM NaCl, 3 mM β-mercaptoethanol, 2% glycerol).

Expression and purification of seleno-methionine PaaK1 was carried out as described for the native PaaK1 with the following changes: the expression vector was transformed into *Escherichia coli* 834 (DE3) (a methionine auxotroph; Novagen) and grown in SelenoMet medium (AthenaES) supplemented with L-seleno-methionine to a final concentration of 40 µg ml^{-1} (AthenaES).

2.4.2 PaaK2

Expression of native PaaK2 was carried out as described for PaaK1, with the following changes to the purification procedure: buffers A, B, and C contained 10% glycerol, and PaaK2 was stored with 3 mM ATP and $MgCl_2$ during thrombin digestion and between purification steps to improve stability.

2.5 Crystallization, Data collection and Structure Solution

2.5.1 PaaK1-ATP Complex

Purified PaaK1 was concentrated to 12 mg ml^{-1} protein, supplemented with 3 mM $MgCl_2$ and ATP, and crystallized using the sitting-drop, vapour diffusion method. Initial crystals of native PaaK1 with 3 mM ATP and $MgCl_2$ were observed in 20% PEG 3350 with 200 mM NaSCN after 1 week at 293K. Optimizations and streak seeding yielded seleno-methionine PaaK1 crystals in 20-25% PEG 3350, 200 mM KSCN, with 5% glycerol to facilitate cryopreservation. A single seleno-methionine PaaK1 crystal measuring 0.2 mm x 0.1 mm x 0.02 mm was looped and stepped into cryo-protectant consisting of reservoir solution supplemented with 3 mM $MgCl_2$, 3 mM ATP, and 10, 15 and 25% glycerol for 30 seconds and flash cooled directly in the cryostream (100K).

Diffraction data were collected on beam line 9-2 at Stanford Synchrotron Radiation Lightsource (SSRL) at 0.9792 Å. Each of the 450 images was collected over an oscillation range of 0.8˚. Diffraction data were processed using iMosflm (Leslie 1992), and scaled and merged to 1.60 Å in Scala (*The CCP4 suite: programs for protein crystallography* 1994; Evans 2006) (Data collection statistics are presented in Table 1) with 5% of reflections set aside for R_{free}. ShelxC/D/E initially searched for 11 heavy atom sites (selenomethionine) and was able to identify 15 sites and determine phases (*The CCP4 suite: programs for protein crystallography* 1994; Sheldrick 2008). Phases of sufficient quality were obtained for automated building using Buccaneer (Cowtan 2006; *The CCP4 suite: programs for protein crystallography* 1994), the remainder of the PaaK1 structure was manually built using COOT (Emsley & Cowtan 2004). ATP and Mg^{2+} coordinating 4 waters were modelled to fit density within the active site. Additional water molecules were automatically added using COOT and visually edited (Emsley & Cowtan 2004). Several molecules of glycerol, β-mercaptoethanol, thiocyanate, and a fragment of a PEG molecule could also be manually modelled. The PaaK1-ATP structure was refined using REFMAC with 5% of reflections set aside for calculation of R_{free} (Murshudov *et al.* 1997) to an R_{cryst} of 16.6% and R_{free} of 19.3%. Stereo-chemical analysis of the refined structure was performed with PROCHECK and SFCHECK in CCP4, with the Ramachandran plot showing excellent stereochemistry with 92.7% of the residues in the favoured conformations and no residues modeled in disallowed orientations (Laskowski *et al.* 1993; Vaguine *et al.* 1999; *The CCP4 suite: programs for protein crystallography* 1994). .

Table 1: Data collection and refinement statistics PaaK1-ATP co-structure

A. Data collection	
Spacegroup	P1
a, b, c (Å)	56.71, 62.47, 78.49
α, β, γ (deg.)	90.97, 109.81, 106.51
Wavelength (Å)	0.9792
Resolution (Å)	40.00-1.60
Measured reflections	482646 (69360)
Unique reflections	121839 (17505)
Redundancy	4.0 (4.0)
Completeness (%)	95.8 (94.1)
$I/\sigma(I)$	8.6 (2.2)
R_{merge} [a]	0.084(0.497)
B. Refinement Statistics	
Resolution range (Å)	29.98-1.60 (1.64-1.60)
R_{cryst} [b]	0.166 (0.302)
R_{free} [c]	0.193 (0.331)
No. of atoms	
Protein (chain A,B)	3344, 3366
Solvent	952
ATP	62
Glycerol	30
Thiocyanate	27
β-mercaptoethanol	20
PEG	12
Mg	6
B-values (Å^2)	
Protein (chain A,B,)	21.33, 18.55
Solvent	34.34
ATP	14.14
Glycerol	34.20
Thiocyanate	44.56
β-mercaptoethanol	55.46
PEG	45.17
Mg	19.78
r.m.s. deviation from ideality	
Bond lengths (Å)	0.009
Bond angles (deg.)	1.199
Ramachandran Statistics	
Most favored	92.7%
Allowed	7.1%
Disallowed	0.0%

Values in parentheses are for the highest resolution shell

[a] $R_{merge} = \sum_{hkl} \sum_i |I_{hkl,i} - [I_{hkl}]| / \sum_{hkl} \sum_i I_{hkl,i}$ where $[I_{hkl}]$ is the is the average of symmetry related observations of a unique reflection

[b] $R_{cryst} = \sum |F_{obs} - F_{calc} / \sum F_{fobs}$, where F_{obs} and F_{calc} are the observed and the calculated structure factors, respectively.

[c] R_{free} is R using 5% of reflections randomly chosen and omitted from refinement

2.5.2 PaaK1-Phenylacetyl Adenylate Complex

PaaK1 was concentrated to 12 mg ml^{-1} protein and incubated for 1 hour with 3 mM $MgCl_2$, ATP, and 5 mM phenylacetic acid prior to crystallization trials using the sitting-drop, vapour diffusion method. Diffraction quality crystals developed over the course of one month but were too large to mount individually and had to be split into several smaller crystals. The condition contained 10% (w/v) Peg 8000, 0.1M Hepes pH 7.5, and 8.0% (v/v) ethylene glycol. With small additions of ethylene glycol to the mother liquor, followed by dehydration, a cryo-protectant solution was obtained. A single crystal was mounted and flash cooled directly in the cryostream (100K). Diffraction data were collected on Rigaku R-axis IV++ area Detector with 0.5˚oscillation for each of the 720 images. Diffraction data to 1.92 Å were processed in iMosflm (Leslie 1992), followed by programs within the CCP4i suite (*The CCP4 suite: programs for protein crystallography* 1994) (Data collection statistics are presented in Table 2).

Structure solution was achieved through molecular replacement in MOLREP using one monomer of the previously determined PaaK1 structure with ligands removed (Vagin & Teplyakov 1997). The structure was manually edited in COOT where mobile regions had been repositioned or disordered in the phenylacetyl adenylate co-structure (Emsley & Cowtan 2004). The phenylacetyl adenylate ligand and accompanying library file (.cif) was generated using the Dundee Prodrg2 Server (Schuttelkopf & van Aalten 2004). As for the previous structure, the PaaK1-phenylacetyl adenylate structure was refined using REFMAC with 5% of reflections set aside for calculation of R_{free} (Murshudov *et al.* 1997). An R_{cryst} of 18.4% and R_{free} of 23.2% were eventually obtained (Murshudov *et al.* 1997).

Stereo-chemical analysis of the refined structure was performed with PROCHECK and SFCHECK in CCP4, with the Ramachandran plot showing excellent stereochemistry with 91.5% of the residues in the favoured conformations and no residues modeled in disallowed orientations (*The CCP4 suite: programs for protein crystallography* 1994; Laskowski *et al.* 1993; Vaguine *et al.* 1999).

Table 2: Data collection and refinement statistics for PaaK1 Phenylacetyl adenylate co-structure

A. Data collection	
Spacegroup	P1
a, b, c (Å)	57.26, 62.38, 76.60
α, β, γ (deg.)	91.14, 108.35, 106.28
Wavelength (Å)	1.5418
Resolution (Å)	27.47-1.92
Measured reflections	262037 (36580)
Unique reflections	68731 (9729)
Redundancy	3.8 (3.8)
Completeness (%)	93.8 (90.8)
I/σ(I)	15.2 (2.7)
R_{merge} [a]	0.056(0.444)
B. Refinement Statistics	
Resolution range (Å)	27.01-1.92 (1.97-1.92)
R_{cryst} [b]	0.1835 (0.325)
R_{free} [c]	0.2318 (0.432)
No. of atoms	
Protein (chain A,B)	3315, 3322
Solvent	665
Phenylacetyl Adenylate	64
Phenylacetate	10
β-mercaptoethanol	12
PEG	17
Mg	4
B-values (Å^2)	
Protein (chain A,B,)	26.50, 32.7
Solvent	38.51
Phenylacetyl Adenylate	23.57
Phenylacetate	62.00
β-mercaptoethanol	42.67
PEG	50.72
Mg	38.22
r.m.s. deviation from ideality	
Bond lengths (Å)	0.006
Bond angles (deg.)	0.861
Ramachandran Statistics	
Most favored	91.5%
Allowed	8.1%
Disallowed	0.0%

Values in parentheses are for the highest resolution shell

[a] $R_{merge}= \sum_{hkl} \sum_i |I_{hkl,i} - [I_{hkl}]| / \sum_{hkl} \sum_i I_{hkl,i}$ where $[I_{hkl}]$ is the is the average of symmetry related observations of a unique reflection

[b] $R_{cryst}=\sum|F_{obs}-F_{calc}/\sum F_{fobs}$, where F_{obs} and F_{calc} are the observed and the calculated structure factors, respectively.

[c] R_{free} is R using 5% of reflections randomly chosen and omitted from refinement

2.5.3 PaaK2-Phenylacetyl Adenylate Complex

PaaK2 was concentrated to 12 mg ml^{-1} protein and incubated for 1 hour with 3 mM $MgCl_2$, ATP, and 5 mM phenylacetic acid prior to crystallization trials using the sitting-drop, vapour diffusion method. Thin, poor quality crystals were obtained in several conditions, however, only optimizations using 17% (w/v) Peg 6000, 0.1M NaHepes pH 7.5, 0.1M KCl, with 2.5% glycerol eventually yielded diffraction quality crystals. A single crystal was looped and stepped slowly into cryo-protectant solution consisting of reservoir solution supplemented with 3 mM $MgCl_2$, 3 mM ATP, 5 mM phenylacetic acid and 10 and 15% glycerol directly cooled on the cryo stream (100K).

Two hundred images each with an oscillation range of 1.0˚ were collected on beam line 9-2 at SSRL. Diffraction data were processed using iMosflm (Leslie 1992), and scaled and merged to 1.90 Å (*The CCP4 suite: programs for protein crystallography* 1994) (Data collection statistics are presented in Table 3). Structure solution was achieved through molecular replacement using one monomer of the previously determined PaaK1 structure with ligands removed. Phases of sufficient quality were obtained for automated building of the majority of the N-terminal domain by Buccaneer, which initially built 708 residues with 661 residues sequenced (Cowtan 2006). The majority of the C-terminal domain required manual building in COOT as initial calculated density for this region was too poor for automated building (Emsley & Cowtan 2004).

The phenylacetyl adenylate intermediate was again modelled into the active site. Subsequent rounds of building and refinement using REFMAC with 5% of reflections set

aside for R_{free} resulted in the PaaK2-adenylate co-structure structure with R_{cryst} 17% of and R_{free} of 21.5% (Murshudov *et al.* 1997). The structure possesses excellent stereochemistry with 92% of all residues in the favoured region of the Ramachandran plot and no residues occupying disallowed regions.

Table 3: Data collection and refinement statistics PaaK2-Phenylacetyl adenylate co-structure

A. Data collection	
Spacegroup	$P2_1$
a, b, c (Å)	69.15, 81.97, 80.89
α, β, γ (deg.)	90.00, 97.27, 90.00
Wavelength (Å)	0.97884
Resolution (Å)	30.00-1.90 (2.0-1.9)
Measured reflections	283830 (31225)
Unique reflections	69914 (9527)
Redundancy	4.1 (3.1)
Completeness (%)	96.0 (92.9)
I/σ(I)	9.2 (3.1)
R_{merge} [a]	0.094 (0.327)
B. Refinement Statistics	
Resolution range (Å)	40.99- 1.90 (1.949-1.90)
R_{cryst} [b]	0.170 (0.236)
R_{free} [c]	0.215 (0.295)
No. of atoms	
Protein (chain A,B)	3339, 3327
Solvent	860
Phenylacetyl Adenylate	64
β-mercaptoethanol	4
PEG	32
Mg	2
B-values (Å^2)	
Protein (chain A,B,)	19.56, 19.40
Solvent	31.06
Phenylacetyl Adenylate	12.14
β-mercaptoethanol	41.84
PEG	48.35
Mg	22.62
r.m.s. deviation from ideality	
Bond lengths (Å)	0.006
Bond angles (deg.)	0.973
Ramachandran Statistics	
Most favored	92.0%
Allowed	7.5%
Disallowed	0.0%

Values in parentheses are for the highest resolution shell

[a] $R_{merge}= \sum_{hkl} \sum_i |I_{hkl,i} - [I_{hkl}]| / \sum_{hkl} \sum_i I_{hkl,i}$ where $[I_{hkl}]$ is the is the average of symmetry related observations of a unique reflection

[b] $R_{cryst}=\sum|F_{obs}-F_{calc}/\sum F_{fobs}$, where F_{obs} and F_{calc} are the observed and the calculated structure factors, respectively.

[c] R_{free} is R using 5% of reflections randomly chosen and omitted from refinement

2.6 Kinetic Determinations for PaaK1 and PaaK2

Enzyme activity measurements for PaaK1 and PaaK2 were conducted using the indirect spectrophotometric assay described by Zeigler *et al*. (1989) adapted to a 96 well plate (Ziegler *et al*. 1989; Bains & Boulanger 2007). Briefly, the assay used three additional enzymes (Myokinase, Pyruvate Kinase, Lactate Dehydrogenase) to link the production of AMP by PaaK1 and PaaK2 to the oxidation of NADH to NAD, which could be measured spectrophotometrically at 365 nm (Figure 11).

The assay utilized 3 μg of PaaK1 or PaaK2 per 200 μL reaction volume. Phenylacetic acid concentrations ranging from 20-1000 μM were tested. Each 200 μL reaction also contained 2 U Myokinase, 4 U pyruvate kinase, 16 U lactate dehydrogenase, 4 mM PEP, 0.4 mM NADH, and 2 mM CoA. Each reaction was started by the addition of 1 mM ATP and the absorbance was read every 5 seconds on the Spectramax Plus^{384} microplate reader (Molecular Devices, CA, USA) spectrophotometer. The change in absorbance over time at 365 nm was calculated for the initial linear region for each reaction. The extinction coefficient of NADH at 365 nm ($3.4 \times 10^3\ M^{-1}\ cm^{-1}$) was used to calculate the rate of oxidation of NADH, which is twice the rate of production of AMP.

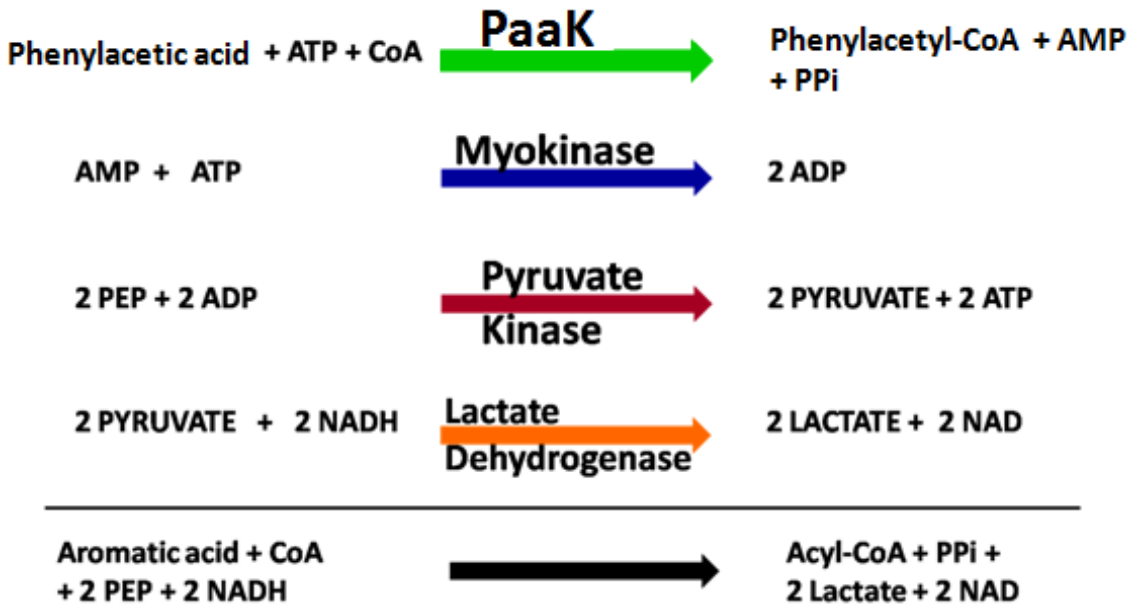


Figure 11: Overall Reaction Scheme for Indirect Activity Assay
Three additional enzymes were used to link the oxidation of NADH to the production of AMP by PaaK1 or PaaK2. The AMP is partially regenerated to ADP through the action of Myokinase, at the expense of a second ATP molecule. Pyruvate Kinase completes the regeneration of both ADP molecules by dephosphorylation of 2 PEP molecules to pyruvate. Lactate dehydrogenase reduces the pyruvate to lactate, oxidizing NADH to NAD in the process. The overall stoichiometry is 2 NAD per 1 molecule AMP produced.

Chapter 3: Crystal structure of PaaK1 reveals novel N-terminal Region and Conserved ATP Interactions

3.1 Introduction

The PAA pathway performs a central metabolic role in many bacteria, facilitating the degradation of several aromatic compounds, such as phenylalanine, as well as environmental pollutants including styrene and ethylbenzene in several *Pseudomonas* sp. (Luengo *et al.* 2001; Corkery *et al.* 1994; del Peso-Santos *et al.* 2006). Most importantly, for bacteria such as *B. cenocepacia,* the PAA pathway is the sole route for phenylalanine degradation (correspondence with Dr. Cardona). Intriguingly, the PAA pathway has also recently been implicated in *B. cenocepacia* pathogenesis (Hunt *et al.* 2004; Law *et al.* 2008). It is plausible that the PAA pathway plays a vital role for satisfying nutritional requirements of this organism during pulmonary infections. However, comprehension of the role of the PAA pathway in *B. cenocepacia* nutrition and pathogenesis will likely require detailed characterization of the enzyme pathways and intermediates.

At present, this novel hybrid aromatic degradation pathway is understood to utilize molecular oxygen for ring epoxidation but also incorporates CoA thioester handles for all pathway intermediates (Ismail *et al.* 2003), akin to classical anaerobic degradation pathways. As such, the first step of the pathway is CoA activation of phenylacetic acid to phenylacetyl CoA, catalyzed by PaaK, a phenylacetate CoA ligase (PCL). Bacterial PCLs, such as PaaK1 of *B. cenocepacia*, are structurally uncharacterized members of the Adenylate forming enzyme superfamily. Interpretation of a high resolution crystal structure of PaaK1 will allow elucidation of structural features of this enzyme necessary

for the activation step of the PAA pathway, not only in *B. cenocepacia*, but for all the Gram positive and Gram negative bacteria who which this novel pathway for degradation of a multitude of aromatic compounds.

Crystal structures from the Adenylate forming enzyme superfamily have revealed a common fold comprised of two α/β domains, a larger N-terminal domain and smaller C-terminal domain joined by a flexible linker. The closest structural homolog to PaaK1 is predicted to be *B. xenovorans* LB400 benzoate CoA ligase (BCL), which performs an analogous role in the benzoate oxidation pathway, a hybrid degradation pathway for benzoic acid (Mohamed *et al.* 2001). While these enzymes catalyze a similar reaction for comparable metabolic pathways, the percent identity between PaaK1 and BCL is remarkably low at 19% and, most interestingly, there is negligible sequence identity over the first ~80 residues of PaaK1 (Appendix A). This considerable lack of identity within the first ~80 residues limits structural predictions of this novel aryl-CoA ligase.

In addition to the 2 domain architecture, short regions of high percent identity or conserved sequence motifs are also retained by superfamily members (Marahiel *et al.* 1997; Chang *et al.* 1997; Babbitt *et al.* 1992). Several of these motifs are also evident in the PaaK1 sequence (Appendix A). Beginning with the phosphate binding motif (P-Loop) at Ser^{93}, PaaK1 contains six of the seven conserved superfamily motifs, five of which are understood to be important for ATP binding and catalysis of the adenylation part reaction. Surprisingly, very few superfamily members have been co-crystallized with ATP, to date. Of the few existing ATP co-structures, all but one are plagued by

problems such as unmodelled peptide regions and absent divalent cations known to be essential for ATP binding. There is a clear need for additional high quality ATP co-structures from this superfamily, which have great potential to further illuminate the ATP-enzyme interactions that are believed to be conserved within this superfamily.

To this end, we have produced a highly ordered co-structure of PaaK1 with ATP, permitting significant observations of key enzyme-substrate interactions with conserved motifs and allowing for comparisons with the few existing ATP bound structures. As the first structural characterization of a bacterial PCL, the novel composition of the N-terminal region, for which there is negligible sequence similarity (Appendix A), is revealed. The crystal structure of PaaK1 in complex with ATP presents a unique opportunity to further explore the canonical adenylate forming enzyme superfamily fold while characterizing a novel N-terminal micro-domain.

3.2 Results

3.2.1 Purification and Crystallization of PaaK1 in Complex with ATP

B. cenocepacia PaaK1, produced recombinantly in *E. coli*, required three purification steps to achieve sufficient homogeneity for crystallization (Figure 12). PaaK1 (MW 48 KDa) eluted at 74 mL from the size exclusion chromatography (SEC) column, corresponding with an apparent MW of 94 KDa, suggesting that PaaK1 forms a dimer in solution. Anion exchange chromatography was incorporated as the final purification step, prior to concentration to 12 mg/mL for crystallization trials. Consistent with the SEC profile, PaaK1 crystallized with two molecules in space group P1 with a Matthews coefficient of 2.49 $Å^3$/Da and a solvent content of approximately 50.0% (v/v).

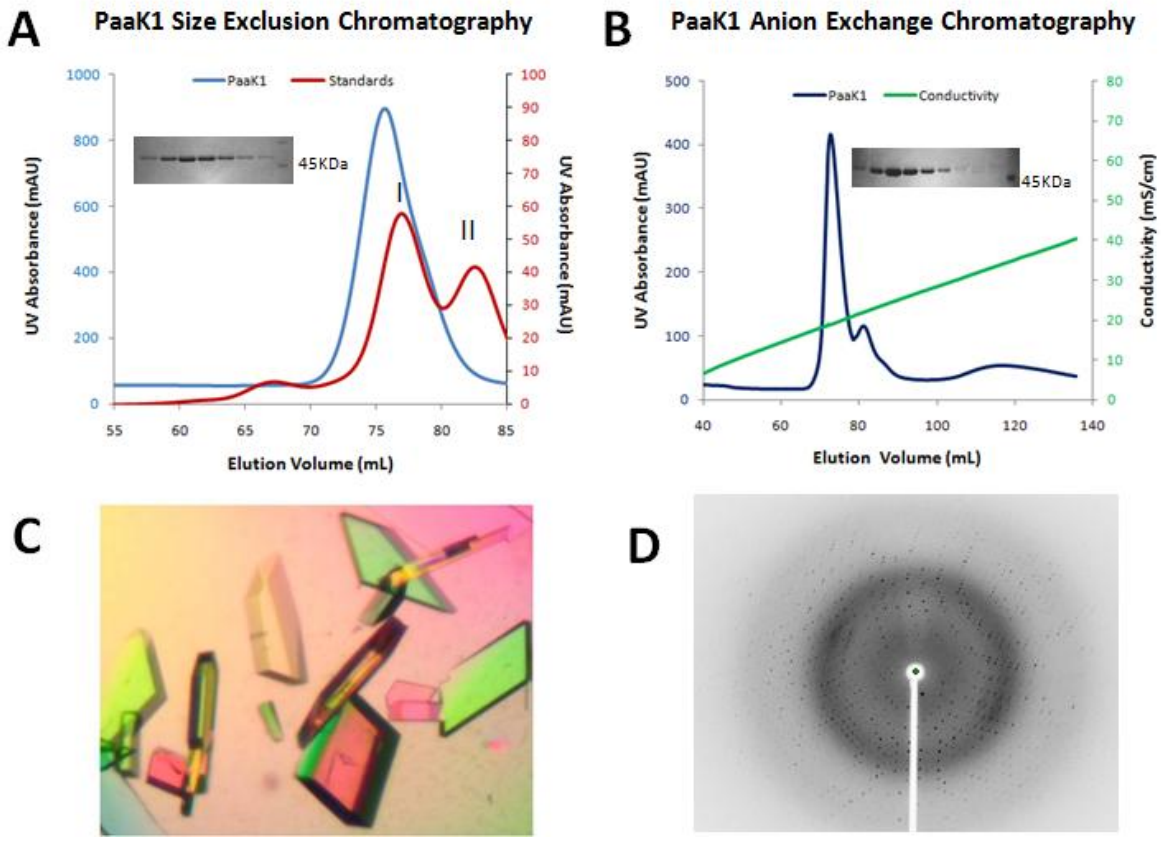


Figure 12: Purification and Crystallization of PaaK1

A. Size Exclusion Chromatography trace of PaaK1 with molecular weight standards: (I) Conalbumin (75 kDa), and (II) Ovalbumin (43 kDa). PaaK1 eluted as an apparent dimer (MW 76 KDa). Corresponding SDS-PAGE gel image demonstrates monomer size (48 KDa) and overall purity. **B**. Anion Exchange Chromatography trace shows PaaK1 elutes at a conductivity of approximately 20 mS/cm, corresponding to ~200 mM NaCl with 20 mM HEPES, 3 mM β-mercaptoethanol, 3% glycerol, pH 8.3, with accompanying SDS-PAGE image. **C**. Selenomethionine derivative PaaK1 crystals required 3 mM $MgCl_2$ and ATP for nucleation and growth. **D**. Diffraction image from selenomethionine derivative PaaK1 crystals.

3.2.2 Solution of PaaK1 structure by SAD phasing

Attempts at solving the PaaK1 structure by molecular replacement were unsuccessful. Several pruned search models were generated, including the *B. xenovorans* benzoate CoA ligase (PDB ID 2V7B), which was predicted to be the best model based on sequence on sequence alignment, and the C-terminal domain of *B. vulgatis* phenylacetate CoA ligase (PDB ID 3GSX). Despite multiple attempts with combinations of these models, no molecular replacement solution was forthcoming. Ultimately the decision was made to produce a selenomethionine derivative of PaaK1 such that the anomalous signal from the selenium could be used to phase the PaaK1 structure. Prior to the production of the selenomethionine derivative, the PaaK1 sequence was inspected to ensure the presence of sufficient methionines for the phasing experiment. A total of 11 methionines are present in the PaaK1 construct (Figure 13) used for crystallization, which is expected to be more than sufficient to phase the ~ 430 residues in the structure.

Selenomethionine derivatized PaaK1 produced irregular, sheet like crystals with excellent birefringence. High quality phases were calculated from the anomalous signal using ShelxC/D/E (Sheldrick 2008). The majority of the model was built by the automated tracing program Buccaneer in the CCP4 suite of programs (Cowtan 2006; *The CCP4 suite: programs for protein crystallography* 1994). The final refined structure begins at either Leu^{7} for chain A or Pro^{4} for Chain B and extends through Lys^{430}. The degree to which PaaK1 was modelled is rare among many of the homologous superfamily structures as loop regions, which often play important roles in catalysis, frequently remain disordered and unmodelled. The structure possesses low temperature factors (Chain A: 21.33 $Å^2$ and Chain B: 18.55 $Å^2$) with good stereochemistry. The majority of

residues (92.7%) lie in the most favourable regions of the Ramachandran plot with no residues in the disallowed regions. The ATP ligand was also modelled with low temperature factors (14.14 Å^2) (Table 1).

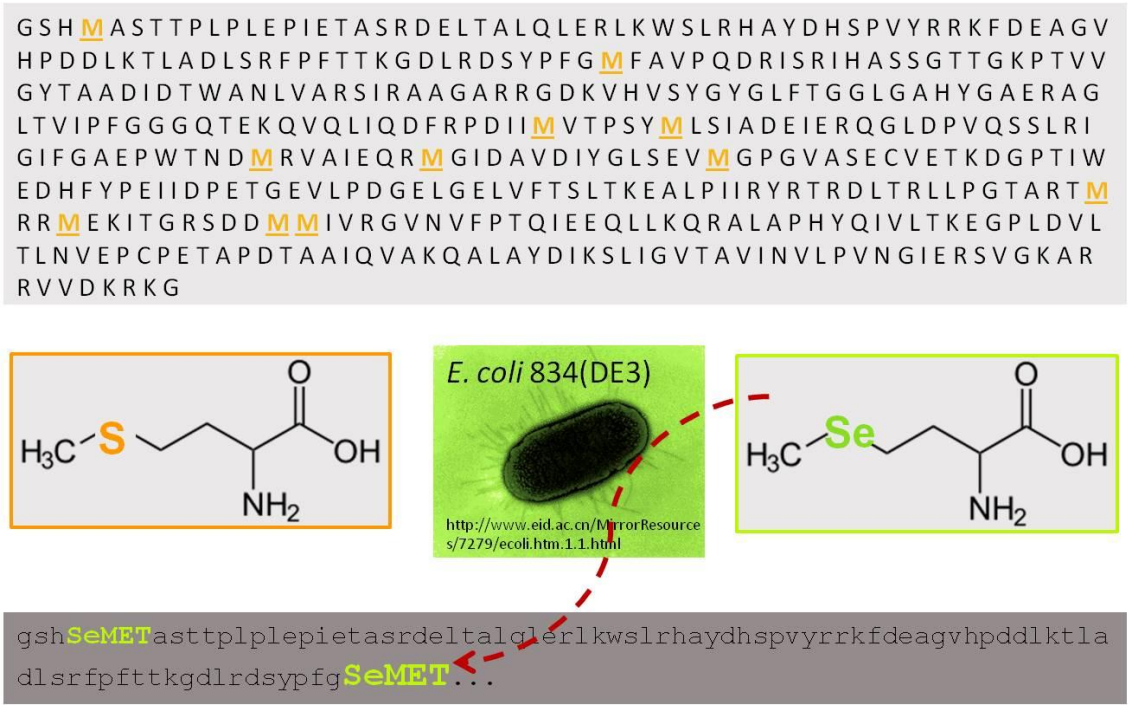


Figure 13: Production of Selenomethionine Derivative PaaK1
Methionine auxotroph *E. coli* 834 (DE3) and minimal media supplemented with selenomethionine were used for production of selenomethionine derivatized PaaK1. The PaaK1 peptide construct shown above contains 11 methionine residues, following His-tag cleavage, making it an excellent candidate for SAD phasing.

3.2.3 PaaK1 Overall Structural Features

PaaK1 was found to adopt two discrete α/β domains, with a larger N-terminal domain (Pro^{4}- Gly^{325}; Figure 14 – purple and green) and smaller C-terminal domain (Met^{331}-Arg^{430}; Figure 14 – blue). The N-terminal domain is comprised of 3 β sheets sandwiched between 9 α helices. Ser^{88}-Ala^{92} and Val^{102}-Tyr^{105} form a double stranded antiparallel β sheet. Phe^{265}-Ile^{270}, Gly^{284}-Ser^{290}, and Asp^{305}- Thr^{307} form a twisted 3 stranded anti-parallel β sheet that is largely shielded from solvent by the connecting loops. Five helices sandwich the distorted 6 stranded β sheet formed by Lys^{131}-Val^{134}, Thr^{157}-Pro^{160}, Ile^{180}-Val^{183}, Ile^{209}-Phe^{212}, Asp^{232}-Tyr^{237}, and Val^{247}-Ser^{249}. A flexible solvated linker joins the N and C-terminal domains. This hinge region, $R^{326}SDDM^{330}$, is well ordered in the PaaK1 structure. The smaller C-terminal domain is comprised of a short, 2 stranded antiparallel β sheet (Met^{331}-Val^{333} and Val^{336}-Val^{338}), and a twisted 4 stranded sheet (Gln^{358}-Glu^{365}, Leu^{367}-Val^{374}, Val^{407}-Leu^{411}, Val^{426}-Asp^{428}) with two α helices (Pro^{340}-Lys^{349}, Thr^{384}-Ile^{402}). The lengthy loop connecting outer strands of the four stranded sheet (Val^{407}-Leu^{411}, Val^{426}-Val^{338}) contains the invariant lysine (Lys^{422}) present in all superfamily members.

The active site of PaaK1 is formed at the interface of the N and C-terminal domains, where, in the case of PaaK1, a well ordered ATP molecule is bound. The C-terminal domain is oriented for the proper positioning of Lys^{422} towards the active site where it is understood to participate in the adenylation part reaction while the domain linker of PaaK1 forms a β turn with a hydrogen bond formed between Arg^{326} (i) carbonyl and Asp^{329} (i+3) main chain amide. Both of these observations are consistent with

superfamily members captured in Conformation 1, poised to catalyze the adenylation part reaction (Wu *et al.* 2009).

Although the general two domain α/β architecture of PaaK1 is consistent with homologous family members, the initial ~80 residues of PaaK form a small microdomain within the N-terminal domain at large. This small bundle of helices (Figure 14-purple) sits between the 3 and 2 stranded β sheets of the N-terminal domain of PaaK1. This region proved to be particularly divergent in structure, with sequence alignments showing negligible sequence similarity to structurally characterized family members (Appendix A).

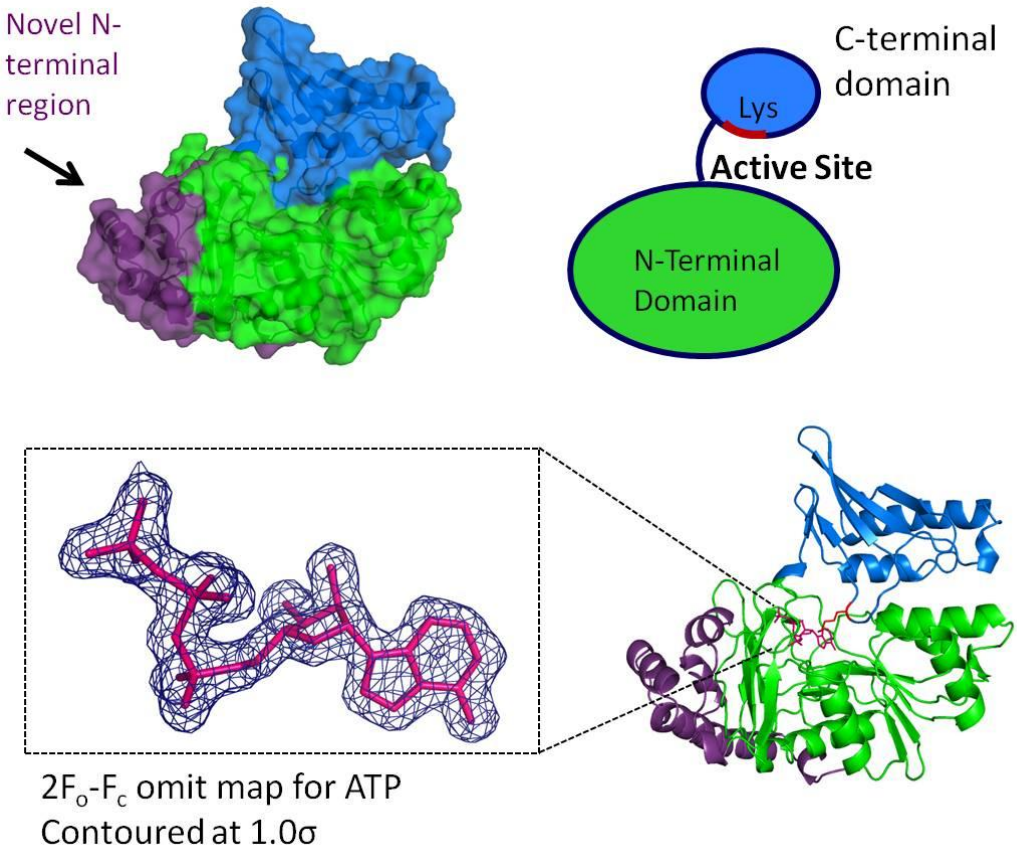


Figure 14: Overall Structure of PaaK1 in Conformation 1.
Larger N-terminal domain is shown in green with the initial 86 residues forming the novel region shown in purple. The smaller C-terminal domain is shown in blue with the active site formed at the domain interface. A well ordered ATP molecule is bound within the active site. PaaK1 has been captured in conformation 1 as the C-terminal domain is oriented as such to direct the conserved lysine towards the active site. Figures were generated using PyMOL (DeLano 2002).

3.2.4 PaaK1 Contains Novel Helical Bundle as N-terminal Microdomain

The remodelling at the extreme N-terminus (residues 1-86) is the most striking divergent feature of the PaaK1 structure. This region showed negligible sequence identity to structurally characterized family members which severely limited structural predictions. The initial 74 residues form a continuous 4 helical bundle approximately 31 x 34 Å, connected by inter-helical loops and coils (Figure 15). Leucine residues of the first (Arg^{15}-His^{36}) and third (Leu^{57}-Arg^{62}) helix, and an additional leucine from a connecting loop Leu^{54} participate in hydrophobic interactions reminiscent of a leucine zipper. Phe^{44} and Phe^{63} are also directed towards this hydrophobic interface, helping to further stabilize the helical bundle. The final few residues of the novel microdomain region (Tyr^{75}-Arg^{86}) form a linker to the initial 2 stranded β sheet. The helical bundle wraps closely around the core N-terminal domain enclosing 1829.2 $Å^2$ buried surface area. The interface of the novel microdomain with the canonical N-terminal domain was more closely examined using the Protein Interfaces, Surfaces and Assemblies Service (PISA), producing a complexation significance score (CSS) of 1.00 (Krissinel 2010). The high complementarity score supports the assertion that the novel N-terminal arrangement exhibited by PaaK1 is not a crystallization artifact. Thirty-one hydrogen bonds and thirteen salt bridges were also predicted to be formed at the interface (Krissinel 2010). Hydrophobic interactions also have a significant presence at the interface; $A^{295}LPII^{299}$ (not shown) of the greater N-terminal domain slot into the groove formed within the large hydrophobic patch at the centre of the helical bundle.

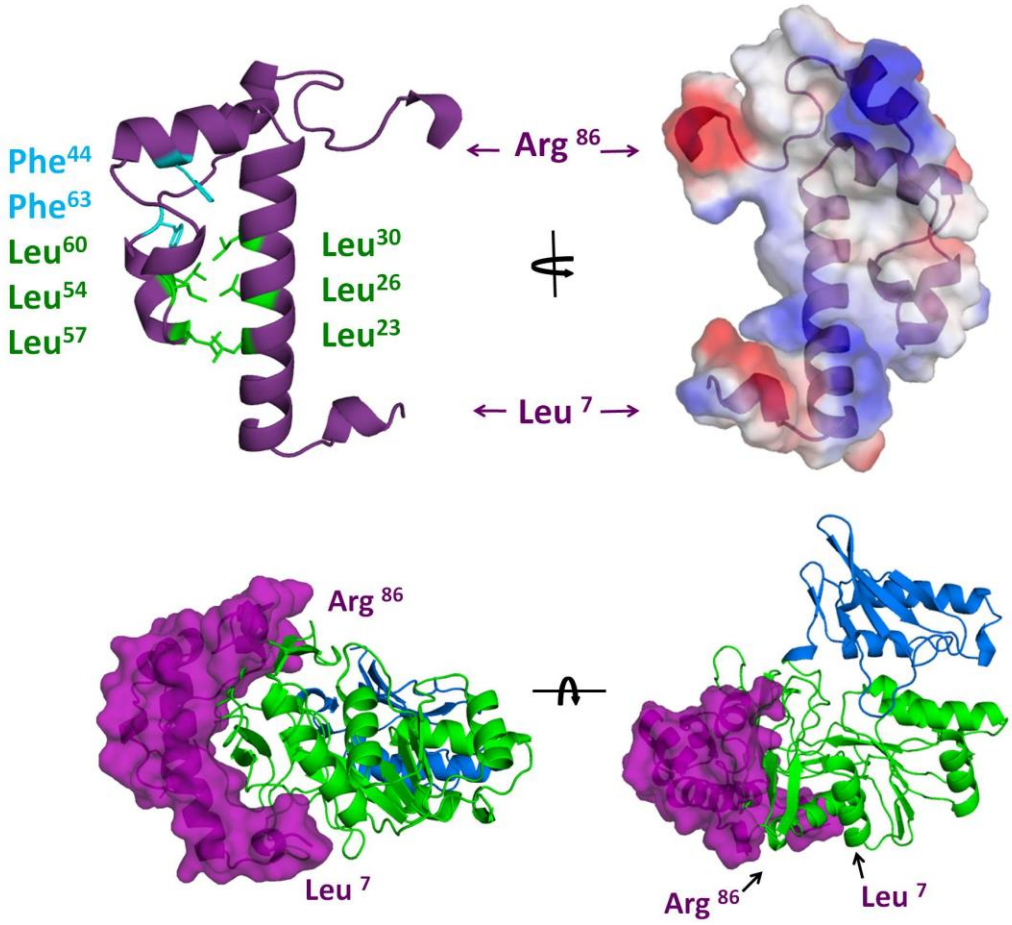


Figure 15: Novel N-terminal Region of PaaK1
The initial 86 residues of PaaK1 form a helical bundle stabilized by hydrophobic interactions (purple). The region wraps tightly around the canonical N-terminal domain (green) with a high degree of complementarity (CSS score = 1.00). A large hydrophobic patch is formed in the centre of the helical bundle, with acidic and basic patches at the edges which are complemented by residues of the canonical N-terminal domain at the interface. Figures were generated using PyMOL (DeLano 2002).

In order to ascertain whether other superfamily members or unrelated protein crystal structures contained a similar arrangement at the N-terminus, the coordinates of the N-terminal domain of PaaK1 were submitted to the DALI server (Holm & Rosenstrom 2010). While the closest structural homologs of PaaK1 were all members of the adenylate forming enzyme superfamily, it became apparent that no structurally characterized family members to date share the same secondary structure arrangement at the extreme N-terminal region as exhibited by PaaK1 (Figure 16). While PaaK1 contains 4 helices located on the alternate face of the N-terminal domain, up to and including residue Ser^{74} (Figure 15 & 16), homologous superfamily members all appear to contain alternating α helices and β strands beginning at the N-terminus (Figure 16).

The structural overlays with homologous family members make it visibly clear that while PaaK1 possesses an unusual microdomain at the extreme N-terminus that replaces the usual topology of alternating strands and helices, however, the core region of the N-terminal domain responsible for enzymatic activity remains intact (Figure 16). Despite possessing low overall sequence identity to superfamily members, PaaK1 retains the familial structure of the N-terminal domain from Ser^{88} onwards. The core region of PaaK1 shows the closest structural homology with benzoate CoA ligase 2V7B. This core region, residues Ser^{88}-Arg^{326} of PaaK1 overlay visibly well with 2V7B, with an r.m.s.d. of 1.72 Å over 141 C-alphas (Figure 16).

One possible consequence of this remodelling may be unusual instability of many bacterial PCLs purified, to date. During the purification of PaaK1, many challenges and

frustrations were encountered and were mainly due to the enzyme's apparent instability and intolerance of temperatures above 4˚ C or short term storage in standard buffers. Apart from *T. thermophilus* PCL (Erb *et al.* 2008), all bacterial PCLs purified to date have been remarkably labile, necessitating the presence of glycerol even for short term stability during purification and up to 20% glycerol during long term storage at -70˚ C (Martinez-Blanco *et al.* 1990; Mohamed & Fuchs 1993; Mohamed *et al.* 2001). Similar observations have not been made for homologous family members which leads one to consider whether the structural reorganization at the N-terminal domain may account for such extreme lability.

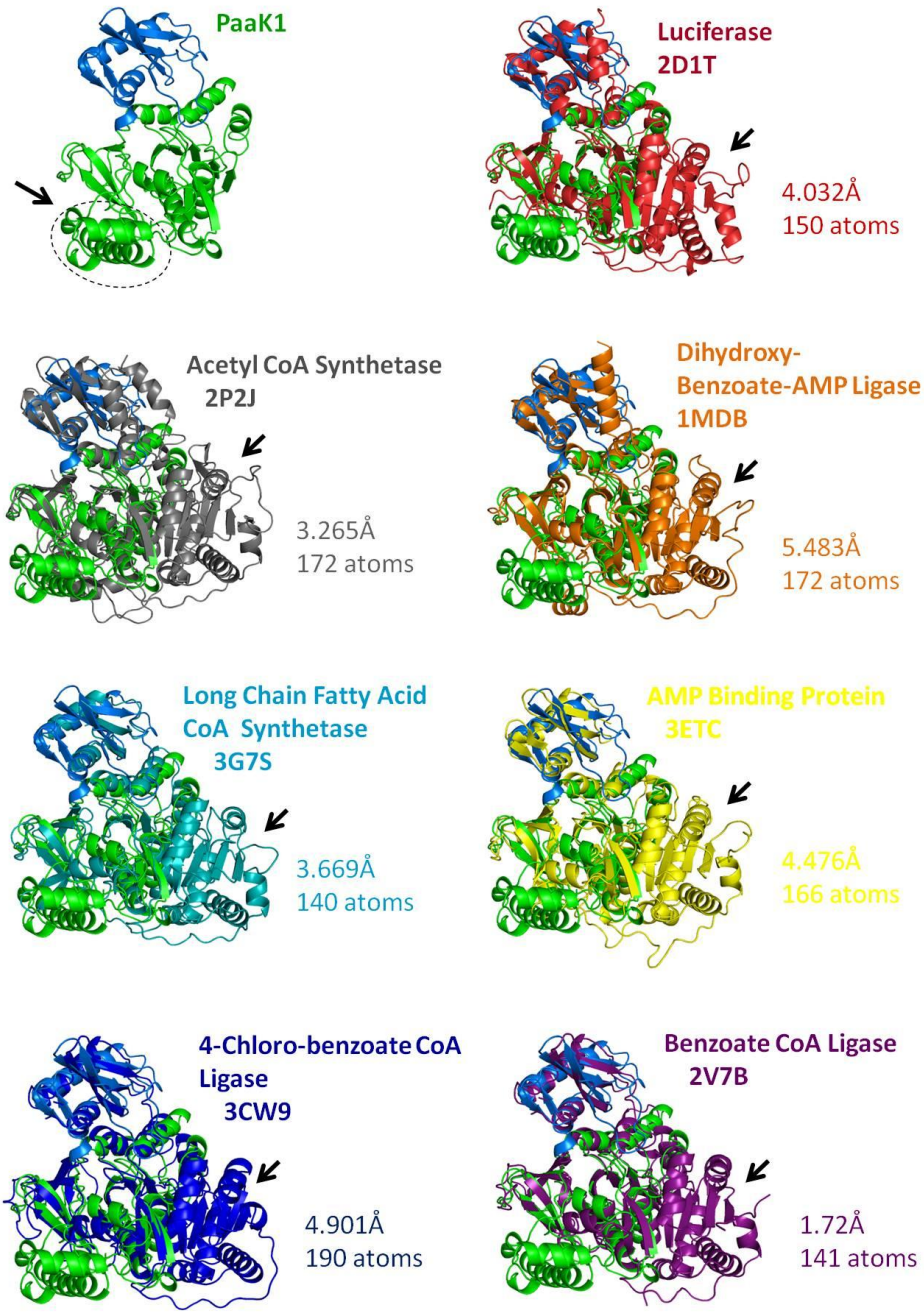


Figure 16: Structural Overlays of Adenylate Forming Enzyme Superfamily Representatives from DALI Server Search

The N and C-terminal domains were overlayed separately with PaaK1 in order to investigate if any other family members shared a similar organization of the N-terminal domain. R.m.s.d. values were calculated for C-alphas. The PaaK1 N-terminal domain contains a novel helix rich region not shared by any structurally characterized superfamily members, to date. Overlay figures were produced using PyMOL (DeLano 2002).

3.2.5 Dimerization of PaaK1 Restores β Sheet Environment of Strands Flanking the P-Loop

Interesting interactions take place at the dimerization interface of PaaK1, where an eight stranded mixed β sheet is formed through interactions between Chain A and Chain B. This high degree of complementarity at the dimerization interface allows for strands from both Chain A and Chain B form a continuous β sheet that flows between the two separate monomers. Intriguingly, the dimerization interaction appears to reconstitute a structural element observed in monomers of homologous family members that had been lost to PaaK1 as a result of restructuring at the N-terminus.

While the strands flanking the P-loop of PaaK1 (residues Ser^{88}-Ala^{92} and Val^{102}-Tyr^{105}) stand alone within the monomer structure, forming just a two stranded β sheet (Figure 17), the strands flanking the P-loop of homologous family members are part of a larger sheet extending through the N-terminal domain. These accompanying strands are absent in PaaK1, as they are a component of the classical N-terminal domain composition not possessed by PaaK1. However, dimerization of PaaK1 incorporates two strands flanking the P-loop of Chain A with an additional six strands contributed by Chain B. As each monomer contributes identically to the dimerization interface, the two anti-parallel strands of Chain B interact identically with the distorted six stranded β sheet of Chain A. It follows that the additional two hydrogen bonds contributed by PaaK1 dimerization may help to stabilize the ‘two stranded’ β sheet that is otherwise incorporated into a larger sheet in homologous family members, perhaps ameliorating the loss of the classical arrangement at the N-terminal domain.

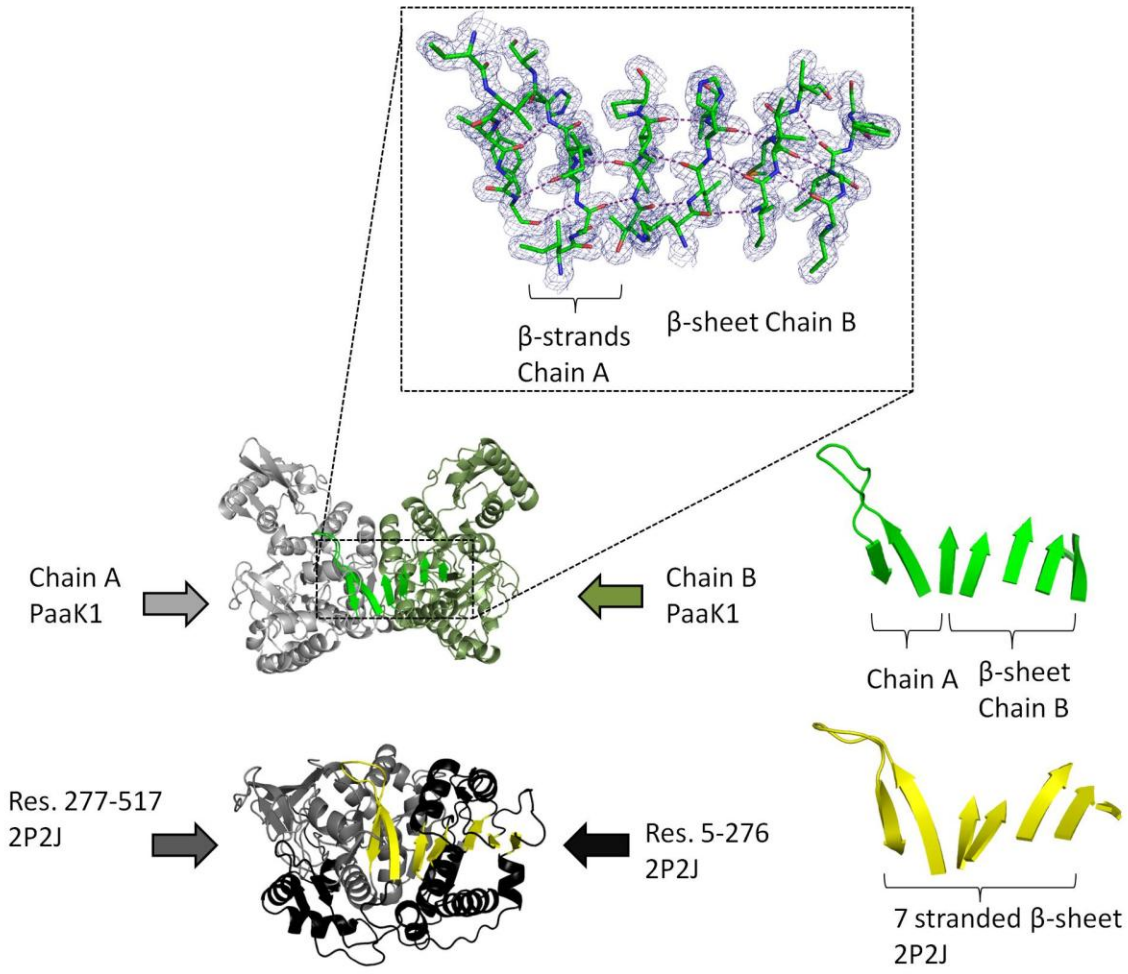


Figure 17: Dimerization of PaaK1 restores β sheet Environment of Strands Adjacent to the P-Loop

$2F_o$-F_c electron density map shown in blue mesh for β strands of P-loop (Chain A) and flanking β strands shown with inter-chain hydrogen bonds marked in purple. PaaK1 dimers are shown with Chain A in grey, Chain B forest green, and P-loop and flanking β strands highlighted in green. Acetyl CoA Synthetase 2P2J core N-terminal region shown in grey, initial portion of N-terminal not overlaying with PaaK1 shown in black with P-loop and flanking β strands shown yellow. Figure was generated using PyMOL (DeLano 2002).

3.2.6 ATP Bound PaaK1 Reveals Rarely Observed Protein-Substrate Interactions with Conserved Residues and Motifs

The P-loop of PaaK1, $S^{93}SGTTGKPT^{101}$, can be seen enveloping the β and γ phosphates of ATP (Figure 18). Thr^{96} & Thr^{97} form hydrogen bonds with the β and γ phosphates respectively, while Ser^{94} interacts with the α phosphate. The 2A ACS structure (PDB ID 3C5E *to be published*), is so far the only other homologous ATP co-structure to possess an ordered P-Loop. In the case of 2A ACS, Thr^{224} and Ser^{225} make comparable hydrogen bonds with the β and γ phosphates with an additional hydrogen bond formed between Lys^{229} and the γ phosphate of ATP. The side chain of the corresponding lysine (Lys^{104}) of PaaK1 is directed towards the bulk solvent, so it appears that the additional hydrogen bond contributed by the lysine is not necessary for ordering of the P-loop in the case of PaaK1. Ser^{245} of PaaK1 also forms a hydrogen bond with the α phosphate of PaaK1. Although not part of the P-loop itself, this residue (often a Thr) has been observed to form hydrogen bonds with the α phosphate in homologous structures containing the acyl-adenylate (Wu *et al.* 2008). In addition to hydrogen bonding with Thr^{96} and Thr^{97} residues of the P-Loop of PaaK1, the β and γ phosphates coordinate a single Mg^{2+} ion. The Mg^{2+} coordinates four water molecules in addition to the phosphates. A conserved glutamate residue (Glu^{241}), part of the conserved ϕ(G/W)x(A/T)E motif (where ϕ represents an aromatic residue) (Gulick 2009), forms hydrogen bonds with two of the water molecules. The β phosphate is also stabilized through bidentate hydrogen bonds with charge contribution to the side chain of Arg^{326} in the PaaK1 structure. Arg^{326} is a member of the interdomain linker, recognized by the sequence motif $Rx(D/K)x_6G$ and appears to be strictly conserved (Gulick 2009).

Additional conserved residues are involved in stabilizing the ribose moiety of ATP in the PaaK1 structure. Asp^{305} forms bidentate hydrogen bonds with the ribose hydroxyls, while Lys^{422} is directed towards the oxygen of the ribose of the ATP (2.93 Å), potentially forming a second hydrogen bond with the closest oxygen of the β phosphate (3.32 Å). The adenine moiety sits in a largely hydrophobic pocket, defined by the side chains of Ile^{323}, Tyr^{237}, Trp^{217} and the backbone of $Ala^{214}Glu^{215}Pro^{216}$, consistent with observations from both homologous AMP and adenylate intermediate bound structures. Asp^{235} interacts with the primary amine of the adenine moiety forming a small basic patch in the binding pocket.

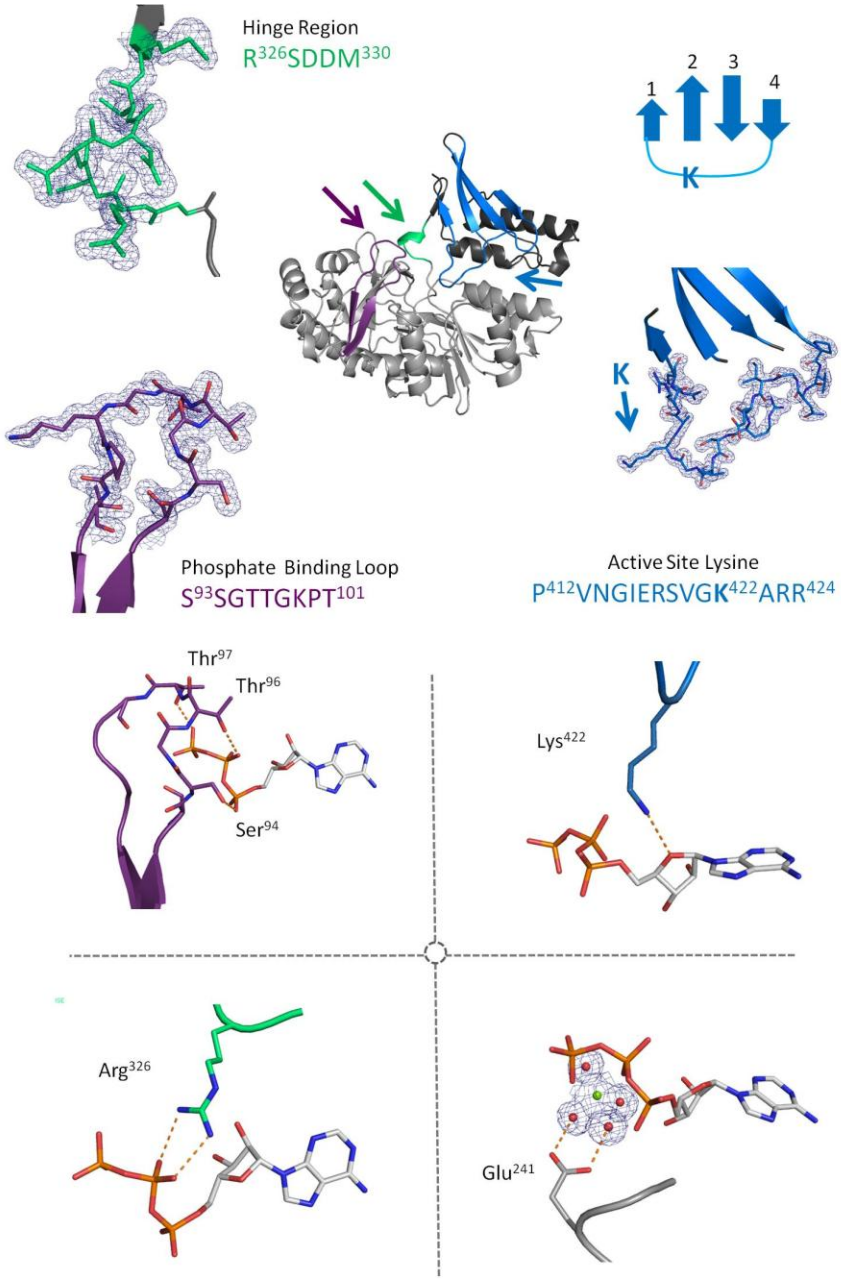


Figure 18: Detailed Assessment of PaaK1-ATP Interactions
Flexible peptide regions are largely involved in mediating ligand binding. The often disordered P-Loop can be seen enveloping the β and γ phosphates of ATP. The flexible domain linker also interacts with the ATP through a conserved arginine residue. An extensive loop between β strands 1 and 4 in the C-terminal domain contains the invariant lysine residue, found in all superfamily members. Additionally, a conserved glutamate residue can be observed hydrogen bonding with two waters, coordinated by the Mg^{2+} cation. Figure was generated using PyMOL (DeLano 2002).

3.3 Discussion

3.3.1 Residues Involved in Substrate Binding During Catalysis

The ATP bound PaaK1 crystal structure reported here provides valuable structural insights into protein ligand interactions involved in ATP binding and Mg^{2+} coordination. While the P-loop is understood to order the β and γ phosphates of ATP, it is surprising that even in the few ATP co-structures within this superfamily, the P-loop often exhibits poor density and remains unmodelled (Osman *et al.* 2009). The PaaK1-ATP co-structure presented here is only the second structure of this superfamily with observable ATP-P-loop interactions, as well as the coordination of the Mg^{2+} cation by the conserved glutamate (Glu^{241}). Other conserved residues, such as the arginine residue found in the interdomain linker, are also observed interacting with the β phosphates of ATP in the PaaK1 structure, roles which differ from those observed in AMP and adenylated intermediate co-structures.

Possible catalytic role of P-loop

In the PaaK1 ATP bound structure, the Thr^{96} & Thr^{97} form hydrogen bonds with the β and γ phosphates respectively, the α phosphate forms hydrogen bonds with Ser^{94}. A catalytic role for the γ phosphate binding threonine was suggested by Wu *et al*. when the $Thr^{161\rightarrow}$A mutant of 4-CBL exhibited a 4000 fold reduction in k_{obs} for the first part reaction (2008). Mutations to the β phosphate binding residue likely would have resulted in a similar decrease in activity. Previous to this work, mutagenesis was performed on the conserved glycine and proline residues of the P-loop (consensus sequence (S/T)(S/T/G)**G**(S/T)T**G**x**P**K) (Chang *et al.* 1997). The Gly mutants exhibited both a 3-4 fold increase in K_M, as well as a 20 and 1400 fold reduction in k_{cat}, respectively, while the

activity of the Pro mutant was immeasurable (Chang *et al.* 1997). Clearly, the Gly and Pro residues play vital structural roles in maintaining the flexibility and torsion of the P-loop, allowing the Thr and Ser functional groups to perform their catalytic roles. These residues are likely responsible for maintaining the β and γ phosphates in a reactive conformation and may interact with the departing pyrophosphate following the adenylation part reaction (Wu *et al.* 2008).

Conserved glutamate and Mg^{2+} cation essential for ATP binding and Adenylation part reaction

The PaaK1 structure contained a Mg^{2+} cation in the active site, coordinating the β and γ phosphate as well as four well ordered water molecules with B-factors less than 16 $Å^2$. This supports the position observed in the few existing ATP co-structures including the 2A ACS co-structure as well as DltA that included a Ca^{2+} cation in place of Mg^{2+}. More to the point, the position of the cation in all three of the ATP containing structures contrasts with existing AMP co-structures, which is placed adjacent to the α phosphate. Interestingly, a similar placement was tentatively used for the Mg^{2+} ion in a *T. thermophilus* long-chain acyl-CoA synthetase structure (PDB ID 1V25) which contained the ATP analog AMP-PNP (Hisanaga *et al.* 2004).

With additional superfamily members crystallized in complex with ATP, it becomes clear that the invariant glutamate plays an initial role in stabilizing the β and γ phosphates. It now seems likely that the biological role of the Mg^{2+} cation and the conserved glutamate involves ATP binding prior to and during the adenylation part reaction. As the role of the conserved Glu in the ATP binding has only recently been observed, there is little

mutagenesis data available. However, our observations support the hypothesis of Osman *et al.* that the conserved Glu residue is responsible for stabilizing the reactive conformation of the phosphate groups, as well as their findings that substitution of said Glu residue for Gln results in a 14-fold reduction of the k_{cat} for the adenylation part reaction (2009). It is interesting that substitution for Gln both increased the K_M approximately 8-fold, and decreased the k_{cat}. As a polar residue, Gln would likely be able to maintain hydrogen bond interactions with the waters coordinated by the Mg^{2+} cation. Perhaps the loss of the negative charged side group alters the coordination geometry slightly to a less productive conformation. It is interesting that in the case of the DltA-ATP co-structure, the researchers found it necessary to replace the catalytically essential Mg^{2+} cation with Ca^{2+} (Osman *et al.* 2009). While the location of the cation is maintained, the water coordination geometry is slightly altered. The positions of the waters and phosphates coordinated by the Mg^{2+} of PaaK1 sit at the points of a near-perfect octahedron, while the distances between the coordinated atoms of the Ca^{2+} cation are greater and the angles are slightly skewed in that not all positions are mutually perpendicular. The researchers also noted that DltA was inactive in the presence of calcium, leading them to predict that the enzyme requires precise geometry as a prerequisite for activity (Osman *et al.* 2009). Substitution of Glu to Gln would likely disrupt such exact geometry, negating any electron withdrawing effects contributed by the Mg^{2+} cation, making the β and γ phosphates a much poorer leaving group for the adenylation reaction.

Conserved hinge-region Arginine plays initial role Adenylation reaction

Similar to the reorientation of the conserved glutamate, the conserved arginine, a member of the interdomain linker, has now also been observed in distinctly different orientations in ATP and adenylated intermediate or AMP bound structures. Previous to the ATP bound structures, the arginine had been observed forming hydrogen bonds with the ribose hydroxyls of the adenylated intermediate (PDB ID 3CW8), or AMP (3CW9). The PaaK1 structure, along with DltA and 2A ACS structures form a growing body of evidence that the conserved arginine plays an initial role stabilizing the phosphate moieties of ATP. Mutational studies by Osman *et al*. demonstrated that substitution of the Arg for Gln resulted in a 20-fold reduction for k_{cat}, leading them to the hypothesis that, like the conserved Glu, the Arg is again involved in promoting the adenylation part reaction (2009). Our findings that Arg^{326} of PaaK1 forms bidentate hydrogen bonds with the β phosphate strongly support this hypothesis. Mutational studies in 4-CBL showed that mutating the equivalent Arg^{400}→Ala resulted in an even greater 300 fold reduction in $k_{1,}$ the rate constant for the first part reaction (Wu *et al*. 2008). While the Gln mutant could likely maintain hydrogen bond interactions with the β-phosphate, the Ala mutant could not, resulting in a much greater decrease in activity as the β-phosphate could not be as well stabilized in the reactive conformation. The slighter loss of activity due to Gln mutation may suggest that the positively charged Arg side chain may also play a role in ameliorating the build-up of negative charge during the first part reaction, perhaps enhancing the ability of the β and γ phosphate to act as a leaving group.

Active site lysine demonstrates range of orientations in Conformation 1

The invariant C-terminal domain lysine residue, essential for the adenylation part reaction (Wu *et al.* 2008), has also been observed in a variety of orientations. In the case of PaaK1 it was directed towards the ribose oxygen of ATP but also within hydrogen bonding distance of the β phosphate, as was the case with the 2A ACS structure. For DltA, the other ATP bound structure, it is directed towards the β phosphate (Osman *et al.* 2009). Alternatively, in AMP bound structures, such as PDB IDs: 1AMU, 1V26, and 2D1Q, the lysine is directed towards the remaining α phosphate. Additionally, the lysine has been observed directed towards the negatively charged carboxyl side chain of structures such as 2V7B, containing only the carboxylate substrate (Bains & Boulanger 2007). The positively charged lysine has been proposed to follow negative charges during the electron transfer from the substrate-carboxylate group to the α phosphate of ATP, stabilizing negative charges during substrate binding and eventually the transition state (Osman *et al.* 2009).

The varieties of substrate dependent orientations of the invariant lysine residue that have now been observed certainly support this theory. Although the lysine was initially proposed to interact with the carboxylate group (Wu *et al.* 2008), the ATP co-structures where the lysine is directed towards the β phosphate strongly suggest that the high density positive charge of the lysine is responsible for the majority of any electron withdrawing effects to the phosphate leaving group. Mutation of Lys→Ala of 4-CBL reduced the rate of catalysis by 1000-fold, while mutation of the conserved Arg→Ala produced only a 300-fold reduction (Wu *et al.* 2008), suggesting that any contribution of the Arg side chain is secondary to the Lys residue.

3.4 Remaining Questions

The PaaK1 structure provided a wealth of information regarding the general architecture of bacterial PCLs, confirming structural reorganization of the N-terminal domain as well as revealing many enzyme-ATP interactions. The invariant glutamate and arginine, along with the P-loop, appear to play an initial role stabilizing the β and γ phosphates and may facilitate their eventual departure. Based on these observations or residue positions, as well as a 48˚ reorientation of the C-terminal domain when comparing the DltA and 2A ACS, Osman *et al*. have also proposed an intriguing *pre-adenylation* state (2009). The researchers observed that the 2A ACS structure contained an ordered P-loop which forms hydrogen bonds with the C-terminal domain. It follows that perhaps this 48˚ reorientation of the C-terminal domain enabled the hydrogen bonds, further stabilizing the P-loop to at last visualize P-loop-ATP interactions. The researchers also state that, because the DltA-ATP structure was essentially isomorphous with a previously obtained DltA-adenylate structure, crystal packing may have restricted C-terminal domain orientation preventing the *pre-adenylation* conformation from being captured (Osman *et al.* 2009). PaaK1 on the other hand, also contains an ordered P-loop, but no hydrogen bonds are observed between the P-loop and the C-terminal domain. Therefore, one of the proposed characteristics of this *pre-adenylation* state, namely interactions between the P-loop and C-terminal domain, do not appear to be essential for ordered P-loop-ATP interactions. However, an interesting trend has emerged whereby the P-loop is often modelled in enzymes captured in the *thioester* forming conformation (PDB IDs: 2P2J, 3B7W, 3CW9), or conformation 2. As a result of the 2A ACS and PaaK1 structures, the

role of the P-loop for phosphate binding in this superfamily is now indisputable. However, an exciting possibility remains that the P-loop may play additional roles stabilizing both conformation 2 as well as the newly proposed *pre-adenylation* conformation.

An additional characteristic of the proposed *pre-adenylation* state is a slight rotation of the C-terminal domain when comparing with adenylate bound structures. In the case of PaaK1 it is very difficult to estimate whether there or not there is a slight rotation of the C-terminal domain of relative to DltA or 2A ACS ATP co-structures. This appears to be mainly due to lower structural homology in the C-terminal domain. Additional PaaK1 structures in complex with the adenylated intermediate could reveal a more subtle reorientation of the C-terminal domain, as well as provide additional insight into the reorientation and additional roles of ATP binding residues following the adenylation part reaction. Furthermore, with only ATP bound within the active site, the aryl substrate binding pocket could not be readily defined; obtaining additional co-structures which divulge the location of the aryl substrate pocket are a prerequisite to comprehending aryl substrate binding for PaaK1 and PaaK2. Functional profiles of the two paralogs also remained to be explored, with the eventual goal of reconciling functional and structural attributes.

Chapter 4: Investigating the Aryl Substrate Binding Pocket and Reaction Kinetics of PaaK1 and PaaK2

4.1 Introduction

Probably the most unusual feature of the PAA pathway in *B. cenocepacia* is that it retains two copies of the initial enzyme, PaaK1 and PaaK2, but just one copy of all subsequent pathway enzymes. Although presently there is no rationale for the presence of two copies of PaaK in *B. cenocepacia,* it is possible that paralogs may possess differing substrate affinities and specific activities to allow controlled entry to the pathway tailored to the bacteria's immediate metabolic needs. Because PaaK1 and PaaK2 catalyze the first and only committed step of the pathway (Mohamed *et al.* 2002) they are perfectly positioned to control the flow of phenylacetic acid into the PAA pathway and the rate at which it is degraded. Furthermore, phenylacetyl-CoA, the product formed by PaaK1 and PaaK2 plays an important regulatory role, acting as the inducer molecule to de-repress transcription of the *paa* genes (correspondence with Dr. Cardona), providing an additional level at which the PaaK paralogs may influence the rate of phenylacetate consumption.

Likewise, for the benzoate degradation pathway in *B. xenovorans* LB400, an analogous hybrid aromatic degradation pathway for the metabolism of benzoate, two entire copies of the pathway are known to be transcribed at different times depending on substrate and phase of growth (Denef *et al.* 2005; Denef *et al.* 2004). The initial enzymes, benzoate CoA ligase C (chromosomal copy) and M (megaplasmid copy) were then determined to possess differing catalytic profiles, allowing the rate of benzoate degradation to be optimized for the growing conditions (Bains & Boulanger 2007). While such

comprehensive transcriptomic details are not known for the PAA pathway in *B. cenocepacia*, a thorough functional comparison of the paralogs may provide some indications as to whether there is a functional rationale for the retention of both copies; detailed structural characterizations may identify the features responsible for any unique functional attributes.

Our hypothesis that the PaaK1 and PaaK2 paralogs possesses differing functional profiles with respect to affinity and specific activity towards phenylacetic acid is in part based on the wide range of kinetic attributes observed for bacterial PCLs. While only a handful of other bacterial phenylacetate CoA ligases have been purified and functionally characterized, there has been a great deal of discrepancies with both catalytic efficiency and substrate specificities. The apparent K_Ms for bacterial phenylacetate CoA ligases range from 14 µM to 16.5 mM, while the specific activities range from 0.48 to 24.4 µmol/min/mg (Mohamed & Fuchs 1993; Martinez-Blanco *et al.* 1990; El-Said Mohamed 2000). Although the location of the aryl pocket for bacterial PCLs has not yet been identified, these results suggest a degree of plasticity in the PCL aryl binding pocket and residues defining it. More to the point, the percent sequence identity for these orthologs ranges from ~50-70%, while PaaK1 and PaaK2 share 69% sequence identity. As PaaK1 and PaaK2 are no more closely related in sequence than other bacterial orthologs, it is not unreasonable to expect differences in kinetic attributes and perhaps in the aryl pocket itself.

In order to concretely identify the aryl substrate binding pocket, crystallization trials for both PaaK1 and PaaK2 were attempted with phenylacetic acid only, as well as in the presence of $MgCl_2$ and ATP, with the goal of obtaining a co-structure with sufficient electron density for placing the phenyl ring. An added benefit of obtaining PaaK structures containing the adenylated intermediate derives from further observations of conserved active site residues and their corresponding roles following the first part reaction. These structural investigations are accompanied by a functional characterization of PaaK1 and PaaK2 with respect to kinetics of phenylacetic acid utilization. Detailed comparisons of the aryl substrate pockets of the paralogs, in conjunction with the kinetic data, are explored in pursuit of a rationale for the presence of two *paaK* copies in *B. cenocepacia.*

4.2 Results

4.2.1 Purification and Crystallization of PaaK1 and PaaK2

PaaK1 and PaaK2 were produced and purified as per the PaaK1 protocol. PaaK2 (MW 48.3KDa) eluted at 77 mL from the SEC column with an apparent molecular weight of 74KD, indicating PaaK2 may form a dimer in solution as was seen for PaaK1 (Figure 19). Prior to crystallization trials, the proteins were incubated with $MgCl_2$, ATP, and phenylacetic acid. Diffraction data from a fragment of a single, large crystal of PaaK1 were collected, processed, scaled and merged to 1.92 Å. PaaK1 was again found to have crystallized in space group P1 with two monomers in the asymmetric unit and a Matthew's coefficient of 2.59 $Å^3$/Da, corresponding to 52.62% solvent. For PaaK2, images were collected from a single crystal and processed, scaled, and merged to 1.9 Å.

Although PaaK2 also crystallized with two monomers in the asymmetric unit, it was found to crystallize in space group $P2_1$ with approximately 48.09% solvent content.

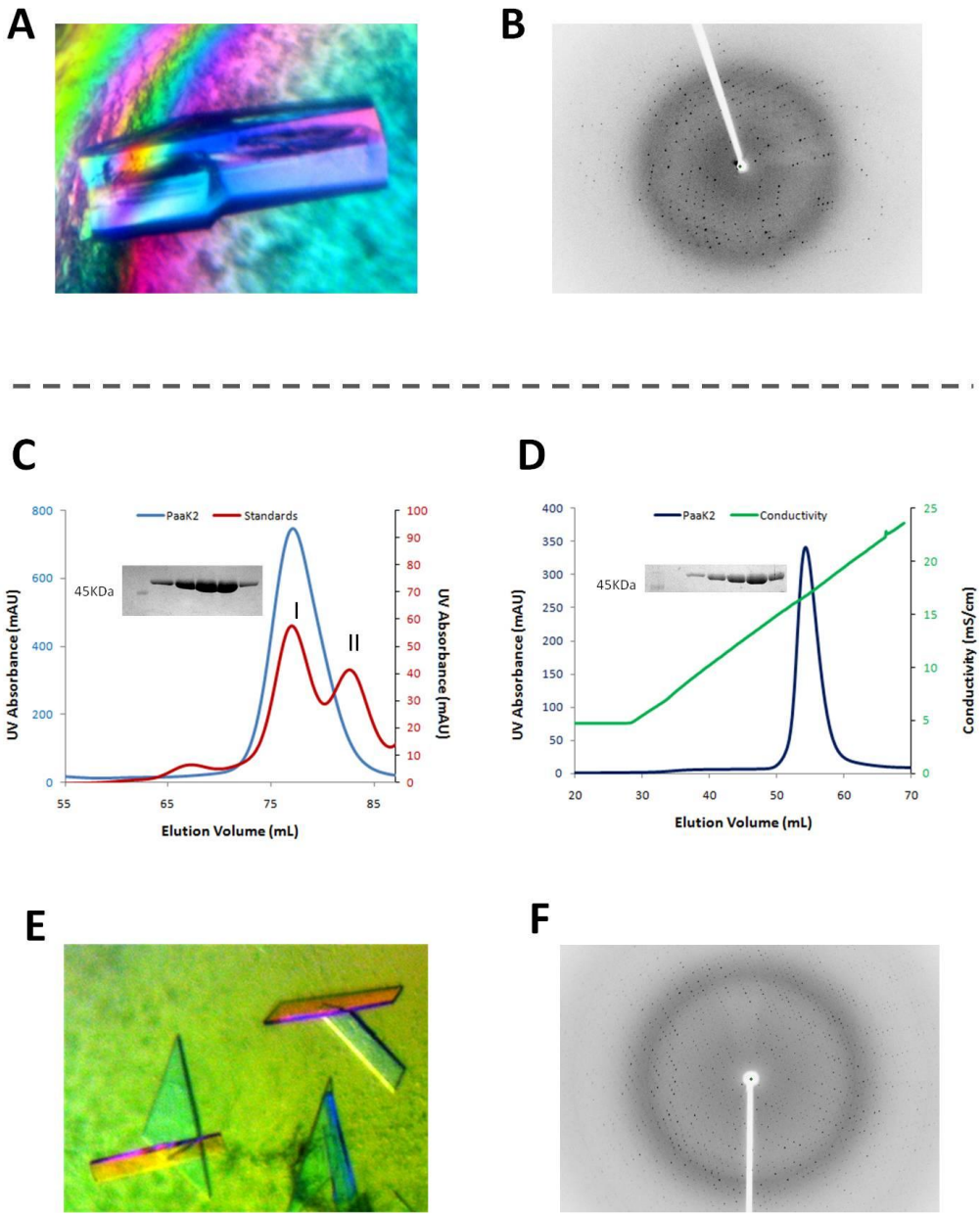


Figure 19: Purification and Crystallization of PaaK1 and PaaK2
A. Crystals of native PaaK1, **B.** Diffraction from native PaaK1 crystals collected on Rigaku R-axis IV++ area Detector at University of Victoria. **C.** Size Exclusion chromatography trace of PaaK2 with standards (I) Conalbumin (75 kDa), and (II) Ovalbumin (43 kDa) with corresponding SDS-PAGE gel image showing monomer size (MW 48.3 KDa) and purity. **D.** Anion exchange chromatography trace shows that PaaK2 elutes at a conductivity of approximately 17 mS/cm. Accompanying SDS-PAGE image demonstrates purity. **E.** Crystals of native PaaK2, post incubation with ATP, $MgCl_2$, and phenylacetic acid. **F.** Diffraction from native PaaK2 crystals collected at SSRL.

4.2.2 Obtaining Phenylacetyl Adenylate Co-structures of PaaK1 and PaaK2

PaaK1 crystallized in the presence of phenylacetic acid, ATP and $MgCl_2$ was essentially isomorphous with the PaaK1 ATP co-structure (Tables 1 and 2) and structure solution was accomplished using coordinates from the PaaK1-ATP dimer with ATP and waters removed. In all, the majority of residues Pro^4-Lys^{432} of chain A and Leu^5-Arg^{430} of chain B, showed clear electron density. Only Thr^{96} & Thr^{97} of chain A and Thr^{96}-Gly^{98} of chain B were disordered and were not modelled, which was likely due to the loss of the β and γ phosphates during the adenylation of phenylacetate. Clear density for the phenylacetyl adenylate intermediate could be observed and the ligand was modelled in with average B-factors of 23.6 $Å^2$ (Table 2). PaaK1 is again observed in conformation 1 with the invariant lysine directed towards the active site (Figure 20).

The PaaK2-phenylacetyl adenylate co-structure was solved by molecular replacement (MR) using the PaaK1-adenylate co-structure as the search model. The MR solution produced sufficient electron density for the N-terminal domain, and density for the phenylacetyl adenylate was immediately apparent. However, the resulting density for the C-terminal domain was initially very poor and it could only be partially modelled. Only after complete modelling of the N-terminal domain accompanied by several rounds of refinement was sufficient additional electron density observed to permit modelling of the C-terminal domain. It soon became apparent that the PaaK2 had already undergone the domain rearrangement and was in conformation 2, poised to catalyze the thioesterification part reaction. Once again, nearly the entire protein is ordered and could be modelled into the density, starting with His^6 and ending with Gln^{338} or Arg^{337} in chains A and B, respectively. Clear electron density corresponding to the phenylacetyl

adenylate is present within the active site and the intermediate was easily modelled with average B-factors of 12.14 $Å^2$ (Figure 20 and Table 3).

4.2.3 General Structural Comparison of PaaK1 and PaaK2

Although captured in different conformations, the N and C-terminal domains of PaaK1 and PaaK2 overlay separately with low r.m.s.d. values. The N-terminal domains (Pro^4-Arg^{326} of PaaK1 and His^6-Gly^{329} of PaaK2) of the phenylacetyl adenylate containing structures overlay with an r.m.s.d. of only 0.48 Å over 294 C-alphas (Figure 21). One difference is readily apparent: while the P-loop of the PaaK1 adenylate structure remains unmodelled, there is excellent density present for the PaaK2 adenylate structure in this region. The C-terminal domain of PaaK2 is rotated approximately 180˚ relative to that of PaaK1, with an additional 40˚ tilt to bring the two larger helices into closer contact with the N-terminal domain. The C-terminal domains, from Ser^{327}-Lys^{431} of PaaK1 and Ser^{331}-Ala^{430} of PaaK2 can be overlayed separately with a slightly higher r.m.s.d. of 1.181 Å over 79 residues (C-alphas). The long flexible loop connecting strands 1 and 4 of the C-terminal domain, which contains the invariant lysine, do not overlay as well which is likely due to the high degree of flexibility of the extended loop. In the PaaK1 adenylate structure, the lysine interacts with the α phosphate of the adenylate intermediate, anchoring it to the active site, while in the Paak2 structure, it projects into the solvent, likely accounting for the differing loop structures.

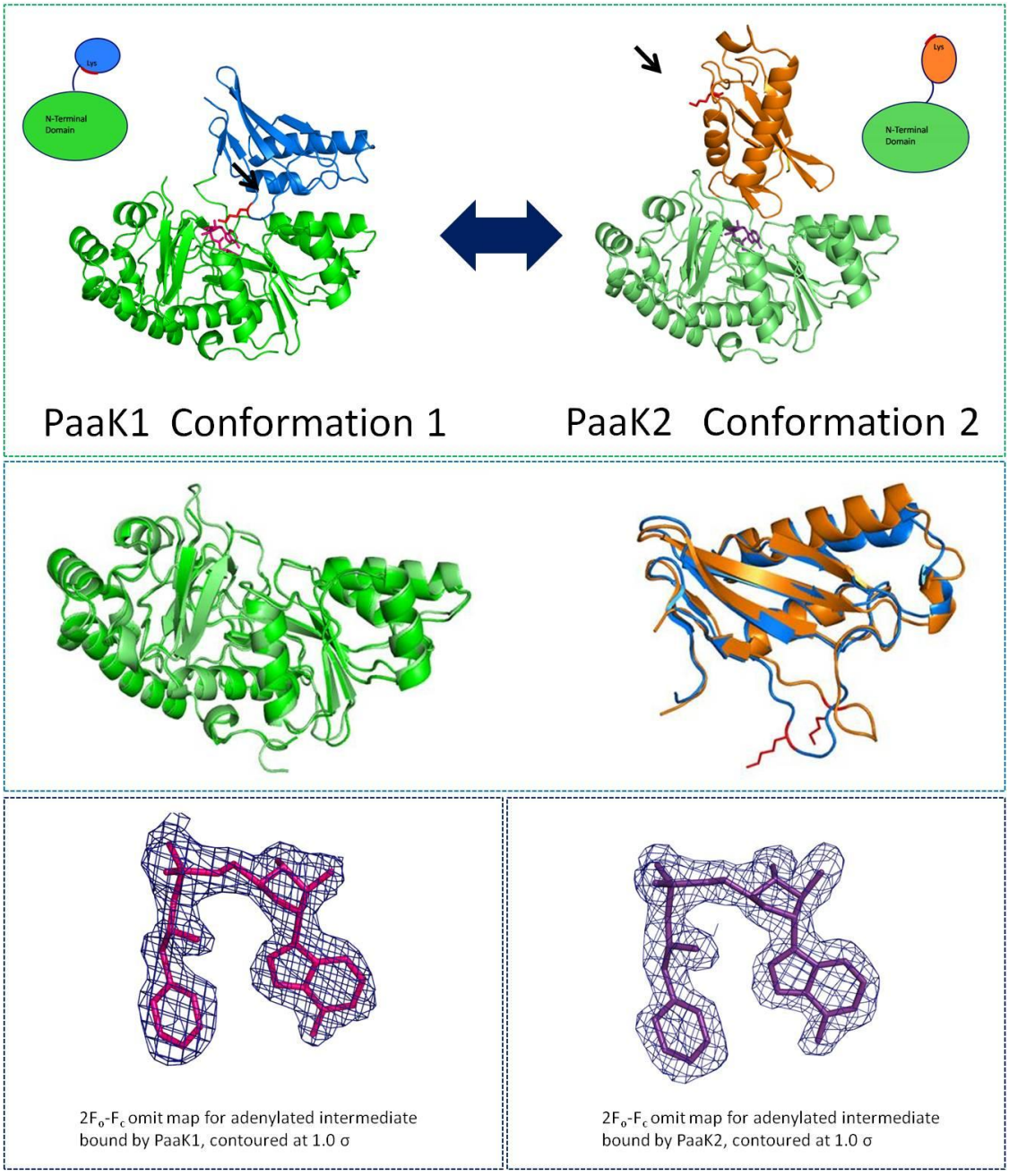


Figure 20: PaaK1 and PaaK2 Phenylacetyl Adenylate co-structures
Top Panel: PaaK1 is in conformation 1 and PaaK2 in conformation 2, having already undergone domain alternation. **Middle Panel**: N-terminal domains overlay with r.m.s.d. of 0.48 Å over 294 C-alphas. C-terminal domains from residues Ser^{327}-Lys^{431} of PaaK1 and Ser^{331}-Ala^{430} of PaaK2 overlay with r.m.s.d of 1.18 Å over 79 C-alphas. **Bottom Panel:** Each contains the phenylacetyl adenylate intermediate within the active site, shown below with the electron density mesh calculated $2F_o$-F_c omit map. Figures were generated using PyMOL (DeLano 2002).

4.2.3 Interactions of Conserved ATP Binding Residues following Adenylation Reaction

Comparisons between PaaK1 captured in complex with ATP and the adenylate co-structures of PaaK1 and PaaK2 reveal that the conserved lysine, arginine, and glutamate have indeed formed new interactions following the departure of the pyrophosphate (Figure 21). Lys^{422}, which was initially directed towards the ribose oxygen, hydrogen bonds with the α phosphate of the PaaK1-adenylate structure. In the case of PaaK2, the invariant lysine is no longer within the active site and is positioned approximately 29 Å from its previous position. Glu^{241} of the PaaK1-adenylate structure has essentially maintained its position from that of the PaaK1-ATP structure. Overlay comparisons of PaaK1-ATP with the PaaK1-adenylate co-structure reveals that the Glu^{241} side chain itself has moved only slightly. Although the Glu side chain is now directed towards the α phosphate, it is 3.77 Å away and likely unable to form hydrogen bond interactions. A similar situation is observed for Glu^{245} of the PaaK2-adenylate structure, which sits 4.16 Å away from the α phosphate. For both PaaK1 and PaaK2, the glutamate forms a salt bridge with a lysine residue. Lys^{68} is 2.73 Å away from Glu^{241} of PaaK1, while Lys^{72} is 2.58 Å away from Glu^{245} of PaaK2, perhaps accounting for the apparent rigidity of the conserved Glu residue. Arg^{326} shows the greatest degree of movement between the ATP and adenylate bound PaaK1 structures. Following the loss of the pyrophosphate, it is reorganized to hydrogen bond with the ribose hydroxyl. Arg^{330} of PaaK2, while still in position to hydrogen bond with the ribose hydroxyls, has again reoriented. This repositioning is likely due to PaaK2 occupying conformation 2. Here, the Arg^{330} side chain forms a salt bridge with Asp^{332}, stabilizing the domain linker for the thioester forming conformation.

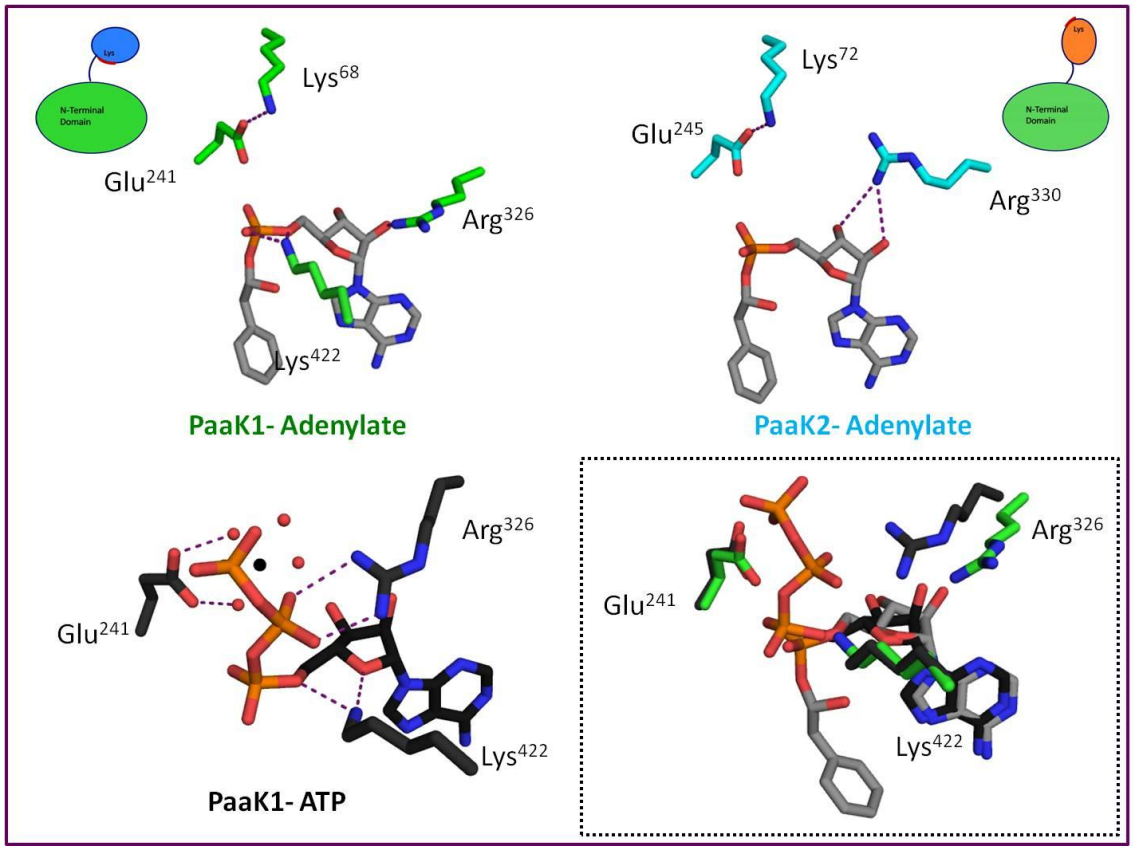
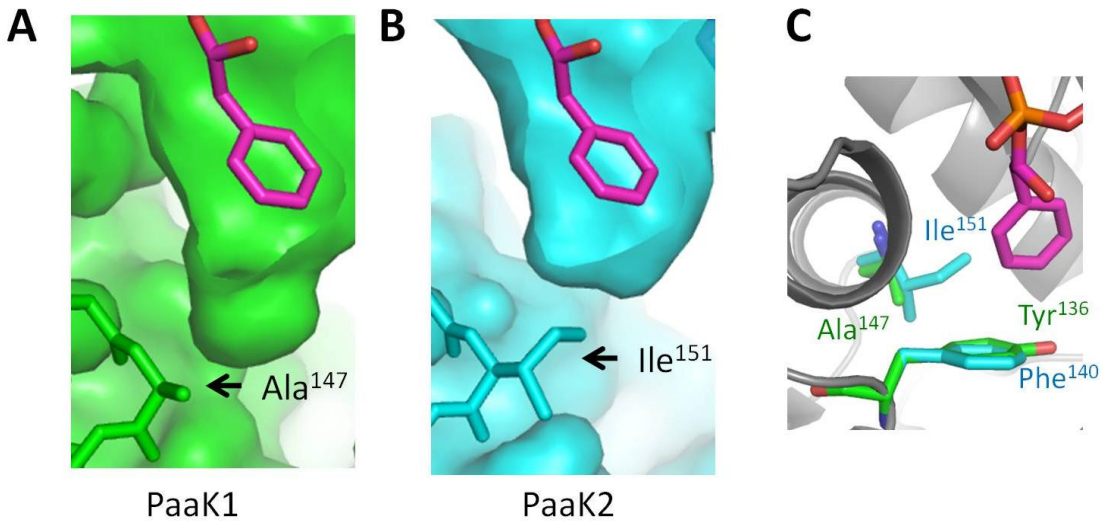


Figure 21: Additional Roles of Conserved Residues Utilized in ATP Binding.
Lys^{422} which initially formed hydrogen bonds with the ribose oxygen and β phosphate of ATP has relocated to the α phosphate in the PaaK1-phenylacetyl adenylate co-structure. As PaaK2 has already undergone domain alternation, the analogous lysine is no longer within the active site. Glu^{241} which coordinated with two waters surrounding the Mg^{2+} ion remains in essentially the same position, forming a salt bridge with Lys^{68}. Glu^{245} of PaaK2 is captured in an identical conformation adjacent to Lys^{72}. Arg^{326} initially formed bidentate hydrogen bonds with the β phosphate, but is now within hydrogen bonding distance of the ribose hydroxyl in the adenylate structures. An overlay of the bound ATP and adenylated intermediate with active site residues of PaaK1demonstrates a static Glu^{241} along with reorientation of Arg^{326} to the ribose hydroxyl. Figures were generated using PyMOL (DeLano 2002).

4.2.4 Establishing the Molecular Determinants of Substrate Affinity

Fortunately, the electron density observed for the adenylated intermediate in both PaaK1 and PaaK2-phenylacetyl adenylate co-structures (Figure 20) finally enabled identification of the pocket (Figure 22). Interestingly, residues defining the pockets are not identical between the two paralogs; most notably, Ala^{147}, which defines the base and side of the PaaK1 pocket, is replaced by Ile^{151} in PaaK2. As a result of this substitution, PyMOL surface representations show that the PaaK1 pocket may contain additional volume, extending below the level of the bottom of the aromatic ring (Figure 22 A) to residues Val^{117} and Met^{182} (DeLano 2002). The PaaK2 pocket on the other hand, shows a much shallower pocket, with the base defined by the bulkier side-chain of Ile^{151} (Figure 22 B). Lastly, Tyr^{136} of PaaK1 is replaced by Phe^{140} in PaaK2. However, these aromatic side chains overlay nearly perfectly and likely don't contribute to differences observed in the binding pocket volume.

While for the most part, the defining resides for the aryl pocket are identical for PaaK1 and PaaK2, the possibility remained that the paralogs may differ in utilization of phenylacetic acid. Even where binding pocket first shell residues are identical, paralogs may exhibit differences in K_M or V_{max} (Bains & Boulanger 2007). In order to test how residue substitutions between the binding pockets of PaaK1 and PaaK2 may contribute to differences phenylacetic acid utilization, a kinetic comparison of the two paralogs was conducted.



D

Residues Defining Aryl Substrate Pocket													
PaaK1	Val 117	Tyr 136	Phe 141	Gly 143	Ala 147	Met 182	Gly 213	Ala 214	Ile 236	Tyr 237	Gly 238	Gly 244	Pro 245
PaaK2	-	Phe 140	Phe 145	Gly 147	Ile 151	-	Gly 217	Ala 218	Ile 240	Tyr 241	Gly 242	Gly 248	Pro 249

Figure 22: Comparison of PaaK1 and PaaK2 Aryl Substrate Binding Pockets
A: Pymol surface predictions of aryl binding pockets for PaaK1 (green). **B**: surface prediction of PaaK2 pocket (light blue). Ala^{147} (PaaK1) and Ile^{151} (PaaK2) are shown positioned to define the base of the pocket. **C:** Unique residues defining PaaK1 and PaaK2 aryl binding pocket shown are superimposed. Ala^{147} & Tyr^{136} (PaaK1) and Ile^{151} & Phe^{140} (PaaK2) are the only binding pocket residues substitutions. **D:** Residues defining aryl substrate binding pocket for PaaK1 and PaaK2 with substitutions highlighted. PaaK1 pocket extends to Val^{117} and Met^{182}. Figure parts A, B, and C were generated using PyMOL (DeLano 2002).

4.2.5 Kinetic comparison of PaaK1 and PaaK2

The high resolution crystal structures of the paralogs revealed that although PaaK1 and PaaK2 share a nearly identical tertiary structure with 69% sequence identity, there are two substitutions in residues defining the aryl binding pocket. Kinetic determinations for the two paralogs were conducted in order to investigate how these differences may contribute to affinity and specific activity towards phenylacetic acid, potentially providing a biochemical rationale for the presence of the two paralogs in *B. cenocepacia*. The rate of catalysis for each paralog was measured using an indirect spectrophotometric assay at varying concentrations of phenylacetic acid while concentrations of ATP and CoA were saturating. Both PaaK1 and PaaK2 displayed Michaelis-Menten kinetics with the rate of catalysis increasing as [phenylacetic acid] was increased, until saturation was reached (Figure 23). Kinetic parameters such as K_M and specific activity for each of the paralogs could then be determined.

PaaK1 and PaaK2 varied most dramatically in the apparent K_M phenylacetic acid; the apparent K_M for PaaK2 (150 µM) was more than double that for PaaK1 (62 µM). Although the paralogs exhibit disparate K_Ms they are not as vastly different as orthologous PCLs characterized to date: 150 µM is still two orders of magnitude lower than the determined K_M determined for *P. putida* (16.5 mM) (Martinez-Blanco *et al.* 1990). The apparent K_M for PaaK1 (62 µM) is in close agreement with the 50 µM K_M determined for *T. thermophilus* PCL (Erb *et al.* 2008), while the *A. evansii* K_M was determined to be just 14 µM (El-Said Mohamed 2000). Clearly, there is a trend towards differing substrate affinities for orthologous PCLs. The specific activities determined for both PaaK1 and PaaK2 (5.0 and 6.0 µmol/min/mg, respectively) are in reasonable

agreement with those determined for other bacterial PCLs, which were between 21 and 24.4 µmol/min/mg (Erb *et al.* 2008; Mohamed & Fuchs 1993; El-Said Mohamed 2000; Martinez-Blanco et al. 1990). However, the specific activity for *A. evansii* PCL was determined to be nearly an order of magnitude smaller (0.48 µmol/min/mg) and is the lowest determined to date (El-Said Mohamed 2000). In addition to being well within the range of previously determined specific activities for PCLs, the specific activities for PaaK1 and PaaK2 are in quite close agreement with each other; any functional rationale for the retention of the paralogous PCLs in *B. cenocepacia* likely stems from differing substrate affinities.

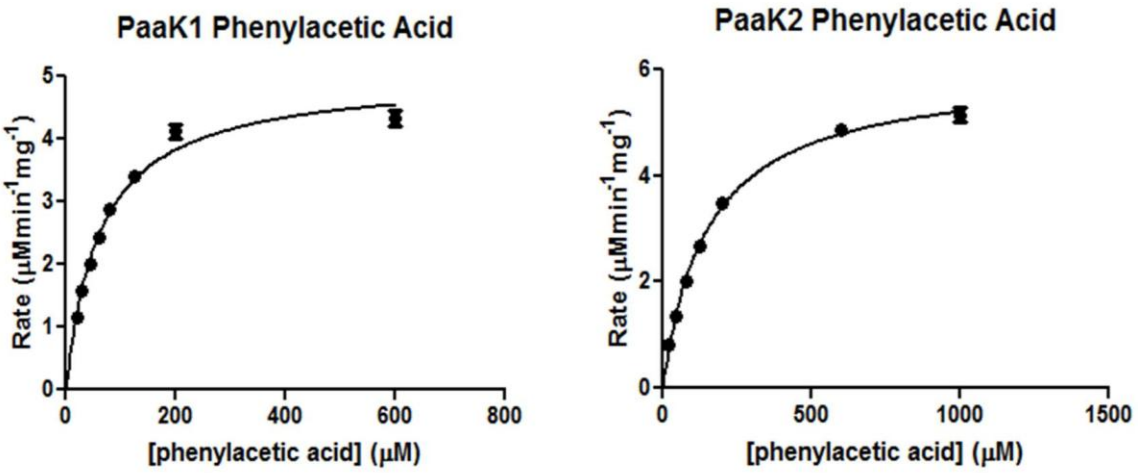


Figure 23: Michaelis-Menten Plot of Initial Reaction Rate vs. [S] for both PaaK1 and PaaK2
Saturating concentrations of ATP and CoA were used while [phenylacetic acid] was varied. Error Bars show SEM of three replicates for each [phenylacetic acid].

Table 4: Apparent K_M and Specific Activity
determined for PaaK1 and PaaK2 for phenylacetic acid with standard error.

	K_M (μM)	Specific Activity (μmol/min/mg)
PaaK1	62 ± 4.3	5.0 ± 0.13
PaaK2	150 ± 7.0	6.0 ± 0.10

4.3 Discussion

Prior to this thesis work, bacterial PCLs remained structurally uncharacterized members of the adenylate forming enzyme superfamily. The initial ATP bound PaaK1 structure revealed that the putative ATP binding residues of the superfamily are indeed conserved, with the novel low-sequence identity region restricted to 3 helices at the N-terminal domain. However, the residues defining the aryl substrate binding site could not be concretely identified from the ATP co-structure alone. The goal of this chapter's work was to examine the possible relocation of residues involved in ATP binding, following the adenylation part reaction, as well as identify the residues which define the pocket in order to compare and contrast the structural and functional attributes of PaaK1 and PaaK2.

4.3.1 A Structural Basis for Substrate Affinity

The functional characterization of the paralogs established that the apparent K_M for PaaK2 was more than double that of PaaK1. The large discrepancy of apparent K_M for phenylacetic acid suggests that there may indeed be a functional rationale behind the retention of an apparently redundant gene. PaaK2 appears to have a lower affinity for phenylacetic acid and requires a much higher concentration before saturation is reached. After inspecting the aryl substrate binding pockets for both paralogs, there is no obvious answer as to the differences in binding affinity. In fact, at first glance, one might expect PaaK2 to better accommodate the aryl ring with the hydrophobic Ile^{151} side chain defining the base of the pocket as opposed to the Ala^{147} of PaaK1, which permits the pocket to extend to Met^{182} and Val^{117}. One observation that may explain the decreased affinity of PaaK2 for phenylacetic acid is that the hydrophobic side chain of Ile^{151} must

be directed towards the polar Thr^{121} in order to accommodate substrate binding. In the case of PaaK1, the Ala^{147} side chain is much smaller, and Val^{117} takes the place of Thr^{121}. These substitutions may contribute to the impressive increase in K_M, but likely do not explain the disparities entirely.

A similar result was described for the paralogous benzoate CoA ligases (BCL) of *B. xenovorans* LB400 (Bains & Boulanger 2007). The BCL paralog encoded on the megaplasmid (BCL_M) was found to be more active towards both benzoate and 2-amino benzoate, despite all residues directly involved in substrate binding being conserved between the paralogs (Bains & Boulanger 2007). It is likely that nearby residues of PaaK1 and PaaK2, although not directly involved in substrate binding, may be involved in forming supporting interactions with residues and side chains that do directly interact with substrate. These second shell residues may also contribute to the energetic of substrate binding and may contribute to surprisingly different affinities of the two paralogs (Bains & Boulanger 2007).

4.3.2 A Remodelled Aryl Binding Pocket for Bacterial PCLs

Upon comparing the aryl binding pockets from aryl activating enzymes crystallized with their respective substrates, it becomes clear that PaaK & PaaK2 use a remodelled pocket to bind the aryl ring in a significantly different orientation (Figure 24). A loop connecting two short strands defines the back of the pocket, with the aromatic substrate tilted forward. The loop is comprised of a short GPGV motif that appears to be conserved among bacterial PCLs, at least the ones that have been shown functional, to

date. Overlays with other aryl activating family members show that the loop protrudes into the ‘typical’ aryl pocket.

In general, it has been found that enzymes in this family with short chain acyl substrates utilize shallow pockets along the surface of the N-terminal domain, while those using medium chain acyl or aryl substrates contain deeper hydrophobic pockets (Shah *et al.* 2009). It was assumed that PaaK1 and PaaK2 would employ a similar deep, hydrophobic binding pocket to those of other aryl CoA ligases, perhaps incrementally deeper than those of benzoate-derivative activating enzymes in order to accommodate the additional carbon linkage of the carboxylic acid group of phenylacetate. However, the structures of PaaK1 and PaaK2 demonstrate that the GPGV loop defines not the base, but one side of the pocket, stacking against the phenyl ring and shifting the pocket towards the adenine moiety. The aryl ring of phenylacetate is also rotated approximately 90˚ relative to benzoate derivatives. The additional CH_2 in the functional group of the phenylacetate compensates for this rotation and reorientation, allowing the carboxyl group to remain in the correct position to react with ATP.

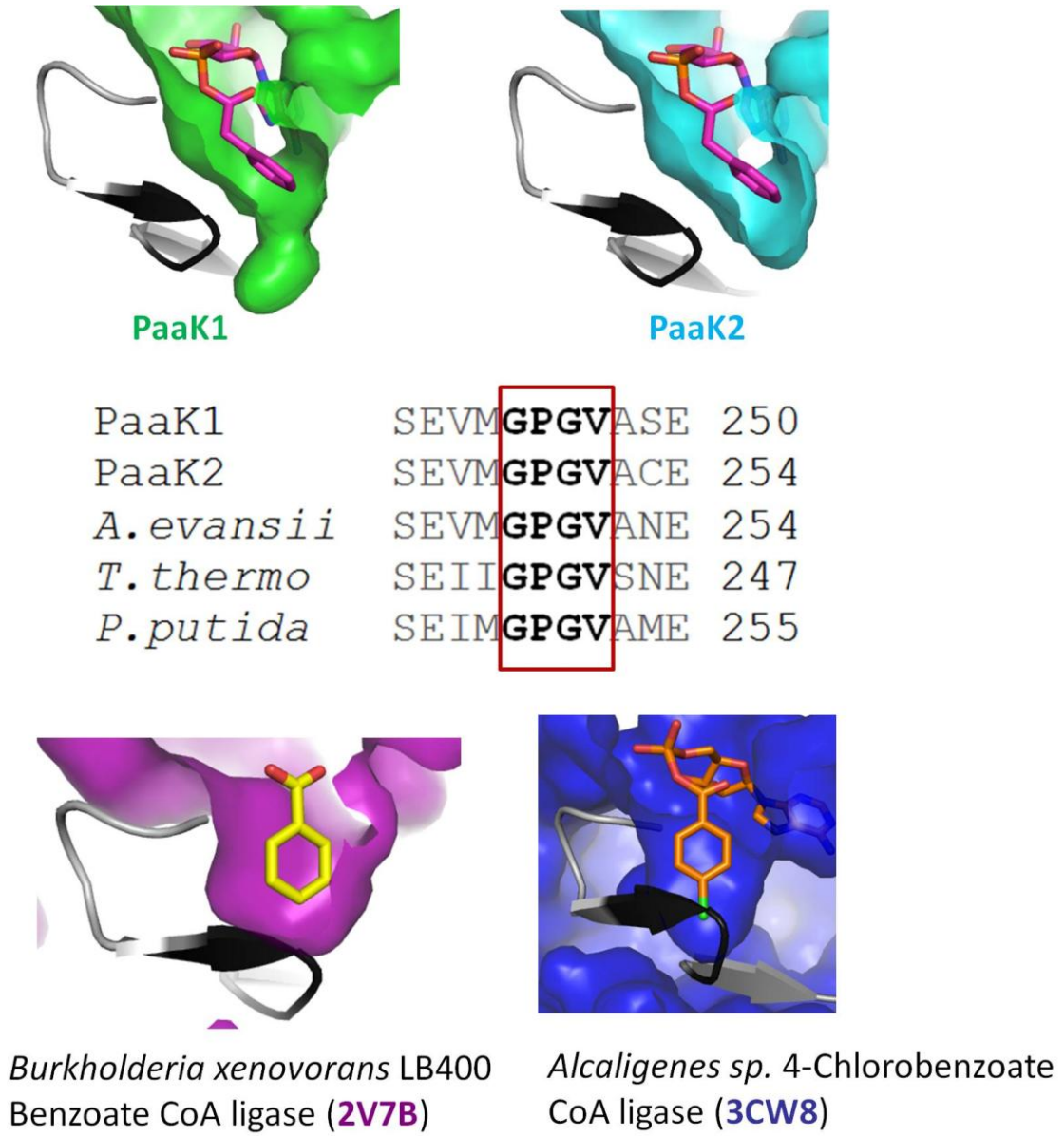


Figure 24: Comparison of Aryl Substrate Binding Pockets for PaaK1, PaaK2, and other Aryl-CoA Ligases

The GPGV loop (Grey, GPGV in black) defines the back of the PaaK aryl binding pocket and appears to be conserved among bacterial phenylacetate CoA ligases. Substrate pockets of other aryl-CoA ligases: *Burkholderia xenovorans* LB400 Benzoate CoA ligase (2V7B) with benzoate and *Alcaligenes sp.* 4-Chlorobenzoate CoA ligase with 4-chlorobenzoyladenylate (3CW8). Protein and ligand representations were generated using PyMOL (DeLano 2002).

4.3.3 PaaK2 Captured in Conformation 2 with Adenylated Phenylacetyl Intermediate

As members of the adenylate forming enzyme superfamily, bacterial phenylacetate CoA ligases such as PaaK1 and PaaK2 utilize a bi-uni-uni-bi ping pong mechanism, incorporating a dynamic rearrangement at the domain interface (Wu *et al.* 2009). Following the adenylation part reaction and prior to the thioesterification of the adenylated intermediate, CoA ligases undergo a large conformational change where the C-terminal domain must rotate approximately 140° relative to the N-terminal domain (Wu *et al.* 2009) (Figure 25). This domain alternation remodels the single active for a second, chemically distinct part reaction; however, the details of what drives the process of this conformational change are not well understood.

The fact that PaaK2 crystallized in conformation 2, containing the phenylacetyl adenylate intermediate but in the absence of CoA, is of much interest. Previously it was proposed that the presence of CoA and its subsequent binding were what drove the enzyme to adopt conformation 2 (Gulick *et al.* 2004); however, apo crystal structures have been obtained in both conformation 1 (PDB IDs 1MDF, 1T5S) and 2 (3B7W, 3ETC). It now appears to be more complicated than simply adding the appropriate ligand in order to coax a particular enzyme into crystallization in one conformation or the other; to date, only 4-chlorobenzoate CoA ligase, 2A ACS, and *T. thermophilus* fatty acid CoA synthetase enzymes have been captured in both conformations.

It has been proposed that certain enzymes may be predisposed to one conformation or another but are then able to rearrange to the productive conformation in the presence of

substrate (Reger *et al.* 2008). It is possible that PaaK2 is one such enzyme with a greater propensity to exist in conformation 2; however, closer examination of crystal packing, however, shows that conformation 1 could not be accommodated within this lattice. In the case of PaaK2, it is possible that crystal packing has influenced the enzyme conformation. Interestingly, the crystals could only be obtained following incubation with $MgCl_2$, ATP, and phenylacetic acid; absence of any of the substrates negated crystal formation. These observations require that PaaK2 adopt conformation 1 in order to catalyze the adenylation of phenylacetate before adopting conformation 2 for crystal formation.

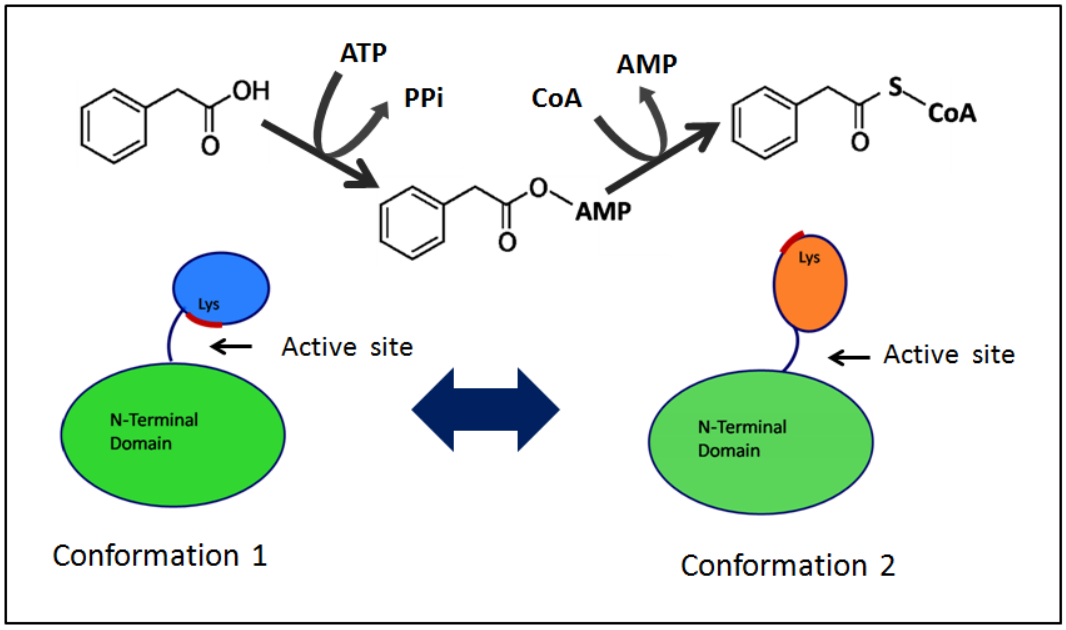


Figure 25: Reaction Mechanism used by PaaK1 & PaaK2
The adenylation of the carboxylate substrate (phenylacetic acid in the case of PaaK1 and PaaK2) is catalyzed while the enzyme is in Conformation 1. The enzyme must release pyrophosphate and undergo a ~140˚ rotation of the C-terminal domain relative to the N-terminal domain (domain alternation) to remodel the active site for CoA binding and the thioesterification reaction catalyzed in Conformation 2.

Chapter 5: Overview of Structural and Functional Findings for PaaK1 and PaaK2

The 1.6 Å crystal structure of PaaK1 represents the first structural characterization of a bacterial PCL. The structure revealed that while PCLs conform to the overall canonical superfamily fold, which consists of a larger N-terminal domain and smaller C-terminal domain joined by a flexible linker, the region of the N-terminal domain containing the initial ~80 residues is significantly different in both sequence and structure from other structurally characterized superfamily members. This novel microdomain arrangement was also observed in the phenylacetyl adenylate PaaK1 and PaaK2 co-structures, suggesting that it is not a crystallization artefact. The biological role and significance of this re-arrangement, however, remains to be determined.

Despite the lack of identity over the first ~80 residues with other superfamily members, PaaK1 and PaaK2 retain the majority of the conserved superfamily motifs within the remainder of the peptide. As one of few ATP co-structures, the 1.6 Å PaaK1 structure revealed initial roles of many conserved residues in ATP binding. Although the general mechanism of domain alternation is now well understood, the finer details at the levels of individual residues especially in the pre-adenylation/ATP bound state, is still somewhat lacking. In the PaaK1-ATP co-structure, the often unmodelled P-loop could be observed enveloping the β and γ phosphates of the co-factor. Most interestingly, the invariant glutamate and arginine, generally only observed in AMP and adenylated intermediate co-structures, now appear to play an initial role stabilizing the β and γ phosphates, perhaps even facilitating their eventual departure.

5.1 Catalytic Roles for Residues in Adenylation Reaction

Based on our observations for the PaaK1-ATP structure, as well as the other two available ATP co-structures, we proposed that it is likely that the invariant C-terminal lysine exerts an electron withdrawing effect at the β phosphates, preparing the pyrophosphate leaving group for the impending nucleophilic attack. However, mutation of the β phosphate binding Arg→Gln also had detrimental effects on catalysis suggesting the positive charge of the arginine side chain also plays an important role. Furthermore, our observations of the near perfect octahedral coordination geometry about the Mg^{2+} cation in the PaaK1 structure may explain why Ca^{2+} is unable to substitute for the catalytically essential Mg^{2+} and why mutation of the invariant Glu→Gln is much less active (Osman *et al.* 2009).

Comparisons with both PaaK1 and PaaK2 captured in complex with phenylacetyl adenylate show conserved interactions similar to what has been observed with homologous family members captured with the adenylated intermediate. These three separate snapshots in time of the reaction process reveal reorganization of the invariant Arg residue following the adenylation reaction, and again following domain alternation. As the domain alternation hypothesis suggests, the C-terminal lysine is removed from the active site by a rotation at the interdomain linker. Interestingly, the active site Glu remains essentially stationary in all three structures and stages of the reaction cycle, stabilized by a salt bridge with a nearby Lys residue. This series of structural studies demonstrates that the observations of conserved active site residues and their respective

ATP interactions are likely typical of all members of the adenylate forming enzyme superfamily, and are not unique to PaaK1. The accumulation of additional ATP bound structures should expand the body of knowledge regarding the multiple roles of individual residues of the active site.

The existence of a *pre-adenylation* state in addition to more generalized 'conformation 1' and 'conformation 2' remains an intriguing possibility. Certainly the active site residues used for substrate binding could differ between a *pre* and *post* adenylation state, as demonstrated by the PaaK1-ATP and PaaK1-adenylate structures. The question remains whether the N and C-terminal domains will pause at a third rotational position relative to one another, and whether this position will be enforced by interactions between the P-loop and C-terminal domain. Comparisons between the *pre* and *post* adenylation structures of PaaK1 show no differences in terms of relative orientation of the two domains. In fact, the two structures overlay in entirety, with an r.m.s.d. of 0.676 Å (381 C-alphas), with no distinguishable difference in C-terminal orientations. As previously discussed, there are no observable interactions between the C-terminal domain and P-loop of PaaK1-ATP structure, while in the adenylate co-structure the P-loop was disordered and unmodelled. In light of these observations, it seems unlikely that there is indeed a third distinct conformation characteristic of a pre-adenylation state, but instead the enzyme undergoes a fluid transition between conformation 1 and 2.

5.2 P-loop Serves and Additional Role stabilizing Conformation 2 in PaaK2

The fact that PaaK2 was captured in conformation 2 provided a complete structural model for the full reaction cycle of bacterial PCLs. Interestingly, the P-loop displayed electron density in both the PaaK2-phenylacetyl adenylate co-structure and PaaK1-ATP co-structure, but could not be built in entirely for the PaaK1-phenylacetyl adenylate co-structure. Most notably, the P-loop ($S^{97}SGTTGKP^{104}$) of PaaK2 adopts a very different conformation than in the PaaK1 ATP structure, interacting with residues from the C-terminal domain (Figure 26). Ser^{98} is seen to hydrogen bond with side chain amine group of Gln^{346} of the C-terminal domain. Further along, the backbone carbonyl of Thr^{100} interacts with the backbone amine of Gln^{346}. Finally, the carboxylate group of the Glu^{349} side chain is within hydrogen bond distance to the backbone amine of Gly^{102}. These interactions are likely mutually stabilizing to both the P-loop as well as the C-terminal domain, stabilizing PaaK2 in conformation 2 for crystal packing. Notably, the P-loop is generally modelled in structures from this superfamily captured in conformation 2 (1PG4, 3CW9, 1V26), strongly suggesting that it may plays an additional role stabilizing the C-terminal domain.

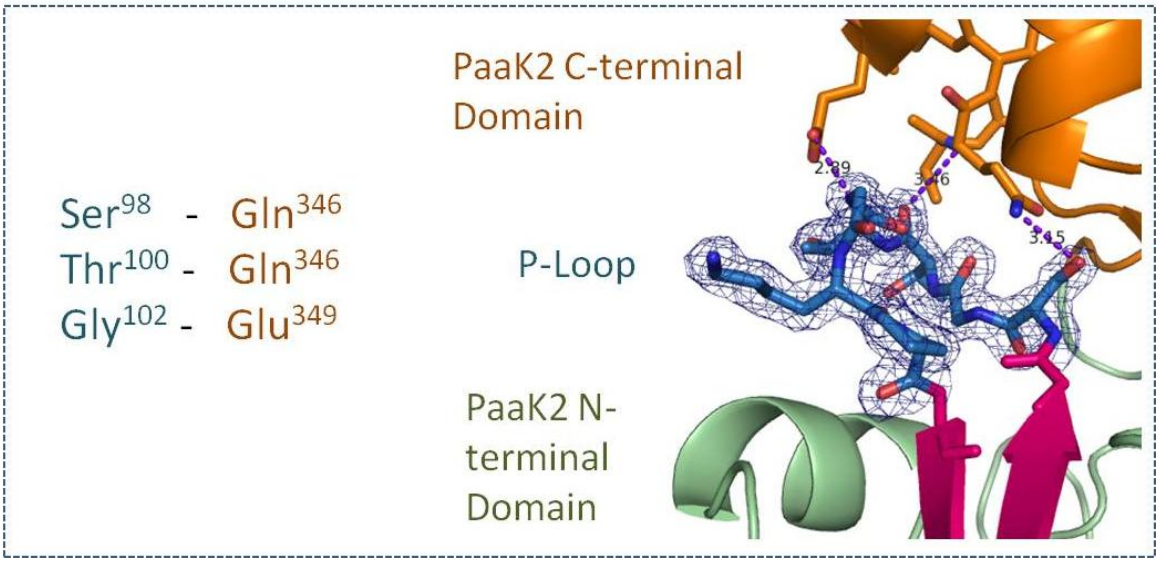


Figure 26: Interface of N and C-terminal Domains of PaaK2
The P-loop ($S^{97}SGTTGKP^{104}$) is well ordered, with blue mesh showing $2F_o$-F_c electron density contoured to 1.0σ. Residues of the P-loop form three hydrogen bonds with residues of C-terminal domain (purple dashes). This figure was produced using PyMOL (DeLano 2002).

5.4 A Functional Rational for PaaK Paralogs

Previously, large scale signature tagged mutagenesis studies identified *paaE* of the PAA pathway as being required for survival *in vivo* of *B. cenocepacia* (Hunt *et al.* 2004). Additional studies using targeted-insertional mutagenesis demonstrated that *paaA* and *paaE* insertional mutants showed attenuated virulence in the *C. elegans* model (Law *et al.* 2008). While the aforementioned insertional mutants showed attenuated growth on phenylacetic acid, insertional mutation to *paaK1* did not impact the rate of growth on phenylacetic acid (Law *et al.* 2008). Despite the fact that the paralogs have now been demonstrated to possess drastically different affinities to phenylacetic acid, such that PaaK2 has a K_M more than double that of PaaK1, it appears that a single functional copy of PaaK is sufficient for growth on phenylacetic acid.

The question remains as to why this less active copy of PaaK has been retained. The *paaK2* gene is adjacent to *paaH* and the two are co-transcribed under the control of the *paaH* operator (Hamlin *et al.* 2009). It has been hypothesized, however, that the three clusters of PAA genes, each with their own promoter, enabling separate regulation of the clusters with different levels of transcription (Hamlin *et al.* 2009). This hypothesis was based on based on reporter assays under the control of the three separate PAA promoter sequences (Hamlin *et al.* 2009). The group found that use of the *paaA* promoter region resulted in approximately 3-fold stronger expression than the *paaH* or *paaZ* promoters (Hamlin *et al.* 2009).

Additionally, the group found that the PAA pathway was subject to catabolic repression when *B. cenocepacia* is grown with alternate carbon sources such as glucose or succinate (Hamlin *et al.* 2009). It is interesting that succinyl-CoA, the CoA thioester of succinate, a final product of the PAA pathway, appears to act as a negative feedback regulator. This observation suggests a more complex regulation of the PAA pathway. Potentially, separate regulation of the two PaaK paralogs may allow *B. cenocepacia* to tailor its usage of phenylacetic acid in the presence of additional carbon sources by using a more (PaaK1) or less (PaaK2) active enzyme when needed.

An additional intriguing possibility for the purpose of PaaK paralogs in *B. cenocepacia* would be for potential differences in CoA utilization. Closer examination of sequence alignments between the proteins (69% identity overall), show that the C-terminal region are much more divergent than the N-terminal domain (Figure 27), which shows 78%

identity and 88% positives over 320 residues (not shown). This raises the possibility that the paralogs could differ in CoA binding ability, resulting in different catalysis rates depending on cellular CoA levels. By comparing the structure of the C-terminal domain of PaaK2 with that of other superfamily members with bound CoA, it is possible to identify residues that would likely be involved in CoA binding. Two separate regions on the C-terminal domain are predicted to be in position to bind CoA; however the first region is entirely conserved between PaaK1 and PaaK2. The second region, on the other hand, shows five residue substitutions between PaaK1 and PaaK2. While it is not possible at this stage to concretely identify any residues used for CoA binding, these differences may contribute to CoA binding and orientation of the C-terminal domain, perhaps effecting the overall rate of reaction and specific activity of the paralogs.

```
C-terminal Sequence Alignment
Identities = 48/108 (44%), Positives = 70/108 (64%), Gaps = 3/108 (2%)

PaaK1  327  SDDMMIVRGVNVFPTQIEEQLLKQRALAPHYQIVLTKEGPLDVLTLNVEPCPETAPDTAA  386
            SDDM+IVRGVNVFP+QIEE ++    L+  +QI L+++G +D L L VE   E A
PaaK2  331  SDDMLIVRGVNVFPSQIEEIVVALPLLSGQFQITLSRDGHMDRLDLAVELRSEAAASVTD  390

PaaK1  387  IQVAKQA--LAYDIKSLIGVTAVINVLPVNGI-ERSVGKARRVVDKRK  432
             + A  A  L + IK+++GV++ + VL   GI   + GKARRV+D+R+
PaaK2  391  GERAALARELQHRIKTMVGVSSGVTVLAAGGIPATATGKARRVIDRRQ  440
```

Figure 27: Sequence Alignment for the C-terminal Domains of PaaK1 and PaaK2.
Highlighted in yellow: conserved lysine. Highlighted in green: putative CoA binding Residues identified through structural overlays of PaaK2 with CoA bound family members also in conformation 2 (PDB ID 3CW9, 2P2F).

5.5 Future Directions

It is intriguing that PaaK1 and PaaK2 exhibit a novel arrangement of the initial 80 residues that has so far not been observed in other superfamily members. It is worth considering that the biochemical consequences of the rearrangement of the N-terminal domain may contribute the instability of PaaK1, PaaK2, and other previously purified bacterial PCLs. One experimental approach to resolve this consideration would be the construction of a PaaK chimera. Substituting the initial 86 residues (up to Ile^{87}) of PaaK1 with the first 184 residues (up to and including Asp^{184}) of *B. xenovorans* Benzoate CoA ligase, which shows the highest similarity at the structural level over the core N-terminal domain, may enhance overall stability. Additionally, the biological role of the rearranged N-terminal domain has also not been established. One intriguing possibility is that the initial 80 residues may be involved in recruiting additional enzymes involved in the PAA pathway. Following the production of phenylacetyl-CoA, the multi-enzyme complex PaaABCDE is putatively responsible for the ring epoxidation (Ferrandez *et al.* 1998; Ismail *et al.* 2003). It is possible that this multi-component enzyme is also able to interact with the initial enzyme. The possibility of PaaK1 and PaaK2 enzymes interacting with additional PAA pathway enzymes, or perhaps even other cellular protein, remains to be investigated. Pull down assays against PaaK1 or PaaK2 could perhaps identify such protein-protein interactions.

In terms of better understanding the role of PaaK1 and PaaK2 within *B. cenocepacia*, it may be worth exploring whether differential regulation may be part of the reason for both gene copies being maintained. As previously mentioned, the three gene clusters each

posses their own operator region with slight differences in the nucleotide sequence. Transcriptomic studies using minimal media supplemented with phenylacetic acid alone or in addition to other carbon sources may reveal different transcription patterns for the two paralogs. Similar studies using synthetic cystic fibrosis medium could ideally reveal expression patterns approximating those of lung infections.

Another possible reason for the presence of both paralogs is the possibility that PaaK1 and PaaK2 differ in their affinity or efficiency for utilizing CoA. While the PaaK2 structure in conformation 2 allowed for identification of residues likely to be in position to interact with CoA, it is impossible at this stage to positively identify whether these residues participate in CoA binding and what roles they may play. A kinetic comparison of the two paralogs should be undertaken with respect to the affinity for CoA, while an additional crystal structure with bound CoA would allow for clear identification of residues involved. As of yet, no crystal screens with either PaaK1 or PaaK2 in presence of CoA have yielded crystal hits. However, as PaaK2 is already in conformation 2, poised to bind CoA and catalyze the thioesterification reaction, soaking or seeding experiments may eventually yield diffraction quality crystals.

While the work presented here has revealed intriguing structural features of bacterial PCLs, specifically a very novel region of the N-terminal domain, the conserved core features of PaaK1 in complex with ATP has contributed to our knowledge of the adenylate forming superfamily at large. The functional characterization of PaaK1 and PaaK2 has shown that there may indeed be a functional rationale for the possession of

both paralogs by *B. cenocepacia*; how this relationship may work in practice within the bacteria remains to be uncovered. Ideally, completion of some of these additional studies will eventually lead to a better understanding of the involvement of the PAA pathway in *B. cenocepacia* metabolism and pathogenesis.

Bibliography

Airth, R. L., Rhodes, W. C. & Mc, E. W., *The function of coenzyme A in luminescence.* Biochimica et Biophysica Acta, 1958. 27(3): p. 519-532.

Alonso, S., Bartolome-Martin, D., del Alamo, M., Diaz, E., Garcia, J. L. & Perera, J., *Genetic characterization of the styrene lower catabolic pathway of Pseudomonas sp. strain Y2.* Gene, 2003. 319: p. 71-83.

Altenschmidt, U. & Fuchs, G., *Anaerobic degradation of toluene in denitrifying Pseudomonas sp.: indication for toluene methylhydroxylation and benzoyl-CoA as central aromatic intermediate.* Archives of Microbiology, 1991. 156(2): p. 152-158.

Babbitt, P. C., Kenyon, G. L., Martin, B. M., Charest, H., Slyvestre, M., Scholten, J. D., Chang, K. H., Liang, P. H. & Dunaway-Mariano, D., *Ancestry of the 4-chlorobenzoate dehalogenase: analysis of amino acid sequence identities among families of acyl:adenyl ligases, enoyl-CoA hydratases/isomerases, and acyl-CoA thioesterases.* Biochemistry, 1992. 31(24): p. 5594-5604.

Bains, J. & Boulanger, M. J., *Biochemical and structural characterization of the paralogous benzoate CoA ligases from Burkholderia xenovorans LB400: defining the entry point into the novel benzoate oxidation (box) pathway.* Journal of Molecular Biology, 2007. 373(4): p. 965-977.

Baldwin, A., Mahenthiralingam, E., Drevinek, P., Vandamme, P., R., G. J., J., W. D., J., L. J., Chiarini, L., Dalmastri, C., Henry, D. A., Speert, D. P., Honeybourne, D., Maiden, M. C. & Dowson, C. G., *Environmental Burkholderia cepacia complex isolates in human infections.* Emerging Infectious Diseases, 2007. 13(3): p. 458-461.

Cao, B., Nagarajan, K. & Loh, K.-C., *Biodegradation of aromatic compounds: current status and opportunities for biomolecular approaches.* Applied Microbiology and Biotechnology, 2009. 85(2): p. 207-228.

The CCP4 suite: programs for protein crystallography. Acta Crystallographica Section D, 1994. 50(Pt 5): p. 760-763.

Chang, K. H., Xiang, H. & Dunaway-Mariano, D., *Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: a site-directed mutagenesis study with the Pseudomonas sp. strain CBS3 4-chlorobenzoate:coenzyme A ligase.* Biochemistry, 1997. 36(50): p. 15650-15659.

Coenye, T., Vandamme, P., Govan, J. R. & LiPuma, J. J., *Taxonomy and identification of the Burkholderia cepacia complex.* Journal of Clinical Microbiology, 2001. 39(10): p. 3427-3436.

Colberg, P. J. & Young, L. Y., *Biodegradation of lignin-derived molecules under anaerobic conditions.* Canadian Journal of Microbiology, 1982. 28(7): p. 886-889.

Copley, S. D. & Crooks, G. P., *Enzymic Dehalogenation of 4-Chlorobenzoyl Coenzyme A in Acinetobacter sp. Strain 4-CB1.* Applied and Environmental Microbiology, 1992. 58(4): p. 1385-1387.

Corkery, D. M., O'Connor, K. E., Buckley, C. M. & Dobson, A. D., *Ethylbenzene degradation by Pseudomonas fluorescens strain CA-4.* FEMS Microbiology Letters, 1994. 124(1): p. 23-27.

Cowtan, K., *The Buccaneer software for automated model building. 1. Tracing protein chains.* Acta Crystallographica Section D, 2006. 62(Pt 9): p. 1002-1011.

Dangel, W., Brackmann, R., Lack, A., Mohamed, M., Koch, J., Oswald, B., Seyfried, B., Tschech, A. & Fuchs, G., *Differential expression of enzyme activities initiating anoxic metabolism of various aromatic compounds via benzoyl-CoA.* Archives of Microbiology, 1991. 155: p. 256-262.

del Peso-Santos, T., Bartolome-Martin, D., Fernandez, C., Alonso, S., Garcia, J. L., Diaz, E., Shingler, V. & Perera, J., *Coregulation by phenylacetyl-coenzyme A-responsive PaaX integrates control of the upper and lower pathways for catabolism of styrene by Pseudomonas sp. strain Y2.* Journal of Bacteriology, 2006. 188(13): p. 4812-4821.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.*

Denef, V. J., Park, J., Tsoi, T. V., Rouillard, J. M., Zhang, H., Wibbenmeyer, J. A., Verstraete, W., Gulari, E., Hashsham, S. A. & Tiedje, J. M., *Biphenyl and benzoate metabolism in a genomic context: outlining genome-wide metabolic networks in Burkholderia xenovorans LB400.* Applied and Environmental Microbiology, 2004. 70(8): p. 4961-4970.

Denef, V. J., Patrauchan, M. A., Florizone, C., Park, J., Tsoi, T. V., Verstraete, W., Tiedje, J. M. & Eltis, L. D., *Growth substrate- and phase-specific expression of biphenyl, benzoate, and C1 metabolic pathways in Burkholderia xenovorans LB400.* Journal of Bacteriology, 2005. 187(23): p. 7996-8005.

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

Elder, D. J. & Kelly, D. J., *The bacterial degradation of benzoic acid and benzenoid compounds under anaerobic conditions: unifying trends and new perspectives.* FEMS Microbiology Reviews, 1994. 13(4): p. 441-468.

Emsley, P. & Cowtan, K., *Coot: model-building tools for molecular graphics.* Acta Crystallographica Section D, 2004. 60(Pt 12 Pt 1): p. 2126-2132.

Erb, T. J., Ismail, W. & Fuchs, G., *Phenylacetate metabolism in thermophiles: characterization of phenylacetate-CoA ligase, the initial enzyme of the hybrid pathway in Thermus thermophilus.* Current Microbiology, 2008. 57(1): p. 27-32.

Evans, P., *Scaling and assessment of data quality.* Acta Crystallographica Section D, 2006. 62(Pt 1): p. 72-82.

Ferrandez, A., Minambres, B., Garcia, B., Olivera, E. R., Luengo, J. M., Garcia, J. L. & Diaz, E., *Catabolism of phenylacetic acid in Escherichia coli. Characterization of a new aerobic hybrid pathway.* Journal of Biological Chemistry, 1998. 273(40): p. 25974-25986.

Fraga, H., Fernandes, D., Fontes, R. & Esteves da Silva, J. C., *Coenzyme A affects firefly luciferase luminescence because it acts as a substrate and not as an allosteric effector.* FEBS Journal, 2005. 272(20): p. 5206-5216.

Gibson, J. & Harwood, C. S., *Metabolic diversity in aromatic compound utilization by anaerobic microbes.* Annual Review of Microbiology, 2002. 56(Journal Article): p. 345-369.

Gulick, A. M., *Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase.* ACS Chemical Biology, 2009. 4(10): p. 811-827.

Gulick, A. M., Lu, X. & Dunaway-Mariano, D., *Crystal structure of 4-chlorobenzoate:CoA ligase/synthetase in the unliganded and aryl substrate-bound states.* Biochemistry, 2004. 43(27): p. 8670-8679.

Gulick, A. M., Starai, V. J., Horswill, A. R., Homick, K. M. & Escalante-Semerena, J. C., *The 1.75 A crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A.* Biochemistry, 2003. 42(10): p. 2866-2873.

Hamlin, J. N., Bloodworth, R. A. & Cardona, S. T., *Regulation of phenylacetic acid degradation genes of Burkholderia cenocepacia K56-2.* BMC Microbiology, 2009. 9: p. 222.

Hartmans, S., van der Werf, M. J. & de Bont, J. A., *Bacterial degradation of styrene involving a novel flavin adenine dinucleotide-dependent styrene monooxygenase.* Applied and Environmental Microbiology, 1990. 56(5): p. 1347-1351.

Harwood, C. S. & Gibson, J., *Uptake of benzoate by Rhodopseudomonas palustris grown anaerobically in light.* Journal of Bacteriology, 1986. 165(2): p. 504-509.

Harwood, C. S. & Parales, R. E., *The beta-ketoadipate pathway and the biology of self-identity.* Annual Review of Microbiology, 1996. 50: p. 553-590.

Hisanaga, Y., Ago, H., Nakagawa, N., Hamada, K., Ida, K., Yamamoto, M., Hori, T., Arii, Y., Sugahara, M., Kuramitsu, S., Yokoyama, S. & Miyano, M., *Structural basis of the substrate-specific two-step catalysis of long chain fatty acyl-CoA synthetase dimer.* Journal of Biological Chemistry, 2004. 279(30): p. 31717-31726.

Holden, M. T., Seth-Smith, H. M., Crossman, L. C., Sebaihia, M., Bentley, S. D., Cerdeno-Tarraga, A. M., Thomson, N. R., Bason, N., Quail, M. A., Sharp, S., Cherevach, I., Churcher, C., Goodhead, I., Hauser, H., Holroyd, N., Mungall, K., Scott, P., Walker, D., White, B., Rose, H., Iversen, P., Mil-Homens, D., Rocha, E. P., Fialho, A. M., Baldwin, A., Dowson, C., Barrell, B. G., Govan, J. R., Vandamme, P., Hart, C. A., Mahenthiralingam, E. & Parkhill, J., *The genome of Burkholderia cenocepacia J2315, an epidemic pathogen of cystic fibrosis patients.* Journal of Bacteriology, 2009. 191(1): p. 261-277.

Holm, L. & Rosenstrom, P., *Dali server: conservation mapping in 3D.* Nucleic Acids Research, 2010. 38 Suppl: p. W545-549.

Hunt, T. A., Kooi, C., Sokol, P. A. & Valvano, M. A., *Identification of Burkholderia cenocepacia genes required for bacterial survival in vivo.* Infection and Immunity, 2004. 72(7): p. 4010-4022.

Ismail, W., El-Said Mohamed, M., Wanner, B. L., Datsenko, K. A., Eisenreich, W., Rohdich, F., Bacher, A. & Fuchs, G., *Functional genomics by NMR spectroscopy. Phenylacetate catabolism in Escherichia coli.* European Journal of Biochemistry / FEBS, 2003. 270(14): p. 3047-3054.

Kichise, T., Hisano, T., Takeda, K. & Miki, K., *Crystal structure of phenylacetic acid degradation protein PaaG from Thermus thermophilus HB8.* Proteins, 2009. 76(4): p. 779-786.

Krissinel, E., *Crystal contacts as nature's docking solutions.* Journal of Computational Chemistry, 2010. 31(1): p. 133-143.

Kunishima, N., Asada, Y., Sugahara, M., Ishijima, J., Nodake, Y., Miyano, M., Kuramitsu, S. & Yokoyama, S., *A novel induced-fit reaction mechanism of asymmetric hot dog thioesterase PAAI.* Journal of Molecular Biology, 2005. 352(1): p. 212-228.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thorton, J. M., *PROCHECK: a program to check the sterochemical quality of protein structures.* Journal of Applied Crystallography, 1993. 26(2): p. 283-291.

Law, R. J., Hamlin, J. N., Sivro, A., McCorrister, S. J., Cardama, G. A. & Cardona, S. T., *A functional phenylacetic acid catabolic pathway is required for full pathogenicity of Burkholderia cenocepacia in the Caenorhabditis elegans host model.* Journal of Bacteriology, 2008. 190(21): p. 7209-7218.

Leslie, A. G. W., *Recent changes to the MOSFLM package for processing film and image plate data* Joint CCP4 + ESF-EAMCB Newsletters on Protein Crystallography, 1992. 26.

Long, M. T., Bartholomew, B. A., Smith, M. J., Trudgill, P. W. & Hopper, D. J., *Enzymology of oxidation of tropic acid to phenylacetic acid in metabolism of atropine by Pseudomonas sp. strain AT3.* Journal of Bacteriology, 1997. 179(4): p. 1044-1050.

Luengo, J. M., Garcia, J. L. & Olivera, E. R., *The phenylacetyl-CoA catabolon: a complex catabolic unit with broad biotechnological applications.* Molecular Microbiology, 2001. 39(6): p. 1434-1442.

Mahenthiralingam, E., Urban, T. A. & Goldberg, J. B., *The multifarious, multireplicon Burkholderia cepacia complex.* Nature Reviews Microbiology, 2005. 3(2): p. 144-156.

Mahenthiralingam, E., Vandamme, P., Campbell, M. E., Henry, D. A., Gravelle, A. M., Wong, L. T., Davidson, A. G., Wilcox, P. G., Nakielna, B. & Speert, D. P., *Infection with Burkholderia cepacia complex genomovars in patients with cystic fibrosis: virulent transmissible strains of genomovar III can replace Burkholderia multivorans.* Clinical Infectious Diseases, 2001. 33(9): p. 1469-1475.

Marahiel, M. A., Stachelhaus, T. & Mootz, H. D., *Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis.* Chemical reviews, 1997. 97(7): p. 2651-2674.

Martinez-Blanco, H., Reglero, A., Rodriguez-Aparicio, L. B. & Luengo, J. M., *Purification and biochemical characterization of phenylacetyl-CoA ligase from Pseudomonas putida. A specific enzyme for the catabolism of phenylacetic acid.* The Journal of Biological Chemistry, 1990. 265(12): p. 7084-7090.

McElroy, W. D., DeLuca, M. & Travis, J., *Molecular uniformity in biological catalyses. The enzymes concerned with firefly luciferin, amino acid, and fatty acid utilization are compared.* Science, 1967. 157(785): p. 150-160.

Mohamed, M.-S., Ismail, W., Heider, J. & Fuchs, G., *Aerobic metabolism of phenylacetic acids in Azoarcus evansii.* Archives of Microbiology, 2002. 178(3): p. 180-192.

Mohamed, M. e.-S. & Fuchs, G., *Purification and characterization of phenylacetate-coenzyme A ligase from a denitrifying Pseudomonas sp., an enzyme involved in the anaerobic degradation of phenylacetate.* Archives of Microbiology, 1993. 159(6): p. 554-562.

Mohamed, M. E., Zaar, A., Ebenau-Jehle, C. & Fuchs, G., *Reinvestigation of a new type of aerobic benzoate metabolism in the proteobacterium Azoarcus evansii.* Journal of Bacteriology, 2001. 183(6): p. 1899-1908.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J., *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallographica Section D, 1997. 53(Pt 3): p. 240-255.

Navarro-Llorens, J. M., Drzyzga, O. & Perera, J., *Genetic analysis of phenylacetic acid catabolism in Arthrobacter oxydans CECT386.* Archives of Microbiology, 2008. 190(1): p. 89-100.

Navarro-Llorens, J. M., Patrauchan, M. A., Stewart, G. R., Davies, J. E., Eltis, L. D. & Mohn, W. W., *Phenylacetate catabolism in Rhodococcus sp. strain RHA1: a central pathway for degradation of aromatic compounds.* Journal of Bacteriology, 2005. 187(13): p. 4497-4504.

Nogales, J., Macchi, R., Franchi, F., Barzaghi, D., Fernandez, C., Garcia, J. L., Bertoni, G. & Diaz, E., *Characterization of the last step of the aerobic phenylacetic acid degradation pathway.* Microbiology, 2007. 153(Pt 2): p. 357-365.

Oba, Y., Ojika, M. & Inouye, S., *Firefly luciferase is a bifunctional enzyme: ATP-dependent monooxygenase and a long chain fatty acyl-CoA synthetase.* FEBS Letters, 2003. 540(1-3): p. 251-254.

Olivera, E. R., Minambres, B., Garcia, B., Muniz, C., Moreno, M. A., Ferrandez, A., Diaz, E., Garcia, J. L. & Luengo, J. M., *Molecular characterization of the phenylacetic acid catabolic pathway in Pseudomonas putida U: the phenylacetyl-CoA catabolon.* Proceedings of the National Academy of Sciences of the United States of America, 1998. 95(11): p. 6419-6424.

Olivera, E. R., Reglero, A., Martinez-Blanco, H., Fernandez-Medarde, A., Moreno, M. A. & Luengo, J. M., *Catabolism of aromatics in Pseudomonas putida U. Formal evidence that phenylacetic acid and 4-hydroxyphenylacetic acid are catabolized by two unrelated pathways.* European Journal of Biochemistry / FEBS, 1994. 221(1): p. 375-381.

Orth, A. B., Denny, M. & Tien, M., *Overproduction of lignin-degrading enzymes by an isolate of Phanerochaete chrysosporium.* Applied and Environmental Microbiology, 1991. 57(9): p. 2591-2596.

Osman, K. T., Du, L., He, Y. & Luo, Y., *Crystal structure of Bacillus cereus D-alanyl carrier protein ligase (DltA) in complex with ATP.* Journal of Molecular Biology, 2009. 388(2): p. 345-355.

Parke, J. L. & Gurian-Sherman, D., *Diversity of the Burkholderia cepacia complex and implications for risk assessment of biological control strains.* Annual Review of Phytopathology, 2001. 39(Generic): p. 225-258.

Parrott, S., Jones, S. & Cooper, R. A., *2-Phenylethylamine catabolism by Escherichia coli K12.* Journal of General Microbiology, 1987. 133(2): p. 347-351.

Reger, A. S., Wu, R., Dunaway-Mariano, D. & Gulick, A. M., *Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase.* Biochemistry, 2008. 47(31): p. 8016-8025.

Schleyer, P. R., *Introduction: aromaticity.* Chemical Reviews, 2001. 101(5): p. 1115-1118.

Schuttelkopf, A. W. & van Aalten, D. M., *PRODRG: a tool for high-throughput crystallography of protein-ligand complexes.* Acta Crystallographica Section D, 2004. 60(Pt 8): p. 1355-1363.

Shah, M. B., Ingram-Smith, C., Cooper, L. L., Qu, J. F., Meng, Y., S., S. K. & Gulick, A. M., *The 2.1 A crystal structure of an acyl-CoA synthetase from Methanosarcina acetivorans reveals an alternate acyl-binding pocket for small branched acyl substrates.* Proteins, 2009. 77(3): p. 685-698.

Sheldrick, G. M., *A short history of SHELX.* Acta Crystallographica Section A, 2008. 64(Pt 1): p. 112-122.

Teufel, R., Mascaraqueb, V., Ismail, W., Voss, M., Perera J., Eisenreich, W., Haehnel, W., & Fuchs, G., *Bacterial phenylalanine and phenylacetate catabolic pathway revealed.* Proceedings of the National Academy of Sciences of the United States of America, 2010. 107(32): p. 14390–14395.

Vagin A, Teplyakov A. 1997. MOLREP: an automated program for molecular replacement. Journal of Applied Crystallography 30:1022-1025.

Vaguine, A. A., Richelle, J. & Wodak, S. J., *SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model.* Acta Crystallographica Section D, 1999. 55(Pt 1): p. 191-205.

Velasco, A., Alonso, S., Garcia, J. L., Perera, J. & Diaz, E., *Genetic and functional analysis of the styrene catabolic cluster of Pseudomonas sp. strain Y2.* J Bacteriol, 1998. 180(5): p. 1063-1071.

Vermis, K., Vandamme, P. A. & Nelis, H. J., *Burkholderia cepacia complex genomovars: utilization of carbon sources, susceptibility to antimicrobial agents and growth on selective media.* Journal of Applied Microbiology, 2003. 95(6): p. 1191-1199.

Wu, R., Cao, J., Lu, X., Reger, A. S., Gulick, A. M. & Dunaway-Mariano, D., *Mechanism of 4-chlorobenzoate:coenzyme a ligase catalysis.* Biochemistry, 2008. 47(31): p. 8026-8039.

Wu, R., Reger, A. S., Lu, X., Gulick, A. M. & Dunaway-Mariano, D., *The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-chlorobenzoate: coenzyme A ligase.* Biochemistry, 2009. 48(19): p. 4115-4125.

Ziegler, K., Buder, R., Winter, J. & Fuchs, G., *Activation of aromatic acids and aerobic 2-aminobenzoate metabolism in a denitrifying Pseudomonas strain.* Archives of Microbiology, 1989. 151(2): p. 171-176.

Appendix A: Sequence Alignment of PaaK1 and Homologous Family Members

Sequence alignment of PaaK1 and homologous family members matched from DALI search colour coded by secondary structure. Red=Helix, Green=Loop, Yellow=strand. Conserved Motifs and residues are boxed. While other family members show alternating helices and strands, PaaK1 aligns extremely poorly for the first ~90 residues and large deletions reflecting an overall absence of β strands until Ser^{88}. Conserved motifs are boxed.